Universidad Nacional de Misiones

Facultad de Ciencias Exactas, Químicas y Naturales

Trabajo Final de Maestría en Tecnologías de la Información

Modelo basado en la Toma Decisiones con Criterios Múltiples para la elección de Metodologías de Data Science

Autora: Ing. Karina Beatriz Eckert

Directora: Dra. Paola Verónica Britos

Año 2019

Dedico este trabajo final de maestría a mis padres y a todas las personas queridas que me han brindado su apoyo y motivación durante este camino.

Resumen

La toma de decisiones estratégica dentro de las organizaciones, se encuentra regida por el correcto procesamiento de los datos generados en sus actividades diarias; esto ha llevado a que áreas como la Ciencia de Datos o Data Science estén en su pleno apogeo. Consecuencia de ello, en los últimos años se ha acrecentado significativamente la cantidad profesionales vinculados a dicha área, los cuales, entre otras aptitudes, deben ser capaces de determinar cuál de las metodologías disponibles se adapta mejor a un proyecto en particular. El objetivo de este trabajo final de maestría es establecer un modelo comparativo de metodologías de Data Science, mediante dos métodos pertenecientes a la Toma de Decisiones con Criterios Múltiples, el Proceso Analítico Jerárquico y el Proceso Analítico Jerárquico Difuso; combinados e integrados con la técnica de la Teoría de la Construcción Personal. El modelo consta de una serie pasos que permiten obtener las ponderaciones finales para cada metodología. Para los dos casos de validación estudiados, la metodología MoProPEI fue la seleccionada. Gracias al desarrollo de esta tesis se pudo comprobar la utilidad y validez del modelo propuesto, así como la integración efectiva de los métodos y la técnica mencionados.

Palabras claves: *Ciencia de Datos, Metodologías, Toma de Decisiones Criterios Múltiples, Proceso Analítico Jerárquico, Proceso Analítico Jerárquico Difuso, Teoría de la Construcción Personal.*

Abstract

Strategic decision making within organizations is mainly governed by the correct processing of data generated in their daily activities; this has led to areas such as Data Science in full swing. As a result, in recent years the number of professionals linked to this area has been increased significantly, which, among other skills, should be determine which of available current methodologies adapts better to a particular project. Final master's degree work objective is to establish a comparative model of Data Science methodologies, using two methods belonging to Multiple Criteria Decision Making, Analytic Hierarchy Process and

Fuzzy Analytic Hierarchy Process; combined and integrated with the technique of Personal Construction Theory. The model consists a steps series that allow obtaining the final weights each methodology. For the two validation cases studied, the MoProPEI methodology was selected. Thanks to thesis development, utility and validity of the proposed model could be verified, as well as effective integration of methods and techniques used.

Keywords: *Data Science, Methodology, Multiple Criteria Decision Making, Analytic Hierarchy Process, Fuzzy Analytic Hierarchy Process, Personal Construction Theory.*

En primer lugar agradezco a mi directora la Dra. Paola V. Britos, del Laboratorio de Informática Aplicada de la Universidad Nacional de Río Negro, por su colaboración y guía en el transcurso de la elaboración de esta tesis; así también por su disposición para validar los casos utilizados.

A la Dra. Florencia Pollo Cattaneo, de la Universidad Tecnológica Nacional Facultad Regional Buenos Aires y al Dr. Hernán Merlino, de la Universidad Nacional de Lanús y la Universidad Tecnológica Nacional Facultad Regional Buenos Aires, por su colaboración como expertos de los casos de validación empleados.

A los docentes de la maestría, de la Facultad de Ciencias Exactas, Químicas y Naturales de la Universidad Nacional de Misiones y de la Facultad de Ciencias Exactas y Naturales y Agrimensura de la Universidad Nacional del Nordeste, por el apoyo en las diferentes instancias del proceso desarrollo de mi carrera de maestría.

A mi familia por su amor, comprensión y motivación constante.

A mis amigos, compañeros, alumnos y a todos aquellos que participaron directa o indirectamente en mi formación.

Y por último, pero no menos importante, agradezco a Dios, por haber guiado mis pasos, darme la fortaleza para poder llegar al final de la meta y por haberme dado la oportunidad de estar rodeada de seres muy especiales.

Índice

I *17*
Introducción *17*
1.1. Objetivos *18*
1.1.1. Objetivo general *18*
1.1.2. Objetivos específicos *18*
1.2. Estructura del documento *18*
II *21*
Estado de la cuestión *21*
2.1. Toma Decisiones con Criterios Múltiples *22*
2.1.1. Proceso Analítico Jerárquico *27*
2.1.2. Proceso Analítico Jerárquico Difuso *33*
2.2. Metodologías de Ciencia de Datos *35*
2.2.1. CRISP-DM *38*
2.2.2. Catalys *38*
2.2.3. MoProPEI *39*
2.3. Educción del Conocimiento *39*
2.3.1. Teoría de la Construcción Personal *40*
III *43*
Problema *43*
3.1. Descripción del problema *44*
IV *47*
Solución *47*
4.1. Solución propuesta *48*
V *63*
Casos de validación *63*
5.1. Generalidades de los casos de validación *64*
5.1.1. Valoración con matrices de dos entradas *65*
5.2. Caso de validación: Rotura de automóviles 0KM durante el transporte a las concesionarias *78*
5.2.1. Valoración con TCP del nivel 4 del CV1 ..... *78*
5.2.2. Completitud de las matrices y establecimiento de prioridades entre criterios del CV1 ..... *90*
5.2.3. Evaluación de consistencia del CV1 ..... *106*
5.2.4. Desarrollo de los métodos de MCDM para el CV1 ..... *115*
5.2.5. Integración de las ponderaciones obtenidas por los métodos AHP y FAHP para el CV1 ..... *159*
5.2.6. Ranking de metodologías de DS para el CV1 ..... *159*
5.3. Caso de validación: Deserción universitaria *160*
5.3.1. Valoración con TCP del nivel 4 del CV2 ..... *160*

5.3.2. Completitud de las matrices y establecimiento de prioridades entre criterios del CV2 ..... *164*
5.3.3. Evaluación de consistencia del CV2 ..... *173*
5.3.4. Desarrollo de los métodos de MCDM para el CV2 ..... *178*
5.3.5. Integración de las ponderaciones obtenidas por los métodos AHP y FAHP para el CV2 ..... *206*
5.3.6. Ranking de metodologías de DS para el CV2 ..... *206*
5.4. Análisis de resultados *207*
VI *209*
Conclusiones y trabajos futuros *209*
6.1. Conclusiones *210*
6.2. Publicaciones *212*
6.3. Líneas futuras de investigación *212*
Bibliografía *215*
Anexos *227*
Anexo I: Proceso Analítico Jerárquico *228*
Anexo II: Proceso Analítico Jerárquico Difuso *236*
Anexo III: Matrices de dos entradas *242*
Anexo IV: Plantilla para definición de los sub-criterios mediante la técnica de la Teoría de la Construcción Personal *254*
Anexo V: Correos de aceptación artículos en prensa *265*

Lista de Tablas

Tabla 4.1: Comparación de criterios por pares *52*
Tabla 4.2: Etiquetas lingüísticas para sub-criterios *53*
Tabla 4.3: Parrilla para valoración de los sub-criterios del cuarto nivel *54*
Tabla 4.4: Matriz de comparación pareada *55*
Tabla 4.5: Priorización del sub-criterio Entorno del Negocio AHP *57*
Tabla 4.6: Prioridades totales de las alternativas con AHP *57*
Tabla 4.7: Matriz triangular difusa de comparación pareada *58*
Tabla 4.8: Cálculo del valor de extensión sintética difusa *58*
Tabla 4.9: Superioridad comparativa *59*
Tabla 4.10: Vector de pesos *59*
Tabla 4.11: Ponderaciones parciales *60*
Tabla 4.12: Prioridades parciales del sub-criterio Entorno del Negocio FAHP *60*
Tabla 4.13: Prioridades finales de las alternativas con FAHP *61*
Tabla 4.14: Prioridades finales de las alternativas *61*
Tabla 5.1: Matriz de dos entradas para criterios principales (nivel 2) *65*
Tabla 5.2: Matriz de dos entradas para sub-criterios Acceso y Uso de los Datos *66*
Tabla 5.3: Matriz de dos entradas para la Portabilidad y demás sub-criterios *66*
Tabla 5.4: Matriz de dos entradas para la Accesibilidad y demás sub-criterios *67*
Tabla 5.5: Matriz de dos entradas para la Diversidad y demás sub-criterios *67*
Tabla 5.6: Matriz de dos entradas para la Diversidad de fuentes de datos y Recursos necesarios *68*
Tabla 5.7: Matriz de dos entradas para la Calidad y demás sub-criterios *68*
Tabla 5.8: Matriz de dos entradas para la Completitud y demás sub-criterios *69*
Tabla 5.9: Matriz de dos entradas para la Funcionalidad y demás sub-criterios *69*
Tabla 5.10: Matriz de dos entradas para el Costo de transformación y Datos riesgosos *70*
Tabla 5.11: Matriz de dos entradas para el criterio Entorno del Negocio y demás criterios *70*
Tabla 5.12: Matriz de dos entradas para los criterios Proyecto Entorno del Negocio y DS en el Proyecto del Negocio *70*
Tabla 5.13: Matriz de dos entradas para la Situación económica circundante y demás sub-criterios *71*
Tabla 5.14: Matriz de dos entradas para la Evaluación/comprensión general de la organización y demás sub-criterios *72*
Tabla 5.15: Matriz de dos entradas para la Definición de los objetivos del negocio y demás sub-criterios *73*
Tabla 5.16: Matriz de dos entradas para la Definición del problema del negocio y demás sub-criterios *73*
Tabla 5.17: Matriz de dos entradas para la Definición de las oportunidades del negocio y demás sub-criterios *74*
Tabla 5.18: Matriz de dos entradas para la Definición de requisitos del negocio y de los criterios de éxito del proyecto *74*
Tabla 5.19: Matriz de dos entradas para Conocimiento de los RRHH del negocio y demás sub-criterios *75*
Tabla 5.20: Matriz de dos entradas para Disponibilidad de los RRHH del negocio y demás sub-criterios *76*
Tabla 5.21: Matriz de dos entradas para Riesgos del problema del negocio y demás sub-criterios *76*

Tabla 5.22: Matriz de dos entradas para Contingencias en relación a los riesgos del problema del negocio y Viabilidad del proyecto 77
Tabla 5.23: Matriz de dos entradas para Objetivos de aplicación de la metodología de DS y demás sub-criterios 77
Tabla 5.24: Matriz de dos entradas para Conocimiento de los RRHH del proyecto de DS y RRHH disponibles del proyecto de DS 78
Tabla 5.25: Parrilla de Acceso a Datos del CV1 80
Tabla 5.26: Parrilla de Uso de los Datos del CV1 82
Tabla 5.27: Parrilla del Entorno del Negocio del CV1 85
Tabla 5.28: Parrilla del Proyecto del Entorno del Negocio del CV1 88
Tabla 5.29: Parrilla de Data Science en el Proyecto del Entorno del Negocio del CV1 90
Tabla 5.30: Matriz pareada del nivel 2 90
Tabla 5.31: Matriz Normalizada y ponderación del nivel 2 91
Tabla 5.32: Matriz pareada del nivel 3 para ED 91
Tabla 5.33: Matriz Normalizada y ponderación del nivel 3 para ED 91
Tabla 5.34: Matriz pareada del nivel 4 para ED-AD 92
Tabla 5.35: Matriz Normalizada y ponderación del nivel 4 para ED-AD 92
Tabla 5.36: Ponderación ED-SC1 en relación a cada alternativa para CV1 92
Tabla 5.37: Ponderación de ED-SC2 en relación a cada alternativa para CV1 93
Tabla 5.38: Ponderación de ED-SC3 en relación a cada alternativa para CV1 93
Tabla 5.39: Ponderación de ED-SC4 en relación a cada alternativa para CV1 93
Tabla 5.40: Ponderación de ED-SC5 en relación a cada alternativa para CV1 94
Tabla 5.41: Matriz pareada del nivel 4 para ED-UD 94
Tabla 5.42: Matriz normalizada y ponderación del nivel 4 para ED-UD 94
Tabla 5.43: Ponderación de ED-SC6 en relación a cada alternativa para CV1 95
Tabla 5.44: Ponderación de ED-SC7 en relación a cada alternativa para CV1 95
Tabla 5.45: Ponderación de ED-SC8 en relación a cada alternativa para CV1 95
Tabla 5.46: Ponderación de ED-SC9 en relación a cada alternativa para CV1 96
Tabla 5.47: Ponderación de ED-SC10 en relación a cada alternativa para CV1 96
Tabla 5.48: Matriz pareada del nivel 3 para CN 97
Tabla 5.49: Matriz normalizada y ponderación del nivel 3 para CN 97
Tabla 5.50: Matriz pareada del nivel 4 para CN-EN 97
Tabla 5.51: Matriz normalizada y ponderación del nivel 4 para CN-EN 98
Tabla 5.52: Ponderación CN-SC1 en relación a cada alternativa para CV1 99
Tabla 5.53: Ponderación CN-SC2 en relación a cada alternativa para CV1 99
Tabla 5.54: Ponderación CN-SC3 en relación a cada alternativa para CV1 100
Tabla 5.55: Ponderación CN-SC4 en relación a cada alternativa para CV1 100
Tabla 5.56: Ponderación CN-SC5 en relación a cada alternativa para CV1 100
Tabla 5.57: Ponderación CN-SC6 en relación a cada alternativa para CV1 101
Tabla 5.58: Ponderación CN-SC7 en relación a cada alternativa para CV1 101
Tabla 5.59: Matriz pareada del nivel 4 para CN-PEN 101
Tabla 5.60: Matriz normalizada y ponderación del nivel 4 para CN-PEN 102
Tabla 5.61: Ponderación CN-SC8 en relación a cada alternativa para CV1 102
Tabla 5.62: Ponderación CN-SC9 en relación a cada alternativa para CV1 103
Tabla 5.63: Ponderación CN-SC10 en relación a cada alternativa para CV1 103
Tabla 5.64: Ponderación CN-SC11 en relación a cada alternativa para CV1 103
Tabla 5.65: Ponderación CN-SC12 en relación a cada alternativa para CV1 104
Tabla 5.66: Matriz pareada del nivel 4 para CN-DSPN 104
Tabla 5.67: Matriz normalizada y ponderación del nivel 4 para CN-DSPN 104

Tabla 5.68: Ponderación de CN-SC13 en relación a cada alternativa para CV1 105
Tabla 5.69: Ponderación de CN-SC14 en relación a cada alternativa para CV1 105
Tabla 5.70: Ponderación de CN-SC15 en relación a cada alternativa para CV1 105
Tabla 5.71: Autovector máximo del ED-AD 106
Tabla 5.72: Evaluación de consistencia del ED-AD 106
Tabla 5.73: Autovector máximo de cada sub-criterio del ED-AD para CV1 107
Tabla 5.74: Evaluación de consistencia de los sub-criterios del ED-AD para CV1 108
Tabla 5.75: Autovector máximo de cada sub-criterio del ED-UD 108
Tabla 5.76: Evaluación de consistencia de los sub-criterios del ED-UD 108
Tabla 5.77: Autovector máximo de cada sub-criterio del ED-UD para CV1 109
Tabla 5.78: Evaluación de consistencia de los sub-criterios del ED-UD para CV1 109
Tabla 5.79: Autovector máximo de cada sub-criterio de la CN 110
Tabla 5.80: Evaluación de consistencia de los sub-criterios de la CN 110
Tabla 5.81: Autovector máximo del CN-EN 110
Tabla 5.82: Evaluación de consistencia CN-EN 111
Tabla 5.83: Autovector máximo de cada sub-criterio del CN-EN para CV1 111
Tabla 5.84: Evaluación de consistencia de los sub-criterios del CN-EN para CV1 112
Tabla 5.85: Autovector máximo del CN-PEN 112
Tabla 5.86: Evaluación de consistencia CN-PEN 113
Tabla 5.87: Autovector máximo de cada sub-criterio del CN-PEN para CV1 113
Tabla 5.88: Evaluación de consistencia de los sub-criterios del CN-PEN para CV1 114
Tabla 5.89: Autovector máximo del CN-DSPN 114
Tabla 5.90: Evaluación de consistencia CN-DSPN 114
Tabla 5.91: Autovector máximo de cada sub-criterio del CN-DSPN para CV1 115
Tabla 5.92: Evaluación de consistencia de los sub-criterios del CN-DSPN para CV1 115
Tabla 5.93: Prioridad local de ED-AD para CV1 116
Tabla 5.94: Prioridad local de ED-UD para CV1 116
Tabla 5.95: Prioridad global de ED-AD para CV1 117
Tabla 5.96: Prioridad global de ED-UD para CV1 117
Tabla 5.97: Prioridad local de CN-EN para CV1 118
Tabla 5.98: Prioridad local de CN-PEN para CV1 118
Tabla 5.99: Prioridad local de CN-DSPN para CV1 119
Tabla 5.100: Prioridad global de CN-EN para CV1 119
Tabla 5.101: Prioridad global de CN-PEN para CV1 119
Tabla 5.102: Prioridad global de CN-DSPN para CV1 120
Tabla 5.103: Prioridades totales con AHP para el CV1 121
Tabla 5.104: Matriz triangular difusa de comparación pareada del nivel 2 122
Tabla 5.105: Matriz triangular difusa de comparación pareada del nivel 3 para ED 122
Tabla 5.106: Matriz triangular difusa de comparación pareada del nivel 4 para ED-AD ... 122
Tabla 5.107: Matriz triangular difusa de comparación pareada de ED-SC1 para CV1 123
Tabla 5.108: Matriz triangular difusa de comparación pareada de ED-SC2 para CV1 123
Tabla 5.109: Matriz triangular difusa de comparación pareada de ED-SC3 para CV1 123
Tabla 5.110: Matriz triangular difusa de comparación pareada de ED-SC4 para CV1 124
Tabla 5.111: Matriz triangular difusa de comparación pareada de ED-SC5 para CV1 124
Tabla 5.112: Matriz triangular difusa de comparación pareada del nivel 4 para ED-UD .. 124
Tabla 5.113: Matriz triangular difusa de comparación pareada de ED-SC6 para CV1 125
Tabla 5.114: Matriz triangular difusa de comparación pareada de ED-SC7 para CV1 125
Tabla 5.115: Matriz triangular difusa de comparación pareada de ED-SC8 para CV1 125
Tabla 5.116: Matriz triangular difusa de comparación pareada de ED-SC9 para CV1 126

Tabla 5.117: Matriz triangular difusa de comparación pareada de ED-SC10 para CV1 126
Tabla 5.118: Matriz triangular difusa de comparación pareada del nivel 3 para CN..... 127
Tabla 5.119: Matriz triangular difusa de comparación pareada del nivel 4 para CN-EN ... 127
Tabla 5.120: Matriz triangular difusa de comparación pareada de CN-SC1 para CV1 128
Tabla 5.121: Matriz triangular difusa de comparación pareada de CN-SC2 para CV1 128
Tabla 5.122: Matriz triangular difusa de comparación pareada de CN-SC3 para CV1 128
Tabla 5.123: Matriz triangular difusa de comparación pareada de CN-SC4 para CV1 128
Tabla 5.124: Matriz triangular difusa de comparación pareada de CN-SC5 para CV1 129
Tabla 5.125: Matriz triangular difusa de comparación pareada de CN-SC6 para CV1 129
Tabla 5.126: Matriz triangular difusa de comparación pareada de CN-SC7 para CV1 129
Tabla 5.127: Matriz triangular difusa de comparación pareada del nivel 4 para CN-PEN. 130
Tabla 5.128: Matriz triangular difusa de comparación pareada a de CN-SC8 para CV1 ... 130
Tabla 5.129: Matriz triangular difusa de comparación pareada de CN-SC9 para CV1 130
Tabla 5.130: Matriz triangular difusa de comparación pareada de CN-SC10 para CV1 131
Tabla 5.131: Matriz triangular difusa de comparación pareada de CN-SC11 para CV1 131
Tabla 5.132: Matriz triangular difusa de comparación pareada de CN-SC12 para CV1 131
Tabla 5.133: Matriz triangular difusa de comparación pareada del nivel 4 para CN-DSPN 132
Tabla 5.134: Matriz triangular difusa de comparación pareada de CN-SC13 para CV1 132
Tabla 5.135: Matriz triangular difusa de comparación pareada de CN-SC14 para CV1 132
Tabla 5.136: Matriz triangular difusa de comparación pareada de CN-SC15 para CV1 132
Tabla 5.137: Extensión sintética difusa del nivel 2 133
Tabla 5.138: Extensión sintética difusa del nivel 3 para ED..... 133
Tabla 5.139: Extensión sintética difusa del nivel 3 para ED-AD..... 134
Tabla 5.140: Extensión sintética difusa de Sub-criterios ED-AD para V1 134
Tabla 5.141: Extensión sintética difusa del nivel 4 para ED-UD 135
Tabla 5.142: Extensión sintética difusa de Sub-criterios ED-UD para CV1 136
Tabla 5.143: Extensión sintética difusa del nivel 3 para CN..... 137
Tabla 5.144: Extensión sintética difusa del nivel 4 para CN-EN..... 137
Tabla 5.145: Extensión sintética difusa de Sub-criterios CN-EN para CV1..... 138
Tabla 5.146: Extensión sintética difusa del nivel 4 ara CN-PEN..... 139
Tabla 5.147: Extensión sintética difusa de Sub-criterios CN-PEN para CV1..... 140
Tabla 5.148: Extensión sintética difusa del nivel 4 para CN-DSPN 140
Tabla 5.149: Extensión sintética difusa de Sub-criterios CN-DSPN para CV1 141
Tabla 5.150: Superioridad comparativa del nivel 2 141
Tabla 5.151: Superioridad comparativa del nivel 3 para ED 142
Tabla 5.152: Superioridad comparativa del nivel 3 para ED-AD..... 142
Tabla 5.153: Superioridad comparativa de Sub-criterios ED-AD para CV1..... 143
Tabla 5.154: Superioridad comparativa del nivel 4 para ED-UD 143
Tabla 5.155: Superioridad comparativa de Sub-criterios ED-UD para CV1 144
Tabla 5.156: Superioridad comparativa del nivel 3 para CN 145
Tabla 5.157: Superioridad comparativa del nivel 4 para CN-EN..... 145
Tabla 5.158: Superioridad comparativa de Sub-criterios CN-EN para CV1..... 146
Tabla 5.159: Superioridad comparativa del nivel 4 para CN-PEN 147
Tabla 5.160: Superioridad comparativa de Sub-criterios CN-PEN para CV1 148
Tabla 5.161: Superioridad comparativa del nivel 4 para CN-DSPN..... 148
Tabla 5.162: Superioridad comparativa de Sub-criterios CN-DSPN para CV1..... 148
Tabla 5.163: Vector de pesos del nivel 2 149
Tabla 5.164: Vector de pesos del nivel 3 para ED 149

Tabla 5.166: Vector de pesos de los Sub-criterios ED-AD para CV1 ..... *150*
Tabla 5.167: Vector de pesos del nivel 4 para ED-UD *150*
Tabla 5.168: Vector de pesos de los Sub-criterios ED-UD para CV1 ..... *150*
Tabla 5.169: Vector de pesos del nivel 3 para CN *151*
Tabla 5.170: Vector de pesos del nivel 4 para CN-EN *151*
Tabla 5.171: Vector de pesos de los Sub-criterios CN-EN para CV1 ..... *151*
Tabla 5.172: Vector de pesos del nivel 4 para CN-PEN *152*
Tabla 5.173: Vector de pesos de los Sub-criterios CN-PEN para CV1 ..... *152*
Tabla 5.174: Vector de pesos del nivel 4 para CN-DSPN *153*
Tabla 5.175: Vector de pesos de los Sub-criterios CN-DSPN para CV1 ..... *153*
Tabla 5.176: Ponderaciones parciales del nivel 2 ..... *153*
Tabla 5.177: Ponderaciones parciales del nivel 3 para ED *154*
Tabla 5.178: Ponderaciones parciales del nivel 4 para ED-AD *154*
Tabla 5.179: Ponderaciones parciales de los Sub-criterios ED-AD para CV1 ..... *154*
Tabla 5.180: Ponderaciones parciales del nivel 4 para ED-UD *155*
Tabla 5.181: Ponderaciones parciales de los Sub-criterios ED-UD para CV1 ..... *155*
Tabla 5.182: Ponderaciones parciales del nivel 3 para CN *156*
Tabla 5.183: Ponderaciones parciales del nivel 4 para CN-EN *156*
Tabla 5.184: Ponderaciones parciales de los Sub-criterios CN-EN para CV1 ..... *157*
Tabla 5.185: Ponderaciones parciales del nivel 4 para CN-PEN *157*
Tabla 5.186: Ponderaciones parciales de los Sub-criterios CN-PEN para CV1 ..... *157*
Tabla 5.187: Ponderaciones parciales del nivel 4 para CN-DSPN *158*
Tabla 5.188: Ponderaciones parciales de los Sub-criterios CN-DSPN para CV1 ..... *158*
Tabla 5.189: Prioridades finales de las alternativas con FAHP para CV1 ..... *159*
Tabla 5.190: Prioridades finales de las alternativas para el CV1 ..... *159*
Tabla 5.191: Parrilla de Acceso a Datos del CV2 ..... *161*
Tabla 5.192: Parrilla de Uso de los Datos del CV2 ..... *161*
Tabla 5.193: Parrilla de Entorno del Negocio del CV2 ..... *162*
Tabla 5.194: Parrilla del Proyecto Entorno del Negocio del CV2 ..... *163*
Tabla 5.195: Parrilla de Data Science del Proyecto Entorno del Negocio del CV2 ..... *163*
Tabla 5.196: Ponderación de ED-SC1 en relación a cada alternativa para CV2 ..... *164*
Tabla 5.197: Ponderación de ED-SC2 en relación a cada alternativa para CV2 ..... *165*
Tabla 5.198: Ponderación de ED-SC3 en relación a cada alternativa para CV2 ..... *165*
Tabla 5.199: Ponderación de ED-SC4 en relación a cada alternativa para CV2 ..... *165*
Tabla 5.200: Ponderación de ED-SC5 en relación a cada alternativa para CV2 ..... *166*
Tabla 5.201: Ponderación de ED-SC6 en relación a cada alternativa para CV2 ..... *166*
Tabla 5.202: Ponderación de ED-SC7 en relación a cada alternativa para CV2 ..... *167*
Tabla 5.203: Ponderación de ED-SC8 en relación a cada alternativa para CV2 ..... *167*
Tabla 5.204: Ponderación de ED-SC9 en relación a cada alternativa para CV2 ..... *167*
Tabla 5.204: Ponderación de ED-SC10 en relación a cada alternativa para CV2 ..... *167*
Tabla 5.206: Ponderación de CN-SC1 en relación a cada alternativa para CV2 ..... *168*
Tabla 5.207: Ponderación de CN-SC2 en relación a cada alternativa para CV2 ..... *168*
Tabla 5.208: Ponderación de CN-SC3 en relación a cada alternativa para CV2 ..... *168*
Tabla 5.209: Ponderación de CN-SC4 en relación a cada alternativa para CV2 ..... *169*
Tabla 5.210: Ponderación de CN-SC5 en relación a cada alternativa para CV2 ..... *169*
Tabla 5.211: Ponderación de CN-SC6 en relación a cada alternativa para CV2 ..... *169*
Tabla 5.212: Ponderación de CN-SC7 en relación a cada alternativa para CV2 ..... *170*
Tabla 5.213: Ponderación de CN-SC8 en relación a cada alternativa para CV2 ..... *170*
Tabla 5.214: Ponderación de CN-SC9 en relación a cada alternativa para CV2 ..... *170*

Tabla 5.215: Ponderación de CN-SC10 en relación a cada alternativa para CV2 171
Tabla 5.216: Ponderación de CN-SC11 en relación a cada alternativa para CV2 171
Tabla 5.217: Ponderación de CN-SC12 en relación a cada alternativa para CV2 171
Tabla 5.218: Ponderación de CN-SC13 en relación a cada alternativa para CV2 172
Tabla 5.219: Ponderación de CN-SC14 en relación a cada alternativa para CV2 172
Tabla 5.220: Ponderación de CN-SC15 en relación a cada alternativa para CV2 173
Tabla 5.221: Autovector máximo de cada sub-criterio del ED-AD para CV2 173
Tabla 5.222: Evaluación de consistencia de los sub-criterios del ED-AD para CV2 174
Tabla 5.223: Autovector máximo de cada sub-criterio del ED-UD para CV2 174
Tabla 5.224: Evaluación de consistencia de los sub-criterios del ED-UD para CV2 175
Tabla 5.225: Autovector máximo de cada sub-criterio del CN-EN para CV2 176
Tabla 5.226: Evaluación de consistencia de los sub-criterios del CN-EN para CV2 176
Tabla 5.227: Autovector máximo de cada sub-criterio del CN-PEN para CV2 177
Tabla 5.228: Evaluación de consistencia de los sub-criterios del CN-PEN para CV2 177
Tabla 5.229: Autovector máximo de cada sub-criterio del CN-DSPN para CV2 178
Tabla 5.230: Evaluación de consistencia de los sub-criterios del CN-DSPN para CV2 178
Tabla 5.231: Prioridad local de ED-AD para CV2 179
Tabla 5.232: Prioridad local de ED-UD para CV2 179
Tabla 5.233: Prioridad global de ED-AD para CV2 180
Tabla 5.234: Prioridad global de ED-UD para CV2 180
Tabla 5.235: Prioridad local de CN-EN para CV2 181
Tabla 5.236: Prioridad local de CN-PEN para CV2 181
Tabla 5.237: Prioridad local de CN-DSPN para CV2 182
Tabla 5.238: Prioridad global de CN-EN para CV2 182
Tabla 5.239: Prioridad global de CN-PEN para CV2 182
Tabla 5.240: Prioridad global de CN-DSPN para CV2 182
Tabla 5.241: Prioridad total con AHP para el CV2 183
Tabla 5.242: Matriz triangular difusa de comparación pareada de ED-SC1 para CV2 184
Tabla 5.243: Matriz triangular difusa de comparación pareada de ED-SC2 para CV2 184
Tabla 5.244: Matriz triangular difusa de comparación pareada de ED-SC3 para CV2 185
Tabla 5.245: Matriz triangular difusa de comparación pareada de ED-SC4 para CV2 185
Tabla 5.246: Matriz Triangular Difusa de Comparación Pareada de ED-SC5 para CV2 ... 185
Tabla 5.247: Matriz triangular difusa de comparación pareada de ED-SC6 para CV2 186
Tabla 5.248: Matriz triangular difusa de comparación pareada de ED-SC7 para CV2 186
Tabla 5.249: Matriz triangular difusa de comparación pareada de ED-SC8 para CV2 186
Tabla 5.250: Matriz Triangular Difusa de Comparación Pareada de ED-SC9 para CV2 ... 186
Tabla 5.251: Matriz triangular difusa de comparación pareada de ED-SC10 para CV2 187
Tabla 5.252: Matriz triangular difusa de comparación pareada de CN-SC1 para CV2 187
Tabla 5.253: Matriz Triangular Difusa de Comparación Pareada de CN-SC2 para CV2 ... 187
Tabla 5.254: Matriz triangular difusa de comparación pareada de CN-SC3 para CV2 188
Tabla 5.255: Matriz triangular difusa de comparación pareada de CN-SC4 para CV2 188
Tabla 5.256: Matriz triangular difusa de comparación pareada de CN-SC5 para CV2 188
Tabla 5.257: Matriz Triangular Difusa de Comparación Pareada de CN-SC6 para CV2 ... 188
Tabla 5.258: Matriz triangular difusa de comparación pareada de CN-SC7 para CV2 189
Tabla 5.259: Matriz triangular difusa de comparación pareada de CN-SC8 para CV2 189
Tabla 5.260: Matriz triangular difusa de comparación pareada de CN-SC9 para CV2 189
Tabla 5.261: Matriz triangular difusa de comparación pareada de CN-SC10 para CV2 190
Tabla 5.262: Matriz triangular difusa de comparación pareada de CN-SC11 para CV2 190
Tabla 5.263: Matriz triangular difusa de comparación pareada de CN-SC12 para CV2 190

Tabla 5.264: Matriz Triangular Difusa de Comparación Pareada de CN-SC13 para CV2 . *190*
Tabla 5.265: Matriz triangular difusa de comparación pareada de CN-SC14 para CV2 .... *191*
Tabla 5.266: Matriz triangular difusa de comparación pareada de CN-SC15 para CV2 .... *191*
Tabla 5.267: Extensión sintética difusa de Sub-criterios ED-AD para CV2 ..... *192*
Tabla 5.268: Extensión sintética difusa de Sub-criterios ED-UD para CV2 ..... *193*
Tabla 5.269: Extensión sintética difusa de Sub-criterios CN-EN para CV2 ..... *194*
Tabla 5.270: Extensión sintética difusa de Sub-criterios CN-PEN para CV2 ..... *195*
Tabla 5.271: Extensión sintética difusa de Sub-criterios CN-DSPN para CV2 ..... *196*
Tabla 5.272: Superioridad comparativa de Sub-criterios ED-AD para CV2 ..... *197*
Tabla 5.273: Superioridad comparativa de Sub-criterios ED-UD para CV2 ..... *197*
Tabla 5.274: Superioridad comparativa de Sub-criterios CN-EN para CV2 ..... *198*
Tabla 5.275: Superioridad comparativa de Sub-criterios CN-PEN para CV2 ..... *198*
Tabla 5.276: Superioridad comparativa de Sub-criterios CN-DSPN para CV2 ..... *199*
Tabla 5.277: Vector de pesos de los Sub-criterios ED-AD para CV2 ..... *199*
Tabla 5.278: Vector de pesos de los Sub-criterios ED-UD para CV2 ..... *200*
Tabla 5.279: Vector de pesos de los Sub-criterios CN-EN para CV2 ..... *201*
Tabla 5.280: Vector de pesos de los Sub-criterios CN-PEN para CV2 ..... *201*
Tabla 5.281: Vector de pesos de los Sub-criterios CN-DSPN para CV2 ..... *202*
Tabla 5.282: Ponderaciones parciales del Sub-criterios ED-AD para CV2 ..... *202*
Tabla 5.283: Ponderaciones parciales del Sub-criterios ED-UD para CV2 ..... *203*
Tabla 5.284: Ponderaciones parciales del Sub-criterios CN-EN para CV2 ..... *204*
Tabla 5.285: Ponderaciones parciales del Sub-criterios CN-PEN para CV2 ..... *204*
Tabla 5.286: Ponderaciones parciales del Sub-criterios CN-DSPN para CV2 ..... *205*
Tabla 5.287: Prioridades finales de las alternativas con FAHP para CV2 ..... *205*
Tabla 5.288: Prioridades finales de las alternativas del CV2 ..... *206*

Lista de Figuras

Figura 4.1: Esquema general del modelo propuesto *49*
Figura 4.2: Estructura jerárquica del modelo propuesto *51*
Figura 4.3: Valoración por niveles *52*
Figura 4.4: Desarrollo de métodos MCDM *56*
Figura 5.1: Ponderaciones globales del ED para el CV1 con AHP *117*
Figura 5.2: Ponderaciones globales de la CN para el CV1 con AHP *120*
Figura 5.3: Ponderaciones globales del ED para el CV1 con FAHP *155*
Figura 5.4: Ponderaciones globales de la CN para el CV1 con FAHP *158*
Figura 5.5: Ranking de prioridades finales para el CV1 ..... *160*
Figura 5.6: Ponderaciones globales del ED para el CV2 con AHP *180*
Figura 5.7: Ponderaciones globales de la CN para el CV2 con AHP *183*
Figura 5.8: Ponderaciones globales del ED para el CV2 con FAHP *203*
Figura 5.9: Ponderaciones globales de la CN para el CV2 con FAHP *205*
Figura 5.10: Ranking de prioridades finales para el CV2 ..... *206*
Figura 5.11: Integración de prioridades finales de los casos de validación *207*
Figura 5.12: Prioridades con AHP y FAHP para CV1 y CV2 ..... *208*

I

Introducción

En este capítulo se presenta primeramente los objetivos (el general y los específicos) del presente Trabajo Final de Maestría en la sección 1.1 y en la sección 1.2 se describe la estructura de la misma.

1.1. Objetivos

1.1.1. Objetivo general

Establecer un modelo basado en la Toma de Decisiones con Criterios Múltiples (MCDM, del inglés Multiple Criteria Decision Making), que sirva como soporte para la toma de decisiones al momento de seleccionar una metodología de Data Science (DS) o Ciencia de Datos.

1.1.2. Objetivos específicos

- Estudiar y analizar métodos de MCDM, en especial el Proceso Analítico Jerárquico (AHP, Analytic Hierarchy Process) y su variante combinado con lógica difusa, denominado Proceso Analítico Jerárquico Difuso (FAHP, Fuzzy Analytic Hierarchy Process).
- Revisar el estado del arte de las metodologías de DS referentes en la actualidad.
- Elaborar un modelo basado en MCDM que incluya criterios utilizando etiquetas lingüísticas para comparar las metodologías de DS a través de la técnica de la Teoría de la Construcción Personal.
- Implementar el modelo integrado con casos de aplicación reales, evaluar y validar los resultados obtenidos con expertos del área.

1.2. Estructura del documento

El presente Trabajo Final de Maestría (TFM) se estructura de la siguiente manera:
En el Capítulo II se encuentra el estado de la cuestión, abordando primeramente la MCDM, los métodos AHP y FAHP; para luego explicar la DS, junto a las metodologías elegidas que son CRISP-DM, Catalys y MoProPEI, y finalmente se expone la Técnica de Construcción Personal dentro de la Educción de Conocimiento, como técnica complementaria al modelo propuesto.

El Capítulo III describe el problema que llevó a la elaboración del presente trabajo.

La solución propuesta se expone en el Capítulo IV, donde se define la secuencia de pasos a seguir para el modelo propuesto.

Los casos de validación, la aplicación del modelo sobre los mismos y los resultados obtenidos, junto a su análisis se encuentran en el Capítulo V.

Finalmente, en el Capítulo VI se presentan las conclusiones arribadas, publicaciones realizadas y líneas futuras de investigación.

II

Estado de la cuestión

En este capítulo se describe el estado de la cuestión del presente TFM; para lo cual se aborda inicialmente la Toma Decisiones con Criterios Múltiples (MCDM) (sección 2.1), para luego tratar los métodos seleccionados, que son el Proceso Analítico Jerárquico (AHP) (sección 2.1.1) y su variante con lógica difusa FAHP (sección 2.1.2). Posteriormente se plantea una visión general de la Ciencia de Datos (Data Science, DS) y sus metodologías contemporáneas (sección 2.2), haciendo hincapié en CRISP-DM (sección 2.2.1), Catalys (sección 2.2.2), y MoProPEI (sección 2.2.3). Finalmente se describe la Técnica de Construcción Personal (sección 2.3.1) dentro de la Educción de Conocimiento (sección 2.3), como técnica complementaria al modelo propuesto que integra AHP y FAHP, para la comparación y elección de las metodologías de DS.

2.1. Toma Decisiones con Criterios Múltiples

La Toma de Decisiones (TD) es un proceso cotidiano para el ser humano, donde diariamente se encuentra ante diferentes situaciones y contextos, que requieren seleccionar una alternativa entre varias. Dicha elección se debe realizar en función a los propios juicios, satisfaciendo la mayor cantidad de necesidades requeridas [1].

Según Dieter [2], tomar una decisión genera una tensión psicológica para la mayoría de las personas. Dicha tensión proviene de dos fuentes: la primera tiene un impacto externalizado, donde el decisor se preocupa por las consecuencias que se derivan de cualquier curso de acción que elija y respecto a la segunda fuente, el decisor teme perder su prestigio y que se vea afectada su autoestima [1].

Para afirmar que una decisión es buena, se debe trazar un objetivo claro de lo que se quiere conseguir, reunir toda la información necesaria y relevante; además se de tener en cuenta las preferencias del decisor [3].

La TD es una actividad esencial dentro de la gestión de proyectos, de cualquier índole, que requiere especial atención, dado que en la actualidad se desarrollan proyectos de gran magnitud y complejidad que acrecientan la importancia de los problemas de decisión; con la dificultad que esto implica para los encargados de tomarlas, a la hora de enfrentarse a problemas complejos y en especial en un entorno cada vez más competitivo, donde las empresas tienen la necesidad de tomar buenas decisiones [4]. Es por ello que dentro de los proyectos se toman decisiones complejas marcadas por la trascendencia que poseen para ellos, por las responsabilidades implicadas para los directores del proyecto, por los agentes

afectados por la decisión tomada y por los diferentes criterios o puntos de vista que hay que tener en cuenta y que a menudo están en conflicto [1].

Antes de tomar cualquier decisión, el conocimiento, la experiencia y los hechos se deben reunir y evaluar en el contexto del problema. Es por ello que, los problemas de TD son procesos compuestos, donde el tomador de decisiones busca seleccionar entre varias alternativas, a partir de una serie de criterios; lo cual conlleva a una comparación y valoración de diversos aspectos; por lo que se requiere de la utilización de herramientas que satisfagan en el mayor grado posible la combinación de los criterios y permitan respaldar la elección de una de las alternativas [5], [6].

El Análisis o Teoría de Decisión con Múltiple Criterios (MCDA o MCDT, del inglés Multiple Criteria Decision Analysis/Theory) es una herramienta sumamente útil para ayudar al decisor durante este proceso de TD. Los métodos propuestos desde esta disciplina permiten de forma ordenada y sistemática, abordar un problema subyacente de subjetividad; ayudando a racionalizar un proceso complejo [1].

El MCDA puede verse como una herramienta analítica de una gran potencialidad en los procesos de ingeniería de sistemas. Existe una superposición de los enfoques multicriterio y sistémico a nivel conceptual y operativo. En un nivel conceptual, cuando los objetivos planteados entran en conflicto y requieren encontrar un compromiso o equilibrio entre ellos. Y a nivel operativo la ingeniería de sistemas puede entender como una secuencia de pasos en las que en todo momento es necesario evaluar y elegir entre diferentes alternativas o criterios [7].

La Toma de Decisiones de Criterios Múltiples (MCDM, del inglés Multiple Criteria Decision Making) es un procedimiento que combina el rendimiento de las alternativas de varios criterios cualitativos y/o cuantitativos y obtiene como resultado una solución de compromiso [8]. Estos métodos son frecuentemente aplicables, en numerosos problemas de la vida real, donde se evalúan conjuntos de alternativas de decisión en función de criterios en conflicto [9]. En [7], [9] y [10] se pueden encontrar aspectos genéricos de la metodología detrás de MCDM y los enfoques más populares para evaluar soluciones alternativas en aplicaciones del mundo real.

Para realizar una correcta selección del método de MCDM a utilizar, se debería considerar un rango de diferentes perspectivas para comprender todos los lados del problema y, cuando sea necesario, considerar interconexiones entre los criterios. Los métodos de MCDM requieren de un procedimiento para estructurar las decisiones, demostrando la compensación entre los

criterios y así ayudar a los encargados de tomar las decisiones a reflexionar, articular y aplicar juicios de valor relacionados con dichas compensaciones de forma satisfactoria, lo que da como resultado sugerencias al considerar alternativas, para estimar el riesgo y la incertidumbre de manera más consistente y razonable, para simplificar la negociación de los criterios y mantener un registro de cómo se toman las decisiones [11], [12]. Algunas aplicaciones del mundo real consideradas como problemas de MCDM, pueden encontrar complicaciones cuando, por ejemplo en los casos aparentemente más simples de atributos cualitativos, la calidad de los datos puede ser una fuente importante de incertidumbre estadística; asimismo, las alternativas se derivan de una amplia gama de opciones, que tienen como objetivo priorizarlas y organizarlas de forma jerárquica. Se debe considerar cuidadosamente el hecho de que diferentes atributos o criterios pueden causar conflictos debido a su grado de integridad, redundancia, reciprocidad e independencia, lo que puede complicar aún más el proceso de TD [9], [12].

Ro y Słowiński [13] ofrecen una serie de preguntas que pueden ayudar a elegir un método MCDM adaptado al contexto de la decisión. Estas preguntas tienen en cuenta varios aspectos del proceso de decisión, presentando las preguntas en un orden jerárquico, desde las más generales y cruciales, a través de otras preguntas pertinentes sobre la agregación de múltiples criterios, hasta las secundarias.

MCDM ha sido y es un área de investigación muy valorada desde hace décadas; surgió en los años 60 y fue expandiéndose rápidamente en todo el mundo; debido principalmente a su naturaleza interdisciplinaria y su aplicación en diferentes problemas reales de decisión con variados criterios y a la formación de grupos de especialistas y sociedades científicas de expertos internacionales (EURO Working Group on Multicriteria Decision Aid, Special Interest Group on MCDM, International Society on Multiple Criteria Decision Making) [14].

El analista puede dudar entre varios tipos de resultados convenientes, entre varios tipos de métodos; incluso si selecciona solo uno, el mismo puede producir más de un resultado y como consecuencia de ello, en muchos casos, puede llevar a una lista corta de más de un método. Una revisión de los métodos MCDM realizada en [13], distingue cinco tipos principales de resultados que podrían considerarse a la hora de seleccionarlos:

- Tipo 1: Asignar un valor numérico (puntaje) a cada acción potencial: Muchos métodos pueden producir este tipo de resultados: AHP (Analytic Hierarchy Process) [15], [16] MAVT (Multi Attribute Value Theory) [17], TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) [18], UTA (UTility Additive) [19],

SMART (Simple Multi Attribute Rating Technique) [20], MAUT (Multi-Attribute Utility Theory) [21], MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique) [22] [23], Choquet Integral [24], función de valor representativo de UTA^{GMS} [25] , entre otros. La aplicación de este tipo de métodos requiere contar con una escala de cada criterio de evaluación, donde la misma posea intervalos. Dichas escalas sirven para interactuar con el tomador de decisiones de una manera específica para cada uno de estos métodos. En algunos contextos, se deben tener en cuenta algunos puntos de vista para los cuales la definición de criterios asociados; además de los datos necesarios para definir el desempeño de cada acción según estos criterios, puede implicar cierta indeterminación. Dicha indeterminación puede manejarse usando modelos probabilísticos o difusos, como Fuzzy AHP (FAHP) [26], [27], [28] y Fuzzy TOPSIS [29], sin embargo, este modelado también puede ser arbitrario en cierta medida. Esta dificultad se puede evitar mediante el análisis de sensibilidad, cuando solo hay uno o dos puntos de vista que necesitan el manejo del impacto de esta indeterminación. No obstante, puede suceder que la diversidad y la importancia de las fuentes de indeterminación dificulten la asignación de un valor numérico a cada una de las acciones potenciales.

- Tipo 2: Clasificar el conjunto de acciones (sin asociar un valor numérico a cada una de ellas) como un orden completo o parcialmente débil: Este tipo de resultado solo puede considerarse si el conjunto A de acciones potenciales se conoce a priori. No es conveniente cuando las acciones potenciales se examinan a medida que llegan. Los métodos relevantes aquí son: ELECTRE III, IV [30], [31], PROMETHEE I y II [32], los métodos de regresión ordinal robusta [33] que producen las clasificaciones necesarias y posibles, como UTA^{GMS} [25], GRIP [34], Extreme Ranking Analysis [35], RUTA [36], ELECTRE^{GKMS} y PROMETHEE^{GKMS} [37], el enfoque basado en dominación basado en el juego en bruto para el ranking (Dominance based Rough Set Approach to ranking) [38] y el enfoque de aprendizaje automático (Machine Learning) [39].
- Tipo 3: Seleccionar un subconjunto de acciones, lo más pequeño posible, en vista de la elección final de una o al principio de algunas acciones: Como en el caso del tipo 2, este tipo de resultado no es conveniente cuando el conjunto A de acciones potenciales no se conoce a priori. El tipo de resultado considerado aquí es conveniente cuando las acciones potenciales se modelan como alternativas, es decir, de tal manera que se

excluye una ejecución conjunta de cualquiera de ellas. Este tipo de resultados se produce directamente mediante métodos tales como: ELECTRE I e IS [30], [31], PROMETHEE V [32] y RUBIS [40]. Además hay que tener en cuenta, que los métodos considerados para resultados del tipo 1 y 2 también se pueden usar en este caso; donde las acciones mejor clasificadas pueden verse como el resultado del tipo 3.

- Tipo 4: Asignar cada acción a una o varias categorías, dado que el conjunto de categorías se ha definido a priori: Puede ser conveniente realizar una ordenación previa cuando se ha incluido un gran número de acciones potenciales en el punto de inicio del proceso de decisión. Tal pre-clasificación también se usa en algunos procedimientos interactivos de optimización multi-objetivo, donde se refiere a una serie de soluciones no dominadas propuestas para la evaluación al tomador de decisiones en cada fase de diálogo [41]. Varios métodos proporcionan este tipo de resultados: los basados en el enfoque de dominación Rough Set [38], PREFDIS [42], UTADIS^{GMS} [33], ELECTRE TRI-B (inicialmente ELECTRE TRI) [30], ELECTRE TRI-C [43], ELECTRE TRI-NC [44], TRINOMFC [45], PROAFTN [46], PAIRCLASS [47], THESEUS [48], entre otros.
- Tipo 5: Proporcionar un subconjunto de acciones potenciales que disfruten de algunas propiedades notables, para servir como base en la siguiente etapa del proceso de ayuda a la decisión: Este tipo de resultado puede ser necesario cuando el conjunto de alternativas contiene una gran cantidad de acciones (más de cien). Por ejemplo, cuando las acciones están definidas por vectores de variables sujetas a algunos cálculos matemáticos y restricciones de programación. En la optimización multi-objetivos, el conjunto A' es un conjunto de acciones no dominadas (también llamadas soluciones eficientes u óptimas de Pareto) o una aproximación de este conjunto. Los fundamentos matemáticos de Wierzbicki [49] han dado una descripción completa y constructiva de la caracterización paramétrica del conjunto de acciones no dominadas; desde un punto de vista práctico. Los algoritmos evolutivos parecieran ser particularmente efectivos para encontrar una buena aproximación del conjunto de acciones no dominadas en la optimización de múltiples objetivos [50]. Este tipo de resultado es adecuado en la optimización interactiva de múltiple objetivos.

Dentro de MCDM existe una gran variedad de métodos, como los descriptos anteriormente. Para este TFM se optó por el Proceso de Análisis Jerárquico (AHP, Analytic Hierarchy Process) [15], [51] y su variante con lógica difusa, Proceso de Análisis Jerárquico Difuso

(FAHP, Fuzzy Analytic Hierarchy Process) [26]; por adecuarse al contexto escogido, que consta de un objetivo concreto, que consiste en la elección de una metodología de Data Science (Ciencia de Datos; conocido popularmente como Minería de Datos (DM, Data Mining) o Ingeniería de Explotación de Información) que se adapte a diversos escenarios, para lo cual se adecuó a una escala definida (tipo 1 [13]); por otra parte, AHP y FAHP son ampliamente estudiados y difundidos en diferentes ámbitos [52], [53], [54]; lo que respalda su utilización.

2.1.1. Proceso Analítico Jerárquico

El Proceso Analítico Jerárquico (AHP) fue propuesto por Thomas Saaty en 1980 [16], [55]. Su idea se fundamenta en que la complejidad inherente a un problema de MCDM, se puede resolver a través de la estructuración jerárquica de los problemas planteados. Precisamente AHP asiste a la TD, mediante la descomposición de un problema complejo en una estructura jerárquica de varios niveles de objetivos, criterios, sub-criterios y alternativas [1], [15], [16], [56], [57].

AHP es una herramienta eficaz para abordar la TD compleja y puede ayudar al encargado de ello, a establecer prioridades y tomar la mejor decisión. Es un método de medición a través de comparaciones por pares y se basa en los juicios de los expertos para derivar en escalas de prioridades. Ayuda a capturar los aspectos subjetivos y objetivos de una decisión, al reducir las decisiones complejas a una serie de comparaciones pareadas y luego sintetizar los resultados. Asimismo, incorpora una técnica útil para verificar la consistencia de las evaluaciones del tomador de decisiones, lo que reduce el sesgo en el proceso de TD [52], [58].

Saaty construyó el método buscando una práctica sistemática para definir prioridades y apoyar la TD compleja [59]. De hecho, la estructura jerárquica de la metodología AHP es capaz de medir y sintetizar una variedad de factores de un proceso complejo de TD de manera jerárquica, lo que simplifica la combinación de las partes en un todo [52].

Estructurar la complejidad, medición y síntesis, son las tres funciones principales del método AHP [59]. Saaty sostiene que para lidiar con la complejidad de un proceso de TD, se necesita identificar los diferentes factores que afectan la decisión y organizarlos en una estructura jerárquica de grupos homogéneos de factores [59]. La razón medida en escala se obtiene comparando los factores de a pares. Las ponderaciones o prioridades en toda la jerarquía se

calculan obteniendo el producto de ellas; de modo que la prioridad de un factor de cada nivel se multiplica por la prioridad del factor con el que se vincula el primero (factor principal). El método es interesante debido a su capacidad para medir y sintetizar la multitud de factores en una jerarquía. Precisamente la supremacía de AHP radica en su capacidad de atribuir un peso relativo a todos los elementos de un problema, tangible o no y construir una jerarquía de su relevancia relativa [52], [59].

Para tomar una buena decisión, el responsable de ello debe definir claramente: el problema, la necesidad y el propósito de la decisión, los criterios y sub-criterios para evaluar las alternativas, las acciones que se deben tomar, así como las partes interesadas y los grupos afectados [60]. Estos criterios y sub-criterios pueden ser tangibles o intangibles; estos últimos no tienen forma de ser medidos en la clasificación de las alternativas. La creación de prioridades para los criterios en sí mismos, con el fin de sopesar las prioridades de las alternativas y sumar todos los criterios para obtener los rangos generales deseados de las alternativas, es una tarea desafiante [52], [60]. Para cumplir con lo mencionado AHP se puede dividir en una serie de pasos [1], [52], [60], [61]:

1. Definición del problema, determinación del tipo de conocimiento buscado y estructuración jerárquica: Aquí se selecciona el problema a analizar entre todos aquellos que se consideran importantes o lo suficientemente complejos como para ser analizados. Al definir y seleccionar un problema es importante hacer explícitos todos los supuestos y la perspectiva por la cual se necesita tomar una decisión. La estructura jerárquica se construye identificando el objetivo principal o meta, luego en los niveles intermedios se definen los criterios y sub-criterios si los hay y en el nivel inferior las alternativas. Una vez que se define el objetivo principal, la estructuración inferior puede hacerse correlacionando a través de un proceso de arriba hacia abajo, de los criterios a las alternativas, o a través del proceso inverso, de las alternativas a los criterios; siendo esta estructura lo suficientemente precisa y detallada, para incluir las principales preocupaciones de los tomadores de decisiones.
2. Construcción de las matrices de comparaciones por pares y establecimiento de las prioridades entre los criterios: Como menciona Saaty [60], cada elemento en un nivel superior se usa para comparar los elementos en el nivel inmediatamente inferior con respecto a él. La comparación se realiza a través de una escala definida, con el propósito de mostrar el grado de importancia o dominancia de un elemento respecto a otro, en relación al criterio o propiedad para el cual se comparan. Dicha escala es

verbal y se usa para medir criterios cuantitativos y cualitativos; mientras que las comparaciones pareadas ayudan a que el análisis sea más preciso.

3. Establecimiento y control de las prioridades locales y globales, mediante el cálculo del peso relativo a los elementos para cada nivel: Este cálculo se realiza en una secuencia de pasos, que de forma resumida incluye, luego de completar la matriz pareada, normalizarla, obtener la prioridad relativa de los criterios, para luego evaluar la consistencia de la matriz; si no lo es, se deberá revisar las comparaciones realizadas y proceder nuevamente los pasos mencionados.
4. Establecimiento de las prioridades totales asociadas a cada alternativa: consiste en calcular las prioridades totales asociadas a cada alternativa, que representan la importancia de las alternativas con respecto al objetivo o meta global; la mejor alternativa es la que tiene el valor más alto de prioridad.
5. Análisis de sensibilidad, comprobación y balance de la decisión: se puede realizar un análisis de sensibilidad que corrobore si los resultados obtenidos son robustos. Verificar si los resultados de la aplicación del método son compatibles con las expectativas y si se identifican problemas, se debería realizar una revisión del proceso anterior. De ser necesario, se puede complementar la inclusión de criterios no identificados o considerados previamente. Es conveniente también registrar todas las razones que respaldan cómo y por qué se tomó la decisión. Estos registros pueden ser útiles para justificar el proceso ante terceros o para reflexionar sobre él en el futuro, lo que permite una mejora continua del proceso de TD.

En el Anexo I se describen los aspectos teóricos vinculados a AHP, como ser sus axiomas, etapas y secuencia de pasos generales.

Las fortalezas de AHP radican en los siguientes aspectos: dentro de las técnicas de MCDM, es una de las pocas que brinda una axiomática teórica; desde el punto de vista práctico, se caracteriza por su buen desempeño; AHP proporciona un modelo fácil de comprender, es flexible y se adapta a una variedad de problemas; incorpora enfoques de sistemas y deductivos para solucionar problemas complejos; pone en manifiesto la forma natural de agrupar elementos semejantes por niveles, lo que facilita su entendimiento; provee de una escala de medición de criterios, que permite establecer prioridades; mediante su poder de síntesis, guía en la estimación total de la utilidad de cada alternativa; permite escoger la mejor alternativa en función de los objetivos, considerando para ello, las prioridades relativas a los factores en un sistema; mediante la repetición del proceso, permite que la persona encargada

de tomar la decisión ajuste su definición del problema y adecue su juicio; posee también la capacidad de sintetizar juicios diferentes, sin exigir un consenso [1].

Como se ha mencionado, AHP es una de las herramientas de MCDM más utilizada por los responsables de la TD e investigadores, por ser simple y poderosa [52], [59], [62], [63]. Ha tenido una muy buena aceptación, pudiendo ser aplicada en diversos ámbitos de la MCDM, como el tecnológico, comercial, educativo, industrial, agrario, desarrollo urbano, entre otros. Algunos de los casos de aplicación, se pueden encontrar en [6], [14], [57], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73]; adicionalmente en [1] y [52] se realizan revisiones sistemáticas de la literatura, donde se mencionan otras investigaciones que han utilizado AHP con éxito para la TD.

Si bien AHP es ampliamente utilizado hasta el día de hoy, no ha estado exento de críticas, incluidos los problemas del significado de consistencia, escala fundamental, los requerimientos de grandes conjuntos de datos, entre otros. Como todas las técnicas de MCDM, AHP tiene aspectos positivos y negativos [1], [74].

La derivación de un vector de prioridad confiable para una matriz de comparación por pares (PCM, del inglés Pair-wise Comparison Matrix), que se recopila a partir de los juicios de los expertos, es un tema importante en AHP. La validez de los vectores de prioridades derivados se basa principalmente en el diseño de métodos de priorización [58]. En las últimas décadas, se han desarrollado muchos métodos de priorización [58], incluido el método de vector propio (EV, EigenVector) creado por Saaty [75], método de los mínimos cuadrados ponderados (WLS, Weighted Least Squares) [76], método de mínimos cuadrados (LS, Least Squares) [77], método de peso propio gradiente (GE, Gradient Eigenweight) y método la distancia mínima (LD, Least Distance) [78], método de mínimos cuadrados logarítmicos (LLS, Logarithmic Least Squares) [79], método de los mínimos cuadrados geométricos (GLS) [74], método de programación de metas (GP, Goal Programming) [80], método de media geométrica (GM, Geometric Mean) [81], método de programación por metas logarítmicas (LGP, Logarithmic Goal Programming) [82], método de programación de preferencias difusas (FPP, Fuzzy Preference Programming) [83], método de descomposición de valores singulares (SVD, Singular Value Decomposition) [84], método de intervalo de prioridad (IP, Interval Priority) [85], método de normalización aditiva (AN, Additive Normalization) [16], [86], método de programación lineal (LP, Linear Programming) [87], método del análisis envolvente de datos (DEA, Data Envelopment Analysis) [88], enfoque de maximización del coeficiente de correlación (CCM, Correlation Coefficient Maximization) [89], procedimiento

de priorización Bayesiano (BP, Bayesian Prioritization) [90], estimación robusta del vector de prioridad utilizando las matrices transformadas [91], estrategia de evolución bi-criterio para derivar la estimación de ponderaciones [92], enfoque heurístico para la derivación del vector de prioridad [93], método de maximización de coseno (CM, Cosine Maximization) [58]. Hay ciertas disputas sobre qué método de asignación de prioridades es mejor, la elección debe ser dictada por el objetivo del análisis, dado que algunos métodos funcionan mejor en algunos casos y otros son mejores en otros. Hasta ahora, el problema de la superioridad relativa de los métodos de priorización aún no está resuelto [58]. Algunos análisis comparativos entre los métodos de priorización comúnmente utilizados se pueden encontrar en [76], [77], [83] y [94].

Otro de los inconvenientes vinculados a AHP, es la limitación en cuanto al tamaño del problema considerado, es decir, al número de alternativas comparadas [95]. Conforme a las consideraciones psicológicas, Miller [96] conjeturó que existe un límite superior en la capacidad humana para procesar información sobre elementos que interactúan simultáneamente con precisión confiable y de validez; este límite es 7 ± 2 elementos a comparar [95], [97]. La medida de inconsistencia aleatoria revela que a medida que aumenta el número de elementos que se comparan, la medida de inconsistencia disminuye lentamente dejando poco espacio para mejorar los juicios y por lo tanto a la consistencia. Para servir tanto a la coherencia como a la redundancia, es mejor mantener el número de elementos en siete o menos [97]. Sin embargo, dada la complejidad de algunos problemas reales, especialmente cuando intervienen aspectos cualitativos o intangibles, no siempre se puede aplicar dicho criterio; es por ello que se han desarrollado procedimientos que aprovechan las medidas relativas y absolutas en los Problemas de Gran Tamaño (PGT), tal es el caso de [98], donde los autores combinan medidas relativas y absolutas, denominadas benchamark o alternativas de referencia; en [95] se presentan dos procedimientos alternativos, basados en la composición de conglomerados, los cuales utilizan medidas relativas para obtener las prioridades de un conjunto grande de alternativas y en [99] utilizan la Teoría del Pesar, con el propósito de alcanzar una escala de razón de las prioridades concernientes a las alternativas del problema, mediante la combinación de aspectos tangibles e intangibles; para ello los autores realizan un análisis para obtener las prioridades aplicando AHP (medidas relativas), utilizando distintos procedimientos de normalización (distributivo e ideal) y síntesis (aditiva y multiplicativa).

En relación a los problemas de inconsistencias en las matrices AHP, Saaty [15], [75] propuso un una técnica útil para verificar la consistencia [52], [58]; aunque con algunas críticas, por lo que se han realizado propuestas de mejoras como en [100] donde se propone un coeficiente de consistencia ordinal y una prueba no paramétrica de tríadas circulares de respuestas de comparación pareadas (inconsistencias); para lo cual considera las situaciones en las que los encuestados solo pueden proporcionar juicios ordinales con confiabilidad y validez. En [101] los autores proponen una prueba de consistencia para la ordinalidad de los elementos en una matriz de comparación pareada en AHP como en una prueba sensorial; utilizando el número de tríadas circulares en una tabla de preferencias sin vínculos. Cuanto mayor sea la calidad del ranking, mayor será su credibilidad, en [102] se determina el número máximo de triadas inconsistentes en el conjunto de comparaciones de pares ordinales con vínculos de tamaño arbitrario; extendiendo así el uso de este índice al AHP y otros métodos cuantitativos basados en la comparación de alternativas en pares. Además en [102] se introduce las nociones de un torneo generalizado y un torneo doble como gráficas que modelan comparaciones ordinales de pares con empates y el conjunto de comparaciones de emparejamientos máximamente inconsistentes con empates, respectivamente; y muestra la relación entre el conjunto más inconsistente de comparaciones por pares con vínculos y el problema de cobertura de conjunto.

Los índices de inconsistencia pueden prestar atención a diferentes aspectos del conjunto de comparaciones por pares, dependiendo de su construcción; en [103] se propone una familia de índices, tratando de combinar sus ventajas, lo que aumenta la expresividad de los elementos de la familia, a su vez que ayuda a compararlos e identificar sus propiedades comunes.

Respecto a las escalas de juicio en la estimación de prioridad en AHP; en [104] se compara y discute la aplicación de diversas escalas de juicio sobre los resultados en un ejemplo práctico propuesto por Saaty [105]; se analiza el impacto de uso las mismas y la consistencia a la escala por defecto; a partir del estudio, los autores mencionan que las escalas de juicio tienen un gran impacto en las prioridades de los criterios, pero no en la clasificación de los mismos. Adicionalmente en [104], se calculan los valores de índice aleatorio necesarios para el cálculo del índice de consistencia en AHP para todas las escalas en cuestión. En [106] se analizan y adecuan varias escalas verbales individualizadas, que fueron aplicadas en un caso de estudio real para seleccionar estrategias de computación en la nube.

En ocasiones la TD debe realizarse en grupos grandes, por lo que se han propuestos alternativas para lograr consensos y consistencia en la decisión final, como los expuestos en [107] y [108].

Por otra parte, en [109] se mide la consistencia de los encargados de la TD utilizando el modelo de nivel de experiencia. Y en [110] los autores proponen una distribución beta para modelar las diferentes preferencias estocásticas de los tomadores de decisiones; utilizando para ello el método de momentos con el propósito de ajustar las diferentes preferencias estocásticas de los tomadores de decisiones y un modelo de programación no lineal para calcular las ponderaciones que maximicen las preferencias, mintiendo el nivel de consistencia. Los tomadores de decisiones tienen conocimientos limitados para especificar sus preferencias sobre comparaciones de múltiples pares. Su diversidad debería modelarse utilizando una distribución de probabilidad única o como conjuntos difusos [110]. Varios de los problemas de MDCM, presentan múltiples criterios cuantitativos y cualitativos, por esta razón, cuando se trata de estos últimos, métodos como AHP pueden no resultar tan eficientes, es por ello que surgieron los métodos híbridos, tales como el FAHP [26] que acopla las bondades de AHP con la lógica difusa [1]. Este método realiza un acercamiento a la TD por parte de los humanos; aunque se debe tener recaudo, dado que puede presentar inconsistencias al obtener el vector propio para AHP a través de la matriz de comparación triangular difusa [27]. FAHP tiene una buena aceptación en diferentes áreas que involucran aspectos cualitativos, como ser la Administración, donde por ejemplo se mide el capital humano, en términos de actitudes, habilidades y capacidades [1].

2.1.2. Proceso Analítico Jerárquico Difuso

Los problemas de TD presentan cierta incertidumbre; donde el lenguaje natural se emplea a menudo para articular el pensamiento y las percepciones subjetivas propias del proceso decisorio. Precisamente, un vocablo puede no tener significancia precisa y estar incorrectamente definido; como consecuencia, si se lo usa como etiqueta para un conjunto, los límites del mismo, al que pertenecen o no los objetos se volverán borrosos (difusos). Para superar este inconveniente, los números (o conjuntos) difusos se incluyen de modo que las variables lingüísticas se manifiesten de manera apropiada [111].

Los conjuntos difusos (fuzzy sets), derivados de la teoría de conjuntos clásicos, fueron introducidos en 1965 por Zadeh [112]. Consisten en una representación matemática de la

ambigüedad del lenguaje ordinario [113], [114]. Se basan en el hecho de que una afirmación no solo es verdadera o falsa, sino que se puede establecer una escala de valores intermedios entre la certeza absoluta o la falsedad [115]. El aglomerado de objetos está caracterizado por una función de pertenencia, admitiendo a cada elemento estar valorado en el intervalo [0,1]; en lugar de en el conjunto {0,1}, como es el caso de la teoría de conjuntos clásica [1], [57], [116].

Dentro de los conjuntos difusos, los términos lingüísticos se representan mediante de funciones de membresía, valoradas en intervalos de los números reales, utilizados para interpretar la imprecisión y ambigüedad del pensamiento humano relacionado a un problema [113], [117]. Expresado de otra manera, la perspectiva borrosa permite modelar el nexo entre las categorías conceptuales los datos y las personas, en el cual los elementos dentro de las clases pueden superponerse, por lo que no se logra determinar la pertenencia o exclusión absoluta de un elemento a una clase en particular [114].

La integración de los conjuntos difusos a AHP; surgió en los años 80 y hasta la actualidad se han desarrollado diferentes métodos para manejar matrices de comparación difusa [27], de los que se destacan los siguientes: Van Laarhoven y Pedrycz [118], en 1983, presentaron un método difuso para seleccionar entre varias alternativas bajo criterios de decisión conflictivos; para ello propusieron el método de los mínimos cuadrados logarítmicos (LLSM, Logarithmic Least Squares Method) para alcanzar los pesos triangulares borrosos de una matriz de comparación difusa triangular. Buckley [26], en 1985, incorporó a AHP una matriz borrosa, donde utilizó el método de la media geométrica para calcular los pesos difusos, con el propósito de integrar la vaguedad en la respuesta de las personas involucradas en la TD y proporcionar una TD con mayor validez [27], [119]. Boender y otros autores [120], en 1989, presentaron una modificación del método LLSM propuesto por Van Laarhoven y Pedrycz [121]. Chang [28], en 1996, planteó un método de análisis extendido, que deriva ponderaciones nítidas para matrices de comparación borrosas [27]. Xu [122], en 2000, presentó el método de prioridad de los mínimos cuadrados difusos (LSM, Least Squares Method) [27]. Csutora y Buckley [123], en 2001, crearon un método llamado Lambda-Max, que es la versión difusa directa del método λ_{max} [27]. Mikhailov [124], en 2003, desarrolló un método de programación de preferencias borrosas (PPM, Preference Programming Method), que deriva pesos nítidos a partir de matrices de comparación difusa [27]. Por otra parte, Wang, Elhag y Hua [125], en 2006, expusieron otro método LLSM borroso modificado [27].

Cada uno de estos método tiene sus fortalezas y debilidades; el método de análisis extendido, propuesto por Chang [28], es uno de los más utilizado en diferentes aplicaciones como las descriptas en [126], [127], [128], [129], [130], [131], [132], [133], [134], [135], entre otras; gracias a su simplicidad computacional. Sin embargo, dicho método tiene ciertas críticas como por ejemplo, que es incapaz para derivar los pesos reales de una matriz de comparación borrosa o nítida; a pesar de las críticas, sigue siendo el método más popular [27], [136].

Desde la perspectiva matemática, existe una variedad de funciones de membresía, de las cuales las más empleadas son la Triangular, Trapezoidal o Gaussiana [1], [113]. Asiduamente se utilizan los números difusos triangulares, para plasmar la imprecisión de los parámetros existentes en la TD [113]. Huang y Wu [119] definieron de forma sintética a FAHP en tres pasos: en primer lugar se utilizan los números borrosos triangulares para trasladar las ideas de los expertos en una matriz de reciprocidad positiva; seguidamente se emplea un método de media geométrica para ponderar los valores parciales para cada alternativa, con la conexión jerárquica establecida y, por último, se usa una función de permanencia por lo que se clasifican las prioridades finales para cada alternativa [137].

En el Anexo II se detallan los pasos formales de FAHP, tenidas en cuenta en la presente TFM.

Existen revisiones de la literatura en relación al uso de FAHP, de las cuales se destacan [111], [136] y [137], las cuales abarcan, en diferentes intervalos, publicaciones efectuadas entre 1994 y 2016. Siendo los sectores de mayor aplicación la industria, manufactura, gobierno, administración, medio ambiente, comercio, ingeniería y educación; y Asia es la líder en este campo, donde los países con mayores artículos publicados son Taiwán, Turquía, Irán y China.

2.2. Metodologías de Ciencia de Datos

Se considera que los datos son el camino para una mejor TD y rentabilidad [138]. Un estudio realizado en [139], señala que cada vez son más las empresas que se basan en datos, las cueles han mejorado su desempeño en medidas objetivas de los resultados financieros y operativos.

En un entorno tan competitivo como el actual, las organizaciones, de cualquier tipo, requieren estrategias para poder adaptarse a los cambios y generar conocimiento a partir de los datos disponibles, es una muy buena estrategia [140]. Para aprovechar al máximo la revolución de los datos, directivos e investigadores deben atender a la importancia y las implicaciones de los datos para la TD [138].

Debido a la evolución de las TIC, los sistemas de almacenamiento pueden acopiar gran cantidad de datos, de diferente índole, a un menudo coste; lo que conlleva a que estos crezcan de manera exponencial en pequeños intervalos de tiempo [140]. Los datos no son recogidos únicamente para el mantenimiento de registros o para validar las hipótesis generadas por los recursos humanos, sino que a medida que el volumen de datos crece, existe la posibilidad de probar las hipótesis que aún no se habían previsto [141]. El crecimiento en cantidad y diversidad de los datos ha llevado a que las herramientas convencionales de administración práctica, no puedan manejar estos conjuntos de datos [138].

Para poder tomar decisiones pertinentes y administrar estos conjuntos de datos potencialmente invaluables, es necesario realizar un análisis exhaustivo sobre los mismos. La Ciencia de Datos es un conjunto de principios fundamentales que apoyan y guían la extracción de información y conocimiento a partir de los datos; incluye diversas metodologías, técnicas, algoritmos y herramientas que facilitan el procesamiento avanzado y automático de los mismos; permitiendo identificar información relevante y estratégica, que a simple vista no es detectada [138], [140], [142].

Con el trascurso de los años y a medida que se ha ido extendiendo, su denominación ha variado de Minería de Datos o Data Mining (DM), Ingeniería de Explotación de Información, a Ciencia de Datos o Data Science (DS); el concepto coincidente aquí es la extracción real del conocimiento a partir de datos y tecnologías que incorporan estos principios [140], [142].

Puede definirse a DS, como la aplicación de métodos cualitativos y cuantitativos para predecir resultados y solucionar problemas relevantes. Con la creciente e inmensa cantidad de datos que existe actualmente, el conocimiento y el análisis del dominio no pueden estar separadas [138], [143].

DS es un campo emergente que combina la pericia de áreas como la informática, estadística, matemática y ciencia del comportamiento, junto al análisis predictivo, con el propósito de pronosticar resultados [144].

El propósito es obtener conocimiento especializado, a partir de resultados representados como modelos o patrones, los cuales son convertidos en información valiosa para la correcta TD. Este tipo de proyectos pueden ser implementados a partir de una situación organizacional o una exploración de los datos disponibles [140], [145], [146].

Existen muchos algoritmos para el procesamiento y extracción de datos, así como una gran cantidad de detalles sobre los métodos aplicados en este campo; pero DS implica mucho más que solo algoritmos de DM. DS proporciona a los profesionales del área una estructura y un

conjunto principios que brindan un marco para tratar sistemáticamente los problemas de extracción de conocimiento útil de los datos; donde los métodos y la metodología para tratar los datos y llevar adelante este tipo de proyectos son trascendentales [138], [142]. Dentro de la ingeniería de sistemas, los proyectos de DS son un tipo especial de proyectos; cuyo objetivo de extraer conocimiento relevante, idóneo y original de los datos [147], [148].

El incremento de profesionales académicos y aplicados vinculados a DS, ha ido creciendo significativamente en los últimos años. Éstos profesionales requieren de habilidades analíticas y un profundo conocimiento del dominio [138], [140], [142].

Existe varias metodologías para el desarrollo de proyectos, en ocasiones la elección no es tarea sencilla, especialmente para los que se inician en el área. Una de las habilidades deseadas para los científicos de datos, es que sean capaces de seleccionar una metodología adecuada para un proyecto en particular [138], [140].

Existen diversas metodologías que han ido perfeccionándose con el correr de los años. En el año 2000 se crearon las primeras metodologías validadas como tal por la comunidad científica, de las que se destacan CRISP-DM [149] y Catalyst (conocida como P^3TQ) [150]. Dichas metodologías se encuentran probadas y validadas por la comunidad científica y de la industria; éstas presentan ciertos problemas, como los relacionados con la gestión de proyectos [147], [151]. Por su parte MoProPEI [152] es una metodología más reciente, creada por profesionales argentinos, que propone mejoras respecto a las anteriores. Algunos estudios comparativo sobre las virtudes y debilidades de determinadas metodologías están disponibles en [145], [146], [153], [154], [155], [156], [157]. Cabe destacar que existen otras metodologías, métodos y procesos creados en los últimos años que comienzan a ser populares en proyectos de DS y Big Data, tal es el caso del Método Unificado de Soluciones Analíticas para Minería de Datos o Analytics Solutions Unified Method - Data Mining (ASUM-DM) [158], Scrum-DM (Agile Data Mining) [159], el Proceso de Equipos en Ciencia de Datos o Team Data Science Process (TDSP) [160], entre otros.

La elección de las metodologías se fundamenta en la pericia y preferencia de los expertos consultados para la validación de los casos reales utilizados en el TFM, que sirven para corroborar la utilidad del modelo propuesto. A continuación se exponen las metodologías escogidas, las cuales son CRISP-DM, Catalys y MoProPEI.

2.2.1. CRISP-DM

Las siglas CRISP-DM deriva de CRoss-Industry Standard Process for Data Mining [149], que significa proceso estándar de la industria para la minería de datos [156], [161].
La metodología reside en un conjunto de tareas descritas en cuatro niveles de abstracción, que van desde lo general a lo particular; organizados de forma jerárquica en fases, tareas generales, tareas especializadas e instancias de procesos [145], [149], [154].
CRISP-DM posee un ciclo de vida organizado en seis fases, que interactúan entre ellas de forma iterativa durante el desarrollo del mismo. Dichas fases son: Comprensión del negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación e Implementación [154], [156].
La metodología suministra un modelo de referencia y guía para el usuario. El primero instaura un conjunto de tareas y actividades para cada fase, junto con sus salidas. Es decir, describe qué debería hacerse en un proyecto de este tipo, pero no especifica cómo llevarlas a cabo. Por otro lado, la guía de usuario ofrece sugerencias prácticas para la ejecución de cada tarea del modelo de referencia [145], [156].
CRISP-DM es considerada como una metodología lo bastante vasta y flexible para ser aplicada en proyectos de diversos tamaños [156], [161].

2.2.2. Catalys

La metodología Catalyst es conocida como P^3TQ (Product, Place, Price, Time, Quantity, en español Producto, Lugar, Precio, Tiempo y Cantidad) y está conformada por dos modelos: el Modelo de Negocio (MN) y el Modelo de Explotación de Información (MEI) o DM [150].
El MN provee una serie de pasos que sirven de guía para desarrollar y construir un modelo que permita identificar un problema u oportunidad en el negocio. Y el MEI brinda una guía pasos para la ejecución de los mismos, de acuerdo al modelo identificado anteriormente (MN) [154].
Cada modelo está compuesto por: (a) Action Boxes o cajas de actividades, que señalan una serie de pasos a desarrollar, (b) Discovery Boxes o cajas de descubrimientos, que proveen acciones de descubrimiento o exploración que incluye resultados asociados, interpretaciones y posibles problemas, necesarios para poder decidir qué hacer en el próximo paso, (c) Technique Boxes o cajas de técnicas, que otorgan información adicional sobre los pasos

recomendados en Action Boxes o Discovery Boxes y (d) Examples Boxes o cajas de ejemplos, que brindan una descripción detallada de cómo usar una técnica determinada. Estas cajas o boxes se requieren dependiendo de las situaciones que se van dando a medida que se avanza en el proyecto, sin que sea de forma secuencial [145], [154].

2.2.3. MoProPEI

La metodología MoProPEI (Modelo de Procesos para Proyectos de Explotación de Información) cuenta con una estructura jerárquica dividido en cuatro niveles: Subprocesos, Fases, Actividades y Tareas; cada uno de los cuales presenta un mayor nivel de especificidad [152].

Los dos subprocesos principales son la Gestión, enfocado la administración y control del proyecto; y el Desarrollo, encauzado a las actividades técnicas. El primer subproceso (Gestión) consta de cinco fases: Iniciación del Proyecto, Planificación del Proyecto, Soporte, Gestión del Control y Calidad, y Gestión de la Entrega. Y el segundo, está compuesto por seis fases: Entendimiento del Dominio, Entendimiento de los Datos, Modelado, Preparación de los Datos, Implementación, y Evaluación y Presentación. La ejecución dichos subprocesos se realizan en paralelo, siendo el primero soporte para las actividades principales del segundo [152].

2.3. Educción del Conocimiento

La Educción o Adquisición de Conocimientos (EC/AC) es el proceso por el cual se recolecta información proveniente de alguna fuente. Este conocimiento debe ser correcto, adecuado y en la cantidad apropiada, dentro un tiempo definido. Es por ello que se debe determinar qué información se necesita, en qué profundidad del tema en particular y qué técnica emplear para adquirir el conocimiento, siendo actividades primordiales de la Ingeniería del Conocimiento (IC) [162].

Se puede resumir AC en tres pasos: Primero, realizar las reuniones iniciales con el propósito de obtener conocimiento general del domino. Segundo, estudiar de documentación del mismo. Y en tercer lugar, el ciclo de educción, donde se interactúa con el experto, en busca de conocimiento más específico [162].

Existen dos tipos de métodos para descubrir lo que saben los expertos; los directos o indirectos, los primeros utilizan preguntas puntuales y directas, en cambio los segundos, buscan colaborar con el experto a ser específicos respecto a sus conocimientos o procesos mentales, que muchas veces no los tiene del todo claros. Dentro de los directos, están las entrevistas y los cuestionarios, y dentro de los indirectos están las técnicas de observación, clasificación de conceptos, la teoría de la construcción personal (emparrillado), entre otras [162]. De las mencionadas, se utilizó la última mencionada, descripta a continuación.

2.3.1. Teoría de la Construcción Personal

La Teoría de la Construcción Personal (TCP) fue propuesta por Kelly [163]. Consiste en una técnica de educción del conocimiento, cuyo propósito consiste en tomar conciencia de las inconsistencias en las propias escalas de valores, dado que cada persona tiene su propia visión del mundo [162], [164].

La fortaleza de TCP radica en que permite modelar la visión interna de cada persona respecto al mundo, sin necesidad de establecer explícitamente lo que es esa visión. Para Kelly cada persona ve el mundo de una manera diferente y cree que esas diferencias pueden expresarse como de representaciones personales. [162], [164], [165].

Para un dominio particular se deben incluir todos los aspectos que el experto considere importante, representados como elementos y su descomposición en características bipolares; los cuales son evaluados [162], [164].

Los datos originados se colocan a una matriz bidimensional o parrilla, donde las columnas representan los elementos y las filas las características. Mediante una escala definida de 1 a N, se completa la matriz; los valores bipolares se ubican en cada extremo de las filas, donde el valor 1 es el opuesto de N. El experto asigna el valor correspondiente en cada celda de la matriz, teniendo en cuenta la escala definida y la intersección elemento y característica que se evalúa [162].

Se considera como un test de clasificación, dividido en cinco etapas [162], [164]:

Donde el primer paso consiste en identificar los elementos representativos dentro del dominio a evaluar. En el segundo paso se identifican las características, como una cualidad atribuida a los elementos inidentificados en el paso anterior. El paso tres, radica en el diseño de la parrilla (pueden ser dicotómicas, clasificatorias o evaluativas), una matriz bidimensional, que vincula los elementos y características identificados. El cuarto paso consiste en la formalización,

donde se clasifican los elementos y las característica por separado. Y por último paso (quinto), se interpreta y analiza los resultados [162].

III

Problema

En este capítulo se reseña el problema tratado en este TFM, que llevó al desarrollo de la misma, mediante la propuesta de un modelo teórico (sección 3.1).

3.1. Descripción del problema

En la actualidad bajo el entorno dinámico en que se encuentra inmersa la sociedad moderna, gracias a las Tecnologías de la Información y las Comunicaciones (TIC), se genera de manera interrumpida un volumen contundente de datos, proveniente de diversas fuentes.

Las organizaciones no son ajenas a esta cuestión, el acceso a información de calidad, confiable y en un tiempo oportuno juega un rol decisivo para la toma de decisiones estratégicas. Exigiendo que posean la capacidad de ser adaptativas y sepan tomar acciones a partir del conocimiento proveniente de los datos. Precisamente los avances de las TIC llevaron a que los sistemas de almacenamiento acrecienten sus capacidades, usufructuando el bajo costo que tiene almacenar los datos del funcionamiento de los procesos o sub sistemas de una organización.

Lo mencionado ha llevado a que áreas como la Ciencia de Datos o Data Science (DS) se encuentre en auge y sea de gran envergadura su aplicación en diversos campos. Una de las principales actividades, es la selección de la metodología más adecuada para guiar un proyecto de dicha índole; aquí la experticia es fundamental, pero muchas veces esta tarea de comparación y elección puede resultar dificultosa, especialmente para los novatos en dicha área. Si bien, se encuentran disponibles algunos estudios comparativos sobre las metodologías existentes (citados en la Sección 2.2) estos no contemplan muchos aspectos que se debieran tener en cuenta; también en ocasiones no se encuentran lo suficientemente validados con casos reales.

Todo problema que requiere tomar decisiones, conlleva a contrastar y evaluar un conjunto de criterios, que permita seleccionar de entre las alternativas disponibles, la más adecuada. Básicamente existen tres tipos de juicios que expresan la importancia para los expertos, estos son la preferencia, la probabilidad y la utilidad, todos ellos basados en la experiencia y/o en la relación costo/beneficios.

Dentro de la Toma de Decisiones con Criterios Múltiples (MCDM), como se mencionó en el capítulo anterior, uno de los métodos más populares es el Proceso Analítico Jerárquico (AHP), el cual descompone el problema en diferentes criterios agrupados bajo una estructura jerárquica, que permite medir de forma precisa y resumir los resultados a partir de los

mismos. En ocasiones, existen ciertas dificultades para establecer un valor exacto, vinculado a la propia incertidumbre de los juicios humanos, una alternativa a ello consiste en definir un rango de valores posibles, esto llevó a que AHP se incorpore las bondades de la teoría de conjuntos difusos, creando así el Proceso Analítico Jerárquico Difuso (FAHP).

Por lo expuesto, el presente TFM busca establecer si mediante la generación de un modelo se podría brindar una base sólida y guiada para comparar de forma válida la elección de metodologías de DS; para lo cual el modelo combina los dos métodos de MCDM mencionados (AHP y FAHP), así como también a la técnica de la Teoría Construcción Personal (TCP) y el uso de etiquetas lingüísticas; estas últimas con el propósito de fortalecer la definición y valoración de criterios en los primeros pasos de los métodos de MCDM.

IV

Solución

El presente capítulo especifica la solución propuesta (sección 4.1), para lo cual se detallan los componentes y el funcionamiento del modelo propuesto, definido en una serie de pasos y ejemplos para su mayor comprensión.

4.1. Solución propuesta

Con el propósito de establecer una base sólida para la selección de metodologías que guíen los proyectos de DS, se propuso un modelo basado en MCDM, a través de los métodos AHP y su variante combinado con lógica difusa (FAHP), junto a la técnica de la Teoría de la Construcción Personal (TCP).

Para la construcción del modelo, se definieron una serie de pasos resumidos en el diagrama de la Figura 4.1; los cuales consisten primeramente en la definición de los criterios y sub-criterios involucrados en proyectos de DS, utilizando para ello la técnica de TCP (paso 1). Seguidamente se construye la estructura jerárquica (paso 2). Para poder valorar los criterios y sub-criterios, dicha actividad se encuentra en color gris (paso 3), dado que involucra dos sub-pasos relacionados a la forma que son valorados (descriptos en el detalle de dicho paso). A partir del paso anterior, se construye y completan las matrices de comparación pareada, estableciendo para ello las prioridades entre los criterios (y sub-criterios) por niveles jerárquicos (paso 4). Es indispensable evaluar si los juicios (valoraciones) emitidas por el o los experto/s poseen una relación de consistencia (paso 5), si esto es así se prosigue con los siguientes pasos, de lo contrario se deberán volver a los pasos 3 y 4, nuevamente evaluando la razón de consistencia para las matrices modificadas. El Desarrollo de los métodos de MCDM (paso 6), se encuentra en color gris debido a que involucra una serie de actividades (pasos) adicionales propio de los métodos AHP y FAHP, que son detallados al abordar dicho paso. A partir de los resultados obtenidos por AHP y FAHP, se integran las ponderaciones obtenidas (paso 7) para presentar una solución global, que luego es ordenada de mayor a menor, como un ranking de metodologías de DS en el paso 8.

Figura 4.1: Esquema general del modelo propuesto

Paso 1. Definición de los criterios con TCP

Como primer paso se identifican los criterios y sub-criterios a evaluar en cada metodología. Éstos se fijan a partir del uso de la técnica de TCP trabajada con expertos del área, las características de las metodologías y a estudios previos.

Paso 2. Construcción de la estructura jerárquica

Como continuación del paso anterior, con la técnica de la TCP se definen los sub-criterios involucrados en las metodologías de DS, que se van agrupando por criterios en diferentes niveles, conformando así la estructura jerárquica del modelo propuesto, visible en la Figura 4.2; donde en el nivel 1 se encuentra el objetivo o meta del mismo, que consiste en la Selección de la Metodología de Data Science más robusta de las alternativas comparadas (mencionadas en la seccipon 2.2).

En el segundo nivel se ubican los criterios principales dentro de los proyectos de estas características, que son el Entendimiento de los Datos y la Comprensión del Negocio.

Seguidamente, en el nivel 3, para el primer criterio (Entendimiento de los Datos) se descompone en dos sub-criterios, que son el acceso a los datos y uso de los mismos; los cuales se especifican con cinco sub-criterios cada uno; para el Acceso de los Datos los sub-criterios son la Portabilidad, la Accesibilidad, la Diversidad, la Diversidad de las fuentes de datos y los Recursos necesarios; y para el sub-criterio Uso de los Datos, la Calidad, la Completitud, la Funcionalidad, los Costos de transformación y los Datos riesgosos.

En la Comprensión del Negocio, existen tres sub-criterios (nivel 3), el Entorno del Negocio, el Proyecto del Entorno del Negocio y Data Science en el Proyecto del Negocio; quienes a su vez se subdividen en siete, cinco y tres sub-criterios respectivamente en el cuarto nivel. Los sub-criterios para el Entorno del Negocio son la Situación económica circundante, la Comprensión general de la organización, la Definición del problema del negocio, la Definición de las oportunidades del negocio, la Definición de requisitos del negocio y la Definición de los criterios de éxito del proyecto del negocio. El Conocimiento de los RRHH del negocio, la Disponibilidad de los RRHH del negocio, los Riesgos del problema del negocio (respecto al proyecto), las Contingencias en relación a los riesgos del problema del negocio y la Viabilidad del proyecto son los sub-criterios del sub-criterio Proyecto del Entorno del Negocio. Por otra parte, los tres sub-criterios de Data Science en el Proyecto del Negocio son los Objetivos de la aplicación de la metodología de DS, el Conocimiento de los RRHH del proyecto de DS y los RRHH disponibles para el proyecto de DS.

Los sub-criterios del último nivel (cuatro) se comparan en función cada metodología, es decir, MoProPEI, CRISP-DM y P^3TQ, siendo estas las alternativas a evaluar; como se aprecia en el quinto nivel de le Figura 4.2.

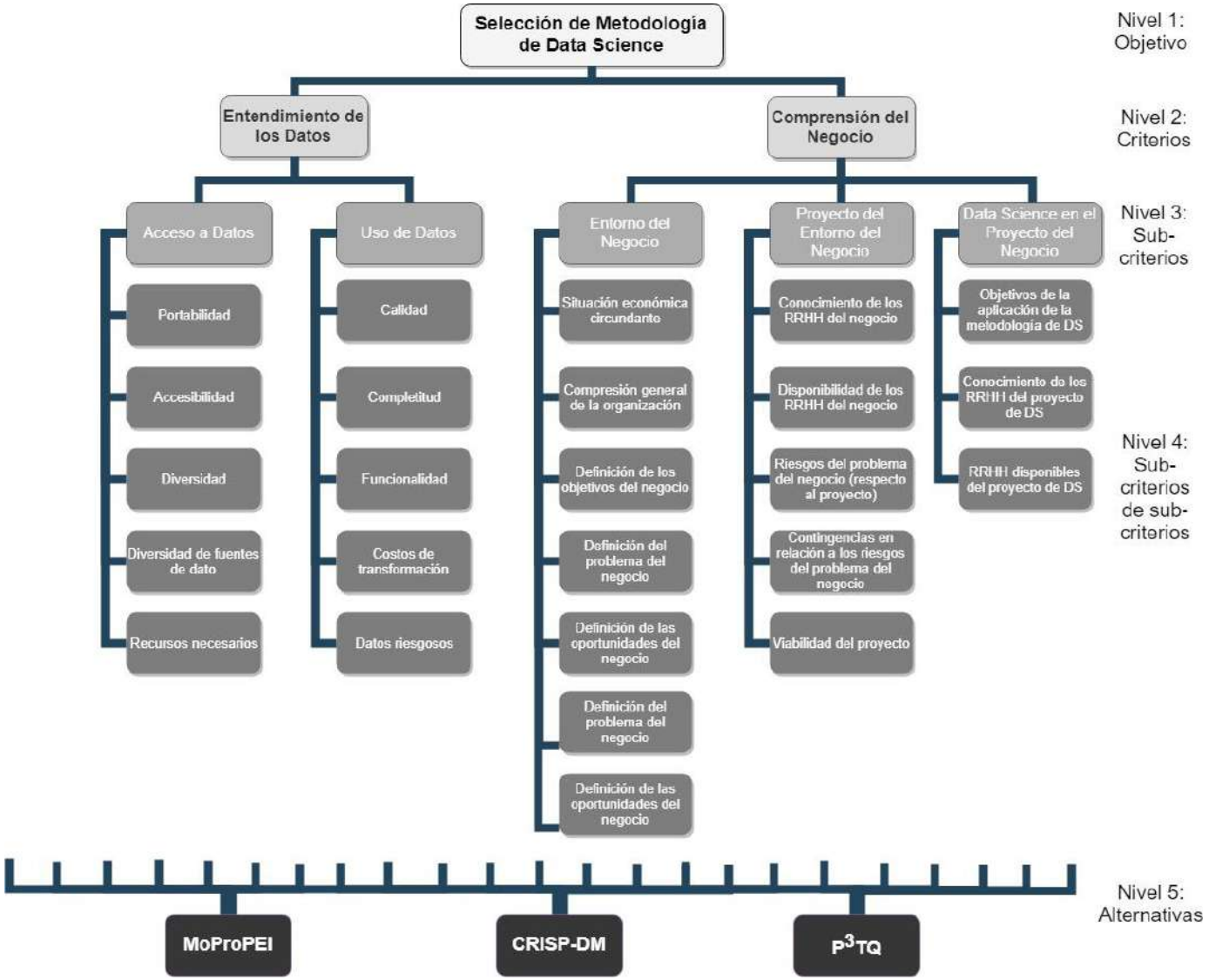

Figura 4.2: Estructura jerárquica del modelo propuesto

Paso 3. Valoración por niveles

Como se aprecia en la Figura 4.1 la valoración se encuentra en color gris, dado que se divide en dos sub-pasos, por un lado se confeccionan unas matrices de dos entradas para los criterios y sub-criterios (Paso 3.1), y por otro, se utiliza la técnica de TCP para indicar la completitud de los sub-criterios del nivel 4 en relación a cada metodología (Paso 3.2); esto se visualiza en la Figura 4.3.

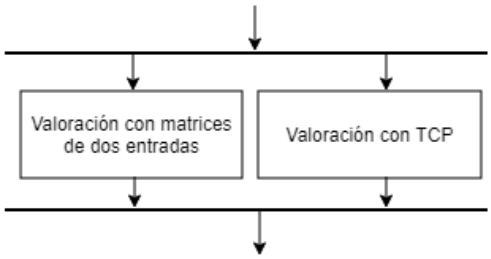

Figura 4.3: Valoración por niveles

Paso 3.1. Valoración con matrices de dos entradas

Se utilizan matrices de dos entradas para los criterios del nivel 2 y sub-criterios del nivel 3 y 4, con el propósito de facilitar las elecciones de los expertos; donde se indica con una *X* el valor correspondiente de preferencia de un criterio respecto a otro, basado en la escala fundamental de Saaty; dichas matrices se encuentran disponibles en el Anexo III.

Tabla 4.1: Comparación de criterios por pares

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
Entendimiento de los Datos					*Comprensión del Negocio*			
				X				

La Tabla 4.1 expone un ejemplo de dichas matrices, donde se evalúan los criterios principales, es decir, el Entendimiento de los Datos y la Comprensión del Negocio, con la misma importancia asignada.

Paso 3.2. Valoración con TCP

Con TCP se establecen las etiquetas lingüísticas, definidas en una escala del 1 al 9, donde el valor 1 indica que no se analiza el sub-criterio; del 2 al 9, representa los valores del intervalo expresados en porcentajes, refiriendo al cumplimiento de ese aspecto dentro del caso de estudio. Dependiendo si el sub-criterio es un aspecto positivo o negativo, los valores de la escala definida entre 2 y 9 pueden estar invertidos. A manera de ejemplo en la Tabla 4.2 se exponen las etiquetas lingüísticas de dos sub-criterios, el primero Accesibilidad, que es un aspecto positivo, por ende los valores (porcentajes) del intervalo se encuentra de manera ascendente; caso contrario es el sub-criterio Diversidad (aspecto negativo), donde los porcentajes expresados en la escala se hallan de forma descendente; el detalle las etiquetas utilizadas para todos los sub-criterios se encuentran en el Anexo IV.

Tabla 4.2: Etiquetas lingüísticas para sub-criterios

Accesibilidad (nivel de accesibilidad/ asequibilidad de los datos)		**Diversidad** (diferentes tipos de datos)	
Escala definida	*Descripción*	*Escala definida*	*Descripción*
1	No se analiza	1	No se analiza
2	1% a 13% de accesibilidad	2	98% a 100% de diversidad
3	14% a 27% de accesibilidad	3	84% a 97% de diversidad
4	28% a 41% de accesibilidad	4	70% a 83% de diversidad
5	42% a 55% de accesibilidad	5	56% a 69% de diversidad
6	56% a 69% de accesibilidad	6	42% a 55% de diversidad
7	70% a 83% de accesibilidad	7	28% a 41% de diversidad
8	84% a 97% de accesibilidad	8	14% a 27% de diversidad
9	98% a 100% de accesibilidad	9	1% a 13% de diversidad

Luego de establecer las definiciones anteriores, se procede a confeccionar una matriz del tipo parrilla, para la cual se fijan los valores bipolares o extremos definidos en la escala anteriormente mencionada; el peor y mejor caso, que representan el valor 1 y 9 respectivamente (ver Anexo IV).

El experto completa la matriz según el valor que corresponda para el caso de estudio, teniendo en cuenta la escala mencionada. A manera de ejemplo, se

exponen algunos sub-criterios con los valores bipolares en cada extremo y la valoración asignada para cada metodología en la Tabla 4.3.

Tabla 4.3: Parrilla para valoración de los sub-criterios del cuarto nivel

	P^{3}TQ	**CRISP-DM**	**MoPro-PEI**	
No se analiza la portabilidad de los datos	5	9	8	De 98% a 100% de portabilidad de los datos
No se analiza la accesibilidad de los datos	5	9	7	De 98% a 100% de accesibilidad de los datos
No se analiza la diversidad de los datos	4	2	3	De 0% a 13% de diversidad de los datos
…	…	…	…	….

Paso 4. Completitud de las matrices y establecimiento de prioridades entre criterios

A partir de las matrices generadas en el paso anterior en los diferentes niveles, se completan las correspondientes según lo planteado por Saaty (ver Anexo I), definiendo su importancia por grupos de criterios o sub-criterios según la jerarquía definida.

En el caso de las matrices de dos entradas generadas para los criterios del nivel 2 y sub-criterios del nivel 3, se efectúan la transferencia directa de las valoraciones de los expertos a los valores correspondientes de las nuevas matrices de comparación pareada.

A partir de las parrillas obtenidas para los sub-criterios del nivel 4, se completan las matrices pareadas correspondiente; tomando para ello la diferencia en valor absoluto entre las valoraciones de a pares, más uno (por ejemplo si dos sub-criterios tienen valoración 5: 5-5=0+1, los dos sub-criterios tienen la misma preferencia o importancia; otro ejemplo, si la valoración para uno fue de 7 y para el otro de 5: 7-5=2+1, el primer sub-criterio tiene una preferencia de 3 por sobre el segundo, es decir se la considera moderadamente más importante); con el propósito de adecuar los valores definidos con las etiquetas lingüísticas a la escala de Saaty, ubicándolos en el lugar que corresponda dentro de la matriz.

En ambos casos dependiendo de las valoraciones fijadas por el experto, se ubican en la parte principal o secundaria de cada matriz pareada, completándola con los valores recíprocos; para luego normalizar dichas matrices y definir las prioridades para cada una de ellas.

En la Tabla 4.4 se puede apreciar la continuidad de lo planteado en la Tabla 4.1, donde se ubicaron los valores asignados anteriormente en las filas y columnas 2 y 3, a partir de las cuales se obtuvo la suma de las columnas para obtener la matriz normalizada y sus respectivas prioridades del nivel que en este ejemplo sencillo es del 50% para cada criterio.

Tabla 4.4: Matriz de comparación pareada

	Entendimiento de los Datos	**Comprensión del Negocio**	**Matriz Normalizada**		**Prioridades**
Entendimiento de los Datos	1	1	0,50	0,50	0,50
Comprensión del Negocio	1	1	0,50	0,50	0,50
Suma columna	2	2	1	1	1

Paso 5. Evaluación de Consistencia

Para evaluar la consistencia en los juicios emitidos por el experto, se analiza la consistencia de cada una de las matrices obtenidas, según la razón o relación de consistencia (CR, Consistency Ratio) establecido por Saaty (ver Anexo I). En los casos que la matriz sea inconsistente, se vuelven a realizar los pasos 3, 4 y el presente (5).

Para el ejemplo planteado en la Tabla 4.4, al ser de dos criterios no tendrá problemas de consistencias; a partir de los 3 criterios en adelante pueden existir conflictos de consistencias, por lo que deben ser evaluados.

Paso 6. Desarrollo métodos de MCDM

Siguiendo el esquema general de la Figura 4.1, se puede visualizar que el desarrollo de los métodos de MCDM, se encuentra en color gris, dado que se realizan una serie de actividades llevadas a cabo en paralelo; por un lado se completa el desarrollo del método AHP (Paso 6.1) y por otro de FAHP (Paso 6.2), tal como se aprecia en la

Figura 4.4, donde a la izquierda se encuentran los pasos del método AHP (restantes) y a la derecha los de FAHP.

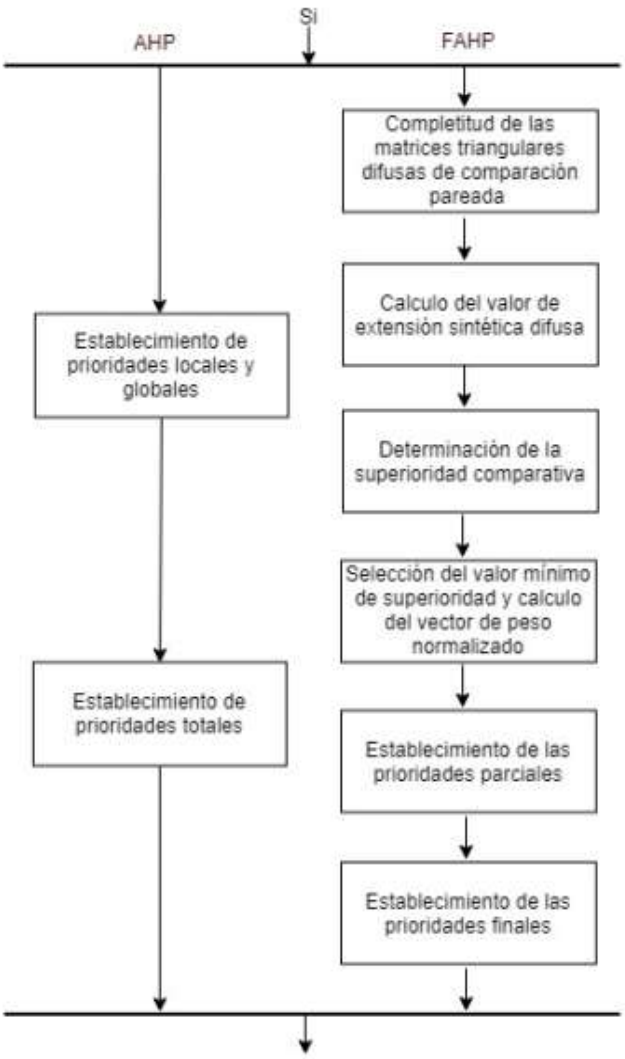

Figura 4.4: Desarrollo de métodos MCDM

Paso 6.1. Proceso Analítico Jerárquico

Siguiendo los pasos del método AHP explicados en Anexo I, se establecen las prioridades locales y globales, mediante el cálculo del peso relativo a los criterios para cada nivel, según el Paso 3 del Anexo I; a partir de las cuales, se obtienen las prioridades totales asociadas a cada alternativa, como se indica en el Paso 4 del mismo anexo; mediante el método de la suma ponderada.

A modo de ejemplo se esquematiza en la Tabla 4.5 las prioridades para el sub-criterio Entorno del Negocio, teniendo en cuenta sus siete sub-criterios y en la

Tabla 4.6 las prioridades totales asociadas a cada alternativa, es decir, la ponderación final obtenida para cada una de las metodologías de DS evaluadas. Dichas tablas se encuentran sin valores asignados dado que éstos estarán sujetos a cada caso de validación en particular (los mismos serán expuestos en el siguiente capítulo).

Tabla 4.5: Priorización del sub-criterio Entorno del Negocio AHP

	P^{3}TQ	**CRISP-DM**	**MoProPEI**	**Pesos (Paso 4)**
Situación económica circundante	0,00	0,00	0,00	0,00
Comprensión general de la organización	0,00	0,00	0,00	0,00
Definición de los objetivos del negocio	0,00	0,00	0,00	0,00
Definición del problema del negocio	0,00	0,00	0,00	0,00
Definición de las oportunidades del negocio	0,00	0,00	0,00	0,00
Definición de requisitos del negocio	0,00	0,00	0,00	0,00
Definición de criterios de éxito del proyecto	0,00	0,00	0,00	0,00
Prioridades Parciales	**0,00**	**0,00**	**0,00**	

Tabla 4.6: Prioridades totales de las alternativas con AHP

	P^{3}TQ	**CRISP-DM**	**MoProPEI**
Entendimiento de los Datos	0,	0,	0,
Comprensión del Negocio	0,	0,	0,
Prioridades Totales	**0,**	**0,**	**0,**

Paso 6.2. Proceso Analítico Jerárquico Difuso

Los pasos necesarios para implementar el método FAHP están descriptos detalladamente en el Anexo II. Para el presente TFM se utilizó la función de

membrecía con números difusos triangulares, para el establecimiento de prioridades entre los criterios y sub-criterio; con el propósito de trasladar las ideas de los expertos en una matriz de reciprocidad positiva; detallada en el Paso 2 del Anexo II. A partir de las métricas de comparación difusa desarrolladas, se establecen las prioridades entre criterios. A modo de ejemplo, en la Tabla 4.7, se exponen los criterios principales de la estructura jerárquica (ver Figura 4.2), donde, como se indicó en el Anexo II: l representan el valor pesimista, m el valor más prometedor y u el valor optimista para cada conjunto difuso. Para este TFM el valor de m proviene de las valoraciones hechas inicialmente en el Paso 3 del modelo; es decir que coinciden con los utilizados con el método AHP.

Tabla 4.7: Matriz triangular difusa de comparación pareada

	Entendimiento de los Datos			**Comprensión del Negocio**		
	l	**m**	**u**	**l**	**m**	**u**
Entendimiento de los Datos	1	1	1	1	1	2
Comprensión del Negocio	$1/2$	$1/1$	$1/1$	1	1	1
Suma columna	1,50	2,00	2,00	2,00	2,00	3,00

Tabla 4.8: Cálculo del valor de extensión sintética difusa

	Ecuación II.4			**Ecuación II.5**			**Ecuación II.3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
Entendimiento de los Datos	2,00	2,00	3,00	3,50	4,00	5,00	0,40	0,50	0,86
Comprensión del Negocio	1,50	2,00	2,00				0,30	0,50	0,57

Luego se calcula el valor de extensión sintética difusa para cada criterio (o sub-criterio), definida en el Paso 3 del Anexo II. Como se aprecia en la Tabla 4.8, dicho valor se obtiene y se representa mediante la Ecuación II.III, junto a las dos ecuaciones (II.IV y II.V) previas necesarias para obtener el valor

mencionado (definido en el Anexo II); el ejemplo expuesto surge de la comparación exhibida en la Tabla 4.7.

Posteriormente se debe determinar la superioridad comparativa de los criterios evaluados, como se indica en el Paso 4 del Anexo II. Para el ejemplo planteado anteriormente (Tablas 4.7 y 4.8) se obtienen los valores expuestos en la Tabla 4.9.

Tabla 4.9: Superioridad comparativa

V(M2>=M1)	
V(C1>=C2)	1,00
V(C2>=C1)	1,00

Paso siguiente, se debe seleccionar el valor mínimo de superioridad; calcular el vector de pesos y normalizarlo para cada criterio, así como se explica en el Paso 5 del Anexo II; lo mencionado se repite en todos los sub-criterios de la jerarquía.

Para el ejemplo de la Tabla 4.9, el mínimo valor es 1. El cálculo del vector de pesos (W´) y la normalización del mismo (W) se muestra en la Tabla 4.10.

Tabla 4.10: Vector de pesos

Vector de pesos	$W' = (1{,}00;1{,}00)^T$	$\sum W' = 2{,}00$
Vector de pesos normalizado	$W = (0{,}5;0{,}50)^T$	

Seguidamente, se establecen las prioridades parciales en los diferentes niveles de la jerarquía y luego de cada alternativa, como se indica en el Paso 6 del Anexo II, donde se calculan los valores normalizados de los criterios con los valores de multiplicación de las alternativas y los sub-criterios estimados.

Tomando la Tabla 4.10, las ponderaciones parciales de ese nivel de forma aislada se representa en la Tabla 4.11, donde se aprecia que ambos criterios tienen una ponderación del 50%.

Tabla 4.11: Ponderaciones parciales

	Pesos
Entendimiento de los Datos	0,50
Comprensión del Negocio	0,50

Como ejemplo la Tabla 4.12 señala las prioridades parciales para el sub-criterio Entorno del Negocio, junto a sus siete sub-criterios; al igual que en el caso de la Tabla 4.5, la misma se encuentra sin cifras asociadas dado que dependerá de cada caso de validación.

Tabla 4.12: Prioridades parciales del sub-criterio Entorno del Negocio FAHP

	P^3TQ	**CRISP-DM**	**MoProPEI**	**Pesos (Paso 5 Anexo II)**
Situación económica circundante	0,00	0,00	0,00	0,00
Comprensión general de la organización	0,00	0,00	0,00	0,00
Definición de los objetivos del negocio	0,00	0,00	0,00	0,00
Definición del problema del negocio	0,00	0,00	0,00	0,00
Definición de las oportunidades del negocio	0,00	0,00	0,00	0,00
Definición de requisitos del negocio	0,00	0,00	0,00	0,00
Definición de criterios de éxito del proyecto	0,00	0,00	0,00	0,00
Prioridades Parciales	**0,00**	**0,00**	**0,00**	

Finalmente, a partir de las prioridades parciales obtenidas de los sub-criterios (por niveles) se alcanza el peso de prioridad final para cada una de las metodologías de DS evaluadas, multiplicándolas con los pesos de los criterios principales (Tabla 4.11), como se establece en el Paso 7 del Anexo II. La Tabla

4.13, esboza las prioridades totales asociadas a cada alternativa (los valores dependerán nuevamente del caso de validación).

Tabla 4.13: Prioridades finales de las alternativas con FAHP

	P^{3}TQ	**CRISP-DM**	**MoProPEI**
Entendimiento de los Datos	0,00	0,00	0,00
Comprensión del Negocio	0,00	0,00	0,00
Prioridades Finales	**0,00**	**0,00**	**0,00**

Paso 7. Integración de las ponderaciones obtenidas por los métodos AHP y FAHP

Una vez obtenidas las ponderaciones finales para cada metodología tanto con AHP, como con FAHP, se presenta de forma resumida la ponderación obtenida para cada metodología, aplicando para ello el promedio de las ponderaciones obtenidas para cada una de ellas. A modo de ejemplo, se representaría como lo expuesto en la Tabla 4.14.

Tabla 4.14: Prioridades finales de las alternativas

	AHP	**FAHP**	**Promedio Ponderado**
P^{3}TQ	0,00	0,00	**0,00**
CRISP-DM	0,00	0,00	**0,00**
MoProPEI	0,00	0,00	**0,00**

Paso 8. Ranking de metodologías de DS

Las ponderaciones finales obtenidas en el paso anterior son ordenadas de mayor a menor, generando un ranking de las metodologías evaluadas.

V

Casos de validación

En este capítulo se detallan las generalidades de validación utilizadas para el modelo basado en MCDM para la selección de metodologías de DS, tratado en la sección 5.1; para luego detallar los dos casos de validación utilizados sobre dominios diferentes, donde se comparan las tres metodologías evaluadas según el caso, por un lado en la sección 5.2 se aborda los causales de rotura de automóviles 0KM durante el transporte a las concesionarias y en la sección 5.3, de la deserción universitaria. Finalmente en la sección 5.4, se realiza un análisis general de los resultados obtenidos a partir de los casos expuestos.

5.1. Generalidades de los casos de validación

Los dos casos de validación empleados tienen como propósito determinar los causales en proyectos verídicos de DS: El primer Caso de Validación (CV1) tuvo como propósito definir los motivos por los cuales se producen roturas de las autopartes de automóviles 0KM (como ser faros, paragolpes, parabrisas, etc.), mientras son transportados mediante los conocidos camiones mosquitos o plataformas porta automóviles, desde la fábrica hasta las diferentes concesionarias del país (por cuestiones de confidencialidad se omite en nombre de la fábrica). Por otra parte, en el segundo caso (CV2) se buscó identificar cuáles son las razones de la deserción universitaria en la carrera de Licenciatura en Sistemas de la Universidad Nacional de Río Negro, para el período comprendido entre 2009 y 2015. La completitud, análisis y corroboración de los casos, fueron realizados por los expertos colaboradores: la Dra. Britos y la Dra. Pollo Cattaneo; cabe aclarar, que las mismas tienen amplia experiencia en numerosos proyectos de DS, ofrecen consultorías al respecto y poseen el grado académico de doctorado.

Como se reveló en el Capítulo IV, primeramente se identificaron los criterios y sub-criterios a evaluar en cada metodología de DS (Paso 1). Los cuales se fueron constituyendo a partir de la técnica de la TCP trabajada con expertos del área, las cualidades de cada metodología y estudios anteriores.

Seguidamente, se construyó la estructura jerárquica del modelo de MCDM (Paso 2); utilizando para ello nuevamente la técnica de la TCP, con la cual se precisaron los sub-criterios comprendidos en las metodologías de DS, agrupándolos por sub-criterios y criterios en diferentes niveles, constituyendo así la estructura jerárquica del modelo propuesto indicada en la Figura 4.2.

A partir de la estructura jerárquica definida, se valoraron los criterios y sub-criterios (Paso 3). Se diseñaron matrices de dos entradas (Paso 3.1), cabe aclarar que los criterios y sub-criterios

fueron evaluados sin tener en cuenta el caso en particular, en cambio los sub-criterios del último nivel en relación a cada metodología (alternativa) son específicos de cada caso, para estos últimos se utilizó la técnica de la TCP para indicar el grado de cumplimiento de dichos sub-criterios, minimizando así problemas de ambigüedad y subjetividad, gracias a la escala brindada y a la evaluación individual de cada uno de ellos (Paso 3.2).

Teniendo en cuenta lo mencionado, en la siguiente sub-sección se exponen las matrices de dos entradas confeccionadas, que no dependen puntualmente del caso de validación en cuestión. Y luego para cada caso de validación se exhiben las etiquetas lingüísticas y parrillas confeccionadas para el nivel 4 de la jerarquía en relación a las tres metodologías de DS y posteriormente se precede a desarrollar los pasos restantes del modelo para cada caso.

5.1.1. Valoración con matrices de dos entradas

Tal como se indica en el Anexo III, para la completitud de las tablas los expertos marcan con una *X* en las filas en color gris, para indicar la preferencia de los criterios o sub-criterios contrastados, teniendo en cuenta la escala fundamental definida por Saaty (disponible en la Tabla I.I del Anexo I). La matriz consta de nueve columnas, la columna central (cinco) indica que los criterios comparados poseen la misma importancia, desde esa columna hacia la izquierda va aumentando la preferencia del criterio situado en dicha parte de la matriz, lo mismo ocurre con el que se localiza a la derecha, solo que en este caso aumenta la importancia hacia la derecha.

Tabla 5.1: Matriz de dos entradas para criterios principales (nivel 2)

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
Entendimiento de los Datos					Comprensión del Negocio			
				X				

Primeramente en la Tabla 5.1 se compararon los criterios principales, que son el Entendimiento de los Datos y la Compresión del Negocio, cuyos criterios poseen la misma importancia dentro de un proyecto de DS, según los expertos.

Tabla 5.2: Matriz de dos entradas para sub-criterios Acceso y Uso de los Datos

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
Acceso a los Datos					Uso de los Datos			
						X		

Siguiendo la jerarquía propuesta en la Figura 4.2, en la rama izquierda, el Entendimiento de los Datos se divide en dos sub-criterios, que son el Acceso a los Datos y el Uso de los Datos, los cuales son valorados en la Tabla 5.2; donde los expertos manifestaron fuerte preferencia en cuanto al uso, por sobre el acceso a los mismos.

Las características o sub-criterios definidos para el Acceso a los Datos, comparados a continuación en las matrices de dos entradas, son:

1. Portabilidad (capacidad de ser transferido).
2. Accesibilidad (nivel de accesibilidad/asequibilidad de los datos).
3. Diversidad (diferentes tipos de datos).
4. Diversidad de fuentes de datos (origen).
5. Recursos necesarios (equipamiento).

Tabla 5.3: Matriz de dos entradas para la Portabilidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C1: Portabilidad					C2: Accesibilidad			
							X	
C1: Portabilidad					C3: Diversidad			
					X			
C1: Portabilidad					C4: Diversidad de fuentes de datos			
						X		
C1: Portabilidad					C5: Recursos necesario			
				X				

Las valoraciones realizadas al sub-criterio Portabilidad respecto los demás sub-criterios, se encuentra en la Tabla 5.3, donde se aprecia que es considerada una característica poco relevante en comparación a las demás, a excepción de los Recursos necesarios que poseen la misma relevancia. La comparación de la Accesibilidad respecto los demás sub-criterios, se expone en la Tabla 5.4, la cual es igual de importante que C3, moderadamente menos que C4 y más que C5. En la Tabla 5.5 se coteja el sub-criterio Diversidad respecto los restantes del mismo nivel y sub-criterio padre, teniendo igual envergadura que C4 y moderadamente más que C5. Y la comparación del sub-criterio Diversidad de las fuentes de datos y los Recursos necesarios, se exhibe en la Tabla 5.6, donde es fuertemente más relevante la primera que la segunda de las características mencionadas.

Tabla 5.4: Matriz de dos entradas para la Accesibilidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C2: Accesibilidad					C3: Diversidad			
				X				
C2: Accesibilidad					C4: Diversidad de fuentes de datos			
					X			
C2: Accesibilidad					C5: Recursos necesario			
			X					

Tabla 5.5: Matriz de dos entradas para la Diversidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C3: Diversidad					C4: Diversidad de fuentes de datos			
				X				
C3: Diversidad					C5: Recursos necesario			
			X					

Tabla 5.6: Matriz de dos entradas para la Diversidad de fuentes de datos y Recursos necesarios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C4: Diversidad de fuentes de datos					C5: Recursos necesario			
		X						

Para el Uso de los Datos, los sub-criterios o características comparadas son las siguientes:

6. Calidad (condiciones de los datos (valores válidos e inexistencia de valores anómalos)).
7. Completitud (grado de completitud de los datos (cantidad de valores faltantes)).
8. Funcionalidad (uso y practicidad de los datos).
9. Costos de transformación (demanda de trabajo para transformarlos/adaptarlos para su uso).
10. Datos riesgosos (datos sensibles, que no deben ser difundidos).

Tabla 5.7: Matriz de dos entradas para la Calidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C6: Calidad					C7: Completitud			
				X				
C6: Calidad					C8: Funcionalidad			
		X						
C6: Calidad					C9: Costos de transformación			
	X							
C6: Calidad					C10: Datos riesgosos			
				X				

En la Tabla 5.7 se confrontan la Calidad (de los datos) con los demás sub-criterios del mismo nivel para el Uso de los Datos, donde se denota su relevancia, en relación a C7 y C10 igualitaria y respecto a C8 y C9, fuerte y muy fuerte. Seguidamente, en la Tabla 5.8 se expone la matriz de dos entradas completadas para el sub-criterio Completitud respecto a los restantes sub-criterios cuya importancia es moderada respecto a C8, fuerte en relación a C9 e igual que C10. La comparación de la característica Funcionalidad (de los datos) en relación a los demás sub-criterios, se muestra en la Tabla 5.9, cuya preferencia es moderada respecto a C9 e igualitaria que C10. Finalmente se comparan en la Tabla 5.10 el Costo de transformación y los Datos riesgosos, siendo este último moderadamente más importante que el primero.

Tabla 5.8: Matriz de dos entradas para la Completitud y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C7: Completitud					C8: Funcionalidad			
			X					
C7: Completitud					C9: Costos de transformación			
		X						
C7: Completitud					C10: Datos riesgosos			
				X				

Tabla 5.9: Matriz de dos entradas para la Funcionalidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C8: Funcionalidad					C9: Costos de transformación			
			X					
C8: Funcionalidad					C10: Datos riesgosos			
				X				

Tabla 5.10: Matriz de dos entradas para el Costo de transformación y Datos riesgosos

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C9: Costos de transformación					C10: Datos riesgosos			
					X			

Una vez finalizada la comparación de los diferentes sub-criterios involucrados en el Entendimiento de los Datos, se prosiguió con la valoración de los sub-criterios dentro de la Comprensión del Negocio.

En las tablas 5.11 y 5.12, se visualizan las matrices de dos entradas de los tres sub-criterios, es decir, Entorno del Negocio, Proyecto del Entorno del Negocio y DS en el Proyecto del Negocio. Se puede apreciar que el último es considerado el más importante de los mencionados.

Tabla 5.11: Matriz de dos entradas para el criterio Entorno del Negocio y demás criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
Entorno del Negocio					Proyecto del Entorno del Negocio			
				X				
Entorno del Negocio					DS en el Proyecto del Negocio			
						X		

Tabla 5.12: Matriz de dos entradas para los criterios Proyecto Entorno del Negocio y DS en el Proyecto del Negocio

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
Proyecto del Entorno del Negocio					DS en el Proyecto del Negocio			
					X			

El Entorno del Negocio, como se indicó anteriormente en paso 2, de la sección 4.1, se divide en siete sub-criterios:

1. Situación económica circundante.
2. Evaluación/comprensión general de la organización.
3. Definición de los objetivos del negocio.
4. Definición del problema del negocio.
5. Definición de las oportunidades del negocio.
6. Definición de requisitos del negocio.
7. Definición de los criterios de éxito del proyecto (desde el negocio).

Tabla 5.13: Matriz de dos entradas para la Situación económica circundante y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C1: Situación económica circundante					C2: Evaluación/comprensión general de la organización			
					X			
C1: Situación económica circundante					C3: Definición de los objetivos del negocio			
					X			
C1: Situación económica circundante					C4: Definición del problema del negocio			
					X			
C1: Situación económica circundante					C5: Definición de las oportunidades del negocio			
					X			
C1: Situación económica circundante					C6: Definición de requisitos del negocio			
						X		
C1: Situación económica circundante					C7: Definición de los criterios de éxito del proyecto (negocio)			
					X			

En primer lugar, se comparó el sub-criterio Situación económica circundante, con los restantes sub-criterios del mismo nivel (y padre); dicha comparación se encuentra en la Tabla 5.13, donde se aprecia que no es considerado un aspecto determinante. Seguidamente, en la Tabla 5.14, se expone las preferencias de los expertos al equiparar la Evaluación o comprensión general de la organización con los otros sub-criterios, donde se ve que poseen la misma importancia. En la Tabla 5.15, al igual que la evaluación anterior (Tabla 5.14), se atribuyen equivalentes relevancia al sub-criterio comparado, en este caso la Definición de los objetivos del negocio respecto a los otros cuatro de dicha tabla. Siguiendo la tendencia de las tablas anteriores, para la Definición del problema del negocio y demás sub-criterios, el experto manifestó que poseen idéntica trascendencia, como se aprecia en la Tabla 5.16.

Tabla 5.14: Matriz de dos entradas para la Evaluación/comprensión general de la organización y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C2: Evaluación/comprensión general de la organización					C3: Definición de los objetivos del negocio			
				X				
C2: Evaluación/comprensión general de la organización					C4: Definición del problema del negocio			
				X				
C2: Evaluación/comprensión general de la organización					C5: Definición de las oportunidades del negocio			
				X				
C2: Evaluación/comprensión general de la organización					C6: Definición de requisitos del negocio			
				X				
C2: Evaluación/comprensión general de la organización					C7: Definición de los criterios de éxito del proyecto (negocio)			
				X				

Tabla 5.15: Matriz de dos entradas para la Definición de los objetivos del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C3: Definición de los objetivos del negocio					C4: Definición del problema del negocio			
				X				
C3: Definición de los objetivos del negocio general de la organización					C5: Definición de las oportunidades del negocio			
				X				
C3: Definición de los objetivos del negocio					C6: Definición de requisitos del negocio			
				X				
C3: Definición de los objetivos del negocio					C7: Definición de los criterios de éxito del proyecto (negocio)			
				X				

Tabla 5.16: Matriz de dos entradas para la Definición del problema del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C4: Definición del problema del negocio					C5: Definición de las oportunidades del negocio			
				X				
C4: Definición del problema del negocio					C6: Definición de requisitos del negocio			
				X				
C4: Definición del problema del negocio					C7: Definición de los criterios de éxito del proyecto (negocio)			
				X				

En la Tabla 5.17 se observa que la Definición de las oportunidades del negocio respecto a la Definición de requisitos del negocio (C6) y a la Definición de los criterios de éxito del proyecto (C7), también poseen igual importancia.

Tabla 5.17: Matriz de dos entradas para la Definición de las oportunidades del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C5: Definición de las oportunidades del negocio					C6: Definición de requisitos del negocio			
				X				
C5: Definición de las oportunidades del negocio					C7: Definición de los criterios de éxito del proyecto (negocio)			
				X				

Al cotejar la Definición de requisitos del negocio y la de los criterios de éxito del proyecto, los expertos indicaron, que el segundo es fuertemente más importante que el primero, como se visualiza en la Tabla 5.18.

Tabla 5.18: Matriz de dos entradas para la Definición de requisitos del negocio y de los criterios de éxito del proyecto

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C6: Definición de requisitos del negocio					C7: Definición de los criterios de éxito del proyecto (negocio)			
						X		

El Proyecto del Entorno del Negocio, se distribuye en 5 sub-criterios, los cuales son:

8. Conocimiento de los RRHH del negocio.
9. Disponibilidad de los RRHH del negocio.
10. Riesgos del problema del negocio (respecto al proyecto).

11. Contingencias en relación a los riesgos del problema del negocio (respecto al proyecto).
12. Viabilidad del proyecto.

La primera característica o sub-criterio que se confronta con los restantes del mismo nivel, es el Conocimiento de los RRHH del negocio. Como se aprecia en la Tabla 5.19, en relación a la característica 9 y 10 posee igual importancia y en relación a C11 y C12, se consideran moderadamente más importantes estos últimos.

Tabla 5.19: Matriz de dos entradas para Conocimiento de los RRHH del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C8: Conocimiento de los RRHH del negocio					C9: Disponibilidad de los RRHH del negocio			
				X				
C8: Conocimiento de los RRHH del negocio					C10: Riesgos del problema del negocio (respecto al proyecto)			
				X				
C8: Conocimiento de los RRHH del negocio					C11: Contingencias en relación a los riesgos del problema del negocio			
					X			
C8: Conocimiento de los RRHH del negocio					C12: Viabilidad del proyecto			
					X			

En la Tabla 5.20 el sub-criterio evaluado es la Disponibilidad de los RRHH del negocio, donde al contrastar con los Riesgos del problema del negocio y las Contingencias en relación a los riesgos del problema del negocio, estos cuentan con una importancia moderada por sobre el primero y en relación a la Viabilidad del Proyecto idéntica.

Tabla 5.20: Matriz de dos entradas para Disponibilidad de los RRHH del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C9: Disponibilidad de los RRHH del negocio					C10: Riesgos del problema del negocio (respecto al proyecto)			
					X			
C9: Disponibilidad de los RRHH del negocio					C11: Contingencias en relación a los riesgos del problema del negocio			
					X			
C9: Disponibilidad de los RRHH del negocio					C12: Viabilidad del proyecto			
				X				

Los expertos consideran que los Riesgos del problema del negocio, respecto a las Contingencias en relación a los riesgos del problema del negocio (C11) y la Viabilidad del Proyecto (C12), disfruta de una importancia equivalente, tal como se muestra en la Tabla 5.21. Al confrontar las Contingencias en relación a los riesgos del problema del negocio y la Viabilidad del proyecto, se indicó que la primera característica es levemente más significativa que la segunda, como se aprecia en la Tabla 5.22.

Tabla 5.21: Matriz de dos entradas para Riesgos del problema del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C10: Riesgos del problema del negocio (respecto al proyecto)					C11: Contingencias en relación a los riesgos del problema del negocio			
				X				
C10: Riesgos del problema del negocio (respecto al proyecto)					C12: Viabilidad del proyecto			
				X				

Tabla 5.22: Matriz de dos entradas para Contingencias en relación a los riesgos del problema del negocio y Viabilidad del proyecto

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C11: Contingencias en relación a los riesgos del problema del negocio					C12: Viabilidad del proyecto			
			X					

Para el sub-criterio Data Science (DS) en el Proyecto del Negocio, se definieron 3 sub-criterios:

13. Objetivos de aplicación de la metodología de DS.
14. Conocimiento de los RRHH del proyecto de DS.
15. RRHH disponibles del proyecto de DS.

Los Objetivos de aplicación de la metodología de DS son considerados fuertemente más relevantes que el Conocimiento de los Recursos Humanos (RRHH) del proyecto de DS y de importancia moderada respecto a la disponibilidad de estos últimos, tal como se aprecia en la Tabla 5.23. Finalmente el conocimiento y la disponibilidad de los RRHH del proyecto de DS son igualmente importantes según lo indicado por los expertos (ver Tabla 5.24).

Tabla 5.23: Matriz de dos entradas para Objetivos de aplicación de la metodología de DS y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C13: Objetivos de aplicación de la metodología de DS					C14: Conocimiento de los RRHH del proyecto de DS			
		X						
C13: Objetivos de aplicación de la metodología de DS					C15: RRHH disponibles del proyecto de DS			
			X					

Tabla 5.24: Matriz de dos entradas para Conocimiento de los RRHH del proyecto de DS y RRHH disponibles del proyecto de DS

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C14: Conocimiento de los RRHH del proyecto de DS					C15: RRHH disponibles del proyecto de DS			
				X				

5.2. Caso de validación: Rotura de automóviles 0KM durante el transporte a las concesionarias

A continuación se describen las tareas realizadas para la obtención de resultados para el caso de validación en cuestión, siguiendo los pasos definidos en el modelo propuesto (ver Capítulo IV).

5.2.1. Valoración con TCP del nivel 4 del CV1

Como se indicó anteriormente, los criterios ubicados en el cuarto nivel de la jerarquía, es decir, las características más específicas evaluadas para proyectos de DS, dependen del caso de validación en particular y la metodología utilizada. Con ayuda de la técnica de TCP se definieron las características, se establecieron las etiquetas lingüísticas y luego se valoraron los sub-criterios de dicho nivel (Paso 3.2, Capítulo IV). A los expertos se les ha brindado tal como se expone en el Anexo IV para su completitud; donde primeramente debían describir el entorno de validación o bien, donde se le indicaba cual era; en este caso son los causales de roturas de automóviles 0KM mientras son transportados desde la fábrica hasta las concesionarias (CV1).

A continuación se describen los sub-criterios, según la estructura jerárquica definida, junto a sus respectivas etiquetas lingüísticas, catalogadas en una escala del 1 al 9, que deben ser tomadas en cuenta para completar las parrillas.

Las características del Acceso a los Datos con sus respectivas etiquetas lingüísticas son las descriptas a continuación:

1. Portabilidad (capacidad de ser transferido)
    1. No se analiza
    2. De 0% a 13% de portabilidad
    3. De 14% a 27% de portabilidad
    4. De 28% a 41% de portabilidad
    5. De 42% a 55% de portabilidad
    6. De 56% a 69% de portabilidad
    7. De 70% a 83% de portabilidad
    8. De 84% a 97% de portabilidad
    9. De 98% a 100% de portabilidad
2. Accesibilidad (nivel de accesibilidad/asequibilidad de los datos)
    1. No se analiza
    2. De 0% a 13% de accesibilidad
    3. De 14% a 27% de accesibilidad
    4. De 28% a 41% de accesibilidad
    5. De 42% a 55% de accesibilidad
    6. De 56% a 69% de accesibilidad
    7. De 70% a 83% de accesibilidad
    8. De 84% a 97% de accesibilidad
    9. De 98% a 100% de accesibilidad
3. Diversidad (diferentes tipos de datos)
    1. No se analiza
    2. De 98% a 100% de diversidad
    3. De 84% a 97% de diversidad
    4. De 70% a 83% de diversidad
    5. De 56% a 69% de diversidad
    6. De 42% a 55% de diversidad
    7. De 28% a 41% de diversidad
    8. De 14% a 27% de diversidad
    9. De 0% a 13% de diversidad
4. Diversidad de fuentes de datos
    1. No se analiza
    2. De 98% a 100% de diversidad de las fuentes

3. De 84% a 97% de diversidad de las fuentes
4. De 70% a 83% de diversidad de las fuentes
5. De 56% a 69% de diversidad de las fuentes
6. De 42% a 55% de diversidad de las fuentes
7. De 28% a 41% de diversidad de las fuentes
8. De 14% a 27% de diversidad de las fuentes
9. De 0% a 13% de diversidad de las fuentes

5. Recursos necesarios (equipamiento)

1. No se analiza
2. De 0% a 13% de los recursos necesarios
3. De 14% a 27% de los recursos necesarios
4. De 28% a 41% de los recursos necesarios
5. De 42% a 55% de los recursos necesarios
6. De 56% a 69% de los recursos necesarios
7. De 70% a 83% de los recursos necesarios
8. De 84% a 97% de los recursos necesarios
9. De 98% a 100% de los recursos necesarios

Tabla 5.25: Parrilla de Acceso a Datos del CV1

		CRISP-DM	**MoPro PEI**	**P^3TQ**	
1.	No se analiza la portabilidad de los datos	5	9	8	De 98% a 100% de portabilidad de los datos
2.	No se analiza la accesibilidad de los datos	5	9	7	De 98% a 100% de accesibilidad de los datos
3.	No se analiza la diversidad de los datos	4	2	3	De 0% a 13% de diversidad de los datos
4.	No se analiza la diversidad de las fuentes de datos	5	2	3	De 0% a 13% de diversidad de la fuentes de datos
5.	No se analiza la disponibilidad de los recursos necesarios	7	9	8	De 98% a 100% de los recursos necesarios

En la Tabla 5.25 se exponen las valoraciones dadas por los expertos para el CV1, según el nivel de cumplimiento de dichas características en cada metodología evaluada.

Los sub-criterios del Uso de los Datos, junto a sus etiquetas lingüísticas se detallan a continuación:

6. Calidad (condiciones de los datos (valores válidos e inexistencia de valores anómalos))
    1. No se analiza
    2. De 0% a 13% de calidad/fiabilidad
    3. De 14% a 27% de calidad/fiabilidad
    4. De 28% a 41% de calidad/fiabilidad
    5. De 42% a 55% de calidad/fiabilidad
    6. De 56% a 69% de calidad/fiabilidad
    7. De 70% a 83% de calidad/fiabilidad
    8. De 84% a 97% de calidad/fiabilidad
    9. De 98% a 100% de calidad/fiabilidad
7. Completitud (grado de completitud de los datos (cantidad de valores faltantes))
    1. No se analiza
    2. De 0% a 13% de completitud
    3. De 14% a 27% de completitud
    4. De 28% a 41% de completitud
    5. De 42% a 55% de completitud
    6. De 56% a 69% de completitud
    7. De 70% a 83% de completitud
    8. De 84% a 97% de completitud
    9. De 98% a 100% de completitud
8. Funcionalidad (uso y practicidad de los datos)
    1. No se analiza
    2. De 0% a 13% de funcionalidad
    3. De 14% a 27% de funcionalidad
    4. De 28% a 41% de funcionalidad
    5. De 42% a 55% de funcionalidad
    6. De 56% a 69% de funcionalidad
    7. De 70% a 83% de funcionalidad

8. De 84% a 97% de funcionalidad
9. De 98% a 100% de funcionalidad

9. Costos de transformación (demanda de trabajo para adaptarlos para su uso)
    1. No se analiza
    2. De 98% a 100% de costos de transformación
    3. De 84% a 97% de costos de transformación
    4. De 70% a 83% de costos de transformación
    5. De 56% a 69% de costos de transformación
    6. De 42% a 55% de costos de transformación
    7. De 28% a 41% de costos de transformación
    8. De 14% a 27% de costos de transformación
    9. De 0% a 13% de costos de transformación
10. Datos riesgosos (datos sensibles, que no deben ser difundidos)
    1. No se analiza
    2. De 98% a 100% de datos riesgosos
    3. De 84% a 97% de datos riesgosos
    4. De 70% a 83% de datos riesgosos
    5. De 56% a 69% de datos riesgosos
    6. De 42% a 55% de datos riesgosos
    7. De 28% a 41% de datos riesgosos
    8. De 14% a 27% de datos riesgosos
    9. De 0% a 13% de datos riesgosos

Tabla 5.26: Parrilla de Uso de los Datos del CV1

		CRISP-DM	**MoPro PEI**	**P^3TQ**	
6.	No se analiza la calidad/fiabilidad de los datos	5	9	7	De 98% a 100% de calidad/fiabilidad de los datos
7.	No se analiza la completitud de los datos	5	9	7	De 98% a 100% de completitud de los datos
8.	No se analiza la funcionalidad de los datos	3	6	4	De 98% a 100% de funcionalidad de los datos

		CRISP-DM	MoPro PEI	P^3TQ	
9.	No se analiza el costo de transformación de los datos	6	4	6	De 0% a 13% de costo de transformación de los datos
10.	No se analiza los riesgos de los de los datos	8	5	6	De 0% a 13% de datos riesgosos

Teniendo en cuentas las características y sus etiquetas mencionadas, los expertos establecieron las valoraciones en relación a las tres metodologías, tal como se exhibe en la Tabla 5.26.

Para el Entorno del Negocio las etiquetas lingüísticas de sus sub-criterios son los siguientes:

1. Situación económica circundante
    1. No se analiza la situación circundante
    2. De 0% a 13% de favorabilidad de la situación circundante
    3. De 14% a 27% de favorabilidad de la situación circundante
    4. De 28% a 41% de favorabilidad de la situación circundante
    5. De 42% a 55% de favorabilidad de la situación circundante
    6. De 56% a 69% de favorabilidad de la situación circundante
    7. De 70% a 83% de favorabilidad de la situación circundante
    8. De 84% a 97% de favorabilidad de la situación circundante
    9. De 98% a 100% de favorabilidad de la situación circundante
2. Evaluación/comprensión general de la organización
    1. No se analiza o realiza una evaluación/comprensión general de la organización
    2. De 0% a 13% de evaluación/comprensión
    3. De 14% a 27% de evaluación/comprensión
    4. De 28% a 41% de evaluación/comprensión
    5. De 42% a 55% de evaluación/comprensión
    6. De 56% a 69% de evaluación/comprensión
    7. De 70% a 83% de evaluación/comprensión
    8. De 84% a 97% de evaluación/comprensión
    9. De 98% a 100% de evaluación/comprensión

3. Definición de los objetivos del negocio
    1. No se analiza o realiza la definición de los objetivos del negocio
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición
4. Definición del problema del negocio
    1. No se analiza o define el problema del negocio
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición
5. Definición de las oportunidades del negocio
    1. No se analizan o definen las oportunidades del negocio
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición
6. Definición de requisitos del negocio
    1. No se analizan o definen de los requisitos
    2. De 0% a 13% de claridad en la definición

3. De 14% a 27% de claridad en la definición
4. De 28% a 41% de claridad en la definición
5. De 42% a 55% de claridad en la definición
6. De 56% a 69% de claridad en la definición
7. De 70% a 83% de claridad en la definición
8. De 84% a 97% de claridad en la definición
9. De 98% a 100% de claridad en la definición

7. Definición de los criterios de éxito del proyecto (desde el negocio)
   1. No se analiza o definen los criterios de éxitos del proyecto (desde el negocio)
   2. De 0% a 13% de claridad en la definición
   3. De 14% a 27% de claridad en la definición
   4. De 28% a 41% de claridad en la definición
   5. De 42% a 55% de claridad en la definición
   6. De 56% a 69% de claridad en la definición
   7. De 70% a 83% de claridad en la definición
   8. De 84% a 97% de claridad en la definición
   9. De 98% a 100% de claridad en la definición

Según el grado de cumplimiento, los expertos asignaron las puntuaciones a cada sub-criterio y metodología, mostradas en la Tabla 5.27.

Tabla 5.27: Parrilla del Entorno del Negocio del CV1

		CRISP-DM	**MoPro PEI**	**P^3TQ**	
1.	No se analiza la situación circundante	7	9	8	De 98% a 100% de favorabilidad de la situación circundante
2.	No se analiza o realiza una evaluación/comprensión general de la organización	6	9	8	De 98% a 100% de evaluación/comprensión general de la organización
3.	No se analiza o realiza la definición de los objetivos del negocio	7	9	8	De 98% a 100% de claridad en la definición de los objetivos del negocio

		CRISP-DM	MoPro PEI	P^3TQ	
4.	No se analiza o define el problema del negocio	7	9	8	De 98% a 100% de claridad en la definición del problema del negocio
5.	No se analizan o definen las oportunidades del negocio	7	9	6	De 98% a 100% de claridad en la definición de las oportunidades del negocio
6.	No se analizan o definen de los requisitos	7	9	7	De 98% a 100% de claridad en la definición de los requisitos
7.	No se analiza o definen los criterios de éxitos del proyecto (desde el negocio)	8	9	7	De 98% a 100% de claridad en la definición de los criterios de éxitos del proyecto (desde el negocio)

Las características del Proyecto del Entorno del Negocio y sus etiquetas se describen a continuación:

8. Conocimiento de los RRHH del negocio
    1. No se analiza el conocimiento de los RRHH del negocio
    2. De 0% a 13% de conocimiento de los RRHH
    3. De 14% a 27% de conocimiento de los RRHH
    4. De 28% a 41% de conocimiento de los RRHH
    5. De 42% a 55% de conocimiento de los RRHH
    6. De 56% a 69% de conocimiento de los RRHH
    7. De 70% a 83% de conocimiento de los RRHH
    8. De 84% a 97% de conocimiento de los RRHH
    9. De 98% a 100% de conocimiento de los RRHH
9. Disponibilidad de los RRHH del negocio
    1. No se analiza la disponibilidad de los RRHH del negocio
    2. De 0% a 13% de disponibilidad de los RRHH
    3. De 14% a 27% de disponibilidad de los RRHH

4. De 28% a 41% de disponibilidad de los RRHH
5. De 42% a 55% de disponibilidad de los RRHH
6. De 56% a 69% de disponibilidad de los RRHH
7. De 70% a 83% de disponibilidad de los RRHH
8. De 84% a 97% de disponibilidad de los RRHH
9. De 98% a 100% de disponibilidad de los RRHH

10. Riesgos del problema del negocio (respecto al proyecto)
    1. No se analizan los riesgos del negocio (respecto al proyecto)
    2. De 0% a 13% de riesgo identificados (Insignificantes)
    3. De 14% a 27% de riesgo identificados
    4. De 28% a 41% de riesgo identificados
    5. De 42% a 55% de riesgo identificados
    6. De 56% a 69% de riesgo identificados
    7. De 70% a 83% de riesgo identificados
    8. De 84% a 97% de riesgo identificados
    9. De 98% a 100% de riesgo identificados (amenazan la supervivencia del proyecto)
11. Contingencias en relación a los riesgos del problema del negocio (del proyecto)
    1. No se analizan contingencias en relación a los riesgos del problema del negocio
    2. De 0% a 13% de contingencias en relación a los riesgos
    3. De 14% a 27% de contingencias en relación a los riesgos
    4. De 28% a 41% de contingencias en relación a los riesgos
    5. De 42% a 55% de contingencias en relación a los riesgos
    6. De 56% a 69% de contingencias en relación a los riesgos
    7. De 70% a 83% de contingencias en relación a los riesgos
    8. De 84% a 97% de contingencias en relación a los riesgos
    9. De 98% a 100% de contingencias en relación a los riesgos
12. Viabilidad del proyecto
    1. No se analiza la viabilidad del proyecto
    2. De 0% a 13% de viabilidad
    3. De 14% a 27% de viabilidad
    4. De 28% a 41% de viabilidad
    5. De 42% a 55% de viabilidad
    6. De 56% a 69% de viabilidad

7. De 70% a 83% de viabilidad
8. De 84% a 97% de viabilidad
9. De 98% a 100% de viabilidad

En la Tabla 5.28 se muestran las valoraciones dadas por los expertos, teniendo en cuenta las características recientemente mencionadas.

Tabla 5.28: Parrilla del Proyecto del Entorno del Negocio del CV1

		CRISP-DM	**MoPro PEI**	**P^3TQ**	
8.	No se analiza el conocimiento de los RRHH del negocio	6	8	7	De 98% a 100% de conocimiento de los RRHH del negocio
9.	No se analiza la disponibilidad de los RRHH del negocio	6	8	7	De 98% a 100% de disponibilidad de los RRHH del negocio
10.	No se analizan los riesgos del negocio (respecto al proyecto)	7	9	7	De 98% a 100% de riesgos identificados del negocio (amenazan la supervivencia del proyecto)
11.	No se analizan contingencias en relación a los riesgos del problema del negocio	7	9	7	De 98% a 100% de contingencias en relación a los riesgos del problema del negocio
12.	No se analiza la viabilidad del proyecto	8	9	6	De 98% a 100% de viabilidad del proyecto

Finalmente las etiquetas lingüísticas para el sub-cirterio Data Science en el Proyecto del Entorno del Negocio, son seguidamente mencionados:

13. Objetivos de aplicación de la metodología de DS
    1. No se analiza o realiza la definición de los objetivos de aplicación de la metodología de DS
    2. De 0% a 13% de claridad en la definición

3. De 14% a 27% de claridad en la definición
4. De 28% a 41% de claridad en la definición
5. De 42% a 55% de claridad en la definición
6. De 56% a 69% de claridad en la definición
7. De 70% a 83% de claridad en la definición
8. De 84% a 97% de claridad en la definición
9. De 98% a 100% de claridad en la definición

14. Conocimiento de los RRHH del proyecto de DS
    1. No se analiza el conocimiento de los RRHH del proyecto de DS
    2. De 0% a 13% de conocimiento de los RRHH
    3. De 14% a 27% de conocimiento de los RRHH
    4. De 28% a 41% de conocimiento de los RRHH
    5. De 42% a 55% de conocimiento de los RRHH
    6. De 56% a 69% de conocimiento de los RRHH
    7. De 70% a 83% de conocimiento de los RRHH
    8. De 84% a 97% de conocimiento de los RRHH
    9. De 98% a 100% de conocimiento de los RRHH
15. RRHH disponibles del proyecto de DS
    1. No se analiza la disponibilidad de los RRHH del proyecto de DS
    2. De 0% a 13% de disponibilidad de los RRHH
    3. De 14% a 27% de disponibilidad de los RRHH
    4. De 28% a 41% de disponibilidad de los RRHH
    5. De 42% a 55% de disponibilidad de los RRHH
    6. De 56% a 69% de disponibilidad de los RRHH
    7. De 70% a 83% de disponibilidad de los RRHH
    8. De 84% a 97% de disponibilidad de los RRHH
    9. De 98% a 100% de disponibilidad de los RRHH

Según el grado de cumplimiento, las puntuaciones fijadas a cada sub-criterio y metodología, están expuestas en la Tabla 5.29.

Tabla 5.29: Parrilla de Data Science en el Proyecto del Entorno del Negocio del CV1

		CRISP-DM	**MoPro PEI**	**P^3TQ**	
13.	No se analiza o realiza la definición de los objetivos de aplicación de la metodología de DS	8	9	7	De 98% a 100% de claridad en la definición de los objetivos de aplicación de la metodología de DS
14.	No se analiza el conocimiento de los RRHH del proyecto de DS	7	9	6	De 98% a 100% de conocimiento de los RRHH del proyecto de DS
15.	No se analiza la disponibilidad de los RRHH del proyecto de DS	7	9	6	De 98% a 100% de disponibilidad de los RRHH del proyecto de DS

5.2.2. Completitud de las matrices y establecimiento de prioridades entre criterios del CV1

Como se señaló en el Paso 4 de modelo propuesto en el Capítulo IV, a partir de las tablas generadas correspondiente a las matrices de dos entradas, como las de tipo parrillas, se completaron y adecuaron a las propuestas por Saaty. Para las correspondientes al primer tipo (matrices de dos entradas), el pasaje fue directo y para las parrillas, se tuvo que calcular la diferencia en valor absoluto por característica según cada metodología tomadas de a par y sumarle uno. Para la ubicación de las valoraciones dadas, dentro de la matriz pareada, según corresponda se ubicó en la parte principal o secundaria de la misma.

Tabla 5.30: Matriz pareada del nivel 2

Criterios principales	**ED**	**CN**
Entendimiento de los Datos (ED)	1	1
Comprensión del Negocio (CN)	1	1
Suma	2,00	2,00

Inicialmente se expone en las Tablas 5.30 y 5.31, las preferencias de los expertos en relación a los criterios del nivel 2, Entendimiento de los Datos (ED) y Comprensión del Negocio (CN), para los cuales los expertos indicaron que ambos atributos poseen la misma relevancia, por lo

que en la Tabla 5.30, 1 es el valor asignado, obteniendo una ponderación del 50% para cada uno, como se aprecia en la Tabla 5.31.

Tabla 5.31: Matriz Normalizada y ponderación del nivel 2

	Matriz Normalizada		**Ponderación**
Entendimiento de los Datos (ED)	0,50	0,50	0,50
Comprensión del Negocio (CN)	0,50	0,50	0,50
		Suma	1,00

Siguiendo la estructura jerárquica en el nivel 3 para el Entendimiento de los Datos (ED) existen dos sub-criterios Acceso a los Datos (ED-AD) y Uso de los Datos (ED-UD), cuyas valoraciones y ponderaciones se visualizan en la Tabla 5.32 y 5.33. Según los expertos, el uso de los datos (ED-UD) es modernamente más relevante que el acceso a los mismos (ED-AD), por lo que sus ponderaciones son de 0,75 y 0,25 respectivamente.

Tabla 5.32: Matriz pareada del nivel 3 para ED

Sub-Criterios de Entendimiento de los Datos	**ED-AD**	**ED-UD**
Acceso a los Datos (ED-AD)	1	1/3
Uso de los Datos (ED-UD)	3	1
Suma	4,00	1,33

Tabla 5.33: Matriz Normalizada y ponderación del nivel 3 para ED

	Matriz Normalizada		**Ponderación**
Acceso a los Datos (ED-AD)	0,25	0,25	0,25
Uso de los Datos (ED-UD)	0,75	0,75	0,75
		Suma	1,00

En cuanto al Acceso a los Datos (ED-AD), el sub-criterio más relevante es la Diversidad de fuentes de datos (ED-SC4), seguido de la Diversidad (ED-SC3), Accesibilidad (ED-SC2) y por último la Portabilidad (ED-SC1) y los Recursos necesarios (ED-SC5), como puede apreciarse en las Tablas 5.34 y 5.35.

Tabla 5.34: Matriz pareada del nivel 4 para ED-AD

Sub-Criterios del Acceso de los Datos	**ED-SC1**	**ED-SC2**	**ED-SC3**	**ED-SC4**	**ED-SC5**
Portabilidad (ED-SC1)	1	1/7	1/3	1/5	1
Accesibilidad (ED-SC2)	7	1	1	1/3	3
Diversidad (ED-SC3)	3	1	1	1	3
Diversidad de fuentes de datos (ED-SC4)	5	3	1	1	5
Recursos necesarios (ED-SC5)	1	1/3	1/3	1/5	1
Suma	17,00	5,48	3,67	2,73	13,00

Tabla 5.35: Matriz Normalizada y ponderación del nivel 4 para ED-AD

	Matriz Normalizada					**Ponderación**
Portabilidad (ED-SC1)	0,06	0,03	0,09	0,07	0,08	0,07
Accesibilidad (ED-SC2)	0,41	0,18	0,27	0,12	0,23	0,24
Diversidad (ED-SC3)	0,18	0,18	0,27	0,37	0,23	0,25
Diversidad de fuentes de datos (ED-SC4)	0,29	0,55	0,27	0,37	0,38	0,37
Recursos necesarios (ED-SC5)	0,06	0,06	0,09	0,07	0,08	0,07
					Suma	1,00

Al analizar cada sub-criterio respecto dentro del ED-AD respecto a las metodologías de DS, representadas como las alternativas de la jerarquía propuesta para el presente TFM, se puede apreciar en las Tablas 5.36 a la 5.40 las ponderaciones obtenidas para cada característica para el CV1.

Tabla 5.36: Ponderación ED-SC1 en relación a cada alternativa para CV1

Portabilidad (ED-SC1)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/5	1/4	0,10	0,12	0,08	0,10
MoProPEI (A2)	5	1	2	0,50	0,59	0,62	0,57
P^3TQ (A3)	4	1/2	1	0,40	0,29	0,31	0,33
Suma	10,00	1,70	3,25	1,00	1,00	1,00	1,00

Para la Portabilidad (ED-SC1), la mayor ponderación obtuvo la metodología MoProPEI (A2), seguida de P^3TQ (A3) y por último CRISP-DM (A1), con 0,57, 0,33 y 0,10 respectivamente, como se puede observar en la Tabla 5.36.

En cuento a la Accesibilidad (ED-SC2), como se expone en la Tabla 5.37, el orden de preferencia es el mismo que en el caso de la característica anterior (ED-SC1), obteniendo una mayor diferencia a favor de la A2.

Tabla 5.37: Ponderación de ED-SC2 en relación a cada alternativa para CV1

Accesibilidad (ED-SC2)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/5	1/3	0,11	0,13	0,08	0,11
MoProPEI (A2)	5	1	3	0,56	0,65	0,69	0,63
P^3TQ (A3)	3	1/3	1	0,33	0,22	0,23	0,26
Suma	9,00	1,53	4,33	1,00	1,00	1,00	1,00

A diferencia que los sub-criterios anteriores, para la Diversidad (ED-SC3), el orden de preferencia se encuentra invertido, es decir, que la metodología que mejor se adapta al CV1 en función a esta características es CRISP-DM (A1), seguida de P3TQ (A3) y por último MoProPEI (A2), con 54%, 30% y 16%, tal como se indica en la Tabla 5.38.

Tabla 5.38: Ponderación de ED-SC3 en relación a cada alternativa para CV1

Diversidad (ED-SC3)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	3	2	0,55	0,50	0,57	0,54
MoProPEI (A2)	1/3	1	1/2	0,18	0,17	0,14	0,16
P^3TQ (A3)	1/2	2	1	0,27	0,33	0,29	0,30
Suma	1,83	6,00	3,50	1,00	1,00	1,00	1,00

Tabla 5.39: Ponderación de ED-SC4 en relación a cada alternativa para CV1

Diversidad de las fuentes de datos (ED-SC4)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/4	1/3	0,13	0,14	0,10	0,12
MoProPEI (A2)	4	1	2	0,50	0,57	0,60	0,56
P^3TQ (A3)	3	1/2	1	0,38	0,29	0,30	0,32
Suma	8,00	1,75	3,33	1,00	1,00	1,00	1,00

Para la Diversidad de las fuentes de datos (ED-SC4) y los Recursos necesarios (ED-SC5), nuevamente la mayor ponderación obtuvo la metodología MoProPEI (A2), luego P^3TQ (A3) y en último lugar CRISP-DM (A1), tal como se observan en las Tablas 5.39 y 5.40.

Tabla 5.40: Ponderación de ED-SC5 en relación a cada alternativa para CV1

Recursos necesarios (ED-SC5)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/2	0,17	0,18	0,14	0,16
MoProPEI (A2)	3	1	2	0,50	0,55	0,57	0,54
P^3TQ (A3)	2	1/2	1	0,33	0,27	0,29	0,30
Suma	6,00	1,83	3,50	1,00	1,00	1,00	1,00

Respecto al Uso de los Datos, tal como se visualiza en las Tablas 5.41 y 5.42, el sub-criterio más relevante es la Calidad (ED-SC6), seguido de la Completitud (ED-SC7), Datos riesgosos (ED-SC10), Funcionalidad (ED-SC8) y por último los Costos de transformación (ED-SC9).

Tabla 5.41: Matriz pareada del nivel 4 para ED-UD

Sub-Criterios del Uso de los Datos	**ED-SC6**	**ED-SC7**	**ED-SC8**	**ED-SC9**	**ED-SC10**
Calidad (ED-SC6)	1	1	5	7	1
Completitud (ED-SC7)	1	1	3	5	1
Funcionalidad (ED-SC8)	1/5	1/3	1	3	1
Costo de transformación (ED-SC9)	1/7	1/5	1/3	1	1/3
Datos riesgosos (ED-SC10)	1	1	1	3	1
Suma	3,34	3,53	10,33	19,00	4,33

Tabla 5.42: Matriz normalizada y ponderación del nivel 4 para ED-UD

	Matriz Normalizada					**Ponderación**
Calidad (ED-SC6)	0,30	0,28	0,48	0,37	0,23	0,33
Completitud (ED-SC7)	0,30	0,28	0,29	0,26	0,23	0,27
Funcionalidad (ED-SC8)	0,06	0,09	0,10	0,16	0,23	0,13
Costo de transformación (ED-SC9)	0,04	0,06	0,03	0,05	0,08	0,05
Datos riesgosos (ED-SC10)	0,30	0,28	0,10	0,16	0,23	0,21
					Suma	1,00

De la Tabla 5.43 a la 5.47 se muestran las ponderaciones obtenidas para cada sub-criterio dentro del ED-UD en relación a las alternativas (metodologías) evaluadas.

Al analizar la Calidad (ED-SC6) y la ponderación de la Completitud (ED-SC7), MoProPEI obtuvo la mayor ponderación (0,63), seguida de P^3TQ (0,26) y por último CRISP-DM (0,11), como se puede observar en las Tablas 5.43 y 5.44. Para la Funcionalidad (ED-SC8), se mantiene la tendencia anterior como se ve en la Tabla 5.45.

Tabla 5.43: Ponderación de ED-SC6 en relación a cada alternativa para CV1

Calidad (ED-SC6)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/5	1/3	0,11	0,13	0,08	0,11
MoProPEI (A2)	5	1	3	0,56	0,65	0,69	0,63
P^3TQ (A3)	3	1/3	1	0,33	0,22	0,23	0,26
Suma	9,00	1,53	4,33	1,00	1,00	1,00	1,00

Tabla 5.44: Ponderación de ED-SC7 en relación a cada alternativa para CV1

Completitud (ED-SC7)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/5	1/3	0,11	0,13	0,08	0,11
MoProPEI (A2)	5	1	3	0,56	0,65	0,69	0,63
P^3TQ (A3)	3	1/3	1	0,33	0,22	0,23	0,26
Suma	9,00	1,53	4,33	1,00	1,00	1,00	1,00

Tabla 5.45: Ponderación de ED-SC8 en relación a cada alternativa para CV1

Funcionalidad (ED-SC8)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/4	1/2	0,14	0,16	0,11	0,14
MoProPEI (A2)	4	1	3	0,57	0,63	0,67	0,62
P^3TQ (A3)	2	1/3	1	0,29	0,21	0,22	0,24
Suma	7	1,58	4,5	1,00	1,00	1,00	1,00

En relación a los Costos de transformación (ED-SC9), las metodologías CRISP-DM (A1) y P^{3}TQ (A3) se encuentran en primer lugar con 43% y MoProPEI (A2) en último lugar con 14%, tal como se muestra en la Tabla 5.46.

Tabla 5.46: Ponderación de ED-SC9 en relación a cada alternativa para CV1

Costos de transformación (ED-SC9)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM **(A1)**	1	3	1	0,43	0,43	0,43	0,43
MoProPEI (A2)	1/3	1	1/3	0,14	0,14	0,14	0,14
P^{3}TQ (A3)	1	3	1	0,43	0,43	0,43	0,43
Suma	2,33	7,00	2,33	1,00	1,00	1,00	1,00

Por último, en la Tabla 5.47 se expone la ponderación de los Datos riesgosos (ED-SC10) en relación a cada metodología, la mejor alternativa es CRISP-DM, luego P^{3}TQ y por último MoProPEI, con el 62, 24 y 24 por ciento respectivamente.

Tabla 5.47: Ponderación de ED-SC10 en relación a cada alternativa para CV1

Datos riesgosos (ED-SC10)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	4	3	0,63	0,57	0,67	0,62
MoProPEI (A2)	1/4	1	1/2	0,16	0,14	0,11	0,14
P^{3}TQ (A3)	1/3	2	1	0,21	0,29	0,22	0,24
Suma	1,58	7,00	4,50	1,00	1,00	1,00	1,00

En el nivel 3, la Comprensión del Negocio (CN), tal como se mencionó antes, se divide en tres sub-criterios: el Entorno del Negocio (EN), Proyecto en el Entorno del Negocio (PEN) y DS en el Proyecto del Negocio (DSPN). Al comparar estos sub-criterios las valoraciones dadas por los expertos son las que se muestran en las Tablas 5.48 y 5.49, a partir de las cuales se puede notar que este último (DSPN), posee mayor relevancia que los dos restantes, es moderadamente más importante que ellos (Tabla 5.48); por lo que su ponderación en dicho nivel es del 60% y para los dos restantes del 20% cada uno (Tabla 5.49), donde la preferencia es igualmente importante entre sí (Tabla 5.48).

Tabla 5.48: Matriz pareada del nivel 3 para CN

Sub-Criterios de Comprensión del Negocio	**CN-EN**	**CN-PEN**	**CN-DSPN**
Entorno del Negocio (CN-EN)	1	1	1/3
Proyecto en el Entorno del Negocio (CN-PEN)	1	1	1/3
DS en el Proyecto del Negocio (CN-DSPN)	3	3	1
Suma	5,00	5,00	1,67

Tabla 5.49: Matriz normalizada y ponderación del nivel 3 para CN

	Matriz Normalizada			**Ponderación**
Entorno del Negocio (CN-EN)	0,20	0,20	0,20	0,20
Proyecto en el Entorno del Negocio (CN-PEN)	0,20	0,20	0,20	0,20
DS en el Proyecto del Negocio (CN-DSPN)	0,60	0,60	0,60	0,60
Suma				1,00

En cuanto al Entorno del Negocio, como se indicó anteriormente, se divide en 7 características, en la Tabla 5.50 se observan los juicios de preferencias emitidos por los expertos a partir de los cuales se obtienen las ponderaciones de la Tabla 5.51, donde se observa que el criterio más relevante es la Definición de criterios de éxito del proyecto (CN-SC7) con una ponderación de 0,19, seguido de la Comprensión general de la organización (CN-SC2), la Definición de los objetivos del negocio (CN-SC3), la Definición del problema del negocio (CN-SC4), la Definición de las oportunidades del negocio (CN-SC5) con 0,16, luego se encuentra la Definición de requisitos del negocio (CN-SC6) con 0,14 y por último Situación económica circundante (CN-SC1) con 0,05 de ponderación.

Tabla 5.50: Matriz pareada del nivel 4 para CN-EN

Sub-Criterios del Entorno del Negocio	**(CN-SC1)**	**(CN-SC2)**	**(CN-SC3)**	**(CN-SC4)**	**(CN-SC5)**	**(CN-SC6)**	**(CN-SC7)**
Situación económica circundante (CN-SC1)	1	1/3	1/3	1/3	1/3	1/3	1/3
Comprensión general de la organización (CN-SC2)	3	1	1	1	1	1	1
Definición de los objetivos del negocio (CN-SC3)	3	1	1	1	1	1	1

Sub-Criterios del Entorno del Negocio	**(CN-SC1)**	**(CN-SC2)**	**(CN-SC3)**	**(CN-SC4)**	**(CN-SC5)**	**(CN-SC6)**	**(CN-SC7)**
Definición del problema del negocio (CN-SC4)	3	1	1	1	1	1	1
Definición de las oportunidades del negocio (CN-SC5)	3	1	1	1	1	1	1
Definición de requisitos del negocio (CN-SC6)	3	1	1	1	1	1	1/3
Definición de criterios de éxito del proyecto (CN-SC7)	3	1	1	1	1	3	1
Suma	19,00	6,33	6,33	6,33	6,33	8,33	5,67

Tabla 5.51: Matriz normalizada y ponderación del nivel 4 para CN-EN

	Matriz Normalizada							**Ponderación**
Situación económica circundante (CN-SC1)	0,05	0,05	0,05	0,05	0,05	0,04	0,06	0,05
Comprensión general de la organización (CN-SC2)	0,16	0,16	0,16	0,16	0,16	0,12	0,18	0,16
Definición de los objetivos del negocio (CN-SC3)	0,16	0,16	0,16	0,16	0,16	0,12	0,18	0,16
Definición del problema del negocio (CN-SC4)	0,16	0,16	0,16	0,16	0,16	0,12	0,18	0,16
Definición de las oportunidades del negocio (CN-SC5)	0,16	0,16	0,16	0,16	0,16	0,12	0,18	0,16
Definición de requisitos del negocio (CN-SC6)	0,16	0,16	0,16	0,16	0,16	0,12	0,06	0,14
Definición de criterios de éxito del proyecto (CN-SC7)	0,16	0,16	0,16	0,16	0,16	0,36	0,18	0,19
							Suma	1,00

Las siguientes tablas (5.52 a 5.58) exponen las ponderaciones de las características mencionadas, a partir del grado de cumplimiento de dichos criterios para el caso de validación (CV1), tal como se indicó anteriormente.

Para la Situación económica circundante (CN-SC1), la ponderación superior corresponde a MoProPEI (57%), seguida de P^{3}TQ (30%) y por último CRISP-DM (16%), como se puede notar en la Tabla 5.52.

Tabla 5.52: Ponderación CN-SC1 en relación a cada alternativa para CV1

Situación económica circundante (CN-SC1)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/2	0,17	0,18	0,14	0,16
MoProPEI (A2)	3	1	2	0,50	0,55	0,57	0,54
P^{3}TQ (A3)	2	1/2	1	0,33	0,27	0,29	0,30
Suma	6,00	1,83	3,50	1,00	1,00	1,00	1,00

En cuento a la Comprensión general de la organización (CN-SC2) en relación a cada metodología, a diferencia que el sub-criterio anterior, en éste la mejor ponderación la obtuvo CRISP-DM con el 62%, luego MoProPEI con 22% y por último P^{3}TQ con 16%, tal como se expone en la Tabla 5.53.

Tabla 5.53: Ponderación CN-SC2 en relación a cada alternativa para CV1

Comprensión general de la organización (CN-SC2)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	4	3	0,63	0,73	0,50	0,62
MoProPEI (A2)	1/4	1	2	0,16	0,18	0,33	0,22
P^{3}TQ (A3)	1/3	1/2	1	0,21	0,09	0,17	0,16
Suma	1,58	5,50	6,00	1,00	1,00	1,00	1,00

Para los sub-criterios CN-SC3 (Definición de los objetivos del negocio) y CN-SC4 (Definición del problema del negocio), MoProPEI logró una ponderación del 54%, P^{3}TQ de 30% y finalmente CRISP-DM de 16% (ver Tablas 5.54 y 5.55).

Tabla 5.54: Ponderación CN-SC3 en relación a cada alternativa para CV1

Definición de los objetivos del negocio (CN-SC3)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/2	0,17	0,18	0,14	0,16
MoProPEI (A2)	3	1	2	0,50	0,55	0,57	0,54
P^3TQ (A3)	2	1/2	1	0,33	0,27	0,29	0,30
Suma	6,00	1,83	3,50	1,00	1,00	1,00	1,00

Tabla 5.55: Ponderación CN-SC4 en relación a cada alternativa para CV1

Definición del problema del negocio (CN-SC4)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/2	0,17	0,18	0,14	0,16
MoProPEI (A2)	3	1	2	0,50	0,55	0,57	0,54
P^3TQ (A3)	2	1/2	1	0,33	0,27	0,29	0,30
Suma	6,00	1,83	3,50	1,00	1,00	1,00	1,00

Como se advierte en la Tabla 5.56, para la Definición de las oportunidades del negocio en el CV1, la metodología MoProPEI es la más adecuada, duplicando la ponderación obtenida en relación a las dos restantes, con un 62% por sobre los 24% y 14% adquiridos de CRISP-DM y P^3TQ.

Tabla 5.56: Ponderación CN-SC5 en relación a cada alternativa para CV1

Definición de las oportunidades del negocio (CN-SC5)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	2	0,22	0,21	0,29	0,24
MoProPEI (A2)	3	1	4	0,67	0,63	0,57	0,62
P^3TQ (A3)	1/2	1/4	1	0,11	0,16	0,14	0,14
Suma	4,50	1,58	7,00	1,00	1,00	1,00	1,00

Como en el caso anterior, MoProPEI cumple en mayor medida con la característica CN-SC6 (Definición de requisitos del negocio), donde alcanzó un 60% y las dos restantes metodologías 20% cada una, tal como puede apreciarse en la Tabla 5.57.

Tabla 5.57: Ponderación CN-SC6 en relación a cada alternativa para CV1

Definición de requisitos del negocio (CN-SC6)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1	0,20	0,20	0,20	0,20
MoProPEI (A2)	3	1	3	0,60	0,60	0,60	0,60
P^3TQ (A3)	1	1/3	1	0,20	0,20	0,20	0,20
Suma	5,00	1,67	5,00	1,00	1,00	1,00	1,00

Finalmente, en la Tabla 5.58, se representan las ponderaciones obtenidas para la Definición de criterios de éxito del proyecto, donde el orden de preferencia ubica a MoProPEI en primer lugar, CRISP-DM en segundo y en tercero a P^3TQ.

Tabla 5.58: Ponderación CN-SC7 en relación a cada alternativa para CV1

Definición de criterios de éxito del proyecto (CN-SC7)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	2	0,29	0,27	0,33	0,30
MoProPEI (A2)	2	1	3	0,57	0,55	0,50	0,54
P^3TQ (A3)	1/2	1/3	1	0,14	0,18	0,17	0,16
Suma	3,50	1,83	6				

Tabla 5.59: Matriz pareada del nivel 4 para CN-PEN

Sub-Criterios del Proyecto del Entorno del Negocio	**CN-SC8**	**CN-SC9**	**CN-SC10**	**CN-SC11**	**CN-SC12**
Conocimiento de los RRHH del negocio (CN-SC8)	1	1	1	1/3	1/3
Disponibilidad de los RRHH del negocio (CN-SC9)	1	1	1/3	1/3	1
Riesgos del problema del negocio (CN-SC10)	1	3	1	1	1
Contingencias en relación a los riesgos del problema del negocio (CN-SC11)	3	3	1	1	3
Viabilidad del proyecto (CN-SC12)	3	1	1	1/3	1
Suma	9,00	9,00	4,33	3,00	6,33

Las valoraciones y ponderaciones del nivel 4, para los cinco sub-criterios dentro del Proyecto del Entorno del Negocio se exponen en las Tablas 5.59 y 5.60, para los cuales según orden

descendente de preferencias se ubicaría primero las Contingencias en relación a los riesgos del problema del negocio (CN-SC11) con 0,34, segundo los Riesgos del problema del negocio (CN-SC10) con 0,23, tercero la Viabilidad del proyecto (CN-SC12) con 0,19, cuarto el Conocimiento de los RRHH del negocio (CN-SC8) con 0,12 y en quinto lugar el Conocimiento de los RRHH del negocio (CN-SC9) con 0,11.

Tabla 5.60: Matriz normalizada y ponderación del nivel 4 para CN-PEN

	Matriz Normalizada					**Ponderación**
Conocimiento de los RRHH del negocio (CN-SC8)	0,11	0,11	0,23	0,11	0,05	0,12
Disponibilidad de los RRHH del negocio (CN-SC9)	0,11	0,11	0,08	0,11	0,16	0,11
Riesgos del problema del negocio (CN-SC10)	0,11	0,33	0,23	0,33	0,16	0,23
Contingencias en relación a los riesgos del problema del negocio (CN-SC11)	0,33	0,33	0,23	0,33	0,47	0,34
Viabilidad del proyecto (CN-SC12)	0,33	0,11	0,23	0,11	0,16	0,19
					Suma	1,00

Tabla 5.61: Ponderación CN-SC8 en relación a cada alternativa para CV1

Conocimiento de los RRHH del negocio (CN-SC8)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/2	0,17	0,18	0,14	0,16
MoProPEI (A2)	3	1	2	0,50	0,55	0,57	0,54
P^3TQ (A3)	2	1/2	1	0,33	0,27	0,29	0,30
Suma	6,00	1,83	3,50	1,00	1,00	1,00	1,00

Al comparar las alternativas (metodologías) con este último nivel de refinamiento de criterios dentro del CN-PEN, el Conocimiento de los RRHH del negocio (CN-SC8) y su disponibilidad (CN-SC9), lograron las mismas ponderaciones, donde MoProPEI tiene 0,54, PT^3Q 0,30 y CRISP-DM 0,16 (ver Tablas 5.61 y 5.62).

Como se puede divisar en las Tablas 5.63 y 5.64, las características Riesgos del problema del negocio (CN-SC10) y las Contingencias en relación a los mismos, poseen un idéntico

comportamiento para el caso de validación (CV1), donde MoProPEI obtuvo el 60% y las otras dos metodologías 20% cada una.

Por último, la Viabilidad del proyecto también indica preferencia de la alternativa 2, por sobre las demás, tal como se aprecia en la Tabla 5.65.

Tabla 5.62: Ponderación CN-SC9 en relación a cada alternativa para CV1

Disponibilidad de los RRHH del negocio (CN-SC9)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/2	0,17	0,18	0,14	0,16
MoProPEI (A2)	3	1	2	0,50	0,55	0,57	0,54
P^{3}TQ (A3)	2	1/2	1	0,33	0,27	0,29	0,30
Suma	6,00	1,83	3,50	1,00	1,00	1,00	1,00

Tabla 5.63: Ponderación CN-SC10 en relación a cada alternativa para CV1

Riesgos del problema del negocio (CN-SC10)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1	0,20	0,20	0,20	0,20
MoProPEI (A2)	3	1	3	0,60	0,60	0,60	0,60
P^{3}TQ (A3)	1	1/3	1	0,20	0,20	0,20	0,20
Suma	5,00	1,67	5,00	1,00	1,00	1,00	1,00

Tabla 5.64: Ponderación CN-SC11 en relación a cada alternativa para CV1

Contingencias en relación a los riesgos del problema del negocio (CN-SC11)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1	0,20	0,20	0,20	0,20
MoProPEI (A2)	3	1	3	0,60	0,60	0,60	0,60
P^{3}TQ (A3)	1	1/3	1	0,20	0,20	0,20	0,20
Suma	5,00	1,67	5,00	1,00	1,00	1,00	1,00

Tabla 5.65: Ponderación CN-SC12 en relación a cada alternativa para CV1

Viabilidad del proyecto (CN-SC12)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	3	0,30	0,29	0,38	0,32
MoProPEI (A2)	2	1	4	0,60	0,57	0,50	0,56
P^{3}TQ (A3)	1/3	1/4	1	0,10	0,14	0,13	0,12
Suma	3,33	1,75	8,00	1,00	1,00	1,00	1,00

Finalmente, el sub-criterio Data Science en el Proyecto del Negocio se desglosa en tres características: Objetivos de la aplicación de la metodología de DS (CN-SC13), Conocimiento de los RRHH del proyecto de DS (CN-SC14) y RRHH disponibles del proyecto de DS (CN-SC15); para los cuales el primero es fuertemente más importante que el segundo y moderadamente más que el tercero, tal como se puede apreciar en la Tabla 5.66, siendo esto reflejado en las ponderaciones de la Tabla 5.67, donde CN-SC13 adquirió 66% de relevancia (peso) .

Tabla 5.66: Matriz pareada del nivel 4 para CN-DSPN

Sub-Criterios de Data Science en el Proyecto del Negocio	**CN-SC13**	**CN-SC14**	**CN-SC15**
Objetivos de la aplicación de la metodología de DS (CN-SC13)	1	5	3
Conocimiento de los RRHH del proyecto de DS (CN-SC14)	1/5	1	1
RRHH disponibles del proyecto de DS (CN-SC15)	1/3	1	1
Suma	1,53	7,00	5,00

Tabla 5.67: Matriz normalizada y ponderación del nivel 4 para CN-DSPN

	Matriz Normalizada			**Ponderación**
Objetivos de la aplicación de la metodología de DS (CN-SC13)	0,65	0,71	0,60	0,66
Conocimiento de los RRHH del proyecto de DS (CN-SC14)	0,13	0,14	0,20	0,16
RRHH disponibles del proyecto de DS (CN-SC15)	0,22	0,14	0,20	0,19
Suma				1,00

Tabla 5.68: Ponderación de CN-SC13 en relación a cada alternativa para CV1

Objetivos de la aplicación de la metodología de DS (CN-SC13)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	2	0,29	0,27	0,33	0,30
MoProPEI (A2)	2	1	3	0,57	0,55	0,50	0,54
P^3TQ (A3)	1/2	1/3	1	0,14	0,18	0,17	0,16
Suma	3,50	1,83	6,00	1,00	1,00	1,00	1,00

Al analizar cada una de estas características por separado, relacionándolas con las metodologías evaluadas, puede percatarse en las Tablas 5.68, 5.69 y 5.70 que para el caso de validación (CV1) MoProPEI adquirió mayor envergadura que la demás, seguido de CRISP-DM y en menor medida P^3TQ.

Tabla 5.69: Ponderación de CN-SC14 en relación a cada alternativa para CV1

Conocimiento de los RRHH del proyecto de DS (CN-SC14)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	2	0,22	0,21	0,29	0,24
MoProPEI (A2)	3	1	4	0,67	0,63	0,57	0,62
P^3TQ (A3)	1/2	1/4	1	0,11	0,16	0,14	0,14
Suma	4,50	1,58	7,00	1,00	1,00	1,00	1,00

Tabla 5.70: Ponderación de CN-SC15 en relación a cada alternativa para CV1

RRHH disponibles del proyecto de DS (CN-SC15)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	2	0,22	0,21	0,29	0,24
MoProPEI (A2)	3	1	4	0,67	0,63	0,57	0,62
P^3TQ (A3)	1/2	1/4	1	0,11	0,16	0,14	0,14
Suma	4,50	1,58	7,00	1,00	1,00	1,00	1,00

5.2.3. Evaluación de consistencia del CV1

Para matrices de dos criterios (o sub-criterios), como lo son los ubicados en el segundo nivel de la estructura jerárquica del problema (ED y CN) no es necesario verificar la consistencia; a partir de las matrices de tres entradas se requiere comprobar la consistencia de los juicios emitidos por los expertos dispuestos en las matrices; a continuación se detallan los controles realizados según las matrices (tablas) expuestas en el paso anterior (5.2.2).

Para el ED, como se mencionó, al ser dos los sub-criterios involucrados (AD y UD) no se realiza un control de consistencia.

En las Tablas 5.71 y 5.72 se exponen los pasos realizados para la evaluación de consistencia de AD. En la Tabla 5.71, WxP corresponde al producto de la matriz (W) de comparación pareada (sin normalizar) por el vector de ponderaciones obtenidas (P). Los valores de la columna cociente se obtiene de dividir el vector de ponderaciones (P) por WxP y así mediante el promedio de la columna se obtiene el autovector máximo $\lambda_{máx}$; cabe destacar que cuanto más cercano es este valor al orden de la matriz más acertados han sido las valoraciones.

Tabla 5.71: Autovector máximo del ED-AD

Sub-criterios ED-AD	WxP	Cociente
Portabilidad (ED-SC1)	0,33	5,04
Accesibilidad (ED-SC2)	1,29	5,27
Diversidad (ED-SC3)	1,27	5,19
Diversidad de fuentes de datos (ED-SC4)	2,04	5,46
Recursos necesarios (ED-SC5)	0,38	5,20
		$\lambda_{máx}$= 5,23

Tabla 5.72: Evaluación de consistencia del ED-AD

$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,058
$\boldsymbol{RI}$	1,115
$\boldsymbol{CR = CI/RI}$	**0,052**

Como se estableció en el paso 3 del Anexo I, la verificación de consistencia se realiza como se indica en la Tabla 5.72; donde n es la cantidad de entradas de la matriz, en este caso 5 y su RI correspondiente es 1,115 (ver Tabla I.II del mismo Anexo). Se puede observar que los

juicios son consistentes, dado que la razón o relación de consistencia (CR, Consistency Ratio) es menor a 0,10.

Una vez evaluada la consistencia de los sub-criterios del AD de manera conjunta, se prosiguió a realizarlo por separado, para lo cual se siguió el mismo procedimiento mencionado, pero ahora de los sub-criterios en relación a las alternativas (metodologías de DS).
En la Tabla 5.73 se visualiza los valores intermedios necesarios para obtener el autovector máximo $\lambda_{máx}$ de los sub-criterios: Portabilidad de los datos (ED-SC1), Accesibilidad (ED-SC2) y la Diversidad de los mismos (ED-SC3), luego la Diversidad de fuentes de datos (ED-SC4) y los Recursos necesarios (ED-SC5).

Tabla 5.73: Autovector máximo de cada sub-criterio del ED-AD para CV1

Sub-criterios ED-AD	Alternativas	WxP	Cociente	$\lambda_{máx}$
ED-SC1	CRISP-DM (A1)	0,30	3,01	3,02
	MoProPEI (A2)	1,73	3,04	
	P^3TQ (A3)	1,01	3,03	
ED-SC2	CRISP-DM (A1)	0,32	3,01	3,04
	MoProPEI (A2)	1,95	3,07	
	P^3TQ (A3)	0,79	3,03	
ED-SC3	CRISP-DM (A1)	1,62	3,01	3,01
	MoProPEI (A2)	0,49	3,00	
	P^3TQ (A3)	0,89	3,01	
ED-SC4	CRISP-DM (A1)	0,37	3,01	3,02
	MoProPEI (A2)	1,69	3,03	
	P^3TQ (A3)	0,97	3,02	
ED-SC5	CRISP-DM (A1)	0,49	3,00	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,89	3,01	

Al evaluar la razón de consistencia de dichos sub-criterios se verificó que los mismos la cumplen, obteniendo un CR menor a 0,10 en todos los casos; tal como se observa la Tabla 5.74. Se puede apreciar que la razón de consistencia de la característica ED-SC1 es de 0,023,

de ED-SC2 es de 0,037, de ED-SC3 es de 0,009, de ED-SC4 es de 0,017 y de ED-SC5 es de 0,009; por tanto son muy consistentes los valores asignados para dichas características.

Tabla 5.74: Evaluación de consistencia de los sub-criterios del ED-AD para CV1

Evaluación de consistencia	ED-SC1	ED-SC2	ED-SC3	ED-SC4	ED-SC5
$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,012	0,019	0,005	0,009	0,005
$\boldsymbol{RI}$	0,525	0,525	0,525	0,525	0,525
$\boldsymbol{CR = CI/RI}$	**0,023**	**0,037**	**0,009**	**0,017**	**0,009**

En cuento al sub-criterio UD (Uso de los Datos), se evaluó primeramente la consistencia de las valoraciones realizada por los expertos de los cinco sub-criterios del UD en conjunto, es decir comparados entre sí, tal como se exponen en las Tablas 5.75 y 5.76, para los cuales el CR obtenido es de 0,057, por tanto se consideran consistentes los valores asignados.

Tabla 5.75: Autovector máximo de cada sub-criterio del ED-UD

Sub-criterios ED-UD	WxP	Cociente
Calidad (ED-SC6)	1,83	5,48
Completitud (ED-SC7)	1,46	5,36
Funcionalidad (ED-SC8)	0,66	5,13
Costo de transformación (ED-SC9)	0,27	5,14
Datos riesgosos (ED-SC10)	1,10	5,17
		$\lambda_{máx}$= 5,26

Tabla 5.76: Evaluación de consistencia de los sub-criterios del ED-UD

$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,064
$\boldsymbol{RI}$	1,115
$\boldsymbol{CR = CI/RI}$	**0,057**

Los pasos y valores intermedios necesarios para obtener el $\lambda_{máx}$ de los sub-criterios: Calidad (ED-SC6), Completitud (ED-SC7), Funcionalidad (ED-SC8), Costo de transformación (ED-SC9) y Datos riesgosos (ED-SC10), se muestran en el Tabla 5.77.

Tabla 5.77: Autovector máximo de cada sub-criterio del ED-UD para CV1

Sub-criterios ED-UD	Alternativas	WxP	Cociente	$\lambda_{máx}$
ED-SC6	CRISP-DM (A1)	1,05	2,97	3,02
	MoProPEI (A2)	1,83	3,08	
	P^3TQ (A3)	0,40	3,01	
ED-SC7	CRISP-DM (A1)	1,33	3,00	3,00
	MoProPEI (A2)	1,33	3,00	
	P^3TQ (A3)	0,44	3,00	
ED-SC8	CRISP-DM (A1)	0,79	2,97	3,02
	MoProPEI (A2)	2,09	3,02	
	P^3TQ (A3)	0,45	3,06	
ED-SC9	CRISP-DM (A1)	1,08	2,87	3,06
	MoProPEI (A2)	2,59	3,24	
	P^3TQ (A3)	0,34	3,08	
ED-SC10	CRISP-DM (A1)	1,20	2,93	3,03
	MoProPEI (A2)	2,05	3,15	
	P^3TQ (A3)	0,35	3,01	

Al evaluar la razón de consistencia de cada uno de las características, tal como se exponen en la Tabla 5.78; se puede ver que en el caso de ED-SC7 (Completitud de los datos) es perfectamente consistente, para las demás características también se logró una consistencia inferior al 10%, siendo todas ellas adecuadas.

Tabla 5.78: Evaluación de consistencia de los sub-criterios del ED-UD para CV1

Evaluación de consistencia	ED-SC6	ED-SC7	ED-SC8	ED-SC9	ED-SC10
$CI = (\lambda_{máx}$-n)/(n-1)	0,009	0,000	0,009	0,031	0,014
RI	0,525	0,525	0,525	0,525	0,525
$CR = CI/RI$	**0,006**	**0,000**	**0,006**	**0,021**	**0,009**

Pasando a la segunda rama de la estructura jerárquica (Figura 4.2) para la CN (Comprensión del Negocio) se evaluó inicialmente la consistencia de sus tres sub-criterios principales (Entorno del Negocio (CN-EN), Proyecto en el Entorno del Negocio (CN-PEN) y DS en el Proyecto del Negocio (CN-DSPN)) y luego en el desglose de cada uno de ellos.

En la Tabla 5.79 se muestra como se obtiene el $\lambda_{máx}$ en la comparación pareada de los sub-criterios de CN y cuál es el la razón de consistencia (CR) obtenida, cuyo valor indica que han sido totalmente coherentes las comparaciones y valores asignados, tal como se indica en la Tabla 5.80.

Tabla 5.79: Autovector máximo de cada sub-criterio de la CN

	WxP	**Cociente**
Entorno del Negocio (CN-EN)	0,60	3,00
Proyecto en el Entorno del Negocio (CN-PEN)	0,60	3,00
DS en el Proyecto del Negocio (CN-DSPN)	1,80	3,00
		$\lambda_{máx}$= 3,00

Tabla 5.80: Evaluación de consistencia de los sub-criterios de la CN

$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,000
$\boldsymbol{RI}$	0,525
$\boldsymbol{CR = CI/RI}$	**0,000**

Tabla 5.81: Autovector máximo del CN-EN

Sub-criterios CN-EN	**WxP**	**Cociente**
Situación económica circundante (CN-SC1)	0,37	7,11
Comprensión general de la organización (CN-SC2)	1,10	7,11
Definición de los objetivos del negocio (CN-SC3)	1,10	7,11
Definición del problema del negocio (CN-SC4)	1,10	7,11
Definición de las oportunidades del negocio (CN-SC5)	1,10	7,11
Definición de requisitos del negocio (CN-SC6)	0,98	7,06
Definición de criterios de éxito del proyecto (CN-SC7)	1,38	7,29
		$\lambda_{máx}$= 7,13

Siguiendo la rama izquierda de CN (ver Figura 4.2), el primer sub-criterio a evaluar es Entorno del Negocio (CN-EN), que se divide en siete características (Situación económica circundante (CN-SC1), Comprensión general de la organización (CN-SC2), Definición de los objetivos del negocio (CN-SC3), Definición del problema del negocio (CN-SC4), Definición de las oportunidades del negocio (CN-SC5), Definición de requisitos del negocio (CN-SC6) y

Definición de criterios de éxito del proyecto (CN-SC7)), para el cual se obtuvo un autovector de 7,13 y una razón de consistencia (CR) de 0,016, siendo consistente la matriz W (Tabla 5.50), tal como se indica en las Tablas 5.81 y 5.82.

Tabla 5.82: Evaluación de consistencia CN-EN

$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,022
$\boldsymbol{RI}$	1,341
$\boldsymbol{CR = CI/RI}$	**0,016**

Tabla 5.83: Autovector máximo de cada sub-criterio del CN-EN para CV1

Sub-criterios CN-EN	Alternativas	WxP	Cociente	$\boldsymbol{\lambda_{máx}}$
CN-SC1	CRISP-DM (A1)	0,49	3,00	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,89	3,01	
CN-SC2	CRISP-DM (A1)	1,99	3,20	3,11
	MoProPEI (A2)	0,69	3,08	
	P^3TQ (A3)	0,47	3,04	
CN-SC3	CRISP-DM (A1)	0,49	3,00	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,89	3,01	
CN-SC4	CRISP-DM (A1)	0,49	3,00	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,89	3,01	
CN-SC5	CRISP-DM (A1)	0,72	3,01	3,02
	MoProPEI (A2)	1,89	3,03	
	P^3TQ (A3)	0,41	3,01	
CN-SC6	CRISP-DM (A1)	0,60	3,00	3,00
	MoProPEI (A2)	1,80	3,00	
	P^3TQ (A3)	0,60	3,00	
CN-SC7	CRISP-DM (A1)	0,89	3,01	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,49	3,00	

Analizada la consistencia de las características de EN en forma conjunta (según la matriz de comparación pareada), se verificaron individualmente según nivel de cumplimiento de cada metodología de DS (alternativa); para lo cual en la Tabla 5.83 se presentan los valores obtenidos para el producto de W (matriz pareada) por P (vector de ponderaciones obtenidas para dicha matriz), el cociente P por WxP y así mediante el promedio de la columna se obtiene el autovector máximo $\lambda_{máx}$; cabe recordar que cuanto más cercano es este valor al orden de la matriz más acertados han sido las valoraciones.

Para la Definición de requisitos del negocio (CN-SC6) se puede notar que $\lambda_{máx}$ es igual al orden de la matriz (Tabla 5.83), por tanto su razón de consistencia es excelente, tal como se puede observar en la Tabla 5.84, donde también se expone la RC de las demás características, todas inferiores a 0,10, siendo éstas apropiadas.

Tabla 5.84: Evaluación de consistencia de los sub-criterios del CN-EN para CV1

Evaluación de consistencia	**CN-SC1**	**CN-SC2**	**CN-SC3**	**CN-SC4**	**CN-SC5**	**CN-SC6**	**CN-SC7**
$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,005	0,055	0,005	0,005	0,009	0,000	0,005
$\boldsymbol{RI}$	0,525	0,525	0,525	0,525	0,525	0,525	0,525
$\boldsymbol{CR = CI/RI}$	**0,009**	**0,104**	**0,009**	**0,009**	**0,017**	**0,000**	**0,009**

Tabla 5.85: Autovector máximo del CN-PEN

Sub-criterios CN-PEN	**WxP**	**Cociente**
Conocimiento de los RRHH del negocio (CN-SC8)	0,65	5,24
Disponibilidad de los RRHH del negocio (CN-SC9)	0,62	5,43
Riesgos del problema del negocio (CN-SC10)	1,23	5,26
Contingencias en relación a los riesgos del problema del negocio (CN-SC11)	1,85	5,43
Viabilidad del proyecto (CN-SC12)	1,02	5,40
		$\lambda_{máx}$= 5,35

El segundo sub-criterio de la CN, es el Proyecto en el Entorno del Negocio (CN-PEN), que se especifica en cinco características (Conocimiento de los RRHH del negocio (CN-SC8), Disponibilidad de los RRHH del negocio (CN-SC9), Riesgos del problema del negocio (CN-

SC10), Contingencias en relación a los riesgos del problema del negocio (CN-SC11) y Viabilidad del proyecto (CN-SC12)), cuyo $\lambda_{máx}$ es 5,35 (Tabla 5.85) y CR 0,079 (Tabla 5.86), por tanto los valores asignados en la matriz pareada correspondiente (Tabla 5.59) son consistentes.

Tabla 5.86: Evaluación de consistencia CN-PEN

$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,088
$\boldsymbol{RI}$	1,115
$\boldsymbol{CR = CI/RI}$	**0,079**

La Tabla 5.87 exhibe la obtención del autovector máximo las cinco características del sub-criterio CN-PEN, en la cual los Riesgos del problema del negocio (CN-SC10) y las Contingencias en relación a los mismos (CN-SC11) poseen el mismo orden que la matriz.

Tabla 5.87: Autovector máximo de cada sub-criterio del CN-PEN para CV1

Sub-criterios CN-PEN	Alternativas	WxP	Cociente	$\lambda_{máx}$
CN-SC8	CRISP-DM (A1)	0,49	3,00	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,89	3,01	
CN-SC9	CRISP-DM (A1)	0,49	3,00	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,89	3,01	
CN-SC10	CRISP-DM (A1)	0,60	3,00	3,00
	MoProPEI (A2)	1,80	3,00	
	P^3TQ (A3)	0,60	3,00	
CN-SC11	CRISP-DM (A1)	0,60	3,00	3,00
	MoProPEI (A2)	1,80	3,00	
	P^3TQ (A3)	0,60	3,00	
CN-SC12	CRISP-DM (A1)	0,97	3,03	3,02
	MoProPEI (A2)	1,69	3,01	
	P^3TQ (A3)	0,37	3,02	

Tabla 5.88: Evaluación de consistencia de los sub-criterios del CN-PEN para CV1

Evaluación de consistencia	CN-SC8	CN-SC9	CN-SC10	CN-SC11	CN-SC12
$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,005	0,005	0,000	0,000	0,009
$\boldsymbol{RI}$	0,525	0,525	0,525	0,525	0,525
$\boldsymbol{CR = CI/RI}$	**0,009**	**0,009**	**0,000**	**0,000**	**0,017**

A partir de los autovectores máximos obtenidos se estimó la razón de consistencia (CR) para cada una de las características mencionadas, visible en la Tabla 5.88; la cual refleja que CN-SC10 y CN-SC11 son totalmente coherentes y las demás características cumplen de manera adecuada.

Tabla 5.89: Autovector máximo del CN-DSPN

Sub-criterios CN- DSPN	WxP	Cociente
Objetivos de la aplicación de la metodología de DS (CN-SC13)	2,00	3,06
Conocimiento de los RRHH del proyecto de DS (CN-SC14)	0,48	3,01
RRHH disponibles del proyecto de DS (CN-SC15)	0,56	3,01
		$\lambda_{máx}$= 3,03

Finalmente, para el tercer sub-criterio de la CN (DS en el Proyecto del Negocio (CN-DSPN)), el autovector máximo obtenido es de 3,03 para las tres características, tal como se puede ver en la Tabla 5.89. En cuanto a la razón de consistencia, tal como se indica en la Tabla 5.90, es consistente (0,028) para la matriz de comparación pareada expuesta en la Tabla 5.66.

Tabla 5.90: Evaluación de consistencia CN-DSPN

$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,015
$\boldsymbol{RI}$	0,525
$\boldsymbol{CR = CI/RI}$	**0,028**

Al analizar las características Objetivos de la aplicación de la metodología de DS (CN-SC13), Conocimiento de los RRHH del proyecto de DS (CN-SC14) y RRHH disponibles del proyecto de DS (CN-SC15) en relación a las tres metodologías evaluadas, se puede ver en las Tablas 5.91 y 5.92 que son adecuadas las valoraciones realizada por el experto.

Tabla 5.91: Autovector máximo de cada sub-criterio del CN-DSPN para CV1

Sub-criterios CN- DSPN	Alternativas	WxP	Cociente	$\lambda_{máx}$
CN-SC13	CRISP-DM (A1)	0,89	3,01	3,01
	MoProPEI (A2)	1,62	3,01	
	P^{3}TQ (A3)	0,49	3,00	
CN-SC14	CRISP-DM (A1)	0,72	3,01	3,02
	MoProPEI (A2)	1,89	3,03	
	P^{3}TQ (A3)	0,41	3,01	
CN-SC15	CRISP-DM (A1)	0,72	3,01	3,02
	MoProPEI (A2)	1,89	3,03	
	P^{3}TQ (A3)	0,41	3,01	

Tabla 5.92: Evaluación de consistencia de los sub-criterios del CN-DSPN para CV1

Evaluación de consistencia	CN-SC13	CN-SC14	CN-SC15
$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,005	0,009	0,009
$\boldsymbol{RI}$	0,525	0,525	0,525
$\boldsymbol{CR = CI/RI}$	**0,009**	**0,017**	**0,017**

5.2.4. Desarrollo de los métodos de MCDM para el CV1

Como se señaló en el Capítulo precedente (IV), el desarrollo de los métodos de MCDM, se divide en dos, por un lado se completa el desarrollo del método AHP y por otro FAHP, tal como se puede notar en la Figura 4.4.

5.2.4.1 Desarrollo de AHP para el CV1

Continuando con los pasos del método AHP (ver Pasos 3 y 4 del Anexo I), se establecieron las prioridades locales y globales, a través del cálculo del peso relativo a los criterios por nivel; las locales corresponden a las ponderaciones sacadas en el paso 5.2.2, a partir de las cuales se obtuvieron las globales. Finalmente, mediante el método de la suma ponderada, se calcularon las prioridades totales asociadas a cada alternativa.

El establecimiento de la prioridad local del Acceso a los Datos (ED-AD) se exponen en la Tabla 5.93, donde se encuentran las ponderaciones de los sub-criterio ED-SC1 a ED-SC5 (ver Tablas 5.36 a 5.37) en relación a cada alternativa y en la última fila la ponderación general de ED-AD (ver Tabla 5.35), a partir de los cuales se obtuvo la prioridad local de ED-AD, para el cual MoProPEI es la metodología preferida con una prioridad de 0,48, seguido de P^3TQ con 0,30 y por último CRISP-DM con 0,22.

En cuanto al Uso de los Datos (ED-UD), la prioridad local calculada se muestra en la Tabla 5.94, donde nuevamente la metodología MoProPEI es la selecta con una prioridad del 50%, por sobre el 26% de P^3TQ y 24% de CRISP-DM.

Tabla 5.93: Prioridad local de ED-AD para CV1

Sub-criterio/ Alternativa	ED-SC1	ED-SC2	ED-SC3	ED-SC4	ED-SC5	Prioridad Local
CRISP-DM (A1)	0,10	0,11	0,54	0,12	0,16	0,22
MoProPEI (A2)	0,57	0,63	0,16	0,56	0,54	0,48
P^3TQ (A3)	0,33	0,26	0,30	0,32	0,30	0,30
Ponderación ED-AD	0,07	0,24	0,25	0,37	0,07	

Tabla 5.94: Prioridad local de ED-UD para CV1

Sub-criterio/ Alternativa	ED-SC6	ED-SC7	ED-SC8	ED-SC9	ED-SC10	Prioridad Local
CRISP-DM (A1)	0,11	0,11	0,14	0,43	0,62	0,24
MoProPEI (A2)	0,63	0,63	0,62	0,14	0,14	0,50
P^3TQ (A3)	0,26	0,26	0,24	0,43	0,24	0,26
Ponderación ED-UD	0,33	0,27	0,13	0,05	0,21	

A partir de las prioridades locales para los sub-criterios ED-AC y ED-UD expuestas en las Tablas 5.93 y 5.94, se calculó la global para cada una de ellas, para la cual se tuvo en cuenta que según los expertos el uso de los datos (ED-UD) es más importante que el acceso a los mismos (ED-AD) tal como se indica en la tercer columna de las Tablas 5.95 y 5.96

Tabla 5.95: Prioridad global de ED-AD para CV1

Sub-criterio/ Alternativa	Prioridad Local	Ponderación ED-AD	Prioridad Global
CRISP-DM (A1)	0,22	0,25	0,06
MoProPEI (A2)	0,48		0,12
P^{3}TQ (A3)	0,30		0,07

Tabla 5.96: Prioridad global de ED-UD para CV1

Sub-criterio/ Alternativa	Prioridad Local	Ponderación ED-UD	Prioridad Global
CRISP-DM (A1)	0,24	0,75	0,18
MoProPEI (A2)	0,50		0,38
P^{3}TQ (A3)	0,26		0,20

Teniendo en cuenta que el Entendimiento de los Datos representa el 50% de importancia al momento de seleccionar o desarrollar un proyecto de DS, la ponderación global de dicho criterio es de 25% para MoProPEI, 14% para P^{3}TQ y del 12% para CRISP-DM. De forma aislada, la rama izquierda de la jerarquía alcanzó los porcentajes presentados en la Figura 5.1 (expresados en un 100%).

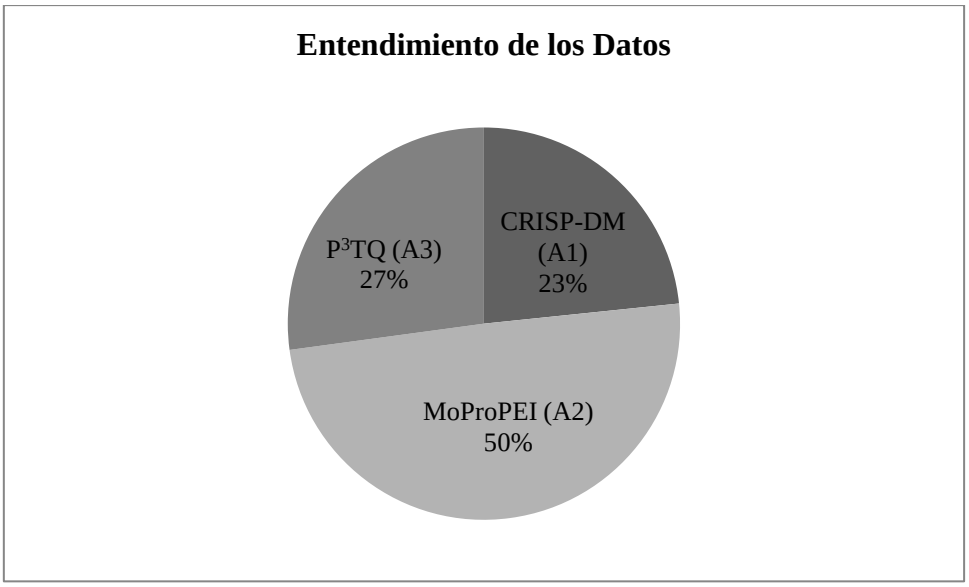

Figura 5.1: Ponderaciones globales del ED para el CV1 con AHP

Para los sub-criterios de la CN se realizó el mismo procedimiento que para ED. Primeramente para CN-EN (Entorno del Negocio), se obtuvieron las prioridades locales disponibles en la Tabla 5.97, donde se representa las ponderaciones de los siete sub-criterio (CN-SC1 a CN-SC7) en relación a cada alternativa y la ponderación general de CN-EN (última fila); como se puede notar en la última columna, MoProPEI es la metodología elegida con una prioridad de 0,51, seguido de CRISP-DM con 0,28 y en último lugar P^3TQ con 0,21.

Tabla 5.97: Prioridad local de CN-EN para CV1

Sub-criterio/ Alternativa	CN-SC1	CN-SC2	CN-SC3	CN-SC4	CN-SC5	CN-SC6	CN-SC7	Prioridad Local
CRISP-DM (A1)	0,16	0,62	0,16	0,16	0,24	0,20	0,30	0,28
MoProPEI (A2)	0,54	0,22	0,54	0,54	0,62	0,60	0,54	0,51
P^3TQ (A3)	0,30	0,16	0,30	0,30	0,14	0,20	0,16	0,21
Ponderación CN-EN	0,05	0,16	0,16	0,16	0,16	0,14	0,19	

Para el sub-criterio CN-PEN (Proyecto en el Entorno del Negocio), la prioridad local alcanzada para sus cinco características (CN-SC8 a CN-SC12) se expone en la Tabla 5.98, en la cual se aprecia que nuevamente MoProPEI es la metodología seleccionada para este sub-criterio y sus características, con una prioridad del 58% y 21% para las dos metodologías restantes.

Tabla 5.98: Prioridad local de CN-PEN para CV1

Sub-criterio/ Alternativa	CN-SC8	CN-SC9	CN-SC10	CN-SC11	CN-SC12	Prioridad Local
CRISP-DM (A1)	0,16	0,16	0,20	0,20	0,32	0,21
MoProPEI (A2)	0,54	0,54	0,60	0,60	0,56	0,58
P^3TQ (A3)	0,30	0,30	0,20	0,20	0,12	0,21
Ponderación CN-PEN	0,12	0,11	0,23	0,34	0,19	

Y en el caso de CN-DSPN (Data Science en el Proyecto del Negocio), las prioridades locales son de 0,57 para MoProPEI, de 0,28 para CRISP-DM y 0,15 para P^3TQ, tal como se puede visualizar en la Tabla 5.99.

Tabla 5.99: Prioridad local de CN-DSPN para CV1

Sub-criterio/ Alternativa	CN-SC13	CN-SC14	CN-SC15	Prioridad Local
CRISP-DM (A1)	0,30	0,24	0,24	0,28
MoProPEI (A2)	0,54	0,62	0,62	0,57
P^3TQ (A3)	0,16	0,14	0,14	0,15
Ponderación CN-DSPN	0,66	0,16	0,19	

Tabla 5.100: Prioridad global de CN-EN para CV1

Sub-criterio/ Alternativa	Prioridad Local	Ponderación CN-EN	Prioridad Global
CRISP-DM (A1)	0,28	0,20	0,06
MoProPEI (A2)	0,51		0,10
P^3TQ (A3)	0,21		0,04

Tabla 5.101: Prioridad global de CN-PEN para CV1

Sub-criterio/ Alternativa	Prioridad Local	Ponderación CN-PEN	Prioridad Global
CRISP-DM (A1)	0,21	0,20	0,04
MoProPEI (A2)	0,58		0,12
P^3TQ (A3)	0,21		0,04

Paso siguiente, se procedió a calcular las prioridades globales de los tres sub-criterios mencionados, en la Tabla 5.100 para CN-EN, en la Tabla 5.101 para CN-PEN y en la Tabla 5.102 para CN-DSPN. Aquí se debe tener en cuenta las preferencias de los expertos para este nivel dentro de la CN, donde los dos primeros sub-criterios poseen una ponderación del 20% y el último (DSPN) del 60%, por tanto en cada tabla la suma de las prioridades globales dará

los porcentajes mencionados y la suma de la cuarta columna de las tres tablas, alcanza el 100% para el criterio Comprensión del Negocio.

Tabla 5.102: Prioridad global de CN-DSPN para CV1

Sub-criterio/ Alternativa	Prioridad Local	Ponderación CN-DSPN	Prioridad Global
CRISP-DM (A1)	0,28	0,60	0,17
MoProPEI (A2)	0,57		0,34
P^{3}TQ (A3)	0,15		0,09

La Comprensión del Negocio, al igual que el Entendimiento de los Datos representa el 50% de importancia al momento de seleccionar una metodología de DS. Una vez obtenidas las prioridades locales y globales en niveles inferiores se calculó la ponderación global de dicho criterio para el cual, nuevamente MoProPEI es la más adecuada con el 28%, en segundo lugar ahora se encuentra CRISP-DM con el 13% y en último lugar P^{3}TQ con el 9%. Al analizar la rama derecha de la estructura jerárquica, la CN obtuvo los porcentajes visibles en la Figura 5.2 (reflejado en un 100%).

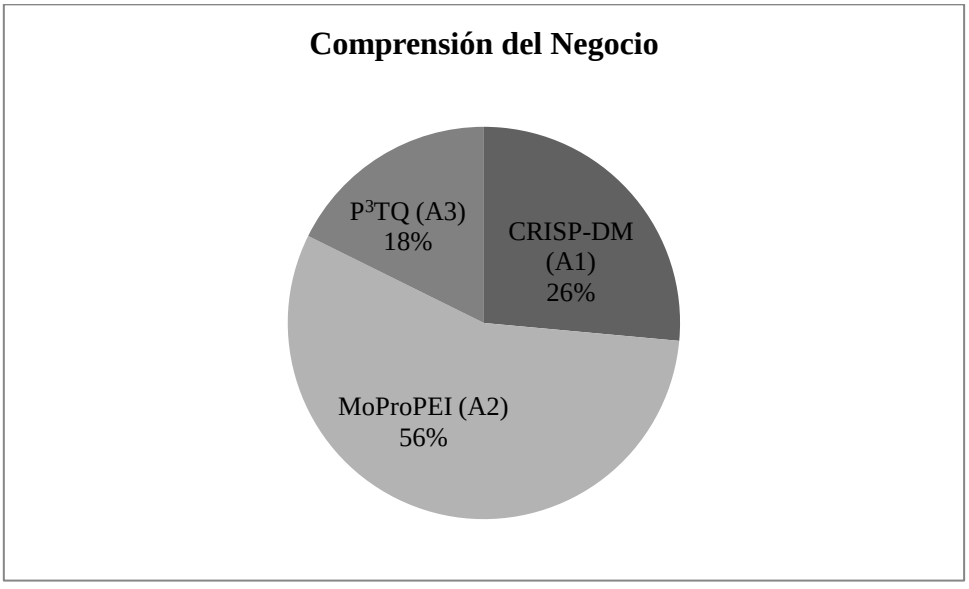

Figura 5.2: Ponderaciones globales de la CN para el CV1 con AHP

Finalmente se realizó el establecimiento de prioridades totales en relación a cada metodología, a partir de las prioridades locales y globales recientemente expuestas. Teniendo en cuenta que

cada criterio del segundo nivel (ED y CN) representa el 50% de trascendencia al momento de desarrollar un proyecto de DS, las prioridades totales obtenidas para el caso de validación 1 (CV1) se exponen en la Tabla 103, donde claramente se denota una significativa adecuación para el caso de la metodología MoProPEI (A2), dado que consiguió una primacía del 53%, quedando en segundo lugar CRISP-DM (A1) con el 25% y postrero P^3TQ (A3) con el 22%.

Tabla 5.103: Prioridades totales con AHP para el CV1

Alternativas	Prioridad Final ED	Prioridad Final CN	Prioridad Total
CRISP-DM (A1)	0,12	0,13	**0,25**
MoProPEI (A2)	0,25	0,28	**0,53**
P^3TQ (A3)	0,14	0,09	**0,22**

5.2.4.2 Desarrollo de FAHP para el CV1

Para el desarrollo del método FAHP se siguió el paso 6.2 del modelo propuesto en el Capítulo IV, el cual se apoya en los pasos descriptos en el Anexo II. Donde, se mencionó que la función de membrecía con números difusos triangulares, es la utilizada para el establecimiento de prioridades entre los criterios y sub-criterios.

Como se indicó en el paso 3 del modelo (Capítulo IV), para el presente trabajo, se utilizó como valor de entrada de m (valor más prometedor) a las matrices de dos entradas y las de tipo parrillas, que fueron adecuadas inicialmente a la escala Saaty, para luego seguir el mismo procedimiento indicado en el primer párrafo de la sección 5.2.2 del actual capitulo. Los valores de l (valor pesimista) y u (valor optimista), se obtuvieron a partir de la escala de conversión difusa triangular visible en la Tabla II.I del Anexo II.

Primeramente en la Tabla 5.104 se expone la matriz triangular difusa de comparación pareada del nivel dos, donde se comparan el Entendimiento de los Datos (ED) y la Comprensión del Negocio (CN), los cuales poseen la misma importancia para los expertos (intersección ED y CN en m) .

En el nivel 3, para los dos sub-criterios del ED y CN, y sus características se realiza el mismo procedimiento. En la Tabla 5.105, se exhibe la matriz triangular difusa de comparación pareada del nivel tres para ED, donde se aprecia que ED-UD es más relevante que ED-AD.

Tabla 5.104: Matriz triangular difusa de comparación pareada del nivel 2

	ED			**CN**		
	l	**m**	**u**	**l**	**m**	**u**
Entendimiento de los Datos (ED)	1	1	1	1	1	2
Comprensión del Negocio (CN)	1/2	1/1	1/1	1	1	1
Suma	1,50	2,00	2,00	2,00	2,00	3,00

Tabla 5.105: Matriz triangular difusa de comparación pareada del nivel 3 para ED

	ED-AD			**ED-UD**		
	l	**m**	**u**	**l**	**m**	**u**
Acceso a los Datos (ED-AD)	1	1	1	1/4	1/3	1/2
Uso de los Datos (ED-UD)	2	3	4	1	1	1
Suma	3,00	4,00	5,00	1,25	1,33	1,50

Tabla 5.106: Matriz triangular difusa de comparación pareada del nivel 4 para ED-AD

	ED-SC1			**ED-SC2**			**ED-SC3**			**ED-SC4**			**ED-SC5**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC1	1	1	1	1/8	1/7	1/6	1/4	1/3	1/2	1/6	1/5	1/4	1	1	2
ED-SC2	6	7	8	1	1	1	1	1	2	1/4	1/3	1/2	2	3	4
ED-SC3	2	3	4	1/2	1	1	1	1	1	1	1	2	2	3	4
ED-SC4	4	5	6	2	3	4	1/2	1	1	1	1	1	4	5	6
ED-SC5	1/2	1	1	1/4	1/3	1/2	1/4	1/3	1/2	1/6	1/5	1/4	1	1	1
Suma	13,50	17,00	20,00	3,88	5,48	6,67	3,00	3,67	5,00	2,58	2,73	4,00	10,00	13,00	17,00

Luego se prosiguió con el desglose de las características de los sub-criterios acceso y uso de los datos (AD y UD). En la Tabla 5.106 se muestra la matriz triangular difusa para

comparación pareada del ED-AD respecto al nivel cuatro, para sus cinco características, la última fila de tabla corresponde a la suma por columna, la cual se utiliza en el paso siguiente (cálculo del valor de extensión sintética difusa).

Desde la Tabla 5.107 a la 5.111, se encuentran las matrices de comparación pareada con función de membrecía difusa triangular de cada característica del AD, donde se aprecia para el CV1 la tendencia de preferencia hacia la alternativa 2 (MoProPEI), excepto para el sub-criterio ED-SC3, que corresponde al sub-criterio Diversidad de los datos.

Tabla 5.107: Matriz triangular difusa de comparación pareada de ED-SC1 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/6	1/5	1/4	1/5	1/4	1/3
MoProPEI (A2)	4	5	6	1	1	1	1	2	3
P^{3}TQ (A3)	3	4	5	1/3	1/2	1	1	1	1
Suma	8,00	10,00	12,00	1,50	1,70	2,25	2,20	3,25	4,33

Tabla 5.108: Matriz triangular difusa de comparación pareada de ED-SC2 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/6	1/5	1/4	1/4	1/3	1/2
MoProPEI (A2)	4	5	6	1	1	1	2	3	4
P^{3}TQ (A3)	2	3	4	1/4	1/3	1/2	1	1	1
Suma	7,00	9,00	11,00	1,00	1,53	1,75	3,25	4,33	5,50

Tabla 5.109: Matriz triangular difusa de comparación pareada de ED-SC3 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	2	3	4	1	2	3
MoProPEI (A2)	1/4	1/3	1/2	1	1	1	1/3	1/2	1
P^{3}TQ (A3)	1/3	1/2	1	1	2	3	1	1	1
Suma	1,58	1,83	2,50	4,00	6,00	8,00	2,33	3,50	5,00

Tabla 5.110: Matriz triangular difusa de comparación pareada de ED-SC4 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/5	1/4	1/3	1/4	1/3	1/2
MoProPEI (A2)	3	4	5	1	1	1	1	2	3
P^{3}TQ (A3)	2	3	4	1/3	1/2	1	1	1	1
Suma	6,00	8,00	10,00	1,53	1,75	2,33	2,25	3,33	4,50

Tabla 5.111: Matriz triangular difusa de comparación pareada de ED-SC5 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/3	1/2	1
MoProPEI (A2)	2	3	4	1	1	1	1	2	3
P^{3}TQ (A3)	1	2	3	1/3	1/2	1	1	1	1
Suma	4,00	6,00	8,00	1,58	1,83	2,50	2,33	3,50	5,00

En cuanto al Uso de los Datos (ED-UD), la matriz triangular difusa alcanzada al comparar sus cinco sub-criterios se presenta en la Tabla 5.112.

Tabla 5.112: Matriz triangular difusa de comparación pareada del nivel 4 para ED-UD

	ED-SC6			ED-SC7			ED-SC8			ED-SC9			ED-SC10		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC6	1	1	1	1	1	2	4	5	6	6	7	8	1	1	2
ED-SC7	1/2	1	1	1	1	1	2	3	4	4	5	6	1	1	2
ED-SC8	1/6	1/5	1/4	1/4	1/3	1/2	1	1	1	2	3	4	1	1	2
ED-SC9	1/8	1/7	1/6	1/6	1/5	1/4	1/4	1/3	1/2	1	1	1	1/4	1/3	1/2
ED-SC10	1/2	1	1	1/2	1	1	1/2	1	1	2	3	4	1	1	1
Suma	2,29	3,34	3,42	2,92	3,53	4,75	7,75	10,33	12,50	15,00	19,00	23,00	4,25	4,33	7,50

Posteriormente en la Tabla 5.113 a la Tabla 5.117, se encuentran los mismos sub-criterios evaluados en relación a la completitud de los mismos respecto a cada metodología.
Para los tres primeros sub-criterios (Calidad, Completitud y Funcionalidad de los datos) la preferencia es en torno a MoProPEI y para los dos últimos (Costo de transformación de los datos y la identificación de datos riesgosos) a CRISP-DM.

Tabla 5.113: Matriz triangular difusa de comparación pareada de ED-SC6 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/6	1/5	1/4	1/4	1/3	1/2
MoProPEI (A2)	4	5	6	1	1	1	2	3	4
P^3TQ (A3)	2	3	4	1/4	1/3	1/2	1	1	1
Suma	7,00	9,00	11,00	1,42	1,53	1,75	3,25	4,33	5,50

Tabla 5.114: Matriz triangular difusa de comparación pareada de ED-SC7 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/6	1/5	1/4	1/4	1/3	1/2
MoProPEI (A2)	4	5	6	1	1	1	2	3	4
P^3TQ (A3)	2	3	4	1/4	1/3	1/2	1	1	1
Suma	7,00	9,00	11,00	1,42	1,53	1,75	3,25	4,33	5,50

Tabla 5.115: Matriz triangular difusa de comparación pareada de ED-SC8 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/5	1/4	1/3	1/3	1/2	1
MoProPEI (A2)	3	4	5	1	1	1	2	3	4
P^3TQ (A3)	1	2	3	1/4	1/3	1/2	1	1	1
Suma	5,00	7,00	9,00	1,45	1,58	1,83	3,33	4,50	6,00

Tabla 5.116: Matriz triangular difusa de comparación pareada de ED-SC9 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	2	3	4	1	1	2
MoProPEI (A2)	1/4	1/3	1/2	1	1	1	1/4	1/3	1/2
P^{3}TQ (A3)	1/2	1	1	2	3	4	1	1	1
Suma	1,75	2,33	2,50	5,00	7,00	9,00	2,25	2,33	3,50

Tabla 5.117: Matriz triangular difusa de comparación pareada de ED-SC10 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	3	4	5	2	3	4
MoProPEI (A2)	1/5	1/4	1/3	1	1	1	1/3	1/2	1
P^{3}TQ (A3)	1/4	1/3	1/2	1	2	3	1	1	1
Suma	1,45	1,58	1,83	5,00	7,00	9,00	3,33	4,50	6,00

Al tomar la rama derecha de la jerarquía visible en la Figura 4.2, la Comprensión del Negocio (CN) posee tres sub-criterios hijos, cuya matriz de comparación difusa triangular se enseña en la Tabla 5.118.

En el caso del sub-criterio Entorno del Negocio (CN-EN), se especifican siete características, que al compararlas entre sí y representarlas como una matriz de comparación pareada difusa triangular, se obtuvieron los valores indicados en la Tabla 5.119.

Al comparar cada una de estas características (Tabla 5.120 a la 5.126), se puede notar que para la característica que refiere a la Comprensión general de la organización (CN-SC2) mejor se adapta la metodología CRISP-DM, seguido de MoProPEI (ver Tabla 5.121), quién para los demás sub-criterios es la que mayor valoración tiene, dado que el experto considera que tiene mayor consideración respecto a esas características, tal como se puede ver en la Tabla 5.120 y en las Tablas 5.122 a la Tabla 5.126.

Tabla 5.118: Matriz triangular difusa de comparación pareada del nivel 3 para CN

	CN-EN			CN-PEN			CN-DSPN		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
Entorno del Negocio (CN-EN)	1	1	1	1	1	2	1/4	1/3	1/2
Proyecto en el Entorno del Negocio (CN-PEN)	1/2	1	1	1	1	1	1/4	1/3	1/2
DS en el Proyecto del Negocio (CN-DSPN)	2	3	4	2	3	4	1	1	1
Suma	3,50	5,00	6,00	4,00	5,00	7,00	1,50	1,67	2,00

Tabla 5.119: Matriz triangular difusa de comparación pareada del nivel 4 para CN-EN

	CN-SC1			CN-SC2			CN-SC3			CN-SC4			CN-SC5			CN-SC6			CN-SC7		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC1	1	1	1	1/4	1/3	1/2	1/4	1/3	1/2	1/4	1/3	1/2	1/4	1/3	1/2	1/4	1/3	1/2	1/4	1/3	1/2
CN-SC2	2	3	4	1	1	1	1	1	2	1	1	2	1	1	2	1	1	2	1	1	2
CN-SC3	2	3	4	1/2	1	1	1	1	1	1	1	2	1	1	2	1	1	2	1	1	2
CN-SC4	2	3	4	1/2	1	1	1/2	1	1	1	1	1	1	1	2	1	1	2	1	1	2
CN-SC5	2	3	4	1/2	1	1	1/2	1	1	1/2	1	1	1	1	1	1	1	2	1	1	2
CN-SC6	2	3	4	1/2	1	1	1/2	1	1	1/2	1	1	1/2	1	1	1	1	1	1/4	1/3	1/2
CN-SC7	2	3	4	1/2	1	1	1/2	1	1	1/2	1	1	1/2	1	1	2	3	4	1	1	1
Suma	13,0	19,0	25,0	3,75	6,33	6,50	4,25	6,33	7,50	4,75	6,33	8,50	5,25	6,33	9,50	7,25	8,33	13,5	5,50	5,67	10,0

Tabla 5.120: Matriz triangular difusa de comparación pareada de CN-SC1 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/3	1/2	1
MoProPEI (A2)	2	3	4	1	1	1	1	2	3
P^{3}TQ (A3)	1	2	3	1/3	1/2	1	1	1	1
Suma	4,00	6,00	8,00	1,58	1,83	2,50	2,33	3,50	5,00

Tabla 5.121: Matriz triangular difusa de comparación pareada de CN-SC2 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	3	4	5	2	3	4
MoProPEI (A2)	1/5	1/4	1/3	1	1	1	1	2	3
P^{3}TQ (A3)	1/4	1/3	1/2	1/3	1/2	1	1	1	1
Suma	1,45	1,58	1,83	4,33	5,50	7,00	4,00	6,00	8,00

Tabla 5.122: Matriz triangular difusa de comparación pareada de CN-SC3 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/3	1/2	1
MoProPEI (A2)	2	3	4	1	1	1	1	2	3
P^{3}TQ (A3)	1	2	3	1/3	1/2	1	1	1	1
Suma	4,00	6,00	8,00	1,58	1,83	2,50	2,33	3,50	5,00

Tabla 5.123: Matriz triangular difusa de comparación pareada de CN-SC4 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/3	1/2	1
MoProPEI (A2)	2	3	4	1	1	1	1	2	3
P^{3}TQ (A3)	1	2	3	1/3	1/2	1	1	1	1
Suma	4,00	6,00	8,00	1,58	1,83	2,50	2,33	3,50	5,00

Tabla 5.124: Matriz triangular difusa de comparación pareada de CN-SC5 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1	2	3
MoProPEI (A2)	2	3	4	1	1	1	3	4	5
P^{3}TQ (A3)	1/3	1/2	1	1/5	1/4	1/3	1	1	1
Suma	3,33	4,50	6,00	1,45	1,58	1,83	5,00	7,00	9,00

Tabla 5.125: Matriz triangular difusa de comparación pareada de CN-SC6 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1	1	2
MoProPEI (A2)	2	3	4	1	1	1	2	3	4
P^{3}TQ (A3)	1/2	1	1	1/4	1/3	1/2	1	1	1
Suma	3,50	5,00	6,00	1,50	1,67	2,00	4,00	5,00	7,00

Tabla 5.126: Matriz triangular difusa de comparación pareada de CN-SC7 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	1	2	3
MoProPEI (A2)	1	2	3	1	1	1	2	3	4
P^{3}TQ (A3)	1/3	1/2	1	1/4	1/3	1/2	1	1	1
Suma	2,33	3,50	5,00	1,58	1,83	2,50	4,00	6,00	8,00

En el cuarto nivel de la estructura jerárquica propuesta, el sub-criterio CN-PEN (Proyecto en el Entorno del Negocio), se divide en cinco características, cuyos valores pesimistas (l), prometedores (m) y optimistas (u), al compararlas entre sí, se encuentran representadas en la Tabla 5.127.

Para las cinco características de CN-PEN, puede percatarse la inclinación clara que hay hacia la segunda alternativa (MoProPEI), tal como se muestran en las Tablas 5.128 a la 5.132.

Tabla 5.127: Matriz triangular difusa de comparación pareada del nivel 4 para CN-PEN

	CN-SC8			**CN-SC9**			**CN-SC10**			**CN-SC11**			**CN-SC12**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC8	1	1	1	1	1	2	1	1	2	1/4	1/3	1/2	1/4	1/3	1/2
CN - SC9	1/2	1	1	1	1	1	1/4	1/3	1/2	1/4	1/3	1/2	1	1	2
CN - SC10	1/2	1	1	2	3	4	1	1	1	1	1	2	1	1	2
CN - SC11	2	3	4	2	3	4	1/2	1	1	1	1	1	2	3	4
CN - SC12	2	3	4	1/2	1	1	1/2	1	1	1/4	1/3	1/2	1	1	1
Suma	6,00	9,00	11,0	6,50	9,00	12,0	3,25	4,33	5,50	2,75	3,00	4,50	5,25	6,33	9,50

Tabla 5.128: Matriz triangular difusa de comparación pareada a de CN-SC8 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/3	1/2	1
MoProPEI (A2)	2	3	4	1	1	1	1	2	3
P^3TQ (A3)	1	2	3	1/3	1/2	1	1	1	1
Suma	4,00	6,00	8,00	1,58	1,83	2,50	2,33	3,50	5,00

Tabla 5.129: Matriz triangular difusa de comparación pareada de CN-SC9 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/3	1/2	1
MoProPEI (A2)	2	3	4	1	1	1	1	2	3
P^3TQ (A3)	1	2	3	1/3	1/2	1	1	1	1
Suma	4,00	6,00	8,00	1,58	1,83	2,50	2,33	3,50	5,00

Tabla 5.130: Matriz triangular difusa de comparación pareada de CN-SC10 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1	1	2
MoProPEI (A2)	2	3	4	1	1	1	2	3	4
P^3TQ (A3)	1/2	1	1	1/4	1/3	1/2	1	1	1
Suma	3,50	5,00	6,00	1,50	1,67	2,00	4,00	5,00	7,00

Tabla 5.131: Matriz triangular difusa de comparación pareada de CN-SC11 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1	1	2
MoProPEI (A2)	2	3	4	1	1	1	2	3	4
P^3TQ (A3)	1/2	1	1	1/4	1/3	1/2	1	1	1
Suma	3,50	5,00	6,00	1,50	1,67	2,00	4,00	5,00	7,00

Tabla 5.132: Matriz triangular difusa de comparación pareada de CN-SC12 para CV1

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	2	3	4
MoProPEI (A2)	1	2	3	1	1	1	3	4	5
P^3TQ (A3)	1/4	1/3	1/2	1/5	1/4	1/3	1	1	1
Suma	2,25	3,33	4,50	1,53	1,75	2,33	6,00	8,00	10,00

Finalmente en la Tabla 5.133 se encuentra la valoración asignada por los expertos, en escala triangular difusa para la comparación de los tres sub-criterios de DSPN y la comparación de los mismos en relación al nivel de cumplimiento de los mismos en cada metodología se exhiben en las Tablas 5.134 a 5.136, nuevamente con una clara preferencia hacia MoProPEI en el caso de validación en cuestión (CV1).

Tabla 5.133: Matriz triangular difusa de comparación pareada del nivel 4 para CN-DSPN

	CN-SC13			**CN-SC14**			**CN-SC15**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC13	1	1	1	4	5	6	2	3	4
CN-SC14	1/6	1/5	¼	1	1	1	1	1	2
CN-SC15	1/4	1/3	½	1/2	1	1	1	1	1
Suma	1,42	1,53	1,75	5,50	7,00	8,00	4,00	5,00	7,00

Tabla 5.134: Matriz triangular difusa de comparación pareada de CN-SC13 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	1	2	3
MoProPEI (A2)	1	2	3	1	1	1	2	3	4
P^{3}TQ (A3)	1/3	1/2	1	1/4	1/3	1/2	1	1	1
Suma	2,33	3,50	5,00	1,58	1,83	2,50	4,00	6,00	8,00

Tabla 5.135: Matriz triangular difusa de comparación pareada de CN-SC14 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/3	1/2	1
MoProPEI (A2)	2	3	4	1	1	1	1	2	3
P^{3}TQ (A3)	1	2	3	1/3	1/2	1	1	1	1
Suma	4,00	6,00	8,00	1,58	1,83	2,50	2,33	3,50	5,00

Tabla 5.136: Matriz triangular difusa de comparación pareada de CN-SC15 para CV1

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1	2	3
MoProPEI (A2)	2	3	4	1	1	1	3	4	5
P^{3}TQ (A3)	1/3	1/2	1	1/5	1/4	1/3	1	1	1
Suma	3,33	4,50	6,00	1,45	1,58	1,83	5,00	7,00	9,00

Continuando con los pasos del modelo propuesto y de FAHP, se calculó el valor de extensión sintética difusa para cada criterio y sub-criterio (ver paso 3 del Anexo II).

Bajo la Ecuación II.III de las siguientes tablas, se encuentran los valores de extensión sintética difusa obtenida a partir de cada matriz de comparación pareada difusa triangular, debajo de la misma, en la columna cuyo encabezado es m (valor más prometedor), se representan las prioridades del nivel (bajo un mismo criterio padre); similares a las prioridades locales en AHP.

En la Tabla 5.137, se puede visualizar para los criterios principales (nivel 2), el valor de extensión sintética difusa (Ecuación II.III), que deriva de las dos columnas anteriores (Ecuaciones II.IV y II.V). La preferencia en este nivel es del 50% para cada criterio.

Tabla 5.137: Extensión sintética difusa del nivel 2

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
Entendimiento de los Datos	2,00	2,00	3,00	3,50	4,00	5,00	0,40	0,50	0,86
Comprensión del Negocio	1,50	2,00	2,00				0,30	0,50	0,57

Para la rama siniestra, se observa en la Tabla 5.138 los valores de extensión sintética difusa al comparar el acceso a los datos (ED-AD) y el uso de los mismos (ED-UD), donde el segundo es más relevante para los expertos que el primero.

Tabla 5.138: Extensión sintética difusa del nivel 3 para ED

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
Acceso a los Datos (ED-AD)	1,25	1,33	1,50	4,25	5,33	6,50	0,19	0,25	0,35
Uso de los Datos (ED-UD)	3,00	4,00	5,00				0,46	0,75	1,18

Seguidamente en la Tabla 5.139 se muestran el valor de extensión sintética difusa de los sub-criterios comparados bajo el criterio Acceso a los Datos (ED-AD). Se puede notar que ED-

SC4 (Diversidad de las fuentes de datos) fue considerado más relevante que los demás sub-criterios y el menos importante es el ED-SC1 (Portabilidad); tal como se indica en los valores pesimistas (l), prometedores (m) y optimistas (u) bajo el encabezado Ecuación II.IV y Ecuación II.III.

Tabla 5.139: Extensión sintética difusa del nivel 3 para ED-AD

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC1	2,54	2,68	3,92	32,96	41,88	52,67	0,05	0,06	0,12
ED-SC2	10,25	12,33	15,50				0,19	0,29	0,47
ED-SC3	6,50	9,00	12,00				0,12	0,21	0,36
ED-SC4	11,50	15,00	18,00				0,22	0,36	0,55
ED-SC5	2,17	2,87	3,25				0,04	0,07	0,10

Tabla 5.140: Extensión sintética difusa de Sub-criterios ED-AD para V1

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC1									
CRISP-DM	1,37	1,45	1,58	11,70	14,95	18,58	0,07	0,10	0,14
MoProPEI	6,00	8,00	10,00				0,32	0,53	0,85
P^{3}TQ	4,33	5,50	7,00				0,23	0,37	0,60
ED-SC2									
CRISP-DM	1,42	1,53	1,75	11,25	14,87	18,25	0,08	0,10	0,16
MoProPEI	7,00	9,00	11,00				0,38	0,61	0,98
P^{3}TQ	3,25	4,33	5,50				0,18	0,29	0,49
ED-SC3									
CRISP-DM	4,00	6,00	8,00	7,92	11,33	15,50	0,26	0,53	1,01
MoProPEI	1,58	1,83	2,50				0,10	0,16	0,32
P^{3}TQ	2,33	3,50	5,00				0,15	0,31	0,63
ED-SC4									
CRISP-DM	1,45	1,58	1,83	9,78	13,08	16,83	0,09	0,12	0,19
MoProPEI	5,00	7,00	9,00				0,30	0,54	0,92
P^{3}TQ	3,33	4,50	6,00				0,20	0,34	0,61

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC5									
CRISP-DM	1,58	1,83	2,50	7,92	11,33	15,50	0,10	0,16	0,32
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	2,33	3,50	5,00				0,15	0,31	0,63

La Tabla 5.140 agrupa las cinco características (sub-criterios) del Acceso a los Datos, en comparación la importancia de los mismos dentro de las metodologías evaluadas; de las cuales MoProPEI es favorecida para la mayoría de las características, excepto para la Accesibilidad de los datos (ED-SC3), donde se recomienda a CRISP-DM, en el caso de P^{3}TQ siempre se encuentra en segundo lugar, independientemente de la característica.

El valor de extensión sintética difusa para ED-UD (Uso de los Datos) se encuentra en la Tabla 5.141, donde según las valoraciones dadas, ED-SC6 (Calidad de los datos) es más importante que los demás sub-criterios, obteniendo un valor prometedor (m) de 0,37 (Ecuación II.III).

Tabla 5.141: Extensión sintética difusa del nivel 4 para ED-UD

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC6	13,00	15,00	19,00	32,21	40,54	51,17	0,25	0,37	0,59
ED-SC7	8,50	11,00	14,00				0,17	0,27	0,43
ED-SC8	4,42	5,53	7,75				0,09	0,14	0,24
ED-SC9	1,79	2,01	2,42				0,04	0,05	0,08
ED-SC10	4,50	7,00	8,00				0,09	0,17	0,25

Tal como se señala en los valores pesimistas (l), prometedores (m) y optimistas (u) bajo el encabezado Ecuación II.III de la Tabla 5.142; para CV1 MoProPEI es más adecuada si se tiene en cuenta las tres primeras características y en el caso de la última referida a los Datos riesgosos (ED-SC10) se favorece a CRISP-DM. La metodología P^{3}TQ queda en segundo lugar para todas las características, excepto para el Costo de transformación de los datos (ED-SC9) donde ocupó el primer puesto junto a CRISP-DM (para m).

Tabla 5.142: Extensión sintética difusa de Sub-criterios ED-UD para CV1

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC6									
CRISP-DM	1,42	1,53	1,75	11,67	14,87	18,25	0,08	0,10	0,15
MoProPEI	7,00	9,00	11,00				0,38	0,61	0,94
P^{3}TQ	3,25	4,33	5,50				0,18	0,29	0,47
ED-SC7									
CRISP-DM	1,42	1,53	1,75	11,67	14,87	18,25	0,08	0,10	0,15
MoProPEI	7,00	9,00	11,00				0,38	0,61	0,94
P^{3}TQ	3,25	4,33	5,50				0,18	0,29	0,47
ED-SC8									
CRISP-DM	1,53	1,75	2,33	9,78	13,08	16,83	0,09	0,13	0,24
MoProPEI	6,00	8,00	10,00				0,36	0,61	1,02
P^{3}TQ	2,25	3,33	4,50				0,13	0,25	0,46
ED-SC9									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	1,50	1,67	2,00				0,10	0,14	0,22
P^{3}TQ	3,50	5,00	6,00				0,23	0,43	0,67
ED-SC10									
CRISP-DM	6,00	8,00	10,00	9,78	13,08	16,83	0,36	0,61	1,02
MoProPEI	1,53	1,75	2,33				0,09	0,13	0,24
P^{3}TQ	2,25	3,33	4,50				0,13	0,25	0,46

Siguiendo la rama diestra, los valores de extensión sintética difusa al comparar Entorno del Negocio (CN-EN), Proyecto en el Entorno del Negocio (CN-PEN) y DS en el Proyecto del Negocio (CN-DSPN) se presenta en la Tabla 5.143, siendo este último sub-criterio el más relevante a la hora de seleccionar una metodología (ver valores bajo la columna titulada Ecuación II.III).

Para los siete criterios de CN-EN, la matriz que representa el valor de extensión sintética difusa de las comparaciones hechas en la Tabla 5.119, se encuentra en la Tabla 5.144, en la cual se refleja para la Ecuación II.III que el último sub-criterio, Definición de las

oportunidades del negocio (CN-SC7), es el más importante según el valor más prometedor (m) con el 19%, en cambio si se toman los valores pesimistas (l) y optimistas (u) el criterio seleccionado es CN-SC2 (Comprensión general de la organización) con 10% y 34% respectivamente.

Tabla 5.143: Extensión sintética difusa del nivel 3 para CN

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
Entorno del Negocio (CN-EN)	2,25	2,33	3,50	9,00	11,67	15,00	0,15	0,20	0,39
Proyecto en el Entorno del Negocio (CN-PEN)	1,75	2,33	2,50				0,12	0,20	0,28
DS en el Proyecto del Negocio (CN-DSPN)	5,00	7,00	9,00				0,33	0,60	1,00

Tabla 5.144: Extensión sintética difusa del nivel 4 para CN-EN

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC1	2,50	3,00	4,00	43,75	58,33	80,50	0,03	0,05	0,09
CN-SC2	8,00	9,00	15,00				0,10	0,15	0,34
CN-SC3	7,50	9,00	14,00				0,09	0,15	0,32
CN-SC4	7,00	9,00	13,00				0,09	0,15	0,30
CN-SC5	6,50	9,00	12,00				0,08	0,15	0,27
CN-SC6	5,25	8,33	9,50				0,07	0,14	0,22
CN-SC7	7,00	11,00	13,00				0,09	0,19	0,30

Al comprar estas siete características y su correspondencia con la contemplación de las mismas dentro de las metodologías, a partir de las matrices triangulares difusas de comparación pareada (Tabla 5.120 a la Tabla 5.126), se generó la Tabla 5.145 con el valor de extensión sintética borrosa de cada una de ellas. La segunda alternativa (MoProPEI) es la más

adecuada para la mayoría de los sub-criterios evaluados para el caso de validación (CV1), menos para el segundo (CN-SC2, Comprensión general de la organización), donde se recomienda utilizar CRISP-DM y MoProPEI en segundo lugar. Para CN-SC1, CN-SC3 y CN-SC4 en segundo lugar se encuentra P^{3}TQ, en CN-SC5 y CN-SC7 CRISP-DM y para CN-SC6 hubo un empate entre estas dos alternativas.

Tabla 5.145: Extensión sintética difusa de Sub-criterios CN-EN para CV1

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC1									
CRISP-DM	1,58	1,83	2,50	7,92	11,33	15,50	0,10	0,16	0,32
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	2,33	3,50	5,00				0,15	0,31	0,63
CN-SC2									
CRISP-DM	6,00	8,00	10,00	9,78	13,08	16,83	0,36	0,61	1,02
MoProPEI	2,20	3,25	4,33				0,13	0,25	0,44
P^{3}TQ	1,58	1,83	2,50				0,09	0,14	0,26
CN-SC3									
CRISP-DM	1,58	1,83	2,50	7,92	11,33	15,50	0,10	0,16	0,32
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	2,33	3,50	5,00				0,15	0,31	0,63
CN-SC4									
CRISP-DM	1,58	1,83	2,50	7,92	11,33	15,50	0,10	0,16	0,32
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	2,33	3,50	5,00				0,15	0,31	0,63
CN-SC5									
CRISP-DM	2,25	3,33	4,50	9,78	13,08	16,83	0,13	0,25	0,46
MoProPEI	6,00	8,00	10,00				0,36	0,61	1,02
P^{3}TQ	1,53	1,75	2,33				0,09	0,13	0,24
CN-SC6									
CRISP-DM	2,25	2,33	3,50	9,00	11,67	15,00	0,15	0,20	0,39
MoProPEI	5,00	7,00	9,00				0,33	0,60	1,00
P^{3}TQ	1,75	2,33	2,50				0,12	0,20	0,28

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
CN-SC7									
CRISP-DM	2,33	3,50	5,00	7,92	11,33	15,50	0,15	0,31	0,63
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^3TQ	1,58	1,83	2,50				0,10	0,16	0,32

En cuento al Proyecto del Entorno del Negocio (CN-PEN), la valoración de extensión sintética borrosa de confrontación de sus sub-criterios se encuentra en la Tabla 5.146, se puede notar que existe una predilección por CN-SC11 (Contingencias en relación a los riesgos del problema del negocio) con 10, 20 y 30 por ciento como valor pesimista (l), conveniente (m) y optimista (u) de la última columna principal (Ecuación II.III).

Tabla 5.146: Extensión sintética difusa del nivel 4 ara CN-PEN

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC8	3,50	3,67	6,00	23,75	31,67	42,50	0,08	0,12	0,25
CN-SC9	3,00	3,67	5,00				0,07	0,12	0,21
CN-SC10	5,50	7,00	10,00				0,13	0,22	0,42
CN-SC11	7,50	11,00	14,00				0,18	0,35	0,59
CN-SC12	4,25	6,33	7,50				0,10	0,20	0,32

En relación de estas cinco características respecto a las alternativas de metodologías evaluadas en el CV1, las tasaciones como extensión difusa sintética de cada una de ellas, se expone en la Tabla 5.147, para las cuales MoProPEI fue la mejor valorada.

Finalmente para determinar la relevancia de los tres sub-criterios de DS en el Proyecto del Negocio (CN-DSPN), la Tabla 148 advierte una importancia de la correcta definición de los objetivos de la aplicación de la metodología (DS CN-SC13) con valores que duplica la suma de los dos sub-criterios restantes.

Tabla 5.147: Extensión sintética difusa de Sub-criterios CN-PEN para CV1

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC8									
CRISP-DM	1,58	1,83	2,50	7,92	11,33	15,50	0,10	0,16	0,32
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	2,33	3,50	5,00				0,15	0,31	0,63
CN-SC9									
CRISP-DM	1,58	1,83	2,50	7,92	11,33	15,50	0,10	0,16	0,32
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	2,33	3,50	5,00				0,15	0,31	0,63
CN-SC10									
CRISP-DM	2,25	2,33	3,50	9,00	11,67	15,00	0,15	0,20	0,39
MoProPEI	5,00	7,00	9,00				0,33	0,60	1,00
P^{3}TQ	1,75	2,33	2,50				0,12	0,20	0,28
CN-SC11									
CRISP-DM	2,25	2,33	3,50	9,00	11,67	15,00	0,15	0,20	0,39
MoProPEI	5,00	7,00	9,00				0,33	0,60	1,00
P^{3}TQ	1,75	2,33	2,50				0,12	0,20	0,28
CN-SC12									
CRISP-DM	3,33	4,50	6,00	9,78	13,08	16,83	0,20	0,34	0,61
MoProPEI	5,00	7,00	9,00				0,30	0,54	0,92
P^{3}TQ	1,45	1,58	1,83				0,09	0,12	0,19

Tabla 5.148: Extensión sintética difusa del nivel 4 para CN-DSPN

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC13	7,00	9,00	11,00	10,92	13,53	16,75	0,42	0,67	1,01
CN-SC14	2,17	2,20	3,25				0,13	0,16	0,30
CN-SC15	1,75	2,33	2,50				0,10	0,17	0,23

En la comparación de CN-SC13, CN-SC14 y CN-SC15 con respecto a CRISP-DM, MoProPEI y P^{3}TQ, la metodología para este nivel, bajo CN-DSPN, la segunda alternativa sobresale por sobre las demás, como se puede apreciar en la Tabla 5.149.

Tabla 5.149: Extensión sintética difusa de Sub-criterios CN-DSPN para CV1

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC13									
CRISP-DM	2,33	3,50	5,00	7,92	11,33	15,50	0,15	0,31	0,63
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	1,58	1,83	2,50				0,10	0,16	0,32
CN-SC14									
CRISP-DM	2,25	3,33	4,50	9,78	13,08	16,83	0,13	0,25	0,46
MoProPEI	6,00	8,00	10,00				0,36	0,61	1,02
P^{3}TQ	1,53	1,75	2,33				0,09	0,13	0,24
CN-SC15									
CRISP-DM	2,25	3,33	4,50	9,78	13,08	16,83	0,13	0,25	0,46
MoProPEI	6,00	8,00	10,00				0,36	0,61	1,02
P^{3}TQ	1,53	1,75	2,33				0,09	0,13	0,24

Luego de calcular el valor de extensión sintética difusa para todos los niveles de la jerarquía, tal como se indicó en el paso 6.2 del modelo propuesto (Capítulo IV), se determinó la superioridad comparativa de los criterios evaluados, como se explica en el paso 4 del Anexo II.

Para los criterios del segundo nivel, la superioridad comparativa se muestra en la Tabla 5.150, donde a cada criterio se le asigna el valor 1,00 (ver Ecuación II.VII y Ecuación II.VIII del Anexo II).

Tabla 5.150: Superioridad comparativa del nivel 2

V(M2>=M1)	
V(ED>=CN)	1,00
V(CN>=ED)	1,00

En el tercer nivel de la jerarquía propuesta (Figura 4.2), la superioridad comparativa obtenida para el criterio Entendimiento de los Datos (ED) y sus sub-criterios (acceso y uso de los datos) se visualiza en la Tabla 5.151; donde se nota la supremacía del UD, por sobre AD.

Tabla 5.151: Superioridad comparativa del nivel 3 para ED

V(M2>=M1)	
V(AD>=UD)	0,00
V(UD>=AD)	1,00

Tabla 5.152: Superioridad comparativa del nivel 3 para ED-AD

V(M2>=M1)	
V(ED-SC1>=ED-SC2)	0,00
V(ED-SC1>=ED-SC3)	0,00
V(ED-SC1>=ED-SC4)	0,00
V(ED-SC1>=ED-SC5)	0,94
V(ED-SC2>=ED-SC1)	1,00
V(ED-SC2>=ED-SC3)	1,00
V(ED-SC2>=ED-SC4)	0,80
V(ED-SC2>=ED-SC5)	1,00
V(ED-SC3>=ED-SC1)	1,00
V(ED-SC3>=ED-SC2)	0,68
V(ED-SC3>=ED-SC4)	0,50
V(ED-SC3>=ED-SC5)	1,00
V(ED-SC4>=ED-SC1)	1,00
V(ED-SC4>=ED-SC2)	1,00
V(ED-SC4>=ED-SC3)	1,00
V(ED-SC4>=ED-SC5)	1,00
V(ED-SC5>=ED-SC1)	1,00
V(ED-SC5>=ED-SC2)	0,00
V(ED-SC5>=ED-SC3)	0,00
V(ED-SC5>=ED-SC4)	0,00

Continuando en la misma rama de la jerarquía, de izquierda a derecha, se determina la superioridad comparativa del AD y su cinco sub-criterios, tal como se observa en la Tabla 5.152.

En la Tabla 5.153 se muestra la superioridad comparativa de los cinco sub-criterios del AD, en relación a cada alternativa.

Tabla 5.153: Superioridad comparativa de Sub-criterios ED-AD para CV1

V(M2>=M1)						
ED-SC1	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,00	V(A2>=A3)	1,00	V(A3>=A2)	0,62
ED-SC2	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,00	V(A2>=A3)	1,00	V(A3>=A2)	0,25
ED-SC3	V(A1>=A2)	1,00	V(A2>=A1)	0,14	V(A3>=A1)	0,63
	V(A1>=A3)	1,00	V(A2>=A3)	0,53	V(A3>=A2)	1,00
ED-SC4	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,00	V(A2>=A3)	1,00	V(A3>=A2)	0,62
ED-SC5	V(A1>=A2)	0,14	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,53	V(A2>=A3)	1,00	V(A3>=A2)	0,63

Para ED-UD y sus cinco sub-criterios, las Tabla 5.154 y 5.155 poseen los valores de preponderancia comparativa calculados.

Tabla5.154: Superioridad comparativa del nivel 4 para ED-UD

V(M2>=M1)	
V(ED-SC6>=ED-SC7)	1,00
V(ED-SC6>=ED-SC8)	1,00
V(ED-SC6>=ED-SC9)	1,00
V(ED-SC6>=ED-SC10)	1,00
V(ED-SC7>=ED-SC6)	0,65
V(ED-SC7>=ED-SC8)	1,00
V(ED-SC7>=ED-SC9)	1,00
V(ED-SC7>=ED-SC10)	1,00

V(M2>=M1)	
V(ED-SC8>=ED-SC6)	0,00
V(ED-SC8>=ED-SC7)	0,36
V(ED-SC8>=ED-SC9)	1,00
V(ED-SC8>=ED-SC10)	0,81
V(ED-SC9>=ED-SC6)	0,00
V(ED-SC9>=ED-SC7)	0,00
V(ED-SC9>=ED-SC8)	0,00
V(ED-SC9>=ED-SC10)	0,00
V(ED-SC10>=ED-SC6)	0,00
V(ED-SC10>=ED-SC7)	0,45
V(ED-SC10>=ED-SC8)	1,00
V(ED-SC10>=ED-SC9)	1,00

Tabla 5.155: Superioridad comparativa de Sub-criterios ED-UD para CV1

V(M2>=M1)						
ED-SC6	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,00	V(A2>=A3)	1,00	V(A3>=A2)	0,22
ED-SC7	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,00	V(A2>=A3)	1,00	V(A3>=A2)	0,22
ED-SC8	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,46	V(A2>=A3)	1,00	V(A3>=A2)	0,22
ED-SC9	V(A1>=A2)	1,00	V(A2>=A1)	0,00	V(A3>=A1)	1,00
	V(A1>=A3)	1,00	V(A2>=A3)	0,00	V(A3>=A2)	1,00
ED-SC10	V(A1>=A2)	1,00	V(A2>=A1)	0,00	V(A3>=A1)	0,22
	V(A1>=A3)	1,00	V(A2>=A3)	0,46	V(A3>=A2)	1,00

Al igual que para el criterio Entendimiento de los Datos, para la Comprensión del Negocio y sus sub-criterios, se evaluó el predominio al compararlos en los diferentes niveles. Para el nivel tres, al comparar los tres sub-criterios generales, la superioridad comparativa se expone en la Tabla 5.156; luego en la Tabla 5.157 de los siete sub-criterios del Entorno del Negocio (CN-EN), en la Tabla 5.159 de los cinco del Proyecto en el Entorno del Negocio (CN-PEN) y en la Tabla 5.161 de los tres de Data Science en el Proyecto del Negocio (CN-DSPN); y de

estos sub-criterios del nivel cuatro en relación a cada metodología se encuentran en las Tablas 5.158, 5.160 y 5.162.

Tabla 5.156: Superioridad comparativa del nivel 3 para CN

V(M2>=M1)	
V(EN>=PEN)	1,00
V(EN>=DSPN)	0,12
V(PEN>=EN)	1,00
V(PEN>=DSPN)	0,00
V(DSPN >=EN)	1,00
V(DSPN >=PEN)	1,00

Tabla 5.157: Superioridad comparativa del nivel 4 para CN-EN

V(M2>=M1)	
V(CN-SC1>=CN-SC2)	0,00
V(CN-SC1>=CN-SC3)	0,00
V(CN-SC1>=CN-SC4)	0,04
V(CN-SC1>=CN-SC5)	0,09
V(CN-SC1>=CN-SC6)	0,22
V(CN-SC1>=CN-SC7)	0,03
V(CN-SC2>=CN-SC1)	1,00
V(CN-SC2>=CN-SC3)	1,00
V(CN-SC2>=CN-SC4)	1,00
V(CN-SC2>=CN-SC5)	1,00
V(CN-SC2>=CN-SC6)	1,00
V(CN-SC2>=CN-SC7)	0,88
V(CN-SC3>=CN-SC1)	1,00
V(CN-SC3>=CN-SC2)	1,00
V(CN-SC3>=CN-SC4)	1,00
V(CN-SC3>=CN-SC5)	1,00
V(CN-SC3>=CN-SC6)	1,00
V(CN-SC3>=CN-SC7)	0,87
V(CN-SC4>=CN-SC1)	1,00

V(M2>=M1)	
V(CN-SC4>=CN-SC2)	1,00
V(CN-SC4>=CN-SC3)	1,00
V(CN-SC4>=CN-SC5)	1,00
V(CN-SC4>=CN-SC6)	1,00
V(CN-SC4>=CN-SC7)	0,86
V(CN-SC5>=CN-SC1)	1,00
V(CN-SC5>=CN-SC2)	1,00
V(CN-SC5>=CN-SC3)	1,00
V(CN-SC5>=CN-SC4)	1,00
V(CN-SC5>=CN-SC6)	1,00
V(CN-SC5>=CN-SC7)	0,85
V(CN-SC6>=CN-SC1)	1,00
V(CN-SC6>=CN-SC2)	0,91
V(CN-SC6>=CN-SC3)	0,92
V(CN-SC6>=CN-SC4)	0,92
V(CN-SC6>=CN-SC5)	0,92
V(CN-SC6>=CN-SC7)	0,74
V(CN-SC7>=CN-SC1)	1,00
V(CN-SC7>=CN-SC2)	1,00
V(CN-SC7>=CN-SC3)	1,00
V(CN-SC7>=CN-SC4)	1,00
V(CN-SC7>=CN-SC5)	1,00
V(CN-SC7>=CN-SC6)	1,00

Tabla 5.158: Superioridad comparativa de Sub-criterios CN-EN para CV1

V(M2>=M1)						
CN-SC1	V(A1>=A2)	0,14	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,53	V(A2>=A3)	1,00	V(A3>=A2)	0,63
CN-SC2	V(A1>=A2)	1,00	V(A2>=A1)	0,19	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,54
CN-SC3	V(A1>=A2)	0,14	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,53	V(A2>=A3)	1,00	V(A3>=A2)	0,63

V(M2>=M1)						
CN-SC4	V(A1>=A2)	0,14	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,53	V(A2>=A3)	1,00	V(A3>=A2)	0,63
CN-SC5	V(A1>=A2)	0,22	V(A2>=A1)	1,00	V(A3>=A1)	0,46
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC6	V(A1>=A2)	0,12	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC7	V(A1>=A2)	0,63	V(A2>=A1)	1,00	V(A3>=A1)	0,53
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,14

Tabla 5.159: Superioridad comparativa del nivel 4 para CN-PEN

V(M2>=M1)	
V(CN-SC8>=CN-SC9)	1,00
V(CN-SC8>=CN-SC10)	0,54
V(CN-SC8>=CN-SC11)	0,25
V(CN-SC8>=CN-SC12)	0,64
V(CN-SC9>=CN-SC8)	1,00
V(CN-SC9>=CN-SC10)	0,44
V(CN-SC9>=CN-SC11)	0,13
V(CN-SC9>=CN-SC12)	0,57
V(CN-SC10>=CN-SC8)	1,00
V(CN-SC10>=CN-SC9)	1,00
V(CN-SC10>=CN-SC11)	0,66
V(CN-SC10>=CN-SC12)	1,00
V(CN-SC11>=CN-SC8)	1,00
V(CN-SC11>=CN-SC9)	1,00
V(CN-SC11>=CN-SC10)	1,00
V(CN-SC11>=CN-SC12)	1,00
V(CN-SC12>=CN-SC8)	1,00
V(CN-SC12>=CN-SC9)	1,00
V(CN-SC12>=CN-SC10)	0,90
V(CN-SC12>=CN-SC11)	0,49

Tabla 5.160: Superioridad comparativa de Sub-criterios CN-PEN para CV1

V(M2>=M1)						
CN-SC8	V(A1>=A2)	0,14	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,53	V(A2>=A3)	1,00	V(A3>=A2)	0,63
CN-SC9	V(A1>=A2)	0,14	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,53	V(A2>=A3)	1,00	V(A3>=A2)	0,63
CN-SC10	V(A1>=A2)	0,12	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC11	V(A1>=A2)	0,12	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC12	V(A1>=A2)	0,62	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00

Tabla 5.161: Superioridad comparativa del nivel 4 para CN-DSPN

V(M2>=M1)	
V(CN-SC13>=CN-SC14)	1,00
V(CN-SC13>=CN-SC15)	1,00
V(CN-SC14>=CN-SC13)	0,00
V(CN-SC14>=CN-SC15)	0,95
V(CN-SC15>=CN-SC13)	0,00
V(CN-SC15>=CN-SC14)	1,00

Tabla 5.162: Superioridad comparativa de Sub-criterios CN-DSPN para CV1

V(M2>=M1)						
CN-SC13	V(A1>=A2)	0,63	V(A2>=A1)	1,00	V(A3>=A1)	0,53
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,14
CN-SC14	V(A1>=A2)	0,22	V(A2>=A1)	1,00	V(A3>=A1)	0,46
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC15	V(A1>=A2)	0,22	V(A2>=A1)	1,00	V(A3>=A1)	0,46
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00

Posterior a calcular la superioridad comparativa de cada criterio y sub-criterio, se escogió el valor mínimo (de superioridad); para luego calcular el vector de pesos y normalizarlo para cada criterio y sub-criterio, tal como se describe en el paso 5 del Anexo II.

Para el segundo nivel, el mínimo valor es 1. El cálculo del vector de pesos (W´) y la normalización del mismo (W) se muestra en la Tabla 5.163, donde se indica que al compararlos los dos criterios, ambos poseen una ponderación del 50%.

Tabla 5.163: Vector de pesos del nivel 2

$W' = (1{,}00; 1{,}00)^T$	$\sum W' = 2{,}00$
$W = (0{,}5; 0{,}50)^T$	

En la Tabla 5.164 se expone el vector de pesos del ED en relación a sus dos sub-criterios (ED-AD y ED-UD).

Tabla 5.164: Vector de pesos del nivel 3 para ED

$W' = (0{,}00; 1{,}00)^T$	$\sum W' = 1{,}00$
$W = (0{,}0; 1{,}00)^T$	

Para el Acceso a los Datos y sus cinco sub-criterios, el vector de pesos obtenido se encuentra en la Tabla 5.165 y el de cada uno de estos sub-criterios respecto a las tres metodologías evaluadas en el CV1, en la Tabla 5.166.

Tabla 5.165: Vector de pesos del nivel 4 para ED-AD

$W' = (0{,}00; 0{,}80; 0{,}50; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}30$
$W = (0{,}0; 0{,}35; 0{,}22; 0{,}43; 0{,}00)^T$	

Tabla 5.166: Vector de pesos de los Sub-criterios ED-AD para CV1

ED-SC1	$W' = (0{,}00; 1{,}00; 0{,}62)^T$	$\sum W' = 1{,}62$
	$W = (0{,}0; 0{,}62; 0{,}38)^T$	
ED-SC2	$W' = (0{,}00; 1{,}00; 0{,}25)^T$	$\sum W' = 1{,}25$
	$W = (0{,}00; 0{,}80; 0{,}20)^T$	
ED-SC3	$W' = (1{,}00; 0{,}14; 0{,}63)^T$	$\sum W' = 1{,}76$
	$W = (0{,}57; 0{,}08; 0{,}35)^T$	
ED-SC4	$W' = (0{,}00; 1{,}00; 0{,}62)^T$	$\sum W' = 1{,}62$
	$W = (0{,}00; 0{,}62; 0{,}38)^T$	
ED-SC5	$W' = (1{,}00; 0{,}14; 0{,}63)^T$	$\sum W' = 1{,}76$
	$W = (0{,}57; 0{,}08; 0{,}35)^T$	

En cuanto al Uso de los Datos y sus cinco características, en la Tabla 5.167 y Tabla 5.168 se visualizan los vectores de pesos, al compararlos entre sí y por separado en relación a cada metodología para el caso.

Tabla 5.167: Vector de pesos del nivel 4 para ED-UD

$W' = (1{,}00; 0{,}65; 0{,}00; 0{,}00; 0{,}00)^T$	$\sum W' = 1{,}65$
$W = (0{,}61; 0{,}39; 0{,}00; 0{,}00; 0{,}00)^T$	

Tabla 5.168: Vector de pesos de los Sub-criterios ED-UD para CV1

ED-SC6	$W' = (0{,}00; 1{,}00; 0{,}22)^T$	$\sum W' = 1{,}22$
	$W = (0{,}00; 0{,}82; 0{,}18)^T$	
ED-SC7	$W' = (0{,}00; 1{,}00; 0{,}22)^T$	$\sum W' = 1{,}22$
	$W = (0{,}00; 0{,}82; 0{,}18)^T$	

ED-SC8	$W' = (0{,}00; 1{,}00; 0{,}22)^T$	$\sum W' = 1{,}22$
	$W = (0{,}00; 0{,}82; 0{,}18)^T$	
ED-SC9	$W' = (0{,}00; 1{,}00; 1{,}00)^T$	$\sum W' = 1{,}62$
	$W = (0{,}50; 0{,}00; 0{,}50)^T$	
ED-SC10	$W' = (0{,}00; 1{,}00; 0{,}22)^T$	$\sum W' = 1{,}22$
	$W = (0{,}00; 0{,}82; 0{,}18)^T$	

El vector de pesos de la Comprensión del Negocio (CN) al cotejar sus tres sub-criterios se visualiza en la Tabla 5.169, en la Tabla 5.170 del Entorno del Negocio y sus siete sub-criterios, en la Tabla 5.172 de los cinco del Proyecto en el Entorno del Negocio y en la Tabla 5.174 de Data Science en el Proyecto del Negocio y sus tres sub-criterios. Los sub-criterios del nivel cuatro contrastados con cada alternativa, obtuvo los vectores expuestos en las Tablas 5.171, 5.173 y 5.175 respectivamente.

Tabla 5.169: Vector de pesos del nivel 3 para CN

$W' = (0{,}12; 0{,}00; 1{,}00)^T$	$\sum W' = 1{,}12$
$W = (0{,}11; 0{,}00; 0{,}89)^T$	

Tabla 5.170: Vector de pesos del nivel 4 para CN-EN

$W' = (0{,}00; 0{,}88; 0{,}87; 0{,}86; 0{,}85; 0{,}74; 1{,}00)^T$	$\sum W' = 5{,}20$
$W = (0{,}00; 0{,}17; 0{,}17; 0{,}17; 0{,}16; 0{,}14; 0{,}19)^T$	

Tabla 5.171: Vector de pesos de los Sub-criterios CN-EN para CV1

CN-SC1	$W' = (0{,}14; 1{,}00; 0{,}63)^T$	$\sum W' = 1{,}76$
	$W = (0{,}08; 0{,}57; 0{,}35)^T$	
CN-SC2	$W' = (1{,}00; 0{,}19; 0{,}00)^T$	$\sum W' = 1{,}19$
	$W = (0{,}84; 0{,}16; 0{,}00)^T$	

CN-SC3	$W' = (0{,}14; 1{,}00; 0{,}63)^T$	$\sum W' = 1{,}76$
	$W = (0{,}08; 0{,}57; 0{,}35)^T$	
CN-SC4	$W' = (0{,}14; 1{,}00; 0{,}63)^T$	$\sum W' = 1{,}76$
	$W = (0{,}08; 0{,}57; 0{,}35)^T$	
CN-SC5	$W' = (0{,}22; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}22$
	$W = (0{,}18; 0{,}82; 0{,}00)^T$	
CN-SC6	$W' = (0{,}12; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}12$
	$W = (0{,}11; 0{,}89; 0{,}00)^T$	
CN-SC7	$W' = (0{,}63; 1{,}00; 0{,}14)^T$	$\sum W' = 1{,}76$
	$W = (0{,}35; 0{,}57; 0{,}08)^T$	

Tabla 5.172: Vector de pesos del nivel 4 para CN-PEN

$W' = (0{,}25; 0{,}13; 0{,}66; 1{,}00; 0{,}49)^T$	$\sum W' = 2{,}52$
$W = (0{,}10; 0{,}05; 0{,}26; 0{,}40; 0{,}19)^T$	

Tabla 5.173: Vector de pesos de los Sub-criterios CN-PEN para CV1

CN-SC8	$W' = (0{,}14; 1{,}00; 0{,}63)^T$	$\sum W' = 1{,}76$
	$W = (0{,}08; 0{,}57; 0{,}35)^T$	
CN-SC9	$W' = (0{,}14; 1{,}00; 0{,}63)^T$	$\sum W' = 1{,}76$
	$W = (0{,}08; 0{,}57; 0{,}35)^T$	
CN-SC10	$W' = (0{,}12; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}12$
	$W = (0{,}11; 0{,}89; 0{,}00)^T$	
CN-SC11	$W' = (0{,}12; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}12$
	$W = (0{,}11; 0{,}89; 0{,}00)^T$	

CN-SC12	$W' = (0{,}62; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}62$
	$W = (0{,}38; 0{,}62; 0{,}00)^T$	

Tabla 5.174: Vector de pesos del nivel 4 para CN-DSPN

$W' = (1{,}00; 0{,}00; 0{,}00)^T$	$\sum W' = 1{,}00$
$W = (1{,}00; 0{,}00; 0{,}00)^T$	

Tabla 5.175: Vector de pesos de los Sub-criterios CN-DSPN para CV1

CN-SC13	$W' = (0{,}63; 1{,}00; 0{,}14)^T$	$\sum W' = 1{,}76$
	$W = (0{,}35; 0{,}57; 0{,}08)^T$	
CN-SC14	$W' = (0{,}22; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}22$
	$W = (0{,}18; 0{,}82; 0{,}00)^T$	
CN-SC15	$W' = (0{,}22; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}22$
	$W = (0{,}18; 0{,}82; 0{,}00)^T$	

Paso subsiguiente, se establecieron las prioridades parciales por nivel jerárquico y de cada alternativa, como se guía en el paso 6 del Anexo II, para las cuales se calcularon los valores normalizados de los criterios con los valores de multiplicación de las alternativas y los sub-criterios estimados.

Como se indicó en la Tabla 5.163, al valorar el experto a los dos criterios principales (ED y CN) con una importancia equivalente, éstos poseen una ponderación del 50% cada uno, tal como se puede apreciar en la Tabla 5.176.

Tabla 5.176: Ponderaciones parciales del nivel 2

	Pesos
Entendimiento de los Datos (ED)	0,50
Comprensión del Negocio (CN)	0,50

Para el Entendimiento de los Datos, la ponderación parcial fue totalmente a favor del Uso de los Datos, como se puede notar en la Tabla 5.177.

Tabla 5.177: Ponderaciones parciales del nivel 3 para ED

	Pesos
Acceso a los Datos (ED-AD)	0,00
Uso de los Datos (ED-UD)	1,00

En las Tablas 5.178 y 5.179, se exponen los pesos obtenidos para los sub-criterios del Acceso de los Datos, primero al compararlos entre sí y luego al hacerlo con las metodologías. La característica más relevante es la Diversidad de fuentes de datos (ED-SC4) con el 43% (Tabla 5.178), para la cual MoProPEI (A2) obtuvo el 62% en el CV1 (Tabla 5.179).

Tabla 5.178: Ponderaciones parciales del nivel 4 para ED-AD

	Pesos
ED-SC1	0,00
ED-SC2	0,35
ED-SC3	0,22
ED-SC4	0,43
ED-SC5	0,00

Tabla 5.179: Ponderaciones parciales de los Sub-criterios ED-AD para CV1

	Pesos		
	A1	**A2**	**A3**
ED-SC1	0,00	0,62	0,38
ED-SC2	0,00	0,80	0,20
ED-SC3	0,57	0,08	0,36
ED-SC4	0,00	0,62	0,38
ED-SC5	0,08	0,57	0,36
Peso ED-AD	**0,12**	**0,56**	**0,31**

En cuanto al Uso de los Datos, las dos primeras características (ED-SC6 y ED-SC7), según los cálculos realizados con FAHP, han sido las importantes (61% y 39% respectivamente),

descartando las demás (ver Tabla 5.180). Para estas, la A2 cumple con un 82% esta característica, seguido de A3 (P^3TQ), tal como se indica en la Tabla 5.181.

Tabla 5.180: Ponderaciones parciales del nivel 4 para ED-UD

	Pesos
ED-SC6	0,61
ED-SC7	0,39
ED-SC8	0,00
ED-SC9	0,00
ED-SC10	0,00

Tabla 5.181: Ponderaciones parciales de los Sub-criterios ED-UD para CV1

	Pesos		
	A1	**A2**	**A3**
ED-SC6	0,00	0,82	0,18
ED-SC7	0,00	0,82	0,18
ED-SC8	0,00	0,82	0,18
ED-SC9	0,50	0,00	0,50
ED-SC10	0,82	0,00	0,18
Peso ED-UD	**0,00**	**0,82**	**0,18**

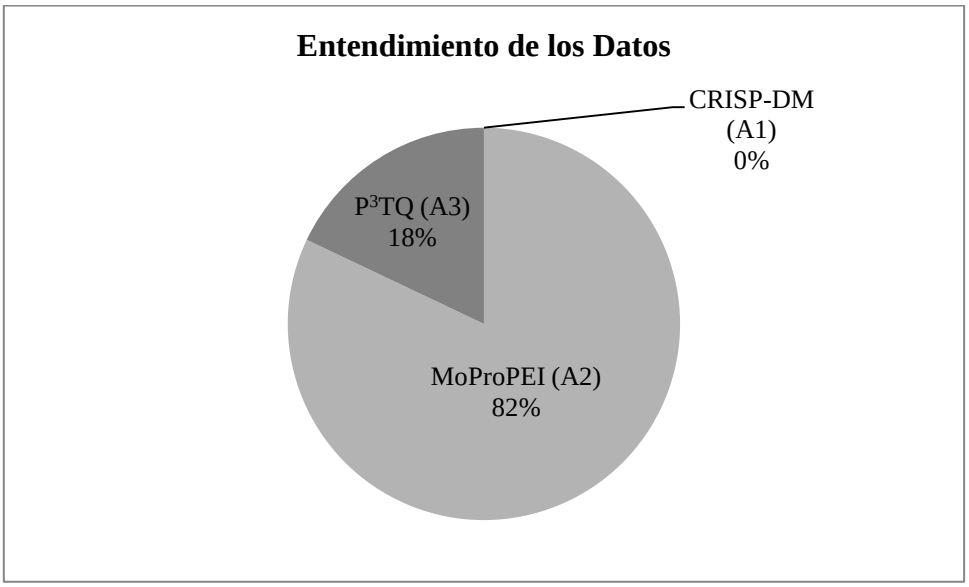

Figura 5.3: Ponderaciones globales del ED para el CV1 con FAHP

A partir de las ponderaciones parciales, se obtuvo las globales para el Entendimiento de los Datos, visible en la Figura 5.3.

Para la Comprensión del Negocio, las ponderaciones alcanzadas se muestran en la Tabla 5.182, donde se nota la trascendencia del sub-criterios DS en el Proyecto del Negocio (CN-DSPN) con un peso de 89%.

Tabla 5.182: Ponderaciones parciales del nivel 3 para CN

	Pesos
Entorno del Negocio (CN-EN)	0,11
Proyecto en el Entorno del Negocio (CN-PEN)	0,00
DS en el Proyecto del Negocio (CN-DSPN)	0,89

Los pesos calculados para los sub-criterios del Entorno del Negocio, confrontados directamente y en relación a las alternativas, se manifiesta en las Tablas 5.183 y 5.184. Las ponderaciones en la primera tabla son equitativas, excepto para la primera (CN-SC1, Situación económica circundante) que es considerada irrelevante con el método FAHP para este caso de validación (CV1). Bajo éste sub-criterio la alternativa 2 es nuevamente la preferida (ver Tabla 5.184).

Tabla 5.183: Ponderaciones parciales del nivel 4 para CN-EN

	Pesos
CN-SC1	0,00
CN-SC2	0,17
CN-SC3	0,17
CN-SC4	0,17
CN-SC5	0,16
CN-SC6	0,14
CN-SC7	0,19

Para el sub-criterio Proyecto del Entorno del Negocio (PEN), la característica de mayor peso se relaciona al plan de contingencia en relación a los riesgos del negocio (ED-SC11), conforme a los cálculos realizados con FAHP, su relevancia es del 40%, tal como se puede notar en la Tabla 5.185. Al mirar las ponderaciones parciales del Sub-criterios PEN, en la

Tabla 5.186, donde MoProPEI es nuevamente la seleccionada para este conjunto de características.

Tabla 5.184: Ponderaciones parciales de los Sub-criterios CN-EN para CV1

	Pesos		
	A1	**A2**	**A3**
CN-SC1	0,08	0,57	0,36
CN-SC2	0,84	0,16	0,00
CN-SC3	0,08	0,57	0,36
CN-SC4	0,08	0,57	0,36
CN-SC5	0,18	0,82	0,00
CN-SC6	0,11	0,89	0,00
CN-SC7	0,36	0,57	0,08
Peso CN-EN	**0,28**	**0,58**	**0,13**

Tabla 5.185: Ponderaciones parciales del nivel 4 para CN-PEN

	Pesos
CN-SC8	0,10
CN-SC9	0,05
CN-SC10	0,26
CN-SC11	0,40
CN-SC12	0,19

Tabla 5.186: Ponderaciones parciales de los Sub-criterios CN-PEN para CV1

	Pesos		
	A1	**A2**	**A3**
CN-SC8	0,08	0,57	0,36
CN-SC9	0,08	0,57	0,36
CN-SC10	0,11	0,89	0,00
CN-SC11	0,11	0,89	0,00
CN-SC12	0,38	0,62	0,00
Peso CN-PEN	**0,16**	**0,79**	**0,05**

En la Tabla 5.187 y Tabla 5.188 se exponen las ponderaciones parciales de los tres sub-criterios de DSPN, donde la característica vinculada a los objetivos de la aplicación de DS (SC13) obtuvo un 100% en comparación a las otras dos características. Para SC13, A2 alcanzó un 57%, seguido de A1 con 36% y A3 con el 8%.

Tabla 5.187: Ponderaciones parciales del nivel 4 para CN-DSPN

	Pesos
CN-SC13	1,00
CN-SC14	0,00
CN-SC15	0,00

Tabla 5.188: Ponderaciones parciales de los Sub-criterios CN-DSPN para CV1

	Pesos		
	A1	**A2**	**A3**
CN-SC13	0,36	0,57	0,08
CN-SC14	0,18	0,82	0,00
CN-SC15	0,18	0,82	0,00
Peso CN-DSPN	**0,36**	**0,57**	**0,08**

En las ponderaciones globales de la rama derecha de la estructura, MoProPEI consiguió un peso del 57%, CRISP-DM del 35% y PT^{3}Q del 8%, tal como se expone en la Figura 5.4.

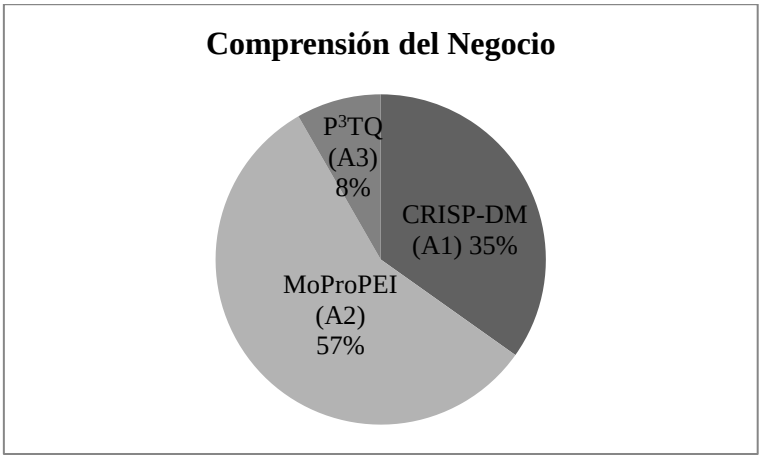

Figura 5.4: Ponderaciones globales de la CN para el CV1 con FAHP

Las prioridades finales logradas por cada alternativa se encuentran en la Tabla 5.189, las mismas se calcularon a partir de las prioridades parciales y globales anteriormente expuestas (siguiendo el paso 7 del Anexo II). Para el caso de validación (CV1) relacionado a los causales de la rotura de automóviles 0KM al ser transportados a las concesionarias, MoProPEI es la metodología más adecuada, dado que obtuvo una prioridad final de 0,69.

Tabla 5.189: Prioridades finales de las alternativas con FAHP para CV1

	CRISP-DM (A1)	**MoProPEI (A2)**	**P^{3}TQ (A3)**
ED	0,00	0,82	0,18
CN	0,35	0,57	0,08
Prioridades Finales	**0,17**	**0,69**	**0,13**

5.2.5. Integración de las ponderaciones obtenidas por los métodos AHP y FAHP para el CV1

En la Tabla 5.190 se exhiben las ponderaciones finales obtenidas para cada metodología, a partir de las ponderaciones finales alcanzadas por cada metodología con los métodos AHP y FAHP, a través del cálculo del promedio de las ponderaciones obtenidas para cada una de ellas.

Tabla 5.190: Prioridades finales de las alternativas para el CV1

Alternativas	**AHP**	**FAHP**	**Prioridades Finales**
CRISP-DM (A1)	0,25	0,17	**0,21**
MoProPEI (A2)	0,53	0,69	**0,61**
P^{3}TQ (A3)	0,22	0,13	**0,18**

5.2.6. Ranking de metodologías de DS para el CV1

Finalmente el ranking de prioridades ubican a la metodología MoProPEI (A2) en primer lugar con el 61% de adecuación para el caso de validación en cuestión (CV1), seguido de CRISP-DM (A1) con el 21% y en último lugar se encuentra P^{3}TQ (A3) con el 18%, tal como se muestra en la Figura 5.5.

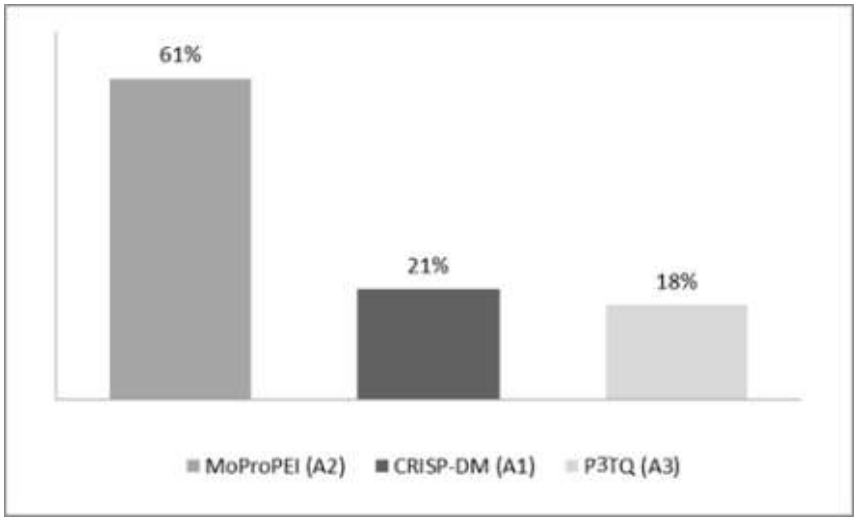

Figura 5.5: Ranking de prioridades finales para el CV1

5.3. Caso de validación: Deserción universitaria

Como se mencionó, el segundo Caso de Validación (CV2) tuvo como propósito determinar las causas de la deserción universitaria. Seguidamente se detallan las tareas realizadas para la obtención de resultados, siguiendo los pasos definidos en el modelo propuesto.

5.3.1. Valoración con TCP del nivel 4 del CV2

Tal como se indicó en la sección 5.2.1., las características del cuarto nivel de la jerarquía fueron evaluadas según el nivel de cumplimiento en cada caso de validación y metodología; para lo cual se recurrió a la técnica de TCP para definir las mismas, establecer las etiquetas lingüísticas y conformar las parrillas que el experto completó (ver Anexo IV).
Las etiquetas lingüísticas, catalogadas en una escala del 1 al 9, que fueron tomadas en cuenta para completar las parrillas, son las mismas que se especificaron en la sección 5.2.1.
Partiendo del Entendimiento de los Datos, en la Tabla 5.191 se muestran las valoraciones dadas por los expertos para el CV2, según las características del Acceso a Datos y las metodologías evaluadas.

Tabla 5.191: Parrilla de Acceso a Datos del CV2

		CRISP-DM	**MoPro PEI**	**P^{3}TQ**	
1.	No se analiza la portabilidad de los datos	8	9	7	De 98% a 100% de portabilidad de los datos
2.	No se analiza la accesibilidad de los datos	6	8	4	De 98% a 100% de accesibilidad de los datos
3.	No se analiza la diversidad de los datos	9	9	7	De 0% a 13% de diversidad de los datos
4.	No se analiza la diversidad de las fuentes de datos	9	9	7	De 0% a 13% de diversidad de la fuentes de datos
5.	No se analiza la disponibilidad de los recursos necesarios	7	9	9	De 98% a 100% de los recursos necesarios

Para los sub-criterios del Uso de los Datos, los expertos establecieron las valoraciones en relación a las tres metodologías, expuestas en la Tabla 5.192.

Tabla 5.192: Parrilla de Uso de los Datos del CV2

		CRISP-DM	**MoPro PEI**	**P^{3}TQ**	
6.	No se analiza la calidad/fiabilidad de los datos	8	9	6	De 98% a 100% de calidad/fiabilidad de los datos
7.	No se analiza la completitud de los datos	9	9	7	De 98% a 100% de completitud de los datos
8.	No se analiza la funcionalidad de los datos	6	8	5	De 98% a 100% de funcionalidad de los datos
9.	No se analiza el costo de transformación de los datos	7	9	4	De 0% a 13% de costo de transformación de los datos
10.	No se analiza los riesgos de los de los datos	8	9	5	De 0% a 13% de datos riesgosos

En cuanto a la Comprensión del Negocio, los expertos asignaron las puntuaciones a cada sub-criterio y metodología, en cuento a su nivel de cumplimiento, expuestas en la Tabla 5.193.

Tabla 5.193: Parrilla de Entorno del Negocio del CV2

		CRISP-DM	**MoPro PEI**	**P^{3}TQ**	
1.	No se analiza la situación circundante	9	9	7	De 98% a 100% de favorabilidad de la situación circundante
2.	No se analiza o realiza una evaluación/comprensión general de la organización	9	9	7	De 98% a 100% de evaluación/comprensión general de la organización
3.	No se analiza o realiza la definición de los objetivos del negocio	9	9	7	De 98% a 100% de claridad en la definición de los objetivos del negocio
4.	No se analiza o define el problema del negocio	9	9	7	De 98% a 100% de claridad en la definición del problema del negocio
5.	No se analizan o definen las oportunidades del negocio	9	9	7	De 98% a 100% de claridad en la definición de las oportunidades del negocio
6.	No se analizan o definen de los requisitos	9	9	7	De 98% a 100% de claridad en la definición de los requisitos
7.	No se analiza o definen los criterios de éxitos del proyecto (desde el negocio)	9	9	7	De 98% a 100% de claridad en la definición de los criterios de éxitos del proyecto (desde el negocio)

Las puntuaciones dadas por los expertos, teniendo en cuenta las características vinculadas el Proyecto del Entorno del Negocio, se exhiben en la Tabla 5.194.

Tabla 5.194: Parrilla del Proyecto Entorno del Negocio del CV2

		CRISP-DM	**MoPro PEI**	**P^{3}TQ**	
8.	No se analiza el conocimiento de los RRHH del negocio	8	9	9	De 98% a 100% de conocimiento de los RRHH del negocio
9.	No se analiza la disponibilidad de los RRHH del negocio	8	9	9	De 98% a 100% de disponibilidad de los RRHH del negocio
10.	No se analizan los riesgos del negocio (respecto al proyecto)	8	9	6	De 98% a 100% de riesgos identificados del negocio (amenazan la supervivencia del proyecto)
11.	No se analizan contingencias en relación a los riesgos del problema del negocio	8	9	6	De 98% a 100% de contingencias en relación a los riesgos del problema del negocio
12.	No se analiza la viabilidad del proyecto	6	9	2	De 98% a 100% de viabilidad del proyecto

Tabla 5.195: Parrilla de Data Science del Proyecto Entorno del Negocio del CV2

		CRISP-DM	**MoPro PEI**	**P^{3}TQ**	
16.	No se analiza o realiza la definición de los objetivos de aplicación de la metodología de DS	9	9	8	De 98% a 100% de claridad en la definición de los objetivos de aplicación de la metodología de DS
17.	No se analiza el conocimiento de los RRHH del proyecto de DS	7	9	9	De 98% a 100% de conocimiento de los RRHH del proyecto de DS
18.	No se analiza la disponibilidad de los RRHH del proyecto de DS	7	9	8	De 98% a 100% de disponibilidad de los RRHH del proyecto de DS

Finalmente según las etiquetas lingüísticas para el sub-cirterio Data Science en el Proyecto del Entorno del Negocio para el CV2, son las presentadas en la Tabla 5.195.

5.3.2. Completitud de las matrices y establecimiento de prioridades entre criterios del CV2

Como se indicó en la sección 5.2.2 y en el paso 4 de modelo propuesto en el Capítulo IV, a partir de las matrices de dos entradas y las parrillas, representadas como tablas; se completaron y adaptaron a las propuestas por Saaty.
Para el presente caso de validación (CV2), al igual que el anterior (CV1), se detallan los pasos del modelo propuesto explicando los valores obtenidos en los diferentes niveles de la estructura jerárquica (ver Figura 4.2) de arriba hacia abajo.
Las preferencias del experto en relación al Entendimiento de los Datos (ED) y Comprensión del Negocio (CN) del nivel se expusieron en la Tabla 5.30, cuya preferencia es igualitaria, es decir que el experto asignó el valor 1, por lo que cada uno de estos sub-criterios poseen una ponderación del 50% como se indicó en la Tabla 5.31.
En el tercer nivel, para la rama izquierda de la jerarquía, los dos sub-criterios de ED, Acceso a los Datos (AD) y Uso de los Datos (UD) las valoraciones y ponderaciones se encuentran en las Tablas 5.32 y 5.33. Con una moderada preferencia de UD, por sobre AD, por lo que las ponderaciones calculadas son del 75% y 25% respectivamente.
Continuando la misma rama, el sub-criterio siniestro (AD), en las Tablas 5.34 y 5.35, se encuentran los valores y ponderaciones obtenidas para las cinco características involucradas.
Ahora bien, al analizar cada característica del AD en relación a las metodologías evaluadas para el CV2, se representan desde la Tabla 5.196 a la 5.200.

Tabla 5.196: Ponderación de ED-SC1 en relación a cada alternativa para CV2

Portabilidad (ED-SC1)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	2	0,29	0,27	0,33	0,30
MoProPEI (A2)	2	1	3	0,57	0,55	0,50	0,54
P^3TQ (A3)	1/2	1/3	1	0,14	0,18	0,17	0,16
Suma	3,50	1,83	6,00	1,00	1,00	1,00	1,00

El sub-criterio Portabilidad (ED-SC1), obtuvo la mayor ponderación para MoProPEI (A2), seguida de CRISP-DM (A1), y por último P^{3}TQ (A3) con 0,54, 0,30 y 0,16 respectivamente, como se puede observar en la Tabla 5.196.

Para la Accesibilidad (ED-SC2), el orden de preferencia es el mismo que en el caso de la característica anterior (ED-SC1), obteniendo una mayor diferencia a favor de la A2 (63%), como se indica en la Tabla 5.197.

Tabla 5.197: Ponderación de ED-SC2 en relación a cada alternativa para CV2

Accesibilidad (ED-SC2)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	3	0,23	0,22	0,33	0,26
MoProPEI (A2)	3	1	5	0,69	0,65	0,56	0,63
P^{3}TQ (A3)	1/3	1/5	1	0,08	0,13	0,11	0,11
Suma	4,33	1,53	9,00	1,00	1,00	1,00	1,00

Tabla 5.198: Ponderación de ED-SC3 en relación a cada alternativa para CV2

Diversidad (ED-SC3)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^{3}TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Tabla 5.199: Ponderación de ED-SC4 en relación a cada alternativa para CV2

Diversidad de las fuentes de datos (ED-SC4)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^{3}TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Para la Diversidad de los datos (EDSC3) y la Diversidad de las fuentes de datos (ED-SC4) las metodologías MoProPEI (A2) y CRISP-DM (A1), fueron las mejores ponderadas con un 43% para cada una, tal como se observa en las Tablas 5.198 y 5.199.

A diferencia que los sub-criterios anteriores, para los Recursos necesarios (ED-SC5), CRISP-DM (A1) y MoProPEI (A2) alcanzaron un 43%, tal como se indica en la Tabla 5.200.

Tabla 5.200: Ponderación de ED-SC5 en relación a cada alternativa para CV2

Recursos necesarios (ED-SC5)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/3	0,14	0,14	0,14	0,14
MoProPEI (A2)	3	1	1	0,43	0,43	0,43	0,43
P^3TQ (A3)	3	1	1	0,43	0,43	0,43	0,43
Suma	7,00	2,33	2,33	1,00	1,00	1,00	1,00

En cuanto al Uso de los Datos, tal como se indicó en las Tablas 5.41 y 5.42, el orden de relevancia de los sub-criterios, ubica en primer lugar a la Calidad (ED-SC6), seguido de la Completitud (ED-SC7), Datos riesgosos (ED-SC10), Funcionalidad (ED-SC8) y por último los Costos de transformación (ED-SC9).

Las ponderaciones obtenidas para cada sub-criterio dentro del ED-UD en relación a las metodologías (alternativas) evaluadas, se presentan desde la Tabla 5.201 a la 5.205. Donde el orden de adecuación de las metodologías para el CV2, ubica primeramente a MoProPEI (A2), seguido de CRISP-DM (A2) y en último lugar P^3TQ (A3). En el caso de la Completitud de los datos (ED-SC7), las alternativas 1 y 2 alcanzaron la misma ponderación del 43%, para las demás características A2 fue superior al 50%.

Tabla 5.201: Ponderación de ED-SC6 en relación a cada alternativa para CV2

Calidad (ED-SC6)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	3	0,30	0,29	0,38	0,32
MoProPEI (A2)	2	1	4	0,60	0,57	0,50	0,56
P^3TQ (A3)	1/3	1/4	1	0,10	0,14	0,13	0,12
Suma	3,33	1,75	8,00	1,00	1,00	1,00	1,00

Tabla 5.202: Ponderación de ED-SC7 en relación a cada alternativa para CV2

Completitud (ED-SC7)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^3TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Tabla 5.203: Ponderación de ED-SC8 en relación a cada alternativa para CV2

Funcionalidad (ED-SC8)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	2	0,22	0,21	0,29	0,24
MoProPEI (A2)	3	1	4	0,67	0,63	0,57	0,62
P^3TQ (A3)	1/2	1/4	1	0,11	0,16	0,14	0,14
Suma	4,50	1,58	7,00	1,00	1,00	1,00	1,00

Tabla 5.204: Ponderación de ED-SC9 en relación a cada alternativa para CV2

Costos de transformación (ED-SC9)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM **(A1)**	1	1/3	4	0,24	0,22	0,36	0,27
MoProPEI (A2)	3	1	6	0,71	0,67	0,55	0,64
P^3TQ (A3)	1/4	1/6	1	0,06	0,11	0,09	0,09
Suma	4,25	1,50	11,00	1,00	1,00	1,00	1,00

Tabla 5.204: Ponderación de ED-SC10 en relación a cada alternativa para CV2

Datos riesgosos (ED-SC10)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	4	0,31	0,29	0,40	0,33
MoProPEI (A2)	2	1	5	0,62	0,59	0,50	0,57
P^3TQ (A3)	1/4	1/5	1	0,08	0,12	0,10	0,10
Suma	3,25	1,70	10,00	1,00	1,00	1,00	1,00

Al situarse en el tercer nivel, la Comprensión del Negocio (CN) se divide en tres sub-criterios: EN, PEN y DSPN. Al compararlos las valoraciones dadas por los expertos son las mismas que se expusieron en las Tablas 5.48 y 5.49, siendo DS en el Proyecto del Negocio (DSPN) la de mayor relevancia con el 60% de ponderación.

Para el Entorno del Negocio y sus siete sub-criterios, en la Tabla 5.50 se mostraron los juicios de preferencias emitidos por el experto y en la Tabla 5.51 las ponderaciones calculadas.

Tabla 5.206: Ponderación de CN-SC1 en relación a cada alternativa para CV2

Situación económica circundante (CN-SC1)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^3TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Tabla 5.207: Ponderación de CN-SC2 en relación a cada alternativa para CV2

Comprensión general de la organización (CN-SC2)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^3TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Tabla 5.208: Ponderación de CN-SC3 en relación a cada alternativa para CV2

Definición de los objetivos del negocio (CN-SC3)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^3TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Para los sub-criterios del EN, las ponderaciones obtenidas para cada una en relación al grado de cumplimiento de las metodologías evaluadas; se exhiben de la Tabla 5.206 a la 5.212. Dado que los expertos valoraron a las dos primeras alternativas con 9 y la última con 7 (ver Tabla 5.193), en este caso de validación (CV2) las ponderaciones obtenidas para las siete características son las mismas, donde tanto CRISP-DM (A1), como MoProPEI (A2) lograron una ponderación del 43% y en el caso de P^{3}TQ del 14%.

Tabla 5.209: Ponderación de CN-SC4 en relación a cada alternativa para CV2

Definición del problema del negocio (CN-SC4)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^{3}TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Tabla 5.210: Ponderación de CN-SC5 en relación a cada alternativa para CV2

Definición de las oportunidades del negocio (CN-SC5)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^{3}TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Tabla 5.211: Ponderación de CN-SC6 en relación a cada alternativa para CV2

Definición de requisitos del negocio (CN-SC6)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^{3}TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00	1,00	1,00	1,00	1,00

Tabla 5.212: Ponderación de CN-SC7 en relación a cada alternativa para CV2

Definición de criterios de éxito del proyecto (CN-SC7)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	3	0,43	0,43	0,43	0,43
MoProPEI (A2)	1	1	3	0,43	0,43	0,43	0,43
P^{3}TQ (A3)	1/3	1/3	1	0,14	0,14	0,14	0,14
Suma	2,33	2,33	7,00				

En las Tablas 5.59 y 5.60 se mostraron las valoraciones y ponderaciones del nivel 4, para los cinco sub-criterios dentro del Proyecto del Entorno del Negocio, siendo estás comunes en ambos casos de validación.

Para el sub-criterio PEN en el caso cuyo propósito es determinar los causales de deserción universitaria (CV2), las ponderaciones realizadas para sus cinco criterios y las tres metodologías evaluadas, se encuentran en las Tablas 5.213 a la 5.217, visibles a continuación.

Tabla 5.213: Ponderación de CN-SC8 en relación a cada alternativa para CV2

Conocimiento de los RRHH del negocio (CN-SC8)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	1/2	0,20	0,20	0,20	0,20
MoProPEI (A2)	2	1	1	0,40	0,40	0,40	0,40
P^{3}TQ (A3)	2	1	1	0,40	0,40	0,40	0,40
Suma	5,00	2,50	2,50	1,00	1,00	1,00	1,00

Tabla 5.214: Ponderación de CN-SC9 en relación a cada alternativa para CV2

Disponibilidad de los RRHH del negocio (CN-SC9)				**Matriz Normalizada**			**Ponderación**
Alternativas	1	1/2	1/2				
CRISP-DM (A1)	2	1	1	0,20	0,20	0,20	0,20
MoProPEI (A2)	2	1	1	0,40	0,40	0,40	0,40
P^{3}TQ (A3)	5,00	2,50	2,50	0,40	0,40	0,40	0,40
Suma	1	1/2	1/2	1,00	1,00	1,00	1,00

Tabla 5.215: Ponderación de CN-SC10 en relación a cada alternativa para CV2

Riesgos del problema del negocio (CN-SC10)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	3	0,30	0,29	0,38	0,32
MoProPEI (A2)	2	1	4	0,60	0,57	0,50	0,56
P^3TQ (A3)	1/3	1/4	1	0,10	0,14	0,13	0,12
Suma	3,33	1,75	8,00	1,00	1,00	1,00	1,00

Tabla 5.216: Ponderación de CN-SC11 en relación a cada alternativa para CV2

Contingencias en relación a los riesgos del problema del negocio (CN-SC11)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/2	3	0,30	0,29	0,38	0,32
MoProPEI (A2)	2	1	4	0,60	0,57	0,50	0,56
P^3TQ (A3)	1/3	1/4	1	0,10	0,14	0,13	0,12
Suma	3,33	1,75	8,00	1,00	1,00	1,00	1,00

Tabla 5.217: Ponderación de CN-SC12 en relación a cada alternativa para CV2

Viabilidad del proyecto (CN-SC12)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/4	5	0,19	0,18	0,36	0,24
MoProPEI (A2)	4	1	8	0,77	0,73	0,57	0,69
P^3TQ (A3)	1/5	1/8	1	0,04	0,09	0,07	0,07
Suma	5,20	1,38	14,00	1,00	1,00	1,00	1,00

Tanto para el sub-criterio Conocimiento de los RRHH del negocio (CN-SC8), como el referido a su disponibilidad (CN-SC9), alcanzaron las mismas ponderaciones, donde MoProPEI y, PT^3Q obtuvieron un 40% y CRISP-DM un 20%, como se divisan en las Tablas 5.213 y 5.214. En cuanto a las características Riesgos del problema del negocio (CN-SC10) y las Contingencias en relación a los mismos, MoProPEI obtuvo una ponderación de 0,56, seguido de CRISP-DM con 0,32 y P^3TQ con 0,12, como se muestra en las Tablas 5.215 y

5.216. La Viabilidad del proyecto opta por la metodología MoProPEI con una ponderación del 69%, tal como se visualiza en la Tabla 5.217.

Para el sub-criterio Data Science en el Proyecto del Negocio y sus tres características las preferencias de los expertos y las ponderaciones obtenidas se indicaron en la Tabla 5.66 y 5.67, respectivamente.

En cuanto la característica Objetivos de la aplicación de la metodología de DS (CN-SC13) y su relación con las metodologías comparadas, MoProPEI y CRISP-DM, lograron una ponderación del 40% cada una, como se muestra en la Tabla 5.218.

Tabla 5.218: Ponderación de CN-SC13 en relación a cada alternativa para CV2

Objetivos de la aplicación de la metodología de DS (CN-SC13)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1	2	0,40	0,40	0,40	0,40
MoProPEI (A2)	1	1	2	0,40	0,40	0,40	0,40
P^{3}TQ (A3)	1/2	1/2	1	0,20	0,20	0,20	0,20
Suma	2,50	2,50	5,00	1,00	1,00	1,00	1,00

Para Conocimiento de los RRHH del proyecto de DS (CN-SC14), en la Tabla 5.219 se exhibe que MoProPEI y P^{3}TQ alcanzaron una ponderación de 0,43.

Tabla 5.219: Ponderación de CN-SC14 en relación a cada alternativa para CV2

Conocimiento de los RRHH del proyecto de DS (CN-SC14)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/3	0,14	0,14	0,14	0,14
MoProPEI (A2)	3	1	1	0,43	0,43	0,43	0,43
P^{3}TQ (A3)	3	1	1	0,43	0,43	0,43	0,43
Suma	7,00	2,33	2,33	1,00	1,00	1,00	1,00

Finalmente para el sub-criterio RRHH disponibles del proyecto de DS (CN-SC15), MoProPEI adquirió mayor envergadura que la demás (54%), seguido de P^{3}TQ (30%) y en menor medida CRISP-DM (16%), como se visualiza en la Tabla 5.220.

Tabla 5.220: Ponderación de CN-SC15 en relación a cada alternativa para CV2

RRHH disponibles del proyecto de DS (CN-SC15)				**Matriz Normalizada**			**Ponderación**
Alternativas	**A1**	**A2**	**A3**				
CRISP-DM (A1)	1	1/3	1/2	0,17	0,18	0,14	0,16
MoProPEI (A2)	3	1	2	0,50	0,55	0,57	0,54
P^{3}TQ (A3)	2	1/2	1	0,33	0,27	0,29	0,30
Suma	6,00	1,83	3,50	1,00	1,00	1,00	1,00

5.3.3. Evaluación de consistencia del CV2

Como se indicó en la sección 5.2.3, para matrices de dos entradas, como los ubicados en el segundo nivel (ED y CN) y los dos los sub-criterios involucrados en el ED (AD y UD); no es necesario verificar la consistencia de los juicios emitidos por los expertos; a partir de las matrices de tres criterios (o sub-criterios) se requiere comprobar la misma.

Tabla 5.221: Autovector máximo de cada sub-criterio del ED-AD para CV2

Sub-criterios ED-AD	**Alternativas**	**WxP**	**Cociente**	$\lambda_{máx}$
ED-SC1	CRISP-DM (A1)	0,89	3,01	3,01
	MoProPEI (A2)	1,62	3,01	
	P^{3}TQ (A3)	0,49	3,00	
ED-SC2	CRISP-DM (A1)	0,79	3,03	3,04
	MoProPEI (A2)	1,95	3,07	
	P^{3}TQ (A3)	0,32	3,01	
ED-SC3	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	
ED-SC4	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	
ED-SC5	CRISP-DM (A1)	0,43	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	1,29	3,00	

En las Tablas 5.71 y 5.72 se exhibieron los pasos realizados para la evaluación de consistencia de los sub-criterios del AD de manera conjunta. Siguiendo el mismo procedimiento, se evaluó la consistencia de los cinco sub-criterios por separado vinculándolos con cada alternativa (metodologías de DS).

Los valores intermedios necesarios para obtener el autovector máximo $\lambda_{máx}$ de los sub-criterios: ED-SC1 (Portabilidad), ED-SC2 (Accesibilidad), ED-SC3 (Diversidad), ED-SC4 (Diversidad de fuentes de datos) y ED-SC5 (Recursos necesarios), se muestran en la Tabla 5.221.

Al evaluar la razón de consistencia de los sub-criterios mencionados, se comprobó que los mismos obtuvieron un CR menor a 0,10; tal como se observa la Tabla 5.222; donde se puede ver que para las últimas tres características es perfecto.

Tabla 5.222: Evaluación de consistencia de los sub-criterios del ED-AD para CV2

Evaluación de consistencia	**ED-SC1**	**ED-SC2**	**ED-SC3**	**ED-SC4**	**ED-SC5**
$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,005	0,019	0,000	0,000	0,000
$\boldsymbol{RI}$	0,525	0,525	0,525	0,525	0,525
$\boldsymbol{CR = CI/RI}$	**0,009**	**0,037**	**0,000**	**0,000**	**0,000**

En relación al sub-criterio Uso de los Datos (UD), en las Tablas 5.75 y 5.76 se encuentra los cálculos intermedios y la razón de consistencia obtenida para sus cinco sub-criterios en conjunto; y en la las Tablas 5.223 y 5.224 cada uno de ellos por separado y en función al nivel de cumplimiento de cada metodología en el caso de validación (CV2).

Tabla 5.223: Autovector máximo de cada sub-criterio del ED-UD para CV2

Sub-criterios ED-UD	**Alternativas**	**WxP**	**Cociente**	$\boldsymbol{\lambda_{máx}}$
ED-SC6	CRISP-DM (A1)	1,05	2,97	3,02
	MoProPEI (A2)	1,83	3,08	
	P^3TQ (A3)	0,40	3,01	
ED-SC7	CRISP-DM (A1)	1,33	3,00	3,00
	MoProPEI (A2)	1,33	3,00	
	P^3TQ (A3)	0,44	3,00	

Sub-criterios ED-UD	Alternativas	WxP	Cociente	$\lambda_{máx}$
ED-SC8	CRISP-DM (A1)	0,79	2,97	3,02
	MoProPEI (A2)	2,09	3,02	
	P^3TQ (A3)	0,45	3,06	
ED-SC9	CRISP-DM (A1)	1,08	2,87	3,06
	MoProPEI (A2)	2,59	3,24	
	P^3TQ (A3)	0,34	3,08	
ED-SC10	CRISP-DM (A1)	1,20	2,93	3,03
	MoProPEI (A2)	2,05	3,15	
	P^3TQ (A3)	0,35	3,01	

En la Tabla 5.224; se puede notar que la característica Completitud de los datos (ED-SC7) es perfectamente consistente y para las demás, inferior a 0,10 siendo todas ellas adecuadas.

Tabla 5.224: Evaluación de consistencia de los sub-criterios del ED-UD para CV2

Evaluación de consistencia	ED-SC6	ED-SC7	ED-SC8	ED-SC9	ED-SC10
$CI = (\lambda_{máx}$-n)/(n-1)	0,009	0,000	0,009	0,031	0,014
RI	0,525	0,525	0,525	0,525	0,525
$CR = CI/RI$	**0,006**	**0,000**	**0,006**	**0,021**	**0,009**

En cuanto a la Comprensión del Negocio (CN), la evaluación de consistencia de sus tres sub-criterios principales (Entorno del Negocio (CN-EN), Proyecto en el Entorno del Negocio (CN-PEN) y DS en el Proyecto del Negocio (CN-DSPN)) en conjunto, se detallaron en las Tablas 5.79 y 5.80, siendo estos totalmente consistentes.

En las Tablas 5.81 y 5.82, se expusieron la evaluación de los siete sub-criterios del Entorno del Negocio (EN), para los cuales se alcanzó un autovector de 7,29 y una razón de consistencia de 0,016.

En la Tabla 5.225 determina para las características: Situación económica circundante (CN-SC1), Comprensión general de la organización (CN-SC2), Definición de los objetivos del negocio (CN-SC3), Definición del problema del negocio (CN-SC4), Definición de las oportunidades del negocio (CN-SC5), Definición de requisitos del negocio (CN-SC6) y Definición de criterios de éxito del proyecto (CN-SC7); que $\lambda_{máx}$ logrado por cada una de

ellas, es igual al orden de la matriz (Tabla 5.225), por lo tanto su razón de consistencia es inmejorable, tal como se puede apreciar en la Tabla 5.226.

Tabla 5.225: Autovector máximo de cada sub-criterio del CN-EN para CV2

Sub-criterios CN-EN	Alternativas	WxP	Cociente	$\lambda_{máx}$
CN-SC1	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	
CN-SC2	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	
CN-SC3	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	
CN-SC4	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	
CN-SC5	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	
CN-SC6	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	
CN-SC7	CRISP-DM (A1)	1,29	3,00	3,00
	MoProPEI (A2)	1,29	3,00	
	P^{3}TQ (A3)	0,43	3,00	

Tabla 5.226: Evaluación de consistencia de los sub-criterios del CN-EN para CV2

Evaluación de consistencia	CN-SC1	CN-SC2	CN-SC3	CN-SC4	CN-SC5	CN-SC6	CN-SC7
$CI = (\lambda_{máx}$-n)/(n-1)	0,000	0,000	0,000	0,000	0,000	0,000	0,000
RI	0,525	0,525	0,525	0,525	0,525	0,525	0,525
$CR = CI/RI$	**0,000**	**0,000**	**0,000**	**0,000**	**0,000**	**0,000**	**0,000**

Tabla 5.227: Autovector máximo de cada sub-criterio del CN-PEN para CV2

Sub-criterios CN-PEN	Alternativas	WxP	Cociente	$\lambda_{máx}$
CN-SC8	CRISP-DM (A1)	0,49	3,00	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,89	3,01	
CN-SC9	CRISP-DM (A1)	0,49	3,00	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,89	3,01	
CN-SC10	CRISP-DM (A1)	0,60	3,00	3,00
	MoProPEI (A2)	1,80	3,00	
	P^3TQ (A3)	0,60	3,00	
CN-SC11	CRISP-DM (A1)	0,60	3,00	3,00
	MoProPEI (A2)	1,80	3,00	
	P^3TQ (A3)	0,60	3,00	
CN-SC12	CRISP-DM (A1)	0,97	3,03	3,02
	MoProPEI (A2)	1,69	3,01	
	P^3TQ (A3)	0,37	3,02	

El Proyecto en el Entorno del Negocio (CN-PEN), se encuentra dividido en cinco sub-criterios (Conocimiento de los RRHH del negocio (CN-SC8), Disponibilidad de los RRHH del negocio (CN-SC9), Riesgos del problema del negocio (CN-SC10), Contingencias en relación a los riesgos del problema del negocio (CN-SC11) y Viabilidad del proyecto (CN-SC12)), cuya estimación de consistencia se presentó en las Tablas 5.85 y 5.86, la cual es inferior al 10%. En la Tabla 5.227 se indica la obtención del autovector máximo de las cinco características mencionadas, a partir del cual se estimó la razón de consistencia (CR) para cada una de ellas, visible en la Tabla 5.228; la cual revela que todas son adecuadas y que CN-SC10 y CN-SC11 son totalmente coherentes.

Tabla 5.228: Evaluación de consistencia de los sub-criterios del CN-PEN para CV2

Evaluación de consistencia	CN-SC8	CN-SC9	CN-SC10	CN-SC11	CN-SC12
$CI = (\lambda_{máx}$-n)/(n-1)	0,005	0,005	0,000	0,000	0,009
RI	0,525	0,525	0,525	0,525	0,525
$CR = CI/RI$	**0,009**	**0,009**	**0,000**	**0,000**	**0,017**

Finalmente, para DS en el Proyecto del Negocio (CN-DSPN), el autovector máximo obtenido para las tres características (Objetivos de la aplicación de la metodología de DS (CN-SC13), Conocimiento de los RRHH del proyecto de DS (CN-SC14) y RRHH disponibles del proyecto de DS (CN-SC15), se expuso en la Tabla 5.89 y su razón de consistencia en la Tabla 5.90. Dichas características evaluadas en función al grado de cumplimiento de CRISP-DM (A1), MoProPEI (A2) y P3TQ (A3), se puede visualizar en las Tablas 5.229 y 5.230, para las cuales son adecuadas las valoraciones realizada por los expertos y cumplen con lo establecido (menor a 0,10).

Tabla 5.229: Autovector máximo de cada sub-criterio del CN-DSPN para CV2

Sub-criterios CN- DSPN	Alternativas	WxP	Cociente	$\lambda_{máx}$
CN-SC13	CRISP-DM (A1)	0,89	3,01	3,01
	MoProPEI (A2)	1,62	3,01	
	P^3TQ (A3)	0,49	3,00	
CN-SC14	CRISP-DM (A1)	0,72	3,01	3,02
	MoProPEI (A2)	1,89	3,03	
	P^3TQ (A3)	0,41	3,01	
CN-SC15	CRISP-DM (A1)	0,72	3,01	3,02
	MoProPEI (A2)	1,89	3,03	
	P^3TQ (A3)	0,41	3,01	

Tabla 5.230: Evaluación de consistencia de los sub-criterios del CN-DSPN para CV2

Evaluación de consistencia	CN-SC13	CN-SC14	CN-SC15
$\boldsymbol{CI = (\lambda_{máx}}$**-n)/(n-1)**	0,005	0,009	0,009
$\boldsymbol{RI}$	0,525	0,525	0,525
$\boldsymbol{CR = CI/RI}$	**0,009**	**0,017**	**0,017**

5.3.4. Desarrollo de los métodos de MCDM para el CV2

Como se indicó en el Capítulo anterior (IV) y en la sección 5.2.4; el desarrollo de los métodos de MCDM, se realiza con el método AHP (pasos faltantes) y con FAHP (ver Figura 4.4).

5.3.4.1 Desarrollo de AHP para el CV2

Como continuidad del método AHP, para el CV2 se establecieron las prioridades locales y globales (ver Anexo I), a través del cálculo del peso relativo a los criterios por nivel; las locales obtenidas en el paso 5.3.3, mediante las cuales se obtuvieron las globales. Finalmente, a través del método de la suma ponderada, se alcanzaron las prioridades totales asociadas a cada alternativa.

En la Tabla 5.231 se muestran las prioridades locales del Acceso a los Datos (ED-AD) en relación a las tres metodologías, calculadas a partir de las ponderaciones de sus cinco características, para el cual MoProPEI es la metodología preferida con el 49%, seguido de CRISP-DM con el 36% y P^3TQ con el 16%.

Para el sub-criterios Uso de los Datos (ED-UD), la prioridad local obtenida se expone en la Tabla 5.232, donde nuevamente la metodología MoProPEI es la selecta con una prioridad del 54% en el CV2.

Tabla 5.231: Prioridad local de ED-AD para CV2

Sub-criterio/ Alternativa	ED-SC1	ED-SC2	ED-SC3	ED-SC4	ED-SC5	Prioridad Local
CRISP-DM (A1)	0,30	0,26	0,43	0,43	0,14	0,36
MoProPEI (A2)	0,54	0,63	0,43	0,43	0,43	0,49
P^3TQ (A3)	0,16	0,11	0,14	0,14	0,43	0,16
Ponderación ED-AD	0,07	0,24	0,25	0,37	0,07	

Tabla 5.232: Prioridad local de ED-UD para CV2

Sub-criterio/ Alternativa	ED-SC6	ED-SC7	ED-SC8	ED-SC9	ED-SC10	Prioridad Local
CRISP-DM (A1)	0,32	0,43	0,24	0,27	0,33	0,34
MoProPEI (A2)	0,56	0,43	0,62	0,64	0,57	0,54
P^3TQ (A3)	0,12	0,14	0,14	0,09	0,10	0,12
Ponderación ED-UD	0,33	0,27	0,13	0,05	0,21	

Las prioridades globales se alcanzaron a partir de las prioridades locales de ED-AC y ED-UD; como se indicó anteriormente el uso de los datos es considerado más relevante que el acceso a los mismos, tal como se expone en las Tablas 5.233 y 5.234.

Tabla 5.233: Prioridad global de ED-AD para CV2

Sub-criterio/ Alternativa	Prioridad Local	Ponderación ED-AD	Prioridad Global
CRISP-DM (A1)	0,36	0,25	0,09
MoProPEI (A2)	0,49		0,12
P^{3}TQ (A3)	0,16		0,04

Tabla 5.234: Prioridad global de ED-UD para CV2

Sub-criterio/ Alternativa	Prioridad Local	Ponderación ED-UD	Prioridad Global
CRISP-DM (A1)	0,34	0,75	0,26
MoProPEI (A2)	0,54		0,40
P^{3}TQ (A3)	0,12		0,09

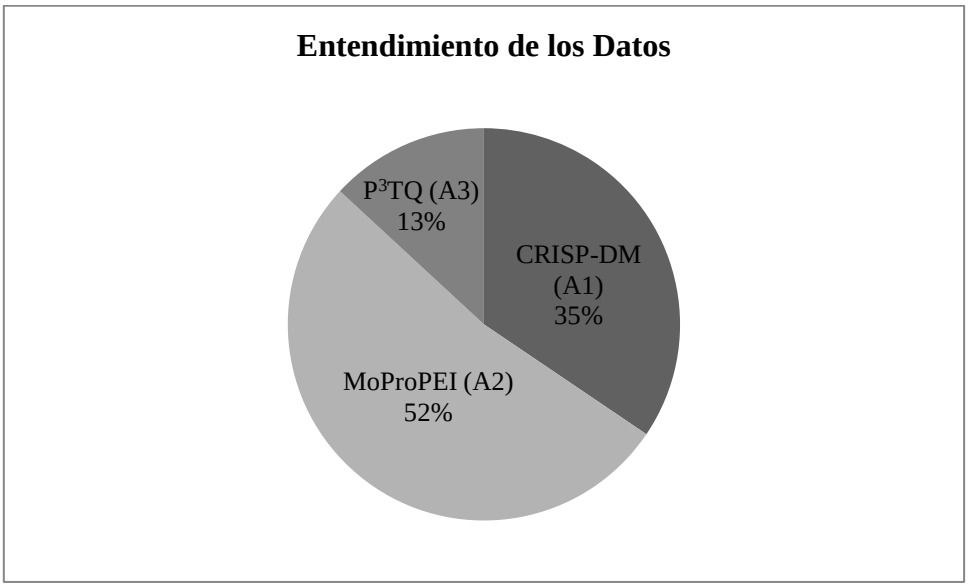

Figura 5.6: Ponderaciones globales del ED para el CV2 con AHP

Al tomar la rama izquierda de la jerarquía propuesta en la Figura 4.2, los porcentajes producidos se visualizan en el gráfico de la Figura 5.6. Como el Entendimiento de los Datos representa el 50% de relevancia en los proyectos de DS, las ponderaciones globales son del 26% para MoProPEI, 17% para CRISP-DM y del 7% para P^3TQ.

Tabla 5.235: Prioridad local de CN-EN para CV2

Sub-criterio/ Alternativa	CN-SC1	CN-SC2	CN-SC3	CN-SC4	CN-SC5	CN-SC6	CN-SC7	Prioridad Local
CRISP-DM (A1)	0,43	0,43	0,43	0,43	0,43	0,43	0,43	0,43
MoProPEI (A2)	0,43	0,43	0,43	0,43	0,43	0,43	0,43	0,43
P^3TQ (A3)	0,14	0,14	0,14	0,14	0,14	0,14	0,14	0,14
Ponderación CN-EN	0,05	0,16	0,16	0,16	0,16	0,14	0,19	

En la Tabla 5.235 se encuentran las prioridades locales del Entorno del Negocio, para las tres alternativas; donde MoProPEI y CRISP-DM obtuvieron el 43% y P^3TQ el 14%. En el caso del Proyecto en el Entorno del Negocio y sus cinco características, las prioridades locales obtenidas se muestran en la Tabla 5.236, donde una vez más MoProPEI es la metodología escogida, con una prioridad de 0,54. Y para DS en el Proyecto del Negocio (DSPN), como se puede ver en la Tabla 5.237**,** las prioridades locales son del 43% para MoProPEI, del 32% para CRISP-DM y del 25% para P^3TQ.

Tabla 5.236: Prioridad local de CN-PEN para CV2

Sub-criterio/ Alternativa	CN-SC8	CN-SC9	CN-SC10	CN-SC11	CN-SC12	Prioridad Local
CRISP-DM (A1)	0,20	0,20	0,32	0,32	0,24	0,28
MoProPEI (A2)	0,40	0,40	0,56	0,56	0,69	0,54
P^3TQ (A3)	0,40	0,40	0,12	0,12	0,07	0,18
Ponderación CN-PEN	0,12	0,11	0,23	0,34	0,19	

Tabla 5.237: Prioridad local de CN-DSPN para CV2

Sub-criterio/ Alternativa	CN-SC13	CN-SC14	CN-SC15	Prioridad Local
CRISP-DM (A1)	0,40	0,14	0,16	0,32
MoProPEI (A2)	0,40	0,43	0,54	0,43
P^{3}TQ (A3)	0,20	0,43	0,30	0,25
Ponderación CN-DSPN	0,66	0,16	0,19	

Tabla 5.238: Prioridad global de CN-EN para CV2

Sub-criterio/ Alternativa	Prioridad Local	Ponderación CN-EN	Prioridad Global
CRISP-DM (A1)	0,43	0,20	0,09
MoProPEI (A2)	0,43		0,09
P^{3}TQ (A3)	0,14		0,03

Tabla 5.239: Prioridad global de CN-PEN para CV2

Sub-criterio/ Alternativa	Prioridad Local	Ponderación CN-PEN	Prioridad Global
CRISP-DM (A1)	0,28	0,20	0,06
MoProPEI (A2)	0,54		0,11
P^{3}TQ (A3)	0,18		0,04

Tabla 5.240: Prioridad global de CN-DSPN para CV2

Sub-criterio/ Alternativa	Prioridad Local	Ponderación CN-DSPN	Prioridad Global
CRISP-DM (A1)	0,32	0,60	0,19
MoProPEI (A2)	0,43		0,26
P^{3}TQ (A3)	0,25		0,15

Las prioridades globales del Entorno del Negocio (EN) se encuentran en la Tabla 5.238, para el cual CRISP-DM (A1) y MoProPEI (A2) cuentan con la misma preponderancia. En cuanto al Proyecto en el EN (PEN), en la Tabla 5.239 se puede ver que A2 es superior a las restantes

y en la Tabla 5.240 para CN-DSPN el predominio es por parte de la segunda alternativa (A2). La suma de todas las prioridades corresponden al 100% de la Comprensión del Negocio, donde EN y PEN representan el 20% cada uno y DSPN el 60%.

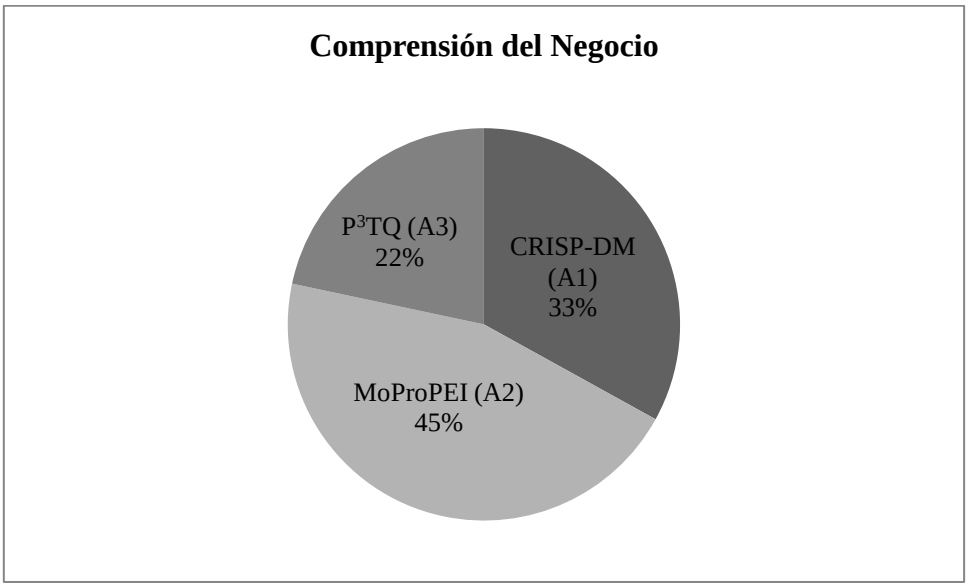

Figura 5.7: Ponderaciones globales de la CN para el CV2 con AHP

Al igual que el Entendimiento de los Datos, la Comprensión del Negocio representa el 50% de trascendencia al momento de elegir una metodología de DS. La ponderación global calculada indica que MoProPEI representa el 23%, CRISP-DM el 17% y P^{3}TQ el 11%; que expresando en un 100%, teniendo en cuenta únicamente la CN, los porcentajes obtenidos son los expuestos en la Figura 5.7.

Tabla 5.241: Prioridad total con AHP para el CV2

Alternativas	Prioridad Final ED	Prioridad Final CN	Prioridad Total
CRISP-DM (A1)	0,17	0,17	0,34
MoProPEI (A2)	0,26	0,23	0,49
P^{3}TQ (A3)	0,07	0,11	0,17

Para finalizar, se realizó el establecimiento de prioridades totales con AHP en relación a cada metodología, a partir de las prioridades mostradas (locales y globales). En la Tabla 5.241 se expone la prioridad total de cada metodología en el proyecto de DS utilizado como caso de

validación (CV2); en dicha tabla se puede evidenciar la preferencia de la segunda alternativa (A2) por sobre las demás, dado que MoProPEI se adecuó en un 49%, luego CRISP-DM (A1) con un 34% y en tercer lugar P^{3}TQ (A3) con un 17%.

5.3.4.2 Desarrollo de FAHP para el CV2

Continuando con el paso 6.2 del modelo propuesto en el Capítulo IV, el desarrollo del método FAHP, el cual se funda en los pasos descriptos en el Anexo II. Se recuerda que se utiliza una función de membrecía con números difusos triangulares; que a partir de las matrices de dos entradas y las de tipo parrillas se asignó el valor más prometedor (m) y valores de l (valor pesimista) y u (valor optimista), se surgen de la escala de conversión difusa triangular disponible en la Tabla II.I del Anexo II.
La matriz triangular difusa de comparación pareada del segundo nivel, donde se confrontan el Entendimiento de los Datos (ED) y la Comprensión del Negocio (CN), se encuentra disponible en la Tabla 5.104, de la sección 5.2.4.2. En la misma sección, en la Tabla 5.105, está la matriz difusa para el ED; para luego continuar con sus dos sub-criterios.

Tabla 5.242: Matriz triangular difusa de comparación pareada de ED-SC1 para CV2

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	1	2	3
MoProPEI (A2)	1	2	3	1	1	1	2	3	4
P^{3}TQ (A3)	1/3	1/2	1	1/4	1/3	1/2	1	1	1
Suma	2,33	3,50	5,00	1,58	1,83	2,50	4,00	6,00	8,00

Tabla 5.243: Matriz triangular difusa de comparación pareada de ED-SC2 para CV2

Alternativas	**A1**			**A2**			**A3**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	2	3	4
MoProPEI (A2)	2	3	4	1	1	1	4	5	6
P^{3}TQ (A3)	1/4	1/3	1/2	1/6	1/5	1/4	1	1	1
Suma	3,25	4,33	5,50	1,42	1,53	1,75	7,00	9,00	11,00

En la Tabla 5.106 se mostró la matriz triangular difusa del Acceso a los Datos (ED-AD) para sus cinco características. Para el CV2, las matrices de comparación pareada difusas de las características del AD, se muestran desde la Tabla 5.242 a la 5.246.

Tabla 5.244: Matriz triangular difusa de comparación pareada de ED-SC3 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^3TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.245: Matriz triangular difusa de comparación pareada de ED-SC4 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^3TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.246: Matriz Triangular Difusa de Comparación Pareada de ED-SC5 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/4	1/3	1/2
MoProPEI (A2)	2	3	4	1	1	1	1	1	2
P^3TQ (A3)	2	3	4	1/2	1	1	1	1	1
Suma	5,00	7,00	9,00	1,75	2,33	2,50	2,25	2,33	3,50

La matriz de comparación pareada difusa del Uso de los Datos (ED-UD) se expuso en la Tabla 5.112. Para el presente caso de validación, al evaluar las cinco características del UD vinculados a la completitud de los mismos respecto a cada metodología, se detallan en la Tabla 5.247 a la Tabla 5.251.

Tabla 5.247: Matriz triangular difusa de comparación pareada de ED-SC6 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	2	3	4
MoProPEI (A2)	1	2	3	1	1	1	3	4	5
P^3TQ (A3)	1/4	1/3	1/2	1/5	1/4	1/3	1	1	1
Suma	2,25	3,33	4,50	1,53	1,75	2,33	6,00	8,00	10,00

Tabla 5.248: Matriz triangular difusa de comparación pareada de ED-SC7 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^3TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.249: Matriz triangular difusa de comparación pareada de ED-SC8 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1	2	3
MoProPEI (A2)	2	3	4	1	1	1	3	4	5
P^3TQ (A3)	1/3	1/2	1	1/5	1/4	1/3	1	1	1
Suma	3,33	4,50	6,00	1,45	1,58	1,83	5,00	7,00	9,00

Tabla 5.250: Matriz Triangular Difusa de Comparación Pareada de ED-SC9 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	3	4	5
MoProPEI (A2)	2	3	4	1	1	1	5	6	7
P^3TQ (A3)	1/5	1/4	1/3	1/7	1/6	1/5	1	1	1
Suma	3,20	4,25	5,33	1,39	1,50	1,70	9,00	11,00	13,00

Tabla 5.251: Matriz triangular difusa de comparación pareada de ED-SC10 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	3	4	5
MoProPEI (A2)	1	2	3	1	1	1	4	5	6
P^3TQ (A3)	1/5	1/4	1/3	1/6	1/5	1/4	1	1	1
Suma	2,20	3,25	4,33	1,50	1,70	2,25	8,00	10,00	12,00

La comparación difusa triangular para la Comprensión del Negocio (CN) y sus tres sub-criterios, se dispuso en la Tabla 5.118, de la sección 5.2.4.2.

Para el primer sub-criterio de CN, el Entorno del Negocio (EN), en la Tabla 5.119 se exhibió la matriz difusa triangular, que compara las siete características del mismo. Estas características en relación a las alternativas evaluadas, son expuestas de la Tabla 5.252 a la 5.258; cuyas valoraciones son equivalentes para las sietes características.

Tabla 5.252: Matriz triangular difusa de comparación pareada de CN-SC1 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^3TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.253: Matriz Triangular Difusa de Comparación Pareada de CN-SC2 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^3TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.254: Matriz triangular difusa de comparación pareada de CN-SC3 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^{3}TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.255: Matriz triangular difusa de comparación pareada de CN-SC4 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^{3}TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.256: Matriz triangular difusa de comparación pareada de CN-SC5 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^{3}TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.257: Matriz Triangular Difusa de Comparación Pareada de CN-SC6 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^{3}TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

Tabla 5.258: Matriz triangular difusa de comparación pareada de CN-SC7 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	2	3	4
MoProPEI (A2)	1/2	1	1	1	1	1	2	3	4
P^{3}TQ (A3)	1/4	1/3	1/2	1/4	1/3	1/2	1	1	1
Suma	1,75	2,33	2,50	2,25	2,33	3,50	5,00	7,00	9,00

En cuento al Proyecto en el Entorno del Negocio, en la Tabla 5.127 se la sección 5.2.4.2 (se recuerda que para dicho nivel, son los mismos valores para ambos casos de validación) se mostró la matriz triangular difusa, para la comparación de a pares de sus cinco sub-criterios; los cuales al ser analizados en relación a las metodologías, se confeccionaron las matrices expuestas en las Tablas 5.259 a la 5.263, donde las valoraciones de SC8 y SC9 por un lado y SC10 y SC11 por otro coinciden, no así para el caso de SC12.

Tabla 5.259: Matriz triangular difusa de comparación pareada de CN-SC8 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	1/3	1/2	1
MoProPEI (A2)	1	2	3	1	1	1	1	1	2
P^{3}TQ (A3)	1	2	3	1/2	1	1	1	1	1
Suma	3,00	5,00	7,00	1,83	2,50	3,00	2,33	2,50	4,00

Tabla 5.260: Matriz triangular difusa de comparación pareada de CN-SC9 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	1/3	1/2	1
MoProPEI (A2)	1	2	3	1	1	1	1	1	2
P^{3}TQ (A3)	1	2	3	1/2	1	1	1	1	1
Suma	3,00	5,00	7,00	1,83	2,50	3,00	2,33	2,50	4,00

Tabla 5.261: Matriz triangular difusa de comparación pareada de CN-SC10 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	2	3	4
MoProPEI (A2)	1	2	3	1	1	1	3	4	5
P^3TQ (A3)	1/4	1/3	1/2	1/5	1/4	1/3	1	1	1
Suma	2,25	3,33	4,50	1,53	1,75	2,33	6,00	8,00	10,00

Tabla 5.262: Matriz triangular difusa de comparación pareada de CN-SC11 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/3	1/2	1	2	3	4
MoProPEI (A2)	1	2	3	1	1	1	3	4	5
P^3TQ (A3)	1/4	1/3	1/2	1/5	1/4	1/3	1	1	1
Suma	2,25	3,33	4,50	1,53	1,75	2,33	6,00	8,00	10,00

Tabla 5.263: Matriz triangular difusa de comparación pareada de CN-SC12 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/5	1/4	1/3	3	5	5
MoProPEI (A2)	3	4	5	1	1	1	7	8	9
P^3TQ (A3)	1/5	1/5	1/3	1/9	1/8	1/7	1	1	1
Suma	4,20	5,20	6,33	1,31	1,38	1,48	11,00	14,00	15,00

Tabla 5.264: Matriz Triangular Difusa de Comparación Pareada de CN-SC13 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1	1	2	1	2	3
MoProPEI (A2)	1/2	1	1	1	1	1	1	2	3
P^3TQ (A3)	1/3	1/2	1	1/3	1/2	1	1	1	1
Suma	1,83	2,50	3,00	2,33	2,50	4,00	3,00	5,00	7,00

Para el último sub-criterio de la CN, DSPN y sus tres características (CN-SC13, CN-SC14 y CN-SC15), la matriz triangular difusa se mostró en la Tabla 5.133; y en las Tablas 5.264, 5.264 y 5.266, se indican las valoraciones pesimistas (l), prometedores (m) y optimistas (u) para el caso de validación en cuestión (CV2).

Tabla 5.265: Matriz triangular difusa de comparación pareada de CN-SC14 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/4	1/3	1/2
MoProPEI (A2)	2	3	4	1	1	1	1	1	2
P^{3}TQ (A3)	2	3	4	1/2	1	1	1	1	1
Suma	5,00	7,00	9,00	1,75	2,33	2,50	2,25	2,33	3,50

Tabla 5.266: Matriz triangular difusa de comparación pareada de CN-SC15 para CV2

Alternativas	A1			A2			A3		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CRISP-DM (A1)	1	1	1	1/4	1/3	1/2	1/3	1/2	1
MoProPEI (A2)	2	3	4	1	1	1	1	2	3
P^{3}TQ (A3)	1	2	3	1/3	1/2	1	1	1	1
Suma	4,00	6,00	8,00	1,58	1,83	2,50	2,33	3,50	5,00

Como paso siguiente, tal como se indicó en el modelo propuesto, se prosiguió con el cálculo de les valores de extensión sintética difusa (consultar paso 3 del Anexo II) de los diferentes criterios de la jerarquía propuesta; expuestos en las consecutivas tablas, donde se encuentran representadas mediante la Ecuación II.III, la cual proviene de las Ecuaciones II.IV y II.V también disponibles en las tablas. Los valores obtenidos en las tablas, derivan de las matrices de comparación pareada difusa triangular expuestas anteriormente.

Partiendo desde la parte superior de la estructura jerárquica (Figura 4.2), para el segundo nivel, el valor de extensión sintética difusa (Ecuación II.III) se reveló en la Tabla 5.137 de la sección 5.2.4.2. En la misma sección, se exhibió el valor de extensión sintética difusa de cinco sub-criterios del Acceso a los Datos (ED-AD) en la Tabla 5.139.

Para el CV2, en la Tabla 5.267 se congrega los cinco sub-criterios del AD, cada uno examinado en relación a las metodologías CRISP-DM (A1), MoProPEI (A2) y P^{3}TQ (A3). Al analizar el valor más prometedor (m) de la última columna (Ecuación II.III), se puede ver que MoProPEI obtuvo siempre la mayor preferencia, con los sub-criterios ED-SC3 (Diversidad de los datos) y ED-SC4 (Diversidad de las fuentes de datos) logró un empate con CRISP-DM y en el caso de ED-SC5 (Recursos necesarios) con P^{3}TQ.

Tabla 5.267: Extensión sintética difusa de Sub-criterios ED-AD para CV2

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC1									
CRISP-DM	2,33	3,50	5,00	7,92	11,33	15,50	0,15	0,31	0,63
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	1,58	1,83	2,50				0,10	0,16	0,32
ED-SC2									
CRISP-DM	3,25	4,33	5,50	11,67	14,87	18,25	0,18	0,29	0,47
MoProPEI	7,00	9,00	11,00				0,38	0,61	0,94
P^{3}TQ	1,42	1,53	1,75				0,08	0,10	0,15
ED-SC3									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22
ED-SC4									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22
ED-SC5									
CRISP-DM	1,50	1,67	2,00	9,00	11,67	15,00	0,10	0,14	0,22
MoProPEI	4,00	5,00	7,00				0,27	0,43	0,78
P^{3}TQ	3,50	5,00	6,00				0,23	0,43	0,67

En relación al Uso de los Datos (ED-UD), en la Tabla 5.141 se expuso el valor de extensión sintética borrosa calculada al comparar sus cinco sub-criterios. Para el presente caso de

validación, en la Tabla 5.268 se indican los valores obtenidos para los sub-criterios (ED-SC6 a ED-SC10); donde para los valores pesimistas (l), prometedores (m) y optimistas (u), a excepción de ED-SC7 (Completitud de los datos) cuyos valores son similares para las dos primeras alternativas, en los demás casos MoProPEI logró sobresalir.

Tabla 5.268: Extensión sintética difusa de Sub-criterios ED-UD para CV2

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
ED-SC6									
CRISP-DM	3,33	4,50	6,00	9,78	13,08	16,83	0,20	0,34	0,61
MoProPEI	5,00	7,00	9,00				0,30	0,54	0,92
P^{3}TQ	1,45	1,58	1,83				0,09	0,12	0,19
ED-SC7									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22
ED-SC8									
CRISP-DM	2,25	3,33	4,50	9,78	13,08	16,83	0,13	0,25	0,46
MoProPEI	6,00	8,00	10,00				0,36	0,61	1,02
P^{3}TQ	1,53	1,75	2,33				0,09	0,13	0,24
ED-SC9									
CRISP-DM	4,25	5,33	6,50	13,59	16,75	20,03	0,21	0,32	0,48
MoProPEI	8,00	10,00	12,00				0,40	0,60	0,88
P^{3}TQ	1,34	1,42	1,53				0,07	0,08	0,11
ED-SC10									
CRISP-DM	4,33	5,50	7,00	11,70	14,95	18,58	0,23	0,37	0,60
MoProPEI	6,00	8,00	10,00				0,32	0,54	0,85
P^{3}TQ	1,37	1,45	1,58				0,07	0,10	0,14

Tomando la rama derecha de la jerarquía, en la Tabla 5.143, se presentaron los valores de extensión sintética difusa para los tres sub-criterios de la Comprensión del Negocio (CN) (CN-EN (Entorno del Negocio), CN-PEN (Proyecto en el Entorno del Negocio) y CN-DSPN (DS en el Proyecto del Negocio)). Luego en la Tabla 5.144, se expusieron los valores (de

extensión sintética difusa) para los siete sub-criterios de CN-EN, en la Tabla 5.146 para los cinco de CN-PEN y en la Tabla 148 para CN-DSPN.

Para el caso de validación en cuestión (CV2), al confrontar las siete características del EN con las tres metodologías, se calculó el valor de extensión sintética borrosa de cada una de ellas visible en la Tabla 5.269; para las características se obtuvieron los mismos valores y al observar el valor más prometedor (m), se puede apreciar que CRISP-DM y MoProPEI tienen la misma adecuación.

Tabla 5.269: Extensión sintética difusa de Sub-criterios CN-EN para CV2

	Ecuación II.IV			**Ecuación II.V**			**Ecuación II.III**		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC1									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22
CN-SC2									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22
CN-SC3									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22
CN-SC4									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22
CN-SC5									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC6									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22
CN-SC7									
CRISP-DM	4,00	5,00	7,00	9,00	11,67	15,00	0,27	0,43	0,78
MoProPEI	3,50	5,00	6,00				0,23	0,43	0,67
P^{3}TQ	1,50	1,67	2,00				0,10	0,14	0,22

En cuanto a las cinco características de PEN y su cumplimiento en las tres metodologías confrontadas, la extensión difusa sintética borrosa de cada una de ellas, se muestra en la Tabla 5.270; donde para cada uno de los sub-criterios MoProPEI obtuvo mejores valoraciones.
Finalmente al comparar los tres sub-criterios de DSPN en relación a CRISP-DM, MoProPEI y P^{3}TQ, se puede apreciar en la Tabla 5.271 que la segunda alternativa sobresale por sobre las demás.

Tabla 5.270: Extensión sintética difusa de Sub-criterios CN-PEN para CV2

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC8									
CRISP-DM	1,67	2,00	3,00	7,17	10,00	14,00	0,12	0,20	0,42
MoProPEI	3,00	4,00	6,00				0,21	0,40	0,84
P^{3}TQ	2,50	4,00	5,00				0,18	0,40	0,70
CN-SC9									
CRISP-DM	1,67	2,00	3,00	7,17	10,00	14,00	0,12	0,20	0,42
MoProPEI	3,00	4,00	6,00				0,21	0,40	0,84
P^{3}TQ	2,50	4,00	5,00				0,18	0,40	0,70
CN-SC10									
CRISP-DM	3,33	4,50	6,00	9,78	13,08	16,83	0,20	0,34	0,61
MoProPEI	5,00	7,00	9,00				0,30	0,54	0,92
P^{3}TQ	1,45	1,58	1,83				0,09	0,12	0,19

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC11									
CRISP-DM	3,33	4,50	6,00	9,78	13,08	16,83	0,20	0,34	0,61
MoProPEI	5,00	7,00	9,00				0,30	0,54	0,92
P^{3}TQ	1,45	1,58	1,83				0,09	0,12	0,19
CN-SC12									
CRISP-DM	4,20	6,25	6,33	16,51	20,58	22,81	0,18	0,30	0,38
MoProPEI	11,00	13,00	15,00				0,48	0,63	0,91
P^{3}TQ	1,31	1,33	1,48				0,06	0,06	0,09

Tabla 5.271: Extensión sintética difusa de Sub-criterios CN-DSPN para CV2

	Ecuación II.IV			Ecuación II.V			Ecuación II.III		
	l	**m**	**u**	**l**	**m**	**u**	**l**	**m**	**u**
CN-SC13									
CRISP-DM	3,00	4,00	6,00	7,17	10,00	14,00	0,21	0,40	0,84
MoProPEI	2,50	4,00	5,00				0,18	0,40	0,70
P^{3}TQ	1,67	2,00	3,00				0,12	0,20	0,42
CN-SC14									
CRISP-DM	1,50	1,67	2,00	9,00	11,67	15,00	0,10	0,14	0,22
MoProPEI	4,00	5,00	7,00				0,27	0,43	0,78
P^{3}TQ	3,50	5,00	6,00				0,23	0,43	0,67
CN-SC15									
CRISP-DM	1,58	1,83	2,50	7,92	11,33	15,50	0,10	0,16	0,32
MoProPEI	4,00	6,00	8,00				0,26	0,53	1,01
P^{3}TQ	2,33	3,50	5,00				0,15	0,31	0,63

Posteriormente al cálculo del valor de extensión sintética difusa para todos los niveles de la jerarquía, se determinó la superioridad comparativa de los criterios evaluados, tal como se señaló en el paso 6.2 del modelo propuesto y se describe en el paso 4 del Anexo II.

Partiendo desde la cúspide y descendiendo por la estructura jerárquica de la Figura 4.2, en la Tabla 5.150 se expuso la superioridad comparativa del segundo nivel.

Para la rama izquierda de la jerarquía, en el tercer nivel, la superior comparativa del AD y UD bajo el ED, se mostró en la Tabla 5.151, y para los mismos (AD y UD) en el cuarto nivel respecto a sus cinco características en la Tabla 5.152 y Tabla 5.154 respectivamente.
Para CV2, la superioridad comparativa de los cinco sub-criterios del AD, en relación a cada alternativa (CRISP-DM, MoProPEI y P^{3}TQ), se exhibe en la Tabla 5.272 y en la Tabla 5.273 para los de UD.

Tabla 5.272: Superioridad comparativa de Sub-criterios ED-AD para CV2

V(M2>=M1)						
ED-SC1	V(A1>=A2)	0,63	V(A2>=A1)	1,00	V(A3>=A1)	0,53
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,14
ED-SC2	V(A1>=A2)	0,22	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
ED-SC3	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
ED-SC4	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
ED-SC5	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,00	V(A2>=A3)	1,00	V(A3>=A2)	1,00

Tabla 5.273: Superioridad comparativa de Sub-criterios ED-UD para CV2

V(M2>=M1)						
ED-SC6	V(A1>=A2)	0,62	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
ED-SC7	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
ED-SC8	V(A1>=A2)	0,22	V(A2>=A1)	1,00	V(A3>=A1)	0,46
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
ED-SC9	V(A1>=A2)	0,22	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
ED-SC10	V(A1>=A2)	0,62	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00

Para la Comprensión del Negocio y sus sub-criterios, se evaluó la superioridad comparativa en los diferentes niveles. Primero, en la Tabla 5.156 se estimó para sus tres sub-criterios principales (nivel tres); luego en la Tabla 5.157 de los siete sub-criterios del EN (Entorno del Negocio), en la Tabla 5.159 de los cinco del PEN (Proyecto en el Entorno del Negocio) y en la Tabla 5.161 de los tres de DSPN (Data Science en el Proyecto del Negocio).

Para el segundo caso de validación, en las Tablas 5.274, 5.275 y 5.276, se presenta el predominio al comparar cada sub-criterio del nivel cuatro en relación a cada metodología.

Tabla 5.274: Superioridad comparativa de Sub-criterios CN-EN para CV2

V(M2>=M1)						
CN-SC1	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC2	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC3	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC4	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC5	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC6	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC7	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00

Tabla 5.275: Superioridad comparativa de Sub-criterios CN-PEN para CV2

V(M2>=M1)						
CN-SC8	V(A1>=A2)	0,51	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,55	V(A2>=A3)	1,00	V(A3>=A2)	0,63
CN-SC9	V(A1>=A2)	0,51	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,55	V(A2>=A3)	1,00	V(A3>=A2)	1,00
CN-SC10	V(A1>=A2)	0,62	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00

CN-SC11	V(A1>=A2)	0,62	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00
CN-SC12	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	0,00
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,00

Tabla 5.276: Superioridad comparativa de Sub-criterios CN-DSPN para CV2

V(M2>=M1)						
CN-SC13	V(A1>=A2)	1,00	V(A2>=A1)	1,00	V(A3>=A1)	0,51
	V(A1>=A3)	1,00	V(A2>=A3)	1,00	V(A3>=A2)	0,55
CN-SC14	V(A1>=A2)	0,00	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,00	V(A2>=A3)	1,00	V(A3>=A2)	1,00
CN-SC15	V(A1>=A2)	0,14	V(A2>=A1)	1,00	V(A3>=A1)	1,00
	V(A1>=A3)	0,53	V(A2>=A3)	1,00	V(A3>=A2)	0,63

Ulterior al cálculo de la superioridad confrontada de cada criterio y sub-criterio, se separó el valor mínimo (de superioridad); a partir del cual se estimó el vector de pesos y se lo normalizó para cada criterio y sub-criterio (descripto en paso 5 del Anexo II).

Descendiendo desde la cima de la jerarquía propuesta, para el segundo nivel, en la Tabla 5.163, se expuso la ponderación del 50% obtenida para el ED y para la CN (W); cuyo valor mínimo fue 1 (W´). Luego en la Tabla 5.164 se presentó el vector de pesos del ED en relación a AD y UD; los cuales en relación a sus cinco sub-criterios se mostró en las Tablas 5.165 y 5.167.

En cuanto a cada uno de los sub-criterios del AD por un lado y por otro del UD, los vectores de pesos alcanzados en relación a las tres metodologías analizadas para el segundo caso de validación (CV2) se muestran en la Tabla 5.277 para el AD y en Tabla 5.278 para el UD.

Tabla 5.277: Vector de pesos de los Sub-criterios ED-AD para CV2

ED-SC1	$W' = (0{,}63; 1{,}00; 0{,}14)^T$	$\sum W' = 1{,}76$
	$W = (0{,}35; 0{,}57; 0{,}08)^T$	
ED-SC2	$W' = (0{,}22; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}22$
	$W = (0{,}18; 0{,}82; 0{,}00)^T$	

ED-SC3	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
ED-SC4	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
ED-SC5	$W' = (0{,}00; 1{,}00; 1{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}00; 0{,}50; 0{,}50)^T$	

Tabla 5.278: Vector de pesos de los Sub-criterios ED-UD para CV2

ED-SC6	$W' = (0{,}62; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}62$
	$W = (0{,}38; 0{,}62; 0{,}00)^T$	
ED-SC7	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
ED-SC8	$W' = (0{,}22; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}22$
	$W = (0{,}18; 0{,}82; 0{,}00)^T$	
ED-SC9	$W' = (0{,}22; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}22$
	$W = (0{,}18; 0{,}82; 0{,}00)^T$	
ED-SC10	$W' = (0{,}62; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}62$
	$W = (0{,}38; 0{,}62; 0{,}00)^T$	

Al igual que para el ED, para la CN (Comprensión del Negocio) se estimó el vector de pesos al confrontar sus tres sub-criterios principales como se mostró en la Tabla 5.169; luego en la Tabla 5.170 del Entorno del Negocio (CN-EN) y sus siete sub-criterios, en la Tabla 5.172 de los cinco del Proyecto en el Entorno del Negocio (CN-PEN) y en la Tabla 5.174 de Data Science en el Proyecto del Negocio (CN-DSPN) y sus tres sub-criterios.

En lo referido al CV2, los sub-criterios del cuarto nivel verificados con CRISP-DM, MoProPEI y P^3TQ, lograron los vectores indicados en las Tablas 5.279 para CN-EN, en la Tabla 5.280 para CN-PEN y en la Tabla 5.281 para CN-DSPN.

Tabla 5.279: Vector de pesos de los Sub-criterios CN-EN para CV2

CN-SC1	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
CN-SC2	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
CN-SC3	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
CN-SC4	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
CN-SC5	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
CN-SC6	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	
CN-SC7	$W' = (1{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}50; 0{,}50; 0{,}00)^T$	

Tabla 5.280: Vector de pesos de los Sub-criterios CN-PEN para CV2

CN-SC8	$W' = (0{,}51; 1{,}00; 1{,}00)^T$	$\sum W' = 2{,}51$
	$W = (0{,}20; 0{,}40; 0{,}40)^T$	
CN-SC9	$W' = (0{,}51; 1{,}00; 1{,}00)^T$	$\sum W' = 2{,}51$
	$W = (0{,}20; 0{,}40; 0{,}40)^T$	
CN-SC10	$W' = (0{,}62; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}62$
	$W = (0{,}38; 0{,}62; 0{,}00)^T$	

CN-SC11	$W' = (0{,}62; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}62$
	$W = (0{,}38; 0{,}62; 0{,}00)^T$	
CN-SC12	$W' = (0{,}00; 1{,}00; 0{,}00)^T$	$\sum W' = 1{,}00$
	$W = (0{,}00; 1{,}00; 0{,}00)^T$	

Tabla 5.281: Vector de pesos de los Sub-criterios CN-DSPN para CV2

CN-SC13	$W' = (1{,}00; 1{,}00; 0{,}51)^T$	$\sum W' = 2{,}51$
	$W = (0{,}40; 0{,}40; 0{,}20)^T$	
CN-SC14	$W' = (0{,}00; 1{,}00; 1{,}00)^T$	$\sum W' = 2{,}00$
	$W = (0{,}00; 0{,}50; 0{,}50)^T$	
CN-SC15	$W' = (0{,}14; 1{,}00; 0{,}63)^T$	$\sum W' = 1{,}76$
	$W = (0{,}08; 0{,}56; 0{,}36)^T$	

Posterior al cálculo de los vectores de peso, se crearon las prioridades parciales en los diferentes niveles de la jerarquía y de cada alternativa (ver paso 6 del Anexo II).

En la Tabla 5.176, se expuso las ponderaciones estimadas para ED y CN, dado que poseen igual relevancia dentro de un proyecto de DS, a ambos corresponde un 50%.

Para el ED, la ponderación parcial fue al 100% para UD, como se mostró en Tabla 5.177. En la Tabla 5.178 se indicó las ponderaciones obtenidas para los sub-criterios del AD al compararlos entre sí y en la Tabla 5.180 para los del UD.

Tabla 5.282: Ponderaciones parciales del Sub-criterios ED-AD para CV2

	Pesos		
	A1	**A2**	**A3**
ED-SC1	0,36	0,57	0,08
ED-SC2	0,18	0,82	0,00
ED-SC3	0,50	0,50	0,00
ED-SC4	0,50	0,50	0,00
ED-SC5	0,00	0,50	0,50
Peso ED-AD	**0,39**	**0,61**	**0,00**

Cada uno de estos sub-criterios evaluados en función de las metodologías lograron las ponderaciones expuestas en la Tabla 5.282 para el AD y en la Tabla 5.283 para el UD, donde en ambos casos la mejor elección es MoProPEI con 0.61 y 0.57, descartando por completo con este método (FAHP) a P^3TQ en este nivel de la jerarquía.

Las ponderaciones globales para el Entendimiento de los Datos, se pueden ver en la Figura 5.8.

Tabla 5.283: Ponderaciones parciales del Sub-criterios ED-UD para CV2

	Pesos		
	A1	**A2**	**A3**
ED-SC6	0,38	0,62	0,00
ED-SC7	0,50	0,50	0,00
ED-SC8	0,18	0,82	0,00
ED-SC9	0,18	0,82	0,00
ED-SC10	0,38	0,62	0,00
Peso ED-UD	**0,43**	**0,57**	**0,00**

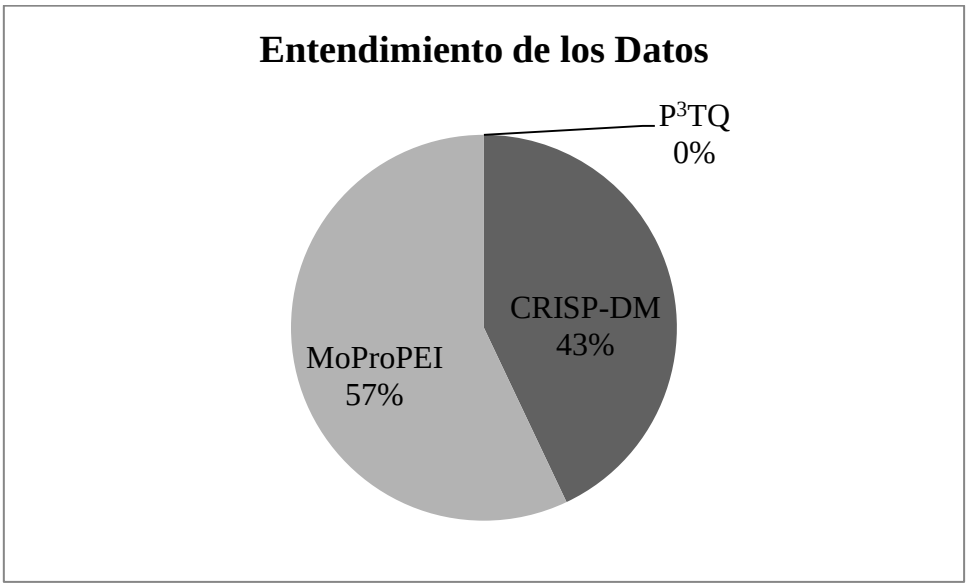

Figura 5.8: Ponderaciones globales del ED para el CV2 con FAHP

Para la Comprensión del Negocio, en la Tabla 5.182, se expusieron las ponderaciones parciales. Para sus tres sub-criterios (Entorno del Negocio (CN-EN), Proyecto en el Entorno

del Negocio (CN-PEN) y DS en el Proyecto del Negocio (CN-DSPN)) en las Tablas 5.183, 5.185 y 5.187 se enseñaron las ponderaciones parciales para cada uno de ellos. Posteriormente, en las Tablas 5.284, 5.285 y 5.286 se visualizan las ponderaciones parciales de los sub-criterios involucrados en CN-EN, CN-PEN y CN-DSPN para el segundo caso de validación (CV2); en las mismas se puede apreciar que para el EN y DSPN, las metodologías CRISP-DM y MoProPEI lograron los mismos porcentajes de adecuación, 50% y 40% respectivamente; y para PEN la más adecuada es MoProPEI con una ponderación parcial de 0,66.

Tabla 5.284: Ponderaciones parciales del Sub-criterios CN-EN para CV2

	Pesos		
	A1	**A2**	**A3**
CN-SC1	0,50	0,50	0,00
CN-SC2	0,50	0,50	0,00
CN-SC3	0,50	0,50	0,00
CN-SC4	0,50	0,50	0,00
CN-SC5	0,50	0,50	0,00
CN-SC6	0,50	0,50	0,00
CN-SC7	0,50	0,50	0,00
Peso CN-EN	**0,50**	**0,50**	**0,00**

Tabla 5.285: Ponderaciones parciales del Sub-criterios CN-PEN para CV2

	Pesos		
	A1	**A2**	**A3**
CN-SC8	0,20	0,40	0,40
CN-SC9	0,20	0,40	0,40
CN-SC10	0,38	0,62	0,00
CN-SC11	0,38	0,62	0,00
CN-SC12	0,00	1,00	0,00
Peso CN-PEN	**0,28**	**0,66**	**0,06**

Tabla 5.286: Ponderaciones parciales del Sub-criterios CN-DSPN para CV2

	Pesos		
	A1	**A2**	**A3**
CN-SC13	0,40	0,40	0,20
CN-SC14	0,00	0,50	0,50
CN-SC15	0,08	0,57	0,36
Peso CN-DSPN	**0,40**	**0,40**	**0,20**

A partir de las ponderaciones parciales para los sub-criterios debajo a CN (haciendo alusión a la Figura 4.2) se calculó las globales para cada metodología, expuestas en la Figura 5.9, donde se puede ver que MoProPEI y CRISP-DM lograron un 41% de adecuación para el CV2 y PT^{3}Q un 18%.

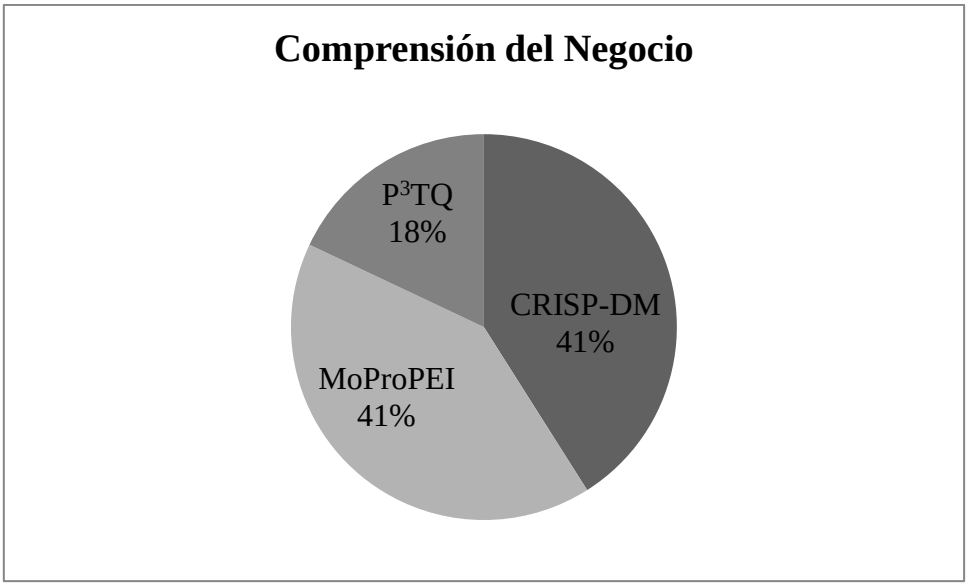

Figura 5.9: Ponderaciones globales de la CN para el CV2 con FAHP

Tabla 5.287: Prioridades finales de las alternativas con FAHP para CV2

	CRISP-DM (A1)	**MoProPEI (A2)**	**P^{3}TQ (A3)**
ED	0,43	0,57	0,00
CN	0,41	0,41	0,18
Prioridades Finales	**0,42**	**0,49**	**0,09**

Por tanto, las prioridades finales alcanzadas por cada alternativa son las que se exhiben en la Tabla 5.287; donde para el segundo caso de validación (CV2) vinculado a los causales de deserción universitaria, la metodología MoProPEI es la más adecuada con el 49% de preferencia, seguida por CRISP-DM con el 42%.

5.3.5. Integración de las ponderaciones obtenidas por los métodos AHP y FAHP para el CV2

Para la integración de las ponderaciones finales obtenidas por los métodos AHP y FAHP para el segundo caso de validación (CV2), se recurrió al promedio de las ponderaciones finales alcanzadas por CRISP-DM (A1), MoProPEI (A2) y P^3TQ (A3) para cada uno de los métodos, dichos valores se exhiben en la Tabla 5.288.

Tabla 5.288: Prioridades finales de las alternativas del CV2

Alternativas	AHP	FAHP	Prioridades Finales
CRISP-DM (A1)	0,34	0,42	**0,38**
MoProPEI (A2)	0,49	0,49	**0,49**
P^3TQ (A3)	0,17	0,09	**0,13**

5.3.6. Ranking de metodologías de DS para el CV2

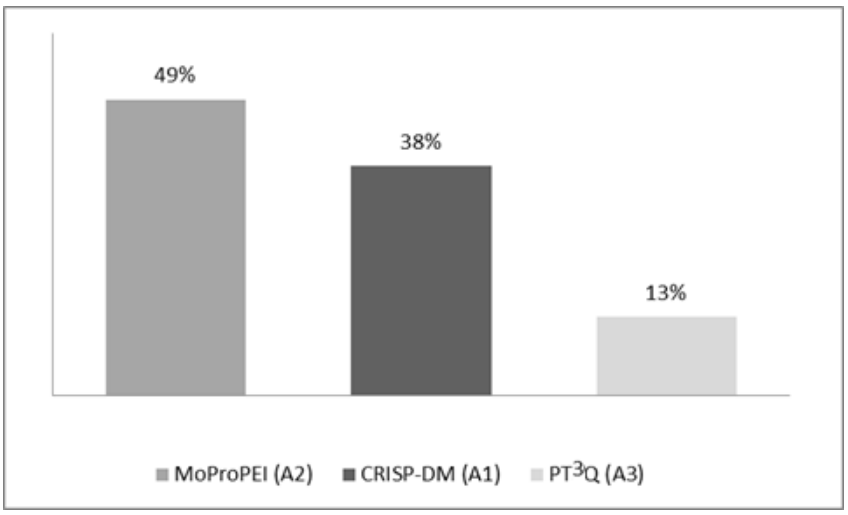

Figura 5.10: Ranking de prioridades finales para el CV2

Para finalizar, se generó el ranking de prioridades, ordenando las ponderaciones obtenidas por cada metodología evaluada en el CV2; para el cual, en primer lugar se ubica MoProPEI (A2) con el 49%, seguido de CRISP-DM (A1) con el 38% y en el último puesto P^{3}TQ (A3) con el 13% de adecuación, tal como se visualiza en la Figura 5.10.

5.4. Análisis de resultados

Como se puedo demostrar en las secciones anteriores, a partir del modelo basado en MCDM propuesto en el Capítulo IV, que brinda soporte para la TD en la elección de metodologías de DS, se han podido desarrollar con éxito los dos casos de validación planteados para este TFM; a partir de los cuales se pudo identificar a MoProPEI como la más adecuada y robusta según los criterios evaluados, tanto para determinar los causales de roturas de automóviles 0KM mientras son transportados desde la fábrica hasta las concesionarias (CV1), como los vinculados a la deserción universitaria (CV2), en segundo lugar se ubicó la metodología CRISP-DM y en última instancia P^{3}TQ en ambos escenarios. Los porcentajes obtenidos para cada caso de validación se resumen en la Figura 5.11. Se puede notar que para el primer caso (CV1), la distancia en porcentajes de adecuación entre una metodología y otra, ha sido más acentuada que para el segundo (CV2).

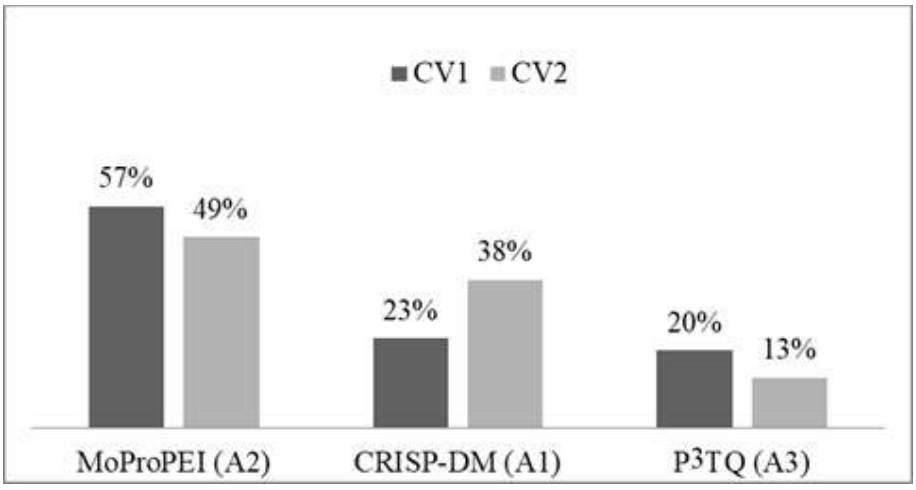

Figura 5.11: Integración de prioridades finales de los casos de validación

Respecto a los resultados obtenidos por los métodos AHP y FAHP, se puedo notar que FAHP tiende a acrecentar las distancias entre las metodologías; dado que discrimina aquellos criterios o sub-criterios considerados menos relevantes, debido al cálculo de superioridad

comparativa y la elección del valor mínimo, por ello que la metodología P^{3}TQ alcanzó porcentajes inferiores para FAHP que AHP, como se muestra en la Figura 5.12, donde se compendian los casos de validación analizados.

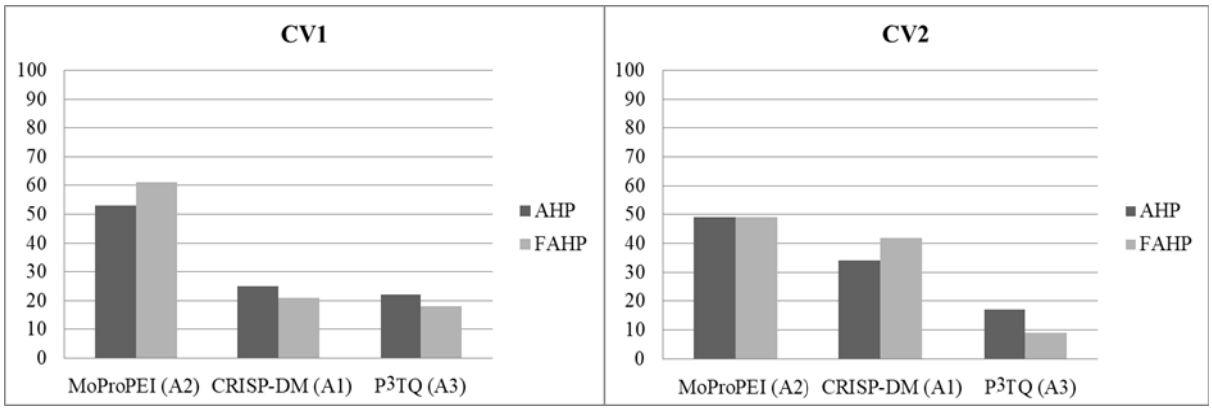

Figura 5.12: Prioridades con AHP y FAHP para CV1 y CV2

Los resultados obtenidos para ambos casos de validación son confiables, ya que han sido corroborados por los expertos, quién inicialmente ha aprobado la estructura jerárquica propuesta (Figura 4.2), así como las etiquetas lingüísticas definidas, también indicó los valores de cumplimiento y preferencia de cada criterio o sub-criterio y finalmente respaldó los resultados obtenidos en cada caso.

VI

Conclusiones y trabajos futuros

Finalmente, en este capítulo se exponen las principales conclusiones arribadas del TFM en la sección 6.1; luego en la sección 6.2 se mencionan las publicaciones realizadas a partir de la misma y las posibles líneas de investigación futuras se describen en la sección 6.3.

6.1. Conclusiones

A partir de la labor realizada en este TFM, se logró fundar un modelo basado en la MCDM con el propósito de comparar metodologías de DS y seleccionar la más robusta, de entre las comparadas, para proyectos de dicha índole; para lo cual, el modelo propuesto integra los métodos AHP y FAHP, así como también a la técnica de TCP; lo cual brinda una base sólida para tomar una decisión respecto a cuál metodología se adecua a cada proyecto de DS.

Luego del trabajo realizado en estado de la cuestión, mediante una revisión exhaustiva, al estudiar y analizar métodos de MCDM existentes, haciendo énfasis en los métodos AHP y FAHP, al revisar el estado actual de las metodologías de DS y de las técnicas de educción de conocimiento como TPC; se pudo apreciar que la integración de estos métodos y técnicas no ha sido llevada a cabo anteriormente, ni evaluadas en el área seleccionado (DS), siendo estos los principales desafíos y contribuciones del TFM.

Mediante la técnica de TCP en colaboración con los expertos, se han definido los criterios y sub-criterios esenciales a tener en cuenta al momento de inicializar un proyecto de DS, a partir de los cuales, se ha diseñado una estructura jerárquica precisa en niveles clasificatorios para los mismos.

A través del trabajo realizado con TPC y expertos, se fijaron las etiquetas lingüísticas para las características (sub-criterios) más específicas dentro de la estructura jerárquica, estas etiquetas se adecuaron según el porcentaje de cumplimiento de cada metodología a una escala de uno a nueve (1 a 9), para poder luego realizar la transición a los métodos de MCDM seleccionados, que se basan precisamente en la escala fundamental propuesta por Saaty y la de conversión difusa. Esta transmutación e integración, no había sido prevista como tal, por lo que implicó un reto; donde primeramente se buscó fusionar el desarrollo completo de TCP a los métodos de MCDM, con lo cual se tuvo ciertas dificultades, tales como ajustar en etapas avanzadas de TCP a la escala fundamental, así como establecer razones de consistencia por debajo a las deseadas (10%); por lo que luego de varias pruebas se definió utilizar las tres primeras etapas de TCP, donde en primer lugar se identificaron los elementos a evaluar, que son precisamente las metodologías de DS, en segundo las características o criterios, para luego construir la

jerarquía del problema y en tercer lugar, se diseñaron las parrillas, que consistieron en una matriz bidimensional en la cual se dispusieron los valores bipolares de las características; según la escala formulada con etiquetas lingüísticas para valorar los criterios (disponible en el Anexo IV). Otro aporte de este trabajo fue la creación de matrices de dos entradas que faciliten la comparación entre criterios, dichas matrices se encuentran en el Anexo III.

A partir de las matrices obtenidas con las parrillas de TCP y las matrices de dos entradas se han desarrollado los métodos AHP y FAHP, si bien el segundo deriva del primero, implicó un desafío la combinación de los mismos, por lo que se definieron los pasos involucrados en el modelo, donde se comenzó con el método AHP hasta comprobar que los valores emitidos por el experto son consistentes, a partir de allí se dividió en actividades paralelas, por un lado se finalizó el desarrollo de AHP hasta obtener las ponderaciones finales para cada metodología en cada caso de validación y por el otro lado, se inició con el desarrollo de FAHP, tomando como valor más prometedor (m) al existente en las matrices de comparación pareada de AHP y completando los demás valores utilizando una función de membrecía triangular difusa y completando los demás pasos de FAHP hasta obtener las ponderaciones finales. Las ponderaciones obtenidas por ambos métodos fueron ordenadas de mayor a menor conformando un ranking y finalmente se calculó el valor promedio de cada metodología, como valor de consenso, lo que permitió indicar cuál es la metodología seleccionada para cada proyecto de DS.

En relación a los casos de validación utilizados para corroborar la efectividad del modelo propuesto, fueron dos casos reales con características disímiles, por un lado relacionado a los causales de roturas de automóviles 0KM durante su traslado a las concesionarias (CV1) y por otro los vinculados a la deserción universitaria en una carrera de grado (CV2), lo cual enriqueció los resultados obtenidos y validó su utilización en diferentes contextos de aplicación. Los resultados arribados fueron evaluados y corroborados por los expertos. Para ambos casos, MoProPEI fue la metodología que mayor grado de cumplimiento y preferencia tuvo en los dos casos estudiados.

Por lo expuesto anteriormente se demostró que el modelo es efectivo para comparar metodologías de DS, inclusive puede ser aplicado a otras áreas.

6.2. Publicaciones

A partir del TFM se confeccionaron artículos enviados a eventos nacionales e internacionales. El Anexo V muestra los correos de aceptación de los artículos en prensa.

- "Modelo basado en la toma decisiones con criterios múltiples para la elección de metodologías de data science", presentado en XX Workshop de Investigadores en Ciencias de la Computación (WICC), Workshop Bases de Datos y Minería de Datos, Corrientes Argentina, abril de 2018, ISBN 978-987-3619-27-4.
- "Model for comparison and selection of Data Science Methodologies through Multiple Criteria Decision Making", enviado al 31st International Conference on Software Engineering & Knowledge Engineering (SEKE), Lisbon Portugal, julio de 2019.
- "Análisis comparativo de metodologías de ciencia de datos utilizando el proceso analítico jerárquico difuso junto a la teoría de la construcción personal", 48 Jornadas Argentinas de Informática (JAIIO). Simposio Argentino de Ciencia de Datos y GRANdes Datos, Salta Argentina, septiembre de 2019.
- "Data science methodologies selection with hierarchical analytical process and personal construction theory", XXV Congreso Argentino de Ciencias de la Computación (CACIC), XVI Workshop Bases de Datos y Minería de Datos (WBDDM), Córdoba Argentina, octubre de 2019.

6.3. Líneas futuras de investigación

Como línea futura de investigación en relación a las aplicaciones del modelo propuesto, se propone el desarrollo de un software que materialice el modelo propuesto, facilitando mediante interfaces gráficas de usuario la completitud de las valoraciones, guiando el desarrollo del mismo, realizando los cálculos necesarios, informando y ajustando los valores en caso de haber inconsistencias en los juicios del experto, obteniendo las ponderaciones en los diferentes niveles de la jerarquía y organizando los porcentajes de adecuación de cada metodología para la correcta TD; para lo cual se podrían utilizar casos de validación adicionales a los utilizados en este TFM.

Asimismo se considera interesante utilizar el modelo propuesto para comparar MoProPEI con otras metodologías de DS, como las mencionadas en la sección 2.2. También se podría ratificar su utilidad en otras áreas de interés, como por ejemplo para la comparación y

elección de metodologías de desarrollo de software, incluso para otros tipos de confrontaciones como método de MCDM; dado que se considera al modelo propuesto lo suficientemente claro y adaptable a otros campos de aplicación.

Finalmente un problema abierto de interés consiste en el estudio de los procesos de TD grupales con análisis de sensibilidad, no abordado en el presente TFM.

Bibliografía

[1] M. del S. García Cascales, «Métodos para la comparación de alternativas mediante un Sistema de Ayuda a la Decisión (S.A.D.) y "Soft Computing"», Tesis de Doctorado, Universidad Politécnica de Cartagena - Departamento de Electrónica, Tecnología de Computadoras y Proyectos, Cartagena, Colombia, 2009.
[2] G. E. Dieter, *Engineering Design: A Materials and Processing Approach*, 1°. Tokyo, Japón: McGraw Hill Higher Education, 1983.
[3] O. G. León, *Tomar decisiones difíciles*, 2°. Madrid, España: McGraw-Hill interamericana, 2001.
[4] F. T. Dweiri y M. M. Kablan, «Using fuzzy decision making for the evaluation of the project management internal efficiency», *Decision Support Systems*, vol. 42, n.º 2, pp. 712-726, nov. 2006.
[5] M. Karanik, S. Gramajo, L. Wanderer, M. Giménez, y D. Carpintero, «Multi-Criteria Decision Model based on AHP and Linguistic Information», *Journal of Computer Science & Technology*, vol. 14, n.º 1, pp. 16-24, abr. 2014.
[6] J. C. Osorio Gómez y J. P. Orejuela Cabrera, «El proceso de análisis jerárquico (AHP) y la toma de decisiones multicriterio. Ejemplo de aplicación.», *Scientia et technica*, vol. 2, n.º 39, ago. 2008.
[7] C. Romero, *Análisis de las Decisiones Multicriterio*, 1°. Madrid, España: Isdefe, 1996.
[8] A. Dadda y I. Ouhbi, «A decision support system for renewable energy plant projects», presentado en 2014 International Conference on Next Generation Networks and Services (NGNS), Casablanca, Morocco, 2014, pp. 356-362.
[9] E. Triantaphyllou y S. H. Mann, «Using the analytic hierarchy process for decision making in engineering applications: Some challenges», *International Journal of Industrial Engineering: Applications and Practice*, vol. 2, n.º 1, pp. 35-44, ene. 1995.
[10] J. R. S. C. Mateo, *Multi-Criteria Analysis in the Renewable Energy Industry*, 1°. Londres, Inglaterra: Springer-Verlag London, 2012.
[11] C. Kahraman y İ. Kaya, «A fuzzy multicriteria methodology for selection among energy alternatives», *Expert Systems with Applications*, vol. 37, n.º 9, pp. 6270-6281, sep. 2010.
[12] A. Kolios, V. Mytilinou, E. Lozano-Minguez, y K. Salonitis, «A Comparative Study of Multiple-Criteria Decision-Making Methods under Stochastic Inputs», *Energies*, vol. 9, n.º 7, p. 566, jul. 2016.
[13] B. Roy y R. Słowiński, «Questions guiding the choice of a multicriteria decision aiding method», *EURO J Decis Process*, vol. 1, n.º 1, pp. 69-97, jun. 2013.
[14] P. J. Herrera Caro, «Toma de decisiones en simulación basadas en la teoría de la decisión multicriterio», Tesis de Maestría, Universidad Complutense - Facultad de Informática - Departamento de Ingeniería del Software e Inteligencia Artificial, Madrid, España, 2008.
[15] T. L. Saaty, «How to make a decision: The analytic hierarchy process», *European Journal of Operational Research*, vol. 48, n.º 1, pp. 9-26, sep. 1990.
[16] T. L. Saaty, «Analytic Hierarchy Process», en *Encyclopedia of Operations Research and Management Science*, S. I. Gass y M. C. Fu, Eds. Boston, MA: Springer US, 2013, pp. 52-64.
[17] R. L. Keeney y H. Raiffa, *Decisions with Multiple Objectives: Preferences and Value Trade-Offs*. Cambridge University Press, 1993.
[18] C.-L. Hwang y K. Yoon, «Methods for Multiple Attribute Decision Making», en *Multiple Attribute Decision Making: Methods and Applications A State-of-the-Art Survey*, C.-L. Hwang y K. Yoon, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1981, pp. 58-191.

[19] E. Jacquet Lagreze y J. Siskos, «Assessing a set of additive utility functions for multicriteria decision-making, the UTA method», *European Journal of Operational Research*, vol. 10, n.º 2, pp. 151-164, jun. 1982.
[20] D. von Winterfeld y W. Edwards, *Decision analysis and behavioral research*, 1°. New York, USA: Cambridge University Press, 1986.
[21] J. S. Dyer, «MAUT — Multiattribute Utility Theory», en *Multiple Criteria Decision Analysis: State of the Art Surveys*, J. Figueira, S. Greco, y M. Ehrogott, Eds. New York, NY: Springer New York, 2005, pp. 265-292.
[22] C. A. B. E. Costa y J.-C. Vansnick, «MACBETH - An Interactive Path Towards the Construction of Cardinal Value Functions», *International Transactions in Operational Research*, vol. 1, n.º 4, pp. 489-500, oct. 1994.
[23] C. A. B. E. Costa, J.-M. de Corte, y J.-C. Vansnick, «MACBETH (Measuring Attractiveness by a Categorical Based Evaluation Technique)», en *Wiley Encyclopedia of Operations Research and Management Science*, American Cancer Society, 2011.
[24] C. Labreuche y M. Grabisch, «The Choquet integral for the aggregation of interval scales in multicriteria decision making», *Fuzzy Sets and Systems*, vol. 137, n.º 1, pp. 11-26, jul. 2003.
[25] S. Greco, M. Kadziński, y R. SŁowiński, «Selection of a representative value function in robust multiple criteria sorting», *Computers & Operations Research*, vol. 38, n.º 11, pp. 1620-1637, nov. 2011.
[26] J. J. Buckley, «Fuzzy hierarchical analysis», *Fuzzy Sets and Systems*, vol. 17, n.º 3, pp. 233-247, dic. 1985.
[27] Y. M. Wang, Y. Luo, y Z. Hua, «On the extent analysis method for fuzzy AHP and its applications», *European Journal of Operational Research*, vol. 186, n.º 2, pp. 735-747, abr. 2008.
[28] D.-Y. Chang, «Applications of the extent analysis method on fuzzy AHP», *European Journal of Operational Research*, vol. 95, n.º 3, pp. 649-655, dic. 1996.
[29] T. C. Wang y H. D. Lee, «Developing a fuzzy TOPSIS approach based on subjective weights and objective weights», *Expert Systems with Applications*, vol. 36, n.º 5, pp. 8980-8985, jul. 2009.
[30] J. Figueira, V. Mousseau, y B. Roy, «Electre Methods», en *Multiple Criteria Decision Analysis: State of the Art Surveys*, J. Figueira, S. Greco, y M. Ehrogott, Eds. New York, NY: Springer New York, 2005, pp. 133-153.
[31] J. R. Figueira, S. Greco, B. Roy, y R. Słowiński, «An Overview of ELECTRE Methods and their Recent Extensions», *Journal of Multi-Criteria Decision Analysis*, vol. 20, n.º 1-2, pp. 61-85, ene. 2013.
[32] J.-P. Brans y B. Mareschal, «Promethee Methods», en *Multiple Criteria Decision Analysis: State of the Art Surveys*, J. Figueira, S. Greco, y M. Ehrogott, Eds. New York, NY: Springer New York, 2005, pp. 163-186.
[33] S. Greco, R. Słowiński, J. R. Figueira, y V. Mousseau, «Robust Ordinal Regression», en *Trends in Multiple Criteria Decision Analysis*, M. Ehrgott, J. R. Figueira, y S. Greco, Eds. Boston, MA: Springer US, 2010, pp. 241-283.
[34] J. R. Figueira, S. Greco, y R. Słowiński, «Building a set of additive value functions representing a reference preorder and intensities of preference: GRIP method», *European Journal of Operational Research*, vol. 195, n.º 2, pp. 460-486, jun. 2009.
[35] M. Kadziński, S. Greco, y R. SŁowiński, «Extreme ranking analysis in robust ordinal regression», *Omega*, vol. 40, n.º 4, pp. 488-501, ago. 2012.

[36] M. Kadziński, S. Greco, y R. Słowiński, «RUTA: A framework for assessing and selecting additive value functions on the basis of rank related requirements», *Omega*, vol. 41, n.º 4, pp. 735-751, ago. 2013.
[37] S. Greco, M. Kadziński, V. Mousseau, y R. Słowiński, «ELECTREGKMS: Robust ordinal regression for outranking methods», *European Journal of Operational Research*, vol. 214, n.º 1, pp. 118-135, oct. 2011.
[38] S. Greco, B. Matarazzo, y R. Slowinski, «Rule-Based Decision Support in Multicriteria Choice and Ranking», en *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, 2001, pp. 29-47.
[39] K. Dembczyński, W. Kotłowski, R. Słowiński, y M. Szeląg, «Learning of Rule Ensembles for Multiple Attribute Ranking Problems», en *Preference Learning*, J. Fürnkranz y E. Hüllermeier, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 217-247.
[40] R. Bisdorff, P. Meyer, y M. Roubens, «RUBIS: a bipolar-valued outranking method for the choice problem», *4OR*, vol. 6, n.º 2, pp. 143-165, jun. 2008.
[41] S. Greco, B. Matarazzo, y R. Słowiński, «Dominance-Based Rough Set Approach to Interactive Multiobjective Optimization», en *Multiobjective Optimization: Interactive and Evolutionary Approaches*, J. Branke, K. Deb, K. Miettinen, y R. Słowiński, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 121-155.
[42] C. Zopounidis y M. Doumpos, «PREFDIS: a multicriteria decision support system for sorting decision problems», *Computers & Operations Research*, vol. 27, n.º 7, pp. 779-797, jun. 2000.
[43] J. Almeida Dias, J. R. Figueira, y B. Roy, «Electre Tri-C: A multiple criteria sorting method based on characteristic reference actions», *European Journal of Operational Research*, vol. 204, n.º 3, pp. 565-580, ago. 2010.
[44] J. Almeida Dias, J. R. Figueira, y B. Roy, «A multiple criteria sorting method where each category is characterized by several reference actions: The Electre Tri-nC method», *European Journal of Operational Research*, vol. 217, n.º 3, pp. 567-579, mar. 2012.
[45] J. Léger y J.-M. Martel, «A multicriteria assignment procedure for a nominal sorting problematic», *European Journal of Operational Research*, vol. 138, n.º 2, pp. 349-364, abr. 2002.
[46] N. Belacel, «Multicriteria assignment method PROAFTN: Methodology and medical application», *European Journal of Operational Research*, vol. 125, n.º 1, pp. 175-183, ago. 2000.
[47] M. Doumpos y C. Zopounidis, «A multicriteria classification approach based on pairwise comparisons», *European Journal of Operational Research*, vol. 158, n.º 2, pp. 378-389, oct. 2004.
[48] E. Fernandez y J. Navarro, «A new approach to multi-criteria sorting based on fuzzy outranking relations: The THESEUS method», *European Journal of Operational Research*, vol. 213, n.º 2, pp. 405-413, sep. 2011.
[49] A. P. Wierzbicki, «On the completeness and constructiveness of parametric characterizations to vector optimization problems», *OR Spektrum*, vol. 8, n.º 2, pp. 73-87, jun. 1986.
[50] K. Deb, «Introduction to Evolutionary Multiobjective Optimization», en *Multiobjective Optimization: Interactive and Evolutionary Approaches*, J. Branke, K. Deb, K. Miettinen, y R. Słowiński, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 59-96.
[51] R. W. Saaty, «The analytic hierarchy process—what it is and how it is used», *Mathematical Modelling*, vol. 9, n.º 3, pp. 161-176, ene. 1987.

[52] R. de F. S. M. Russo y R. Camanho, «Criteria in AHP: A Systematic Review of Literature», *Procedia Computer Science*, vol. 55, pp. 1123-1132, ene. 2015.
[53] Y. Peng, G. Kou, G. Wang, W. Wu, y Y. Shi, «Ensemble of software defect predictors: an ahp-based evaluation method», *Int. J. Info. Tech. Dec. Mak.*, vol. 10, n.º 01, pp. 187-206, ene. 2011.
[54] J. M. Moreno Jiménez, «El Proceso Analítico Jerárquico (AHP). Fundamentos, Metodología y Aplicaciones». 09-oct-2014.
[55] T. L. Saaty, *The analytic hierarchy process*. New York: McGraw-Hill, 1980.
[56] P. T. Harker, «The Art and Science of Decision Making: The Analytic Hierarchy Process», en *The Analytic Hierarchy Process: Applications and Studies*, B. L. Golden, E. A. Wasil, y P. T. Harker, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 1989, pp. 3-36.
[57] L. Vera Montenegro, «Aplicación y Comparación de Metodologías Multicriterio (AHP y Fuzzy Logic) en la Selección de Tecnologías Postcosecha para Pequeños Productores de Cacao», Tesis de Doctorado, Universidad Politécnica de Valencia, Valencia, España, 2014.
[58] G. Kou y C. Lin, «A cosine maximization method for the priority vector derivation in AHP», *European Journal of Operational Research*, vol. 235, n.º 1, pp. 225-232, may 2014.
[59] E. H. Forman y S. I. Gass, «The Analytic Hierarchy Process—An Exposition», *Operations Research*, vol. 49, n.º 4, pp. 469-486, ago. 2001.
[60] T. L. Saaty, «Decision making with the analytic hierarchy process», *International Journal of Services Sciences*, vol. 1, n.º 1, pp. 83-98, ene. 2008.
[61] T. L. Saaty, *Fundamentals of Decision Making and Priority Theory With the Analytic Hierarchy Process*. RWS Publications, 2000.
[62] O. S. Vaidya y S. Kumar, «Analytic hierarchy process: An overview of applications», *European Journal of Operational Research*, vol. 169, n.º 1, pp. 1-29, feb. 2006.
[63] J. Mayor, S. Botero, y J. D. González-Ruiz, «Modelo de decisión multicriterio difuso para la selección de contratistas en proyectos de infraestructura: caso Colombia», *Obras y proyectos*, n.º 20, pp. 56-74, dic. 2016.
[64] J. Wang, C. Li, y S. Zhang, «Big Data Impact Analysis of Smart Grid based on AHP method», en *International Conference on Manufacturing Science and Engineering (ICMSE 2015), Atlantis Press*, 2015, pp. 1487–1490.
[65] M. Lněnička, «AHP Model for the Big Data Analytics Platform Selection», *Acta Informatica Pragensia*, vol. 4, n.º 2, pp. 108-121, 2015.
[66] E. M. Rodríguez, «Aplicación del proceso jerárquico de análisis en la selección de la lozalización de una PYME», *Anuario jurídico y económico escurialense*, n.º 40, pp. 523–542, 2007.
[67] G. C. Rótolo, M. Milo Vaccaro, M. Hoyos Mallqui, S. Bacigaluppo, F. Salvagiotti, y J. Castellarín, «Ponderación de la información generada en la Estación Experimental Agropecuaria Oliveros del INTA (INTA EEA Oliveros) mediante el proceso analítico jerárquico», *RIA. Revista de investigaciones agropecuarias*, vol. 42, n.º 2, pp. 138-143, ago. 2016.
[68] D. A. Mendoza Casseres, «Aplicación de la teoría de decisión multicriterio discreta para ponderar factores en procesos de acreditación de alta calidad», *INGE CUC*, vol. 9, n.º 1, pp. 25-41, 2013.
[69] C. E. Riaño-Luna y M. L. Palomino-Leiva, «Proceso analítico jerárquico para evaluar tres laboratorios virtuales en la educación superior», *Entramado*, vol. 11, n.º 1, pp. 194–204, 2017.

[70] A. Bascetin, «Technical note: An application of the analytic hierarchy process in equipment selection at Orhaneli open pit coal mine», *Mining Technology*, vol. 113, n.º 3, pp. 192-199, sep. 2004.
[71] A. M. A. Bahurmoz, «The Analytic Hierarchy Process at Dar Al-Hekma, Saudi Arabia», *Interfaces*, vol. 33, n.º 4, pp. 70-78, ago. 2003.
[72] E. Gutiérrez López, J. M. Preciado Rodríguez, y J. M. Robles Parra, «Modelo de toma de decisiones para la construcción sustentable de obra pública», *Estudios sociales (Hermosillo, Son.)*, vol. 28, n.º 51, pp. 1-22, jun. 2018.
[73] Y. Zhang y S. Liu, «Analysis of Data Mining Modes of Gymnastics Training Based on AHP Method», *DEStech Transactions on Computer Science and Engineering*, vol. 0, n.º iccae, 2016.
[74] G. Islei y A. G. Lockett, «Judgemental modelling based on geometric least square», *European Journal of Operational Research*, vol. 36, n.º 1, pp. 27-35, jul. 1988.
[75] T. L. Saaty, «A scaling method for priorities in hierarchical structures», *Journal of Mathematical Psychology*, vol. 15, n.º 3, pp. 234-281, jun. 1977.
[76] A. T. W. Chu, R. E. Kalaba, y K. Spingarn, «A comparison of two methods for determining the weights of belonging to fuzzy sets», *J Optim Theory Appl*, vol. 27, n.º 4, pp. 531-538, abr. 1979.
[77] T. L. Saaty y L. G. Vargas, «Comparison of eigenvalue, logarithmic least squares and least squares methods in estimating ratios», *Mathematical Modelling*, vol. 5, n.º 5, pp. 309-324, ene. 1984.
[78] K. O. Cogger y P. L. Yu, «Eigenweight vectors and least-distance approximation for revealed preference in pairwise weight ratios», *J Optim Theory Appl*, vol. 46, n.º 4, pp. 483-491, ago. 1985.
[79] G. Crawford y C. Williams, «A note on the analysis of subjective judgment matrices», *Journal of Mathematical Psychology*, vol. 29, n.º 4, pp. 387-405, dic. 1985.
[80] N. Bryson, «A Goal Programming Method for Generating Priority Vectors», *Journal of the Operational Research Society*, vol. 46, n.º 5, pp. 641-648, may 1995.
[81] J. Barzilai, «Deriving weights from pairwise comparison matrices», *Journal of the Operational Research Society*, vol. 48, n.º 12, pp. 1226-1232, dic. 1997.
[82] N. (Kweku-M. Bryson y A. Joseph, «Generating consensus priority point vectors: a logarithmic goal programming approach», *Computers & Operations Research*, vol. 26, n.º 6, pp. 637-643, may 1999.
[83] L. Mikhailov y M. G. Singh, «Comparison analysis of methods for deriving priorities in the analytic hierarchy process», en *IEEE SMC'99 Conference Proceedings. 1999 IEEE International Conference on Systems, Man, and Cybernetics (Cat. No.99CH37028)*, 1999, vol. 1, pp. 1037-1042 vol.1.
[84] S. I. Gass y T. Rapcsák, «Singular value decomposition in AHP», *European Journal of Operational Research*, vol. 154, n.º 3, pp. 573-584, may 2004.
[85] K. Sugihara, H. Ishii, y H. Tanaka, «Interval priorities in AHP by interval regression analysis», *European Journal of Operational Research*, vol. 158, n.º 3, pp. 745-754, nov. 2004.
[86] B. Srdjevic, «Combining different prioritization methods in the analytic hierarchy process synthesis», *Computers & Operations Research*, vol. 32, n.º 7, pp. 1897-1919, jul. 2005.
[87] B. Chandran, B. Golden, y E. Wasil, «Linear programming models for estimating weights in the analytic hierarchy process», *Computers & Operations Research*, vol. 32, n.º 9, pp. 2235-2254, sep. 2005.

[88] R. Ramanathan, «Data envelopment analysis for weight derivation and aggregation in the analytic hierarchy process», *Computers & Operations Research*, vol. 33, n.º 5, pp. 1289-1307, may 2006.
[89] Y. M. Wang, C. Parkan, y Y. Luo, «Priority estimation in the AHP through maximization of correlation coefficient», *Applied Mathematical Modelling*, vol. 31, n.º 12, pp. 2711-2718, dic. 2007.
[90] A. Altuzarra, J. M. Moreno-Jiménez, y M. Salvador, «A Bayesian priorization procedure for AHP-group decision making», *European Journal of Operational Research*, vol. 182, n.º 1, pp. 367-382, oct. 2007.
[91] S. Lipovetsky y W. M. Conklin, «Robust estimation of priorities in the AHP», *European Journal of Operational Research*, vol. 137, n.º 1, pp. 110-122, feb. 2002.
[92] B. Srdjevic y Z. Srdjevic, «Bi-criteria evolution strategy in estimating weights from the AHP ratio-scale matrices», *Applied Mathematics and Computation*, vol. 218, n.º 4, pp. 1254-1266, oct. 2011.
[93] C. Lin, G. Kou, y D. Ergu, «A heuristic approach for deriving the priority vector in AHP», *Applied Mathematical Modelling*, vol. 37, n.º 8, pp. 5828-5836, abr. 2013.
[94] B. Golany y M. Kress, «A multicriteria evaluation of methods for obtaining weights from ratio-scale matrices», *European Journal of Operational Research*, vol. 69, n.º 2, pp. 210-220, sep. 1993.
[95] M. T. Escobar Urmeneta y J. M. Moreno Jiménez, «Problemas de gran tamaño en el proceso analítico jerárquico», *Estudios de economía aplicada*, n.º 8, pp. 25-40, sep. 1997.
[96] G. A. Miller, «The magical number seven, plus or minus two: some limits on our capacity for processing information», *Psychological Review*, vol. 63, n.º 2, pp. 81-97, 1956.
[97] T. L. Saaty y M. S. Ozdemir, «Why the magic number seven plus or minus two», *Mathematical and Computer Modelling*, vol. 38, n.º 3, pp. 233-244, ago. 2003.
[98] W. C. Wedley, E. U. Choo, y B. Schoner, «Benchmark Measurement : Between Relative and Absolute», presentado en Fourth International Symposium on the Analytic Hierarchy Process, Vancouver, Canadá, 1996, pp. 335-345.
[99] J. M. Moreno Jiménez y M. T. Escobar Urmeneta, «El pesar en el proceso analítico jerárquico», *Estudios de Economía Aplicada*, vol. 14, pp. 95-115, 2000.
[100] R. E. Jensen y T. E. Hicks, «Ordinal data AHP analysis: A proposed coefficient of consistency and a nonparametric test», *Mathematical and Computer Modelling*, vol. 17, n.º 4, pp. 135-150, feb. 1993.
[101] Y. Iida, «Ordinality consistency test about items and notation of a pairwise comparison matrix in AHP», en *Proceedings of the international symposium on the analytic hierarchy process*, 2009.
[102] K. Kułakowski, «Inconsistency in the ordinal pairwise comparisons method with and without ties», *European Journal of Operational Research*, vol. 270, n.º 1, pp. 314-327, oct. 2018.
[103] K. Kulakowski y J. Szybowski, «The new triad based inconsistency indices for pairwise comparisons», *Procedia Computer Science*, vol. 35, pp. 1132–1137, 2014.
[104] J. Franek y A. Kresta, «Judgment Scales and Consistency Measure in AHP», *Procedia Economics and Finance*, vol. 12, pp. 164-173, ene. 2014.
[105] T. L. Saaty, «Decision-making with the AHP: Why is the principal eigenvector necessary», *European Journal of Operational Research*, vol. 145, n.º 1, pp. 85-91, feb. 2003.

[106] B. R. Meesariganda y A. Ishizaka, «Mapping verbal AHP scale to numerical scale for cloud computing strategy selection», *Applied Soft Computing*, vol. 53, pp. 111-118, abr. 2017.
[107] B. Blagojevic, B. Srdjevic, Z. Srdjevic, y T. Zoranovic, «Deriving Weights of the Decision Makers Using AHP Group Consistency Measures», *Fundamenta Informaticae*, vol. 144, n.º 3-4, pp. 383-395, ene. 2016.
[108] M. Ivanco, G. Hou, y J. Michaeli, «Sensitivity analysis method to address user disparities in the analytic hierarchy process», *Expert Systems with Applications*, vol. 90, pp. 111-126, dic. 2017.
[109] F. Rozi y E. Herowati, «Supplier selection and consistency level measurement of decision maker using AHP method and expertise level model», *MATEC Web Conf.*, vol. 204, p. 02013, 2018.
[110] E. R. Jalao, T. Wu, y D. Shunk, «A stochastic AHP decision making methodology for imprecise preferences», *Information Sciences*, vol. 270, pp. 192-203, jun. 2014.
[111] A. Mardani, A. Jusoh, y E. K. Zavadskas, «Fuzzy multiple criteria decision-making techniques and applications – Two decades review from 1994 to 2014», *Expert Systems with Applications*, vol. 42, n.º 8, pp. 4126-4148, may 2015.
[112] L. A. Zadeh, «Fuzzy sets», *Information and control*, vol. 8, n.º 3, pp. 338–353, 1965.
[113] C. A. Yajure, «Comparación de los métodos multicriterio AHP y AHP Difuso en la selección de la mejor tecnología para la producción de energía eléctrica a partir del carbón mineral.», *Scientia Et Technica*, vol. 20, n.º 3, pp. 255-260, 2015.
[114] R. Díaz Mora y J. G. Piña, «Uso de AHP y conjuntos difusos para mejorar la toma de decisiones. Caso: selección de empresas contratistas de construcción en la administración pública venezolana», en *7th Latin American and Caribbean Conference for Engineering and Technology*, San Cristóbal, Venezuela, 2009.
[115] A. Suárez Rodríguez, «Utilización de la lógica difusa en la estimación del riesgo en proyectos», en *XIII Congreso Internacional de Ingeniería de Proyectos*, Badajoz, España, 2009, pp. 2091-2102.
[116] P. Ponce Cruz, *Inteligencia artificial con aplicaciones a la ingeniería*. Marcombo, 2011.
[117] M. B. Javanbarg, C. Scawthorn, J. Kiyono, y B. Shahbodaghkhan, «Fuzzy AHP-based multicriteria decision making systems using particle swarm optimization», *Expert Systems with Applications*, vol. 39, n.º 1, pp. 960-966, ene. 2012.
[118] P. J. M. van Laarhoven y W. Pedrycz, «A fuzzy extension of Saaty's priority theory», *Fuzzy Sets and Systems*, vol. 11, n.º 1, pp. 229-241, ene. 1983.
[119] L.-C. Huang y R. Y.-H. Wu, «Applying fuzzy analytic hierarchy process in the managerial talent assessment model – an empirical study in Taiwan's semiconductor industry», *International Journal of Technology Management*, vol. 30, n.º 1-2, pp. 105-130, ene. 2005.
[120] C. G. E. Boender, J. G. de Graan, y F. A. Lootsma, «Multi-criteria decision analysis with fuzzy pairwise comparisons», *Fuzzy Sets and Systems*, vol. 29, n.º 2, pp. 133-143, ene. 1989.
[121] L. Wang, J. Chu, y J. Wu, «Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process», *International Journal of Production Economics*, vol. 107, n.º 1, pp. 151-163, may 2007.
[122] R. Xu, «Fuzzy least-squares priority method in the analytic hierarchy process», *Fuzzy Sets and Systems*, vol. 112, n.º 3, pp. 395-404, jun. 2000.
[123] R. Csutora y J. J. Buckley, «Fuzzy hierarchical analysis: the Lambda-Max method», *Fuzzy Sets and Systems*, vol. 120, n.º 2, pp. 181-195, jun. 2001.

[124] L. Mikhailov, «Deriving priorities from fuzzy pairwise comparison judgements», *Fuzzy sets and systems*, vol. 134, n.º 3, pp. 365–385, 2003.
[125] Y. M. Wang, T. M. S. Elhag, y Z. Hua, «A modified fuzzy logarithmic least squares method for fuzzy analytic hierarchy process», *Fuzzy Sets and Systems*, vol. 157, n.º 23, pp. 3055-3071, dic. 2006.
[126] C. E. Bozdağ, C. Kahraman, y D. Ruan, «Fuzzy group decision making for selection among computer integrated manufacturing systems», *Computers in Industry*, vol. 51, n.º 1, pp. 13-29, may 2003.
[127] E. Tolga, M. L. Demircan, y C. Kahraman, «Operating system selection using fuzzy replacement analysis and analytic hierarchy process», *International Journal of Production Economics*, vol. 97, n.º 1, pp. 89-117, jul. 2005.
[128] F. Tüysüz y C. Kahraman, «Project risk evaluation using a fuzzy analytic hierarchy process: An application to information technology projects», *International Journal of Intelligent Systems*, vol. 21, n.º 6, pp. 559-584, 2006.
[129] G. Büyüközkan, C. Kahraman, y D. Ruan, «A fuzzy multi-criteria decision approach for software development strategy selection», *International Journal of General Systems*, vol. 33, n.º 2-3, pp. 259-280, abr. 2004.
[130] F. T. Bozbura, A. Beskese, y C. Kahraman, «Prioritization of human capital measurement indicators using fuzzy AHP», *Expert Systems with Applications*, vol. 32, n.º 4, pp. 1100-1112, may 2007.
[131] G. H. Enríquez, F. G. Viejó, S. C. Paez, y D. Z. Vera, «Proceso Analítico Jerárquico Difuso en la selección de variables para la evaluación de la resiliencia en zonas afectadas por desastres», *Ingeniería Industrial. Actualidad y Nuevas Tendencias*, vol. V, n.º 16, pp. 45-66, 2016.
[132] A. Gnanavelbabu y P. Arunagiri, «Ranking of MUDA using AHP and Fuzzy AHP algorithm», *Materials Today: Proceedings*, vol. 5, n.º 5, Part 2, pp. 13406-13412, ene. 2018.
[133] C. Kahraman, U. Cebeci, y D. Ruan, «Multi-attribute comparison of catering service companies using fuzzy AHP: The case of Turkey», *International Journal of Production Economics*, vol. 87, n.º 2, pp. 171-184, ene. 2004.
[134] K. Mandic, B. Delibasic, S. Knezevic, y S. Benkovic, «Analysis of the financial parameters of Serbian banks through the application of the fuzzy AHP and TOPSIS methods», *Economic Modelling*, vol. 43, pp. 30-37, dic. 2014.
[135] J.-F. Chen, H.-N. Hsieh, y Q. H. Do, «Evaluating teaching performance based on fuzzy AHP and comprehensive evaluation approach», *Applied Soft Computing*, vol. 28, pp. 100-108, mar. 2015.
[136] S. Kubler, J. Robert, W. Derigent, A. Voisin, y Y. Le Traon, «A state-of the-art survey & testbed of fuzzy AHP (FAHP) applications», *Expert Systems with Applications*, vol. 65, pp. 398-422, dic. 2016.
[137] P. A. Barajas Díaz, V. A. Cloquell Ballester, J. L. García Alcaraz, y A. A. Alvarado Iniesta, «El Proceso Jerárquico Analítico y Lógica Difusa: Sus Aplicaciones», en *Congreso Internacional de Investigación*, Celaya Guanajuato México, 2012, vol. 4, pp. 249–254.
[138] M. A. Waller y S. E. Fawcett, «Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management», *Journal of Business Logistics*, vol. 34, n.º 2, pp. 77-84, 2013.
[139] A. McAfee y E. Brynjolfsson, «Big data: the management revolution», *Harv Bus Rev*, vol. 90, n.º 10, pp. 60-68, oct. 2012.

[140] K. Eckert y P. V. Britos, «Modelo basado en la toma decisiones con criterios múltiples para la elección de metodologías de data science», presentado en XX Workshop de Investigadores en Ciencias de la Computación, 2018.
[141] R. Agarwal y V. Dhar, «Editorial-Big Data, Data Science, and Analytics: The Opportunity and Challenge for IS Research», *Information Systems Research*, vol. 25, n.º 3, pp. 443-448, sep. 2014.
[142] F. Provost y T. Fawcett, «Data Science and its Relationship to Big Data and Data-Driven Decision Making», *Big Data*, vol. 1, n.º 1, pp. 51-59, feb. 2013.
[143] T. Schoenherr y C. Speier-Pero, «Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential», *Journal of Business Logistics*, vol. 36, n.º 1, pp. 120-132, 2015.
[144] B. T. Hazen, C. A. Boone, J. D. Ezell, y L. A. Jones-Farmer, «Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications», *International Journal of Production Economics*, vol. 154, pp. 72-80, ago. 2014.
[145] J. M. Moine, «Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo», Tesis de Maestría, Facultad de Informática, 2013.
[146] M. T. Rodríguez Montequín, J. V. Álvarez Cabal, J. M. Mesa Fernández, y A. González Valdés, «Metodologías para la realización de proyectos de Data Mining», presentado en VII Congreso Internacional de Ingeniería de Proyectos, Pamplona España, 2003, pp. 257-265.
[147] P. Pytel, P. Britos, y R. García Martínez, «Proposal and Validation of a feasibility Model for Information Mining Projects», presentado en 25th International Conference on Software Engineering and Knowledge Engineering, Boston, USA, pp. 33-88.
[148] J. Schiefer, and S. Kapoor, y P. Chowdhary, «Process information factory: a data management approach for enhancing business process intelligence», en *Proceedings. IEEE International Conference on e-Commerce Technology, 2004. CEC 2004.*, 2004, pp. 162-169.
[149] P. Chapman *et al.*, «CRISP-DM 1.0: Step-by-Step Data Mining Guide». Edited by SPSS, 2000.
[150] D. Pyle, *Business Modeling and Data Mining*, 1st ed. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2003.
[151] J. Á. Vanrell, R. A. Bertone, y R. García Martínez, «Modelo de proceso de operación para proyectos de explotación de información», presentado en XVI Congreso Argentino de Ciencias de la Computación, 2010.
[152] S. Martins, P. Pesado, y R. García Martínez, «Propuesta de Modelo de Procesos para una Ingeniería de Explotación de Información: MoProPEI», *Revista Latinoamericana de Ingenieria de Software*, vol. 2, n.º 5, pp. 313-332, 2014.
[153] A. I. R. L. Azevedo y M. F. Santos, «KDD, SEMMA and CRISP-DM: a parallel overview», *IADS - DM*, 2008.
[154] P. V. Britos, «Procesos de explotación de información basados en sistemas inteligentes», Tesis, Facultad de Informática, 2008.
[155] J. M. Moine, S. E. Gordillo, y A. S. Haedo, «Análisis comparativo de metodologías para la gestión de proyectos de minería de datos», presentado en XVII Congreso Argentino de Ciencias de la Computación, 2011.
[156] J. C. Giraldo Mejia y J. A. Jiménez Builes, «Caracterización del proceso de obtención de conocimiento y algunas metodologías para crear proyectos de minería de datos», *Revista Latinoamericana de Ingeniería de Software*, 2013.

[157] H. J. G. Palacios, G. A. H. Pantoja, A. A. M. Navarro, I. M. A. Puetaman, y R. A. J. Toledo, «Comparative between CRISP-DM and SEMMA for data cleaning of MODIS products in a study of land use and land cover change», en *2016 IEEE 11th Colombian Computing Conference (CCC)*, 2016, pp. 1-9.

[158] IBM, «Analytics solutions unified method - Implementations with agile principles». IBM, 2016.

[159] D. R. P. Nogueira, «Agile Data Mining: uma metodologia ágil para o desenvolvimento de projetos de data mining», Maestría, Faculdade de Engenharia da Universidade do Porto, Brasil, 2014.

[160] W. A. R. Roald Bradley Severtson, «What is the Team Data Science Process?» [En línea]. Disponible en: https://docs.microsoft.com/en-us/azure/machine-learning/team-data-science-process/overview. [Accedido: 19-nov-2019].

[161] E. Fernández, H. Merlino, M. A. Ochoa, E. Diez, P. V. Britos, y R. García Martínez, «Gestión asistida de documentos en una metodología de explotación de información», presentado en XI Congreso Argentino de Ciencias de la Computación, 2005.

[162] R. García Martínez y P. V. Britos, *Ingenieria de Sistemas Expertos*. Nueva Librería, 2004.

[163] G. Kelly, *The Psychology of Personal Constructs : Volume One: Theory and Personality*. Routledge, 2002.

[164] P. Britos, B. Rossi, y R. García Martínez, «Notas sobre didáctica de las etapas de formalización y análisis de resultados de la técnica de emparrillado. Un Ejemplo», en *Proceedings del V Congreso Internacional de Ingeniería Informática*, 1999, pp. 200–209.

[165] T. Butt, *George Kelly: The Psychology of Personal Constructs*. Macmillan International Higher Education, 2008.

[166] B. Liu y F. Kong, «Research and application of sidewall stability prediction method based on analytic hierarchy process and fuzzy integrative evaluation method», *Natural Science*, vol. 4, n.º 02, p. 142, 2012.

[167] J. Aguarón y J. M. Moreno Jiménez, «The geometric consistency index: Approximated thresholds», *European Journal of Operational Research*, vol. 147, n.º 1, pp. 137-145, may 2003.

[168] P. McCauley-Bell y A. B. Badiru, «Fuzzy modeling and analytic hierarchy processing-means to quantify risk levels associated with occupational injuries. II. The development of a fuzzy rule-based model for the prediction of injury», *IEEE Transactions on Fuzzy Systems*, vol. 4, n.º 2, pp. 132-138, may 1996.

[169] S. Tyagi, S. Agrawal, K. Yang, y H. Ying, «An extended Fuzzy-AHP approach to rank the influences of socialization-externalization-combination-internalization modes on the development phase», *Applied Soft Computing*, vol. 52, pp. 505-518, mar. 2017.

Anexos

Anexo I: Proceso Analítico Jerárquico

I.I. Axiomas

AHP consta de cuatro axiomas básicos [1], [6]:

1. Reciprocidad: Las comparaciones y preferencias del decisor se realizan de a pares; garantizando un análisis bidireccional. Si frente a un criterio, una alternativa A es x veces mejor o preferida que B, entonces B es 1/x veces preferida que A.
2. Homogeneidad: Las preferencias se representan mediante una escala definida; por lo que los elementos comparados no deben diferir en mucho según las características establecidas.
3. Independencia: Los juicios de prioridades de los elementos en una jerarquía son independientes de los elementos del nivel inferior (subcriterios y/o alternativas).
4. Expectativas: Se asume que la jerarquía es completa, dado el propósito u objetivo de la TD.

I. II. Etapas

La estructura de AHP, se divide en cuatro etapas básicas [1], [54], [61]:

1. Modelización: Consiste en la construcción de la estructura jerárquica, donde se identifican en diferentes niveles los aspectos relevantes del problema; siendo ésta representativa, completa y concreta al problema en cuestión. Es la parte creativa del proceso.
2. Valorización: Se identifican las preferencias mediante matrices de comparación pareadas.
3. Priorización y síntesis: Se suministran las distintas prioridades consideradas en la resolución del problema; las cuales se ajustan a una escala definida y permite comparar aspectos tangibles e intangibles. Existen tres tipos de prioridades:
    a. Locales: Se encuentran en un mismo nivel y bajo el mismo nodo.
    b. Globales: Importancia de los elementos individuales respecto al objetivo o meta global. El cálculo se realiza multiplicando los diferentes pesos de la jerarquía (por rama), desde el elemento inferior (hoja) hasta el objetivo global (nodo raíz).

c. Final o total: El método comúnmente utilizado en AHP, es agregar por cada alternativa las prioridades globales de las diferentes ramas; equivalente a la suma ponderada.

4. Análisis de sensibilidad: Se puede examinar el grado de sensibilidad de los resultados obtenidos, realizando cambios en las prioridades de los criterios principales.

I.II: Secuencia de pasos generales

La secuencia de pasos a seguir para concretar las mencionadas etapas son los siguientes [1], [61], [15], [54], [57], [77], [166] (las mismas fueron mencionadas resumidamente en apartado 2.1.1):

Paso 1: Definición del problema, determinación del tipo de conocimiento buscado y estructuración del problema como una jerarquía

Primeramente se debe modelar el problema de decisión a resolver como una estructura jerárquica; como se puede ver en la Figura I.I:

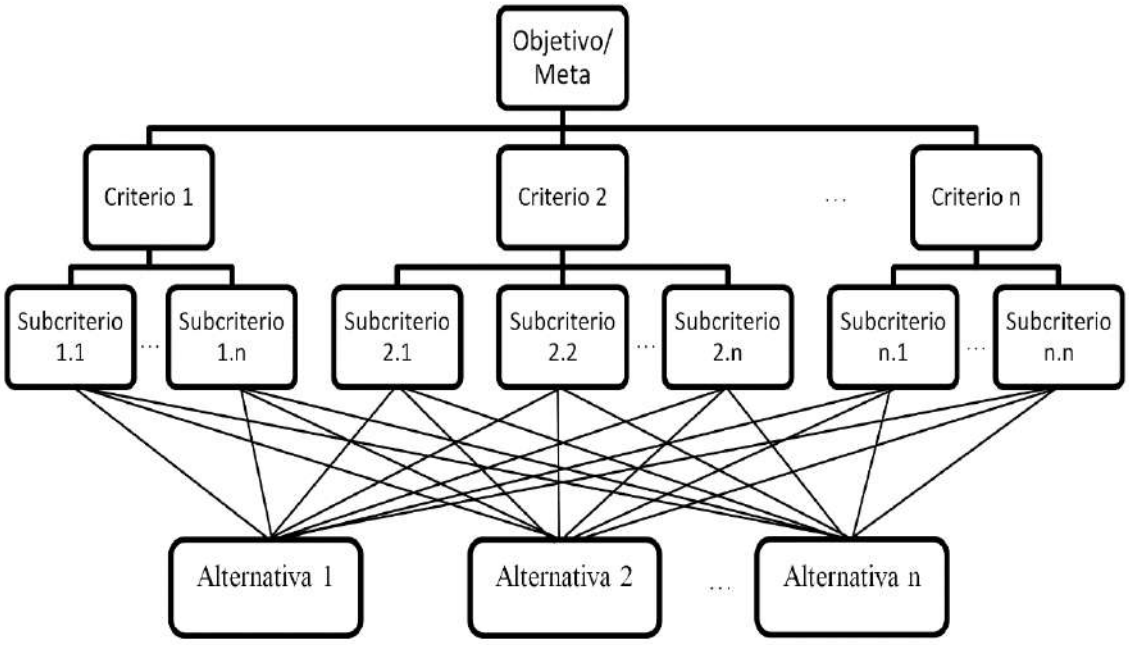

Figura I.I: Estructura Jerárquica AHP

En el nivel superior de la jerarquía se encuentra el objetivo o meta que se desea alcanzar. El problema de decisión consiste en elegir la alternativa que mejor contribuya al objetivo o meta propuesta.

En el segundo nivel se encuentran los criterios que atañen a aspectos tales como atributos, objetivos o parámetros a partir de los cuales se justifican, transforman y argumentan las preferencias. Se recomienda que los criterios sean no redundantes, completos para cumplir con todos los objetivos, que permitan la simplificación del proceso de evaluación mediante la descomposición en partes del problema y lo más pequeños posible sin perder aspectos importantes del problema. Puede darse el caso en que los criterios se puedan o deban descomponer en sub-criterios formando otra jerarquía descendente. Para la construcción de la jerarquía, el problema se debe analizar en profundidad para representarlo de forma completa y global, considerando el entorno que rodea al problema.

En el último nivel de la jerarquía se encuentran las alternativas.

Paso 2: Establecimiento de las prioridades entre los criterios

Cuando se mide algo con respecto a una propiedad, usualmente se usa una escala conocida para ese propósito. Existen dos estrategias para asignar un valor numérico a cada criterio que se represente, estas son la asignación directa e indirecta. La primera se elabora mediante una escala definida que correlaciona valores de grado con valores numéricos y la segunda se realiza a través de técnicas de comparaciones entre criterios.

En este paso en particular, se busca construir el vector de prioridades o pesos que evalúe la importancia relativa de un criterio respecto a otro. Se usa una asignación de pesos indirecta, lo cual implica un análisis exhaustivo; realizando comparaciones pareadas entre criterios y empleando los conceptos matemáticos de autovector/valor y autovector/valor propio.

AHP permite centrarse en el juicio por separado en cada uno de los criterios o prioridades esenciales para tomar una decisión acertada. Razón por la cual las comparaciones pareadas en combinación con la estructura jerárquica son tan útiles para derivar mediciones.

Para la comparación de unos criterios C_i $(i = 1,2, \dots n)$ se establecen los pesos w_i asociados a dichos criterios. La comparación se expresa mediante la matriz W y se busca encontrar un vector de prioridades (Ecuación I.I):

$$w = (w_1 , w_2 , \dots , w_n)^T \tag{I.I}$$

Para lo cual se plantea la siguiente ecuación (Ecuación I.II):

$$\begin{bmatrix} w_1/w_1 & w_1/w_2 & \dots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \dots & w_2/w_n \\ \vdots & \vdots & \ddots & \\ w_n/w_1 & w_n/w_2 & \dots & w_n/w_n \end{bmatrix} . \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix} = n . \begin{bmatrix} w_1 \\ w_2 \\ \dots \\ w_n \end{bmatrix} \quad \text{(I.II)}$$

Donde w_i corresponde a los pesos de los criterios.

La matriz de la Ecuación I.III:

$$W = (w_{ij}), w = \left(\frac{w_i}{w_j}\right), \text{para } i, j = 1,2, \dots \quad \text{(I.III)}$$

Tiene entradas positivas en todas partes y satisface la propiedad recíproca (Ecuación I.IV):

$$C_{ij} . C_{ji} = 1, \text{para } i, j = 1,2, \dots n. \quad \text{(I.IV)}$$

La matriz W cuenta con un único autovector distinto de cero; igual a la dimensión n de la matriz y el vector propio asociado es w; con lo que el problema del autovector se reduce a (Ecuación I.V):

$$W.w = n.w \quad \text{(I.V)}$$

Con un modo distributivo (Ecuación I.VI):

$$\sum_j w_j = 1 \quad \text{(I.VI)}$$

La suma de los criterios o elementos de la matriz W para la fila i se obtiene mediante la siguiente ecuación (Ecuación I.VII):

$$w_i . \sum_{j=1}^{n} \frac{1}{w_j} \quad \text{(I.VII)}$$

Y la suma para la columna j, se puede definir como (Ecuación I.VIII):

$$\frac{1}{w_j}.\sum_{i=1}^{n} w_i = \frac{1}{w_j} \quad \text{(I.VIII)}$$

De modo que si se normaliza la suma de columnas se obtiene el vector w; característica de la matriz W que se emplean en el análisis de perturbaciones.

Se debe tener en cuenta el vector de prioridades w se puede encontrar mediante la normalización de cualquier columna de W, dado que cada columna de la matriz W es múltiplo constante de w. Asimismo, se dice que W es consistente porque satisface la condición de la Ecuación I.IX:

$$w_{ik}.w_{kj} = w_{ij} \quad \text{para todo } i,j,k \quad \text{(I.IX)}$$

Tabla I.I: Escala fundamental de comparación pareada

Escala numérica	Escala verbal	Explicación
1	Igual importancia o preferencia	Dos criterios contribuyen por igual al objetivo
2	De igual a moderadamente más preferido	
3	Importancia moderada de un elemento sobre otro	La experiencia y el juicio están a favor de un criterio sobre otro
4	De moderadamente a una importancia fuerte	
5	Importancia fuerte de un elemento sobre otro	Un criterio es fuertemente favorecido respecto a otro
6	De fuerte a muy fuertemente preferido	
7	Importancia muy fuerte de un elemento sobre otro	Un criterio es muy dominante o fuertemente preferido sobre otro
8	De muy fuerte a extremadamente preferido	
9	Importancia extrema de un elemento sobre otro	Extremadamente más preferido. La evidencia que favorece un criterio sobre otro es la más alta posible orden de afirmación

Al establecer las prioridades a través de la comparación pareada, se busca definir los pesos relativos de los criterios, dichos valores numéricos son positivos e indican la

importancia o prioridad relativa entre el criterio C_i y el criterio C_j respecto al elemento del nivel inmediatamente superior en la jerarquía del problema. Para ello se utiliza una escala previamente establecida, la más difundida es la escala fundamental propuesta por Saaty, visible en la Tabla I.I; los valores pares de la escala se utilizan de compromiso entre un valor y otro. La numeración de la escala representa la proporción de importancia de un elemento respecto a otro, en relación a un criterio u objetivo que tienen en común.

Haciendo uso de la escala fundamental, el decisor debe determinar los pesos asignados a cada criterio, completando para ello la matriz R, donde r_{ij} representa la prioridad relativa entre el criterio C_i y el criterio C_j respecto al objetivo global. La matriz obtenida se representa de la siguiente manera (Ecuación I.X):

$$R=\begin{bmatrix} 1 & r_{12} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2n} \\ \vdots & \vdots & \ddots & 1 \\ r_{n1} & r_{n2} & & \end{bmatrix} \quad \text{(I.X)}$$

Donde $r_{ij}.r_{ji} = 1$, satisfaciendo la propiedad recíproca. Para una matriz de estas características, se cumple que el autovector máximo $\lambda_{máx}$ es un número real positivo y que existe un vector propio Z, cuyos elementos asociados a este autovector son positivos. A posteriori se debe normalizar el autovector para que la suma sea la unidad.

Paso 3: Establecimiento y control de las prioridades locales y globales

El decisor debe ser concienzudo con sus valoraciones, ya que un error de incoherencia en el proceso de comparación pareada, genera una matriz y un autovector poco representativo; producto de una contradicción al incumplir la transitividad de los valores. Supóngase que se realizó unas valoraciones en las que $C_1 > C_2$ y $C_2 > C_3$; por transitividad se saca que $C_1 > C_3$; pero puede ocurrir que se haya indicado que $C_3 > C_1$. Como se mencionó en párrafos anteriores, para que la matriz sea consistente debe cumplir con la Ecuación I.XI:

$$w_{ij}.w_{jk} = w_{ik} \quad \text{para todo } i,j,k \quad \text{(I.XI)}$$

Saaty propone la razón o relación de consistencia (CR, Consistency Ratio) para evaluar la coherencia en los juicios emitidos por el decisor. CR se expresa según la siguiente ecuación (Ecuación I.XII):

$$CR = \frac{CI}{RI} \tag{I.XII}$$

Donde se calcula el cociente entre CI, que es el índice de consistencia (Consistency Index) y RI, que corresponde al índice de aleatoriedad (Random Index).
CI se define como (Ecuación I.XIII):

$$CI = \frac{\lambda_{max} - n}{n - 1} \tag{I.XIII}$$

Donde λ_{max} es el máximo valor propio de la matriz y n su orden.
RI es una medida que se utiliza para mejorar la consistencia de los juicios teniendo en cuenta la dimensión de la matriz. Mediante la simulación de 100.000 matrices recíprocas [167] generadas aleatoriamente utilizando la escala de Saaty (Tabla I.I); se definió el RI medio, representado en la Tabla I.II.

Tabla I.II: Índice de consistencia aleatorio (RI) en función al tamaño de la matriz (n)

n	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
RI	0	0	0.525	0.882	1.115	1.252	1.341	1.404
n	9	10	11	12	13	14	15	16
RI	1.452	1.484	1.513	1.535	1.555	1.570	1.583	1.595

Para una consistencia adecuada de la matriz, Saaty indicó que el CR debe ser menor o igual al 10% ($CR = 0{,}10$); es decir que cuando más cercano a 0, mayor es la consistencia de los juicios emitido; por el contrario contrario, si CR es mayor $0{,}10$ se debería rever los juicios emitidos para el problema dado, aunque siendo más tolerables se puede tener una buena estimación y resultados hasta un 15% de tolerancia ($CR = 0{,}15$).

Si el problema modelado, requiere una descomposición de ciertos criterios en sub-criterios, se debe calcular previamente el peso global asociados a los mismos. El procedimiento es igual al especificado anteriormente, empero en este caso se deberán realizar las comparaciones entre pares de sub-criterios para determinar la importancia

relativa respecto al criterio inmediatamente superior en la jerarquía. Precisamente se denomina prioridad local, al cálculo del vector de pesos asociados a un conjunto de sub-criterios respecto a su criterio superior o padre.

Para el cálculo de la importancia relativa global de cada sub-criterio respecto al objetivo del problema, se calcula el producto de los diferentes pesos de cada uno de los sub-criterios y criterios, recorriendo la jerarquía desde la parte inferior hasta la cúspide de la misma, este procedimiento se denomina composición jerárquica.

Paso 4: Establecimiento de las prioridades totales asociadas a cada alternativa

Saaty propone el uso del método de la suma ponderada; que consiste en encontrar el vector global de prioridades p, que agrega las prioridades obtenidas de los criterios y alternativas. Los componentes p_i del vector corresponden a las prioridades totales asociadas a cada alternativa A_i, reflejando el valor total que tiene para el decisor cada alternativa; para lo mencionado se puede recurrir a la siguiente expresión (Ecuación I.XIV):

$$p_i = \sum_{j=1}^{n} (w_j \, . \, r_{ij}) \quad i = 1, 2, \dots, m \tag{I.XIV}$$

Donde w_j corresponde a los pesos asociados a cada uno de los criterios que se consideran y r_{ij} son los componentes de la matriz normalizada. Para resolver el problema de decisión y determinar la mejor alternativa de todas, que será la de suma ponderada mayor, basta con la ordenar las alternativas en base a los valores $p_1, p_2, \dots, p_m$.

Anexo II: Proceso Analítico Jerárquico Difuso

Como el Proceso Analítico Jerárquico Difuso (FAHP) es una extensión del método AHP, respeta los cuatro axiomas detallados en el Anexo I, empero incorporando aspectos de la lógica difusa.

II.I. Secuencia de pasos

Paso 1: Definición del problema como una estructura jerárquica

Al igual que el método AHP, primeramente se modela el problema de decisión a resolver como una estructura jerárquica, ver el paso 1 de la secuencia de pasos del Anexo I.

Paso 2: Establecimiento de prioridades entre los criterios

Como en el caso del método AHP, se deben establecer las prioridades de los criterios comparándolos de a pares. Pero a diferencia del método AHP, con FAHP se debe utilizar una función de membresía; como se indicó en el apartado 2.1.2., existen diversas funciones de membresía como la Triangular, Gaussiana y Trapezoidal [1], [113].

Zadeh [112] desarrolló la teoría de los conjuntos difusos, la cual representa los juicios humanos matemáticamente, para capturar la imprecisión y vaguedad de la información suministrada. Puede manejar la incertidumbre de manera efectiva si las evaluaciones difusas coinciden con las variaciones en la información de del entrada tomador de decisiones [168], [169]. Un conjunto difuso se caracteriza por una función de membresía; para calcular el valor de membresía (μ) se puede utilizar la siguiente ecuación (Ecuación II.I) [133], [169]:

$$\mu(x|\widetilde{M})=\begin{bmatrix} 0 & x<1 \\ \dfrac{(x-l)}{(m-l)} & (l\leq x\leq m) \\ \dfrac{(u-x)}{(u-m)} & (m\leq x\leq u) \\ 0 & x>u \end{bmatrix} \quad \text{(II.I)}$$

Donde l, m y u representan el valor pesimista (más pequeño posible), el valor más prometedor y el valor optimista (más grande posible) para un conjunto difuso. Estas estimaciones de tres puntos (o tripletas) se convierten en la distribución triangular; formando un número borroso triangular (TFN, Triangular Fuzzy Number) de un conjunto difuso, que se representa al poner ~ a la letra M [133], [169].

Para desarrollar métricas de comparación difusa, en la Figura II.I se muestra la función de membresía triangular y TFN para las comparaciones pareadas a partir de la cual se define la escala de conversión difusa triangular utilizada para la presente tesis; la cual es representada en la Tabla II.I.

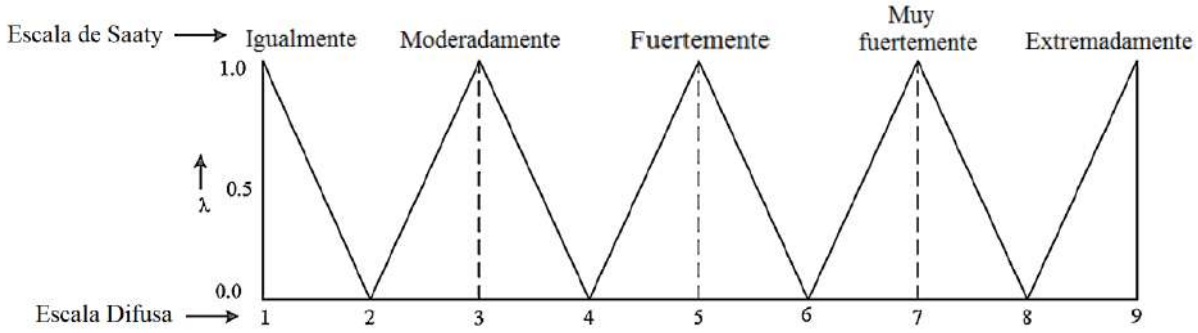

Figura II.I: Función de membresía triangular en comparaciones pareadas [169].

Tabla II.I: Escala de conversión difusa triangular

Escala lingüística	Escala difusa triangular	Escala recíproca difusa triangular
Igualmente importante	(1,1,2)	(1/2,1,1)
Moderadamente más importante	(2,3,4)	(1/4.1/3,1/2)
Fuertemente importante	(4,5,6)	(1/6.1/5,1/4)
Muy fuertemente importante	(6,7,8)	(1/8.1/7,1/6)
Extremadamente importante	(8,9,9)	(1/9.1/9,1/8)

Se solicita al responsable de la toma de decisiones que facilite la comparación de un criterio con respecto a otro en términos lingüísticos, teniendo en cuenta el objetivo general como en el caso de AHP. Para la presente tesis se tomaron como entrada de ambos métodos (AHP y FAHP) las valoraciones obtenidas a partir de las plantillas

para definición de los sub-criterios mediante la técnica de la Teoría de la Construcción Personal (ver Anexo III) y las matrices de dos entradas (ver Apéndice D).

Sea $X = \{x_1, x_2, x_3, \dots, x_n\}$ un conjunto de objetos y $U = \{u_1, u_2, u_3, \dots, u_m\}$ un conjunto de objetivos, se desarrolla el análisis ampliado para cada uno de los valores de los objetos, acorde al método de análisis extendido de Chang [28]; de este modo se pueden obtener para cada objetivo g_i. Por lo tanto, los valores de análisis extendido de m se pueden obtener con la siguiente notación (Ecuación II.II) [131], [133]:

$$M_{gi}^1, M_{gi}^1, \dots, M_{gi}^1, \qquad \text{i=1,2,3,…,n} \tag{II.II}$$

Donde todo $M_{gi}^j (j = 1,2,3, \dots m)$ son TFN.

Paso 3: Cálculo del valor de extensión sintética difusa

A partir de este paso se describen los cuatros pasos esenciales definidos por el método de análisis extendido de Chang [28].

Para el criterio de confirmación con respecto a cualquier objeto (o) es estimado por la Ecuación II.III [28], [131], [133], [169]:

$$S_i = \sum_{j=1}^{m} M_{gi}^j \otimes \left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^j\right]^{-1} \tag{II.III}$$

Donde $\sum_{j=1}^{m} M_{gi}^j$ se obtiene a partir de la operación de adición borrosa de m valores del análisis extendido para una matriz en particular, como se expone en la Ecuación II.IV [28], [131], [133], [169]:

$$\sum_{j=1}^{m} M_{gi}^j = \left(\sum_{j=1}^{m} l_j, \sum_{j=1}^{m} m_j, \sum_{j=1}^{m} u_j\right) \tag{II.IV}$$

$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^j\right]^{-1}$, se obtiene a partir de la operación de adición borrosa de los valores M_{gi}^j $(j = 1,2,3, \dots, m)$, mediante la siguiente ecuación (Ecuación II.V) [28], [131], [133], [169]:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^{j} = \left(\sum_{j=1}^{n} l_j\,, \sum_{j=1}^{n} m_j\,, \sum_{j=1}^{n} u_j\right) \tag{II.V}$$

Posteriormente se calcula el vector inverso con la siguiente ecuación (Ecuación II.VI) [28], [131], [133], [169]:

$$\left[\sum_{i=1}^{n}\sum_{j=1}^{m} M_{gi}^{j}\right]^{-1} = \left(\frac{1}{\sum_{j=1}^{n} u_i}, \frac{1}{\sum_{j=1}^{n} m_i}, \frac{1}{\sum_{j=1}^{n} l_i}\right) \tag{II.VI}$$

Paso 4: Determinación de la superioridad comparativa

El grado de posibilidad de que $M_2 = (l_2, m_2, u_2) \geq M_1 = (l_1, m_1, u_1)$ se define según la Ecuación II.VII como [28], [131], [133], [169]:

$$V(M_2 \geq M_1) = sup_{y \geq x}[\min(\mu_{M_1}(x), \mu_{M_2}(y))] \tag{II.VII}$$

Pudiendo ser expresado de manera equivalente según la siguiente ecuación (Ecuación II.VIII) [28], [131], [133], [169]:

$$V(M_2 \geq M_1) = hgt(M_1 \cap M_2) = \mu_{M_2}(d)$$
$$= \begin{bmatrix} 1, & si\ m_2 \geq m_1 \\ 0, & si\ l_1 \geq u_2 \\ \dfrac{l_1 - u_2}{(m_2 - u_2) - (m_1 - l_1)}, & de\ otra\ manera \end{bmatrix} \tag{II.VIII}$$

Donde d es la ordenada del punto de intersección más alta del punto D situado entre μ_{M_1} y μ_{M_2} en la Figura II.II. Se requiere los valores de $V(M_1 \geq M_2)$ y $V(M_2 \geq M_1)$ para comparar M_1 y M_2 [28], [129], [131], [133].

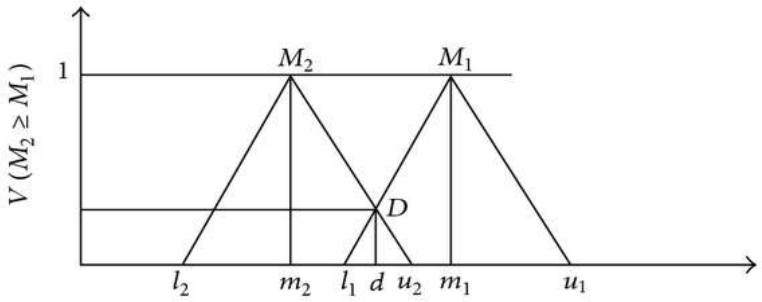

Figura II.II: Intersección entre M_1 y M_2 [28]

Paso 5: Selección del valor mínimo de superioridad y cálculo del vector de peso normalizado

El grado de posibilidad de que un número difuso convexo sea mayor que k números convexos se define mediante la siguiente ecuación (Ecuación II.IX) [28], [131], [133], [169]:

$$V(M \geq M_1, M_2, \dots, M_k) = V[(M \geq M_1)\ y\ (M \geq M_2)\ y \dots y\ (M \geq M_k)] = \min\ V(M \geq M_i)\ , \qquad i = 1,2,3, \dots, k \tag{II.IX}$$

Asumiendo que (Ecuación II.X) [28], [131], [133], [169]:

$$d'(A_i) = minV(S_i \geq S_k) \quad para\ k = 1,2,3, \dots, n; k \neq i \tag{II.X}$$

El peso del vector está dado por la Ecuación II.XI [28], [131], [133], [169]:

$$W' = (d'(A_1), d'(A_2), \dots,\ d'(A_n))^T \tag{II.XI}$$

Donde $A_i (i = 1,2,3, \dots, n)$ son n elementos.

La normalización del vector se expresa mediante la siguiente fórmula (Ecuación II.XII) [28], [131], [133], [169]:

$$W = (d(A_1), d(A_2), \dots, (A_n))^T \tag{II.XII}$$

Donde W no es un número difuso y representa el conjunto de pesos para cada matriz.

Se repite los pasos 2 al 5 para determinar el vector peso normalizado de cada sub-criterio en respuesta a los criterios. Además, se estima el peso normalizado de las alternativas correspondientes a cada sub-criterio [169].

Paso 6: Establecimiento de las prioridades parciales

Para estimar el peso de las prioridades parciales de cada alternativa, se multiplican los valores del vector de peso normalizados de cada alternativa con los de los sub-criterios por separado [28], [169].

Paso 7: Establecimiento de las prioridades totales asociadas a cada alternativa

Finalmente, al agregar las ponderaciones por alternativa multiplicadas por las ponderaciones de los criterios correspondientes, se obtiene una puntuación final para cada alternativa candidata. Esto se obtiene calculando el producto de los valores normalizados de los criterios con los valores de multiplicación de la alternativa y sub-criterios estimados en el paso 6 y se los agrega para estimar el peso final de prioridad. El modo con mayor valor de peso de prioridad final tiene la mayor influencia en la fase de desarrollo [28], [169].

Anexo III: Matrices de dos entradas

Teniendo en cuenta la escala de Saaty, marcar con *X* los renglones (filas) en gris, según corresponda en cada tabla (III.I a III.XXIV).

III.I. Criterios principales

Tabla III.I: Matriz de dos entradas para criterios principales (nivel 2)

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
Entendimiento de los Datos					Comprensión del Negocio			

III.II. Entendimiento de los Datos

Tabla III.II: Matriz de dos entradas para criterios Acceso y Uso de los Datos

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
Acceso a los Datos					Uso de los Datos			

Características:

Acceso a los Datos

1. Portabilidad (capacidad de ser transferido)
2. Accesibilidad (nivel de accesibilidad/asequibilidad de los datos)
3. Diversidad (diferentes tipos de datos)
4. Diversidad de fuentes de datos
5. Recursos necesarios (equipamiento)

Uso de Datos

6. Calidad (condiciones de los datos (valores válidos e inexistencia de valores anómalos))
7. Completitud (grado de completitud de los datos (cantidad de valores faltantes))
8. Funcionalidad (uso y practicidad de los datos)
9. Costos de transformación (demanda de trabajo para transformarlos/adaptarlos para su uso)
10. Datos riesgosos (datos sensibles, que no deben ser difundidos)

Acceso a los Datos

Tabla III.III: Matriz de dos entradas para la Portabilidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C1: Portabilidad					C2: Accesibilidad			
C1: Portabilidad					C3: Diversidad			
C1: Portabilidad					C4: Diversidad de fuentes de datos			
C1: Portabilidad					C5: Recursos necesario			

Tabla III.IV: Matriz de dos entradas para la Accesibilidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C2: Accesibilidad					C3: Diversidad			
C2: Accesibilidad					C4: Diversidad de fuentes de datos			

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C2: Accesibilidad					C5: Recursos necesario			

Tabla III.V: Matriz de dos entradas para la Diversidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C3: Diversidad					C4: Diversidad de fuentes de datos			
C3: Diversidad					C5: Recursos necesario			

Tabla III.VI: Matriz de dos entradas para la Diversidad de fuentes de datos y Recursos necesarios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C4: Diversidad de fuentes de datos					C5: Recursos necesario			

Uso de los Datos

Tabla III.VII: Matriz de dos entradas para la Calidad y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C6: Calidad					C7: Completitud			

Extrema importancia: 9	Importancia muy fuerte :7	Importancia fuerte :5	Importancia moderada: 3	Igual importancia: 1	Importancia moderada: 3	Importancia fuerte :5	Importancia muy fuerte :7	Extrema importancia: 9
C6: Calidad					C8: Funcionalidad			
C6: Calidad					C9: Costos de transformación			
C6: Calidad					C10: Datos riesgosos			

Tabla III.VIII: Matriz de dos entradas para la Completitud y demás sub-criterios

Extrema importancia: 9	Importancia muy fuerte :7	Importancia fuerte :5	Importancia moderada: 3	Igual importancia: 1	Importancia moderada: 3	Importancia fuerte :5	Importancia muy fuerte :7	Extrema importancia: 9
C7: Completitud					C8: Funcionalidad			
C7: Completitud					C9: Costos de transformación			
C7: Completitud					C10: Datos riesgosos			

Tabla III.IX: Matriz de dos entradas para la Funcionalidad y demás sub-criterios

Extrema importancia: 9	Importancia muy fuerte :7	Importancia fuerte :5	Importancia moderada: 3	Igual importancia: 1	Importancia moderada: 3	Importancia fuerte :5	Importancia muy fuerte :7	Extrema importancia: 9
C8: Funcionalidad					C9: Costos de transformación			
C8: Funcionalidad					C10: Datos riesgosos			

Tabla III.X: Matriz de dos entradas para el Costo de transformación y Datos riesgosos

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C9: Costos de transformación					C10: Datos riesgosos			

III.III. Comprensión del Negocio

Tabla III.XI: Matriz de dos entradas para el criterio Entorno del Negocio y demás criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
Entorno del Negocio					Proyecto del Entorno del Negocio			
Entorno del Negocio					Data Science en el Proyecto del Negocio			

Tabla III.XII: Matriz de dos entradas para los criterios Proyecto Entorno del Negocio y Data Science en el Proyecto del Negocio

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
Proyecto del Entorno del Negocio					Data Science en el Proyecto del Negocio			

Características:

Entorno del Negocio

1. Situación económica circundante
2. Evaluación/comprensión general de la organización

3. Definición de los objetivos del negocio
4. Definición del problema del negocio
5. Definición de las oportunidades del negocio
6. Definición de requisitos del negocio
7. Definición de los criterios de éxito del proyecto (desde el negocio)

Proyecto del Entorno del Negocio

8. Conocimiento de los RRHH del negocio
9. Disponibilidad de los RRHH del negocio
10. Riesgos del problema del negocio (respecto al proyecto)
11. Contingencias en relación a los riesgos del problema del negocio (respecto al proyecto)
12. Viabilidad del proyecto

Data Science en el Proyecto del Negocio

13. Objetivos de aplicación de la metodología de DS
14. Conocimiento de los RRHH del proyecto de DS
15. RRHH disponibles del proyecto de DS

Entorno del Negocio

Tabla III.XIII: Matriz de dos entradas para la Situación económica circundante y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C1: Situación económica circundante					C2: Evaluación/comprensión general de la organización			
C1: Situación económica circundante					C3: Definición de los objetivos del negocio			
C1: Situación económica circundante					C4: Definición del problema del negocio			

Extrema importancia: 9	Importancia muy fuerte :7	Importancia fuerte :5	Importancia moderada: 3	Igual importancia: 1	Importancia moderada: 3	Importancia fuerte :5	Importancia muy fuerte :7	Extrema importancia: 9
C1: Situación económica circundante					C5: Definición de las oportunidades del negocio			
C1: Situación económica circundante					C6: Definición de requisitos del negocio			
C1: Situación económica circundante					C7: Definición de los criterios de éxito del proyecto (negocio)			

Tabla III.XIV: Matriz de dos entradas para la Evaluación/comprensión general de la organización y demás sub-criterios

Extrema importancia: 9	Importancia muy fuerte :7	Importancia fuerte :5	Importancia moderada: 3	Igual importancia: 1	Importancia moderada: 3	Importancia fuerte :5	Importancia muy fuerte :7	Extrema importancia: 9
C2: Evaluación/comprensión general de la organización					C3: Definición de los objetivos del negocio			
C2: Evaluación/comprensión general de la organización					C4: Definición del problema del negocio			
C2: Evaluación/comprensión general de la organización					C5: Definición de las oportunidades del negocio			
C2: Evaluación/comprensión general de la organización					C6: Definición de requisitos del negocio			
C2: Evaluación/comprensión general de la organización					C7: Definición de los criterios de éxito del proyecto (negocio)			

Tabla III.XV: Matriz de dos entradas para la Definición de los objetivos del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C3: Definición de los objetivos del negocio					C4: Definición del problema del negocio			
C3: Definición de los objetivos del negocio general de la organización					C5: Definición de las oportunidades del negocio			
C3: Definición de los objetivos del negocio					C6: Definición de requisitos del negocio			
C3: Definición de los objetivos del negocio					C7: Definición de los criterios de éxito del proyecto (negocio)			

Tabla III.XVI: Matriz de dos entradas para la Definición del problema del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C4: Definición del problema del negocio					C5: Definición de las oportunidades del negocio			
C4: Definición del problema del negocio					C6: Definición de requisitos del negocio			
C4: Definición del problema del negocio					C7: Definición de los criterios de éxito del proyecto (negocio)			

Tabla III.XVII: Matriz de dos entradas para la Definición de las oportunidades del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C5: Definición de las oportunidades del negocio					C6: Definición de requisitos del negocio			
C5: Definición de las oportunidades del negocio					C7: Definición de los criterios de éxito del proyecto (negocio)			

Tabla III.XVIII: Matriz de dos entradas para la Definición de requisitos del negocio y de los criterios de éxito del proyecto

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C6: Definición de requisitos del negocio					C7: Definición de los criterios de éxito del proyecto (negocio)			

Proyecto del Entorno del Negocio

Tabla III.XIX: Matriz de dos entradas para Conocimiento de los RRHH del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C8: Conocimiento de los RRHH del negocio					C9: Disponibilidad de los RRHH del negocio			
C8: Conocimiento de los RRHH del negocio					C10: Riesgos del problema del negocio (respecto al proyecto)			

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C8: Conocimiento de los RRHH del negocio					C11: Contingencias en relación a los riesgos del problema del negocio			
C8: Conocimiento de los RRHH del negocio					C12: Viabilidad del proyecto			

Tabla III.XX: Matriz de dos entradas para Disponibilidad de los RRHH del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte :7**	**Importancia fuerte :5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte :5**	**Importancia muy fuerte :7**	**Extrema importancia: 9**
C9: Disponibilidad de los RRHH del negocio					C10: Riesgos del problema del negocio (respecto al proyecto)			
C9: Disponibilidad de los RRHH del negocio					C11: Contingencias en relación a los riesgos del problema del negocio			
C9: Disponibilidad de los RRHH del negocio					C12: Viabilidad del proyecto			

Tabla III.XXI: Matriz de dos entradas para Riesgos del problema del negocio y demás sub-criterios

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C10: Riesgos del problema del negocio (respecto al proyecto)					C11: Contingencias en relación a los riesgos del problema del negocio			

Extrema importancia: 9	Importancia muy fuerte: 7	Importancia fuerte: 5	Importancia moderada: 3	Igual importancia: 1	Importancia moderada: 3	Importancia fuerte: 5	Importancia muy fuerte: 7	Extrema importancia: 9
C10: Riesgos del problema del negocio (respecto al proyecto)					C12: Viabilidad del proyecto			

Tabla III.XXII: Matriz de dos entradas para Contingencias en relación a los riesgos del problema del negocio y Viabilidad del proyecto

Extrema importancia: 9	Importancia muy fuerte: 7	Importancia fuerte: 5	Importancia moderada: 3	Igual importancia: 1	Importancia moderada: 3	Importancia fuerte: 5	Importancia muy fuerte: 7	Extrema importancia: 9
C11: Contingencias en relación a los riesgos del problema del negocio					C12: Viabilidad del proyecto			

Data Science (DS) en el Proyecto del Negocio

Tabla III.XXIII: Matriz de dos entradas para Objetivos de aplicación de la metodología de DS y demás sub-criterios

Extrema importancia: 9	Importancia muy fuerte: 7	Importancia fuerte: 5	Importancia moderada: 3	Igual importancia: 1	Importancia moderada: 3	Importancia fuerte: 5	Importancia muy fuerte: 7	Extrema importancia: 9
C13: Objetivos de aplicación de la metodología de DS					C14: Conocimiento de los RRHH del proyecto de DS			
C13: Objetivos de aplicación de la metodología de DS					C15: RRHH disponibles del proyecto de DS			

Tabla III.XXIV: Matriz de dos entradas para Conocimiento de los RRHH del proyecto de DS y RRHH disponibles del proyecto de DS

Extrema importancia: 9	**Importancia muy fuerte: 7**	**Importancia fuerte: 5**	**Importancia moderada: 3**	**Igual importancia: 1**	**Importancia moderada: 3**	**Importancia fuerte: 5**	**Importancia muy fuerte: 7**	**Extrema importancia: 9**
C14: Conocimiento de los RRHH del proyecto de DS					C15: RRHH disponibles del proyecto de DS			

Anexo IV: Plantilla para definición de los sub-criterios mediante la técnica de la Teoría de la Construcción Personal

A continuación se detalla la plantilla diseñada y utilizada con la técnica de TCP para la evaluación de los sub-criterios del cuarto nivel de la jerarquía propuesta (Fig. 4.1).

Entorno: (Se deberá describir de forma resumida el caso de validación a evaluar).
Dadas los siguientes sub-criterios y sus respectivas etiquetas lingüísticas, enumeradas del 1 al 9, se deberá completar las Tablas IV.I a IV.V, correspondientes a la Parrilla del Entendimiento de los Datos y Comprensión del Negocio respectivamente.

IV.I. Entendimiento de los Datos

Acceso a los Datos

1. Portabilidad (capacidad de ser transferido)
    1. No se analiza
    2. De 0% a 13% de portabilidad
    3. De 14% a 27% de portabilidad
    4. De 28% a 41% de portabilidad
    5. De 42% a 55% de portabilidad
    6. De 56% a 69% de portabilidad
    7. De 70% a 83% de portabilidad
    8. De 84% a 97% de portabilidad
    9. De 98% a 100% de portabilidad
2. Accesibilidad (nivel de accesibilidad/asequibilidad de los datos)
    1. No se analiza
    2. De 0% a 13% de accesibilidad
    3. De 14% a 27% de accesibilidad
    4. De 28% a 41% de accesibilidad
    5. De 42% a 55% de accesibilidad
    6. De 56% a 69% de accesibilidad
    7. De 70% a 83% de accesibilidad
    8. De 84% a 97% de accesibilidad
    9. De 98% a 100% de accesibilidad

3. Diversidad (diferentes tipos de datos)
    1. No se analiza
    2. De 98% a 100% de diversidad
    3. De 84% a 97% de diversidad
    4. De 70% a 83% de diversidad
    5. De 56% a 69% de diversidad
    6. De 42% a 55% de diversidad
    7. De 28% a 41% de diversidad
    8. De 14% a 27% de diversidad
    9. De 0% a 13% de diversidad
4. Diversidad de fuentes de datos
    1. No se analiza
    2. De 98% a 100% de diversidad de las fuentes
    3. De 84% a 97% de diversidad de las fuentes
    4. De 70% a 83% de diversidad de las fuentes
    5. De 56% a 69% de diversidad de las fuentes
    6. De 42% a 55% de diversidad de las fuentes
    7. De 28% a 41% de diversidad de las fuentes
    8. De 14% a 27% de diversidad de las fuentes
    9. De 0% a 13% de diversidad de las fuentes
5. Recursos necesarios (equipamiento)
    1. No se analiza
    2. De 0% a 13% de los recursos necesarios
    3. De 14% a 27% de los recursos necesarios
    4. De 28% a 41% de los recursos necesarios
    5. De 42% a 55% de los recursos necesarios
    6. De 56% a 69% de los recursos necesarios
    7. De 70% a 83% de los recursos necesarios
    8. De 84% a 97% de los recursos necesarios
    9. De 98% a 100% de los recursos necesarios

Tabla IV.I: Parrilla de Acceso a Datos

		CRISP-DM	**MoPro PEI**	**P^{3}TQ**	
1.	No se analiza la portabilidad de los datos				De 98% a 100% de portabilidad de los datos
2.	No se analiza la accesibilidad de los datos				De 98% a 100% de accesibilidad de los datos
3.	No se analiza la diversidad de los datos				De 0% a 13% de diversidad de los datos
4.	No se analiza la diversidad de las fuentes de datos				De 0% a 13% de diversidad de la fuentes de datos
5.	No se analiza la disponibilidad de los recursos necesarios				De 98% a 100% de los recursos necesarios

Uso de los Datos

6. Calidad (condiciones de los datos (valores válidos e inexistencia de valores anómalos))
    1. No se analiza
    2. De 0% a 13% de calidad/fiabilidad
    3. De 14% a 27% de calidad/fiabilidad
    4. De 28% a 41% de calidad/fiabilidad
    5. De 42% a 55% de calidad/fiabilidad
    6. De 56% a 69% de calidad/fiabilidad
    7. De 70% a 83% de calidad/fiabilidad
    8. De 84% a 97% de calidad/fiabilidad
    9. De 98% a 100% de calidad/fiabilidad
7. Completitud (grado de completitud de los datos (cantidad de valores faltantes))
    1. No se analiza
    2. De 0% a 13% de completitud
    3. De 14% a 27% de completitud
    4. De 28% a 41% de completitud
    5. De 42% a 55% de completitud
    6. De 56% a 69% de completitud
    7. De 70% a 83% de completitud

8. De 84% a 97% de completitud
9. De 98% a 100% de completitud

8. Funcionalidad (uso y practicidad de los datos)
    1. No se analiza
    2. De 0% a 13% de funcionalidad
    3. De 14% a 27% de funcionalidad
    4. De 28% a 41% de funcionalidad
    5. De 42% a 55% de funcionalidad
    6. De 56% a 69% de funcionalidad
    7. De 70% a 83% de funcionalidad
    8. De 84% a 97% de funcionalidad
    9. De 98% a 100% de funcionalidad

9. Costos de transformación (demanda de trabajo para transformarlos/adaptarlos para su uso)
    1. No se analiza
    2. De 98% a 100% de costos de transformación
    3. De 84% a 97% de costos de transformación
    4. De 70% a 83% de costos de transformación
    5. De 56% a 69% de costos de transformación
    6. De 42% a 55% de costos de transformación
    7. De 28% a 41% de costos de transformación
    8. De 14% a 27% de costos de transformación
    9. De 0% a 13% de costos de transformación

10. Datos riesgosos (datos sensibles, que no deben ser difundidos)
    1. No se analiza
    2. De 98% a 100% de datos riesgosos
    3. De 84% a 97% de datos riesgosos
    4. De 70% a 83% de datos riesgosos
    5. De 56% a 69% de datos riesgosos
    6. De 42% a 55% de datos riesgosos
    7. De 28% a 41% de datos riesgosos
    8. De 14% a 27% de datos riesgosos
    9. De 0% a 13% de datos riesgosos

Tabla IV.II: Parrilla de Uso de los Datos

		CRISP-DM	**MoPro PEI**	**P^3TQ**	
6.	No se analiza la calidad/fiabilidad de los datos				De 98% a 100% de calidad/fiabilidad de los datos
7.	No se analiza la completitud de los datos				De 98% a 100% de completitud de los datos
8.	No se analiza la funcionalidad de los datos				De 98% a 100% de funcionalidad de los datos
9.	No se analiza el costo de transformación de los datos				De 0% a 13% de costo de transformación de los datos
10.	No se analiza los riesgos de los de los datos				De 0% a 13% de datos riesgosos

IV.II. Comprensión del Negocio

Entorno del Negocio

1. Situación económica circundante
    1. No se analiza la situación circundante
    2. De 0% a 13% de favorabilidad de la situación circundante
    3. De 14% a 27% de favorabilidad de la situación circundante
    4. De 28% a 41% de favorabilidad de la situación circundante
    5. De 42% a 55% de favorabilidad de la situación circundante
    6. De 56% a 69% de favorabilidad de la situación circundante
    7. De 70% a 83% de favorabilidad de la situación circundante
    8. De 84% a 97% de favorabilidad de la situación circundante
    9. De 98% a 100% de favorabilidad de la situación circundante
2. Evaluación/comprensión general de la organización
    10. No se analiza o realiza una evaluación/comprensión general de la organización
    1. De 0% a 13% de evaluación/comprensión
    2. De 14% a 27% de evaluación/comprensión
    3. De 28% a 41% de evaluación/comprensión
    4. De 42% a 55% de evaluación/comprensión

5. De 56% a 69% de evaluación/comprensión
6. De 70% a 83% de evaluación/comprensión
7. De 84% a 97% de evaluación/comprensión
8. De 98% a 100% de evaluación/comprensión

3. Definición de los objetivos del negocio
    1. No se analiza o realiza la definición de los objetivos del negocio
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición
4. Definición del problema del negocio
    1. No se analiza o define el problema del negocio
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición
5. Definición de las oportunidades del negocio
    1. No se analizan o definen las oportunidades del negocio
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición

6. Definición de requisitos del negocio
    1. No se analizan o definen de los requisitos
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición
7. Definición de los criterios de éxito del proyecto (desde el negocio)
    1. No se analiza o definen los criterios de éxitos del proyecto (desde el negocio)
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición

Tabla IV.III: Parrilla del Entorno del Negocio

		CRISP-DM	**MoPro PEI**	**P^{3}TQ**	
1.	No se analiza la situación circundante				De 98% a 100% de favorabilidad de la situación circundante
2.	No se analiza o realiza una evaluación/comprensión general de la organización				De 98% a 100% de evaluación/comprensión general de la organización
3.	No se analiza o realiza la definición de los objetivos del negocio				De 98% a 100% de claridad en la definición de los objetivos del negocio

		CRISP-DM	MoPro PEI	P^{3}TQ	
4.	No se analiza o define el problema del negocio				De 98% a 100% de claridad en la definición del problema del negocio
5.	No se analizan o definen las oportunidades del negocio				De 98% a 100% de claridad en la definición de las oportunidades del negocio
6.	No se analizan o definen de los requisitos				De 98% a 100% de claridad en la definición de los requisitos
7.	No se analiza o definen los criterios de éxitos del proyecto (desde el negocio)				De 98% a 100% de claridad en la definición de los criterios de éxitos del proyecto (desde el negocio)

Proyecto del Entorno del Negocio

8. Conocimiento de los RRHH del negocio
    1. No se analiza el conocimiento de los RRHH del negocio
    2. De 0% a 13% de conocimiento de los RRHH
    3. De 14% a 27% de conocimiento de los RRHH
    4. De 28% a 41% de conocimiento de los RRHH
    5. De 42% a 55% de conocimiento de los RRHH
    6. De 56% a 69% de conocimiento de los RRHH
    7. De 70% a 83% de conocimiento de los RRHH
    8. De 84% a 97% de conocimiento de los RRHH
    9. De 98% a 100% de conocimiento de los RRHH
9. Disponibilidad de los RRHH del negocio
    1. No se analiza la disponibilidad de los RRHH del negocio
    2. De 0% a 13% de disponibilidad de los RRHH
    3. De 14% a 27% de disponibilidad de los RRHH
    4. De 28% a 41% de disponibilidad de los RRHH
    5. De 42% a 55% de disponibilidad de los RRHH
    6. De 56% a 69% de disponibilidad de los RRHH
    7. De 70% a 83% de disponibilidad de los RRHH
    8. De 84% a 97% de disponibilidad de los RRHH

9. De 98% a 100% de disponibilidad de los RRHH

10. Riesgos del problema del negocio (respecto al proyecto)
    1. No se analizan los riesgos del negocio (respecto al proyecto)
    2. De 0% a 13% de riesgo identificados (Insignificantes)
    3. De 14% a 27% de riesgo identificados
    4. De 28% a 41% de riesgo identificados
    5. De 42% a 55% de riesgo identificados
    6. De 56% a 69% de riesgo identificados
    7. De 70% a 83% de riesgo identificados
    8. De 84% a 97% de riesgo identificados
    9. De 98% a 100% de riesgo identificados (Catastróficos, amenazan la supervivencia del proyecto)

11. Contingencias en relación a los riesgos del problema del negocio (respecto al proyecto)
    1. No se analizan contingencias en relación a los riesgos del problema del negocio
    2. De 0% a 13% de contingencias en relación a los riesgos
    3. De 14% a 27% de contingencias en relación a los riesgos
    4. De 28% a 41% de contingencias en relación a los riesgos
    5. De 42% a 55% de contingencias en relación a los riesgos
    6. De 56% a 69% de contingencias en relación a los riesgos
    7. De 70% a 83% de contingencias en relación a los riesgos
    8. De 84% a 97% de contingencias en relación a los riesgos
    9. De 98% a 100% de contingencias en relación a los riesgos

12. Viabilidad del proyecto
    1. No se analiza la viabilidad del proyecto
    2. De 0% a 13% de viabilidad
    3. De 14% a 27% de viabilidad
    4. De 28% a 41% de viabilidad
    5. De 42% a 55% de viabilidad
    6. De 56% a 69% de viabilidad
    7. De 70% a 83% de viabilidad
    8. De 84% a 97% de viabilidad
    9. De 98% a 100% de viabilidad

Tabla IV.IV: Parrilla del Proyecto del Entorno del Negocio

		CRISP-DM	**MoPro PEI**	**P^{3}TQ**	
8.	No se analiza el conocimiento de los RRHH del negocio				De 98% a 100% de conocimiento de los RRHH del negocio
9.	No se analiza la disponibilidad de los RRHH del negocio				De 98% a 100% de disponibilidad de los RRHH del negocio
10.	No se analizan los riesgos del negocio (respecto al proyecto)				De 98% a 100% de riesgos identificados del negocio (amenazan la supervivencia)
11.	No se analizan contingencias en relación a los riesgos del problema del negocio				De 98% a 100% de contingencias en relación a los riesgos del problema del negocio
12.	No se analiza la viabilidad del proyecto				De 98% a 100% de viabilidad del proyecto

Data Science en el Proyecto del Entorno del Negocio

13. Objetivos de aplicación de la metodología de DS
    1. No se analiza o realiza la definición de los objetivos de aplicación de la metodología de DS
    2. De 0% a 13% de claridad en la definición
    3. De 14% a 27% de claridad en la definición
    4. De 28% a 41% de claridad en la definición
    5. De 42% a 55% de claridad en la definición
    6. De 56% a 69% de claridad en la definición
    7. De 70% a 83% de claridad en la definición
    8. De 84% a 97% de claridad en la definición
    9. De 98% a 100% de claridad en la definición
14. Conocimiento de los RRHH del proyecto de DS
    1. No se analiza el conocimiento de los RRHH del proyecto de DS
    2. De 0% a 13% de conocimiento de los RRHH
    3. De 14% a 27% de conocimiento de los RRHH

4. De 28% a 41% de conocimiento de los RRHH
5. De 42% a 55% de conocimiento de los RRHH
6. De 56% a 69% de conocimiento de los RRHH
7. De 70% a 83% de conocimiento de los RRHH
8. De 84% a 97% de conocimiento de los RRHH
9. De 98% a 100% de conocimiento de los RRHH

15. RRHH disponibles del proyecto de DS
    1. No se analiza la disponibilidad de los RRHH del proyecto de DS
    2. De 0% a 13% de disponibilidad de los RRHH
    3. De 14% a 27% de disponibilidad de los RRHH
    4. De 28% a 41% de disponibilidad de los RRHH
    5. De 42% a 55% de disponibilidad de los RRHH
    6. De 56% a 69% de disponibilidad de los RRHH
    7. De 70% a 83% de disponibilidad de los RRHH
    8. De 84% a 97% de disponibilidad de los RRHH
    9. De 98% a 100% de disponibilidad de los RRHH

Tabla IV.V: Parrilla de Data Science en el Proyecto del Entorno del Negocio

		CRISP-DM	**MoPro PEI**	**P^3TQ**	
13.	No se analiza o realiza la definición de los objetivos de aplicación de la metodología de DS				De 98% a 100% de claridad en la definición de los objetivos de aplicación de la metodología de DS
14.	No se analiza el conocimiento de los RRHH del proyecto de DS				De 98% a 100% de conocimiento de los RRHH del proyecto de DS
15.	No se analiza la disponibilidad de los RRHH del proyecto de DS				De 98% a 100% de disponibilidad de los RRHH del proyecto de DS

Anexo V: Correos de aceptación artículos en prensa

A continuación, se copian los correos de aceptación de los artículos enviados a los eventos que no han sido publicados:

V.I. Software Engineering & Knowledge Engineering (SEKE)

SEKE2019 <seke2019@easychair.org>

Dear SEKE2019 Author Karina B. Eckert:

The SEKE2019 program committee is pleased to inform you that your paper, "*Model for comparison and selection of Data Science Methodologies through Multiple Criteria Decision Making*", is accepted as a REGULAR paper of the 2019 International Conference on Software Engineering and Knowledge Engineering (SEKE 2019).

All papers accepted for presentation in the SEKE 2019 conference will be published in the SEKE2019 Proceedings in one volume. Papers submitted to the conference are subject to the same rigorous review and therefore are equal in quality. The page limit of a REGULAR paper is SIX pages, but you can purchase two extra pages so that the absolute page limit is EIGHT pages.
Please follow the author's instructions to include the DOI number at the lower left corner of the first page of your manuscript, so that your paper can be properly indexed by Cross-Ref.
You are invited to give an oral presentation and your presentation time is TWENTY FIVE plus FIVE minutes for Q&A. A computer-controlled projector will be provided for your presentation.
In order to be indexed all papers must be orally presented at the conference. In case you cannot attend the conference in person you may designate a colleague to present your paper or ask the conference to help you find a qualified attendee to present your paper by providing a small honorarium to that person. In case you are willing to present other author's papers please contact the conference.
The acceptance rate of regular papers for SEKE2019 is 39%.
All the information related to paper submission and registration can be found at:
http://ksiresearchorg.ipage.com/seke/seke19.html

In particular, you will find the following information useful:
(a) AUTHOR GUIDELINES/INSTRUCTIONS FOR PROCEEDINGS MANUSCRIPTS
http://ksiresearchorg.ipage.com/seke/seke19author.html
(b) SEKE 2019 CONFERENCE ONLINE REGISTRATION FORM
http://ksiresearchorg.ipage.com/seke/seke19regis.html
(c) KSI ONLINE COPYRIGHT FORM
http://ksiresearchorg.ipage.com/seke/seke19regis.html
(d) FREQUENTLY ASKED QUESTIONS AND ANSWERS
http://ksiresearchorg.ipage.com/seke/seke19faq.html

The reviews of your paper will be attached to this letter. Please note the May 10, 2019 deadline for electronic submission of papers and payment of registration fee. This is a firm, fixed deadline, and late papers may not be included in the proceedings. We are sorry that registration online, by fax or e-mail, not accompanied by full payment, will not be accepted. Papers will be published in the proceedings and included in the program, only when at least one of the authors registers. For authors of multiple papers, each paper must be accompanied by a separate registration with full payment. Just sending the registration form itself does not constitute registration. Some authors may ask their institutions to pay their registration fee. Since processing of institutional payment always takes a long time, please first pay the registration fee yourself by credit card or personal check. Later, when we receive your institution's payment, your personal payment will be refunded to you by check.
Your final paper must be submitted electronically through EasyChair website.
File format is PDF, WITHOUT page header, footer or page number on any page of your paper. The deadline for electronic submission is May 10, 2019.
Submitting a paper electronically will expedite publication of your paper, however you must register by May 10, otherwise your paper may not be included in the proceedings.
A selected number of the regular papers will be invited for submission to IJSEKE, provided that the paper is published in the SEKE2019 proceedings and also presented at the SEKE2019 conference. Invitation decisions will be made on August 1, 2019.

Congratulations, and see you at SEKE 2019!

Prof. Oscar Pereira
Conference Chair of SEKE 2019
Angelo Perkusich
Program Committee Chair of SEKE 2019

---------- REVIEW 1 ----------
PAPER: 103
TITLE: Model for comparison and selection of Data Science Methodologies through Multiple Criteria Decision Making AUTHORS: Karina B. Eckert and Paola V. Britos
Overall evaluation: 3 (strong accept)
---------- Overall evaluation ----------
1. Please check the grammar and spelling before submitting the final version. For example, in the first paragraph of the introduction, you wanted to write 'therefore'; but it is written as 'The afore'.
2. It is recommended to add another paragraph for future research in the Conclusion, where you should mention the limits of your research and your future plan to solve them.
---------- REVIEW 2 ----------
PAPER: 103
TITLE: Model for comparison and selection of Data Science Methodologies through Multiple Criteria Decision Making
AUTHORS: Karina B. Eckert and Paola V. Britos
Overall evaluation: -1 (weak reject)
---------- Overall evaluation ----------
The paper proposes a model for selecting methodologies in data science projects. The

model is based on the so-called multiple criteria decision making. The model consists of a systematic process that a decision maker can use to make the selection.
The proposed approach is reasonable. However, the novelty seems to be limited, both in terms of techical challenges that have to be addrssed, and the main ideas to address the challenges. Also, the experiments are described in a way that is difficult to understand. Some important apsects of the experiments are not given, e.g., sujects, metrics, experimental procedure, i.e., how you have obtained the data presented in the paper, and what are the numbers mean?
---------- REVIEW 3 ----------
PAPER: 103
TITLE: Model for comparison and selection of Data Science Methodologies through Multiple Criteria Decision Making
AUTHORS: Karina B. Eckert and Paola V. Britos
Overall evaluation: 0 (borderline paper)
---------- Overall evaluation ----------
This paper presented a decision making model for the multiple criteria decision making (MCDM) that integrates AHP and FAHP. The model was validated by two cases and data was collected and analyzed. By the conclusion, the authors concluded MoProPEI was the most adequate and complete methodology for the evaluated criteria and sub criteria. The strength of this paper is a model was presented and validated in the case study in AHP and FAHP. The paper is well organized and clearly presented. However, not sure if this paper fits for the domain of this conference.

V.II. 48 Jornadas Argentinas de Informática (JAIIO) - Simposio Argentino de Ciencia de Datos y GRANdes Datos

Dr Carlos Sarraute <charles@grandata.com>

Estimados/as,

Felicidades, su trabajo "Análisis comparativo de metodologías de ciencia de datos utilizando el proceso analítico jerárquico difuso junto a la teoría de la construcción personal" ha sido aceptado para ser presentado en 48º Jornadas Argentinas de Informática que tendrá lugar del 16 al 20 de Septiembre 2019 en Ciudad de Salta.

Por favor tengan en cuenta los comentarios de los revisores a la hora de preparar la versión "camera-ready", la cual deberá ser subida al Sistema antes del día 17 de Julio de 2019.

Gracias por considerar su participación en este evento.

Dr Carlos Sarraute

Grandata Labs

charles@grandata.com

Revisor/a A:

Relevancia:

Aceptable

Presentación:

Bueno

Originalidad:

Bueno

Claridad:

Muy Bueno

Referencias:

Muy Bueno

Longitud:

Muy Bueno

Potencial / Impacto / Valor para I + D:

Aceptable

Evaluación General (obligatorio - es el que se toma como evaluación):

Bueno (recomiendo que se acepte)

Comentarios detallados para los autores del trabajo:

El trabajo presenta un análisis comparativo de tres metodologías estandarizadas de ciencias de datos. Este análisis se realizó mediante las técnicas de proceso analítico difuso y teoría de la construcción personal. El trabajo es consistente y puede leerse bien. A continuación se agregan algunos comentarios al respecto.

- Faltan detalles sobre quienes son los expertos (como los eligieron, cuantos, en qué tipo de tareas se desempeñan, etc)

- Sería interesante contar con una hipótesis de por qué hay mayor diferencia para el caso C1 respecto al C2.

- La frase "Los problemas de DM son generalmente inciertos;" no parece correcta. Mal condicionados? Con cierta variabilidad?.

-Conjuntos difusos (no números)

- Sección 4, no queda claro a que se refiere con "descomposición en características bipolares;" que aparece bastante más adelante en el texto.

-Falta presentacion y cita del término "teoría de la construccion personal".

Cuestiones menores:

-Quedan párrafos muy cortos, a veces con una sola frase, conectar

-Existe(s) diversos métodos...

- Falta extension del acrónimo FAHP

Revisor/a B:

Relevancia:

Aceptable

Presentación:

Bueno

Originalidad:

Muy débil

Claridad:

Aceptable

Referencias:

Bueno

Longitud:

Aceptable

Potencial / Impacto / Valor para I + D:

Débil

Evaluación General (obligatorio - es el que se toma como evaluación):

Límite (está en el límite de lo aceptable)

Comentarios detallados para los autores del trabajo:

En el trabajo realizan un analisis comparativo entre metodologías que se usan en ciencias de datos para definir un procedimiento que ayude de forma estructurada a comparar los criterios a tener en cuenta para tomar decisiones. Deciden comparar las metodologías CRISP-DM, Catalyst y MoProPEI, las cuales indican que son las mas relevantes para la comunidad cientifica. En el presente trabajo, incluso mencionan que MoProPEI es la más reciente y fue propuesta como mejora a las anteriores. Para comparer las metodologias utilizaron el Proceso Analítico Jerárquico Difuso y la técnica de Teoría de la Construcción Personal. Las metodologías fueron evaluadas en dos casos reales. En el procedimiento de comparación, tomaron varias decisiones, las cuales fueron basadas en valoraciones de expertos. Concluyendo que MoProPEI es la metodología mas robusta en los casos analizados. Esta conclusión era bastante de esperar, dado que dicen que es la más reciente y fue propuesta como mejora a las anteriores. El trabajo, tiene como objetivo realizar un análisis comparativo, y para poder realizarlo muchas desiciones tuvieron que ser tomadas. En el trabajo describen las metodologias ultilizadas y para compararlas utilizan dos técnicas y son aplicadas a dos ejemplos. Pero, considero que si el objetivo del trabajo es de analisis comparativo de metodologias ya existentes, es recomendable realizar un análisis más riguroso. Concuerdo, con su comentario final en el que dicen que queda la validación con casos adicionales.

48º Jornadas Argentinas de Informática

AGRANDA 2019 - Simposio Argentino de Ciencia de Datos y GRANdes DAtos

http://sgc.sadio.org.ar/sgc/index.php/48JAIIO/AGRANDA2019/index

V.III. XXV Congreso Argentino de Ciencias de la Computación (CACIC) - XVI Workshop Bases de Datos y Minería de Datos (WBDDM

CACIC2019 <cacic2019@dc.exa.unrc.edu.ar>

Estimado Colega:

Su artículo 12443 - "Data science methodologies selection with hierarchical analytical process and personal construction theory" ha sido aceptado para su exposición y publicación en CACIC2019 (Argentina).

Se requiere la inscripción al menos de 1 autor en el Congreso para su publicación en el CACIC2019.

Mas información en http://dc.exa.unrc.edu.ar/cacic2019

Saludos cordiales,

Comité Organizador

CACIC2019

Evaluación número 1:

Recomendación parcial: Aceptar (Strong Accept)

Calificación global: 8

Conocimiento del tema por el evaluador: 8

Comentarios: En este artículo se estudian 3 metodologías de ciencias de datos (P3TQ CRISP-DM y MoProPEI) para establecer cuál de ellas es la más adecuada en casos reales de aplicación. Se propone un modelo de análisis y se estudian dos casos reales: causas de roturas de autos nuevos cuando se transportan de la fábrica a las concesionarias y causas de la deserción universitaria. El artículo aporta una contribución interesante en el área de Ciencias de Datos. Está en condiciones de ser aceptado.

Evaluación número 2:

Recomendación parcial: Aceptar (Strong Accept)

Calificación global: 8

Conocimiento del tema por el evaluador: 7

Comentarios: Se sugiere ordenar las referencias bibliográficas.