



---

### Importancia del Aprendizaje Automático utilizando técnicas de minería de datos en la gestión académica

*Julia Patricia Melo Morín<sup>1</sup>, Eric Álvarez Baltierra<sup>1</sup>, Gwendolyne Barrera Silva<sup>1</sup>.*

<sup>1</sup>Instituto Tecnológico Superior de Pánuco

patricia.melo@itspanuco.edu.mx

#### RESUMEN

El aprendizaje automático es el proceso en que las computadoras desarrollan el reconocimiento de patrones, o la capacidad de aprender continuamente y hacer predicciones basadas en datos; una forma de generar patrones de comportamiento es mediante la aplicación de técnicas de minería de datos.

El aprendizaje automático es parte de la inteligencia artificial que se centra en el desarrollo de programas informáticos generando resultados que pueden cambiar cuando se exponen a nuevos datos. Es importante aplicar en las instituciones educativas, soluciones que ayuden a la transformación digital de las mismas para proporcionar un cambio favorable en la búsqueda de estrategias que garantizan los fines administrativos, investigativos y académicos necesarios al momento de valorar la calidad de las instituciones educativas.

Éste artículo describe una propuesta del uso del aprendizaje automático, aplicando técnicas de minería de datos en el Instituto Tecnológico Superior de Pánuco como apoyo en la gestión académica.

**Palabras clave:** Aprendizaje Automático, Minería de Datos, Gestión Académica, Patrones.

## INTRODUCCIÓN

La aplicación de métodos y técnicas de minería de datos en la educación, está documentados en una gran cantidad de artículos, incluyendo minería de datos y aprendizaje automático; psicometría y otras áreas de la estadística, visualización de la información y el modelamiento computacional (Baker R. & Yacef, K., 2009).

En el ámbito educativo la aplicación de la minería de datos como técnica de análisis se ubica en el entorno del sistema educativo y en particular en la educación superior. Las bases de datos que se utilizan en los sistemas educativos permiten disponer de una gran cantidad de información, tanto de los estudiantes, trabajadores, departamentos, universidades, etc. (Torrado, 2011).

Romero y Ventura (2007), clasifican el trabajo de la minería de datos educativa en las categorías de estadística y visualización, así como minería de datos web. Baker (2010) clasifica el trabajo en la minería de datos educativa como Predicción, Agrupación, Relación, Visualización de los datos para el juicio humano y Descubrimiento con modelos. Romero, Ventura, Espejo y Hervás (2008) llegaron a la conclusión de que el modelo de clasificador es apropiado para el uso educativo en términos de precisión y claridad para la toma de decisiones.

Las instituciones educativas cuentan con muchos datos que se generan en distintas áreas tanto a nivel administrativo, académico e investigativo, pero en la mayoría de los casos no cuentan con información que permita la toma de decisiones oportunas. Es prioritario identificar el conocimiento oculto en la información de los estudiantes, para tener la posibilidad de encontrar patrones de comportamiento, entender, observar, identificar, administrar y responder cuestionamientos que permitan beneficiar el rendimiento académico de los mismos.

Aplicaciones basadas en el aprendizaje automático (Machine Learning) proporcionan a las instituciones, herramientas de gestión que permita brindar servicios de calidad, permitiendo la reducción del tiempo al momento de integrar, explotar y analizar datos, transformándose en información que permita a los directivos tomar decisiones basadas en hechos comprobables y medidos acerca del desempeño de la institución y de sus estudiantes.

Este artículo describe la necesidad de implementación herramientas del aprendizaje automático en las instituciones de nivel superior, tal como el Instituto Tecnológico Superior de Pánuco (ITSP), como apoyo a la obtención de patrones de comportamiento del rendimiento académico.

## Estado del Arte.

Hay una gran cantidad de artículos que describen el uso del aprendizaje automático y técnicas de minería de datos en la educación, tanto a nivel internacional como en México, describiendo sus aplicaciones y usos para la mejora en la gestión académica.

Entre los sistemas que aplican la inteligencia de negocios se encuentran el Sistema de Gestión de la Nueva Universidad (SIGENU), para apoyar el desempeño de los procesos docentes en la educación superior en Cuba, así como el Sistema Universitario Nacional (SIU) de la República Argentina que tiene como objetivo contribuir a mejorar la gestión de las instituciones desde la integridad de la información, la optimización de recursos y la posibilidad de trazar proyecciones y estrategias educativas (Díaz M. & Boza L., 2007) Universitas XXI e Inteligencia Institucional es un sistema orientado al análisis y toma de decisiones que conceptualiza la realidad de la universidad, integrando datos generados de los diferentes sistemas de gestión existentes en la institución en un único almacén de datos consolidados. Este sistema tiene la capacidad de emitir los informes estadísticos del centro y permite detectar de errores ocultos en la calidad de los datos (OCU, 2018).

La Universidad Nacional del Noroeste de la provincia de Buenos Aires (UNNOBA), utiliza la tecnología aplicada a la educación, para diseñar e implementar estrategias tecnológicas para minimizar la deserción; utilizando la minería de datos. Se define un modelo de intervención áulica virtual que permite diseñar e implementar estrategias tecnológicas (Russo, C, et al, 2018).

Istvan, R. M., & Lasagna, V. (2018), describen que en Argentina en la Universidad UTN FRLP, exploran las técnicas de minería de datos para responder a las demandas de la institución, desarrollando un sistema de detección temprana de posibles desertores y gestión tutorial, para elevar la eficiencia terminal de la institución. Han formado un sistema de indicadores de detección temprana de problemáticas del aprendizaje y dificultades en el trayecto de formación académica, utilizando la minería de datos para para la detección de patrones individuales.

En Colombia se creó el programa de ingeniería SARA (Sistema de Acompañamiento para el Rendimiento Académico), apoyado con el proceso de descubrimiento del conocimiento en base de datos, para reducir los índices de deserción estudiantil en la Educación Superior, definiendo estrategias que ayuden en la intervención de la vida académica de los estudiantes, a través de la inclusión de técnicas de minería de datos para identificar patrones que permitan caracterizar o predecir posibles casos de deserción dentro del programa (Calvache-Fernandez, L.C et al, 2018).

Rivero P. (2014), describe que las técnicas de aprendizaje automático están basadas sobre un modelo explícito o implícito establecido que posibilita categorizar los patrones

analizados; los datos están etiquetados para entrenar el modelo de comportamiento. Entre los modelos utilizados en las técnicas del aprendizaje automático son Redes Bayesianas, Modelos de Markov, Redes Neuronales, Técnicas de lógica difusa, Algoritmos genéticos, Agrupamiento y detección de outlier.

En México en la Universidad Politécnica de Puebla (UPP) realizaron investigaciones para la construcción de un modelo que ayude a predecir las causas de reprobación, tomando en cuenta diversos aspectos de los estudiantes actuales y anteriores, entre ellos, los datos del historial académico, problemas personales y psicológicos. La finalidad es la obtención del modelo predictivo para indicar mediante una interfaz gráfica las materias que generan mayor índice de reprobación (Ramos, et al, 2010).

En otra investigación se utilizaron técnicas de minería de datos para detectar, cuáles son los factores que más influyen para que los estudiantes de enseñanza media o secundaria fracasen, es decir, suspendan o abandonen sus estudios en el Programa II de la Unidad Académica Preparatoria de la Universidad Autónoma de Zacatecas (UAPUAZ) de México (Vera, C. M., Morales, C. R., & Soto, S. V., 2012).

En instituciones de nivel superior se encuentra una investigación realizada en la Universidad Tecnológica de Izúcar de Matamoros, donde aplicando técnicas de minería de datos predicen la deserción escolar, tomando como base de análisis los datos del estudio socioeconómico del EXANI-II, elaborado por el CENEVAL (Orea, S. V., Vargas, A. S., & Alonso, M. G. ,2005). Para el desarrollo del proyecto, se siguió el proceso KDD, el cual marca el proyecto de minería de datos en una serie de fases definidas: integración y recopilación; selección, limpieza y transformación; minería de datos; evaluación e interpretación, y difusión y uso (Hernández et al., 2004).

## **Planteamiento del problema**

El rendimiento académico de los estudiantes es una preocupación constante en la mayoría de las instituciones educativas y éste constituye una de las cuestiones fundamentales a la hora de abordar el tema de la calidad de la Enseñanza Superior (De Miguel Díaz, et al, 2002), debido a que es un indicador que permite una aproximación a la realidad educativa (Vargas, 2007).

De acuerdo a la Asociación Nacional de Universidades e Instituciones de Educación Superior (ANUIES) uno de los problemas en la Educación Superior en México es la baja eficiencia terminal de los alumnos que ingresan, ya sea por problemas de deserción, reprobación o falta de titulación, considerando que de cada 100 estudiantes, solo 60 egresan y de los cuales solo 20 se titulan; y de acuerdo a la Organización Para la Cooperación y el Desarrollo Económico (OCDE) en 2013 aproximadamente un 22% de los estudiantes de educación superior en México se gradúan en educación terciaria en programas de grado .

Los organismos internacionales toman en cuenta indicadores como reprobación, rezago, aprobación, cobertura y deserción, entre otros, para evaluar y determinar los resultados del proceso educativo de una institución. Dentro de los indicadores establecidos como metas a nivel nacional, se encuentran el indicador de deserción, de reprobación y el de eficiencia terminal.

Existen muchos factores que pueden afectar al rendimiento académico, dichos factores pueden ser cualitativos (tipo psicológico y sociológico) y cuantitativos, que proporcionan datos precisos sobre los resultados académicos tanto de los estudiantes como de la institución.

El análisis del rendimiento académico y del fracaso escolar para las instituciones de educación superior es de gran interés así como la determinación de los múltiples factores que pueden influir en los mismos; con la finalidad de generar información que permita diseñar acciones encaminadas a lograr un mejor trayecto del alumno en la institución educativa.

Las instituciones de educación superior en la búsqueda de la calidad educativa se someten a diferentes evaluaciones de acreditación de planes y programas de estudio, así como la necesidad de la implementación de normas de calidad; dichas evaluaciones incluyen indicadores para determinar el grado de cumplimiento de las actividades de la gestión académica del estudiante.

Es importante y urgente que las instituciones apliquen el aprendizaje automático utilizando minería de datos educativa empleando herramientas tecnológicas, algoritmos y estrategias de análisis de información utilizadas por la inteligencia artificial, pero dentro de un contexto educativo, dónde se resuelvan problemas que mejoren el proceso enseñanza-aprendizaje, proporcionando un análisis estadístico de una evaluación, programa curricular o una unidad de aprendizaje, mediante modelos predictivos generados por métodos y técnicas de la minería de datos tradicional, así también, el mejoramiento de software educativo mejorando las capacidades para individualizar las experiencias de aprendizaje de los estudiantes.

Si las instituciones de nivel superior pueden identificar los factores de bajo rendimiento de los estudiantes y son capaz de predecir el comportamiento de los alumnos, este conocimiento puede ayudarles a tomar acciones proactivas que permitan mejorar el rendimiento de los estudiantes.

## **DESARROLLO**

Para aplicar el aprendizaje automático en el Instituto Tecnológico Superior de Pánuco (ITSP), se propone implantar un sistema experto que aplique técnicas de minería de datos para generar un modelo que determine el rendimiento académico de los estudiantes para apoyar la inteligencia de negocio en instituciones educativas y así

prevenir la deserción y reprobación escolar. El desarrollo de un sistema experto es un impacto positivo en el ámbito escolar que permite aumentar el desempeño de las instituciones educativas.

En el ITSP es una preocupación el problema del fracaso escolar, así como la determinación de los múltiples factores que pueden influir en él mismo, el principal interés para el ITSP es la generación de un sistema que permita identificar posibles alertas de deserción y/o reprobación de los estudiantes, para así implementar estrategias que permitan reducir esta problemática.

Comprender y analizar los factores involucrados para el rendimiento deficiente de los estudiantes es un proceso complejo y oculto en el pasado y presente de la información que indican el rendimiento académico y el comportamiento de los estudiantes. Potentes herramientas son necesarias para analizar y predecir el rendimiento de los estudiantes desde el punto de vista científico y una de las herramientas tecnológicas disponibles es la minería de datos.

Al combinar las trayectorias estudiantiles observadas mediante el aprendizaje automático en las instituciones es posible predecir (proyectar) el desempeño de un alumno. La utilización de herramientas informáticas al proceso educativo de los alumnos, permitirá la gestión inteligente de grandes volúmenes de datos, siendo posible el procesamiento y extracción de información que permita la inteligencia de negocios para la toma de decisiones en el ámbito de la gestión académica; determinando los factores que afectan la deserción y/o reprobación de los alumnos y así ofrecer información que permita realizar acciones para evitar la deserción o reprobación del estudiante, favoreciendo la inteligencia de la institución educativa.

En el ITSP se genera una gran cantidad de información desde el ingreso, permanencia y egreso de los estudiantes tanto de nivel superior como de posgrado, la cual no se analiza para tener una integración de la misma que permitan la identificación de patrones que permita prever problemáticas de deserción. Es necesaria la integración de herramienta o conjunto de herramientas que permitan analizar todos los flujos de información, que permitan ser un elemento para medir la calidad educativa que se está ofreciendo.

### **Metodología propuesta**

La propuesta de aplicación en el ITSP requiere de una investigación de tipo aplicada que analicen diferentes técnicas de aplicación del aprendizaje automático utilizando técnicas de minería de datos para la incorporación de una o varias de ellas en el desarrollo de la aplicación que apoye a la gestión académica.

El desarrollo de la propuesta aplicará las características del método científico para que sea congruente, confiable y fundamentada en las teorías relativas al objeto de estudio,

ya que se requiere que los resultados tengan el grado máximo de confiabilidad. Se pretende realizar una investigación documental sobre las investigaciones realizadas en este ámbito.

El diseño experimental se definirá en base a una secuencia de pasos planeados de antemano para asegurar que los datos que se obtengan permitan un análisis objetivo que lleve a conclusiones válidas. El estudio se dividirá en las siguientes etapas:

- Comprensión y planteamiento del problema.
- Elección de factores o variables.
- Selección de variables óptimas.
- Análisis de datos
- Presentación de resultados.

El diseño de la investigación seguirá las siguientes fases.

- Identificación del problema y los elementos involucrados.
- Revisión del estado de arte relacionado con rendimiento académico y minería de datos.
- Revisión del estado del arte de aplicaciones de gestión académica, aprendizaje automático en el área educativa.
- Identificación de técnicas de minería de datos educativa.
- Análisis de las características de cada una de las técnicas identificadas.
- Comparar características para la elección de la técnica más adecuada.
- Identificación de los factores relacionados con el rendimiento académico
- Análisis estadístico de los datos para identificar patrones de comportamiento inicial.
- Aplicación de la metodología de Conocimiento en Base de datos en el proceso de la minería de datos.
- Integración del almacén de los datos y limpieza de los datos.
- Aplicación de algoritmos de clasificación, asociación y agrupamiento para la obtención de patrones de comportamiento.
- Analizar y obtener un modelo de los datos.
- Presentar la interpretación del modelo obtenido través de un sistema informático.
- Interacción con el software de Weka en la aplicación y uso de las técnicas de Minería de Datos.
- Presentar los resultados a través de un sistema informático que permita la inteligencia de negocios de la institución educativa, permitiendo la interacción y generación de información a partir de nuevos datos.

Las fuentes empleadas, para la realización de la investigación serán:

- Entrevistas a directivos en los diferentes departamentos académicos
- Bases estadísticas, almacenes de datos y documentales oficiales
- Publicaciones de referencia nacional e internacional.

## CONCLUSIONES

Es de gran importancia proporcionar a las instituciones educativas de herramientas que permitan la toma de decisiones de manera eficaz; el uso de las tecnologías es de gran ayuda para generar modelos que analicen y obtengan resultados con grandes volúmenes de datos y a velocidades mayores que el ser humano pueda realizar.

La minería de datos educativa y el aprendizaje automático es urgente aplicarlo en las diferentes instituciones, ya que permite descubrir el conocimiento oculto en los datos y así permitir tomar acciones que apoyen a prevenir y detectar problemas educativos, como el rendimiento académico de los estudiantes.

## LECTURA CITADA

Alcover R. et al (2007). Análisis del rendimiento académico en los estudios de informática de la Universidad Politécnica de Valencia aplicando técnicas de minería de datos. Consultado el 16 de Marzo de 2016 en <http://bioinfo.uib.es/~joemiro/aenui/procJenui/Jen2007/alanal.pdf>

Baker, R. and Yacef, K. (2009). The State of Educational Data mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*,

Calvache-Fernandez, L. C., Álvarez-Vallejo, V., & Triviño-Arbelaez, J. I. (2018). Proceso KDD como apoyo a las estrategias del proyecto SARA (Sistema de Acompañamiento para el Rendimiento Académico). *Revista Educación en Ingeniería*, 13(26), 82-89.

Dapozo, G. N., Porcel, E., López, M. V., Bogado, V. S. and Bargiela, R. (2006). Aplicación de minería de datos con una herramienta de software libre en la evaluación del rendimiento académico de los alumnos de la carrera de Sistemas de la FACENA-UNNE. In VIII Workshop de Investigadores en Ciencias de la Computación.

De Miguel Díaz, M., Urquijo, P. A., Blanco, J. M. A., Escorza, T. E., Espinar, S. R. and García, J. V. (2002). Evaluación del rendimiento en la enseñanza superior. Comparación de resultados entre alumnos procedentes de la LOGSE y del COU. *revista de Investigación Educativa*, 20(2), 357-383.

Díaz M. & Boza L. (2007). Módulo de Recuperación Web de Información Docente del Sistema de Gestión de la Nueva Universidad "SIGENU". Facultad de Ingeniería Informática, Instituto Superior Politécnico José Antonio Echeverría. Trabajo de diploma para optar por el título de Ingeniería en Informática.

- Hernández J., Ramirez M. and Ferri C. (2004). *Introducción a la Minería de Datos*. Ed. Pearson, México.
- Istvan, R. M., & Lasagna, V. (2018). Modelo predictivo para la determinación de riesgo de deserción universitaria para ingresantes de la UTN FRLP. In XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste).
- OCDE (2006). *Higher education: quality, equity and efficiency*. Consultado el 28 de Julio de 2016, en [http://www.oecd.org/site/0,3407,en\\_21571361\\_36507471\\_1\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/site/0,3407,en_21571361_36507471_1_1_1_1_1,00.html)
- OCU (2018). Oficina de Cooperación Universitaria (OCU) Universitat XXI - Inteligencia Institucional. Fecha de acceso: 25/03/2018; Disponible en <http://www.ocu.es/>.
- Orea, S. V., Vargas, A. S., and Alonso, M. G. (2005). Minería de datos: predicción de la deserción escolar mediante el algoritmo de árboles de decisión y el algoritmo de los k vecinos más cercanos. *Ene*, 779(73), 33.
- Ramos, E. R., et al (2010). Modelo predictivo para la determinación de causas de reprobación mediante minería de datos. In II Conferencia Conjunta Iberoamericana sobre Tecnologías para el aprendizaje-CcITA (pp. 48-55).
- Robles R. and Martínez L. J. F. (2007). La reprobación de Matemáticas desde la perspectiva del alumno, el docente y la academia. *Psicología y Educación*, 1(1), 97-104
- Rivero Pérez, Jorge Luis. (2014). Técnicas de aprendizaje automático para la detección de intrusos en redes de computadoras. *Revista Cubana de Ciencias Informáticas*, 8(4), 52-73. Recuperado en 03 de septiembre de 2018, de [http://scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S2227-18992014000400003&lng=es&tlng=es](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992014000400003&lng=es&tlng=es).
- Romero C. and Ventura S. (2007). "Educational data mining: A survey from 1995 to 2005," *Expert Systems with Applications* 33 (Science Direct).
- Romero, C.; Ventura S.; Espejo, P. and Hervás C. (2008). "Data Mining Algorithms to Classify Students." In EDM, pp. 8-17.
- Russo, C. C., Ahmad, T., Adó, M., Lencina, P., Serrano, E., Rodríguez, M., & Piernalini, R. (2018). Tecnologías aplicadas a educación en UNNOBA. In XX Workshop de Investigadores en Ciencias de la Computación (WICC 2018, Universidad Nacional del Nordeste). Identificador URI <http://sedici.unlp.edu.ar/handle/10915/68658>
- Tejada, J. M. and Arias, F. (2005). Prácticas organizacionales y el compromiso de los trabajadores hacia la organización. *Enseñanza e Investigación en Psicología*, 10-295-309. Recuperado de <http://www.redalyc.org/articulo.oa?id=29210206>.
- Torrado, M. (2011) *Minería de datos aplicados a la educación*. Barcelona: Universidad de Barcelona. Deposito Digital <http://hdl.handle.net/2445/19862>

Vargas, G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista educación*, 31(1), 43-63

Vera, C. M., Morales, C. R. and Soto, S. V. (2012). Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos. *Revista Iberoamericana de Tecnologías del/da Aprendizaje/Aprendizagem*, 109.