

# Minería de datos para la toma de decisiones en la unidad de nivelación y admisión universitaria ecuatoriana

Data mining for decision-making in the ecuadorian  
university leveling and admission unit

**María Uvidia Fassler**

Universidad Nacional de Chimborazo (Ecuador)  
Escuela Superior Politécnica de Chimborazo  
muvidia@unach.edu.ec

**Andrés Cisneros Barahona**

Universidad Nacional de Chimborazo (Ecuador)  
Escuela Superior Politécnica de Chimborazo  
(Ecuador)

**Pablo Méndez Naranjo**

Universidad Nacional de Chimborazo (Ecuador)  
Escuela Superior Politécnica de Chimborazo  
(Ecuador)

**Henry Villa Yáñez**

Universidad Nacional de Chimborazo (Ecuador)  
Escuela Superior Politécnica de Chimborazo  
(Ecuador)

**Revista Cumbres Vol.4 N°2**

Versión impresa ISSN 1390-9541

Versión electrónica ISSN 1390-3365

<http://investigacion.utmachala.edu.ec/revistas/index.php/Cumbres>

## RESUMEN

Data Mining o minería de datos (DM) fue aplicado en este trabajo de investigación, donde a partir de la selección de algoritmos y análisis de información, se pudieron obtener patrones que una vez observados y examinados se convirtieron en conocimiento para la toma de decisiones en la Unidad de Nivelación y Admisión de la Escuela Superior Politécnica de Chimborazo (ESPOCH). Dichos datos fueron generados desde el año 2012 en la ESPOCH en estricto cumplimiento a la normativa vigente a la fecha y dicho proceso generó una importante cantidad de información sin procesar y que no ha aportado a la toma de decisiones. Mediante la aplicación de minería de datos en Waikato Environment for Knowledge Analysis (WEKA) se pudieron analizar algoritmos de predicción de clasificación (árboles de decisión y redes neuronales) y regresión (regresión lineal y optimización de secuencia mínima) permitiendo el conocimiento de la realidad, cuyo objetivo de la investigación fue generar conocimiento de tendencias de postulaciones por área, género y años, hasta llegar a conocer predicciones. Finalmente, mediante el análisis de parámetros estadísticos se determinaron los mejores algoritmos que aseguraron la confiabilidad en la información y generaron conocimiento para la toma de decisiones académica, siendo éstas: redes Bayesianas y optimización de secuencia mínima.

**Palabras clave:** Minería de datos, WEKA, Universidad Ecuatoriana

## ABSTRACT

Data Mining (DM) was applied in this research work, where from the selection of algorithms and information analysis, it was possible to obtain patterns that once observed and examined became knowledge for decision making in the Leveling and Admission Unit of the Higher Polytechnic School of Chimborazo (ESPOCH). These data were generated since 2012 ESPOCH in strict compliance with the regulations in force to date, adopted the policies issued by the SNNA, this process generated an important amount of information without processing and has not contributed to the decision making. Through the application of data mining in Waikato Environment for Knowledge Analysis (WEKA) it was possible to analyze Classification Prediction algorithms (Decision Trees and Neural Networks) and Regression (Linear Regression and Minimum Sequence Optimization) allowing the knowledge of reality, whose objective of the research was to generate knowledge of trends of applications by area, gender and years, until getting to know predictions. Finally, through the analysis of statistical parameters, the best algorithms were determined that ensured the reliability in the information and generated knowledge for the academic decision-making, being these: Bayesian Networks and Minimum Sequence Optimization.

**Keywords:** Data Mining, WEKA, Ecuadorian University.

## INTRODUCCIÓN

Los datos almacenados son un tesoro para las organizaciones, es en donde se guardan las interacciones pasadas con los clientes, la contabilidad de sus procesos internos, además representan la memoria de la organización. Tener memoria no es suficiente, hay que pasar a la acción inteligente sobre los datos para extraer la información que almacenan, este es el objetivo de la minería de datos (Aluja, 2001).

La Ley Orgánica de Educación Superior (LOES) en Art. 81 contempla que el ingreso a las instituciones de educación superior públicas está regulado a través del Sistema Nacional de Nivelación y Admisión (SNNA), para lo cual la ESPOCH implementa estos procesos a nivel institucional. La gran cantidad de datos creados desde el año 2012 por la Unidad de Nivelación y Admisión (UNAE) de la ESPOCH, no han sido procesados mediante el análisis de técnicas de DM, para lo cual se generó información para la toma de decisiones en el tiempo y momento precisos, considerando según Aluja (2001), que no existe la técnica más inteligente, sino formas inteligentes de utilizar una técnica y que cada uno utiliza de forma inteligente aquello que conoce.

Según Hernández, Ferrari y Ramírez (2004), DM implica un proceso de descubrimiento de nuevas y significativas relaciones, patrones y tendencias que son observadas al examinar grandes cantidades de información. La minería de datos es el núcleo del proceso KDD (Descubrimiento de Conocimiento en Base de datos, Knowledge Discovery in Data Bases) que permitirá identificar patrones válidos, útiles y comprensibles, además estructuras que conviertan los datos en información útil para alcanzar un trabajo organizado y permitir la implementación de acciones pertinentes (Juan, Moine, Gordillo, Ana, & Haedo, 2011). Dentro de KDD hay dos objetivos principales en la minería de datos: la predicción y la descripción. Predicción se refiere a la supervisión de minería de datos (técnicas probadas), mientras que descriptiva de minería de datos incluye los aspectos sin supervisión (técnicas no probadas y validadas) y visualización de minería de datos (Bernabeu, 2010).

“DM se refiere al acto de extraer patrones o modelos a partir de los datos” (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). Así el valor real de los datos reside en la información que se puede extraer de ellos, información que ayude a tomar decisiones o mejorar nuestra comprensión de los fenómenos que nos rodean. Hoy, más que nunca, los métodos analíticos avanzados son el arma secreta de muchos negocios exitosos (Asencios, 2004).

### Data Mining

DM se define como aquel proceso que consiste en extraer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos. En ese sentido, la tarea fundamental de la minería de datos es encontrar modelos inteligibles a partir de los datos recogidos (Hernández, Ferrari & Ramírez, 2004)

Según Han y Kamber (2006) DM se refiere a la extracción de conocimiento o minería de grandes cantidades de datos. Ahora bien, al vincular las ideas de datos y extracción, minería de datos se ha convertido en una opción más popular.

Según Gonzalez (2005), DM es el proceso por el cual se genera un modelo útil para la predicción. Dicho modelo se construye teniendo como fundamento los datos que se encuentran en una base de datos, a los cuales se le ha aplicado algún algoritmo justamente con el fin de plantear un modelo.

En conclusión, se podría decir que DM es un proceso que integra los datos de diferentes fuentes para, posteriormente, extraer un importante conocimiento, es decir, identificar información trascendente, valiosa y útil, a partir de lo cual las instituciones van a poder tomar alguna significativa decisión (Daza, 2016).

Una variedad de modelos informáticos analíticos se ha utilizado en DM. Los tipos de modelo estándar en la minería de datos incluyen la regresión (regresión normal, para la predicción, la regresión logística para la clasificación), redes neuronales, y los árboles de decisión, siendo estas técnicas las más conocidas (Olson & Delen, 2008).

De acuerdo a Maimon y Rokach (2010), DM está clasificado como se muestra en la Figura 1.

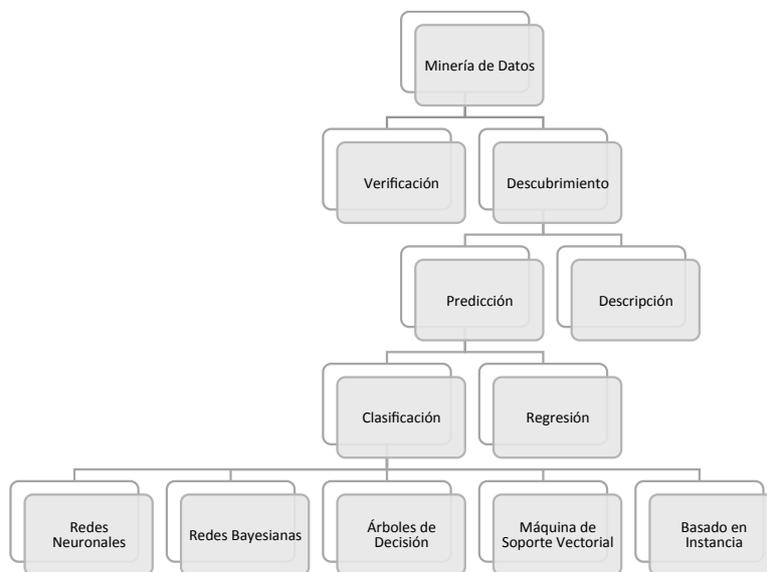


Figura 1. Clasificación de técnicas DM  
 Fuente: Data Mining and Knowledge Discovery Handbook, (2010)

Implementar DM permite analizar factores de influencia en determinados procesos, predecir o estimar variables o comportamientos futuros, segmentar o agrupar ítems similares, además de obtener secuencias de eventos que provocan comportamientos específicos, siendo su principal ventaja inferir en comportamientos, modelos, relaciones y estimaciones de los datos, para poder desarrollar predicciones sobre los mismos, sin la necesidad de contar con patrones o reglas preestablecidas, permitiendo tomar decisiones proactivas y basadas en un conocimiento acabado de la información. Es muy importante tener en cuenta que en las herramientas OLAP (On Line Analytic Processing) y en los reportes y consultas, el análisis parte de una pregunta o hipótesis generada por los usuarios, en cambio DM permite generar estas hipótesis.

Los sistemas DM se desarrollan bajo lenguajes de última generación basados en Inteligencia Artificial y utilizan métodos matemáticos tales como:

- Redes Neuronales.
- Sistemas Expertos.
- Programación Genética.
- Árboles de Decisión.

### **Clasificación de algoritmos de DM**

#### **Árboles de Decisión**

Son estructuras de forma de árbol que representan conjuntos de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos, las cuales explican el comportamiento de una variable con relación a otras, y pueden traducirse fácilmente en reglas de negocio. Son utilizados con finalidad predictiva y de clasificación. Por ejemplo, los árboles de decisión pueden emplearse para:

- Optimizar respuestas de campañas.
- Identificar clientes potenciales.
- Realizar evaluación de riesgos.

#### **Redes bayesianas**

Según Aluja (2001), consiste en representar todos los posibles sucesos en los que se está interesado mediante un grafo de probabilidades condicionales de transición entre sucesos. Puede codificarse a partir del conocimiento de un experto o puede ser inferido a partir de los datos. Permite establecer relaciones causales y efectuar predicciones.

#### **Regresión Lineal**

Técnica estadística para determinar la relación entre variables. Permite predecir a partir de un muestreo de datos aleatorio. Se adapta a una amplia variedad de situaciones. La regresión ajustada con el error cuadrático medio más bajo se elige como el modelo final (González & García, 2010).

Al aplicar el análisis de funciones automáticamente se genera un modelo de regresión lineal de predicción. La precisión del modelo generado depende en gran manera de la cantidad de datos que se manejen, así, la exactitud de la predicción es directamente proporcional al número de datos disponibles (Lewandowski, 2015).

#### **Optimización de Secuencia Mínima (SMOreg)**

Es el algoritmo de optimización de secuencia mínima desarrollado por John C. Esta implementación sustituye todos los valores que faltan y transforma atributos nominales en otros binarios. También normaliza todos los atributos por defecto (los coeficientes a la salida están basados en los datos normalizados, no en los originales) (González & García, 2010).

### Conocimiento

De acuerdo a la definición de Harlan Cleveland (1985) el conocimiento es la “información organizada, interiorizada por uno, integrada con todo lo que se conoce, desde la experiencia, el estudio o la intuición y por lo tanto es útil para guiar la vida y el trabajo”, mientras que para Alejandro Pavez (2000), este es “las creencias cognitivas, confirmadas, experimentadas y contextualizadas del conocedor sobre el objeto a conocer, las cuales estarán condicionadas por el entorno, y serán potenciadas y sistematizadas por las capacidades de dicho conocedor, las que establecen las bases para la acción objetiva y la generación de valor”.

Acorde a estas definiciones, el conocimiento con una base en la experiencia, el entorno e información permite que sea el apoyo a la toma de decisiones, además el elemento más importante que una vez adquirido permite repotencializar, planificar y generar acciones de perfeccionamiento del área donde se aplique. Lo trascendental del conocimiento es que este sea real, es decir, obtenido de fuentes válidas y consistentes que aseguren un conocimiento adecuado.

### MATERIALES Y MÉTODOS

El diseño del trabajo de investigación y desarrollo fue del tipo cuasi experimental. Se escogieron los algoritmos de DM y se analizó el conocimiento generado para la toma de decisiones

Para el desarrollo del proceso de DM se utilizó como fuente de información un Data Warehouse (DW) y WEKA como se muestra en la Tabla 1.

HERRAMIENTA TECNOLÓGICA	NOMBRE
Sistema de Gestión de Base de Datos - Data Warehouse (Fuente de datos)	PostgreSQL 9.3
Data Mining	WEKA Developer 3.7.13

Tabla 1. Herramientas tecnológicas utilizadas

### RESULTADOS Y DISCUSIÓN

En la UNAE fue importante conocer su realidad, por tal motivo se consideraron dos escenarios. Primero se conoció la realidad de las postulaciones de los estudiantes que obtienen un cupo para la ESPOCH y sus reglas del negocio para tomar decisiones, siendo el escenario A para los algoritmos de clasificación: Árboles de Decisión y Redes Bayesianas. Y el escenario B para conocer la predicción de la cantidad de estudiantes que serán parte de la institución en los próximos años, es por eso se analizaron algoritmos de regresión: Regresión Lineal y Optimización de Secuencia Mínima, como se muestra en la Figura 2, utilizando la herramienta WEKA Developer 3.7.13.

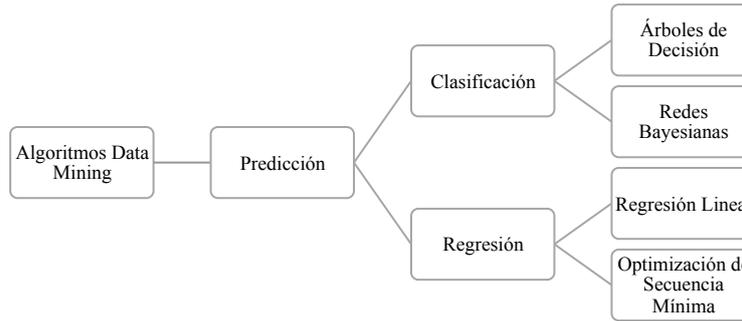


Figura 2. Elección del algoritmo de DM para la UNAE - ESPOCH

Además, para el análisis de los patrones obtenidos como resultado de cada uno de los algoritmos, se consideraron los siguientes indicadores:

- Instancias correctamente clasificadas
- Kappa: Índice que compara la coincidencia entre varios expertos con el nivel de coincidencia que se pondría por casualidad. Resultados:
- +1: total coincidencia
- Valores de 0: No más coincidencia que la que puede esperarse por casualidad
- -1: total desacuerdo
- Error absoluto: Diferencia entre los valores obtenidos frente al valor exacto.
- Matriz de Confusión

### ESCENARIO A

**Algoritmo:** CLASIFICACIÓN

**Algoritmos de Clasificación:** Árboles de Decisión y Redes Bayesianas

**Datos analizados:** Estudiantes por año, área de estudio de las carreras y género

**Atributos:** Año, Área, Género (Nombres de las columnas, en la base de datos se utilizaron nombres sin caracteres especiales debido a que el idioma base es el inglés)

Resultado mostrado en la Figura 3, que representan las áreas de conocimiento de las carreras en articulación con la clasificación CINE Unesco y la cantidad de estudiantes clasificados por género.

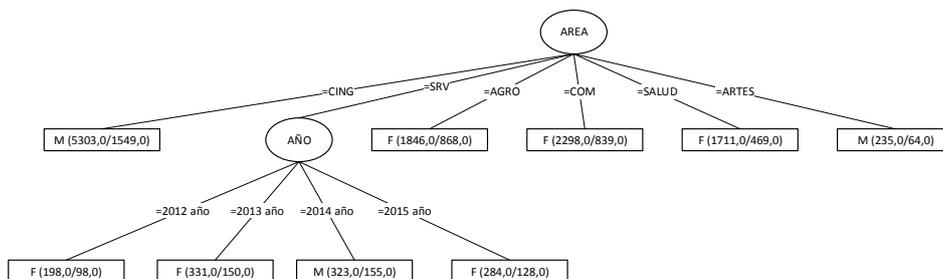


Figura 3. Resultado del algoritmo Árbol de Decisión

**Donde:**

CING: Ciencias e Ingeniería; SRV: Servicios; AGRO: Agricultura; COM: Comercial; SALUD: Salud; ARTES: Artes y Humanidades; M: Masculino; F: Femenino.

Las hojas del árbol de decisión muestran los focos de información que permiten conocer la distribución de las áreas de los cursos de nivelación con la cantidad de estudiantes tanto del género femenino y masculino, permitiendo conocer que género representa mayor cantidad.

Al aplicar la técnica de redes bayesianas con los mismos datos, se puede observar la información mediante grafos, siendo más complicado obtener la información, ya que se debe ingresar a cada nodo.

**Análisis de patrones:**

En la Tabla II se muestra el análisis de patrones del escenario A, que es la información que se obtuvo directamente de la ejecución de WEKA (el idioma base es el inglés). Para realizar la comparación de las dos técnicas, se resaltan los valores más importantes.

RESULTADOS DE ÁRBOLES DE DECISIÓN		RESULTADOS DE REDES BAYESIANAS	
Scheme: weka.classifiers.trees.J48 -C 0,25 -M 2		Scheme: weka.classifiers.bayes.BayesNet -D -Q weka.classifiers.bayes.net.search.local.K2 --P 1 -S BAYES	
Correctly Classified Instances	8160	Correctly Classified Instances	8194
	65,1289 %		65,4003 %
Incorrectly Classified Instances	4369	Incorrectly Classified Instances	4335
	34.8711 %		34.5997 %
Kappa statistic	0,3058	Kappa statistic	0,3118
Mean absolute error	0,4408	Mean absolute error	0.4402
Root mean squared error	0,4697	Root mean squared error	0.4695
Relative absolute error	88,4776 %	Relative absolute error	88,3741 %
Root relative squared error %	94,1072	Root relative squared error	94,071 %
Coverage of cases (0.95 level)	100 %	Coverage of cases (0.95 level)	100 %
Mean rel. region size (0.95 level)	100 %	Mean rel. region size (0.95 level)	100 %
Total Number of Instances	12529	=== Confusion Matrix ===	
		a b <-- classified as3	
=== Confusion Matrix ===		4002 2643   a = M	
	a b <-- classified as	1692 4192   b = F	
	4028 2617   a = M		
	1752 4132   b = F		

Tabla 2. Análisis de patrones del escenario A

Al realizar una comparación entre algoritmos empleados se deduce que:

- La mayor cantidad de instancias correctamente clasificadas del total de 12529 registros, lo realiza el algoritmo de Redes Bayesianas con un total de 8194, que representa el 65,4003%.
- El índice Kappa del algoritmo Redes Bayesianas se acerca más a 1, es decir, con 0,3118 es el que más coincidencia entre variables presenta, siendo este un punto aceptable.
- El error absoluto es de 0,4402 en las Redes Bayesianas que mientras menor sea es mejor ya que existe menos diferencia con el promedio exacto.
- Las diagonales de la Matriz de Confusión de las Redes Bayesianas y Árboles de Decisión, tienen casi el mismo equilibrio, por lo que en ambos casos son valores aceptables.

**Resultado 1:** El algoritmo de Clasificación - Redes Bayesianas, es el que mejor se acopla al manejo de información de la UNAE - ESPOCH, aunque Árboles de Decisión tiene valores muy cercanos al otro algoritmo.

### ESCENARIO B

**Algoritmo:** REGRESIÓN

**Algoritmos de Regresión:** Regresión Lineal y Optimización de Secuencia Mínima

**Datos analizados:** Estudiantes por año

**Atributos:** Año, Estudiantes

Resultado mostrado en la Figura 4 y Figura 5.

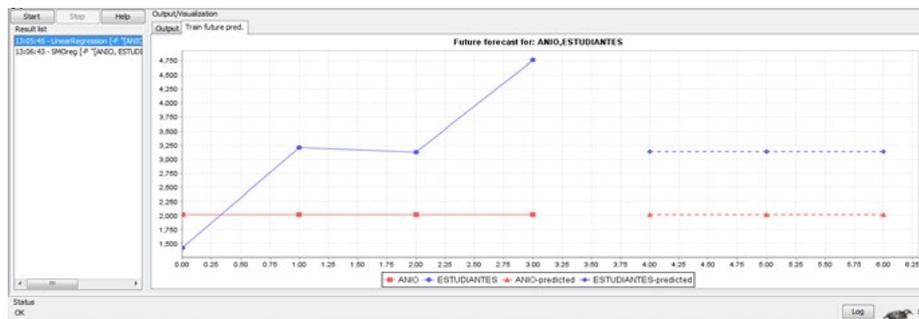


Figura 4. Resultado del algoritmo Regresión Lineal

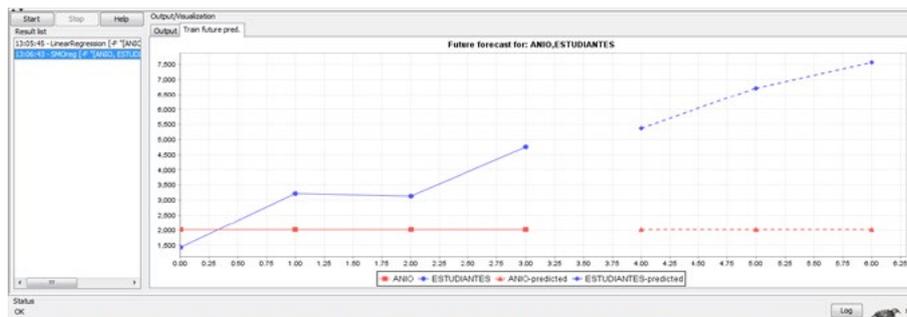


Figura 5. Resultado del algoritmo Optimización de Secuencia Mínima

**Análisis de patrones:**

En la Tabla III se muestra el análisis de patrones del escenario B, que es la información que se obtuvo directamente de la ejecución de WEKA (el idioma base es el inglés). Para realizar la comparación de las dos técnicas, se resaltan los valores más importantes.

RESULTADOS DE REGRESIÓN LINEAL			RESULTADOS DE OPTIMIZACIÓN DE SECUENCIA MÍNIMA		
Scheme: LinearRegression -S 0 -R 1.0E-8 -num-decimal-places 4			Scheme: SMOreg -C 1.0 -N 0 -I "RegSMOImproved -T 0.001 -V -P 1.0E-12 -L 0.001 -W 1" -K "PolyKernel -E 1.0 -C 250007"		
			Number of kernel evaluations: 10 (87,952% cached)		
inst#	AÑO ESTUDIANTES		inst#	AÑO ESTUDIANTES	
2012	2012	1422	2012	2012	1422
2013	2013	3212	2013	2013	3212
2014	2014	3130	2014	2014	3130
2015	2015	4765	2015	2015	4765
2016*	2016,0002	3132,25	2016*	2016,0029	5379,4973
2017*	2017,0006	3132,25	2017*	2017,0081	6712,3302
2018*	2018,0011	3132,25	2018*	2018,0142	7576,2995

Tabla 3. Análisis de patrones del escenario B

Al realizar el análisis entre los dos algoritmos de regresión, es evidente notar que la optimización de secuencia mínima presenta una proyección de datos, pero que estos no son tan aceptables debido a que presentan un crecimiento grande en la cantidad de datos, pero para tomar esta decisión también se deben considerar otras variables externas. Por lo que la predicción puede ser medianamente aceptable. Mientras que los resultados de Regresión Lineal mantienen una constante, que también podría ser considerada como una predicción válida, pero con menos probabilidad de que ocurra.

**Resultado 2:** Al ser algoritmos utilizados para la predicción, y mostrar un 87,952% de confiabilidad, el algoritmo de Optimización de Secuencia Mínima, es el algoritmo que se puede tomar como referencia para el crecimiento de la cantidad de estudiantes en la ESPOCH, para los siguientes 3 años.

*Conocimiento Descubierto:*

En base a los algoritmos aplicados es preciso presentar el conocimiento descubierto en los datos de la UNA - ESPOCH.

Conocimientos:

- La mayor cantidad de estudiantes de la ESPOCH en los años 2012, 2013, 2014 y 2015, están en el área de Ingeniería, donde de los 5303 estudiantes, 1549 fueron del género femenino.
- En el área de Servicios, en el año 2014 existió mayor cantidad de estudiantes, siendo 155 mujeres de los 323 estudiantes.
- El área Comercial en los últimos cuatro años ha generado el segundo

lugar con mayor cantidad de estudiantes, siendo 2298, con tan sólo 839 mujeres.

- En el área de Artes la mayoría de los estudiantes son hombres.
- El área con el tercer lugar de cantidad de estudiantes es Agricultura con un total de 1846, donde tan sólo hubo 868 hombres.
- La cantidad de estudiantes hombres en el universo global de estudiantes representa un 53%, siendo mayor, pero que muestra también que la diferencia entre estos dos géneros casi se está equiparando.
- El año en donde existió mayor cantidad de estudiantes mujeres fue el 2015 con un 37.2%
- El área con mayor cantidad de estudiantes hombres y mujeres es el área de Ingeniería, mientras que el segundo lugar con mayor cantidad de estudiantes mujeres es el área Comercial y para los estudiantes hombres el área de Agricultura.
- La ESPOCH se proyecta a tener mayor cantidad de estudiantes en los próximos 3 años

Al aplicar los algoritmos de DM, es evidente como se solucionó la falta de información que la UNAE tenía. Mediante entrevistas a los coordinadores General, Académico y Administrativo, que fueron el personal del nivel estratégico que toman decisiones y son usuarios directos de la información, posterior a la aplicación y explicación de los resultados obtenidos en DM, se obtuvieron los siguientes puntos considerables mencionados por el personal.

- El poseer información consistente de los procesos académicos de Nivelación permite tomar decisiones y plantear estrategias para que el rendimiento de los estudiantes mejore.
- El conocer información directa de los estudiantes, como los lugares de donde provienen, género y discapacidades, permite conocer de mejor forma la realidad institucional, creando programas de bienestar y vinculación que mejoren el transcurso del estudiante en la ESPOCH.
- Conocer la tendencia al momento de la selección de las carreras de la institución permiten potenciar las más escogidas y buscar estrategias para las menos postuladas.
- Mediante estas entrevistas se pudo demostrar que se generó conocimiento para la toma de decisiones mediante información apropiada, correcta, consistente.

## CONCLUSIONES

El proyecto de investigación y desarrollo cumplió el objetivo principal de generación de conocimientos mediante el análisis de algoritmos de DM que se acoplaron a la realidad de la UNAE - ESPOCH, siendo el algoritmo de clasificación de Redes Bayesianas el más apropiado para los datos, y de los algoritmos de regresión; el algoritmo de Optimización de Secuencia Mínima, tanto en postulaciones como predicciones a tres años, brindando la posibilidad de contar con información correcta, consistente y adecuada para la toma de decisiones.

Toda esta información permite plantear estrategias para que el rendimiento de los estudiantes mejore, además, al conocer los lugares de donde provienen, género y discapacidades, permite saber de mejor forma la realidad institucional, creando programas de bienestar y vinculación que mejoren el transcurso del estudiante en la ESPOCH. Además, se permite la construcción de infraestructura y aseguramiento de la calidad en la educación superior, siendo el principal soporte para la toma de decisiones.

## REFERENCIAS BIBLIOGRÁFICAS

- Aluja, T. (2001). La Minería de Datos, entre la Estadística y la Inteligencia Artificial. *Qüestiió*, 25(3), 479-498.
- Ancell Trueba, R. (2013). Aportaciones de las redes bayesianas en meteorología. Predicción probabilística de precipitación. Tesis de Doctorado, 228. Retrieved from [http://www.aemet.es/documentos/es/conocermas/publicaciones/redes\\_bayesianas/Predicc\\_probab\\_precipitac.pdf](http://www.aemet.es/documentos/es/conocermas/publicaciones/redes_bayesianas/Predicc_probab_precipitac.pdf)
- Asencios, V. (2004). Data Mining y el descubrimiento del conocimiento. *Industrial Data*, 83-86. Obtenido de <http://www.redalyc.org/resumen.oa?id=81670213>
- Bernabeu, R. (2010). *Hefesto*. Obtenido de <http://www.dataprix.com/data-warehousing-y-metodologia-hefesto/hefesto-metodologia-propia-para-la-construccion-un-data-warehhttp://www.dataprix.com/data-warehousing-y-metodologia-hefesto/ii-hefesto-metodologia-propia-para-la-construccion-un-data-wa>
- Camana Fiallos, R. (2012). Aplicación de Técnicas de Minería de Datos para la Indagación y Estudio de Resultados Electorales. *CienciAmérica: Revista de Divulgación Científica de La Universidad Tecnológica Indoamérica*, 1, 85-94. Obtenido de <https://dialnet.unirioja.es/servlet/articulo?codigo=6163757>
- Daza, A. (2016). *Data Mining, Minería de datos*. Perú: Editorial Macro.
- Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *17(3)*, 37. doi:<http://doi.org/10.1609/aimag.v17i3.1230>
- González, C. B., & García, F. (2010). Práctica Final INTELIGENCIA EN COMUNICACIONES Minería de Dato. En *Predicción de las condiciones meteorológicas*.
- González, L. (2005). Una arquitectura para el análisis de información que integra procesamiento analítico en línea con minería de datos. En *Tesis para optar por el grado de Maestro en Ciencias con especialidad en Ingeniería en Sistemas Computacionales*. Universidad de las Américas Puebla, Puebla.
- Han, J., & Kamber, M. (2006). *Data mining concepts and techniques*. New York: Elsevier.
- Hernández, J., Ferrari, C., & Ramírez, M. (2004). *Introducción a la minería de datos*. Madrid: Pearson Education.
- Juan, I., Moine, M., Gordillo, D., Ana, D., & Haedo, S. (2011). *Proyectos de minería de datos*.
- Lewandowski, C. (2015). *The Effects of Brief Mindfulness Intervention*

- on Acute Pain Experience: An Examination of Individual Difference. doi:<http://doi.org/10.1017/CBO9781107415324.004>
- Lopez-Yáñez, I., Márquez, C. Y., Nieto, O. C., & Cruz, A. J. A. (2011). Predicción de la concentración de contaminantes atmosféricos basada en un clasificador asociativo de patrones. *Revista Facultad de Ingeniería*, (60), 20–30.
- Maimon, O., & Rokach, L. (2010). Data Mining and Knowledge Discovery Handbook. En *In Data Mining and Knowledge Discovery Handbook* (págs. 22–38). doi:[http://doi.org/10.1007/0-387-25465-x\\_2](http://doi.org/10.1007/0-387-25465-x_2)
- Olson, D., & Delen, D. (2008). *Advanced Data Mining Techniques*. doi:<http://doi.org/10.1007/978-3-540-76917-0>
- PostgreSQL. (2016). *Sobre PostgreSQL*. Obtenido de [http://www.postgresql.org/es/sobre\\_postgresql](http://www.postgresql.org/es/sobre_postgresql)
- Thapelo, S. (2014). Técnicas de aprendizaje automatizado para el pronóstico de temperaturas mínimas en el Centro Meteorológico de Villa Clara, Santa Clara. Retrieved from <http://dspace.uclv.edu.cu/handle/123456789/7341>
- Valero, S., Salvador, A., & García, M. (2010). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. *Recursos Digitales*. <https://doi.org/74420>
- Weka. (2016). *Weka 3: Data Mining Software in Java*. Obtenido de <http://www.cs.waikato.ac.nz/ml/weka/>

