

Proceso analítico jerárquico difuso extendido para análisis comparativo de metodologías de ciencia de datos

Karina B. Eckert¹⁻² and Paola V. Britos³

¹ Universidad Gastón Dachary, Posadas, Misiones, Argentina

² Universidad Nacional de Misiones, Posadas, Misiones, Argentina

³ Laboratorio de Computación Aplicada, Universidad Nacional de Río Negro, El Bolsón, Río Negro, Argentina

karinaeck@gmail.com, pbritos@unrn.edu.ar

Resumen. En entornos cambiantes y altamente competitivos, la correcta toma de decisiones es clave en las organizaciones, quienes basan sus decisiones en gran parte a partir de los datos disponibles. El objetivo del presente trabajo presentar un modelo extendido que involucra el Proceso Analítico Jerárquico Difuso y la técnica de Teoría de la Construcción Personal para realizar un análisis comparativo de las metodologías de Ciencia de Datos CRISP-DM, Catalyst y MoProPEI. Siguiendo los pasos del modelo, se definieron los criterios y sub-criterios a partir de la pericia de los expertos y estudios previos. Las metodologías fueron evaluadas a partir de dos casos de validación reales según el grado de completitud de los criterios y sub-criterios involucrados; donde se pudo identificar a MoProPEI como la metodología más robusta. Demostrando así la utilidad del modelo propuesto.

Palabras claves: Metodologías de Ciencia de datos, Proceso Analítico Jerárquico Difuso, Teoría de la Construcción Personal.

1 Introducción

La Toma de Decisiones (DM, del inglés Decision Making) es una actividad fundamental e inherente a la gestión de proyectos; donde los encargados de tomarlas se encuentran cada vez más presionados por la relevancia que esto implica, en especial en entornos cada vez más competitivos, donde existe una necesidad clara de tomar buenas decisiones [1]. La Toma de Decisiones con Múltiples Criterios (MCDM, Multiple Criteria Decision Making) es un procedimiento que combina el rendimiento de las alternativas de varios criterios y obtiene como resultado una solución de compromiso [2]. Existen diversos métodos de MCDM, estos tienen en común que definen un procedimiento para estructurar y comparar los criterios a tener en cuenta para tomar las decisiones, de tal forma que guían a los encargados de hacerlo a considerar y establecer preferencias de una forma reflexiva y articulada [3], [4]. Muchos de los problemas de MCDM presentan múltiples criterios cuantitativos y/o cualitativos, por esta razón, cuando se trata de estos últimos, los métodos como el popular Proceso Analítico Jerárquico (AHP, Analytic Hierarchy Process) [5], pueden resultar un tanto inefi-

cientes. Para superar estas limitaciones, surgieron los métodos híbridos, tales como el Proceso Analítico Jerárquico Difuso (FAHP, Fuzzy Analytic Hierarchy Process) [6] que combina las bondades de AHP con la lógica difusa [7].

La gran cantidad de datos disponibles a partir de los sistemas informáticos, son de gran utilidad para la correcta DM, por lo que se debe prestar especial atención a la relevancia e implicancia de los mismos dentro de las organizaciones [8], [9]. A causa de ello, la cantidad de compañías que basan sus decisiones en los datos, ha ido creciendo en los últimos años, mejorando el desempeño en términos operativos y financieros [10]. Precisamente el área encargado de ello se denomina Ciencia de Datos o Data Science (DS), que permite tomar decisiones oportunas a partir de los datos disponibles en las organizaciones, incluso de su entorno. Dentro de DS uno de los aspectos importantes a tener en cuenta al momento de encarar un proyecto de esta índole, es seleccionar la metodología que mejor se adapte al mismo. Es por ello que este trabajo tiene como propósito realizar un análisis comparativo de las tres metodologías consideradas como más robustas; donde el contraste exhaustivo de las mismas es realizado con el método FAHP, junto a la Teoría de la Construcción Personal (TCP), que es una técnica para educir conocimiento.

El presente trabajo propone una variación del modelo propuesto por en [11] y [12], para el cual se incorporan conceptos de la lógica difusa, definiendo un modelo que utiliza FAHP y TCP, y sirve para comparar metodologías de DS.

Para lo mencionado, se estructuró el artículo de la siguiente manera: en la Sección 2 se describe DS y se mencionan las metodologías seleccionadas para el estudio. Luego en la Sección 3, se explica FAHP y en la Sección 4 TCP. Posteriormente se describe la propuesta comparativa en la Sección 5, y en la Sección 6, se exponen resultados con casos reales. Para finalizar, en la Sección 7, se presentan las principales conclusiones.

2 Ciencia de Datos

DS es un término derivado de Minería de Datos (Data Mining) o Ingeniería de Explotación de Información (Information Mining), que ha ido transformándose con el correr de los años. El aspecto coincidente entre estas denominaciones es la extracción de conocimiento a partir de datos y el conjunto de principios fundamentales, que apoyan y guían la extracción del mismo, a través de una metodología para tal fin, incluye una variedad de algoritmos, técnicas y herramientas que facilitan el procesamiento detallado y automático de los mismos, permitiendo identificar conocimiento útil imperceptible a simple vista [8], [9], [13].

DS combina la experiencia de áreas como la informática, estadística, matemática, ciencia del comportamiento y análisis predictivo; cuyo propósito final es obtener conocimiento especializado para tomar decisiones oportunas, a partir de los resultados representados como patrones o modelos [9], [14], [15], [16]. Los principios fundamentales de DS proporcionan a los profesionales del área, un marco para tratar sistemáticamente los problemas de extracción de conocimiento de datos; guiado por métodos y metodologías específicos para cada caso [8], [13].

De las metodologías validadas como tal por la comunidad científica, las más relevantes son CRISP-DM [17], [18] y Catalyst (conocida como P³TQ) [18], [19] y recientemente MoProPEI [20] empieza a ser reconocida. Dichas metodologías son las evaluadas en el presente artículo.

3 Proceso Analítico Jerárquico Difuso

Los problemas de DM pueden tener cierta variabilidad; para la evaluación, juicio y decisión, el lenguaje natural se emplea a menudo para articular el pensamiento y las percepciones subjetivas. Los términos pueden no tener un significado claro y correctamente definido; como consecuencia, si se utiliza como una etiqueta para un conjunto, los límites del conjunto al que pertenecen o no los objetos se volverán borrosos (difusos). Para superar este inconveniente, se incluyeron los conjuntos o números difusos, de manera que las variables lingüísticas se expresen de forma adecuada [21].

Los conjuntos difusos o borrosos (fuzzy sets), fueron introducidos en 1965 por Zadeh [22], derivados de la teoría de conjuntos clásicos. Son una representación matemática de la vaguedad del lenguaje ordinario [23], [24]. Se basa en el hecho de que una afirmación no solo puede ser verdadera o falsa, sino que se puede establecer una escala de valores intermedios entre la certeza absoluta o la falsedad [25]. El grupo de objetos está caracterizado por una función de membresía, admitiendo pertenencias valoradas en el intervalo $[0,1]$ a cada objeto; en lugar de $\{0,1\}$, como es el caso de la teoría conjuntista clásica [7], [26], [27].

Para interpretar la imprecisión o vaguedad del pensamiento humano relacionado a un problema, en los conjuntos difusos los términos lingüísticos se representan mediante funciones de membresía, valoradas en intervalos de los números reales [23], [28]. Dicho de otra manera, el encuadre difuso permite modelar el nexo entre las categorías conceptuales del ser humano y los datos, los elementos dentro de las clases pueden solaparse, por lo que no se puede determinar la pertenencia o exclusión absoluta de un elemento a una clase en particular [24].

A partir de los años 80, se empezaron a integrar los conjuntos difusos a AHP; hasta el momento se han desarrollado diversos métodos para manejar matrices de comparación difusa [29], de los que se destacan los siguientes: En 1983 Van Laarhoven y Pedrycz [30], presentaron un método difuso para elegir entre varias alternativas bajo criterios de decisión conflictivos; para ello propusieron el método de los mínimos cuadrados logarítmicos (LLSM, Logarithmic Least Squares Method) para obtener los pesos triangulares difusos de una matriz de comparación borrosa triangular. En 1985 Buckley [6] incorporó a AHP una matriz borrosa, 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 [29], [31]. En 1989, Boender y otros autores [32] presentaron una modificación del método LLSM propuesto por Van Laarhoven y Pedrycz [33]. En 1996 Chang [34] propuso un método de análisis extendido, que deriva ponderaciones nítidas para matrices de comparación difusas [29]. Xu [35] en el 2000, presentó el método de prioridad de los mínimos cuadrados difusos (LSM, Least Squares Method) [29]. En

2001, Csutora y Buckley [36] crearon un método llamado Lambda-Max, que es la versión borrosa directa del método λ_{max} [29]. Mikhailov en 2003 [37] desarrolló un método de programación de preferencias difusas (PPM, Preference Programming Method), que deriva pesos nítidos a partir de matrices de comparación borrosa [29]. Por otra parte, en 2006 Wang, Elhag y Hua [38] presentaron otro método LLSM borroso modificado [29]. Cada método tiene sus fortalezas y debilidades; el método de análisis extendido, propuesto por Chang [34], es uno de los más empleado en diferentes aplicaciones como las expuestas en [39], [40], [41], [42] entre otras; gracias a su simplicidad computacional.

Desde la óptica matemática existe una diversidad de funciones de membresía, de las cuales las más utilizadas son la función Triangular, Trapezoidal o Gaussiana [7], [23]. Habitualmente se emplean los números difusos triangulares, para plasmar la vaguedad de los parámetros presentes en la DM [23]. El FAHP fue definido en tres pasos por Huang y Wu [31]: en primer lugar se usan números difusos triangulares para trasladar las ideas de los expertos en una matriz de reciprocidad positiva; luego se utiliza un método de media geométrica para ponderar los valores parciales para cada alternativa, con la conexión jerárquica establecida y, finalmente, una función de permanencia por lo que se clasificaron las prioridades para cada alternativa [43].

En la Fig. 1 se muestra la función de membresía comparaciones pareadas a partir de la cual se define la escala de conversión difusa triangular utilizada y representada en la Tabla 1.

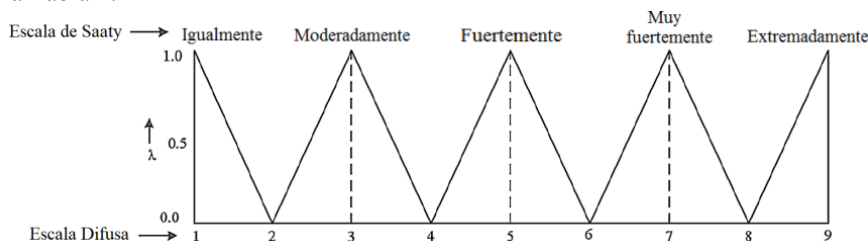


Fig. 1. Función de membresía comparaciones pareadas [44].

Tabla 1. 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) |

4 Teoría de la Construcción Personal

La Teoría de la Construcción Personal (TCP), consiste en una técnica de educación del conocimiento, que ayuda a tomar conciencia de las inconsistencias de las propias escalas de valores, dividida en cinco etapas. Para cada dominio en particular se incluyen los aspectos que los expertos consideren importante, representados como elementos y su descomposición en características bipolares (ejemplo: completo/incompleto, etc.), a partir de los cuales son evaluados; lo que provoca un doble análisis por parte de los decisores, dado que deben resaltar una de las propiedades y a su vez, calcular las distancias desde los dos puntos de vista (completo e incompleto) [45], [46].

5 Desarrollo del modelo y análisis comparativo

Como se mencionó en la sección 1, el artículo se basa en el modelo propuesto en [11] y [12], proponiendo una variación del mismo, para el cual se incorporan conceptos de la lógica difusa mediante método FAHP integrado a la técnica TCP y es utilizado para la selección de metodologías de DS.

A partir de las características de las metodologías seleccionadas, estudios previos y la opinión de expertos del área, se seleccionaron las metodologías de DS para el estudio, mencionadas en el apartado 2. Cabe aclarar, que los expertos elegidos tienen experiencia en diversos proyectos de DS, ofrecen consultorías al respecto y tienen el grado académico de doctorado.

El modelo propuesto para el análisis comparativo de las metodologías de DS, utilizando FAHP y TCP, fue definido en tres de pasos como se indica en la Fig.2 y descritos en los siguientes sub-apartados.



Fig. 2. Pasos generales del modelo propuesto.

5.1 Definición de criterios y construcción de la estructura jerárquica

Utilizando la técnica de TCP se definieron los sub-criterios involucrados en las metodologías, que luego fueron agrupándose por criterios en diferentes niveles, conformando así la jerarquía del problema planteado [11], [12], visible en la Fig. 3.

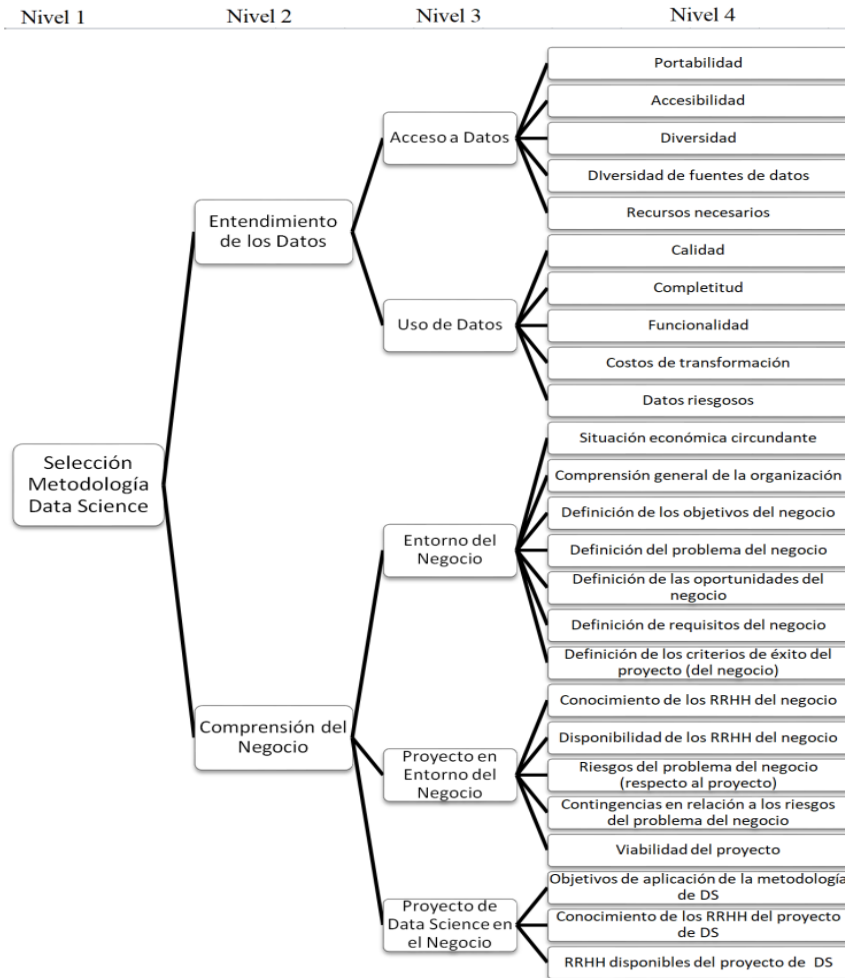


Fig. 3. Estructura jerárquica para la selección de la mejor metodología de DS.

En el primer nivel se define el objetivo propuesto (selección de metodología de DS); en el segundo nivel se definen los dos áreas principales de las metodologías y proyectos de DS, que son el entendimiento de los datos y la comprensión del negocio; estas a su vez en el tercer nivel se dividen en acceso y uso de los datos por un lado (entendimiento de los datos), con cinco sub-criterios cada uno en el nivel 4; y por otro, la comprensión del negocio, en el entorno del negocio, proyecto del entorno del negocio y DS en el proyecto del entorno del negocio del nivel 3, con siete, cinco y

tres sub-criterios respectivamente en nivel 4; todos los sub-criterios del último nivel se comparan en función a cada metodología evaluada.

5.2 Completitud de las preferencias por criterios

Mediante TCP se definió el grado de cumplimiento de los sub-criterio involucrados en las metodologías de DS. Para lo cual se fijó una escala del 1 al 9, donde el valor 1 indica el incumplimiento y no análisis del sub-criterio, del 2 al 9, expresa intervalos de porcentajes según el grado de cumplimiento del sub-criterio dentro del caso de estudio. Dependiendo si el sub-criterio es un aspecto positivo o negativo los intervalos de la escala definida entre 2 y 9 pueden estar invertidos; el ejemplo de la Tabla 2 es considerado un aspecto positivo, por lo que la escala expresada en porcentajes se encuentra de manera creciente.

Tabla 2. Etiquetas lingüísticas para sub-criterios.

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

Tabla 3. Parrilla para evaluar completitud de los sub-criterios.

| | P3TQ | CRISP-DM | MoProPEI | |
|---|-------------|-----------------|-----------------|---|
| No se analiza la calidad/fiabilidad de los datos | 5 | 9 | 7 | De 98% a 100% de calidad/fiabilidad de los datos |
| No se analiza el costo de transformación de los datos | 6 | 4 | 6 | De 0% a 13% de costo de transformación de los datos |
| ... | ... | ... | ... | |

Siguiendo la técnica de TCP, se elaboraron matrices del tipo parrilla, con 5 columnas, en la primera y última (1 y 5), se expone la descripción del peor y mejor caso (valores bipolares) de cada sub-criterio, que concuerda con las etiquetas lingüísticas 1 y 9 (ver ejemplo de la Tabla 2), en las columnas intermedias (2, 3 y 4) corresponde a cada metodología de DS (P³TQ, CRISP-DM y MoProPEI). Los expertos completan las matrices según el valor que considere para cada sub-criterio, teniendo en cuenta la

escala establecida en la Tabla 2. A manera de ejemplo, en la Tabla 3 se exponen dos sub-criterios con sus valores extremos y ponderación asignada para cada metodología.

A partir de las valoraciones de los expertos, se confeccionaron las matrices pareadas de FAHP con función de membresía triangular (ver Fig. 1); para las cuales el valor medio se obtuvo calculando la diferencia en valor absoluto entre las valoraciones de a pares +1, adecuando así los valores a la escala fundamental de Saaty y ubicándolos en el lugar que corresponda dentro de la matriz y completando los valores contiguos según la escala difusa de la Tabla 1.

Adicionalmente, se elaboraron otras matrices de comparaciones de a dúo para facilitar las elecciones de los expertos de los demás criterios de los niveles superiores.

La evaluación de consistencia se estimó con el Coeficiente de Consistencia (CR, Consistency Ratio) propuesto por Saaty [47]; para luego proceder con el desarrollo de FAHP.

5.3 Ponderación de los criterios y elección de la mejor metodología

Siguiendo los pasos del método FAHP [31], [43], [44]: se ponderaron los valores parciales para cada metodología, según la conexión jerárquica establecida. Para lo cual se desarrollaron las métricas de comparación difusa, se calculó el valor de extensión sintética difusa, se seleccionó el valor mínimo de superioridad comparativa; se calculó el vector de pesos y se normalizó para cada criterio y sub-criterio dentro de la jerarquía.

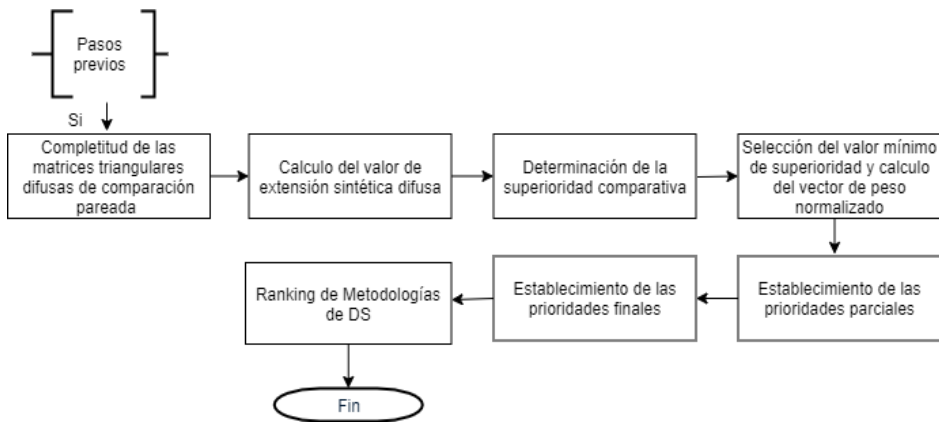


Fig. 4. Ponderación y elección mediante FAHP.

Para finalizar se calculó la función de pertenencia por lo que se clasifican las prioridades para cada alternativa; se estimaron los valores normalizados de los criterios con los valores obtenidos de multiplicar las alternativas y los sub-criterios estimados, a partir de los cuales se calcularon los pesos de prioridades finales para cada metodología, los pasos descriptos se pueden apreciar a través del fragmento de diagrama de flujo de la Fig. 4, los pasos previos refieren a los paso 1 y 2 de la Fig. 2 (descriptos en las sub-secciones anteriores, desde la definición de criterios, construcción de la es-

estructura jerárquica (5.1), hasta la evaluación de consistencia en la completitud de las preferencias por criterios (5.2)).

6 Evaluación de resultados

Para la evaluación de modelo aplicado sobre las metodologías de DS se tomaron dos casos de validación reales, validados con expertos, el primero tiene como propósito determinar los causales de roturas de automóviles 0KM mientras son transportados desde la fábrica hasta las concesionarias (C1) y el segundo, los causales de la deserción universitaria (C2). A continuación, se exponen los resultados principales obtenidos para ambos escenarios, que se corresponde al último paso definido en el modelo (descripto en la sección 5.3).

En las Tablas 4 y 5 se exponen los resultados finales para los sub-criterios del tercer nivel por caso de validación. En la primera tabla (4) se encuentran las ponderaciones obtenidas para el acceso y uso de los datos, los mejores valores son para MoProPEI en ambos escenarios y en segundo lugar varia de un caso al otro, para C1 es P³TQ y para C2 CRISP-DM. En la Tabla 5 se pueden apreciar las valoraciones para los sub-criterios de la comprensión del negocio, entorno del negocio, proyecto del entorno del negocio y proyecto de DS en el negocio, donde nuevamente las preferencias se inclinan hacia la metodología MoProPEI.

Tabla 4. Preferencias en sub-criterios del entendimiento de los datos.

| Caso de validación | Criterio | P ³ TQ | CRISP-DM | MoProPEI |
|--------------------|----------------|-------------------|----------|----------|
| C1 | Acceso a datos | 0,314 | 0,124 | 0,561 |
| | Uso de datos | 0,179 | 0,000 | 0,821 |
| C2 | Acceso a datos | 0,000 | 0,389 | 0,611 |
| | Uso de datos | 0,000 | 0,430 | 0,570 |

Tabla 5. Preferencias en sub-criterios de la comprensión del negocio.

| Caso de validación | Criterio | P ³ TQ | CRISP-DM | MoProPEI |
|--------------------|----------------------------------|-------------------|----------|----------|
| C1 | Entorno del negocio | 0,133 | 0,282 | 0,585 |
| | Proyecto del entorno del negocio | 0,053 | 0,157 | 0,790 |
| | Proyecto de DS en el negocio | 0,077 | 0,356 | 0,567 |
| | Entorno del negocio | 0,000 | 0,500 | 0,500 |
| C2 | Proyecto del entorno del negocio | 0,059 | 0,283 | 0,658 |
| | Proyecto de DS en el negocio | 0,202 | 0,399 | 0,399 |

A partir de las preferencias obtenidas con los sub-criterios de los niveles inferiores y las ponderaciones dentro de cada criterio, se obtuvieron las ponderaciones finales para estos criterios, visibles en la Tabla 6.

Tabla 6. Preferencias en criterios principales.

| Caso de validación | Criterio | P ³ TQ | CRISP-DM | MoProPEI |
|--------------------|----------------------------|-------------------|----------|----------|
| C1 | Entendimiento de los datos | 0,179 | 0,000 | 0,821 |
| | Comprensión del negocio | 0,083 | 0,348 | 0,569 |
| C2 | Entendimiento de los datos | 0,000 | 0,389 | 0,611 |
| | Comprensión del negocio | 0,000 | 0,430 | 0,570 |

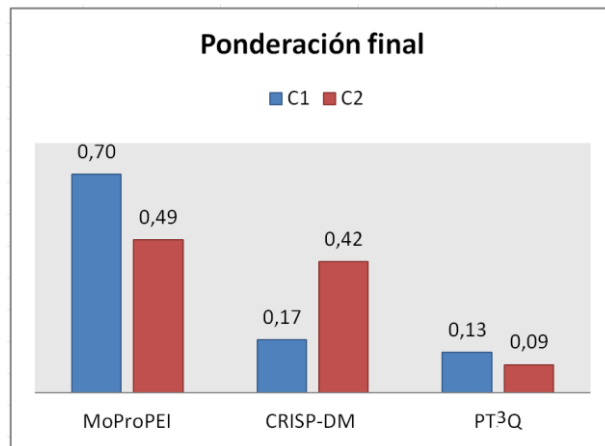


Fig. 5. Preferencias finales para cada caso de validación.

Finalmente, en la Fig. 5 se exponen las preferencias finales obtenidas para cada metodología, donde se puede apreciar que MoProPEI obtuvo una ponderación del 70% para el primer caso de validación (C1) y 49% para el segundo caso (C2), siendo estas las mejores para cada escenario. En segundo lugar se encuentra CRISP-DM con un 17% y 42% y en el último lugar P³TQ con 13% y 9% para C1 y C2 respectivamente. Esto evidencia que dado los criterios y sub-criterios evaluados, la metodología MoProPEI es la más adecuada y completa para determinar los causales de roturas de automóviles 0KM mientras son transportados desde la fábrica hasta las concesionarias y para los causales de deserción universitaria. Para el primer caso, la diferencia entre una metodología y otra fue más notable, como se aprecia en la figura.

Los expertos consultados respaldan los resultados obtenidos y la aplicabilidad del modelo, ratificando así su utilidad.

7 Conclusiones

El artículo tuvo como objetivo presentar el modelo basado en MCDM de que integra al método FAHP y a la técnica de Teoría de la Construcción Personal para realizar un análisis comparativo exhaustivo de metodologías de DS seleccionadas, con el propósito de seleccionar la más adecuada para proyectos de este tipo, para lo cual se definieron y evaluaron dos casos de validación reales, los cuales fueron descriptos en los apartados anteriores.

El modelo permite establecer una estructura jerárquica que esquematiza de forma precisa los criterios y sub-criterios a tener en cuenta al momento de encarar un proyecto de DS. Haciendo uso de las etiquetas lingüísticas y las matrices del tipo parrilla se establecieron los niveles de cumplimiento de los sub-criterios dentro de cada metodología, evitando así ambigüedades. Cabe destacar que si bien, el modelo fue implementado para evaluar comparativamente metodologías de DS, en particular las tres seleccionadas, el modelo resulta lo suficientemente flexible para su utilización con otras metodologías de DS e incluso en otras áreas, ya sea para comparar metodologías (por ejemplo, de desarrollo de software) o bien otros proyectos u aspectos como un método de MCDM.

A partir del desarrollo del trabajo, según la implementación propuesta y el análisis de sus resultados se identificó que la metodología MoProPEI es la más robusta según los criterios evaluados tanto para determinar los causales de roturas de automóviles 0KM mientras son transportados desde la fábrica hasta las concesionarias (C1), como los causales de la deserción universitaria (C2). Más allá de comparar las tres metodologías seleccionadas para el estudio, se buscó validar el modelo propuesto, el cual puede ser utilizado para comparar otras metodologías de DS, e incluso de otras áreas.

Como trabajo futuro se busca desarrollar un software que implemente el modelo propuesto, facilitando así la carga de valores, realización de cálculos y obtención de ponderaciones finales. También se propone la integración con otro método de MCDM y validación en otras áreas.

Referencias

1. Dweiri, F.T., Kablan, M.M.: Using fuzzy decision making for the evaluation of the project management internal efficiency. *Decision Support Systems*. 42, 712–726 (2006). <https://doi.org/10.1016/j.dss.2005.04.001>.
2. Dadda, A., Ouhbi, I.: A decision support system for renewable energy plant projects. Presented at the 2014 International Conference on Next Generation Networks and Services (NGNS), Casablanca, Morocco May 30 (2014).
3. Kolios, A., Mytilinou, V., Lozano-Minguez, E., Salonitis, K.: A Comparative Study of Multiple-Criteria Decision-Making Methods under Stochastic Inputs. *Energies*. 9, 566 (2016). <https://doi.org/10.3390/en9070566>.
4. Kahraman, C., Kaya, İ.: A fuzzy multicriteria methodology for selection among energy alternatives. *Expert Systems with Applications*. 37, 6270–6281 (2010). <https://doi.org/10.1016/j.eswa.2010.02.095>.
5. Saaty, T.L.: *The analytic hierarchy process*. McGraw-Hill, New York (1980).

6. Buckley, J.J.: Fuzzy hierarchical analysis. *Fuzzy Sets and Systems*. 17, 233–247 (1985). [https://doi.org/10.1016/0165-0114\(85\)90090-9](https://doi.org/10.1016/0165-0114(85)90090-9).
7. García Cascales, M. del S.: Métodos para la comparación de alternativas mediante un Sistema de Ayuda a la Decisión (S.A.D.) y “Soft Computing,” (2009).
8. Waller, M.A., Fawcett, S.E.: Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. *Journal of Business Logistics*. 34, 77–84 (2013). <https://doi.org/10.1111/jbl.12010>.
9. Eckert, K., Britos, P.V.: Modelo basado en la toma de decisiones con criterios múltiples para la elección de metodologías de data science. Presented at the XX Workshop de Investigadores en Ciencias de la Computación (2018).
10. McAfee, A., Brynjolfsson, E.: Big data: the management revolution. *Harv Bus Rev*. 90, 60–68 (2012).
11. Eckert, K.B., Britos, P.V.: Data science methodologies selection with hierarchical analytical process and personal construction theory. Presented at the XXV Congreso Argentino de Ciencias de la Computación (CACIC) , Río Cuarto, Córdoba, Argentina October 13 (2019).
12. Eckert, K., Britos, P.V.: Proposed extended analytic hierarchical process for selecting data science methodologies. *Propuesta del Proceso Analítico Jerárquico Extendido para la Selección de Metodologías de Ciencias de Datos*. 21, no. 1, (2021).
13. Provost, F., Fawcett, T.: Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*. 1, 51–59 (2013). <https://doi.org/10.1089/big.2013.1508>.
14. Hazen, B.T., Boone, C.A., Ezell, J.D., Jones-Farmer, L.A.: 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*. 154, 72–80 (2014). <https://doi.org/10.1016/j.ijpe.2014.04.018>.
15. Rodríguez Montequín, M.T., Álvarez Cabal, J.V., Mesa Fernández, J.M., González Valdés, A.: Metodologías para la realización de proyectos de Data Mining. Presented at the VII Congreso Internacional de Ingeniería de Proyectos , Pamplona España (2003).
16. Moine, J.M.: Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo, <http://hdl.handle.net/10915/29582>, (2013).
17. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRISP-DM 1.0: Step-by-Step Data Mining Guide, <http://tinyurl.com/crispdm>, (2000).
18. Britos, P.V.: Procesos de explotación de información basados en sistemas inteligentes, <http://hdl.handle.net/10915/4142>, (2008).
19. Pyle, D.: *Business Modeling and Data Mining*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA (2003).
20. Martins, S., Pesado, P., García Martínez, R.: Propuesta de Modelo de Procesos para una Ingeniería de Explotación de Información: MoProPEI. *Revista Latinoamericana de Ingeniería de Software*. 2, 313–332 (2014).

21. Mardani, A., Jusoh, A., Zavadskas, E.K.: Fuzzy multiple criteria decision-making techniques and applications – Two decades review from 1994 to 2014. *Expert Systems with Applications*. 42, 4126–4148 (2015). <https://doi.org/10.1016/j.eswa.2015.01.003>.
22. Zadeh, L.A.: Fuzzy sets. *Information and control*. 8, 338–353 (1965).
23. Yajure, C.A.: 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*. 20, 255–260 (2015).
24. Díaz Mora, R., Piña, J.G.: Uso de AHP y conjuntos difusos para mejorar la toma de decisiones. Caso: selección de empresas contratistas de construcción en la administración pública venezolana. In: *7th Latin American and Caribbean Conference for Engineering and Technology*. , San Cristóbal, Venezuela (2009).
25. Suárez Rodríguez, A.: Utilización de la lógica difusa en la estimación del riesgo en proyectos. In: *XIII Congreso Internacional de Ingeniería de Proyectos*. pp. 2091–2102. , Badajoz, España (2009).
26. Vera Montenegro, L.: Aplicación y Comparación de Metodologías Multicriterio (AHP y Fuzzy Logic) en la Selección de Tecnologías Postcosecha para Pequeños Productores de Cacao, (2014).
27. Ponce Cruz, P.: *Inteligencia artificial con aplicaciones a la ingeniería*. Marcombo (2011).
28. Javanbarg, M.B., Scawthorn, C., Kiyono, J., Shahbodaghkhan, B.: Fuzzy AHP-based multicriteria decision making systems using particle swarm optimization. *Expert Systems with Applications*. 39, 960–966 (2012). <https://doi.org/10.1016/j.eswa.2011.07.095>.
29. Wang, Y.M., Luo, Y., Hua, Z.: On the extent analysis method for fuzzy AHP and its applications. *European Journal of Operational Research*. 186, 735–747 (2008). <https://doi.org/10.1016/j.ejor.2007.01.050>.
30. van Laarhoven, P.J.M., Pedrycz, W.: A fuzzy extension of Saaty’s priority theory. *Fuzzy Sets and Systems*. 11, 229–241 (1983). [https://doi.org/10.1016/S0165-0114\(83\)80082-7](https://doi.org/10.1016/S0165-0114(83)80082-7).
31. Huang, L.-C., Wu, R.Y.-H.: Applying fuzzy analytic hierarchy process in the managerial talent assessment model – an empirical study in Taiwan’s semiconductor industry. *International Journal of Technology Management*. 30, 105–130 (2005). <https://doi.org/10.1504/IJTM.2005.006347>.
32. Boender, C.G.E., de Graan, J.G., Lootsma, F.A.: Multi-criteria decision analysis with fuzzy pairwise comparisons. *Fuzzy Sets and Systems*. 29, 133–143 (1989). [https://doi.org/10.1016/0165-0114\(89\)90187-5](https://doi.org/10.1016/0165-0114(89)90187-5).
33. Wang, L., Chu, J., Wu, J.: Selection of optimum maintenance strategies based on a fuzzy analytic hierarchy process. *International Journal of Production Economics*. 107, 151–163 (2007). <https://doi.org/10.1016/j.ijpe.2006.08.005>.
34. Chang, D.-Y.: Applications of the extent analysis method on fuzzy AHP. *European Journal of Operational Research*. 95, 649–655 (1996). [https://doi.org/10.1016/0377-2217\(95\)00300-2](https://doi.org/10.1016/0377-2217(95)00300-2).

35. Xu, R.: Fuzzy least-squares priority method in the analytic hierarchy process. *Fuzzy Sets and Systems*. 112, 395–404 (2000). [https://doi.org/10.1016/S0165-0114\(97\)00376-X](https://doi.org/10.1016/S0165-0114(97)00376-X).
36. Csutora, R., Buckley, J.J.: Fuzzy hierarchical analysis: the Lambda-Max method. *Fuzzy Sets and Systems*. 120, 181–195 (2001). [https://doi.org/10.1016/S0165-0114\(99\)00155-4](https://doi.org/10.1016/S0165-0114(99)00155-4).
37. Mikhailov, L.: Deriving priorities from fuzzy pairwise comparison judgements. *Fuzzy sets and systems*. 134, 365–385 (2003).
38. Wang, Y.M., Elhag, T.M.S., Hua, Z.: A modified fuzzy logarithmic least squares method for fuzzy analytic hierarchy process. *Fuzzy Sets and Systems*. 157, 3055–3071 (2006). <https://doi.org/10.1016/j.fss.2006.08.010>.
39. Tüysüz, F., Kahraman, C.: Project risk evaluation using a fuzzy analytic hierarchy process: An application to information technology projects. *International Journal of Intelligent Systems*. 21, 559–584 (2006). <https://doi.org/10.1002/int.20148>.
40. Büyüközkan, G., Kahraman, C., Ruan, D.: A fuzzy multi-criteria decision approach for software development strategy selection. *International Journal of General Systems*. 33, 259–280 (2004). <https://doi.org/10.1080/03081070310001633581>.
41. Gnanavelbabu, A., Arunagiri, P.: Ranking of MUDA using AHP and Fuzzy AHP algorithm. *Materials Today: Proceedings*. 5, 13406–13412 (2018). <https://doi.org/10.1016/j.matpr.2018.02.334>.
42. Chen, J.-F., Hsieh, H.-N., Do, Q.H.: Evaluating teaching performance based on fuzzy AHP and comprehensive evaluation approach. *Applied Soft Computing*. 28, 100–108 (2015). <https://doi.org/10.1016/j.asoc.2014.11.050>.
43. Barajas Díaz, P.A., Cloquell Ballester, V.A., García Alcaraz, J.L., Alvarado Iniesta, A.A.: El Proceso Jerárquico Analítico y Lógica Difusa: Sus Aplicaciones. In: *Congreso Internacional de Investigación*. pp. 249–254. Academia Journals, Celaya Guanajuato México (2012).
44. Tyagi, S., Agrawal, S., Yang, K., Ying, H.: An extended Fuzzy-AHP approach to rank the influences of socialization-externalization-combination-internalization modes on the development phase. *Applied Soft Computing*. 52, 505–518 (2017). <https://doi.org/10.1016/j.asoc.2016.10.017>.
45. Britos, P., Rossi, B., García Martínez, R.: Notas sobre didáctica de las etapas de formalización y análisis de resultados de la técnica de emparillado. Un Ejemplo. In: *Proceedings del V Congreso Internacional de Ingeniería Informática*. pp. 200–209 (1999).
46. García Martínez, R., Britos, P.V.: *Ingeniería de Sistemas Expertos*. Nueva Librería (2004).
47. Saaty, T.L.: How to make a decision: The analytic hierarchy process. *European Journal of Operational Research*. 48, 9–26 (1990). [https://doi.org/10.1016/0377-2217\(90\)90057-I](https://doi.org/10.1016/0377-2217(90)90057-I).