



EL LORO HUASTEKO
Órgano de Divulgación Científica y Tecnológica del
Instituto Tecnológico Superior de Pánuco

Memorias del 1er Seminario de Investigación Multidisciplinario Interinstitucional 2019
Instituto Tecnológico Superior de Pánuco
Instituto Tecnológico Superior de Ébano

Aplicación del aprendizaje automático en la educación

Melo Morín Julia Patricia
Santana Esparza Gil
Ahumada Cervantes Ma. De Los Ángeles
Email autor corresponsal:
Área de participación:

Instituto Tecnológico Superior de Pánuco
Instituto Tecnológico Superior de Pánuco
Instituto Tecnológico Superior de Pánuco
patricia.melo@itspanuco.edu.mx
Investigación Educativa

RESUMEN

La tecnología informática constituye una herramienta fundamental en las grandes organizaciones y permite la optimización de diferentes procesos. La utilización de la tecnología permite convertir grandes volúmenes de datos en experiencia, conocimiento y sabiduría, que permita a las organizaciones y directivos tomar decisiones oportunas.

En la educación el aprendizaje automático (Machine Learning) ha tomado gran importancia, permitiendo la toma de decisiones de forma oportuna para abatir una de las grandes preocupaciones en la educación: el rendimiento académico de los alumnos.

El aprendizaje automático proporciona 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 el proceso de aplicación 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.

Palabras claves: Aprendizaje Automático, Rendimiento Académico, Minería de Datos.

ABSTRACT

Computer technology is a fundamental tool in large organizations and allows the optimization of different processes. The use of technology allows converting large volumes of data into experience, knowledge and wisdom, which allows organizations and managers to make timely decisions.

In education, machine learning has taken on great importance, allowing timely decision making to address one of the major concerns in education: the academic performance of students.

Machine learning provides institutions with management tools that allow providing quality services, allowing the reduction of time when integrating, exploiting and analyzing data, transforming into information that allows managers to make decisions based on verifiable and measured facts about the performance of the institution and its students.

This article describes the process of applying machine learning in higher level institutions, such as the Instituto Tecnológico Superior de Pánuco (ITSP), as support for obtaining patterns of academic performance behavior.

Key words: Machine Learning, Academic Performance, Data Mining.

INTRODUCCIÓN

De acuerdo a De Miguel Díaz, et al [2002], el rendimiento académico de los alumnos constituye uno de los elementos fundamentales cuando se habla de calidad en la Enseñanza Superior, ya que es un indicador que permite una aproximación a la realidad educativa [Vargas, 2007].

Existen muchos factores que afectan al abandono escolar y al rendimiento académico, tales como la falta de dinero para la adquisición de materiales escolares, transporte o inscripción; falta de interés en el estudio, ya sea por clases deficientes, embarazo u obligaciones familiares; o inclusive el nivel académico de los padres influyen en el éxito o fracaso de los alumnos; dichos factores proporcionan datos precisos sobre los resultados académicos tanto de los estudiantes como de la institución [Melo, 2018].

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 necesario indagar en el conocimiento oculto de los datos para identificar los factores que afectan al bajo rendimiento, para encontrar y predecir patrones de comportamiento que den respuestas a cuestionamientos de las causas que permitan tomar decisiones proactivas que mejoren el rendimiento de los estudiantes.

El aprendizaje automático es parte de la inteligencia artificial que se centra en el desarrollo de programas informáticos que permiten el reconocimiento de patrones o la capacidad de aprender continuamente y hacer predicciones basadas en datos, que pueden cambiar cuando se exponen a nuevos datos. Es importante aplicar en las instituciones educativas, como los tecnológicos, aplicaciones que apoyadas de la tecnología generen un cambio propicio para planear acciones que garanticen el desarrollo sostenible de la nación y permitan evaluar exitosamente la calidad de las actividades académicas realizadas.

METODOLOGÍA

Para llevar a cabo la aplicación del aprendizaje automático se integraron diferentes almacenes de datos que describen resultados históricos académicamente de los alumnos desde el año 2007 hasta el 2018, en el Instituto Tecnológico Superior de Pánuco y aplicando la metodología de minería de datos KDD, para integrar cada etapa de acuerdo al almacén de datos. La Figura 1, indica las etapas de la extracción y descubrimiento de conocimiento de la metodología KDD.

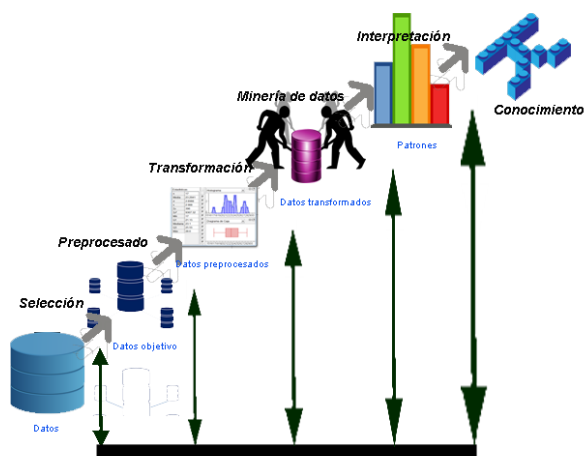


Figura 1. Etapas para el proceso de extracción de conocimiento. Fuente: Basado en Fayyad et al (1996) y Gómez (1998).

Los datos existentes de los estudiantes desde su ingreso, permanencia y egreso permitió la integración del almacén de los datos, las diferentes técnicas predictivas y deductivas del aprendizaje automático con minería de datos generó patrones de comportamientos que a través de un sistema experto basado en inteligencia de negocio permite procesar, analizar, visualizar y acceder a resultados en forma automática y eficiente que apoye a la toma de decisiones oportuna.

Se aplicaron las diferentes técnicas de minería de datos utilizando el software de WEKA, en base a los resultados de las diferentes técnicas de minería de datos se generó un modelo predictivo que permitió ser la base de conocimiento para la realización del sistema experto.

Se identificaron factores y variables que describen los índices del rendimiento académico de los alumnos, principalmente reprobación y deserción. Se generaron reglas de comportamiento utilizando el software Weka, que es un software que permite aplicar diferentes algoritmos predictivos y deductivos de la minería de datos.

RESULTADOS Y DISCUSIÓN

Como resultado se generó un sistema experto que basado en reglas generadas por las diferentes técnicas de extracción de datos, mediante algoritmos de clasificación de Minería de datos que clasificó en un porcentaje del 97% de instancias correctamente y una estadística Kappa de 0.9208.

Estas reglas obtuvieron características en cada uno de los atributos, así como el porcentaje de certeza de cada una de ellas, para identificar los factores que ponen en riesgo a un estudiante de desertar su carrera profesional. El algoritmo Part para el caso de estudio, generó 143 reglas que fueron implementadas en el sistema experto, la ejecución del algoritmo PART se muestra en la Figura 2.

```

Classifier output
ansiedad
Test mode: 10-fold cross-validation
=== Classifier model (full training set) ===

PART decision list
-----
REG_PROC = 3 AND
CVE_PROC = 188126 AND
ESCO_PAD = 2: 2 (134.0/5.0)

VERSION = 999 AND
vive = 8: 2 (380.0/17.0)

dictamen_lenguaje = 2 AND
CVE_PROC = 188126: 2 (102.0/1.0)

dictamen_lenguaje = -1: 2 (51.0/1.0)

dictamen_lenguaje = 2 AND
ICNE = 940: 2 (12.0/1.0)

dictamen_lenguaje = 2 AND
ICNE = 1006: 2 (11.0)

dictamen_lenguaje = 2 AND
ICNE = 910: 2 (11.0)
    
```

Figura 2. Ejecución del algoritmo Part. Fuente: Software Weka, ejecución propia.

En el algoritmo Part las reglas se expresan mediante sentencias lógicas que relacionan la descripción, mediante variables de un objeto (condiciones) y las clases de decisión.

Los elementos básicos que formaron parte de la arquitectura del sistema experto llevado a cabo fueron la base de conocimientos, el motor de inferencia, la base de datos y la interfaz con el usuario. Se creó un sistema experto en línea, que permite a los usuarios interactuar y, por lo tanto, permitir la toma de decisiones en función de los resultados presentados. La Figura 3 muestra el proceso involucrado en la realización del sistema experto.

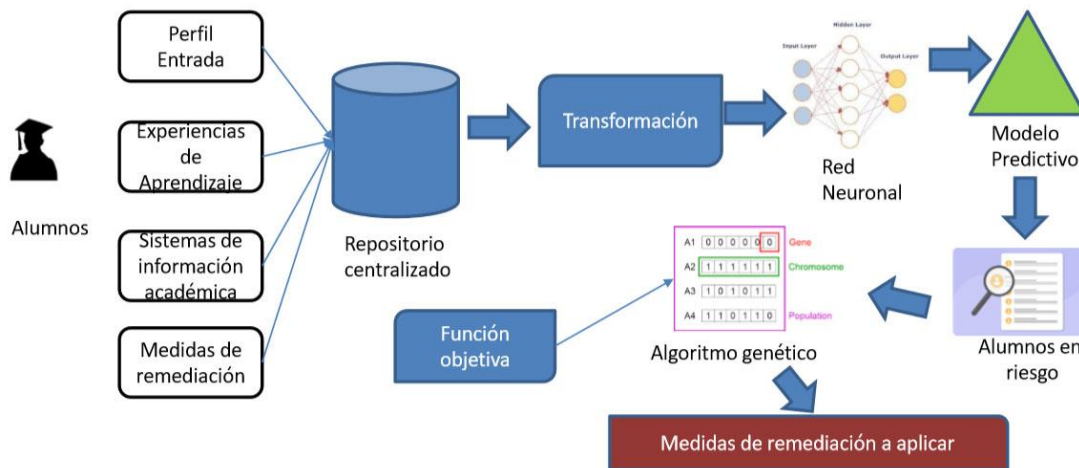


Figura 3. Proceso de aplicación del aprendizaje automático en la generación del sistema experto.

El sistema experto se basó en modelos predictivos de minería de datos que permite alertar de situaciones donde el estudiante se encuentre en riesgo de no concluir su formación profesional, o identificar en base a los datos históricos previamente analizados si cuenta con características similares a los alumnos con bajo rendimiento, para que pueda ser atendido y apoyado con los diferentes programas institucionales que cuente el instituto.

El sistema experto fue validado sus resultados y complementado con una aplicación desarrollado en Lenguaje R, con librerías que permita la simulación de datos actuales con datos históricos. La Figura 4, muestra la imagen de una red neuronal generada a partir de los datos procesados del almacén de datos, utilizando el lenguaje R.

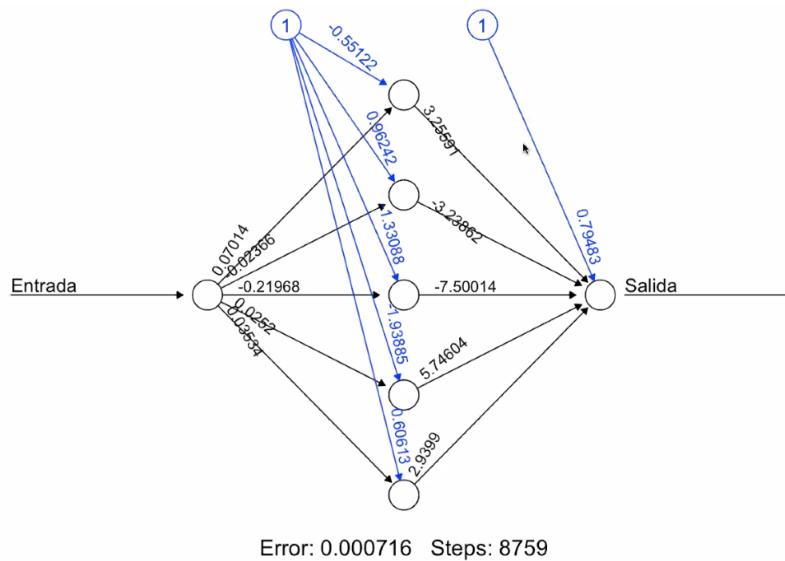


Figura 4. Red neuronal, utilizando el lenguaje R.

El aprendizaje automático utilizando Minería de datos Educativa permite aplicar 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.

La interfaz gráfica del sistema experto en línea se muestra en la Figura 5.

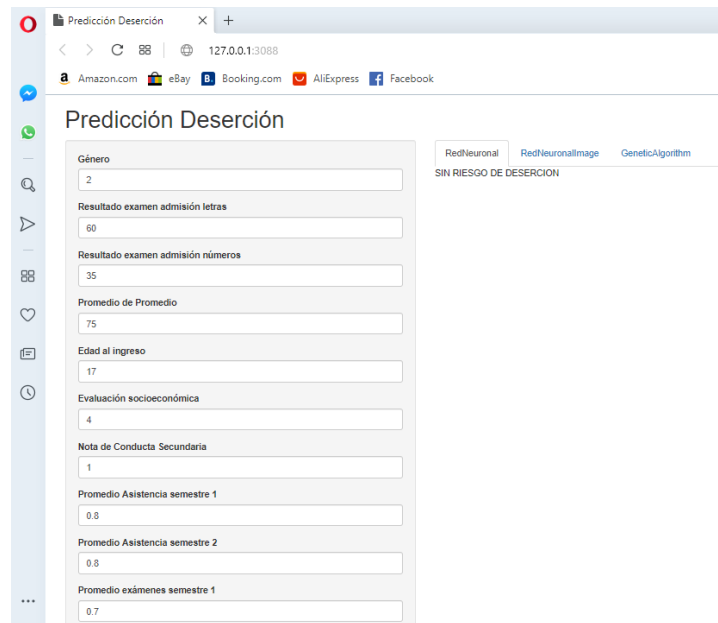


Figura 5. Sistema experto resultante.

La realización de esta investigación apoyándose 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.

TRABAJO A FUTURO

Los datos que forman parte del almacén de datos son los resultantes del EXANNI-II, como resultados del examen de admisión, se continuará con la investigación para integrar datos de calificaciones de los estudiantes.

CONCLUSIONES

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. El conocer el comportamiento e influencia de ciertos factores en la deserción, reprobación o conclusión exitosa de sus estudios en los estudiantes del ITSP, permiten planear con anticipación las estrategias que permitan apoyar en la toma de decisiones para los diferentes departamentos académicos.

La minería de datos permite mostrar datos ocultos que a simple vista es difícil identificarlos, por lo que es útil aplicarse en cualquier ámbito o problemática.

La realización de sistemas expertos permite implementar las reglas obtenidas para indicar en base a los factores identificados el tratamiento ideal para cada estudiante y permitir dar el acompañamiento necesario a los alumnos que lo requieren para que concluyan sus estudios satisfactoriamente.

AGRADECIMIENTOS

Agradezco al Instituto Tecnológico Superior de Pánuco el proporcionarme los datos necesarios para llevar a cabo esta investigación, a través de los diferentes departamentos académicos, que permitieron aportar conocimiento e inteligencia de negocios al ITSP.

Así como al TecNM por el financiamiento proporcionado en la realización de esta investigación.

REFERENCIAS

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>

Aler, R. (2009). Tutorial de Weka. Descargado el 10 de Enero de 2016 en <http://ocw.uc3m.es/ingenieria-informatica/herramientas-de-la-inteligencia-artificial/contenidos/transparencias/TutorialWeka.pdf>

Avendaño O.W. (1990). "La Reprobación Escolar". Ed. DESE. Madrid, España.

Baker, R. (2010). Data mining for education. *International encyclopedia of education*, 7, 112-118.

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

- Britos, P. V. (2005). *Minería de Datos* (1ª. ed.). Buenos Aires: Nueva Librería.
- Castillo M. et al., (2003). Hábitos de estudio, perfil de egreso e índices de reprobación y deserción en alumnos de nuevo ingreso del CBTA Núm. 1. Consultado el 01 de Agosto de 2016 en <http://74.125.47.132/search?q=cache:pfqwZLKQgPQJ:redexperimental.gob.mx/descargar.php%3Fid%3D375+AUIES+%2B+reprobaci%C3%B3n&cd=7&hl=es&ct=clnk&gl=mx>.
- Dapozo, G. N., Porcel, E., López, M. V., Bogado, V. S., & 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.
- 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.
- Díaz Barriga, A. (1993). *El examen: textos para su historia y debate*. CESU/UNAM. México.
- Hadju B. (2012). *Minería de datos en educación*. Consultado el 17 de Marzo de 2016 en <https://biankahajdu.com/2012/01/25/mineria-de-datos-en-educacion/>
- Hernández J., Ramírez M. & Ferri C. (2005). *Introducción a la Minería de Datos*. Ed. Pearson. Madrid España.
- Jiménez, M. (2000). *Competencia social: intervención preventiva en la escuela*. Infancia y Sociedad.
- Martínez-Otero P. V. (1997). *Los a Los adolescentes ante el estudio: causas y consecuencias del r rendimiento académico*. Editorial Fundamentos. Consultado en línea en https://books.google.com.mx/books?id=G_eWnliRpQgC en Marzo 25, 2018.
- Melo-Morín, J. P. Ahumada-Cervantes M. A., Ponce-Rivera, E. (2018). Comparación de los métodos predictivos de minería de datos para determinar los factores del rendimiento académico. *Revista de Tecnología Informática*. Año 1-1: 1-11
- Millon, T. (1997). *Inventario Millon de Estilos de Personalidad*. Buenos Aires: Paidós.
- Muñoz, M.L. (1993). *Estudio comparativo de algunos factores que inciden en el rendimiento académico en una población de estudiantes de niveles medio superior y superior*. México. Tesis de Maestría en Psicología. Universidad Iberoamericana.
- Nghe, N. ; Janecek, P. & Haddawy P. (2007). "A comparative analysis of techniques for predicting academic performance." *Frontiers In Education Conference-Global Engineering: Knowledge Without Borders, Opportunities Without Passports, 2007. FIE'07. 37th Annual. IEEE*.
- 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
- Orea, S. V., Vargas, A. S., & 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.
- Pérez M. (2015). *Minería de Datos a través de ejemplos*. Ed. Alfaomega. México D.F.
- 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).
- Riquelme, J. C., Roberto, R., & Gilbert, K. (2006). *Minería de Datos: Conceptos y Tendencias*. Inteligencia Artificial. *Revista Iberoamericana de Inteligencia Artificial*, primavera , 11-18.
- Romero C. & 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. & Hervás C. (2008). "Data Mining Algorithms to Classify Students." In EDM, pp. 8-17.

Romero C., Ventura S. & Hervas C. (2005). Estado actual de la aplicación de la minería de datos a los sistemas de enseñanza basada en web. Actas del III Taller Nacional de Minería de Datos y Aprendizaje, TAMIDA2005,49, 56.

OCU (2018). Oficina de Cooperación Universitaria (OCU) Universitat XXI - Inteligencia Institucional. Fecha de acceso: 25/03/2018; Disponible en <http://www.ocu.es/>.

Torrado, M. (2011) Minería de datos aplicados a la educación. Barcelona: Universidad de Barcelona. Deposito Digital <http://hdl.handle.net/2445/19862>

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

Witten, I. H., & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

Woolfolk, A. (1995). Psicología Educativa. México: Prentice Hall Hispanoamericana, S.A.