

TRABAJO ORIGINAL

Modelo predictivo para estimar la deserción de estudiantes en la carrera de medicina de la Universidad Nacional de La Matanza

Predictive model to estimate medical school dropouts at Universidad Nacional de La Matanza

Modelo predictivo para estimar o abandono de estudiantes no curso de medicina da Universidad Nacional de La Matanza

Hugo Fernán Milione ^{1*}, Diego Fernández ¹, Claudio Antonio Ortiz ¹, Blanca Giménez Prieto ¹, Micaela Sabrina Magariños ¹, Jennifer Sabrina Appeceix ¹

Recibido: 27 de marzo de 2023. Aceptado: 24 de mayo de 2023.

¿Qué se sabe?

La deserción es entendida como la salida temprana o la falla de un estudiante en completar un programa de estudio en el que se inscribió. Se sabe que produce “costos” con diferentes alcances y consecuencias para el alumno. Una de las formas de predecirla puede ser mediante la utilización de modelos de inteligencia artificial a través del denominado aprendizaje automático.

¿Qué aporta este trabajo?

Conocer el alcance de un modelo de inteligencia artificial para determinar si el conjunto de los datos obtenidos al ingreso de cada alumno a la carrera de medicina (socio-culturales-familiares) tienen relación con la deserción voluntaria y, si así fuera, cuáles son las de mayor asociación.

Resumen

Introducción: La aplicación de un modelo de Inteligencia Artificial (IA) puede determinar el conjunto de datos socio económico financieros obtenidos al ingreso de cada estudiante a la carrera de medicina de la Universidad Nacional de La Matanza y su relación con la deserción voluntaria.

Materiales y métodos: Se llevó a cabo un estudio longitudinal y analítico. Las herramientas de observación y análisis fueron: 1. datos socioeconómico-familiares que surgen de una encuesta voluntaria a todos los estudiantes de medicina al ingreso a la carrera en el período 2012-2018. 2. datos del rendimiento académico de los estudiantes durante la carrera. Se aplicó un programa de IA denominado Orange Canvas a los contenidos de las bases de datos antes mencionadas, que consiste en una herramienta de minería de datos que proporciona una serie de complementos para el análisis, visualización y modelado de datos.

Resultados: El modelo “Random Forest” del programa logró un 72% de precisión, incrementándose al 76,7% con el modelo Test and Score. Finalmente, con el modelo KNN la precisión para predecir deserción ascendió al 87,2%.

Conclusión: Después de realizar todas las pruebas al conjunto de datos de la encuesta inicial y el conjunto de datos de las materias aprobadas por el estudiante surge que el modelo KNN de la herramienta de IA alcanza el mejor nivel de predicción (87,2%) para predecir deserción.

Palabras clave: abandono escolar, inteligencia artificial, valor predictivo de las pruebas

* Correspondencia: Hugo Milione, hmilione@unlam.edu.ar

¹ Universidad Nacional de La Matanza, Departamento de Ciencias de la Salud, San Justo, Argentina.



Esta obra está bajo una Licencia Creative Commons Atribución 4.0 Internacional (CC BY 4.0).

Abstract

Introduction: The application of an Artificial Intelligence (AI) model can determine the set of socioeconomic-financial data obtained at the entrance of each student to the medicine program at Universidad Nacional de La Matanza and its relationship with voluntary dropouts.

Materials and methods: A longitudinal and analytical study was carried out. Observation and analysis tools were: 1. Socioeconomic-family data from a voluntary survey of all medical students upon admission to the program in the period 2012-2018. 2. Data on the academic performance of the students during the program. An AI program called Orange Canvas was applied on the contents of the above-mentioned databases, which is a data mining tool that provides a series of complements for data analysis, visualization and modelling.

Results: The “Random Forest” model of the AI program achieved a 72% accuracy, which increased to 76.7% with the Test and Score model. Finally, with the KNN model, the accuracy to predict dropouts rose to 87,2%.

Conclusion: After performing all the tests on the data set of the initial survey and the dataset of the subjects each student passed, the KNN model of the AI tool reached the highest prediction level (87,2%) for dropouts.

Keywords: student dropouts, artificial intelligence, predictive value of tests.

Resumo

Introdução: A aplicação de um modelo de Inteligência Artificial (IA) pode determinar o conjunto de dados socioeconômico e financeiros obtidos na entrada de cada aluno no curso de medicina da Universidade Nacional de La Matanza e sua relação com o abandono voluntário.

Materiais e métodos: Foi realizado um estudo longitudinal e analítico. Os instrumentos de observação e análise foram: 1. Dados socioeconômicos-familiares resultantes de um inquérito voluntário a todos os estudantes de medicina no momento do ingresso ao curso no período 2012-2018. 2. Dados sobre o rendimento acadêmico dos alunos durante a carreira. Ao conteúdo das bases de dados supracitadas foi aplicado um programa de IA denominado Orange Canvas, que consiste em uma ferramenta de mineração de dados que fornece uma série de complementos para análise, visualização e modelagem de dados.

Resultados: O modelo “Random Forest” do programa alcançou 72% de acerto, aumentando para 76,7% com o modelo Test and Score. Finalmente, com o modelo KNN, a precisão para prever o abandono aumentou para 87,2%.

Conclusão: Após realizar todos os testes no conjunto de dados da pesquisa inicial e no conjunto de dados das disciplinas aprovadas pelo aluno, verifica-se que o modelo KNN da ferramenta de IA atinge o melhor nível de predição (87,2%) para prever abandono.

Palavras-chave: evasão escolar, inteligência artificial, valor preditivo dos testes.

Fuentes de financiamiento:

Las autoras declaran no tener ninguna afiliación financiera ni participación en ninguna organización comercial que tenga un interés financiero directo en cualquier asunto incluido en este manuscrito. Programación CyTMA2, Código C2SAL-044, Secretaría de Ciencia y Tecnología, Departamento de Ciencias de la Salud, Universidad Nacional de La Matanza.

Conflicto de intereses:

Las y los autores declaran no tener ningún conflicto de intereses.

Introducción

La deserción es entendida como la salida temprana de un estudiante de un programa de estudio en el que se inscribió, la misma puede ser de carácter voluntario o involuntario: la deserción es voluntaria a través de su renuncia formal o por el abandono no informado del estudiante y es involuntaria, cuando por decisión institucional el estudiante no puede seguir sus estudios por razones académicas o disciplinarias.

La deserción fue y es motivo de estudio de varias disciplinas: la psicología, la economía, la historia, la sociología, la gestión educativa y desde no hace mucho tiempo, la minería de datos. Todas ellas hacen su aporte al entendimiento de este fenómeno que produce costos con diferentes alcances y consecuencias: la frustración y la deuda financiera que genera al estudiante, el congelamiento del financiamiento a la institución educacional, la pérdida de una vacante que pudo ser utilizada por otro estudiante que sí finalizaría el programa, el estancamiento del desarrollo educacional de la sociedad y la disminución de un profesional al país, mayormente en aquellas profesiones más demandadas.

Una de las áreas de investigación antes mencionadas, la Inteligencia Artificial (IA), es una rama de estudio de la informática que se enfoca en la creación de sistemas y programas capaces de realizar tareas que requieren de inteligencia humana. La IA se basa en la búsqueda de algoritmos y modelos que permitan a las máquinas aprender, razonar, planificar, reconocer patrones, entender lenguaje natural y tomar decisiones de manera autónoma.

Dentro de la IA existen diferentes enfoques y técnicas, como por ejemplo el aprendizaje automático (Machine Learning), el procesamiento del lenguaje natural (NLP), la visión por computadora y la robótica, entre otros. Estas técnicas permiten a los sistemas de IA adquirir conocimiento a través de serie de datos y experiencias, y utilizar ese conocimiento para realizar diferentes tareas, la IA no busca reemplazar completamente a los seres humanos, no hay dudas que contribuye a complementar y mejorar las capacidades humanas a través de la automatización de tareas que requieren inteligencia para agilizar procesos, tomar decisiones más precisas y proporcionar soluciones más eficientes.

Si bien la definición de deserción estudiantil es debatible, existe consenso con respecto a que el abandono voluntario puede ser explicado por diferentes categorías de variables: socioeconómicas, individuales, institucionales y académicas que, a su vez, pueden ser analizadas teniendo en cuenta diferentes modelos:

1. El modelo psicológico: sostiene que los rasgos de la personalidad diferencian a los estudiantes que terminan sus estudios regulares de aquellos que no lo logran (1).
2. El modelo sociológico: resalta la influencia de factores externos al individuo, que influyen en el rendimiento y en la probabilidad de que el estudiante decida abandonar sus estudios (2).
3. El modelo económico: considera el costo beneficio de trabajo vs estudio y la influencia de los subsidios para evitar la deserción (3-6).
4. El modelo organizacional: mantiene que la deserción depende de las características de la organización en la integración social, y más particularmente en el abandono de los estudiantes ingresantes (7).
5. El modelo de interacción: explica el proceso de permanencia en la educación superior como una función del grado de ajuste entre el estudiante y la institución, a partir de la integración académica y social (8).

En cuanto a los modelos teóricos para analizar la deserción, podemos mencionar los siguientes: (1) modelo basado en la teoría del suicidio, (2) modelo basado en la teoría del intercambio y (3) el modelo basado en el modelo de productividad del ambiente laboral. El modelo del suicidio establece que la deserción sería un resultado de una falla en la integración del individuo con su entorno educacional y familiar, afectando fuertemente al estudiante, ya que ambos los exponen a influencias, expectativas y demandas que podrían afectar tanto en su integración social como con sus pares afectando el rendimiento académico (9).

La teoría del intercambio sostiene que en general los seres humanos evitan aquellas conductas que les generan costos de algún tipo y buscan los beneficios fundamentalmente en las relaciones e interacciones con sus pares y la institución educacional (10). A partir de ello, los estudiantes solo se mantendrían en el programa que se inscribieron cuando los beneficios percibidos superen el esfuerzo, dedicación y otros costos personales; cualquier situación de desequilibrio, podría desencadenar en una deserción (8). En cuanto al basado en el modelo de productividad del ambiente laboral, las variables relacionadas con los antecedentes del estudiante, tales como estado socioeconómico, desempeño académico previo y residencia actual, impactarían en los determinantes organizacionales o bien, de la institución educacional y estos a su vez, en la decisión de desertar (11). En la Tabla 1 describimos algunas variables y sus relaciones según este autor.

El mismo autor, años más tarde agrega como variables a considerar las sociodemográficas como el género, edad y etnia del estudiante. El estudio planteó la heterogeneidad que existe en el cuerpo estudiantil, lo que se traduce en que para el fenómeno de deserción es importante considerar características propias del estudiante (12).

En la Carrera de Medicina de la Universidad Nacional de la Matanza, la tasa de deserción global de pregrado es del 21%, teniendo en cuenta 1751 ingresantes totales y 1270 activos al año 2019¹.

Tabla 1. Lista de variables de antecedentes planteadas por Bean et al. (11)

Variable	Definición
Desempeño previo	Grado en que el estudiante ha demostrado sus logros académicos previos
Estatus Socioeconómico	Grado en que los padres del estudiante han logrado como estatus a través de la ocupación familiar
Residente en el Estado	Si el estudiante es un residente del estado en donde está la institución educativa.
Distancia a casa	Distancia de su residencia actual a la casa de sus padres
Tamaño de la ciudad	Tamaño de la ciudad donde el estudiante paso la mayor parte de su tiempo durante su crecimiento

Materiales y métodos

La metodología adoptada para la realización de este estudio fue longitudinal y analítica.

Utilizamos como herramientas de observación y análisis: 1. Los resultados de los datos socioeconómico-familiares que surgen de una encuesta voluntaria que se realiza los estudiantes de la carrera de medicina al momento del ingreso: los datos fueron obtenidos de las encuestas anónimas y voluntarias que se realiza a todos los estudiantes de medicina al ingreso a la carrera, desde el año 2012 hasta la actualidad; 2. Datos acerca del rendimiento académico los estudiantes en los sucesivos años hasta la pasantía final obligatoria (correspondientes a las cohortes 2012-2013-2014 que son los alumnos que han cumplido la totalidad de la carrera) obtenidos de la base de datos de la Universidad a través del Sistema de Gestión Académica Guarani, a partir de algoritmos de búsqueda definidos específicamente para obtener los datos necesarios.

1. La secuencia metodológica fue la siguiente:

A. Armado de una base de datos relacional en MS Access para combinar sets de datos provenientes de la encuesta y de rendimiento académico.

B. Armado de set de datos para establecer si X estudiante ha abandonado.

Origen: Encuesta de ingresantes a la carrera de medicina desde el año 2012, que dispone de las siguientes variables:

1. Apellido
2. Nombre
3. DNI
4. Mail
5. Celular
6. Teléfono de línea casa
7. Género
8. Fecha de nacimiento
9. Edad
10. Lugar de nacimiento
11. País de nacimiento
12. Donde reside/residirá en el transcurso de este año?
13. Si escogió otra, cual
14. Si trabaja
15. Cantidad de hs. semanales que trabaja aproximadamente:

1. Fuente: Sistema de Gestión Académica Guarani. SIU Guarani. UNLaM

16. El horario de trabajo que cumple
17. Si prevé trabajar
18. Escolaridad Primaria
19. Si hizo la primaria en el municipio de La Matanza
20. Escolaridad secundaria
21. Si hizo la secundaria en el municipio de La Matanza
22. Tipo de secundaria

C. Armado de set de datos para establecer rendimiento académico.

Origen: datos exportados del Sistema Guaraní donde constan los resultados de todas las materias cursadas por los alumnos desde el inicio de la carrera.

1. Apellido y nombres
2. DNI
3. Cohorte de ingreso
4. Condición regular
5. Código de la materia
6. Nombre de la materia
7. Fecha
8. Tipo de nota
9. Nota
10. Resultado
11. Fecha vigencia
12. Condición vigente
13. Condición adeudada
14. Año de egreso

2. La herramienta de Inteligencia Artificial utilizada:

A partir de los contenidos de las bases de datos antes mencionadas, se utilizó un programa de Inteligencia Artificial denominado *Orange Canvas*, que es una herramienta de minería de datos² que proporciona una serie de complementos para el análisis, visualización y modelado de datos. En esta investigación se utilizó esta herramienta para implementar una serie de modelos predictivos, para tratar de determinar cuál es la probabilidad de predecir si un estudiante deserta o no. Los modelos utilizados fueron:

REGRESIÓN LOGÍSTICA

Este modelo estadístico se utiliza como clasificador, utilizando una función logística para modelar una variable dependiente binaria. Es decir, se utiliza para predecir la probabilidad de que exista determinado evento de manera dicotómica, con resultados tales como: si/no, pasa/no pasa. Esta variable indicadora arroja valores con resultados “0” y “1”

NAIVE BAYES

Al igual que el anterior, este modelo se utiliza como clasificador probabilístico y se basa en el Teorema de Bayes, con supuestos de independencias entre características. Este teorema describe la probabilidad de ocurrencia de un evento basándose en el conocimiento de sus características y condiciones previas.

K-Nearest Neighbors (KNN)

El algoritmo KNN es un método de aprendizaje automático no paramétrico utilizado para clasificación y regresión. En ambos casos, la entrada consta de los K ejemplos de entrenamiento más cercanos en el espacio de características y utiliza su promedio como predicción.

RED NEURONAL

Consiste en un grupo de unidades llamadas neuronas artificiales, que están conectadas entre sí para transmitir señales. La información de entrada pasa a través de la red neuronal (en la que se realizan varias operaciones) para producir el valor de salida. Cada neurona está conectada a otras neuronas a través de enlaces. En estos enlaces, el valor de salida de la neurona anterior se multiplica por el valor de peso.

El Flujo de Trabajo con la herramienta *Orange Canvas* consistió en 5 pasos (Fig. 1) *Ver en página 38*

Paso 1: se vinculó el set de datos analizado (datos de la encuesta y rendimiento académico), fundamental para indicar cuál es la variable que vamos a tratar de predecir.

Paso 2: se tomó un set de datos y se entrenó a los modelos predictivos de la herramienta para testear el nivel de precisión.

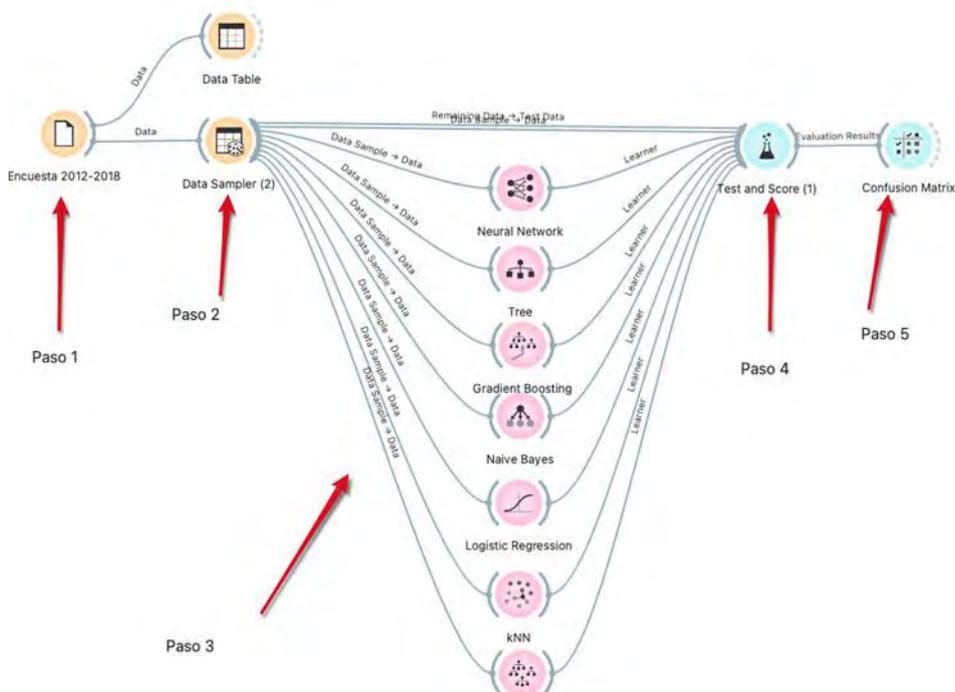
Paso 3: se dividieron los datos en un set de entrenamiento y un set de datos de prueba utilizando el elemento {dataSampler}, técnica utilizada en el análisis de datos para seleccionar una muestra representativa de un conjunto de datos más grande.

Paso 4: una vez entrenado los modelos se verificó el grado de precisión que tiene cada modelo predictivo con ese conjunto de datos.

Paso 5: se utilizó la Matriz de Confusión (Confusión Matrix), herramienta visual utilizada en el campo de la inteligencia artificial y el aprendizaje automático para evaluar el rendimiento de un modelo de clasificación, para reforzar el paso anterior y así evaluar el grado de precisión que tiene el modelo.

2. La minería de datos, también conocida como descubrimiento de conocimiento en bases de datos (KDD, por sus siglas en inglés), es el proceso de extracción de información valiosa, patrones y conocimientos significativos a partir de grandes conjuntos de datos.

Figura 1. Flujo de Trabajo de la herramienta Orange Canvas



Resultados

Para la obtención de los resultados se realizaron 7 pruebas con las siguientes conclusiones:

Prueba 1: los modelos predictivos utilizados a partir del set de datos con los resultados de las encuestas realizadas entre los años 2012 a 2018, tienen la capacidad de predecir en un 57,1% si el estudiante desertará o no. Con lo cual, se concluye que son insuficientes para establecer una herramienta de predicción confiable.

Prueba 2: debido al insuficiente rendimiento obtenido en la primera prueba, se decidió indagar si existían otras variables que pudieran explicar y predecir mejor la deserción del estudiante. Para esto, se realizó una tabla dinámica con el set de datos “Rendimiento Académico”, contando la cantidad de veces que el estudiante ha desaprobado una materia. Las materias figuran como columnas (campos) y la cantidad de veces reprobadas figuran como números del 1 al N, y si nunca la han desaprobado se obtiene el valor 0.

Con este set de datos se realizaron las mismas pruebas que para la prueba N°1, utilizando un Data Sampler para dividir los datos en dos grupos: un 79% para entrenar al modelo y el 21% restante para corroborar el nivel de predicción.

En esta segunda prueba se obtuvo que el nivel de precisión, con el modelo “Random Forest” alcanzó el 72%, lo que constituye una mejora en el nivel de precisión en comparación con el primer

set de datos. Esto indicaría que el rendimiento académico, más precisamente la cantidad de veces que un estudiante desaproveba una materia, incide en la deserción de la carrera.

En la “Confusion Matrix” se observa que con “Random Forest” (algoritmo de aprendizaje automático que se utiliza en problemas de clasificación y regresión) se alcanza una probabilidad de predicción del 72% para aquellos estudiantes que han desertado la carrera.

Conclusión de la prueba:

Prueba 3: en esta prueba, con el mismo conjunto de datos del “Rendimiento académico”, se realizó una nueva tabla dinámica considerando, para una misma asignatura, si el estudiante había aprobado o no la misma.

A través de la herramienta Test and Score, se observó que, en este caso, el nivel de predicción se incrementó al 76,7%, lo cual mejora con respecto a la Prueba N° 2.

Finalmente, con la herramienta “Confusion Matrix”, se corrobora que el “Ascenso de Gradiente” resultó el mejor modelo, con un 76,7%. Concluimos que la cantidad de materias aprobadas por el estudiante predice aún mejor que la encuesta inicial y que la cantidad de veces que reprueba una materia la precisión en la predicción de la deserción de un estudiante.

Prueba 4: se combinaron dos sets de datos: el de encuestas de in-

greso de la Prueba 1 y el set de datos de “Rendimiento Académico” número de materias desaprobadas de la Prueba 2. Es decir, el propósito de la prueba fue testear si los datos de la encuesta al ingreso y las materias desaprobadas resultan un mejor predictor de la deserción de los estudiantes que cualquiera de los conjuntos de datos por separado. Para ello se utilizó la herramienta “Unión de datos” que, utilizando el DNI, une los conjuntos de datos para establecer una nueva tabla de datos.

Esta nueva tabla de datos tiene 58 *features* (variables) y 1323 instancias. Con este conjunto de datos, se evidenció que el nivel de precisión desciende al 73,2%. Es decir, que este conjunto de datos tiene menos capacidad de predicción que el conjunto de datos de materias aprobadas (Prueba 3).

Prueba 5: se combinaron los datos de la encuesta de inicio con las materias aprobadas, para verificar si de esta manera mejora el nivel de predicción.

Para esto, se utilizó la misma herramienta que en la Prueba N°4, para unir dos sets de datos mediante el número de DNI. Este nuevo set de datos se compone de 58 variables y 1323 instancias. En este caso, se observa que a través del modelo KNN, la precisión para predecir deserción ascendió al 87,2%, concluyendo que la unión entre el set de datos de encuestas al inicio y la cantidad de materias aprobadas por estudiante es la mejor combinación para predecir la deserción del estudiante.

Prueba 6: se determinó el nivel de precisión en la predicción

que tiene la combinación del set de datos cantidad de veces que los estudiantes desaprobaron una materia con la cantidad de materias aprobadas. Este nuevo set de datos se compone de 82 variables se obtuvo una precisión en la predicción del 82,1%, lo que representa un descenso en el nivel de predicción. Las variables obtenidas de la combinación de los datos la cantidad de veces que un estudiante desaprobó una materia con las materias aprobadas por el mismo estudiante resultó menos preciso que el conjunto de datos utilizado en la Prueba N°5 (encuestas de inicio + desaprobados).

Prueba 7: se combinaron los tres sets de datos: los datos de la encuesta de inicio, los datos del rendimiento académico de cantidad de veces que los estudiantes desaprobaron una materia, y la cantidad de materias aprobadas por el estudiante. El resultado de esta unión arroja una tabla con 99 variables y 1422 instancias. El nivel de precisión alcanzado fue solo del 81%, es decir, menos predictivo que la encuesta más la cantidad de materias aprobadas.

Conclusión de todas las pruebas realizadas:

Del resultado de todas las pruebas realizadas surge que, si tomamos solo el set de datos de la encuesta de inicio (hipótesis inicial del trabajo), el nivel de predicción logrado es del 57,1%; ahora, si al conjunto de datos de la encuesta le sumamos el set de datos de las materias aprobadas por el estudiante el nivel de predicción es de 87,2% logrando el mejor nivel para predecir la deserción del estudiante (Tabla 2).

Tabla 2. Relación set de datos y nivel de precisión logrado

Prueba	Set de Datos	Mejor Modelo	Mejor Nivel de Precisión
1	Set de datos 1: Encuestas al Inicio	Gradient Boosting	57,1%
2	Set de datos 2: Rendimiento Académico / Cantidad de veces que desaprobaron una materia	Random Forest	72,0%
3	Set de datos 3: Rendimiento académico / Materias Aprobadas	Gradient Boosting	76,7%
4	Set de datos 4: Unión de set de datos 1 y el set 2	Gradient Boosting	73,2%
5	Set de datos 5: Unión del set de datos 1 y el set 3	KNN	87,2%
6	Set de Datos 6: Unión del set de datos 2 y el set 3	Random Forest	82,1%
7	Set de datos 7: Unión de los sets de datos 1, 2 y 3	Gradient Boosting	81,0%

Conclusiones

De las distintas pruebas realizadas con la herramienta de Inteligencia Artificial utilizada podemos afirmar que, contrario a nuestra hipótesis inicial, el conjunto de datos proveniente de las respuestas a las encuestas de inicio, por sí solas, no es suficientemente preciso para predecir la deserción del estudiante.

Ahora bien, de la unión de los datos provenientes de las encuestas al ingreso y el rendimiento académico (cantidad de materias aprobadas antes de perder regularidad) utilizando nuestra herramienta de IA (y luego de varias pruebas realizadas), la precisión en la predicción de deserción llega hasta un 87,2%. Este dato nos plantea como nueva pregunta de investigación, cuál es el rol del rendimiento académico general y es-

pecífico de la currícula de medicina de la Universidad Nacional de la Matanza en la deserción de los alumnos, como factor independiente de las variables socioeconómicas ya analizadas en este trabajo.

Creemos importante continuar con la aplicación institucional de una metodología de análisis utilizando esta herramienta de minería de datos de información, en el proceso académico de los alumnos ingresantes y en su proceso durante los dos primeros años de la carrera, en pos de ofrecer herramientas para un acercamiento a la comprensión del fenómeno de la deserción universitaria, y así dirigir acciones de las políticas departamentales que busquen modificar la naturaleza y condiciones académicas que lo posibilitan. De esta forma se mejorará la experiencia educativa estudiantil y el uso de los recursos institucionales.

Contribución de los/as autores:

HFM, DF, CAO, BGP, MSM y JSA: búsqueda bibliográfica, diseño del estudio, recolección de datos, análisis de datos y preparación del manuscrito.

Referencias

1. Fishbein M, Ajzen I. Attitudes toward objects as predictors of simple and multiple behavioural criteria. *Psychological Review*. 1975. 81:59-74.
2. Durkheim E. *Suicide: A study in sociology* (Internet) New York: Free Press. Disponible en <https://www.gacbe.ac.in/images/E%20books/Durkheim%20-%20Suicide%20-%20A%20study%20in%20sociology.pdf>
3. Cabrera A, Nora A, Castañeda M. The role of finances in the persistence process: A structural model. *Res High Educ*. 1992. 33(5):303-336.
4. Cabrera A, Nora A, Castañeda M. College Persistence: Structural Equations Modelling Test of Integrated Model of Student Retention. *J Higher Educ*. 1993. 64(2):123-320.
5. Bernal E, Cabrera A, Terenzini P. The relationship between race and socioeconomic status (SES): Implications for institutional research and admissions policies. *Removing Vestiges: Research-based strategies to promote inclusion*. American Association of Community Colleges. 2000. (3):6-19.
6. St. John E, Cabrera A, Aker N, Asker E. Economic influences on persistence. In: J. M. Braxton. *Reworking the student departure puzzle: new theory and research on college student retention*. Nashville: Vanderbilt University Press. 2000. 29-47
7. Berger J, Milem J. Organizational Behavior in Higher Education and Student Outcomes. In: J. Smart (Ed.), *Higher Education: Handbook of theory and research*. 2000. (15):268-338.
8. Tinto V. Limits of theory and practice of student attrition. *J Higher Educ*. 1982. 3(6):687-700.
9. Spady WG. Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange*. 1970. (1):64-85.
10. Nye J. Independence and Interdependence. *Foreign Policy*. Spring. 1976. (22):130-161.
11. Bean J. Student Attrition, Intentions and Confidence: *Res High Educ*. 1980. (17):291-320.
12. Bean JP, Metzner BS. A conceptual model non-traditional undergraduate student attrition. *Rev Educ Res*. 1985. 55(4):485-540.