
MINERÍA DE DATOS EDUCATIVOS: UNA PROPUESTA DE INNOVACIÓN EN LA INSPECCIÓN EDUCATIVA.

José Francisco Álvarez Aguilar

Profesor de educación secundaria. Especialidad informática. Córdoba.

Ingeniero Técnico en Informática de Sistemas (UCO).

Grado en ingeniería informática, mención computación (UCO)

Máster Oficial en sistemas inteligentes (UCO)

Resumen

El objetivo del presente trabajo es dar a conocer las diferentes formas de obtención de modelos de predicción basados en datos educativos utilizando técnicas de minería de datos con el fin de intentar predecir el fracaso educativo en etapas tempranas, proponiendo a su vez a la Inspección Educativa la posibilidad de utilizar dichos modelos para poder asesorar con la suficiente antelación a los centros educativos en los que se detecten posibles fracasos educativos

Palabras clave: *Función Inspectora, minería de datos educativos, predicción de fracaso escolar.*

Abstract

The objective of this work is to present the different ways of obtaining prediction models based on educational data using data mining techniques in order to try to predict school failure in early stages, proposing, in turn, to the Educational Inspectorate the possibility of using these models in order to be able to give advice in sufficient time to educational centers where possible school failures are detected.

Keywords: *Inspection function, educational data mining, school failure prediction.*

1. INTRODUCCIÓN

En los últimos años ha surgido una creciente preocupación por el aumento del fracaso escolar en nuestro país debido a los resultados obtenidos en los test de evaluación del alumnado como el informe PISA (Programa Internacional para la Evaluación de Estudiantes),

en el que nuestro país queda muy por debajo de la media internacional, los indicadores homologados de la extinta Agencia Andaluza de Evaluación Educativa (AGAEVE), que comparan cómo de buenos son los resultados de un Instituto de Educación Secundaria con respecto a otros centros del mismo entorno socio-económico y con la media de la provincia, o las pruebas de diagnóstico, enmarcadas dentro del plan de evaluación del sistema educativo andaluz, que evalúa el nivel de conocimiento y habilidades de los alumnos de un nivel educativo en unas materias determinadas.

Además de esto, el creciente abandono escolar en los centros de educación hace que nos preguntemos qué está sucediendo y cómo podríamos anticiparnos a este fracaso.

Debemos empezar por el hecho de que el fracaso escolar existe, y como tal, es nuestra responsabilidad buscar solución al mismo, usando las herramientas que encontremos para atajarlo.

Existen diferentes mecanismos para intentar solucionar el fracaso escolar, pero todos ellos se suelen aplicar generalmente a posteriori. Y es en este punto en el que tal vez debamos hacer uso de modelos de predicción del fracaso escolar.

Puede sonar un tanto fantástico, o arriesgado y pretencioso, el intentar predecir el fracaso escolar, dado que está relacionado íntimamente con la subjetividad propia de cada alumno, pero recordemos que día a día hacemos uso de modelos de predicción, como, por ejemplo, predicción de meteorología, con mayor o menor acierto en diferentes ocasiones, y se hacen teniendo en cuenta una cantidad de datos cada día mayor.

E incluso, en las cadenas de Centros Comerciales aplican técnicas de minería de datos para predecir el comportamiento de los consumidores, basadas en todas las compras que se realizan en todos los centros (contamos con el clásico ejemplo de la cadena Wal-Mart con el hallazgo de la relación entre pañales de bebé y cerveza: encontraron que los compradores de cerveza y pañales eran varones de entre 25 y 35 años, que solían comprar estos productos conjuntamente los viernes por la tarde. Tomaron la decisión de colocar las cervezas cerca de los pañales y aumentaron entre un diez y un quince por ciento tanto las ventas de cervezas como de pañales).

Por otro lado, en la Comunidad Autónoma de Andalucía disponemos de *Séneca*, una Plataforma para la Gestión del Sistema Educativo Andaluz, la cual maneja una cantidad de datos ingente. Tener muchísimos datos puede hacer que la extracción de conclusiones sea una tarea compleja y abrumadora.

2. OBJETIVO

El objetivo del presente trabajo es dar a conocer las diferentes formas de obtención de modelos de predicción basados en datos educativos utilizando técnicas de minería de datos con el fin de intentar predecir el fracaso educativo en etapas tempranas.

2.1 QUIÉN PUEDE HACERLO.

Un experto en educación junto con un ingeniero informático con conocimientos en técnicas de minería de datos.

2.2 QUÉ SE NECESITA.

Datos personales del alumnado y, por evaluaciones, sus calificaciones numéricas y faltas de asistencia, junto con los datos de filiación de Centro. Cuanto más heterogéneos y diversos, mejor.

2.3 QUÉ PODEMOS OBTENER.

Modelos de predicción de fracaso educativo en etapas tempranas.

2.4 QUIÉN PUEDE INTERPRETAR LOS RESULTADOS.

Un experto en educación junto con un ingeniero informático con conocimientos en técnicas de minería de datos.

3. TÉCNICAS

La Minería de Datos es un proceso que intenta descubrir patrones de comportamiento en grandes volúmenes de conjuntos de datos y la posterior interpretación de dichos patrones con el fin de predecir acciones futuras.

La Minería de Datos Educativos es, por tanto, la minería de datos aplicada en el ámbito de la enseñanza. Esta Minería de Datos necesita datos. Estos datos solo nos dicen lo que ya ha sucedido, pero no nos dan un por qué ni un cómo.

La Minería de Datos Educativos proporciona un marco de trabajo en el que, a través de los datos educativos y la aplicación de diferentes algoritmos, poder obtener modelos de comportamiento, siendo en nuestro caso, modelos de fracaso educativo los que nos interesan.

Existen numerosos estudios consistentes en el desarrollo y la aplicación de métodos que exploran la información de las bases de datos de educación y el uso de técnicas que ayudan a comprender mejor a los estudiantes y los entornos en los que aprenden [1, 2].

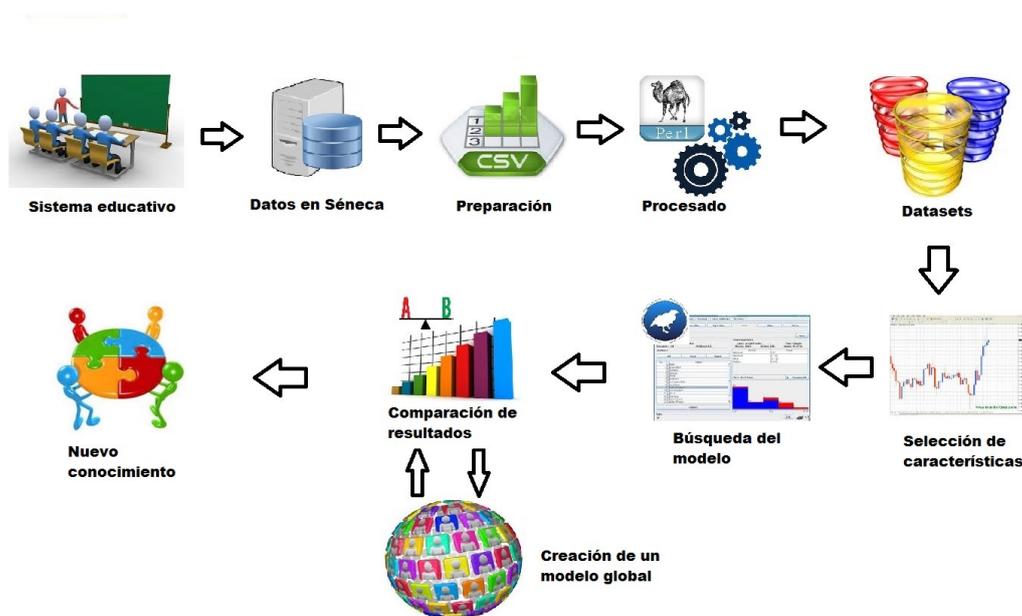


Ilustración 1: Flujo de trabajo de la Minería de Datos Educativos

La Minería de Datos Educativos trata de abordar el problema del abandono escolar [3, 4] o predicción de los resultados académicos [5, 6, 7, 8] por medio de técnicas clásicas de Minería de Datos como la asociación, clasificación, regresión o agrupamiento.

La Minería de Datos Educativos se ha utilizado con éxito en la creación de modelos de predicción del rendimiento escolar y del abandono, pero hay que resaltar que la mayoría de los estudios realizados se han llevado a cabo sobre datos de niveles de educación superior, entornos universitarios [9, 10, 11] o datos generados en entornos de e-learning o sistemas adaptativos [2].

Son pocos los estudios encontrados encaminados a la detección precoz y prevención del fracaso escolar en estudios previos a la educación Universitaria [7, 4, 12], en nuestro país la Educación Secundaria Obligatoria (ESO) [13, 14].

3.1 MÉTODOS.

Para realizar los experimentos para la obtención de modelos de clasificación se debe utilizar el entorno de WEKA, en el que se deben comparar diez algoritmos de distintos tipos: bayesianos, redes neuronales, árboles de decisión, algoritmos basados en reglas...

Los métodos recomendados son los siguientes:

- NaiveBayes: Clasificador Bayesiano probabilístico fundamentado en el teorema de Bayes junto con hipótesis simplificadoras que asumen la independencia entre todas las variables de la muestra. Da como resultado la probabilidad de pertenencia a una clase.
- Sequential Minimal Optimization (SMO): Algoritmo basado en la implementación de las Máquinas vector-soporte (SVMs) consistente en resolver subproblemas del problema inicial.

para disminuir el coste computacional. Es uno de los métodos basado en SVMs más extendido.

- Perceptrón Multicapa (MLP): Red neuronal artificial, evolución del Perceptrón Simple, que resuelve problemas de clasificación no separables linealmente, incluyendo una o varias capas ocultas.

- Radial Basis Function Network (RBFNetwork): Red neuronal artificial con una capa oculta, donde en cada neurona se utiliza una función de base radial con forma gaussiana.

- K-Nearest Neighbor (K-NN) (Ibk): Algoritmo de los K vecinos más cercanos. Este clasificador almacena una tabla con los ejemplos disponibles y la clase a la que pertenecen y ante un nuevo caso a clasificar calcula su distancia con respecto a los ya conocidos y lo clasifica en la misma clase que los k-vecinos más cercanos. En contrapartida a su eficiencia, la implementación Ibk disponible en WEKA, no genera un modelo clasificador

- RIPPER (JRip): Algoritmo basado en el clasificador RIPPER. Implementa una técnica basada en un aprendizaje de reglas proposicionales mediante la repetición de una poda incremental para la reducción del error.

- One Rule (OneR): Algoritmo basado en reglas. Construye y usa un clasificador con una única regla.

- C4.5 J48: Algoritmo usado para generar un árbol de decisión de la misma forma que lo hace ID3, pero usando el concepto de entropía de información.

- PART: Algoritmo que genera una lista de decisión a partir de un árbol C4.5. Desde el paradigma «divide y vencerás» crea un árbol C4.5 de forma parcial, y en cada iteración selecciona la mejor hoja para construir una de sus reglas.

- Adaptive Boosting (AdaBoost) (AdaBoostM1): Clasificador nominal basado en ensemble. Es un meta-algoritmo que trata de obtener un clasificador fuerte mediante la combinación de varios débiles. Puede tener una tendencia al sobreaprendizaje.

3.2 ANÁLISIS ESTADÍSTICO.

Para determinar si existen diferencias significativas entre los resultados de los modelos en cada momento elegido para la predicción se deben utilizar distintas técnicas de análisis estadístico, como el Test de Friedman, y medidas de accuracy, kappa de Cohen, Media Geométrica y test de Wilcoxon.

4. ANÁLISIS DE UN ENTORNO

Al realizar diferentes trabajos de investigación sobre cuatro promociones de la ESO en cuatro centros de secundaria de la campiña cordobesa (IES Francisco de los Ríos, IES Miguel Crespo, IES Profesor Tierno Galván, IES Maestro Eloy Vaquero), recogiendo datos previamente anonimizados, gracias a la colaboración del equipo directivo de los centros, tales como las calificaciones obtenidas, las faltas de asistencia, los retrasos, y el entorno

familiar, al aplicar técnicas de minería de datos educativos basadas en algoritmos de generación de reglas, el modelo de predicción nos arrojó el siguiente patrón de comportamiento: «el alumnado que suspende la asignatura de Lengua Castellana y Literatura en 1º de ESO, tiene un 87% de probabilidad de no obtener el título de la ESO», entre otros muchos patrones, algunos más significativos que otros.

Cabe decir que la regla anterior es una generalización aplicable a los cuatro centros. Generalmente se pueden obtener reglas más complejas, dependiendo del algoritmo usado, al estilo de «el alumnado que suspende las asignaturas de Geografía e Historia, Educación Física y Matemáticas en la 2ª evaluación en 1º de ESO y, además, pertenece a una familia monoparental, no obtiene el título de la ESO».

5. ACTUACIÓN DE LA INSPECCIÓN DE EDUCACIÓN

Un modelo de predicción sin aplicación no sirve de nada. La utilidad reside en aplicarlo a los nuevos alumnos de 1º de ESO, y, si detectamos que hay alumnado que es clasificado como posible fracaso escolar, es en ellos y ellas en los que hay que centrar una actuación pedagógica.

Uniendo estas técnicas con los datos de los que se dispone en Séneca, se podrían obtener modelos de predicción del fracaso escolar enmarcados en diferentes visiones: edad, localidad, entorno familiar, etc.

La información obtenida sería de gran interés, no sólo para el alumnado, sino también para el profesorado, expertos en pedagogía e inspectores de educación, que podrían obtener, con la mayor anticipación, posibles candidatos a abandonar los estudios o a fracasar en su titulación, y medidas para afrontar este problema, conociendo qué, cuándo y a quién tienen que ofrecerle esos recursos de apoyo o refuerzo.

Mi propuesta consiste, por tanto, en que la Inspección Educativa, con la ayuda técnica apropiada, en este caso, un ingeniero en informática con conocimientos en minería de datos, podría obtener diferentes modelos de predicción del fracaso escolar a diferentes niveles (autonómico, provincial, local e incluso por centro educativo) y, haciendo uso de sus funciones de supervisión, evaluación y asesoramiento, hacer partícipes de ese conocimiento al profesorado, emitiendo los informes que estime oportunos para que se puedan acometer, de la manera más efectiva y eficiente posible, los posibles casos de fracaso escolar futuro.

Se preguntaba durante el XX Encuentro Nacional de Inspectores de Educación celebrado en Sevilla en octubre de 2019 Doña Adela Cortina «¿Cómo lograr el progreso moral en la vida contemporánea, marcada por las nuevas tecnologías y neurociencias?». En mi humilde opinión, la tecnología es y debe ser una ayuda, y no debe tomarse nunca como un sustituto. Creo que se puede lograr cierto progreso con la ayuda de la tecnología aplicada correctamente.

Tal y como ya se decía en el preámbulo del Real Decreto de 30 de marzo de 1849, atribuido a Gil de Zárate, «sin la Inspección la Administración nada ve, nada sabe, nada puede remediar». Mi intención, con esta propuesta, es que tanto la Administración como la Inspección pueda ver más, saber más, y, por tanto, remediar más.

CONCLUSIONES.

Aplicando técnicas de Minería de Datos Educativos, la Inspección Educativa podría conseguir las siguientes mejoras en su práctica:

- Enfocado en el eje funcional de Supervisión: se obtendrían datos para la mejora la organización de los centros.
- Enfocado en el eje funcional de Evaluación: se obtendría una visión general y particular del origen del fracaso escolar.
- Enfocado en el eje funcional de Asesoramiento e Información: se podría asesorar al profesorado sobre las predicciones observadas antes de que se produzca el fracaso escolar.

REFERENCIAS

- [1] C. Romero, S. Ventura, M. Pechenizkiy, and R. S. J. Baker. Handbook of Educational Data Mining. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series. Taylor & Francis, 2011. ISBN 9781439804582.
- [2] C. Romero and S. Ventura. Educational data mining: A review of the state of the art. IEEE Transactions on Systems Man and Cybernetics Part C-Applications and Reviews, 40(6):601–618, 2010. Times Cited: 20.
- [3] J. D. Finn and D. A. Rock. Academic success among students at risk for school failure. Journal of Applied Psychology, 82(2):221–234, 1997.
- [4] L.S. Affendey, I.H.M. Paris, N. Mustapha, M.N. Sulaiman, and Z. Muda. Ranking of influencing factors in predicting students' academic performance. Information Technology Journal, 9(4):832–837, 2010.
- [5] P. Cortez and A. Silva. Using data mining to predict secondary school student performance. In Proceeding of the 15th European Concurrent Engineering Conference/5th Future Business Technology Conference, pages 5–12, Porto, Portugal, 2008.
- [6] M.S.B.M.a Azmi and I.H.B.M.b Paris. Academic performance prediction based on voting technique. pages 24–27, 2011.
- [7] P. Gu and Q. Zhou. Student performances prediction based on improved c4.5 decision tree algorithm. Advances in Intelligent and Soft Computing, 146 AISC:1–8, 2012.

- [8] S.A.H.a Basha, A.b Govardhan, S.c Viswanadha Raju, and N.d Sultana. A comparative analysis of prediction techniques for predicting graduate rate of university. *European Journal of Scientific Research*, 46(2):186–193, 2010.
- [9] K. Barker, T. Trafalis, and T. R. Rhoads. Learning from student data. 2004 IEEE Systems & Information Engineering Design Symposium. 2004.
- [10] K. Bunkar, U.K. Singh, B. Pandya, and R. Bunkar. Data mining: Prediction for performance improvement of graduate students using classification. In *Wireless and Optical Communications Networks (WOCN), 2012 Ninth International Conference on*, pages 1–5, Sept 2012. doi: 10.1109/WOCN.2012.6335530.
- [11] S. Anupama Kumar and M.N. Vijayalakshmi. Mining of student academic evaluation records in higher education. In *Recent Advances in Computing and Software Systems (RACSS), 2012 International Conference on*, pages 67–70, April 2012. doi: 10.1109/RACSS.2012.6212699.
- [12] Carlos Márquez-Vera, Cristóbal Romero, and Sebastián Ventura Soto. Predicción del fracaso escolar mediante técnicas de minería de datos. *IEEE-RITA*, 7(3):109–117, 2012. URL <http://dblp.uni-trier.de/db/journals/ieee-rita/ieee-rita7.html#VeraRS12>.
- [13] M.A. Jiménez and Sebastián Ventura. Minería de datos para la detección precoz del fracaso escolar en secundaria. Master's thesis, Universidad de Córdoba, 2013.
- [14] María Teresa Acisclos García and Sebastián Ventura. Predicción del fracaso escolar mediante técnicas de minería de datos: Aplicación a Centros de Secundaria en la Campiña de Córdoba. Master's thesis, Universidad de Córdoba, 20

LEGISLACIÓN.

Orden de 13 de julio de 2007, por la que se desarrolla la organización y el funcionamiento de la inspección educativa de Andalucía. BOJS de 2/08/2007