

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

Karina B. Eckert<sup>1</sup>, Paola V. Britos<sup>2</sup>

<sup>1</sup>Facultad de Ciencias Exactas, Químicas y Naturales, Universidad Nacional de Misiones

<sup>2</sup>Universidad Nacional de Río Negro

<sup>1</sup>karinaeck@gmail.com, <sup>2</sup>pbritos@unrn.edu.ar

## RESUMEN

La capacidad de almacenamiento de datos generados por las organizaciones ha aumentado de manera significativa en las últimas décadas y poder analizarlos de manera adecuada, genera un factor estratégico para la Toma de Decisiones (TD). Se puede observar que en los últimos años se ha incrementado la cantidad de profesionales vinculados a la Ciencia de Datos (Data Science) y una de las habilidades requeridas es conocer y manejar a la perfección las metodologías disponibles y determinar cuál se adapta mejor para cada proyecto. Es por ello que al existir una diversidad de metodologías propuestas para el desarrollo de proyectos, en ocasiones la elección no es tarea sencilla, especialmente para los que se inician en el área. La presente investigación tiene como objetivo establecer un modelo basado en la Toma de Decisiones Multicriterios (o criterios múltiples), a través de métodos como el Proceso Analítico Jerárquico y su variante combinado con Lógica Difusa, con el propósito de establecer una base sólida para la selección de metodologías que guíen los proyectos de Minería de Datos o Data Science.

**Palabras clave:** Data Science, Metodologías de Minería de Datos, Toma de Decisiones Multicriterios, Proceso Analítico Jerárquico, Proceso Analítico Jerárquico Difuso.

## CONTEXTO

El presente trabajo de investigación se desarrolla como trabajo final de posgrado de la Maestría en Tecnologías de Información, del tipo interinstitucional convenida entre la Facultad de Ciencias Exactas, Químicas y Naturales (FCEQyN) de la Universidad Nacional de Misiones (UNaM) y a la Facultad de Ciencias Exactas y Naturales y Agrimensura (FaCENA) de la Universidad Nacional del Nordeste (UNNE). El trabajo final de maestría se ha sido iniciado en la FCEQyN de la UNaM a mediados del año 2017 y se encuentra en etapa de desarrollo.

## 1. INTRODUCCIÓN

Con el avance de las Tecnologías de la Información y las Comunicaciones (TICs), los sistemas de almacenamiento poseen cada vez mayor capacidad de procesamiento, alto manejo y producción de datos y bajo costo del mismo; implicando que estos sistemas crezcan hasta límites inesperados. Es por ello, que es de trascendental que las organizaciones posean la capacidad de ser adaptativas y generar acciones a partir del conocimiento de la información almacenada.

En los últimos años han surgido una serie metodologías, procesos y herramientas, basadas en técnicas que facilitan el procesamiento avanzado de datos y permiten realizar un análisis en profundidad de los mismos de

manera automática; lo cual permite detectar la información oculta de los datos que a simple vista no son detectados y utilizarla de manera estratégica; esto es posible mediante la Ingeniería de Explotación de Información, popularmente conocida desde hace unas décadas como Minería de Datos o Data Mining (DM) y en la actualidad como Data Science (DS) [1].

Las técnicas de DM aportan información especializada destinada a favorecer una correcta TD en base a los resultados obtenidos a través de fuentes internas y concretas de las organizaciones, esto es, basarse en experiencias, datos, condiciones, entre otros factores. El resultado de este proceso es un conjunto de modelos o patrones, los cuales serán convertidos en información valiosa para la toma de decisiones. Los proyectos de DM pueden ser llevados a cabo en distintos escenarios, a partir de una situación organizacional o una exploración de los datos disponibles [1], [2], [3].

Al momento de seleccionar cuál de las metodologías disponibles actualmente utilizar para encarar un proyecto de DS, en ocasiones no es tarea sencilla, si bien existen algunos estudios comparativos sobre las mismas como los expuestos en [2], [3], [4], [5], [6], [7]; se propone como alternativa la utilización de la Teoría de la Toma de Decisiones Multicriterio (MCDM, Multiple Criteria Decision Making) [8], como soporte para la TD al momento de seleccionar una metodología de DS (o DM).

Los problemas de TD son procesos complejos en los cuales intervienen múltiples criterios; implican reconocimiento, análisis y evaluación de diversos aspectos; por lo cual es necesario utilizar herramientas que permitan discernir entre estos para obtener una solución que satisfaga en mejor grado la combinación de alternativas posibles [9]. Una de estas herramientas, es el Proceso Analítico Jerárquico (AHP, Analytic Hierarchy Process), creado por Saaty [10]. Los expertos hacen tres tipos generales de juicios para expresar su importancia, preferencia o probabilidad y

utilización, para elegir lo mejor entre las alternativas planteada; basan estos juicios en el conocimiento, en la memoria o en analizar los beneficios, costos y riesgos; sin normas se comparan las alternativas en lugar de calificarlas y las comparaciones deben estar dentro de un rango admisible de consistencia, cabe aclarar que AHP incluye tanto los métodos de calificación, como los de comparación. Por tanto, AHP permite estructurar, medir y sintetizar criterios para toma de decisiones con múltiple criterios [10], [11], [12]. El mismo ha sido aplicado ampliamente en la solución de una gran variedad de problemas, entre los cuales se pueden mencionar los presentados en [13], [14], [15], [16], [17], [18], [19], [20], [21].

Para seleccionar la mejor alternativa, AHP requiere que las comparaciones de los criterios y alternativas estén representadas en un número exacto [10], [11], usando una escala de nueve puntos, la cual representa las preferencias de los encargados de tomar las decisiones entre diferentes alternativas. Si bien la escala es simple y fácil de utilizar, no tiene en cuenta la incertidumbre asociada a los juicios humanos. Tal como afirmó Büyüközkan [22], "los responsables de la toma de decisiones generalmente se sienten mejor presentando sus juicios como un rango, en lugar de dar un valor exacto y fijo y lo anteriores porque él, ella o ellos son incapaces de explicar sus preferencias, dado el carácter difuso de comparación". Buckley [23] incorporó una matriz borrosa (difusa) al método AHP para representar la ambigüedad en las respuestas de las personas involucradas en la TD, lo cual proporciona un análisis para la TD con mayor validez. Huang y Wu [24] han expuesto que, "con la ayuda de la teoría de conjuntos difusos, se resuelven algunos defectos encontrados en el método tradicional de AHP, tales como la aplicación de escalas limitadas para la explicación de las consideraciones de expertos, la correlación entre los atributos para la TD, la imprecisión, la ambigüedad y la incertidumbre de expertos para encontrar los valores de las comparaciones" [24], [25]. Al igual que AHP tradicional,

Proceso Analítico Jerárquico Difuso (FAHP, Fuzzy Analytic Hierarchy Process) ha sido aplicado en diferentes escenarios, como los expuestos en [18], [26], [27], [28], [29].

## 2. LÍNEAS DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN

Los ejes de la investigación están relacionados a la aplicación de la teoría de MCDM, más precisamente AHP y FAHP, las cuales son ampliamente aplicadas para la selección de la mejor alternativa de las existentes, en función a la selección de múltiples criterios, para los cuales se utilizan etiquetas lingüísticas, que permiten evitar ambigüedades en las respuestas de los expertos. El modelo propuesto servirá como soporte en la toma de decisiones a la hora de seleccionar metodologías de Data Science (o DM), siendo este otro de los ejes de la investigación.

## 3. RESULTADOS Y OBJETIVOS

### Objetivo General

Establecer un modelo basado en la MCDM, que sirva como soporte para la TD al momento de seleccionar una metodología de DS.

### Objetivos Específicos

- Estudiar y analizar métodos de MCDM, como ser el AHP y FAHP.
- Revisar el estado del arte de las metodologías de DS (o DM) referentes en la actualidad.
- Definir criterios utilizando etiquetas lingüísticas para comparar de las metodologías de DS.
- Implementar el MCDM integrado a las etiquetas lingüísticas del caso de aplicación, evaluar y validar los resultados obtenidos con expertos del área.

## 4. FORMACIÓN DE RECURSOS HUMANOS

La línea de investigación está inmersa en el desarrollo de un trabajo final de la Maestría en Tecnología de la Información dictada en la FCEQyN de la UNaM a cargo de la Ing. Karina B. Eckert y dirigida por la Dra. Paola V. Britos. Se prevé la incorporación de un alumno de grado como parte del equipo de trabajo, que pueda realizar su trabajo final de carrera dentro de la presente línea de investigación.

## 5. REFERENCIAS

- [1] K. B. Eckert and R. Suénaga, “Determinación del perfil académico a través del proceso de KDD”, ed. Editorial Académica Española, 2016.
- [2] J. M. Moine, “Metodologías para el descubrimiento de conocimiento en bases de datos: un estudio comparativo”, Tesis de Maestría, Facultad de Informática Universidad Nacional de La Plata, La Plata, Argentina, 2013.
- [3] M. T. Rodríguez Montequín, J. V. Álvarez Cabal, J. M. Mesa Fernández and A. González Valdés, “Metodologías para la realización de proyectos de Data Mining”, in VII Congreso Internacional de Ingeniería de Proyectos, pp. 257-265, 2005.
- [4] J. M. Moine, S. Gordillo and A. S. Haedo, “Análisis comparativo de metodologías para la gestión de proyectos de minería de datos”, in XVII Congreso Argentino de Ciencias de la Computación (CACIC), pp. 931-938, 2011.
- [5] A. Azevedo and M. F. Santos, “KDD, SEMMA and CRISP-DM: A parallel overview”, in IADIS European Conference on Data Mining, Amsterdam, 2008.
- [6] J. Giraldo Mejía and J., 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, vol. 1, no. 2, pp. 42-44, 2013.
- [7] P. Britos, “Procesos de explotación de información basados en sistemas inteligentes”, Tesis de Doctorado, Facultad

- de Informática Universidad Nacional de La Plata, La Plata, Argentina, 2008.
- [8] B. Roy Kolios, and R. Słowiński, “Questions guiding the choice of a multicriteria decision aiding method”, *EURO J Decis Process*, vol. 1, pp. 69-97, 2013.
- [9] M. Karanik, S. Gramajo, L. Wanderer, M. Giménez and D. Carpintero, “Multi-Criteria Decision Model based on AHP and Linguistic Information”, *Journal of Computer Science & Technology (JCS&T)*, vol. 14, no. 1, pp. 16-24, 2014.
- [10] T. L. Saaty, “How to make a decision: The analytic hierarchy process”, *European Journal of Operational Research*, vol. 48, no. 1, pp. 9-26, September 1990.
- [11] T. L. Saaty, “Decision making with the analytic hierarchy process”, *Int. J. Services Sciences*, vol. 1, no. 1, 2008.
- [12] V. Subramaniam and K. Lee, “Dynamic selection of dispatching rules for job shop scheduling”, *Production Planning and control*, vol. 11, no. 1, pp. 73-81, 2000.
- [13] J. C. Osorio Gómez and J. P. Orejuela Cabrera, “El proceso de análisis jerárquico (AHP) y la toma de decisiones multicriterio. Ejemplo de aplicación”, *Scientia et technica*, vol. 14, no. 39, 2008.
- [14] L. M. Ramírez, J. M. Flores and A. T. Vargas, “Análisis de Toma de Decisiones en Incubadoras Empresariales Mexicanas: un Modelo Basado en Jerarquización Analítica”, *CIAIQ 2017*, vol. 3, 2017.
- [15] M. A. Asma Bahurmoz, “The Analytic Hierarchy Process at Dar Al-Hekma, Saudi Arabia”, *Interfaces*, vol. 33, no. 4, pp. 70–78, 2003.
- [16] A. Başçetin, “An application of the analytic hierarchy process in equipment selection at Orhaneli open pit coal mine”, *Mining Technology*, vol. 113, no. 3, pp. 192 – 199, 2013.
- [17] D. Mendoza Casseres, “Aplicación de la teoría de decisión multicriterio discreta para ponderar factores en procesos de acreditación de alta calidad”, *Inge Cuc*, vol. 9, no. 1, pp. 25-41, 2013.
- [18] L. Vera Montenegro, “Aplicación y comparación de metodologías multicriterio (AHP y Fuzzi Logic) en la selección de tecnología postcosecha para pequeños productores de cacao”, Tesis de Doctorado, Universidad Politécnica de Valencia, Valencia, España, 2014.
- [19] E. Martínez Rodríguez, “Aplicación del proceso jerárquico de análisis en la selección de la localización de una PYME”, *Anuario Jurídico y Económico Escurialense*, pp. 523-542, 2007.
- [20] M. Lněnička, “AHP Model for the Big Data Analytics Platform Selection”, *Informatica Pragensia*, vol. 4, no. 2, pp. 108-121, 2015.
- [21] J. Wang, C. Li and S. Zhang, “Big Data Impact Analysis of Smart Grid based on AHP method”, *International Conference on Manufacturing Science and Engineering (ICMSE 2015)*, Atlantis Press, pp. 1487-1490, 2015.
- [22] G. Büyüközkan, C. Kabraman and D. Ruan, “A fuzzymulti-criteriadecision approach for software development strategy selection”, *Intern. Journal of General Systems*, vol. 33, no. 2, pp. 259-280, 2004.
- [23] J. J. Buckley, “Fuzzy Hierarchical Analysis”, *Fuzzy sets and systems*, vol.17, no.3, pp.233-247, 1985.
- [24] L. Huang and R. 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*, pp. 105-130, 2005.
- [25] G. Herrera Enríquez, F. Guevara Viejó, S. Castillo Paez and D. Zambrano 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, no. 16, pp. 45-66. 2016.

- [26] P. A. Barajas Díaz, V. A. Cloquell Ballester, J. Luisarcía Alcaraz and A. Alvarado Iniesta, “El Proceso Jerárquico Analítico y Lógica Difusa: Sus Aplicaciones”, in Congreso Internacional de investigación - Celaya, Gto. Academia Journals, vol. 4, no.3, 249-254, 2012.
- [27] 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 Año XX*, vol. 20, no. 3, pp.255-260, 2015.
- [28] R. Díaz, J. G. Piña, D. B. Ríos and M. P. Serafin, “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 Seventh LACCEI Latin American and Caribbean Conference for Engineering and Technology (LACCEI’2009) - Energy and Technology for the Americas: Education, Innovation, Technology and Practice, 2009.