

Propuesta de un Modelo Predictivo utilizando Aprendizaje Profundo para el análisis de deserción estudiantil en Universidades Colombianas Virtuales

Martínez, Julio César^{1(*)}; Mateus, Sandra Patricia¹

¹Politécnico Colombiano Jaime Isaza Cadavid, Facultad de Ingeniería, Medellín, Colombia

RESUMEN

La deserción estudiantil es una realidad en el país y es un fenómeno complejo. En este artículo, se propone un modelo predictivo que sirva como apoyo a las Universidades colombianas para el análisis de la deserción en estudiantes, principalmente, en programas de pregrado en modalidad virtual. Un modelo predictivo, puede ayudar a las organizaciones a generar ganancias y evitar pérdidas futuras, tomando datos históricos y con ellos arrojar resultados esperados para ser analizados y apoyar en las decisiones. Este modelo se desarrolla, tomando eventos históricos con distintas variables de tipo social, académico, personal, laboral, ingresos a las plataformas e-learning, etc. y posteriormente, a estas variables se le aplican algoritmos de aprendizaje profundo. Se espera la predicción de probabilidades de deserción de cada estudiante, posteriormente, con esta información se puede alertar y aplicar medidas preventivas tempranas con la población estudiantil.

Palabras clave: aprendizaje Profundo; instituciones educativas; e-Learning; deserción estudiantil; modelos predictivos.

Proposal of a predictive model using deep learning for dropout student analysis in virtual Colombian Universities

ABSTRACT

Student dropout is a reality and a complex phenomenon in the country. In this paper, a predictive model is proposed that serves as a support to the colombian universities for dropout analysis on students, main, on virtual undergraduate programs. A predictive model, can help to organizations generate profits and avoid future losses, taking historical data so show results expected to be analyzed and support on decisions. This model is developed, taking historical events with different variables of type: social, academic, personal, work, login to e-learning platforms, etc, next, this variables deep learning algorithms are applied. The dropout probabilities prediction for each student is expected, then, can be alerted to apply early preventive measures with the student population

Keywords: deep learning; educational institutions; electronic learning; student dropout; predictive models.

Recibido: 13/04/2020 - Aceptado: 29/04/2020
Correspondencia: juliomartinez54172@elpoli.edu.co

INTRODUCCIÓN

Según Mortis y Lozoya (2005), en los programas educativos virtuales se presenta un alto índice de deserción. Estos autores afirman, que esta problemática es una de las grandes preocupaciones de investigadores, planificadores educacionales y entidades educativas que desarrollan o desean desarrollar programas basados en la modalidad virtual. Estos autores, también definen, cuatro categorías de deserción que pueden ser identificadas en la virtualidad: primero, están los alumnos que se matriculan en programas pero que no participan; segundo están los alumnos que se retiran luego de algún tiempo; en tercer lugar, se encuentran los estudiantes que participan, pero no cumplen con todas las tareas, lo cual los inhabilita para presentar los exámenes finales; y por último, están los alumnos que cumplen y participan cumpliendo con sus tareas, pero reprueban cursos.

Con base en lo anterior, se han creado e implementado modelos predictivos en algunas Instituciones de Educación Superior, utilizando distintas técnicas y algoritmos como, por ejemplo: minería de datos, árboles de decisión, el algoritmo ID3 (en inglés, Induction Decision Trees), etc. Amaya y Barrientos (2014), proponen un modelo predictivo de deserción estudiantil utilizando técnicas de minería de datos, caracterizando a los estudiantes de una universidad con el objetivo de poder predecir la probabilidad de deserción de los estudiantes con diferentes algoritmos para clasificar variables. Por su parte Quintana y Trinidad (2013), presentan una comparación de algoritmos de clasificación para la predicción de la deserción escolar en un centro universitario y con ello, determinar el algoritmo más adecuado, basados en la precisión de la clasificación. Patiño y Cardona (2012), desarrollan un estudio sobre los niveles de deserción en Colombia y Latinoamérica para identificar los factores que inciden en el aumento de deserción en las universidades, pero no establece un modelo para predecir dicha deserción e investiga sobre deserción en modalidad tradicional, no virtual. En este sentido, Cuji y Gavilanes (2017), proponen la construcción de un modelo para pronosticar la probabilidad de abandono por parte de los estudiantes de su programa académico universitario, pero no son estudiantes en modalidad virtual y utiliza técnicas de clasificación basadas en árboles de decisión. (Vila y Granda, 2018) realizan un trabajo semejante en una Universidad de Ecuador, pero usando técnicas de minería de datos. Y finalmente, Forero y Piñeros (2019), muestran un

caso de estudio de identificación de ingeniería industrial en riesgo de deserción en ciertos semestres académicos, técnicas de aprendizaje automático.

Por su parte dentro de estos modelos predictivos, se encuentran las técnicas de aprendizaje profundo supervisado, la cual consiste en emular el aprendizaje humano con el fin de obtener ciertos conocimientos a partir de un conjunto de datos ejemplos, un conjunto de entrenamiento es enviado como entrada al sistema durante la fase de entrenamiento, cada entrada es etiquetada con un valor de salida deseado y de esta forma dicho sistema sabe cómo es la salida cuando llega una entrada; y es la que se implementa en este trabajo, dado que no todas las Instituciones de Educación Superior, cuentan con una modelo de este tipo, por lo que se diseñará una propuesta que ayude a la toma de decisiones en la prevención de deserción estudiantil y que sirva para analizar este fenómeno.

Este artículo está organizado de la siguiente manera: en la sección 2, se presenta el marco referencial; luego en la sección 3, la metodología; y, por último, las conclusiones.

MARCO CONCEPTUAL

En este apartado se presenta un acercamiento general de los conceptos vinculados al desarrollo del presente artículo. Se inicia introduciendo el concepto de deserción estudiantil, siguiendo con lo que es un modelo predictivo; por último, el concepto de aprendizaje profundo y tensorflow.

Modelos predictivos

Los modelos predictivos son modelos de la relación entre el rendimiento específico de una unidad en una muestra y uno o más atributos o características conocidos de la unidad. El objeto del modelo es evaluar la probabilidad de que una unidad similar en una muestra diferente exhiba un comportamiento específico. Esta categoría abarca modelos que se encuentran en muchas áreas, como el marketing, donde buscan patrones de datos ocultos para responder preguntas sobre el desempeño del cliente, como los modelos de detección de fraude. Los modelos predictivos a menudo, ejecutan cálculos durante las transacciones en curso, por ejemplo, para evaluar el riesgo o la oportunidad de un cliente o transacción

en particular, de forma que aporte conocimiento a la hora de tomar una decisión. Con los avances en la velocidad de computación, los sistemas de modelado de agentes individuales han sido capaces de simular el comportamiento humano o reacciones ante estímulos o escenarios específicos (Espino, 2017).

Deserción

Uno de los primeros desafíos que presenta una investigación sobre deserción es la conceptualización y delimitación del fenómeno y del sujeto que deserta, esto es, el desertor. La literatura ofrece diferentes conceptualizaciones que representan paradigmas o filosofías diferentes sobre la problemática (Rodríguez y Londoño, 2011).

Tinto (1989), uno de los autores más referenciados en la literatura sobre deserción, explica la manera diversa como la denominación de desertor ha venido siendo tratada por diferentes autores, a tal punto de llegar a influenciar directamente el análisis de las causas y el carácter real de la deserción en la educación superior. Así, mientras que para unos estudiosos la probabilidad de desertar guarda una relación directa con la poca capacidad del estudiante, para otros es lo contrario, o sea, la probabilidad tiende a ser más alta mientras más brillante sea el estudiante.

Rodríguez y Londoño (2011), también mencionan los tipos de abandono o deserción de la siguiente manera: Con respecto al tiempo, al espacio y al individuo.

- Con respecto al tiempo: precoz, temprana y tardía.
- Con respecto al espacio: institucional; al mismo tiempo: externa o total, interna (migración interna a la IES)
- Con respecto al individuo: estudiantiles, académicos.

En (Díaz, 2008), se resume los determinantes de la deserción estudiantil, como se muestra en la Figura 1.



Figura 1. Estado del arte de los determinantes de la deserción estudiantil según (Díaz, 2008)

Aprendizaje Profundo

El aprendizaje profundo, es una técnica basada en Redes Neuronales Artificiales. Está surgiendo en los últimos años, como una herramienta poderosa para el aprendizaje automático, que promete remodelar el futuro de la Inteligencia Artificial. Las rápidas mejoras en el poder computacional, el almacenamiento rápido de datos y la paralelización también han contribuido a la rápida incorporación de la tecnología, además de su poder predictivo y su capacidad, para generar características de alto nivel optimizadas automáticamente e interpretación semántica a partir de los datos de entrada (Ravi, 2017).

Tensorflow

TensorFlow™ es una biblioteca de software libre propiedad de Google, que se utiliza para realizar cálculos numéricos mediante diagramas de flujo de datos. Los nodos de los diagramas representan operaciones matemáticas y las aristas reflejan las matrices de datos multidimensionales (tensores) comunicadas entre ellas (Tensorflow, 2020).

METODOLOGÍA

A continuación, se mencionan los conceptos y características más importantes ligadas al modelo para entender su proceso de desarrollo:

Arquitectura de alto nivel de la propuesta

En la Figura 2, se muestra una vista de la propuesta de arquitectura tecnológica de alto nivel de la propuesta a desarrollar.

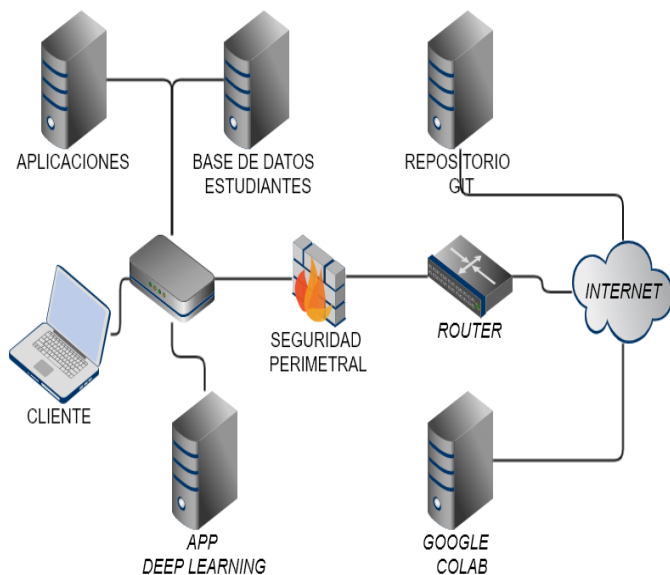


Figura 2. Arquitectura tecnológica de alto nivel propuesta

En la anterior Figura, se puede ver la relación entre los componentes físicos y lógicos de la propuesta del modelo y su comunicación con los demás sistemas de la Institución de Educación Superior Virtual, en donde se puede relacionar el funcionamiento de lo que será al implementarlo. Se cuenta con un servidor de aplicaciones y base de datos en donde se encuentran los datos del Sistema Gestor de Aprendizaje (LMS, Learning Manager System) y una base de datos de los estudiantes con su información completa lo cual es necesario; un servidor para el versionamiento del código para la propuesta y Google Colab para probar y ejecutar código en la nube. El servidor de aprendizaje profundo, aloja la aplicación, que a su vez cuenta con servicios web que se “consumirán” para solicitar y recibir información al respecto.

Roadmapping

En la Figura 3, se presenta un Roadmapping de la propuesta; un Roadmapping es una herramienta de planeación de proyectos para estimar los recursos con los que se cuenta o faltan para analizar la viabilidad.

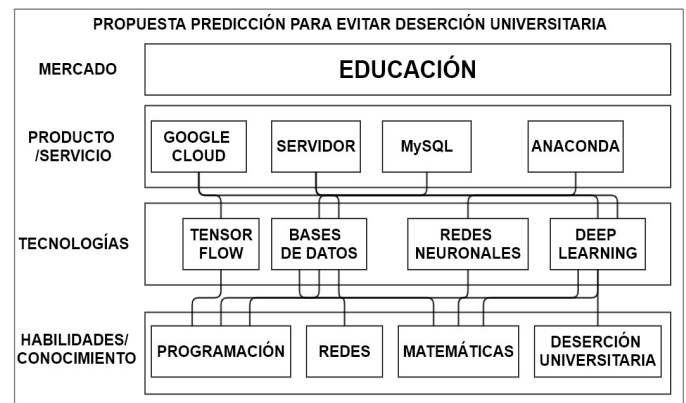


Figura 3. Roadmapping planificación del proyecto

En esta Figura se relacionan: el mercado al que se expondrá el producto final, el producto / servicio, las tecnologías a usar y las habilidades o conocimientos necesarios para realizarlo.

Modelo de deserción

Existen tipos de datos categóricos, por ejemplo: género, estado civil, situación laboral, atención del tutor hacia los estudiantes, percepción del estudiante hacia el tutor virtual, entre otros; y numéricos (continuos y discretos), por ejemplo: edad, peso, número de hijos, número de hermanos, número de veces de ingreso al campus virtual semanal (estudiante y tutor), número de visitas semanales promedio a páginas en el campus virtual, tiempos de respuesta promedio del tutor al estudiante, etc. (Díaz, 2008)

- Las variables o atributos que se extraerán y se tendrán en cuenta para el modelo de aprendizaje profundo son de tipo numérico, Factores Individuales: edad.
- Factores académicos: promedio estudiantil, promedio ICFES para ingreso, ingreso al campus virtual semanal por parte del estudiante.
- Factores internos a la Universidad: costo semestral.
- Factores socioeconómicos: estrato, ingresos económicos.

Se toman factores de tipo numérico para realizar la propuesta inicial del modelo de predicción en el que las “variables de entrada” o etiquetas sean numéricas para emplear un modelo de Aprendizaje

je profundo Supervisado de Regresión, en el que se obtengan dígitos como resultado. Esto no excluye la forma en transformar variables no numéricas a un valor numérico para procesamiento, pero como propuesta inicial se realizará con el tipo de valores en mención.

Los factores tecnológicos como: acceso a internet y familiaridad con las TIC, son muy importantes en el presente contexto, pues se trata de educación virtual, pero no se tendrá en cuenta como variable en el modelo. Otros factores que no se incluirán en esta propuesta inicial, pero que son importantes son: estado civil y situación laboral de los padres.

Lo anterior, se resume en la Figura 4.

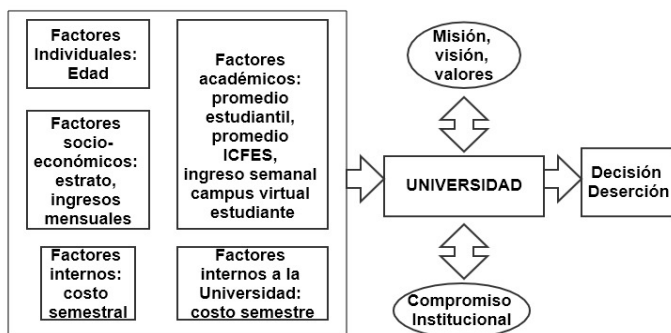


Figura 4. Modelo análisis de deserción propuesto

A continuación, se muestra también la Figura 5, donde se puede apreciar de forma ilustrativa el proceso del modelo de aprendizaje profundo utilizado para el modelo predictivo propuesto, aquí se define la relación entre las “características o atributos” y las “etiquetas”; se hace la definición del objetivo, normalización de datos que existan a lugar; a continuación la creación de un conjunto de datos de entrenamiento y de pruebas, para llegar a la creación del modelo, el entrenamiento y finalmente las predicciones a partir de los subprocessos anteriores.

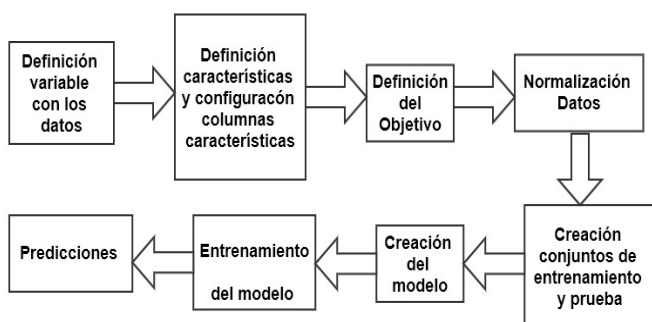


Figura 5. Proceso de la propuesta Modelo de aprendizaje profundo

Los resultados de las predicciones halladas por el modelo con las variables que se mencionaron anteriormente y con el procesamiento de datos históricos, se tienen en cuenta para tomar una decisión con los estudiantes que de una u otra forma obtengan probabilidades altas de deserción, no dejando de lado a los que tengan menos probabilidades, puesto que el fenómeno de deserción no se considera determinístico. Un estudiante con un buen promedio, ingresos continuos al campus, cumplimientos en sus actividades, acompañamiento docente permanente, puede llegar a desertar en cualquier momento por alguna condición o situación repentina; también puede ocurrir un caso totalmente contrario. Las predicciones servirán de apoyo al área correspondiente en las universidades, por ejemplo, bienestar estudiantil y demás dependencias que de forma directa atienden este tipo de situaciones, para posibles ayudas mediante programas de prevención, atención y seguimiento continuo a su situación y así mitigar la tasa de deserción estudiantil en los individuos que desde diferentes lugares asisten de forma virtual a sus clases.

CONCLUSIONES

Se ha diseñado la propuesta de un modelo de aprendizaje profundo para la predicción de deserción estudiantil en las instituciones de educación superior en modalidad virtual. Con datos históricos sobre situaciones de deserción o permanencia estudiantil en la Institución, se puede establecer aproximaciones sobre predicciones que describan este fenómeno. Un modelo predictivo que analiza fenómenos de deserción estudiantil puede servir de ayuda a la comunidad estudiantil y al ente educativo.

El modelo pretende analizar y hacer recomendaciones, con el objetivo de apoyar la toma decisiones apropiadas y a tiempo con respecto a la tasa de deserción estudiantil principalmente en el ámbito de la educación virtual en instituciones de educación superior, siendo un proyecto continuo y a largo plazo. Contribuyendo con los estudiantes, a que tengan una mejor calidad de vida y rendimiento académico. Dependiendo de las predicciones que arroje el modelo sobre los estudiantes, se podrán tomar medidas preventivas tempranas por parte de bienestar estudiantil y demás dependencias de la institución que estén relacionadas con este proceso, en lo que se pueden promover programas de refuerzo, créditos estudiantiles, charlas, entre otras.

Se puede realizar un afinamiento del modelo incluyendo atributos adicionales tanto de tipo categórico como numérico. Y mejorar la arquitectura de la red neuronal, asignando capas adicionales; y con esto lograr resultados de mayor precisión y más cercano a la realidad. Incluso, se puede comparar con un modelo utilizando otra técnica de inteligencia artificial y comparar los resultados y la eficiencia.

REFERENCIAS

- Díaz, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. (U. C. Concepción, Ed.) Estudios Pedagógicos, XXXIV. <http://dx.doi.org/10.4067/S0718-07052008000200004>
- Quintana, M., Trinidad, J., & Morales, S. (2013). Análisis Comparativo de Algoritmos de Minería de Datos para Predecir la Deserción Escolar. *Advances in Computing Science* 67, 13-22.
- Ravi, D. (2017). Deep Learning for Health Informatics. *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 4-21. <https://doi.org/10.1109/JBHI.2016.2636665>
- Vila, D., Cisneros S., Granda, P., Ortega, C., Posso-Yépez M., & García-Santillán, I. (2019) Detection of Desertion Patterns in University Students Using Data Mining Techniques: A Case Study. (eds) *Technology Trends. CITT 2018. Communications in Computer and Information Science*, vol 895. https://doi.org/10.1007/978-3-030-05532-5_31
- Zea L.D.F., Reina Y.F.P., & Molano J.I.R. (2019) Machine Learning for the Identification of Students at Risk of Academic Desertion. (eds) *Learning Technology for Education Challenges. LTEC 2019. Communications in Computer and Information Science*, vol 1011. https://doi.org/10.1007/978-3-030-20798-4_40