

41

MODELO PREDICTIVO DEL PROGRESO EN EL APRENDIZAJE DE LOS ESTUDIANTES DE UNIMINUTO APLICANDO TÉCNICAS DE MACHINE LEARNING

PREDICTIVE MODEL OF THE LEARNING PROGRESS OF UNIMINUTO STUDENTS APPLYING MACHINE LEARNING TECHNIQUES

Elmer Bautista Cañón¹

E-mail: elmer.bautista@uniminuto.edu

ORCID: <https://orcid.org/0000-0001-9476-8106>

Jenny E. Quirama Salamanca¹

E-mail: jquirama@uniminuto.edu

ORCID: <https://orcid.org/0000-0002-0800-0043>

Edilfonso Bautista Cañón¹

E-mail: edilfonso.bautista@uniminuto.edu

ORCID: <https://orcid.org/0000-0002-6191-3829>

¹ Corporación Universitaria UNIMINUTO. Colombia.

Cita sugerida (APA, séptima edición)

Bautista Cañón, E., Quirama-Salamanca, J. E., & Bautista Cañón, E. (2021). Modelo predictivo del progreso en el aprendizaje de los estudiantes de UNIMINUTO aplicando técnicas de machine learning. *Revista Conrado*, 17(83), 305-310.

RESUMEN

Esta investigación tiene como propósito validar la hipótesis donde se propone explicar la relación del progreso del aprendizaje de los estudiantes de UNIMINUTO en las pruebas de salida, con respecto a las variables de las pruebas de entrada a la educación superior, a partir del desarrollo de un modelo predictivo de aprendizaje supervisado. Para el desarrollo del proyecto se estableció una metodología con 4 fases, análisis descriptivo de los datos para 10505 instancias con 69 variables, análisis predictivo, análisis prescriptivo y aplicaciones del modelo. Como resultado se desarrollaron tres modelos predictivos con los algoritmos de regresión logística, máquinas de vector soporte y redes neuronales, los cuales mostraron una eficiencia cercana al 75% y la evaluación de las métricas de precisión, recall y f1, con porcentajes de eficiencia similares. En conclusión, se logró desarrollar 3 modelos predictivos del progreso del aprendizaje de los estudiantes de UNIMINUTO, a partir de la información suministrada por el ICFES, que relaciona las variables de entrada con las variables de salida de la educación superior en Colombia.

Palabras clave:

Aprendizaje supervisado, modelo predictivo, regresión logística, máquinas de vector soporte, redes neuronales, valor agregado educación.

ABSTRACT

The purpose of this research is to validate the hypothesis where it is proposed to explain the relationship of the learning progress of UNIMINUTO students in the exit tests, with respect to the variables of the entrance tests to higher education, based on the development of a supervised learning predictive model. For the development of the project, a methodology with 4 phases was established, descriptive analysis of the data for 10,505 instances with 69 variables, predictive analysis, prescriptive analysis and applications of the model. As a result, three predictive models were developed with logistic regression algorithms, support vector machines and neural networks, which showed an efficiency close to 75% and the evaluation of the precision metrics, recall and f1, with similar percentages of efficiency. In conclusion, it was possible to develop 3 predictive models of the learning progress of UNIMINUTO students, based on the information provided by ICFES, which relates the input variables with the output variables of higher education in Colombia.

Keywords:

Supervised learning, predictive model, logistic regression, support vector machines, neural networks, value added education.

INTRODUCCIÓN

El Ministerio de Educación Nacional de Colombia (MEN) ha implementado las pruebas estandarizadas para medir los niveles de aprendizaje de entrada y salida en la educación superior, con el fin de determinar el valor que agregó la Institución de Educación Superior (IES) al proceso de aprendizaje de sus estudiantes. El rol que cumplen los Estados, ya sea promoviendo los cambios y/o controlando la calidad con la que se ofrece la IES, constituyen tendencias que conforman el marco bajo los que surgen los procesos de evaluación de la calidad de los procesos universitarios (Feitó Cespón & Pérez de Armas 2018).

La comprobación y evaluación de los conocimientos y las habilidades, es denominado como evaluación del aprendizaje (De La Hoz, et al., 2019; Arteaga Valdés, et al., 2018). La formulación de las preguntas de las pruebas estandarizadas utiliza categorías de las dimensiones cognitivas de la taxonomía de Anderson & Krathwohl (2001), en orden ascendente son: recordar, comprender, aplicar, analizar, evaluar y crear (Cunha & Miller, 2014; Contreras Bravo, et al., 2021; Minte Münzenmayer, et al., 2021).

El valor agregado se puede definir como la diferencia entre el resultado de aprendizaje obtenido y lo que se espera podría obtener el estudiante en función de sus características personales (familiares, contexto). Esto ha llevado a generar grandes cantidades de información de los estudiantes de las instituciones de educación superior - IES (Touria, et al., 2021), la cual el Instituto Colombiano para la Evaluación de la Educación (ICFES), comparte de manera periódica con las IES. El principal desafío para las IES es analizar los datos disponibles y extraer información útil que genere nuevo conocimiento para la gestión académica utilizando herramientas tecnológicas como la minería de datos (Kabakchieva, 2013). El análisis de big data que utilizan técnicas de ML puede ayudar a los usuarios a detectar patrones o anomalías (Lokanan, et al., 2019). De igual manera, el desarrollo de modelos predictivos con diferentes algoritmos de ML, permiten establecer proyecciones, con las cuales se pueden identificar oportunidades de mejora. Los modelos se evalúan mediante métricas, para establecer la eficiencia de cada algoritmo (Jara, et al., 2011). Los modelos basados en ML extraen expresiones matemáticas o encuentran relaciones empíricas entre las variables de entrada y salida, los cuales han sido ampliamente utilizado modelos predictivos (Shamshirband, et al., 2020).

Este artículo se centra en el desarrollo de un modelo predictivo del progreso en el aprendizaje de los estudiantes de UNIMINUTO en las competencias de razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés, dentro de las pruebas estandarizadas de salida de la educación superior, a partir de la información suministrada por el ICFES, aplicando técnicas de ML con tres algoritmos de aprendizaje automático, donde se evalúan las métricas de eficiencia.

El objetivo principal de este artículo es describir la metodología para desarrollar modelos predictivos asociados a los procesos de enseñanza-aprendizaje en las IES, que permita, desde la gestión académico-administrativa de los programas, llevar a cabo una intervención temprana y focalizada, a partir de las características psicosociales de los estudiantes, logrando así, el cierre de las brechas de los aprendizajes.

El artículo está dividido en cuatro fases de acuerdo a la metodología propuesta. La primera fase es el análisis descriptivo de los datos que han sido suministrados por el ICFES. La segunda fase es el análisis predictivo, donde se determinan las variables dependientes e independientes a utilizar en los modelos, se seleccionan los modelos de aprendizaje supervisado (regresión logística, máquinas de vector soporte y redes neuronales), se realiza en entrenamiento y pruebas de los modelos, así como la evaluación de las métricas seleccionadas (accuracy, precisión, recall y f1). La tercera fase es el análisis prescriptivo donde se evalúan los modelos y se plantean las posibles estrategias a seguir para cerrar la brecha de aprendizaje de los estudiantes, de acuerdo a los resultados de los modelos. La cuarta fase es la aplicación del modelo en actividades de gestión académico administrativa en UNIMINUTO.

MATERIALES Y MÉTODOS

En la investigación se realiza un estudio piloto experimental de tipo descriptivo, correlacional y predictivo, donde se describen las tecnologías aplicadas en cada etapa del proyecto, de acuerdo con la metodología propuesta, que para el caso son: análisis descriptivo de los datos (resultados en las pruebas estandarizadas saber pro 2016-2018 y la prueba saber 11 correspondiente a cada estudiante de UNIMINUTO); análisis predictivo con la aplicación de las técnicas de aprendizaje supervisado de ML con los algoritmos de regresión logística, máquinas de vector soporte (SVM siglas en inglés) y redes neuronales (RRNN), para determinar mediante las métricas establecidas, el modelo predictivo más eficiente para predecir el progreso en el aprendizaje de los estudiantes de UNIMINUTO; análisis prescriptivo donde se realiza la evaluación de los resultados del modelo predictivo y se orientan estrategias para cerrar la brecha del aprendizaje de los estudiantes de UNIMINUTO; finalmente las recomendaciones para aplicar este tipo de modelos en los procesos de gestión académico-administrativa a nivel programa académico.

Durante el desarrollo de la investigación se utilizaron una serie de herramientas tecnológicas para el análisis descriptivo, análisis predictivo y análisis prescriptivo del proyecto. La primera fase de la metodología propuesta para la investigación hace referencia al análisis descriptivo de los datos en las pruebas estandarizadas por MEN para la entrada y la salida de la educación superior. La información suministrada por el ICFES se refiere a los resultados

de los estudiantes de UNIMINUTO en las competencias de razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés. A continuación, en la tabla 1, se presentan las características del conjunto de datos de los resultados de las pruebas estandarizadas de entrada (saber 11) y salida (saber Pro) de los estudiantes de UNIMINUTO.

La hipótesis que se plantea en la presente investigación es de tipo causal predictiva, donde se propone explicar la relación del progreso del aprendizaje de los estudiantes de UNIMINUTO en las pruebas de salida, con respecto a las variables de las pruebas de entrada a la educación superior, para anticipar los resultados que tendrán los estudiantes que se encuentran desarrollando sus estudios y con ello, llevar a cabo una intervención temprana y focalizada, logrando así, el cierre de las brechas de los aprendizajes.

RESULTADOS Y DISCUSIÓN

En la tabla 1, se presentan las características del conjunto de datos de los resultados de las pruebas estandarizadas de entrada (saber 11) y salida (saber Pro) de los estudiantes de UNIMINUTO.

Tabla 1. Características del conjunto de datos Saber 11 vs Saber Pro UNIMINUTO.

Año Data	Atributos (observaciones)	Instancias (variables)	Tipo variables	Distribución variables			
				char	num	PO-SIXct	Logi
2018	4278	69	Multivariado	45	23	1	0
2017	3598	69	Multivariado	45	23	1	0
2016	2629	69	Multivariado	45	23	1	0

En la figura 1 se muestra la distribución de los resultados por competencia (razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés) en las pruebas de entrada (saber 11) y de salida (saber pro - 2018). Para el análisis los resultados han sido normalizados en una escala de 0 a 100, donde 50 puntos es el referente de la media nacional.

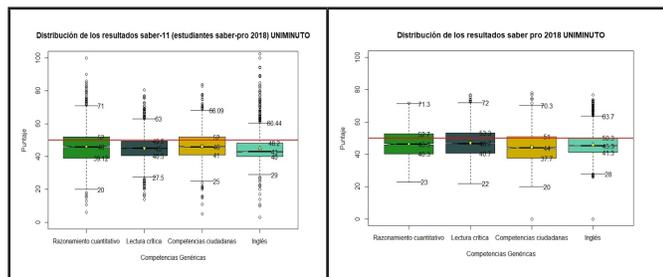


Figura 1. Distribución de los resultados por competencia saber 11 y saber pro 2018 UNIMINUTO.

En las figuras 2 y 3 se muestran los resultados correspondientes a los años 2017 y 2016 respectivamente.

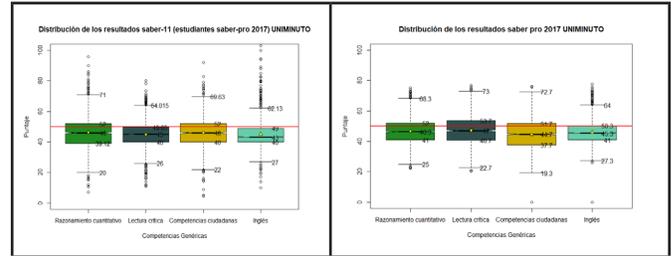


Figura 2. Distribución de los resultados por competencia saber 11 y saber pro 2017 UNIMINUTO.

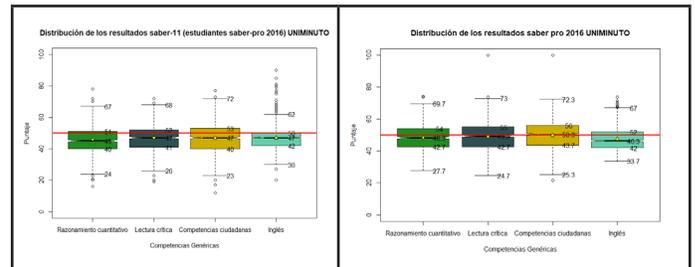


Figura 3. Distribución de los resultados por competencia saber 11 y saber pro 2016 UNIMINUTO.

De la información se puede evidenciar que hay una tendencia en las pruebas de entrada y de salida en cada competencia, de estar por debajo de la media nacional (50 puntos – línea color rojo). De igual manera para las pruebas saber pro se observan que todos los tercetos cuartiles alcanzan la media nacional.

Para el análisis predictivo se desarrolla a partir de la medición del crecimiento en las pruebas saber pro, modelados índices, definición de variables para los modelos predictivos,

Para el análisis predictivo, inicialmente se determinó el crecimiento en las pruebas saber pro, el cual se puede describir como la probabilidad de que un estudiante se encuentre en la zona de crecimiento como se define en la ecuación (1)

$$p(\text{Zona de Crecimiento}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \dots + \beta_k X_k)}} \quad (1)$$

Donde X_i para $i=1, \dots, k, k \in \mathbb{N}$, representa un factor significativo del modelo, es decir que el crecimiento del estudiante i sea mayor que uno ($C_i > 1$) como se indica en la ecuación (2)

$$C_i = (Psproi, Ps11i) = \frac{Psproi}{Ps11i} \quad (2)$$

C_i : crecimiento del estudiante

$Psproi$: transformación del resultado del estudiante en la prueba saber pro.

$Ps11i$: transformación del resultado del estudiante en la prueba saber 11.

Este modelo se aplicó para determinar la distribución de crecimiento en los resultados de las pruebas saber pro 2016 - 2018 de los estudiantes de UNIMINUTO. Los resultados del crecimiento en las pruebas saber pro 2018 se muestran en la figura 4.

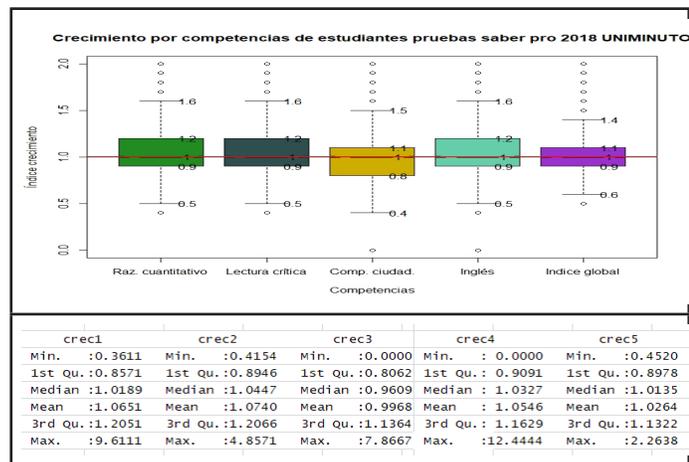


Figura 4. Resumen distribución de crecimiento por competencias pruebas saber pro 2018 UNIMINUTO.

A continuación, se aplica el modelo de crecimiento en los resultados de las pruebas saber pro 2016 - 2018 de los estudiantes de UNIMINUTO por grupos de referencia. En la figura 5, se muestra la distribución de crecimiento de la competencia razonamiento cuantitativo por grupo de referencia para el año 2018.

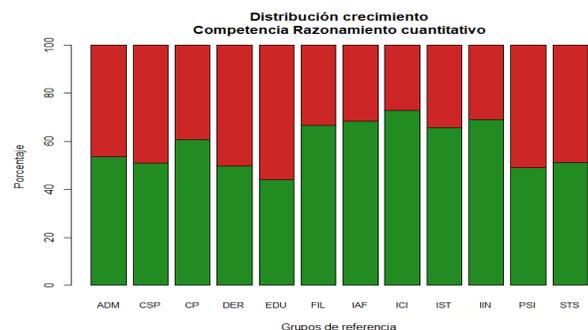


Figura 5. Distribución de crecimiento competencia cuantitativa por grupos de referencia 2018

A partir de las variables sociodemográficas, socioeconómicas, familiares, académicas y el índice socioeconómico (INSE) determinado por el ICFES, por cada estudiante se determinaron índices, a partir de las metodologías MCA (Análisis de Correspondencias Múltiples) y PCA (Análisis de Componentes Principales), utilizando R (entorno y lenguaje de programación con enfoque al análisis estadístico) para ser utilizados en los modelos de aprendizaje automático, los cuales no funcionan con variables cualitativas.

A continuación, se definen las variables para los modelos predictivos, consolidando la información con un total de 10505 atributos con 16 instancias independientes y 5 instancias dependientes como entrada al sistema y, que se utilizarán para los modelos de

aprendizaje automático (regresión logística, redes neuronales y máquinas de vector soporte).

Se realiza el entrenamiento y prueba de los modelos predictivos de aprendizaje automático de regresión logística, máquinas de vector soporte y redes neuronales. Para efectos de medición de la eficiencia, se definen las métricas de precisión, recall, F1 y análisis ROC.

La regresión logística es una técnica de clasificación de ML común para resolver problemas de clasificación binaria. Es un método estadístico cuyo resultado o variable objetivo es de naturaleza dicotómica.

La competencia con mejores p valores (menores al 5%), son las competencias ciudadanas y el puntaje global con 12 variables cada una; les sigue lectura crítica con 10 variables y la de menores valores es la de inglés con 8 variables relevantes. Las máquinas de vector soporte se utilizan para la solución de diferentes tipos de problemas como los de clasificación. En este proyecto se utiliza la técnica de máquinas de vector soporte para diseñar un modelo a partir de la data que contiene 8815 observaciones y 16 variables independientes.

El modelo se configura separando los datos de entrenamiento (train set) con un porcentaje de la data del 70% y prueba (test set) con un porcentaje de la data del 30%. Una vez separada la data, se entrena el modelo (fit) con los datos de entrenamiento (x_train, y_train) para realizar una clasificación mediante la clasificación de vectores de soporte (svc en svm de Scikit-Learn), utilizando el kernel linear. Las redes neuronales se utilizan para la solución de diferentes tipos de problemas como los de clasificación. En este proyecto se utiliza la técnica de redes neuronales para diseñar un modelo predictivo.

El comportamiento de las métricas para lectura crítica muestra una precisión para el 0 del 78% y para el 1 del 71%; la métrica de recall muestra para el 0 un valor del 75% y para el 1 un valor del 74%; para el F1 score el valor para el 0 es del 77% y el valor para el 1 es del 72%. Las métricas para las competencias ciudadanas muestran una precisión para el 0 del 75% y para el 1 del 75%; la métrica de recall muestra para el 0 un valor del 82% y para el 1 un valor del 66%; para el F1 score el valor para el 0 es del 79% y el valor para el 1 es del 70%. Las métricas para inglés muestran una precisión para el 0 del 79% y para el 1 del 80%; la métrica de recall muestra para el 0 un valor del 76% y para el 1 un valor del 81%; para el F1 score el valor para el 0 es del 77% y el valor para el 1 es del 81%. Las métricas para el puntaje global muestran una precisión para el 0 del 71% y para el 1 del 81%; la métrica de recall muestra para el 0 un valor del 80% y para el 1 un valor del 72%; para el F1 score el valor para el 0 es del 75% y el valor para el 1 es del 76%.

La evaluación de las métricas precisión, recall y F1 score de los modelos predictivos, nos permite establecer cual es la técnica de ML más eficiente entre regresión logística, redes neuronales y máquinas de vector soporte. La data de entrada a los modelos tiene 8815 observaciones, 16 variables independientes y 5 variables dependientes (dicotómicas) (Tabla 2).

Tabla 2. Resultados métricas por competencia modelo redes neuronales.

Técnica	Competencia	Salida	Accu.	Precision	Recall	F1	
Regresión logística	Razonamiento cuantitativo	0	0.76111	0.76000	0.75000	0.75000	
		1		0.76000	0.78000	0.75000	
	Lectura Crítica	0	0.73270	0.71000	0.68000	0.70000	
		1		0.75000	0.78000	0.76000	
	Competencias Ciudadanas	0	0.75200	0.76000	0.80000	0.78000	
		1		0.74000	0.69000	0.71000	
	Inglés	0	0.79050	0.78000	0.76000	0.77000	
		1		0.80000	0.82000	0.81000	
	Puntaje global	0	0.76140	0.75000	0.74000	0.74000	
		1		0.78000	0.78000	0.78000	
	Redes neuronales	Razonamiento cuantitativo	0	0.76260	0.77000	0.74000	0.75000
			1		0.75000	0.79000	0.77000
Lectura Crítica		0	0.74520	0.78000	0.75000	0.77000	
		1		0.71000	0.74000	0.72000	
Competencias Ciudadanas		0	0.75120	0.75000	0.82000	0.79000	
		1		0.75000	0.66000	0.70000	
Inglés		0	0.78980	0.78000	0.76000	0.77000	
		1		0.80000	0.81000	0.81000	
Puntaje global		0	0.75730	0.71000	0.80000	0.75000	
		1		0.81000	0.72000	0.76000	
Máquinas de vector soporte		Razonamiento cuantitativo	0	0.77470	0.80000	0.75000	0.77000
			1		0.75000	0.80000	0.78000
	Lectura Crítica	0	0.73190	0.72000	0.67000	0.69000	
		1		0.74000	0.78000	0.76000	
	Competencias Ciudadanas	0	0.74440	0.76000	0.78000	0.77000	
		1		0.72000	0.70000	0.71000	
	Inglés	0	0.77500	0.76000	0.75000	0.75000	
		1		0.78000	0.80000	0.79000	
	Puntaje global	0	0.76900	0.77000	0.72000	0.74000	
		1		0.77000	0.81000	0.79000	

De acuerdo con los resultados evidenciados se encuentra que en exactitud hay un promedio de eficiencia cercano al 75% y los modelos no presentan una diferencia significativa. En precisión, recall y f1, el modelo de máquinas de vector soporte tiene la mejor eficiencia, mientras que en las demás competencias y métricas son las redes neuronales las que tienen mejor eficiencia, pero no por márgenes significativos. En términos generales el comportamiento de los modelos tiene un comportamiento muy

similar y aun cuando los mejores resultados los tienen las redes neuronales, cualquier modelo predictivo se puede aplicar con este tipo de datos.

A partir de los resultados obtenidos con los modelos predictivos del progreso en el aprendizaje de los estudiantes de UNIMINUTO en las pruebas estandarizadas saber pro, con una eficiencia de cerca del 75%, se pueden establecer las siguientes estrategias para lograr el cierre de las brechas de aprendizaje de los estudiantes.

- Realizar una caracterización al ingreso de estudiantes a UNIMINUTO, que permita obtener el 100% de los datos de estudiantes en lo referente a la prueba saber 11.
- Aplicar el modelo predictivo a la información de los estudiantes que aún no presentan las pruebas de salida saber pro, a partir de los resultados de las pruebas de entrada sabe 11.
- Con los resultados del modelo se pueden establecer estrategias para el cierre de la brecha de aprendizaje por competencia, a corto, mediano y largo plazo.
- Se pueden implementar pruebas estándar para ser aplicadas al 25% y 50% de los créditos de los estudiantes, para ir ajustando las estrategias.
- El modelo permite agregar nueva información de carácter institucional como calificaciones de los cursos asociados a las competencias genéricas de razonamiento cuantitativo, lectura crítica, competencias ciudadanas e inglés, con el fin de ajustar el modelo y mejorar la eficiencia del mismo.
- Con las pruebas estándar institucionales es posible validar la eficiencia de las estrategias planteadas para mejorar el resultado en las pruebas saber pro.
- El modelo se aplica cada vez que se tenga nueva información, con lo cual la gestión académico-administrativa de UNIMINUTO contará con una importante herramienta para gestionar estrategias de mejora continua.

Los modelos predictivos del progreso en el aprendizaje de los estudiantes de UNIMINUTO en las pruebas de salida de la educación superior denominadas saber pro, pueden tener varios campos de acción dentro de las facultades, los programas, las áreas de formación y los cursos, como lo son:

- Establecer un modelo predictivo para determinar posibles deserciones tempranas de estudiantes de primer año.
- Establecer un modelo predictivo para determinar el progreso en áreas específicas como matemáticas o inglés.
- Establecer un modelo predictivo para determinar el progreso de estudiantes de un programa académico en específico.
- Establecer un modelo predictivo para determinar el progreso en una ruta de aprendizaje específica.
- Establecer un modelo predictivo para determinar el progreso con un profesor específico.
- Establecer un modelo predictivo para determinar el progreso los resultados de aprendizaje esperados a nivel programa, área o curso.

Durante la investigación se lograron desarrollar tres modelos predictivos que explican la relación del progreso del aprendizaje de los estudiantes de UNIMINUTO en las pruebas de salida,

con respecto a las pruebas de entrada a la educación superior, con una eficiencia cercana al 75%, lo que denota la correlación de las variables de estudio. Los modelos permiten anticipar los resultados que tendrán los estudiantes que se encuentran desarrollando sus estudios. Con esta información la gestión académico-administrativa tiene las herramientas para establecer estrategias que permitan realizar una intervención temprana y focalizada, logrando así, el cierre de las brechas de los aprendizajes. El modelo se puede ajustar para obtener resultados a nivel rectoría, facultad, centro regional, o programa académico.

Como recomendación, UNIMINUTO podría desarrollar una caracterización para obtener la información de los resultados de las pruebas de entrada y salida del 100% de sus estudiantes, toda vez que el ICFES no tiene un modelo eficiente para relacionar los datos de las pruebas saber 11 con el respectivo resultado de las pruebas saber pro, lo que generó una reducción significativa del universo de estudio.

Los modelos predictivos se deben ajustar con nueva información que incluyan instancias o variables nuevas, revisando la eficiencia del modelo. Para mejorar la eficiencia del modelo se pueden adicionar variables de tipo institucional como las calificaciones obtenidas en los cursos, horas de laboratorios disponibles para los estudiantes, tiempo de uso de los recursos tecnológicos disponibles, áreas disponibles para actividades curriculares, entre otras.

A partir de los resultados de la investigación, se puede proponer una hipótesis que relacione la deserción de estudiantes con respecto a las variables institucionales, que permitan anticipar los resultados y con ello llevar a cabo estrategias de permanencia focalizadas en UNIMINUTO.

En conclusión, se desarrolló una metodología para desarrollar modelos predictivos del progreso del aprendizaje de estudiantes a partir de la información de las pruebas de entrada y salida de la educación superior suministrada por el ICFES, con eficiencias del 75% en la predicción.

CONCLUSIONES

Con el análisis descriptivo se alcanzaron 10505 instancias con 69 variables correspondientes a la información de ingreso y salida de la educación superior de estudiantes de UNIMINUTO entre el 2016 y el 2018.

Con el análisis predictivo, se lograron analizar las variables categóricas a partir de las metodologías MCA y PCA. Se seleccionaron 3 modelos de aprendizaje supervisado con los algoritmos regresión logística, máquinas de vector soporte y redes neuronales. Se entrenaron los modelos y a partir de las métricas seleccionadas (accuracy, precisión, recall y f1) se evaluaron las eficiencias de los modelos, encontrando que el de redes neuronales fue el más eficiente de los tres modelos, pero no de forma significativo.

Con el análisis prescriptivo se brinda la proyección que tiene este tipo de herramientas tecnológicas en los procesos académico administrativos de UNIMINUTO, asociados a la realidad tecnológica que vivimos actualmente.

Finalmente, con las aplicaciones del modelo, se orienta a UNIMINUTO sobre las posibilidades que se tienen con este tipo de herramientas en aspectos como la permanencia académica.

REFERENCIAS BIBLIOGRÁFICAS

- Arteaga Valdés, E., León Capote, M. Á., & Del Sol Martínez, J. L. (2018). La clase de Matemática en la Educación Superior con un enfoque problémico. *Revista Conrado*, 14(64), 63-71.
- Contreras Bravo, L. E., Rodríguez Molano, J. I., & Fuentes López, H. J. (2021). Analítica académica: nuevas herramientas aplicadas a la educación. *Revista Boletín Redipe*, 10(3), 137-158.
- Cunha, J. M., & Miller, T. (2014). Measuring value-added in higher education: Possibilities and limitations in the use of administrative data. *Economics of Education Review*, 42, 64-77.
- De La Hoz, E. J., Fontalvo, E. J. & De La Hoz, T.J. (2019). Metodología de Aprendizaje Automático para la Clasificación y Predicción de Usuarios en Ambientes Virtuales de Educación. *Información Tecnológica*, 30(1), 247-254.
- Feitó Cespón, M., & Pérez de Armas, M. (2018). Lecciones aprendidas desde la praxis: proceso de evaluación y acreditación de carrera de ingeniería industrial en la Universidad de Cienfuegos. *Revista Conrado*, 14(64), 104-113.
- Jara, J., Chacón, M. A., Zelaya, G. (2011). Evaluación empírica de tres métodos de aprendizaje automático para clasificar automáticamente diagnósticos de neoplasias. *Ingeniare*, 19(3), 359-368.
- Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies*, 13(1), 61-72.
- Lara, A.A. 2020. «Prediction of psychosocial occupational risk in urban transport applying machine learning techniques.» *Revista Iberica de Sistemas e Tecnologias de Informacao* 153-165.
- Lokanan, M., Tran, V., & Vuong, N. H. (2019). Detecting anomalies in financial statements using machine learning algorithm: The case of vietnamese listed firms. .» *Asian Journal of Accounting Research*, 4(2), 181-201.
- Minte Münzenmayer, A., Sepúlveda Obreque, A., Obando, D. I., & Díaz Levicoy, D. (2021). ¿Qué nivel de razonamiento promueven las preguntas de las pruebas escritas en la educación básica chilena? *Revista Conrado*, 17(83), 375-381.
- Shamshirband, S., Mosavi, A., Rabczuk, T., Nabipour, N., & Kwok-wing, C. (2020). Prediction of significant wave height; comparison between nested grid numerical model, and machine learning models of artificial neural networks, extreme learning and support vector machines. *Engineering Applications of Computational Fluid Mechanics*, 14(1), 805-817.
- Touria, H., Faouzia, B., & Nawal, S. (2021). Student Profile Modeling Using Boosting Algorithms. *International Journal of Web-Based Learning and Teaching Technologies*, 17(5), 1-13.