

Diseño de un modelo matemático para estimar el efecto independiente de los problemas de salud mental en la deserción universitaria temprana

Jesús Miguel Sepúlveda Suárez

Trabajo de Investigación para optar al título de Magister en Matemática Aplicada

Directora:

Lina María Vera Cala

PhD. Epidemiología

Codirector:

Jorge Andrés Niño García

Médico Psiquiatra de Enlace

Universidad Industrial de Santander

Facultad de Ciencias

Escuela de física

Maestría en Matemática Aplicada

Bucaramanga

2024

Contenido

	Pág.
Introducción	10
1. Planteamiento del problema.....	12
2. Hipótesis	15
3. Pregunta de investigación	15
4. Marco teórico	16
4.1 Conceptualización.....	16
4.2 Factores que contribuyen a la deserción	18
4.2.1 Factores individuales	19
4.2.2 Factores académicos	20
4.2.3 Factores institucionales	20
4.2.4 Factores socioeconómicos	21
4.2.5 Factores de salud mental.....	22
4.3 Modelos matemáticos	27
4.3.1 Modelos determinísticos y estocásticos	27
4.3.2 Modelos continuos y discretos.....	29
4.3.3 Modelos lineales y no lineales	31
4.3.4 Modelos de optimización.....	32
4.3.5 Modelos de simulación	33
4.3.6 Modelos de redes	35
4.3.7 Modelos de probabilidad y estadística.....	36

4.3.8 Modelos de computación 38

4.3 Análisis Multivariante (AM)..... 48

4.3.1 Objetivos del análisis multivariante..... 49

4.3.2 Categorías de técnicas multivariantes 49

5. Estado del arte..... 53

6. Objetivos 57

6.1 Objetivo general..... 57

6.2 Objetivos Específicos..... 57

7. Metodología 58

7.1 Diseño del estudio..... 58

7.2 Población..... 58

7.3 Criterios de Inclusión 59

7.4 Criterios de Exclusión..... 59

7.5 Muestra 59

7.6 Instrumentos..... 61

7.7 Aspectos Éticos..... 62

7.8 Recolección y Preparación de Datos..... 63

7.9 Definición de Variables y Objetivo 64

7.10 Limitaciones Potenciales 66

7.11 Red neuronal 67

7.11.1 División del conjunto de datos..... 67

7.11.2 Diseño de la arquitectura de la red neuronal..... 68

7.11.3 Definición de función de pérdida y sobreajuste..... 69

7.11.4 Optimización y entrenamiento del modelo	70
7.11.5 Validación y ajuste de hiperparametros	72
7.12 Regresión logística.....	72
8. Resultados.....	73
8.1 Descripción de la población.....	73
8.2 Análisis bivariado	75
8.3 Análisis Multivariado.....	77
8.4 Evaluación del rendimiento de la red neuronal artificial	79
8.5 Combinación de la red neuronal artificial y la regresión logística	83
9. Discusión.....	83
12. Conclusiones y Recomendaciones	90
Referencias Bibliográficas	92
Apéndices.....	102

Lista de Tablas

	Pág.
Tabla 1. Determinantes de la deserción	21
Tabla 2. Variables	64
Tabla 3. Arquitectura de la red neuronal artificial	68
Tabla 4. Variables asociadas a la deserción	76
Tabla 5. Variables del modelo final	78

Lista de Figuras

	Pág.
Figura 1. Estudiantes inscritos, admitidos y absorción.....	14
Figura 2. Deserción.....	14
Figura 3. Modelo de McCulloch-Pitts para una neurona artificial	39
Figura 4. Ejemplo de arquitectura de una RNA.....	40
Figura 5. Diseño de una neurona artificial.....	42
Figura 6. Función sigmoide	43
Figura 7. Función Rectified Linear Unit.....	44
Figura 8. Función tangente hiperbólica	45
Figura 9. Técnicas de análisis multivariantes	50
Figura 10. Flujograma de estudios incluidos	55
Figura 11. Matriz de confusión.....	80
Figura 12. Pérdida de entrenamiento y validación.....	81
Figura 13. Precisión de entrenamiento y validación.....	82

Lista de Apéndices

	Pág.
Apéndice A. Esquema de extracción de datos	102
Apéndice B. Factores de la deserción	107
Apéndice C. Tabla de programas académicos	108
Apéndice D. Tabla de puesto de ingreso	109
Apéndice E. Tabla de variables en el análisis univariado.....	112

Resumen

Título: Diseño de un modelo matemático para estimar el efecto independiente de los problemas de salud mental en la deserción universitaria temprana *

Autor: Jesús Miguel Sepúlveda Suárez **

Palabras Clave: Regresión logística, redes neuronales artificiales, deserción universitaria

Descripción:

En Colombia, la deserción universitaria afecta a uno de cada dos estudiantes que ingresan a instituciones de educación superior (IES), generando pérdidas significativas en ingresos. Según la OCDE, cada año adicional de escolaridad puede aumentar el PIB per cápita entre 4 y 7%. Para abordar este problema, se llevó a cabo un estudio utilizando Regresión Logística (RL) y Redes Neuronales Artificiales (RNN) para predecir la deserción.

Se utilizó información recopilada por la Dirección de Bienestar Estudiantil de la UIS, enfocándose en estudiantes de nuevo ingreso entre 2015-2 y 2017-2, con un seguimiento de 4 semestres. La base de datos incluyó 4273 estudiantes y 26 variables predictoras, con una edad promedio de 18,1 años. El 74,24% residía en el Área Metropolitana de Bucaramanga y un 32,59% estaba en riesgo de depresión, mientras que el 40,78% presentaba riesgo de ansiedad.

La RL identificó las variables más significativas para la deserción: procedencia fuera del AMB, sexo masculino, disfuncionalidad familiar, baja autoconfianza y automotivación en el aprendizaje, insatisfacción con la vida, y no ser beneficiario de servicios de comedores de la UIS. Se desarrolló un modelo predictivo con RNN que logró una precisión del 87%. Finalmente, se combinó la RL y RNN, obteniendo una precisión del 84% y simplificando las variables para su implementación efectiva.

* Trabajo de Investigación

** Facultad de Ciencias Escuela de física Maestría en Matemática Aplicada Directora: Lina María Vera Cala PhD.
Epidemiología Codirector: Jorge Andrés Niño García Médico Psiquiatra de Enlace

Abstract

Title: Design of a mathematical model to estimate the independent effect of mental health issues on early university dropout*.

Author: Jesús Miguel Sepúlveda Suárez**

Key Words: Logistic Regression, artificial neural networks, university dropouts

Description:

In Colombia, university dropout affects one in two students entering higher education institutions (IES), resulting in significant income losses. According to the OECD, each additional year of schooling can increase GDP per capita by 4 to 7%. To address this issue, a study was conducted using Logistic Regression (LR) and Artificial Neural Networks (ANN) to predict dropout rates.

Data collected by the UIS Student Welfare Office focused on first-year students from 2015-2 to 2017-2, tracking them over four semesters. The database included 4,273 students and 26 predictive variables, with an average age of 18.1 years. Of these, 74.24% lived in the Bucaramanga Metropolitan Area, and 32.59% were at risk of depression, while 40.78% showed signs of anxiety.

LR identified the most significant variables for dropout: coming from outside the metropolitan area, being male, family dysfunction, low self-confidence, lack of self-motivation in learning, life dissatisfaction, and not being a beneficiary of UIS cafeteria services. An ANN predictive model achieved 87% accuracy. Finally, combining LR and ANN resulted in an 84% accuracy rate while simplifying the variables for effective implementation.

* Work of investigation

** Facultad de Ciencias Escuela de física Maestría en Matemática Aplicada Directora: Lina María Vera Cala PhD. Epidemiología Codirector: Jorge Andrés Niño García Médico Psiquiatra de Enlace

Introducción

Uno de los principales retos de las instituciones de educación superior es generar estrategias de retención universitaria, ya que, pese al esfuerzo del Gobierno Nacional y Regional por ampliar la cobertura, las tasas de deserción universitaria continúan siendo altas, estimándose que uno de cada dos estudiantes se desvincula del sistema educativo (MEN, 2009).

En este contexto de suma gravedad por el impacto negativo a nivel individual, familiar y social, se han realizado un gran número de investigaciones con el objetivo de cuantificar el fenómeno de la deserción universitaria y los factores asociados, que se han clasificado en cuatro grandes grupos: individuales, académicos, institucionales y socioeconómicos, sin embargo, es poco el énfasis que se ha hecho en los problemas de salud mental, considerando que el mayor porcentaje de estudiantes que ingresan a la universidad se encuentran en la transición de la adolescencia a la vida adulta, una población vulnerable a desarrollar trastornos del afecto y trastorno por uso de sustancias psicoactivas (Paula, 2016).

Es claro que el fenómeno de la deserción es el resultado de la confluencia de múltiples factores, por lo cual resulta indispensable apoyarnos en los modelos matemáticos para comprender el peso de las variables y la relación entre ellas con el abandono de la educación universitaria. Por lo que inicialmente se realizó una revisión de alcance para identificar y describir los modelos matemáticos que han sido utilizados para predecir la deserción estudiantil en los estudiantes de educación superior, que nos permita tener un mejor entendimiento acerca de la utilización de los modelos matemáticos en la deserción estudiantil. Tanto las redes neuronales artificiales como los

modelos de regresión logística son eficientes para abordar el fenómeno de la deserción estudiantil debido a sus capacidades de modelado y adaptabilidad a datos complejos.

Las redes neuronales artificiales, son capaces de capturar relaciones no lineales y patrones complejos en conjuntos de datos. Dada la naturaleza multifacética de los factores que contribuyen a la deserción estudiantil, las redes neuronales pueden aprender representaciones jerárquicas que van más allá de las capacidades de los modelos lineales. Además, la capacidad de adaptación de las redes neuronales a grandes volúmenes de datos y su habilidad para identificar interacciones entre variables hacen que sean idóneas para problemas complejos como la deserción estudiantil.

Por otro lado, la regresión logística, siendo un modelo lineal generalizado, es efectiva para problemas de clasificación binaria, como la predicción de deserción. Su naturaleza logística permite interpretar las probabilidades de eventos, lo que facilita la identificación de factores predictivos.

En resumen, tanto la regresión logística como las redes neuronales artificiales ofrecen herramientas poderosas y complementarias para abordar el fenómeno de la deserción estudiantil.

A partir de la base que nos proporcionó la revisión de alcance se llevó a cabo un análisis detallado de la información disponible a través de los datos obtenidos en la evaluación multifactorial que lleva a cabo la Dirección de Bienestar Estudiantil a los estudiantes que ingresan a la universidad, permitiéndonos seleccionar las variables que serían de utilidad para los modelos matemáticos. Posteriormente se realizó un proceso de refinamiento de la base de datos para optimizar la eficiencia y precisión de nuestro enfoque analítico, así como una depuración de las variables de interés con base en los resultados de la regresión logística para optimizar la red neuronal.

1. Planteamiento del problema

La deserción universitaria es un fenómeno que ha estado presente en las instituciones de educación superior (IES) desde hace varias décadas, que se había normalizado y se le restaba importancia, justificándose erróneamente, como una prueba de exigencia y calidad. Sin embargo, hoy se entiende que este fenómeno constituye una muestra de insuficiencia en el aseguramiento de la calidad de la educación y genera costos adicionales no solo para los estudiantes y las instituciones de educación superior, sino para toda la sociedad.

La experiencia mundial muestra la existencia de una estrecha correlación entre el nivel de desarrollo de los países, en su sentido amplio, con la fortaleza de sus sistemas educativos y de investigación científica y tecnológica. Según estudios de la Organización para la Cooperación y el Desarrollo Económico (OCDE), un año adicional de escolaridad incrementa el producto interno bruto (PIB) per cápita de un país entre 4 y 7% (Torres et al., 2015). En ese sentido, que cerca del 50% de los estudiantes universitarios no culminen sus estudios, se convierte en un gran problema para la sociedad colombiana, por el impacto negativo en el desarrollo de la nación, que implica la necesidad de profundizar la comprensión del fenómeno para establecer una solución.

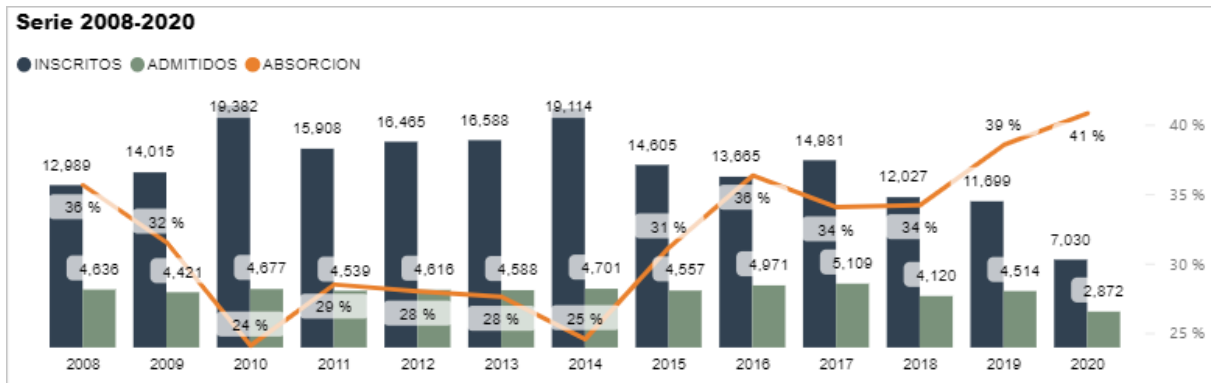
En los informes presentados por la Unidad de Información y Análisis Estadístico (UIAES) de la UIS se evidencia una reducción de la tasa de deserción durante la última década ya sea por retiro voluntario, PFU o cualquier otro motivo. Observando que el punto más alto de deserción fue en el 2012 con un 7,9%, reduciendo esta tasa hasta un 3,5% en el 2019.

Esto refleja el impacto positivo de las estrategias de retención que ha implantado la universidad en cuatro dimensiones: socioeconómica, cognitiva, académica y social afectiva. Entre

las que se incluyen programas de acompañamiento (PAMRA, MIDAS, ASAE, MANSA, FPC), auxilias estudiantiles, formación permanente de tutores, residencias, servicio de comedores. Sin embargo, la mayoría de ellas se han basado en componentes socioeconómicos y académicos, que son los más reportados en la literatura, dejando a un lado las condiciones de salud de los estudiantes, y específicamente la salud mental, que es pilar fundamental para el desempeño adecuado en cualquier actividad que realice el ser humano (Bean y Eaton, 2001). Por lo tanto, debemos tener en cuenta que la población que ingresa a la institución se compone principalmente de adolescentes y adultos jóvenes, que se encuentran en un momento de su trayectoria de vida de alta vulnerabilidad para desarrollar problemas de salud mental, entre ellos el consumo de sustancias psicoactivas, que pueden influir negativamente en el desempeño académico (Paula, 2016). Por ello se hace necesario establecer la asociación entre el problema de deserción temprana y la existencia de problemas de salud mental. Así mismo, la contribución independiente de los problemas de salud mental en la deserción temprana, teniendo en cuenta la presencia de otras variables relacionadas con las dimensiones previamente mencionadas, a través del diseño de un modelo matemático multivariado explicatorio de este fenómeno. El conocimiento generado podrá ser usado para orientar el desarrollo de nuevas estrategias de retención universitaria que mejoren esta problemática al interior de la UIS y puedan ser usados como sustrato en otras IES del País.

Figura 1.

Estudiantes inscritos, admitidos y absorción

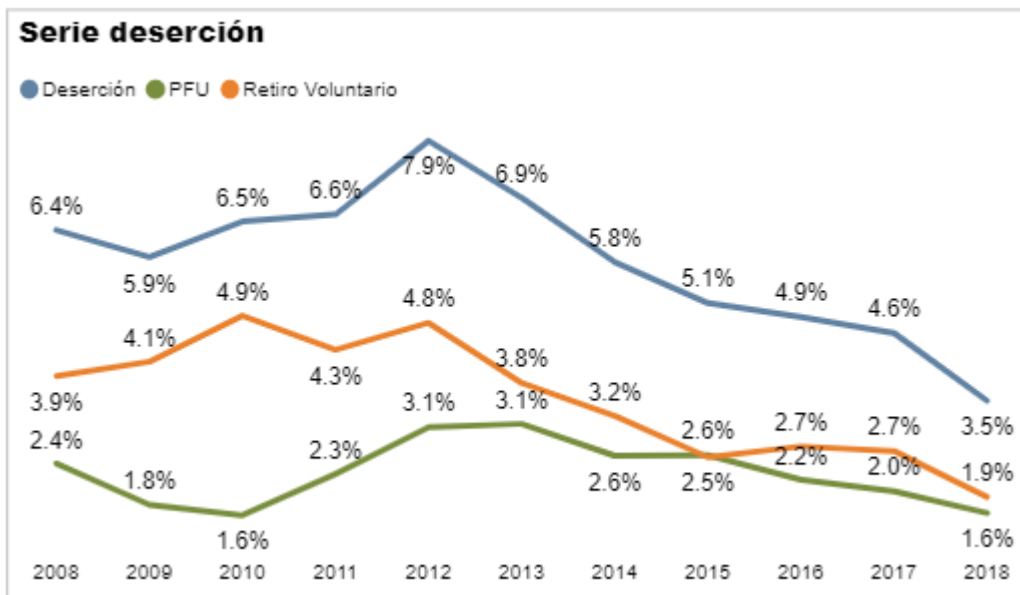


Nota. Tomado de UIS (2019) UIS Cifras

<https://www.uis.edu.co/planeacion/documentos/uisencifras/2019/index.html>

Figura 2.

Deserción



Nota. Tomado de UIS (2019) UIS Cifras

<https://www.uis.edu.co/planeacion/documentos/uisencifras/2019/index.html>

2. Hipótesis

Teniendo en cuenta que los factores asociados a la deserción son de diversa índole, se plantea una combinación de dos modelos matemáticos, en primer lugar, se construirá una red neuronal con todas las variables de entrada, posteriormente se llevará a cabo una regresión logística con un modelo multivariado, para finalmente depurar la red neuronal con las variables que estuvieron asociadas en la regresión logística para ver si mejora la precisión de la red neuronal.

3. Pregunta de investigación

¿Cuáles son los factores explicativos de la deserción temprana de los estudiantes que ingresan a un programa presencial en la Universidad Industrial de Santander?

4. Marco teórico

4.1 Conceptualización

La definición de deserción establecida por el Ministerio de Educación Nacional (MEN, 2009) en Colombia tiene en cuenta la inactividad prolongada de un estudiante en términos académicos. Esta definición refleja la idea de que la deserción no es simplemente un evento puntual, sino un proceso que se desarrolla a lo largo del tiempo. Para nuestro estudio, nos centramos en la deserción universitaria temprana que es aquella que ocurre en los primeros dos años del programa académico (MEN, 2009).

En ese sentido, la comprensión de los factores que influyen en la deserción universitaria es esencial para identificar los estudiantes que se encuentran en riesgo desde el momento que ingresan a la universidad, para implementar estrategias efectivas que fomenten la retención y el éxito estudiantil. Como parte de esos esfuerzos, es importante resaltar los aportes de los modelos sociológicos desarrollados por Spady, Tinto y Bean, en las décadas de 1970 y 1980, proporcionaron un marco teórico inicial para analizar los aspectos complejos que rodean la deserción.

Spady (1970), abordó la deserción desde una perspectiva amplia, considerando factores externos a la institución que pueden afectar la permanencia de un estudiante en la universidad. Tinto (1998), por su parte, enfatizó la importancia de la integración social y académica en la retención, al sugerir, que los estudiantes que se sienten conectados tanto social como académicamente con la institución, tienen más probabilidades de completar sus estudios.

Finalmente, Bean (1985) destaca la importancia de la adaptación del estudiante al entorno universitario, y cómo las dificultades en este proceso pueden llevar a la deserción.

Ahora bien, en el contexto de Colombia, las estadísticas presentadas por el programa PISA en 2015 revelaron un bajo rendimiento en áreas clave como matemáticas, lectura crítica y ciencias. Este bajo rendimiento podría influir negativamente en la percepción de los estudiantes sobre su propia competencia académica y, en última instancia, afectar su autoconcepto y la motivación para continuar con sus estudios. En consecuencia, es importante abordar, no sólo los factores internos de la institución, sino también las influencias externas a los que han sido expuestos los estudiantes durante la formación académica previa que también pueden contribuir a la deserción universitaria (Universidad de los Andes, 2006).

En ese orden de ideas, las habilidades para la vida que se van desarrollando simultáneamente con la adquisición de conocimientos disciplinares, son esenciales para el éxito universitario, ya que le permiten al estudiante desarrollar una actitud positiva hacia la salud y afrontar las exigencias de la vida cotidiana posibilitando el mayor desarrollo del capital humano. Como respuesta a estas necesidades manifiestas en la universidad se ha avanzado en la actualización del modelo pedagógico, que le da un mayor peso al componente electivo y de formación integral en desarrollo de la malla curricular de los programas académicos (Universidad Industrial de Santander, 2021). De esta manera, se propende para que los estudiantes no solo desarrollen competencias disciplinares, sino que también deben desarrollar inteligencia social, habilidades cognitivas e inteligencia emocional, que les permitan interactuar de manera asertiva con su entorno sociocultural y asuman mayor responsabilidad social que contribuyan a la transformación social como se ha planteado desde la Organización Mundial de la Salud (Organización Mundial de la Salud, 1997).

La clasificación de la deserción en función del tiempo y el espacio proporciona una visión más detallada de los momentos y contextos en los que los estudiantes enfrentan mayores desafíos para mantener su compromiso académico. Al analizar las etapas críticas que experimenta un estudiante con riesgo de deserción, tanto en el momento de admisión como durante los primeros semestres, se denota la importancia de brindar un apoyo sólido desde el principio de la experiencia universitaria. Es aquí, donde los programas de orientación y tutoría pueden desempeñar un papel fundamental en el proceso de adaptación de los estudiantes a su nuevo entorno (Universidad de La Salle, 2003).

En conclusión, el análisis de la deserción estudiantil abarca múltiples factores, como se menciona desde los modelos teóricos iniciales, además de considerar las condiciones socioeconómicas, culturales y educativas específicas de cada país. Abordar este fenómeno requiere un enfoque holístico, que considere tanto los factores internos como los externos, para generar estrategias de apoyo y retención que ayuden a los estudiantes a superar desafíos y asegurar su éxito académico.

4.2 Factores que contribuyen a la deserción

El Ministerio de Educación Nacional (MEN) en su informe del año 2014 sostiene que la deserción estudiantil es un fenómeno complejo que involucra diversos factores interrelacionados, es decir; que es el resultado de la interacción de múltiples categorías de factores y no del efecto individual de una categoría particular.

Castaño, Gallón, Gómez y Vásquez (2007, citados por MEN 2009), realizaron un resumen del estado del arte de los factores determinantes de la deserción agrupando las variables asociadas

en cuatro categorías: Factores individuales, Factores académicos, Factores institucionales y Factores socioeconómicos. Ver Tabla 1

4.2.1 Factores individuales

Los factores individuales juegan un papel fundamental en la decisión de un estudiante de abandonar sus estudios universitarios. Los rasgos de personalidad, la motivación, la adaptación al entorno académico y social, y la autoeficiencia académica, pueden influir en la probabilidad de abandonar los estudios. La sensación de pertenencia y la identificación con la institución también son factores clave que afectan la retención.

Adicionalmente, la personalidad del estudiante también juega un papel esencial, especialmente considerando que muchos estudiantes ingresan a la educación superior en la etapa de la adolescencia, una fase en la que aún se están definiendo aspectos de la personalidad. Esto puede resultar en estrategias de afrontamiento inadecuadas que dificultan la adaptación a la vida universitaria. Tales dificultades, pueden llevar a sentimientos de frustración, síntomas depresivos y ansiosos, lo que a su vez puede afectar negativamente el rendimiento académico y la retención en la institución (Aragón, 2011).

De igual forma, la transición desde la enseñanza media a la educación superior se considera una etapa crítica en la vida de un estudiante. Los primeros cortes del primer semestre a menudo presentan dificultades, ya que el estudiante abandona un entorno familiar para adentrarse en uno completamente nuevo. El éxito de esta transición depende en gran medida de cómo el estudiante se adapta al nuevo entorno y cumple con las responsabilidades académicas y obligaciones institucionales (Pérez et al., 2020).

4.2.2 Factores académicos

El rendimiento académico, las calificaciones y la satisfacción con el programa de estudio, son componentes cruciales que están relacionados con la deserción. Los estudiantes que enfrentan dificultades en cursos específicos o que no ven relevancia de los contenidos con su formación, tienen más probabilidades de considerar el abandono del programa académico. Además, la falta de claridad de los objetivos académicos, también pueden influir en la deserción (Universidad Nacional de Colombia, 2002).

4.2.3 Factores institucionales

En el análisis del fenómeno de la deserción estudiantil, una de las variables cruciales es la consideración de los factores que motivan a un estudiante a elegir una Institución de Educación Superior (IES) y un programa académico específico. La experiencia puede variar significativamente según si un individuo escoge una institución por afinidad con sus valores y objetivos, o simplemente por la falta de otras opciones disponibles o motivos económicos. Esta elección inicial puede influir de manera significativa en la deserción precoz o temprana, ya que un estudiante que no comparte la visión y la misión institucional puede tener dificultades para adaptarse y enfrentar los desafíos iniciales (Song et al., 2012).

Como se puede apreciar, el entorno institucional desempeña un rol significativo en la retención estudiantil. La calidad de la enseñanza, el apoyo académico, las oportunidades de participación en actividades extracurriculares y la accesibilidad a recursos de apoyo, impactan la decisión de un estudiante de permanecer en la universidad. Las políticas institucionales en cuanto

a la flexibilidad de la malla curricular, el acompañamiento a estudiantes y la comunicación eficiente, también son determinantes (Universidad Pontificia Bolivariana, seccional Bucaramanga, 2005).

4.2.4 Factores socioeconómicos

Los aspectos socioeconómicos, como el nivel de ingresos familiar, la situación laboral del estudiante y las restricciones económicas, tienen una relación directa con la deserción estudiantil. Los estudiantes que enfrentan dificultades financieras pueden verse obligados a abandonar sus estudios con el fin de buscar un empleo, o colaborar en la actividad económica familiar (Palacio et al., 2020). Además, el acceso a becas y programas de apoyo financiero puede influir en la retención estudiantil.

Tabla 1.

Determinantes de la deserción

Determinantes de la Deserción			
Individuales	Académicos	Institucionales	Socioeconomicos
Edad, género y estado civil	Orientación profesional	Normalidad académica	Estrato
Posición dentro de los hermanos	Tipo de colegio	Becas y formas de financiamiento	Situación laboral
Entorno familiar	Rendimiento académico	Recursos universitarios	Situación laboral de los padres e ingresos
Calamidad y problemas de salud	Calidad del programa	Orden público	Dependencia económica
Integración social	Métodos de estudio	Entorno político	Personas a cargo
Incompatibilidad horaria con	Resultado examen de ingreso	Nivel de interacción personal con profesores	Nivel educativo de los padres

Determinantes de la Deserción			
Individuales	Académicos	Institucionales	Socioeconomicos
actividades extraacadémicas			
Expectativas no satisfechas	Insatisfacción con el programa y otros factores	Apoyo académico	Entorno macroeconómico del país
Embarazo	Número de materias	Apoyo psicológico	

4.2.5 Factores de salud mental

La salud mental es un aspecto esencial del bienestar de los individuos, y su importancia en la población universitaria no puede ser subestimada. La universidad no solo cumple un papel formativo desde una perspectiva instrumental y técnica en el desarrollo de futuros profesionales, también es vital para el desarrollo de habilidades sociales y estrategias de adaptabilidad. La transición a la vida universitaria puede traer consigo una serie de desafíos y presiones únicas que afectan la salud mental de los estudiantes (Alarcón, 2019). La elevada exigencia académica, la presión por obtener buenos resultados, el desplazamiento desde otras ciudades y la integración en nuevos entornos sociales y culturales pueden representar un proceso disruptivo para muchos individuos por lo que es crucial reconocer la importancia de cuidar y promover la salud mental en este grupo demográfico, dado que puede influir de manera significativa en su éxito académico, sus relaciones interpersonales y su calidad de vida en general.

En un estudio llevado a cabo Bélgica, en una muestra de 4921 estudiantes universitarios de primer año, aproximadamente uno de cada tres estudiantes reportó haber tenido problemas de salud mental durante el último año, los cuales tenían un menor rendimiento académico al final del periodo académico (Bruffaerts et al., 2018). En Colombia, en un estudio de revisión narrativa, la

prevalencia de depresión osciló entre 7,4 y 9,6%, y en cuatro de los estudios incluidos se reportó una asociación positiva de depresión con bajo rendimiento académico (García et al., 2019).

4.2.5.1 Depresión. La depresión es uno de los problemas de salud mental más común entre los estudiantes universitarios. El estrés por el rendimiento y la adaptación a un nuevo entorno pueden precipitar sentimientos de tristeza, minusvalía, pérdida de la motivación y fatiga. La sobrecarga de trabajo, los horarios ajustados y la sensación de competir constantemente, también pueden aumentar el riesgo de depresión en esta población (Silva-Ramos et al., 2020).

En Perú, un estudio llevado a cabo en la Universidad Cayetano Heredia, con una muestra de 378 estudiantes de diferentes semestres de medicina se encontró una prevalencia de síntomas de corte depresivo de 29.9%. (Osada et al., 2010).

En Colombia, en la encuesta nacional de salud mental de 2015 se documentó una prevalencia de vida en adolescentes entre los 12 y los 17 años para cualquier trastorno depresivo del 2,4% y en adultos jóvenes entre los 18 y los 44 años del 7,3%, pero en los estudios llevados a cabo en universitarios se evidencia una prevalencia mucho mayor como se ha reportado en un estudio realizado en la Universidad CES de Medellín en una muestra de 1344 estudiantes entre los 16 y los 37 años, encontrando hasta una prevalencia de un 47,3 %, donde además se encontró que la mayoría de estudiantes que padecen episodios depresivos severos tenían problemas académicos moderados o severos. (Gutiérrez Rodas et al., 2010).

4.2.5.2 Ansiedad. La ansiedad es otro problema significativo en la población universitaria. Las altas expectativas académicas, la presión por los exámenes, las demandas sociales y la incertidumbre sobre el futuro, pueden generar niveles elevados de ansiedad. Los trastornos de

ansiedad, como el trastorno de ansiedad generalizada y los ataques de pánico, son frecuentes en adolescentes y adultos jóvenes, y pueden interferir con el rendimiento académico y la calidad de vida de los estudiantes (Jadue, 2001).

Adicionalmente, la depresión y la ansiedad pueden acompañarse de alteraciones en la atención sostenida, debido a la presencia de pensamientos automáticos negativos o cogniciones catastróficas respectivamente que pueden distraer a los estudiantes de sus actividades académicas. Dicha incapacidad para enfocarse en el trabajo escolar puede conllevar un menor rendimiento en exámenes por una menor comprensión de los conceptos clave y las dificultades para completar las tareas. Además, estos problemas de salud mental están asociados con una disminución de la motivación y el interés en las actividades diarias. Asimismo, los estudiantes que padecen problemas de salud mental presentan mayor ausentismo, dejan de participar en los debates académicos, se vuelven inactivos y dedican menos tiempo a la preparación académica, dando como resultado un desempeño académico inferior y un menor compromiso con los estudios (Cuamba et al., 2020).

En la encuesta nacional de salud mental, se reportó que la prevalencia de vida para los trastornos de ansiedad en adolescentes entre los 12 y los 17 años era del 10,2%, siendo el más la fobia social. En la población universitaria, se evidencia una prevalencia mayor de ansiedad en un estudio realizado en la Universidad de Santa Marta en el que participaron 1349 estudiantes de distintos programas, encontrándose una sintomatología ansiosa en un 34% de la muestra estudiada (Ceballos et al., 2013). De igual manera, en una muestra de 200 estudiantes de la Universidad Cooperativa de Colombia sede Medellín, se documentó una prevalencia de ansiedad del 58%. (Cardona-Arias et al., 2015).

4.2.5.3 Estrés. El estrés es una respuesta natural a las demandas y desafíos de la vida universitaria. Sin embargo, cuando los individuos se exponen a situaciones de estrés crónico y se agotan los mecanismos de adaptación, pueden ocasionar un impacto negativo en la salud mental. Los estudiantes a menudo enfrentan múltiples presiones, incluidas la carga académica, presiones sociales y dificultades económicas. Cuando la información percibida del evento estresante no logra ser procesada de forma adecuada, puede dar lugar a expresiones psicopatológicas como ansiedad, depresión o alteraciones conductuales (Silva-Ramos et al., 2020).

De la misma manera, el estrés crónico y la ansiedad pueden interferir con la formación de nuevas conexiones neuronales, lo que dificulta el proceso de aprendizaje, y, además, genera fatiga mental y emocional, lo que puede disminuir la capacidad de recuperar información durante exámenes y evaluaciones (Duval et al., 2010). Un estudio en la universidad autónoma de Aguascalientes con una muestra de 255 estudiantes se observó que el 86,3% de la población presentaron un nivel de estrés moderado (Silva-Ramos et al., 2020).

Los problemas de salud mental también pueden hacer que las tareas académicas sean más desafiantes de completar en el tiempo asignado. La falta de energía, la procrastinación y la falta de concentración pueden llevar a retrasos en la finalización de tareas y proyectos. Este retraso puede afectar la calidad del trabajo y reflejarse negativamente en la calificación final (Silva-Ramos et al., 2020).

4.2.5.4 Trastornos alimentarios. Se ha observado que los hábitos alimentarios saludables se correlacionan con un mejor rendimiento académico (Ibarra Mora et al., 2019). En la población universitaria, se ha observado un incremento de la prevalencia de trastornos de la conducta alimentaria, como la anorexia nerviosa, la bulimia y la ortorexia. La presión por mantener ciertos

estándares de apariencia, la comparación con compañeros y la búsqueda de un sentido de control, pueden contribuir a la aparición de estos trastornos (Salazar-Rendón et al., 2018). En ese sentido, los estudiantes con malos hábitos alimentarios pueden llegar a experimentar dificultades para concentrarse o falta de energía para desarrollar las actividades académicas, debido a que se encuentran desnutridos o malnutridos.

De acuerdo con la Encuesta Nacional de Salud Mental (ENSM, 2015), la población adolescente, el 9,5% de los hombres y el 8,8% de las mujeres encuestadas reportaron algún tipo de comportamiento de riesgo para trastornos de la conducta alimentaria.

4.2.5.5 Trastorno por uso de sustancias psicoactivas. En Colombia, la Encuesta Nacional de Salud Mental (ENSM, 2015) y el Estudio Nacional de Estupefacientes (ENCSPA, 2019) ha revelado una reducción en la edad de inicio del consumo de sustancias psicoactivas, por lo que el impacto es mayor teniendo en cuenta que se está afectando el desarrollo del sistema nervioso central, y las consecuencias pueden perdurar para toda la vida. Como se ha mencionado la transición de la adolescencia a la vida adulta es una etapa de vulnerabilidad, ya que las personas tienden a tomar decisiones de manera impulsiva, por lo que no perciben los riesgos derivados del consumo de psicoactivos, además en esta etapa se describen mayor prevalencia de ansiedad social y recurren a las sustancias psicoactivas por su efecto entactógeno para facilitar la socialización con los compañeros que en los primeros semestres son personas desconocidos. La exposición a las sustancias psicoactivas incrementa el riesgo de desarrollar trastornos mentales y acentuar rasgos desadaptativos de personalidad que pueden tener un impacto negativo en el rendimiento académico, social y familiar, contribuyendo a la deserción universitaria (Navalón y Ruiz-Callado, 2017).

En resumen, la deserción estudiantil es un fenómeno complejo y multifacético. Los factores individuales, la personalidad, el entorno socioeconómico y el consumo de sustancias, entre otros, interactúan de manera única en cada estudiante y pueden influir en su decisión de abandonar la educación superior. Por lo tanto, abordar este problema requiere de estrategias holísticas que consideren la variedad de factores y circunstancias que rodean a cada estudiante.

4.3 Modelos matemáticos

El modelamiento matemático, es el proceso de usar ecuaciones y conceptos matemáticos, para crear una representación simplificada de los sistemas del mundo real, con el propósito de comprender, predecir y tomar decisiones sobre su comportamiento.

Los modelos matemáticos se clasifican en diversas categorías según su estructura, función y aplicabilidad. Estas categorías permiten agruparlos en función de sus características y las clases de problemas que son capaces de abordar. A continuación, se presenta una descripción de las principales clasificaciones de modelos matemáticos y ejemplos de los problemas en los que se utilizan:

4.3.1 Modelos determinísticos y estocásticos

4.3.1.1 Determinísticos. Un modelo determinista es un modelo matemático donde las mismas entradas o condiciones iniciales producirán invariablemente las mismas salidas o resultados, no contemplándose la existencia de azar, o incertidumbre en el proceso modelada mediante dicho modelo (Insúa et al., 2004).

Un ejemplo de estos modelos son los de crecimiento poblacional que es de la forma:

$$P(t) = P_0 \cdot e^{rt} \tag{1}$$

donde:

- $P(t)$ es la población en el tiempo t
- P_0 es la población inicial
- r es la tasa de crecimiento per cápita
- e es la base del logaritmo natural

En este modelo, la población crece de manera proporcional a su tamaño actual, lo que da lugar a un crecimiento exponencial.

4.3.1.2 Estocásticos. Un modelo estocástico es un tipo de modelo matemático que incorpora elementos de azar o aleatoriedad en su estructura para representar la variabilidad inherente a un sistema.

En un modelo estocástico, las variables pueden seguir distribuciones de probabilidad y la evolución del sistema se describe en términos de probabilidades. Esto refleja la naturaleza inherentemente incierta de ciertos fenómenos y permite capturar mejor la variabilidad observada en el mundo real (Insúa et al., 2004).

Un ejemplo de estos modelos es la simulación de Monte Carlo que utiliza números aleatorios para modelar fenómenos y realizar análisis probabilísticos y es de la forma:

$$E = \frac{1}{N} \sum_{i=1}^N f(X_i) \tag{2}$$

donde:

- E es la estimación del resultado deseado

- N es el número total de simulaciones
- $f(X_i)$ es el valor de la función o resultado que se quiere estimar en la i -ésima simulación.
- X_i son variables aleatorias que representan las condiciones o escenarios de cada simulación.

La precisión de la estimación mejora a medida que se aumenta el número de simulaciones (N). En muchos casos, la convergencia al valor real o teórico es regida por la ley de los grandes números.

4.3.2 Modelos continuos y discretos

4.3.2.1 Continuos. Los modelos continuos son representaciones matemáticas de sistemas o fenómenos que varían de manera suave y continua en el tiempo o en el espacio. Estos modelos son utilizados para describir procesos que no experimentan cambios abruptos, sino que exhiben variaciones suaves y sin interrupciones. La continuidad se refiere a la ausencia de saltos o discontinuidades en las funciones matemáticas que describen el comportamiento del sistema.

Un modelo continuo puede tomar varias formas, pero generalmente se expresa mediante ecuaciones diferenciales o integrales. Estas ecuaciones describen cómo cambian las variables del sistema en función del tiempo o de otra variable independiente de manera continua (de la Horra Navarro, 2018).

Un ejemplo de estos modelos es la ecuación de onda que describe la propagación de ondas a lo largo del tiempo y el espacio, la ecuación de onda unidimensional es de la forma:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \cdot \frac{\partial^2 u}{\partial x^2} \quad (3)$$

donde u es la función de onda, t es el tiempo, x es la posición y c es la velocidad de propagación.

4.3.2.2 Discretos. Los modelos discretos son representaciones matemáticas de sistemas o fenómenos que evolucionan en pasos discretos en lugar de manera continua en el tiempo o el espacio. Estos modelos describen sistemas que experimentan cambios o eventos en momentos específicos y discretos, en contraste con los modelos continuos, que representan cambios de manera suave y continua.

En un modelo discreto, el tiempo y, en algunos casos, el espacio, se dividen en unidades discretas o intervalos. La evolución del sistema ocurre en pasos discretos, y las variables del sistema cambian solo en puntos específicos en el tiempo. Los modelos discretos son particularmente útiles para describir sistemas que tienen comportamientos discretos o eventos distintos (González Manteiga, 2003).

Un ejemplo es el modelo de crecimiento exponencial discreto que se interpreta de la forma:

$$P_{t+1} = P_t + r \cdot P_t \quad (4)$$

Donde P_t es la población al tiempo t y r es la tasa de crecimiento per cápita. Esta ecuación indica que la población al tiempo $t + 1$ es igual a la población al tiempo t más el crecimiento proporcional $r \cdot P_t$.

4.3.3 Modelos lineales y no lineales

4.3.3.1 Lineales. Los modelos lineales son un conjunto de técnicas estadísticas y matemáticas utilizadas para describir la relación entre una variable dependiente y una o más variables independientes. La característica fundamental de estos modelos es que asumen una relación lineal entre las variables, lo que implica que los cambios en la variable dependiente son proporcionales a los cambios en las variables independientes, con una constante llamada coeficiente (de la Horra Navarro, 2018).

La forma general de un modelo lineal simple puede expresarse como:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \epsilon \quad (5)$$

Donde:

- Y es la variable dependiente
- X_1, X_2, \dots, X_n son las variables independientes
- $\beta_0, \beta_1, \dots, \beta_n$ son los coeficientes que representan la pendiente de la relación entre las variables.
- ϵ es el término de error, que captura la variabilidad no explicada por el modelo.

4.3.3.2 No lineales. Los modelos no lineales son un conjunto de técnicas y enfoques que se utilizan para describir relaciones entre variables donde la relación no sigue una forma lineal. A diferencia de los modelos lineales, que asumen una relación directamente proporcional entre las variables, los modelos no lineales permiten capturar patrones más complejos y no lineales en los datos.

La forma general de un modelo no lineal puede ser más compleja y variada, dependiendo de la naturaleza específica del problema y de los datos. Algunos ejemplos de modelos no lineales incluyen polinomios de grados superiores, funciones exponenciales, logarítmicas, trigonométricas, sigmoidales, entre otros (de la Horra Navarro, 2018).

Por ejemplo, un modelo no lineal simple podría tener la forma:

$$Y = \beta_0 + \beta_1 X_1^2 + \beta_2 \text{sen}(X_2) + \epsilon \quad (6)$$

Donde:

- Y es la variable dependiente
- X_1 y X_2 son variables independientes
- β_0, β_1 y β_2 son coeficiente
- ϵ es el término de error

La flexibilidad de los modelos no lineales permite capturar comportamientos más complejos y adaptarse mejor a situaciones en las que las relaciones son intrínsecamente no lineales.

4.3.4 Modelos de optimización

Los modelos de optimización son representaciones matemáticas o computacionales que buscan encontrar la mejor solución posible para un problema específico, sujeto a ciertas restricciones. Estos problemas suelen implicar la maximización o minimización de una función objetivo, que representa la cantidad que se desea optimizar, y están sujetos a limitaciones o restricciones que deben cumplirse. El objetivo es encontrar un conjunto de valores para las variables de decisión que optimice la función objetivo dentro de los límites establecidos por las restricciones (Morales et al., 2011).

Un ejemplo de un modelo de optimización podría ser la maximización del área de un rectángulo, teniendo una cuerda de longitud fija L , y que se quiere utilizar para construir un rectángulo.

Variables de decisión:

l : Longitud de un lado del rectángulo (la mitad de la longitud de la cuerda).

w : Longitud del otro lado del rectángulo (la otra mitad de la longitud de la cuerda).

Función objetivo para maximizar:

$$Area = l \times w \quad (7)$$

Restricciones:

- La longitud total de la cuerda es L :

$$2l + 2w = L \quad (8)$$

- Las longitudes de los lados del rectángulo deben ser positivas:

$$l > 0 ; w > 0 \quad (9)$$

En este problema, se desea encontrar los valores óptimos de l y w que maximizan el área del rectángulo, sujeto a la restricción de que la longitud total de la cuerda es constante.

4.3.5 Modelos de simulación

Los modelos de simulación son herramientas que buscan reproducir el comportamiento de sistemas del mundo real mediante representaciones matemáticas o computacionales. Estos modelos permiten experimentar virtualmente con el sistema, observando cómo evolucionan las variables a lo largo del tiempo bajo diferentes condiciones. La simulación ofrece la posibilidad de analizar el comportamiento dinámico, entender interacciones complejas y evaluar escenarios

variados sin realizar cambios directos en el sistema real. Este enfoque flexible y experimental facilita la toma de decisiones informadas en diversos campos, desde ingeniería y ciencia hasta planificación urbana y gestión de operaciones (Cristina et al., 2017).

Una aplicación de estos modelos es la simulación de una red de comunicación como por ejemplo optimizar la eficiencia de transmisión de datos en una red de computadoras.

Construcción del modelo

- **VARIABLES DE DECISIÓN**

- *B*: Ancho de banda disponible en la red
- *L*: Longitud de los paquetes de datos transmitidos
- *R*: Tasa de error en la transmisión
- *D*: Retardo de propagación en la red

- **RESTRICCIONES**

- $B > 0$: El ancho de banda debe ser positivo
- $L > 0$: La longitud de los paquetes debe ser positiva
- $0 \leq R \leq 1$: La tasa de error está entre 0% y 100%
- $D \geq 0$: El retardo de propagación debe ser no negativo

- **Función objetivo**

- Maximizar la tasa de transferencia de datos efectiva:

$$\text{Maximizar } \frac{B \cdot (1 - R)}{L + D} \tag{10}$$

Este modelo de simulación podría utilizarse para optimizar la configuración de una red de computadoras, considerando diferentes factores que afectan la transmisión de datos. La simulación permitiría probar diversas configuraciones y parámetros para lograr una transmisión de datos más eficiente en la red.

4.3.6 Modelos de redes

Hacen referencia a representaciones matemáticas o computacionales de las interconexiones y relaciones entre individuos o entidades en una red social. Estos modelos buscan capturar la estructura y dinámica de las interacciones sociales dentro de una comunidad en línea.

Un ejemplo son los modelos de redes sociales que se pueden representar de la siguiente forma:

Nodos

$$V = \{Miguel, Jorge, Lina\}$$

Enlaces o conexiones

$$E = \{(Miguel, Jorge), (Lina, Miguel), (Jorge, Lina)\}$$

Matriz de adyacencia

La matriz A donde $A_{ij} = 1$ si hay una conexión entre i y j , y $A_{ij} = 0$ de lo contrario

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Vector de grado

d_i es el grado del nodo i , es decir, el número de conexiones que tiene

$$d = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Centralidad de grado

$C_d(i)$ es la centralidad de grado del nodo i , normalizada por el máximo grado posible en el grafo

$$C_d(i) = \frac{d_i}{\max(d)} \quad (11)$$

Cercanía

$C_c(i)$ es la cercanía del nodo i , inversamente proporcional a la suma de las distancias a todos los demás nodos

$$C_c(i) = \frac{1}{\sum_{j \neq i} d(i, j)} \quad (12)$$

Estas fórmulas representan medidas comunes en el análisis de redes sociales. La matriz de adyacencia muestra las conexiones entre nodos, el vector de grado refleja cuántas conexiones tiene cada nodo, la centralidad de grado mide la importancia relativa de cada nodo en función de su grado, y la cercanía evalúa cuán cerca está un nodo de los demás en términos de distancias de conexión.

Estas métricas pueden proporcionar información sobre la estructura y la importancia de los nodos en una red social y son utilizadas en análisis más detallados de la topología y dinámicas de la red (de la Horra Navarro, 2018).

4.3.7 Modelos de probabilidad y estadística

Los modelos de probabilidad y estadística son herramientas matemáticas y conceptuales utilizadas para describir, analizar y comprender la variabilidad, la incertidumbre y las relaciones dentro de conjuntos de datos o fenómenos específicos.

4.3.7.1 Modelos de probabilidad. Representan la posibilidad numérica de eventos aleatorios y se basan en la teoría de la probabilidad. Estos modelos cuantifican la incertidumbre y permiten la evaluación de escenarios futuros (Peña, 2014).

Un ejemplo de estos modelos es el lanzamiento de un dado, supongamos que queremos modelar el lanzamiento de un dado justo de seis caras. En este caso, el modelo de probabilidad se basa en la idea de que cada cara tiene la misma probabilidad de aparecer. El modelo sería de la forma:

- **Espacio muestral (S)**

El conjunto de todos los posibles resultados del lanzamiento del dado. En este caso sería,
 $S = \{1,2,3,4,5,6\}$

- **Probabilidad de cada resultado**

Dado que el dado es justo, la probabilidad de cada resultado individual es $P(1) = P(2) = P(3) = P(4) = P(5) = P(6) = \frac{1}{6}$

- **Evento**

Se puede definir el evento, como la probabilidad de obtener un número par. El evento E sería $E = \{2,4,6\}$.

La probabilidad de este evento es $P(E) = P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$

Este modelo de probabilidad refleja la igualdad de oportunidades para cada resultado individual en el lanzamiento de un dado justo. Proporciona una base para calcular la probabilidad de eventos específicos relacionados con los resultados del dado, y es un ejemplo fundamental de un modelo de probabilidad discreta.

4.3.7.2 Estadística. Se enfoca en la recopilación, análisis e interpretación de datos observados. Los modelos estadísticos describen patrones, tendencias y relaciones en los datos, y son fundamentales para realizar inferencias y tomar decisiones informadas (Peña, 2014).

Un ejemplo conocido de estos modelos es la distribución de Poisson que describe la probabilidad de que ocurran un número específico de eventos raros en un intervalo fijo de tiempo o espacio. La función de probabilidad de Poisson es:

$$P(x = k) = \frac{e^{-\lambda} \lambda^k}{k!} \quad (13)$$

Donde k es el número de eventos que se quieren moldear y λ es la tasa promedio de ocurrencia de eventos por unidad de tiempo o espacio.

4.3.8 Modelos de computación

Los modelos de computación son descripciones abstractas y teóricas que representan el funcionamiento de los sistemas computacionales y la ejecución de algoritmos. Estos modelos proporcionan marcos conceptuales que permiten comprender, analizar y diseñar procesos de cálculo. Se utilizan en diversas disciplinas y abarcan desde máquinas abstractas, como la Máquina de Turing, hasta modelos de computación cuántica y sistemas distribuidos. Estos modelos ofrecen herramientas para explorar las capacidades y limitaciones de los sistemas computacionales, así como para desarrollar algoritmos eficientes y resolver problemas en el ámbito de la informática y la ciencia computacional.

Los problemas que se abordan por medio de modelos matemáticos son diversos y se encuentran en una amplia gama de campos, como la física, la biología, la economía, la ingeniería, la medicina y más. Estos modelos permiten comprender, predecir y tomar decisiones informadas

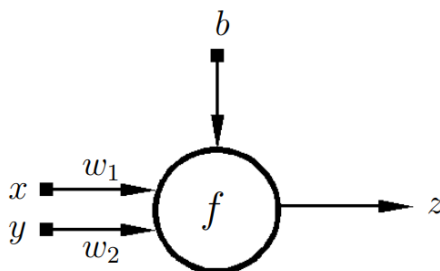
en situaciones complejas y realistas, lo que los convierte en herramientas esenciales en la resolución de problemas en diferentes disciplinas (de la Horra Navarro, 2018).

4.3.8.1 Redes neuronales artificiales. La investigación sobre las redes neuronales ha captado la atención de diversos grupos de estudio. Inicialmente, se abordó desde una perspectiva biológica, con el propósito de entender el funcionamiento del cerebro y los procesos neuronales asociados. Más tarde, se extendió a la creación de sistemas "inteligentes" capaces de llevar a cabo una variedad de actividades. Desde sus inicios, las redes neuronales artificiales han demostrado su utilidad en aplicaciones específicas. Este éxito ha impulsado, especialmente en fechas recientes, un renovado interés en explorar los fundamentos teóricos que gobiernan la dinámica de estos sistemas (Prieto et al., 2020).

El modelo neuronal introducido por Warren S. McCulloch y Walter Pitts en 1943 se destaca como el pionero de los modelos neuronales modernos. Este modelo ha servido como punto de partida fundamental para el desarrollo de numerosos modelos neuronales contemporáneos. Además, se ha convertido en un referente esencial para la evaluación y comparación del rendimiento de otros modelos en el ámbito de las redes neuronales.

Figura 3.

Modelo de McCulloch-Pitts para una neurona artificial



Las entradas x e y son el estímulo que la neurona artificial recibe del entorno que la rodea, y la salida Z es la respuesta a tal estímulo. La neurona se adapta al medio circundante y aprende de él modificando el valor de sus pesos sinápticos W_1 y W_2 y su término aditivo b . Estos son conocidos como los parámetros libres del modelo, pues los mismos pueden ser modificados y adaptados para realizar una tarea determinada. En este modelo, la salida neuronal Z está dada por:

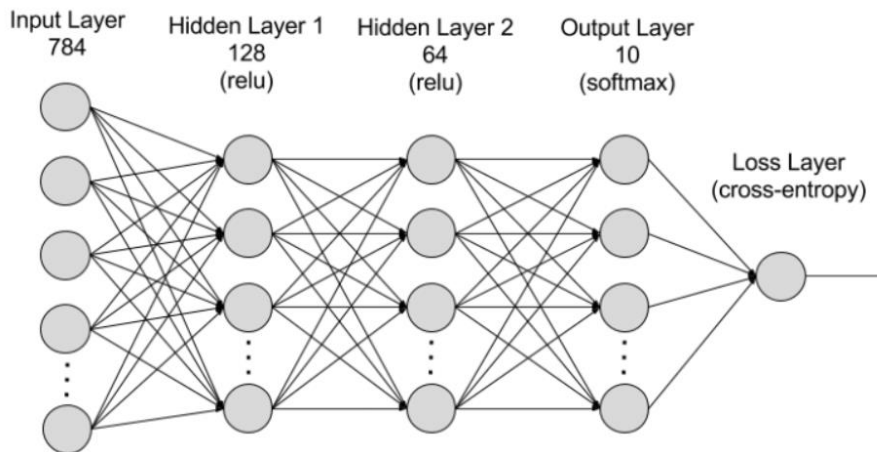
$$Z = F(W_1x + W_2y + b) \tag{14}$$

La función de activación F es seleccionada de acuerdo con la tarea realizada por la neurona.

Las redes neuronales artificiales son un tipo de modelo de aprendizaje automático, que busca imitar el funcionamiento de las redes neuronales biológicas presentes en el cerebro humano. Estas redes son capaces de aprender patrones y relaciones complejas a partir de datos, lo que las convierte en una herramienta poderosa para tareas como clasificación, reconocimiento de patrones, procesamiento de lenguaje natural y más.

Figura 4.

Ejemplo de arquitectura de una RNA



Nota. Tomado de: Amazon (s.f.) *Neural network* <https://aws.amazon.com/es/what-is/neural-network/>

4.3.8.1.1 Estructuras de las redes neuronales artificiales. Estos sistemas capaces de procesar gran cantidad de información mediante modelos matemáticos (Yen y Langari, 2000) poseen varios grados de conectividad cuyo ajuste determina el desempeño de la red neuronal.

La configuración inicial de la RNA está compuesta por una matriz X , de vectores con n entradas $X_p = (X_{p1}, X_{p2}, \dots, X_{pN})^T$, que a su vez tienen una estructura de pesos sinápticos (W_{ij}^h) que son utilizados para la ponderación de los valores de entrada en la configuración interna de la red. A su vez, esta configuración se compone también de un umbral o sesgo (θ_j^h) que representa que tan predispuesta está la neurona a disparar un 1 o un 0 independientemente de los pesos, como se observa en (15), el cual se aplica para reducir la entrada a la función de activación.

$$net_{pj}^h = \sum_{i=1}^N W_{ji}^h X_{pi} + \theta_j^h \quad (15)$$

Otro componente de una RNA es el operador matemático de suma. Este realiza la operación de suma de las variables de entrada, multiplicadas por sus respectivos pesos. Por último, existe la función de activación o transferencia (I_{pj}) donde se define la salida de la neurona (net_{pk}^o, O_{pk}) . Para el proceso de aprendizaje de la red neuronal es conocido como algoritmo de aprendizaje, como se expresa en (16) a (18) y el objetivo del algoritmo de aprendizaje es modificar los pesos para optimizar la función de activación que permite decidir si la neurona debe activarse o no con base a su entrada acumulada (Fullér, 1995).

$$I_{pj} = f_j^h(net_{pj}^h) \quad (16)$$

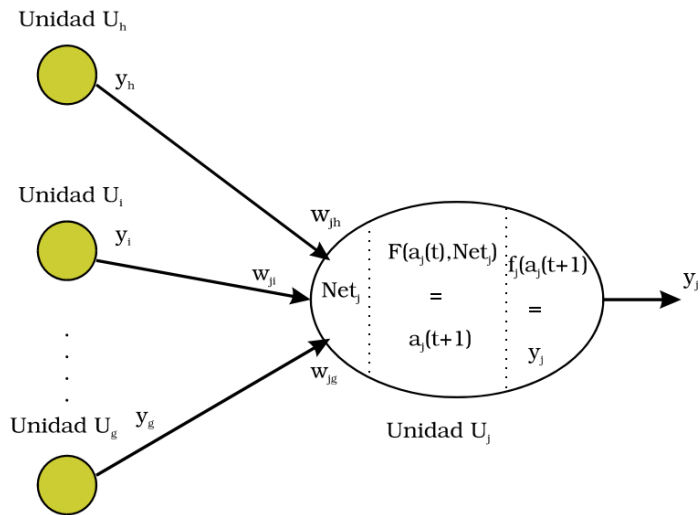
$$net_{pk}^o = \sum_{j=1}^L W_{kj}^o X_{pj} + \theta_k^o \quad (17)$$

$$O_{pk} = f_k^o(net_{pk}^o) \quad (18)$$

Los aspectos más característicos de las RNA son las estructuras de conexiones. El comportamiento de la red depende en gran medida del desempeño de la matriz o vector de pesos.

Figura 5.

Diseño de una neurona artificial



Nota. Tomado de: Hilera, J. R. & Martínez Hernando, M. (1995) *Redes neuronales artificiales : fundamentos, modelos y aplicaciones*
https://www.researchgate.net/publication/44343683_Redес_neuronales_artificiales_fundamentos_modelos_y_aplicaciones_Jose_Ramon_Hilera_Gonzalez_Victor_Jose_Martinez_Hernando

Funciones de activación

Las funciones de activación desempeñan un papel fundamental en el funcionamiento de las redes neuronales. Cada neurona en la red utiliza una función de activación para calcular su salida en función de la entrada que recibe.

Existen dos categorías principales de funciones de activación: lineales y no lineales. Las funciones lineales simplifican la salida multiplicando la entrada por un peso y sumando un término

de sesgo. Por otro lado, las funciones no lineales, más complejas, permiten que la red modele relaciones no lineales en los datos.

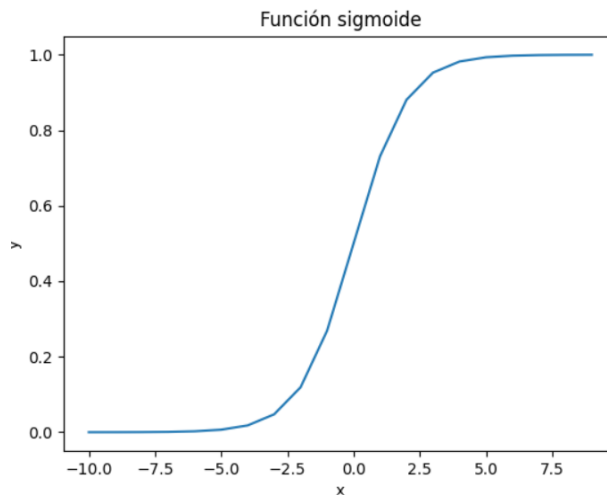
La elección de la función de activación repercute significativamente en la capacidad de la red para aprender y generalizar a nuevos datos. Algunas funciones de activación comunes incluyen la función sigmoide, la función ReLU y la función tangente hiperbólica. Es esencial seleccionar la función adecuada según la naturaleza del problema para optimizar el rendimiento y la capacidad de adaptación de la red neuronal (González y Hernando, 1995).

Función Sigmoide

Esta función que se expresa como $f(x) = \frac{1}{(1+e^{-x})}$ tiene una forma de “S” y su rango de salida está entre 0 y 1. Es utilizada normalmente a menudo en problemas de clasificación binaria, es decir, donde la salida debe ser 0 o 1.

Figura 6.

Función sigmoide



Nota. Tomado de: Huet, P. (2023) *Qué son las redes neuronales y sus aplicaciones*

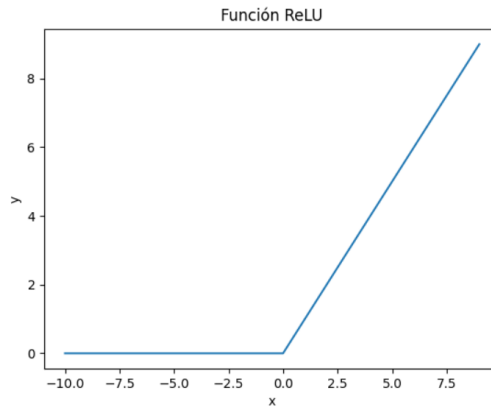
<https://openwebinars.net/blog/que-son-las-redes-neuronales-y-sus-aplicaciones/>

Función ReLU

La función de activación ReLU (Rectified Linear Unit) es la función más utilizada en las redes neuronales modernas. Su forma matemática es $f(x) = \max(0, x)$, lo que significa que la salida es 0 para valores negativos y lineal para valores positivos. La ventaja principal de esta función es que proporciona un valor cero verdadero para las entradas negativas, pero comportándose de manera similar a la lineal. Está siendo actualmente usada de manera habitual en el aprendizaje profundo, donde acelera el proceso de aprendizaje en los casos de aprendizaje representacional.

Figura 7.

Función Rectified Linear Unit



Nota. Tomado de: Huet, P. (2023) *Qué son las redes neuronales y sus aplicaciones* <https://openwebinars.net/blog/que-son-las-redes-neuronales-y-sus-aplicaciones/>

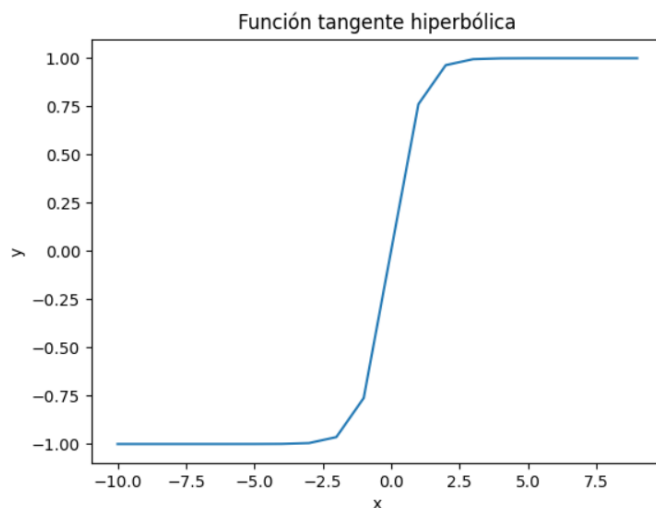
Función tangente hiperbólica

La función de activación tanh es similar a la función sigmoïdal, pero su rango de salida está entre -1 y 1. Su forma matemática es $f(x) = \frac{(e^x - e^{-x})}{(e^x + 1)}$. Esta función resulta adecuada en los

casos de problemas de clasificación binaria y también para problemas de regresión donde las salidas pueden ser negativas.

Figura 8.

Función tangente hiperbólica



Nota. Tomado de: Huet, P. (2023) *Qué son las redes neuronales y sus aplicaciones* <https://openwebinars.net/blog/que-son-las-redes-neuronales-y-sus-aplicaciones/>

4.3.8.1.2 Aprendizaje y entrenamiento. El aprendizaje en una red neuronal se logra mediante el ajuste de los pesos de las conexiones. Durante la fase de entrenamiento, la red se expone a un conjunto de datos de entrada junto con las salidas deseadas. A medida que la red compara sus salidas con las salidas deseadas, esta ajusta los pesos para minimizar la diferencia entre ellas utilizando algoritmos de optimización, como el descenso de gradiente.

La finalidad de esta fase de entrenamiento es que la red aprenda a generalizar a partir de ejemplos, de modo que pueda hacer predicciones precisas sobre datos que no ha visto previamente.

Los métodos de aprendizaje se clasifican en dos categorías:

Aprendizaje supervisado

Es un caso de entrenamiento conociendo el resultado final. Se presentan dos vectores (entrada y salida deseada). La red compara la salida que produce con la salida esperada, ajustando sus pesos para disminuir el error. Este proceso se repite de forma iterativa hasta que la disparidad entre la salida obtenida y la deseada alcanza un nivel aceptablemente bajo. Con la incorporación de n parejas de este tipo, se construye un conjunto de entrenamiento (Quesada et al., 1994).

El aprendizaje supervisado se suele dividir a su vez en dos subcategorías:

Aprendizaje Estructural: Se refiere a la búsqueda de la conexión óptima o afinidad posible entre la entrada y la salida para cada pareja de patrones individuales.

Aprendizaje Temporal: Este concepto se centra en la captura de patrones esenciales para alcanzar un resultado final. En el aprendizaje temporal, la respuesta actual de la red se ve influenciada por las entradas y respuestas previas. A diferencia del aprendizaje estructural, en este caso, existe una clara dependencia temporal en la dinámica de la red neuronal.

Aprendizaje no supervisado

Es un caso de entrenamiento sin conocer la salida esperada y solo usa información local durante todo el proceso de aprendizaje. Es un modelo más cercano al sistema biológico, ya que no utiliza el vector de salida esperada, y solo hay vectores de entrada en el conjunto de entrenamiento. El algoritmo modifica los pesos de forma que las salidas sean consistentes, es decir, que, a entradas muy parecidas, la red compute la misma salida. Las salidas se asocian a las entradas de acuerdo con el proceso de entrenamiento. El proceso extrae características, abstrayendo las propiedades colectivas subyacentes del conjunto de entrenamiento, y agrupa por clases de similitudes (Martín y Sanz, 2007).

4.3.8.1.3 Aplicaciones. Las redes neuronales se pueden aplicar a diferentes sectores como los siguientes:

Reconocimiento de Patrones

Las redes neuronales desempeñan un papel fundamental en la identificación de patrones en extensos conjuntos de datos. Por ejemplo, en el ámbito del reconocimiento de voz, estas redes son capaces de transformar ondas sonoras en texto, analizando los patrones del sonido para discernir las palabras pronunciadas y generar transcripciones precisas.

En el caso de la clasificación de imágenes, las redes neuronales demuestran su eficacia al reconocer características específicas, como formas y colores, permitiendo la clasificación en diversas categorías. Podrían, por ejemplo, analizar imágenes de animales para categorizarlas como "perros", "gatos", "aves", entre otras.

Procesamiento de Lenguaje Natural

Las redes neuronales se destacan en el procesamiento del lenguaje humano, destacando en la traducción automática y la generación de texto. En la traducción, estas redes aprenden a traducir idiomas al analizar patrones lingüísticos, mientras que en la generación de texto pueden crear contenido coherente y relevante, escribiendo historias o artículos atractivos para un público específico. Además, son utilizadas para analizar consultas en motores de búsqueda y proporcionar resultados pertinentes, mejorando la experiencia del usuario.

Visión por Computadora

Las redes neuronales son esenciales para la identificación de objetos en imágenes y videos, siendo valiosas en aplicaciones como la detección de objetos en tiempo real y el seguimiento de objetos. Pueden analizar imágenes en tiempo real para identificar objetos y proporcionar

información detallada. Por ejemplo, en la seguridad, estas redes pueden detectar objetos sospechosos o analizar imágenes de automóviles para ofrecer detalles sobre marcas y modelos.

Predicción y Toma de Decisiones

Gracias a su capacidad para reconocer patrones, las redes neuronales son herramientas poderosas en la predicción y toma de decisiones. En el comercio electrónico, analizan datos de clientes para prever preferencias y recomendar productos relevantes. Asimismo, en el análisis de riesgos, pueden predecir riesgos potenciales, evaluando datos de pacientes para anticipar la probabilidad de enfermedades como la diabetes o enfermedades cardíacas.

En conjunto, estas aplicaciones destacan la versatilidad y la capacidad transformadora de las redes neuronales en diversos sectores, desde el reconocimiento de patrones hasta la toma de decisiones avanzada (González y Hernando, 1995).

4.3 Análisis Multivariante (AM)

El análisis multivariante, según la definición proporcionada por Cuadras (2007), es una rama fundamental en la estadística y el análisis de datos que se dedica a estudiar, analizar, representar e interpretar conjuntos de datos resultantes de observar múltiples variables estadísticas en una muestra de individuos. En esta metodología, las variables observadas están interconectadas y presentan correlaciones, sin que ninguna de ellas prevalezca sobre las demás. La esencia del análisis multivariante radica en el enfoque multidimensional de la información estadística, donde conceptos geométricos, cálculos matriciales y distribuciones multivariantes desempeñan un papel crucial.

4.3.1 Objetivos del análisis multivariante

El análisis multivariante persigue dos objetivos principales que pueden resumirse de la siguiente manera:

Síntesis de Datos Multivariantes: Proporcionar métodos que permitan estudiar conjuntos de datos con múltiples variables en su totalidad, algo que el análisis estadístico unidimensional o bidimensional no logra alcanzar. Esta síntesis permite una comprensión profunda de la estructura de los datos y de las relaciones complejas entre las variables.

Apoyo a la Toma de Decisiones: Facilitar al analista o investigador la toma de decisiones óptimas en contextos específicos, utilizando la información derivada del análisis de datos multivariantes. Estos métodos ayudan a extraer información valiosa e identificar patrones ocultos que pueden influir en la toma de decisiones informadas.

4.3.2 Categorías de técnicas multivariantes

Las técnicas multivariantes se dividen en tres grupos principales, cada uno con enfoques y objetivos únicos:

Métodos de Dependencia: Estos métodos asumen que las variables se pueden dividir en dos grupos: variables dependientes y variables independientes. El propósito de estas técnicas es determinar si el conjunto de variables independientes tiene algún impacto en las variables dependientes y, en caso de ser afirmativo, expone cómo se manifiesta esa relación.

Métodos de Interdependencia: En contraste con los métodos de dependencia, estas técnicas no hacen una distinción entre variables dependientes e independientes. En su lugar, buscan

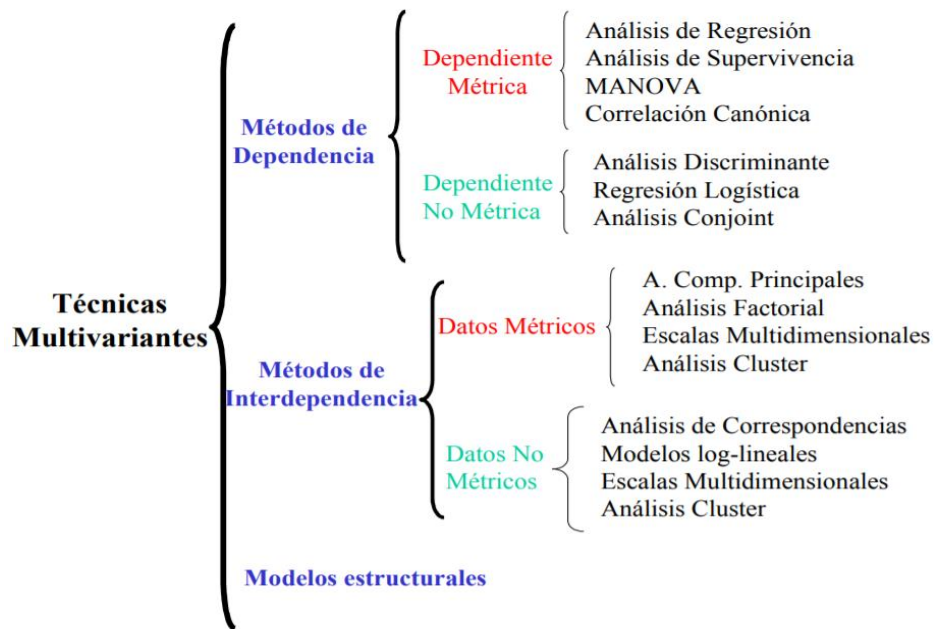
identificar las relaciones entre las variables y entender la naturaleza de esas conexiones, explorando cómo están relacionadas, por qué lo están y en qué medida.

Modelos Estructurales: Estos métodos también consideran la división entre variables dependientes e independientes. No obstante, su enfoque se centra en analizar no sólo cómo las variables independientes influyen en las dependientes, sino también en comprender cómo las variables de ambos grupos se entrelazan y se influyen mutuamente.

En conjunto, el análisis multivariante se posiciona como una herramienta poderosa para explorar la complejidad de los datos que involucran múltiples variables. A través de sus enfoques, busca descifrar relaciones, patrones y estructuras subyacentes en los datos, proporcionando información valiosa para la toma de decisiones informadas en diversos campos académicos y profesionales (Mateos-Aparicio y Hernández Estrada, 2021).

Figura 9.

Técnicas de análisis multivariantes



4.3.2.1 Regresión logística. La regresión logística se destaca como una herramienta estadística expresiva y versátil ampliamente utilizada en el análisis de datos en campos como la clínica y la epidemiología. Su origen se remonta a la década de los sesenta (Cornfield et al., 1961), siendo Walter y Duncan quienes la emplearon en su forma actual en 1967, para estimar la probabilidad de ocurrencia de un evento en función de variables específicas. Su popularidad y aplicación se universalizaron a partir de principios de los ochenta, principalmente impulsadas por los avances en las capacidades informáticas disponibles desde entonces.

La regresión logística es una herramienta estadística utilizada para investigar y entender la relación entre una variable dependiente cualitativa y una o más variables explicativas independientes, conocidas como covariables. Este enfoque se aplica en situaciones donde la variable dependiente es dicotómica (dos categorías) o tiene más de dos categorías. Dependiendo de la situación, se pueden utilizar la regresión logística binaria (para variables dependientes dicotómicas) o la regresión logística multinomial (para variables dependientes con más de dos categorías).

La regresión logística se basa en el concepto de probabilidades y se adapta especialmente bien a situaciones en las que la variable dependiente no es una variable continua, sino una variable cualitativa. Las covariables, que pueden ser cualitativas o cuantitativas, se utilizan para predecir la probabilidad de que la variable dependiente tome un valor específico.

Cuando se trabajan con covariables cualitativas dicotómicas, se utilizan variables dummy (variables ficticias) para representar cada categoría. Esto permite que cada categoría influya en el modelo de manera individual. La función logística es la función principal en la regresión logística y tiene la forma:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (19)$$

Donde:

- $P(Y = 1)$ es la probabilidad condicional de que la variable dependiente Y sea 1
- e es la base del logaritmo natural
- β_0 es la intersección de la regresión
- $\beta_1, \beta_2, \dots, \beta_n$ son los coeficientes de regresión asociados con las variables

independientes x_1, x_2, \dots, x_n respectivamente

La función logística transforma la combinación lineal de las variables independientes ponderadas por los coeficientes en una probabilidad que varía entre 0 y 1. La interpretación de los coeficientes β es importante, ya que indican cómo cambia la probabilidad logarítmica de éxito para un cambio unitario en la variable correspondiente, manteniendo constantes las demás variables.

4.3.2.1.1 Objetivos de la Regresión Logística. La regresión logística tiene tres objetivos principales:

Cuantificar la Importancia de las Relaciones: Permite medir cuánto influyen las covariables en la variable dependiente. Los coeficientes en el modelo logístico indican el cambio en el logaritmo de las probabilidades a medida que las covariables cambian.

Identificar Interacción y Confusión: Ayuda a detectar si hay interacciones o confusión entre las covariables en relación con la variable dependiente. Esto implica comprender cómo las variables explicativas pueden influir entre sí y cómo pueden confundir la relación con la variable dependiente.

Clasificación de Individuos: Además de analizar relaciones, la regresión logística también se utiliza para clasificar individuos en categorías específicas de la variable dependiente. Esta clasificación se basa en las probabilidades calculadas por el modelo.

En conclusión, la regresión logística es una técnica valiosa para analizar y predecir relaciones entre variables cualitativas y cuantitativas. Al desentrañar patrones y probabilidades, esta herramienta proporciona información valiosa para la toma de decisiones y la comprensión de fenómenos en diversas disciplinas (Ayçaguer y Utra, 2004).

5. Estado del arte

Para la realización de este proyecto, se llevó a cabo una revisión de alcance con el propósito de identificar y describir los modelos matemáticos que han sido empleados para predecir la deserción estudiantil en el ámbito de la educación superior. Este proceso tiene como objetivo principal mejorar la comprensión de cómo los modelos matemáticos se aplican en el contexto de la deserción universitaria.

La revisión de alcance se desarrolló siguiendo la metodología establecida por el Instituto Joanna Briggs (Tricco et al., 2018) y abarcó una amplia gama de estudios observacionales analíticos, que incluyeron investigaciones de cohortes tanto prospectivas como retrospectivas, así como estudios experimentales no encuadrados en el marco epidemiológico. Estos estudios involucraron a estudiantes de educación superior, independientemente de su programa académico y modalidad de estudio.

La estrategia de búsqueda está enfocada en identificar estudios en inglés o español, publicados y no publicados (literatura gris) y se realizó en tres pasos. Inicialmente se realizó una búsqueda en PubMed. Las palabras de texto contenidas en los títulos y resúmenes de los artículos relevantes identificados en la búsqueda inicial, así como los términos indexados usados para describir estos, fueron utilizados para realizar la búsqueda de literatura relevante en las bases de datos consideradas en esta revisión. La lista de referencias de los todos los estudios incluidos se evaluó para identificar estudios adicionales no identificados en las búsquedas.

Los estudios publicados y no publicados fueron buscados en las siguientes bases de datos: PubMed, Science direct, Proquest, ERIC, Open Dissertations y Google Scholar.

Continuando con la búsqueda, todas las citas que se identificaron fueron cargados en el software (Zotero) para eliminar los duplicados. Todas las citas fueron importadas dentro the JBI System for the Unified Management, Assessment and Review of Information (JBI SUMARI) (JBI, Adelaide, Australia) (Piper, 2019). Para esta revisión se evaluaron los títulos y resúmenes de 100 estudios escogidos de manera aleatoria acorde a los criterios de inclusión propuestos. Se recuperaron los textos completos de los estudios potenciales para la revisión y se realizó una evaluación en detalle de estos teniendo en cuenta los criterios de inclusión. Todo el proceso de selección fue realizado por dos revisores independientes, en caso de encontrar desacuerdo fueron resueltos por discusión o con un revisor adicional. Los resultados de la búsqueda y del proceso de inclusión fueron presentados en el diagrama de flujo de “PRISMA-ScR” (Tricco et al., 2018).

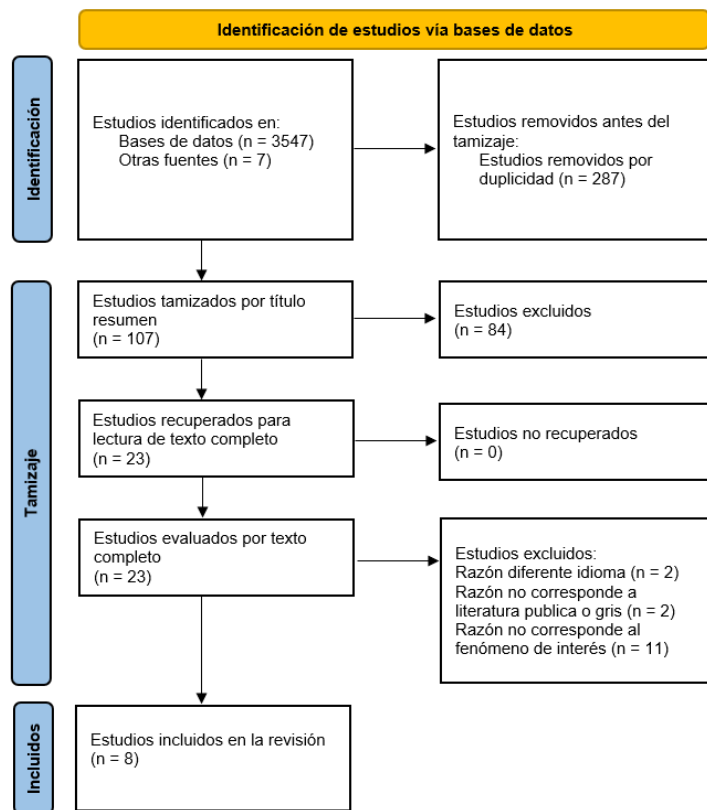
La información de los estudios encontrados en la revisión de alcance fue extraída por un revisor, un segundo revisor verificó esta. Para la extracción de los datos se utilizó una planilla en Excel (Anexo A), la información a extraer incluye detalles acerca de participantes, concepto,

contexto, tipo de estudio, métodos, variables explicativas, forma de medición, precisión de los modelos y principales hallazgos.

Se recuperaron un total de 3554 estudios, de estos, 3547 fueron identificados en las bases de datos y, 7 estudios, fueron encontrados mediante referencias cruzadas. 287 estudios fueron removidos por duplicidad. A partir de la muestra aleatoria de 100 estudios, se recuperaron 23 estudios para la lectura de texto completo. Finalmente, se incluyeron 8 estudios para la extracción de datos (Figura 2).

Figura 10.

Flujograma de estudios incluidos



Esta revisión de alcance sienta las bases para un análisis detallado y completo de los modelos matemáticos aplicados en la predicción de la deserción estudiantil en la educación superior. Los modelos matemáticos identificados para predecir la deserción estudiantil fueron: árboles de decisión, modelos probabilísticos de redes bayesianas, regresión logística, Adda-Boost, Naive Bayes, bosques aleatorios, máquinas de soporte vectorial, cluster ensembles y k vecinos más próximos (Anexo A).

Se identificaron 44 diferentes tipos de variables que se usaron para predecir la deserción estudiantil mediante los modelos matemáticos, estas variables fueron clasificadas en 4 categorías: Individual (16), Académica (15), Institucional (5) y Socioeconómica (8) (Anexo B). De todos los estudios incluidos sólo dos estudios realizaron la denominación de las variables.

De los ocho estudios incluidos, tres reportaron las variables explicativas que tienen un mayor peso/importancia en los modelos matemáticos para predecir la deserción estudiantil. Estas fueron: Adicción a las redes sociales, dependencia emocional, estado marital, la edad, GPA ponderado, GPA acumulativo, tiempo directo de estudio, puntuación SAT y fecha de nacimiento.

Finalmente, es relevante mencionar que, entre los modelos matemáticos identificados, aquellos que han logrado una precisión superior al 90% incluyen la regresión logística, los árboles de decisión, los bosques aleatorios y los cluster ensembles. Los resultados de dichos modelos matemáticos proporcionan una comprensión más profunda del fenómeno de la deserción estudiantil.

6. Objetivos

6.1 Objetivo general

Establecer la asociación entre los problemas de salud mental al ingreso a la Universidad y la deserción estudiantil temprana en los estudiantes de pregrado presencial en cualquiera de las sedes de la Universidad Industrial de Santander (UIS).

6.2 Objetivos Específicos

Estimar la incidencia de deserción temprana en los estudiantes de la UIS.

Establecer la prevalencia de los problemas de salud mental al momento del ingreso de los estudiantes de la UIS.

Identificar la prevalencia de otros determinantes al momento del ingreso que afectan la permanencia del estudiante en su programa académico.

Calcular la asociación independiente entre los problemas de salud mental y la incidencia de deserción temprana en los estudiantes de la UIS.

Diseñar un modelo predictivo de deserción temprana en estudiantes universitarios.

7. Metodología

7.1 Diseño del estudio

Se realizó un estudio observacional analítico de tipo cohorte retrospectivo, para la conformación de la muestra se incluyó la información recolectada de los estudiantes admitidos en la Universidad Industrial de Santander en el semestre 2015-2 hasta 2017-2, para seguirlos durante los primeros 4 semestres hasta la ocurrencia del evento de interés: quedar por fuera de la universidad por bajo rendimiento académico, cancelación definitiva de matrícula, haber quedado excluido por dos o más semestres cancelados de manera consecutiva, retiro definitivo mayor a tres períodos académicos, y cancelación definitiva de matrícula.

7.2 Población

Se obtuvieron los datos de 6605 estudiantes a través de la Dirección de Bienestar Estudiantil siendo estos admitidos en la universidad desde el segundo periodo académico de 2015 hasta el segundo periodo académico de 2017, considerando que se requieren dos años de seguimiento teniendo en cuenta la definición de deserción temprana.

7.3 Criterios de Inclusión

Estar matriculado en los semestres 2015-2 a 2017-2 en algún programa de pregrado presencial de las sedes de la UIS.

7.4 Criterios de Exclusión

No se incluyeron estudiantes admitidos después del segundo semestre de 2017 considerando que se requería un seguimiento de cuatro semestres, en el primer periodo académico de 2020 por motivo de la pandemia por Covid-19 se modificaron los escenarios de aprendizaje y el modelo pedagógico a las declaradas en los planes de estudios de los programas académicos.

7.5 Muestra

Se calculó un tamaño de 854 con una confiabilidad del 95%, un poder del 80%, y una razón de no expuesto a expuesto de 3:1, considerando que la prevalencia estimada de una pobre salud mental en estudiantes universitarios es del 24%. De este grupo, el 32% desertaron, y, por el contrario, aquellos con una buena salud mental presentaron una deserción del 22% (Hjorth et al., 2016).

Este tamaño de muestra se calculó de la siguiente forma:

Kelsey

$$n_1 = \frac{(Z_{\alpha} + Z_{\beta})^2 \times (pq) \times (r - 1)}{r(p_1 - p_2)^2} \approx 194 \quad (20)$$

$$n_2 = r \cdot n_1 \approx 581 \quad (21)$$

$$Total = 775 \quad (22)$$

Fleiss

$$n_1 = \frac{\left[\frac{Z_\alpha}{2} \sqrt{(r+1)\underline{p}\underline{q}} + Z_\beta \sqrt{rp_1q_1 + p_2q_2} \right]^2}{r(p_1 - p_2)^2} \approx 201 \quad (23)$$

$$n_2 = rn_1 \approx 601 \quad (24)$$

$$Total = 802 \quad (25)$$

$$\underline{p} = \frac{p_1 + rp_2}{r+1} \quad (26)$$

$$\underline{q} = 1 - \underline{p} \quad (27)$$

Fleiss con CC

$$n_1 = \frac{n_1}{4} \left[1 + \sqrt{1 + \frac{2(r+1)}{n_1 r (|p_2 - p_1|)}} \right] \approx 214 \quad (28)$$

$$n_2 = rn_1 \approx 640 \quad (29)$$

$$Total = 854 \quad (30)$$

Donde:

- n_1 es el número de expuestos
- n_2 es el número de no expuestos
- $Z_{\frac{\alpha}{2}}$ es el valor crítico de la distribución normal estándar asociado al nivel de confianza
- Z_β es el valor crítico de la distribución normal estándar asociado al poder deseado
- p_1 es la proporción estimada de la característica de interés en la población expuesta
- p_2 es la proporción estimada de la característica de interés en la población no expuesta
- r es Razón de no expuesto a expuesto

El cálculo de los valores críticos para el nivel de confianza y poder deseado se calculan de la forma:

$$Z_{\frac{\alpha}{2}} = \Phi^{-1}(0,95) \approx 1,96 \quad (31)$$

$$Z_{\beta} = \Phi^{-1}(1 - 0,8) \approx 0,84 \quad (32)$$

- Φ^{-1} es la función inversa de la función de distribución acumulativa normal estándar
- P es la probabilidad acumulativa asociada al valor crítico.
- β es el nivel de significancia, y $1-\beta$ es el poder estadístico

Finalmente se construyó la base de datos con la información obtenida de 4273 estudiantes que tenían la información completa de las encuestas aplicadas por la Dirección de Bienestar Estudiantil en la evaluación multifactorial que realizan al ingreso durante la semana de inducción a la vida universitaria. Dándonos un mayor poder estadístico para detectar diferencias o asociaciones entre las variables predictoras (Kelsey, 1996).

7.6 Instrumentos

Debido a que la información obtenida fue por medio de la base de datos que genera la dirección de bienestar estudiantil al momento del ingreso de los estudiantes, no se tienen las encuestas o test utilizados sino solo los resultados de estos. Se tiene conocimiento que todas las encuestas o test aplicados están validados para la recolección de información.

7.7 Aspectos Éticos

La investigación que se llevó a cabo se rige estrictamente por los sólidos principios éticos que han sido establecidos en el informe de Belmont. Estos principios fundamentales incluyen el respeto por la dignidad y los derechos de todas las personas involucradas en la investigación, la búsqueda del beneficio, evitando causar daño (principio de no maleficencia), y asegurando que la distribución de los beneficios y las cargas de la investigación sea justa.

Asimismo, tomamos en consideración los estándares éticos específicos para la investigación médica en seres humanos que se encuentran descritos en la Declaración de Helsinki, en su versión actualizada de 2013. Esto garantiza que nuestra investigación cumpla con las pautas internacionales más rigurosas para la ética en la investigación científica.

En cumplimiento con la Resolución 008430 del 4 de octubre de 1993, la cual establece las normas científicas, técnicas y administrativas para la investigación en salud en Colombia, esta investigación se clasifica como "sin riesgo" según el artículo 11, numeral b del primer capítulo. Esto significa que no se someterá a los participantes a ningún tipo de riesgo y que toda la información necesaria se obtuvo de manera ética y legal desde la base de datos de la Dirección de Bienestar Estudiantil de la Universidad Industrial de Santander. No se llevó a cabo ningún tipo de contacto directo con los participantes, garantizando así su privacidad y confidencialidad en todo momento.

La utilización de bases de datos para almacenar la información constituye una medida eficaz para minimizar el riesgo de divulgación de datos sensibles de los participantes. Esta precaución adicional asegura que la privacidad y confidencialidad de la información se mantengan en todo momento.

Cabe destacar que la realización de este proyecto estuvo sujeta a la aprobación y aval del Comité de Ética en Investigaciones Científicas de la Universidad Industrial de Santander (UIS) el día 7 de mayo del 2021 en el acta N° 08. Esta instancia garantiza que todas las fases y procedimientos de la investigación se ajusten a los más altos estándares éticos y legales, reforzando así la integridad y validez de los resultados obtenidos.

7.8 Recolección y Preparación de Datos

A partir de la revisión de alcance realizada, conseguimos obtener un entendimiento más profundo acerca de las variables que ejercen una mayor influencia en el fenómeno de la deserción estudiantil a nivel universitario. Esto nos proporcionó una base sólida para llevar a cabo un análisis detallado de la información disponible a través de la dirección de bienestar estudiantil, permitiéndonos seleccionar con precisión las variables que serían implementadas en nuestros modelos matemáticos.

Posteriormente, nos sumergimos en un proceso de refinamiento de la base de datos, dado que la información inicial se encontraba dispersa. Esta labor fue esencial para poder agrupar y estructurar la información de manera coherente y legible para su posterior análisis con los modelos matemáticos. La estructuración de la información se llevó a cabo por medio de etiquetas como se muestra en la (tabla 2), por último, se realizó la agrupación de esta información en un solo archivo debido a que las variables se encontraban en múltiples ficheros.

Este paso estratégico no solo facilitó la interpretación de los resultados, sino que también optimizó la eficiencia y precisión de nuestro enfoque analítico.

7.9 Definición de Variables y Objetivo

Variable dependiente

Deserción temprana: El estudiante que, siendo admitido, abandona la universidad en los 4 primeros semestres.

Tabla 2.

Variables

Variable	Definición	Naturaleza	Escala	Valor
Variables sociodemográficas				
Género	Condición orgánica que distingue ser mujer o ser hombre	Cualitativa Dicotómica	Nominal	Masculino o femenino
Edad	Tiempo transcurrido desde el nacimiento del individuo	Cualitativa Politómica	Nominal	Cualquier número entero
Variable socioeconómica				
Porcentaje de descuento	Cuanto porcentaje de descuento tiene en la matrícula	Cualitativa Politómica	Nominal	0 - 100%
Auxiliaturas	Tiene auxiliaturas con la universidad	Cualitativa Dicotómica	Nominal	Si o No
Servicio de comedores	Cuántos servicios de comedores está recibiendo	Cualitativa Politómica	Nominal	0, 1, 2 o 3
Municipio de procedencia	Se encuentra en el área metropolitana de Bucaramanga	Cualitativa Dicotómica	Nominal	Si o No
Riesgo socioeconómico	Tiene riesgo socioeconómico según el test	Cualitativa Dicotómica	Nominal	Si o No
Variables explicatorias				
Depresión	Grado de depresión con el test de Zung	Cualitativa Politómica	Ordinal	Ausencia de depresión

Variable	Definición	Naturaleza	Escala	Valor
				Depresión leve Depresión moderada Depresión grave
Test de Cage	Nivel de alcoholismo	Cualitativa Politómica	Nominal	0, 1, 2, 3 o 4
Satisfacción y vida	Se encuentra satisfecho con la vida	Cualitativa Dicotómica	Nominal	Si o No
Competencia lectora	Tiene riesgo en la comprensión lectora	Cualitativa Dicotómica	Nominal	Si o No
Ansiedad	Grado de ansiedad con el test de Zung	Cualitativa Politómica	Ordinal	Ausencia de ansiedad, Ansiedad leve, Ansiedad moderada, Ansiedad grave
Funcionalidad familiar	Grado de funcionalidad familiar con el test APGAR	Cualitativa Politómica	Ordinal	Normal Disfunción leve Disfunción moderada Disfunción severa
Admitidos por cohorte	Cantidad de admitidos en la carrera	Cuantitativa Politómica	Nominal	Cualquier número entero
Puesto de ingreso	Puesto con el que se ingresa a la universidad	Cuantitativa Politómica	Nominal	Cualquier número entero
Colegio	Tipo de colegio del cual proviene	Cualitativa Dicotómica	Nominal	Público o privado
Programa académico	Programa al que se matricula	Cualitativa Politómica	Ordinal	Programa académico que están cursando
Programa opcional	Si el programa que está estudiando fue su primera o segunda opción	Cualitativa Dicotómica	Nominal	Primera o segunda opción
Capacidad de consulta	Malos hábitos al momento de consultar	Cualitativa Dicotómica	Nominal	Si o No

Variable	Definición	Naturaleza	Escala	Valor
Automotivación para el aprendizaje	Malos hábitos de motivación en el aprendizaje	Cualitativa Dicotómica	Nominal	Si o No
Autoconfianza en el aprendizaje	Tiene autoconfianza en su forma de aprender	Cualitativa Dicotómica	Nominal	Si o No
Estrategias de aprendizaje	Tiene estrategias de aprendizaje	Cualitativa Dicotómica	Nominal	Si o No
Cuidado de sí mismo	Malos hábitos de cuidado personal	Cualitativa Dicotómica	Nominal	Si o No
Relaciones interpersonales	Malos hábitos de comunicación	Cualitativa Dicotómica	Nominal	Si o No
Eficiencia personal	Malos hábitos de rendimiento	Cualitativa Dicotómica	Nominal	Si o No
Hábitos saludables del sueño	Malos hábitos del sueño	Cualitativa Dicotómica	Nominal	Si o No
Estado académico	Estado académico en el que se encuentra el estudiante	Cualitativa Politómica	Ordinal	Condicional Normal PFU

7.10 Limitaciones Potenciales

- Debido a que el estudio es de cohorte retrospectiva, se basa en la disponibilidad y fidelidad de la información que se encuentra en la base de datos otorgada por la Dirección de Bienestar Estudiantil.
- Al no poder controlar el registro exacto de los datos se puede introducir un sesgo de información, dado que los estudiantes pueden no tener el compromiso de responder las encuestas auto aplicadas con la responsabilidad que se requiere, como también pueda ser que no entiendan las preguntas respecto a las encuestas, o que los profesionales de la Dirección de Bienestar Estudiantil no registren la información adecuadamente.

- La presencia de los problemas de salud mental se determina por escalas de tamizaje, pero estas no se confrontan con el patrón de oro que corresponde a la entrevista por psiquiatría. Sin embargo, las escalas que han sido seleccionadas por BE fueron adaptadas y validadas en nuestro país, y son útiles como instrumentos de tamizaje.

- Realizando el estudio para determinar la asociación entre los problemas de salud mental y la deserción, no se encontró una asociación estadísticamente significativa. Lo anterior, debido a que, en el análisis con la base de datos suministrada por la Dirección de Bienestar Estudiantil, no se cuenta con una muestra lo suficientemente alta para llegar a un IC del 95% y un poder del 80%. Puesto que necesitaríamos una muestra por encima de 4900 estudiantes y solo se cuenta con 4273, esto nos deja con un poder del 71%.

7.11 Red neuronal

7.11.1 División del conjunto de datos

Posterior al refinamiento de la base de datos, se realizó una división de esta en tres conjuntos: entrenamiento (73%), validación (15%) y prueba (12%). Siguiendo las recomendaciones de la documentación de Keras y TensorFlow, donde se establecen los porcentajes de la información para entrenamiento (60-80%), validación (10-20%) y prueba (10-20%). Esta división de la base de datos se realizó haciendo uso del método “train_test_split()” que permite dividir la información de un conjunto de datos de manera aleatoria.

Los conjuntos de entrenamiento están compuestos por 26 variables independientes y la variable dependiente (deserción).

- El conjunto de datos de entrenamiento es utilizado para enseñar al modelo a realizar predicción o clasificaciones basadas en patrones y relaciones que se encuentran en la base de datos, asignando de esta manera el peso, o el grado de influencia de las variables en el modelo.
- El conjunto de datos de validación fue sometido a múltiples iteraciones, las cuales permitieron evaluar los hiperparámetros propuestos (número de capas), y así ajustarlos con el fin de lograr un mayor rendimiento en el modelo.
- El conjunto de datos de prueba fue esencial para comprobar que el modelo se desempeña de manera eficiente utilizando datos completamente nuevos, certificando así la red neuronal con una precisión por encima del 80%.

Adicionalmente, la lectura de estos conjuntos de datos se realizó por medio de la librería “pandas” en python que facilita la lectura para la RNA.

7.11.2 Diseño de la arquitectura de la red neuronal

A partir de la documentación de Keras y TensorFlow, se recopilaron las mejores prácticas en el diseño de arquitecturas para redes neuronales predictoras. De lo anterior se resaltan los siguientes parámetros clave:

Tabla 3.

Arquitectura de la red neuronal artificial

Red neuronal secuencial	Una red neuronal secuencial es un tipo de arquitectura de red neuronal en la que las neuronas están organizadas en capas consecutivas y la información fluye en una sola dirección, de la entrada a la salida.
-------------------------	--

Función de activación ReLu	La función de activación ReLU en una red neuronal retorna cero para valores negativos y el mismo valor para valores positivos. Ayuda a introducir no linealidad y es eficiente en el cálculo.
Función de activación Sigmoid	La función de activación Sigmoide transforma cualquier valor real en un rango entre 0 y 1, es comúnmente usada en capas de salida para problemas de clasificación binaria.
Parámetro Dropout	Ayuda a prevenir el sobreajuste durante el entrenamiento al "apagar" aleatoriamente un porcentaje de neuronas en cada iteración

```
#Arquitectura de la red neuronal
model = Sequential()
model.add(Dense(units=128, input_dim=variables, activation='relu')) #Capa de entrada
model.add(Dropout(0.3))
model.add(Dense(units=128, activation='relu')) #Capas ocultas
model.add(Dropout(0.3))
model.add(Dense(units=64, activation='relu')) #Capas ocultas
model.add(Dense(1, activation='sigmoid')) #Capa de salida
```

7.11.3 Definición de función de perdida y sobreajuste

Debido a que este modelo de datos está planteado para un problema de clasificación binaria, se utiliza la función de pérdida “BinaryCrossentropy”, ya esta función de pérdida mide la discrepancia entre las distribuciones de probabilidad reales y predichas de un modelo de clasificación binaria.

La fórmula matemática para la “BinaryCrossentropy” es:

$$BC = -\frac{1}{N} \sum_{i=1}^N [y_i \cdot \log(\hat{y}_i) + (1 - y_i) \cdot \log(1 - \hat{y}_i)] \tag{33}$$

Donde:

- N es el número total de muestras en el conjunto de datos

- y_i es la etiqueta real de la i -ésima muestra (0 ó 1)
- \hat{y}_i es la predicción del modelo para la i -ésima muestra, una probabilidad entre 0 y 1

La “BinaryCrossentropy” penaliza fuertemente las predicciones incorrectas. Si la etiqueta real (y_i) es 1, la pérdida se penaliza más si la predicción del modelo (\hat{y}_i) está más cerca de 0 que de 1. Similarmente, si la etiqueta real es 0, la pérdida se penaliza más si la predicción está más cerca de 1 que de 0. Esta penalización afecta la actualización de los pesos durante el entrenamiento.

7.11.4 Optimización y entrenamiento del modelo

Con el objetivo de minimizar la función de pérdida del modelo se implementó el algoritmo de optimización “Adam”. Esto, dado que combina técnicas de otros optimizadores que permite ajustar la tasa de aprendizaje de forma adaptativa para cada parámetro. Los componentes claves de este algoritmo de optimización son:

- **Momentos (Momentum):** Incorpora una especie de inercia al ajuste de los pesos. Ayuda a suavizar el proceso de actualización, permitiendo que los pesos se actualicen no solo en función del gradiente actual, sino también teniendo en cuenta el gradiente acumulado de pasos anteriores.

- **RMSprop (Root Mean Square Propagation):** Normaliza los gradientes dividiéndolos por la raíz cuadrada de una media móvil de los cuadrados de los gradientes anteriores. Esto ayuda a adaptar las tasas de aprendizaje de las diferentes características para mejorar la convergencia.

- **Corrección de sesgo (Bias Correction):** Mitiga el problema del sesgo en los primeros pasos de entrenamiento. Al inicio, los momentos y la media cuadrática pueden estar sesgados hacia

cero. Adam utiliza una corrección de sesgo para compensar este efecto y garantizar que los momentos y la media cuadrática sean representativos.

Los gradientes indican la dirección y la magnitud en la que la función de pérdida cambia más rápidamente con respecto a cada peso individual.

La actualización de los pesos en Adam se realiza mediante la fórmula general:

$$\theta_{t+1} = \theta_t - \frac{n}{\sqrt{\hat{v}_t + \epsilon}} \cdot \hat{m}_t \tag{34}$$

Donde:

- θ_{t+1} son los pesos actualizados
- θ_t son los pesos actuales
- n es la tasa de aprendizaje
- \hat{m}_t es la estimación del momento
- \hat{v}_t es la estimación de la media cuadrática
- ϵ es un término de suavizado para evitar la división por cero.

Lo cual lo hace efectivo en una variedad de problemas de optimización. Adicionalmente usamos el parámetro “accuracy” para que el modelo tenga la proporción de predicciones correctas al momento de realizar el entrenamiento.

La fórmula que utiliza para calcular la precisión es:

$$Precisión = \frac{\text{Número de precisiones correctas}}{\text{Número de precisiones incorrectas}} \tag{35}$$

Finalmente se realizó el entrenamiento de la red neuronal bajo los parámetros definidos, alimentando el modelo con el conjunto de datos de entrenamiento.

```
#ajustes de la red antes del entrenamiento
model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

7.11.5 Validación y ajuste de hiperparámetros

Utilizando el conjunto de datos de validación, se realizaron múltiples iteraciones para evaluar los hiperparámetros propuestos en el modelo inicial. De esta manera, se ajustaron el número de capas de la red y se obtuvo un mayor rendimiento en el tiempo de aprendizaje del modelo. Cabe resaltar que el proceso de entrenamiento y validación es realizado simultáneamente de manera que se pudo analizar la precisión de la red neuronal.

```
#entrenamiento red neuronal
model.fit(training_data, y_train, epochs=50, batch_size=64, validation_data=(vali_data, y_vali))
```

El parámetro “epochs” se dejó establecido en 50 debido a que no había una mejora significativa en el proceso de entrenamiento de la RNA para un número de ciclos superiores, así mismo el “batch_size” que se refiere al número de muestras de datos que se utilizan en una iteración durante el proceso de entrenamiento del modelo. Durante el entrenamiento, los datos se dividen en lotes más pequeños (también llamados "batches"), y el modelo se actualiza con cada lote, en lugar de actualizar los pesos del modelo después de cada muestra individual. Este valor se dejó establecido en 64 después de múltiples pruebas ajustando este parámetro, donde este valor nos genera mejores resultados en el modelo final.

7.12 Regresión logística

Para llevar a cabo la regresión logística, se inició con un análisis descriptivo para conocer el comportamiento de cada una de las variables del estudio. Para ello, se calcularon las medidas de tendencia central y sus correspondientes medidas de dispersión en las variables cualitativas de

interés, teniendo en cuenta la distribución de estas. Así mismo, para las variables cuantitativas se calcularon frecuencias relativas y sus respectivos intervalos de confianza (IC) del 95%.

Seguido de esto, se hizo un análisis bivariado para evaluar la asociación cruda (no ajustada) entre las potenciales variables explicativas y la presencia del desenlace desfavorable, deserción, al final de este. Y, a partir de lo anterior, se determinó la razón relativa (OR) para medir el efecto de cada factor de riesgo sobre la deserción. Finalmente, se realizó un análisis multivariado para estimar la asociación ajustada entre las variables explicativas y la deserción. Las variables que se incluyeron en la construcción del modelo se seleccionaron teniendo en cuenta su significancia estadística en el análisis bivariado y su rol en la explicación del desenlace, en el modelo final se incluyeron aquellas con una $p < 0,05$ y aquellas de importancia para explicar la asociación de estudio, de acuerdo con lo reportado en estudios previos. Este análisis se hizo con el software Stata v14.0.

8. Resultados

8.1 Descripción de la población

La población observada para este estudio es de 4273 de los cuales el 34,3% tienen un desenlace desfavorable en sus estudios académicos, posteriormente se hace una descripción de las variables asociadas a la deserción en este estudio. La descripción detallada por género de las variables asociadas a la deserción se encuentra en el anexo E.

Las variables observadas en la población, asociadas a la deserción, se han categorizado en cuatro grupos como: Individuales, académicas, institucionales y socioeconómicas.

El grupo de variables individuales son aquellas que hacen alusión a las características intrínsecas del individuo a nivel físico y psicológico. En este orden de ideas, la población se compone de un 40,68 % para el sexo femenino y un 59,32% masculino, así mismo, el promedio de edad de la población estudiada es de 18,1 años, y el 74,24% de los estudiantes residen en el área metropolitana de Bucaramanga.

En relación con el autocuidado, alrededor del 54,84% de los estudiantes considera tener hábitos saludables de sueño; el 39,29% tiene un buen cuidado de sí mismo; el 54,63% tiene buenas relaciones interpersonales y el 66,20% de la población posee una buena eficiencia personal. Además, se observa que solo el 9,95% de los estudiantes no se encuentran satisfechos con su vida en general.

Continuando con las variables correspondientes a la salud mental de los estudiantes, usando el test de Zung para evaluar la probabilidad de depresión y ansiedad en la población, se encontró que el 32,59% están en riesgo de depresión y el 40,78% están en riesgo de ansiedad. Por otra parte, el test de Cage nos permitió observar que solo el 5,45% de los estudiantes tiene alto riesgo de presentar un trastorno por uso de alcohol. Asimismo, usando la herramienta del apgar familiar se encontró que el 20,37% de los estudiantes perciben una disfuncionalidad en su núcleo familiar.

Por otra parte, el grupo de variables académicas son aquellas relacionadas a preferencias profesionales y habilidades académicas. Con base en lo anterior, el 44,38% de los estudiantes no considera tener automotivación para el aprendizaje y el 63,55% no considera tener autoconfianza de aprendizaje. Siguiendo esta línea de análisis, de la población estudiada el 47,66% de los estudiantes desarrollaron una buena competencia lectora, y el 55,36% una excelente capacidad de

consulta, no obstante, como contraparte, es preocupante que el 47,8% de los estudiantes carecen de estrategias de aprendizaje.

También, se analizó el ingreso a la universidad teniendo en cuenta si fue su primera o segunda opción de carrera profesional al momento de la inscripción, el 44,45% de los estudiantes ingresaron a su segunda opción de programa académico y con respecto a la procedencia de instituciones educativas de bachillerato se observó que el 32,94% son de instituciones privadas. Adicionalmente, la descripción de las variables programa académico, puesto de ingreso y admitidos por cohorte se encuentran en los anexos C y D

En cuanto a las variables categorizadas como institucionales que son aquellas relacionadas con los beneficios que presenta la universidad a los estudiantes. De lo antes planteado, se observó que las herramientas de ayuda económica para los estudiantes como las auxilias, el servicio de comedores y el porcentaje de descuento en la matrícula universitaria, solo tuvieron una cobertura del 5,38%, 17,6% y 4,96%, respectivamente.

Finalmente, para el grupo de variables socioeconómicas que se evaluaron por medio de un solo indicador que engloba esta categoría, se encontró que solo el 3,68% de los estudiantes tiene un alto riesgo socioeconómico.

8.2 Análisis bivariado

Se realizó un análisis bivariado para observar la significancia de las variables asociadas a la deserción. En la siguiente tabla se muestran los resultados del análisis bivariado (resultados crudos) para todas las variables analizadas.

Tabla 4.

Variables asociadas a la deserción

Variable	OR crudo	IC95%	Valor de p
Género	1,4400	1,31 - 1,57	0,000
Edad	0,9986	0,97 - 1,01	0,895
Porcentaje de descuento	0,9637	0,93 - 0,99	0,008
Auxiliaturas	0,9170	0,66 - 1,00	0,060
Servicio de comedores	1,0067	0,95 - 1,06	0,819
Municipio de procedencia	0,6366	0,58 - 0,69	0,000
Riesgo socioeconómico	1,199	0,98 - 1,45	0,066
Depresión	1,0523	1,02 - 1,08	0,001
Test de Cage	1,1094	0,93 - 1,31	0,232
Satisfacción y vida	1,1492	1,08 - 1,21	0,000
Competencia lectora	0,9914	0,94 - 1,04	0,751
Ansiedad	1,0450	1,01 - 1,07	0,005
Funcionalidad familiar	1,2809	1,16 - 1,40	0,000
Admitidos por cohorte	1,3736	1,14 - 1,64	0,001
Puesto de ingreso	1,0031	1,001 - 1,004	0,000
Colegio	1,0097	0,92 - 1,10	0,828
Programa académico	1,0326	1,01 - 1,05	0,001
Programa opcional	1,0880	1,00 - 1,18	0,050
Capacidad de consulta	1,0121	0,96 - 1,05	0,593
Automotivación para el aprendizaje	1,1075	1,07 - 1,13	0,000
Autoconfianza en el aprendizaje	0,9430	0,88 - 1,00	0,086
Estrategias de aprendizaje	1,0834	1,05 - 1,11	0,000
Cuidado de sí mismo	1,0713	1,03 - 1,10	0,000
Relaciones interpersonales	1,0779	1,04 - 1,11	0,000
Eficiencia personal	1,0953	1,06 - 1,12	0,000
Hábitos saludables del sueño	1,0620	1,02 - 1,09	0,000

A partir del análisis bivariado se puede observar que las variables tener servicio de comedores, admitidos por cohorte, puesto de ingreso, programa opcional, porcentaje de descuento, municipio de procedencia, género, funcionalidad familiar, ansiedad, depresión, eficiencia personal, automotivación para el aprendizaje, autoconfianza en el aprendizaje, hábitos saludables del sueño, cuidado de sí mismo, relaciones interpersonales, satisfacción y vida, capacidad de

consulta, riesgo socioeconómico y el programa académico tienen una asociación estadísticamente significativa con la deserción universitaria temprana, dado que tienen un $P < 0,05$.

8.3 Análisis Multivariado

Con base en los resultados del análisis bivariado, se seleccionaron las variables con una significancia estadística $\leq 0,25$, además de incluir la variable de interés test de Cage citada en la revisión bibliográfica.

Posteriormente se fueron retirando las variables con una $P > 0,05$, de manera secuencial, para evaluar el efecto de confusión sobre las otras variables. Las variables que no confunden el efecto de las demás, se retiraron definitivamente del modelo.

Las variables admitidas por cohorte, capacidad de consulta, porcentaje de descuento, cuidado de sí mismo, autoconfianza en el aprendizaje y hábitos saludables del sueño no eran significativas ni confusoras sobre las otras variables, por lo cual se retiraron del modelo. La variable riesgo socioeconómico tampoco fue significativa por lo cual se decidió retirarla.

Adicionalmente, se definieron algunas variables de la siguiente manera: el servicio de comedores se evaluó según la cantidad de comidas diarias que recibe cada estudiante; eficiencia personal, automotivación para el aprendizaje, satisfacción y vida, se midieron en función de la presencia o ausencia de dicho riesgo en cada estudiante. Por último, la variable programa académico, se evaluó de acuerdo con 32 valores que representan cada carrera universitaria dentro de la base de datos.

Esta definición de las variables se hizo debido a que se mejoraba la significancia de estas.

$$\begin{aligned}
 P \text{ Deserción} = & -4,71 + 1,38 * \text{foráneo} + 0,95 * \text{sexo} + 0,26 * \text{funfam} \\
 & + 0,36 * \text{eficiencia} + 0,43 * \text{automotive} + 1,06 \\
 & * \text{satisfacción} + 0,39 * \text{opcarrera} + 0,01 * \text{ingreso} - 0,59 \\
 & * \text{comedores}
 \end{aligned}
 \tag{36}$$

Debido a que no se encontraron interacciones entre las variables estudiadas en la revisión bibliográfica realizada, se probaron diversas interacciones que se observaron en el análisis bivariado, pero ninguna de estas resultó estadísticamente significativa al agregarlas al modelo preliminar por lo cual se mantuvo como modelo final.

En la construcción del modelo multivariado, se usó como referencia la carrera de medicina, ya que en esta fue donde se presentó el menor riesgo de deserción y presentaba un buen tamaño de muestra.

Tabla 5.

Variables del modelo final

Variable	OR	Valor de P	Valor de Z	IC
Procedencia fuera del AMB	3,986	0,000	12,82	3.22734 - 4.92544
Sexo masculino	2,586	0,000	7,39	2.01003 - 3.32833
Disfuncionalidad familiar	1,304	0,043	2,03	1.00865 - 1.68678
Mala eficiencia personal	1,438	0,006	2,73	1.10838 - 1.86785
Pésima automotivación en el aprendizaje	1,552	0,000	3,74	1.23327 - 1.95452
Alto riesgo de insatisfacción para la vida	2,904	0,010	2,59	1.29625 - 6.50682
Segunda opción de ingreso al programa	1,487	0,004	2,86	1.13329 - 1.95343
Puesto de ingreso al programa (por cada puesto que se aleja del primer puesto se va incrementando el riesgo)	1,017	0,000	7,33	1.01307 - 1.02273

Variable	OR	Valor de P	Valor de Z	IC
Beneficio de 3 comidas en comedores	0,553	0,022	-2,29	0,33326 - 0,95877

8.4 Evaluación del rendimiento de la red neuronal artificial

Al someter el modelo a un conjunto de datos de prueba independiente, se llevó a cabo una evaluación exhaustiva de su precisión y eficiencia. Se usó el método “classification_report” que nos proporciona un informe detallado de varias métricas de evaluación de modelos de clasificación. Estas métricas se basan en la matriz de confusión y ofrecen una visión completa del rendimiento del modelo en términos de precisión, sensibilidad (recall) y precisión global (accuracy). Definiendo la precisión como la proporción de individuos predichos como "desertores" que realmente lo son. Un alto valor de precisión indica que la mayoría de los casos predichos como deserción realmente abandonaron. La sensibilidad como la proporción de individuos reales que abandonaron y que fueron correctamente identificados por el modelo. Un alto valor de sensibilidad indica que el modelo es capaz de identificar la mayoría de los casos de deserción. Y por último la precisión global que nos da la presión del modelo de la RNA.

Figura 11.

Matriz de confusión

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Nota. Tomado de: Book down (s.f.) *Metricas* https://bookdown.org/f_izco/BDC-POC/metricas.html

Estas métricas se calculan de la siguiente manera:

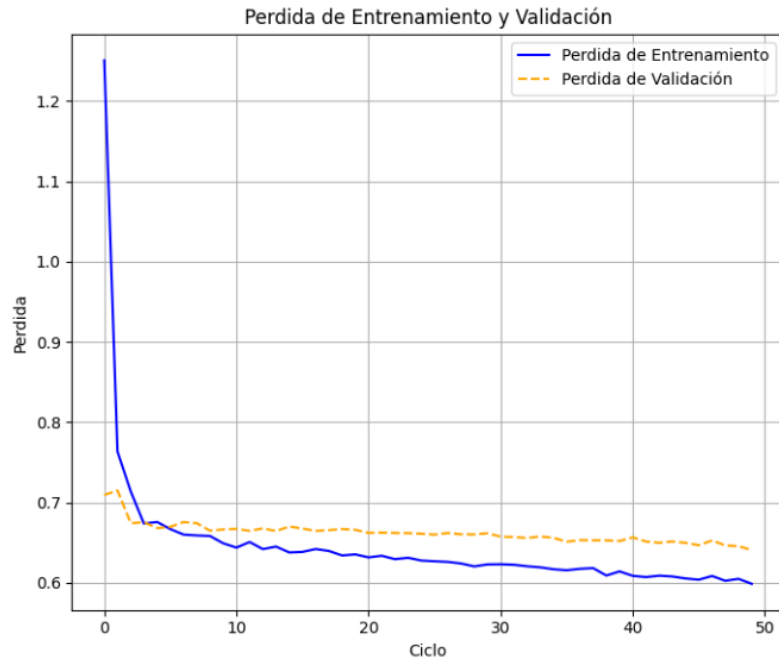
$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{37}$$

$$precisión = \frac{TP}{TP + FP} \tag{38}$$

$$sensibilidad = \frac{TP}{TP + FN} \tag{39}$$

Figura 12.

Perdida de entrenamiento y validación



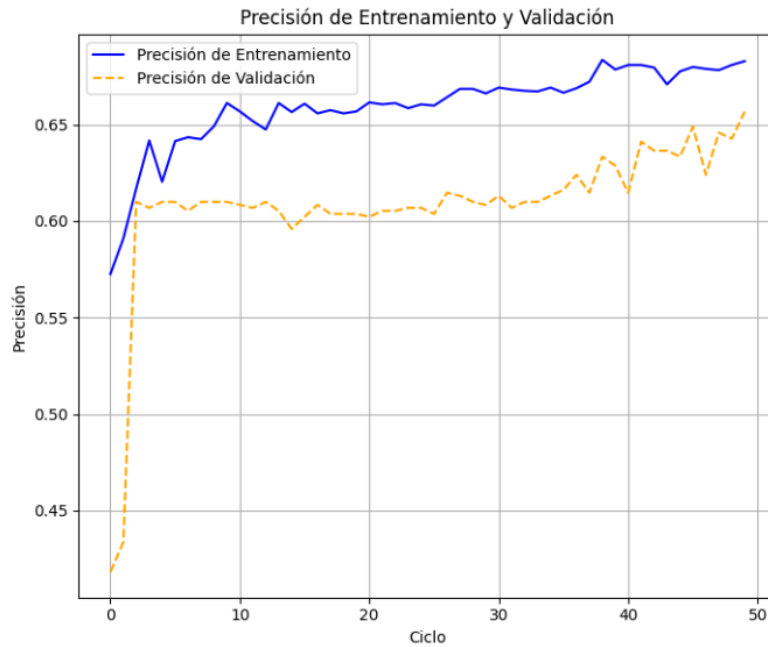
Adicionalmente de las siguientes gráficas, se pueden observar los siguientes puntos:

- Con la gráfica 6 se puede observar como la pérdida del modelo cambia durante el entrenamiento y la validación por cada ciclo, si la pérdida de entrenamiento y validación disminuyen, esto nos indica que el modelo está bien generalizado y no está sobreajustándose. Sin embargo, si la pérdida de validación está aumentando y la de entrenamiento está disminuyendo, puede ser un signo de un sobreajuste.

- Con la gráfica 7 se puede observar como la precisión del modelo cambia durante el entrenamiento y la validación en cada ciclo, que nos muestra que tanto la precisión de entrenamiento y validación deben estar en aumento, si esto fuera al contrario nos indica que existe un sobreajuste en el modelo

Figura 13.

Precisión de entrenamiento y validación



Con base en esos resultados se planteaban diferentes hiperparámetros de la RNA para optimizar su entrenamiento y rendimiento, estableciendo finalmente los hiperparámetros descritos anteriormente.

Los resultados revelaron una notable precisión superior al 85%, lo cual respalda de manera sólida el desempeño del modelo. Esta validación representa un paso fundamental hacia la implementación efectiva del modelo en situaciones reales, proporcionando una sólida confianza en su capacidad para tomar decisiones precisas y útiles en el mundo práctico.

```
accuracy = accuracy_score(y_prueba, y_pred.round())  
print(f'Precisión: {accuracy*100}')
```

```
Precisión: 87.57637474541752
```

8.5 Combinación de la red neuronal artificial y la regresión logística

Se propone realizar una combinación de la red neuronal artificial y la regresión logística, utilizando las variables del modelo final de la regresión logística como variables de entrada de la red neuronal, manteniendo la arquitectura e hiperparámetros seleccionados previamente. Con lo cual se observó una reducción de la precisión de la red neuronal artificial de 87% a 84%.

```
[ ] accuracy = accuracy_score(y_prueba, y_pred.round())  
print(f'Precisión: {accuracy*100}')
```

```
Precisión: 84.72505091649694
```

9. Discusión

A través de la creación del modelo de red neuronal que logró predecir la deserción universitaria con una precisión del 87%, nos sugiere que las redes neuronales son una herramienta prometedora para abordar el desafío de la deserción estudiantil, ya que pueden capturar relaciones complejas y no lineales entre una variedad de factores predictivos y el abandono de los estudios (Reyes et al., 2019). Además, el hecho de que el modelo haya alcanzado una tasa de precisión tan alta indica su capacidad para identificar eficazmente a los estudiantes en riesgo de desertar, lo que permite a la institución educativa intervenir de manera proactiva y brindar el apoyo necesario para mejorar las tasas de retención. Sin embargo, es importante reconocer que ningún modelo es perfecto y que siempre existe la posibilidad de errores de predicción. Por lo tanto, es crucial realizar

una validación rigurosa del modelo en diferentes conjuntos de datos y entornos educativos para garantizar su robustez y generalización. En última instancia, este modelo de red neuronal destaca el potencial de la inteligencia artificial y el aprendizaje automático para abordar desafíos complejos en el ámbito educativo y mejorar los resultados estudiantiles.

Dado que la regresión logística es especialmente adecuada para problemas de clasificación binaria, como la deserción estudiantil, donde el objetivo es predecir si un estudiante abandonará o no sus estudios. Ya que, la naturaleza binaria de la variable dependiente se ajusta perfectamente al marco de la regresión logística, lo que facilita la interpretación de los resultados. Esto nos deja entender cómo cada variable predictora influye en la probabilidad de deserción estudiantil. Los coeficientes estimados nos indican la dirección y la magnitud de la relación entre las variables independientes y la variable dependiente, lo que proporciona información valiosa sobre los factores que contribuyen a la deserción.

Además, la regresión logística es un modelo computacionalmente eficiente, lo que lo hace adecuado para grandes conjuntos de datos de estudiantes. Esto es importante en entornos educativos donde se recopilan grandes cantidades de información sobre los estudiantes, y se requiere un análisis rápido y efectivo para identificar aquellos en riesgo de deserción. Lo que la convierte en una herramienta valiosa para identificar y abordar este importante problema en el ámbito educativo.

A partir del análisis de regresión logística, se pudo concluir que los estudiantes que procedían fuera del Área Metropolitana de Bucaramanga presentaron un mayor riesgo de deserción temprana, lo cual también se ha descrito en otros estudios por el papel que cumplen el soporte familiar y la integración social en la adaptación a la vida universitaria, ya que esta etapa conlleva cambios significativos en la vida de una persona debido a las nuevas condiciones en las que se ve

inmersa, debido a que implica mayores responsabilidades, como el desarrollo de autonomía e independencia. La exposición a nuevos contextos en una etapa de vulnerabilidad como es la transición de la adolescencia a la adultez trae consigo la modificación de comportamiento y el modelamiento de ciertas dimensiones de la personalidad (Gómez et al., 2012). Un estudio llevado a cabo en una universidad del noroeste de Tailandia mostró que los estudiantes que se encontraban alejados del apoyo de su familia; definido como provenir de una provincia diferente a donde estaba ubicada la institución educativa, tenían 1,32 veces más probabilidades de desertar cada semestre, con una mayor deserción en el primer año que llegaba al 20%. Este fenómeno puede explicarse debido a que el soporte familiar está fuertemente asociado con experiencias de menor estrés emocional y sentimientos de soledad, así como mejor autoestima y autoeficacia académica (Sosu y Pheunpha, 2019). Por otra parte, cuando se inicia la vida universitaria los estudiantes procedentes fuera del AMB no cuentan con otras redes de apoyo que puedan brindar soporte inmediato. Además, otros factores contextuales que pueden influir negativamente en la integración es la falta de familiaridad con el entorno y la cultura. Así mismo, durante la transición del colegio a la universidad, los estudiantes cambian los patrones de alimentación que se caracteriza por la baja ingesta de frutas y verduras, así como el incremento de la ingesta de comidas rápidas o alimentos ultra procesados, que obedece a múltiples factores como: el poco tiempo para prepararse las comidas por las jornadas de estudio, el escaso conocimiento y pocas habilidades culinarias, y los problemas económicos que limitan la posibilidad de costearse una residencia con servicio de comidas (Pinzón, 2020). Estos patrones de alimentación están asociados al desarrollo de enfermedades cardiovasculares y metabólicas como la obesidad y la diabetes. Ahora bien, si tomamos en cuenta que estas condiciones médicas se asocian con estados proinflamatorios que afectan el sistema nervioso central y por ende deterioran las funciones cognitivas, es probable que

los estudiantes tengan menor desempeño académico, como se evidencia en un estudio llevado a cabo en estudiantes de secundaria de la ciudad de Mérida, en donde los adolescentes entre 12 y 15 años que presentaban sobrepeso u obesidad tuvieron un menor rendimiento académico (Salazar-Rendón et al., 2018). Por otra parte, cocinar también implica dedicar tiempo que podría destinarse al estudio (Ruiz-Casanova et al., 2021).

Por otra parte, se evidencio que los hombres tenían mayor riesgo de desertar en los primeros dos años, estos hallazgos coinciden con los reportados en un estudio llevado en Hungría en el cual explicaban este fenómeno debido a que los hombres se orientan a conseguir empleo y ganar dinero lo más pronto posible, mientras que a las mujeres se alentaba a continuar a la educación superior (Pusztai et al., 2022). Una situación similar ocurre en nuestro contexto, en donde la mayoría de los hombres asumen el rol de ser los principales proveedores económicos (Zárate-Rueda y Marcela, 2009). Si tenemos en cuenta que la mayoría de los estudiantes provienen de los estratos 1, 2 y 3 se infiere que tienen un bajo poder adquisitivo, por lo que es probable que muchos de nuestros estudiantes requieren ubicarse laboralmente para sufragar la manutención y contribuir con los gastos del hogar, que los lleva a comprometer el tiempo destinado para las actividades académicas en trabajar para proveer el sustento económico a sus familias, impactando negativamente en el rendimiento académico.

Adicionalmente, se aprecia que los estudiantes que percibían disfunción familiar presentaban un incremento del riesgo 1,3 veces mayor de deserción temprana, como se ha observado en otros estudios en donde los estudiantes con dificultades familiares o sociales tienen mayor riesgo de abandonar la universidad en el primer año académico, debido a que la disfunción familiar puede influir negativamente de diversas maneras cuando no brindan el soporte emocional y económico que requiere el estudiante (González-Ramírez y Pedraza-Navarro, 2016). En un

estudio cualitativo llevado a cabo con estudiantes adolescentes entre los 14 y 17 años en la región de la Araucanía en Chile, se encontró que una de las causas más decisivas para abandonar los estudios obedecía a la ocurrencia de dos elementos; los problemas económicos familiares y la falta de apoyo de los adultos responsables (Peña et al., 2016). Estos hallazgos concuerdan con los resultados obtenidos en un estudio de cohortes que evaluó el efecto acumulativo de la vulnerabilidad académica con la proximidad de contar con apoyo familiar, evidenciando que los estudiantes sin soporte familiar que presentaban bajas calificaciones tenían una probabilidad del 30% de abandonar la universidad en el segundo año (Sosu y Pheunpha, 2019). Por otra parte, cuando se analiza el efecto de las transiciones familiares, se observa que los estudiantes se ven afectados negativamente por el divorcio o la separación de los padres durante la escolarización (Song et al., 2012). En ese orden de ideas, (Ríos et al., 2020). realizaron un estudio que evidenció que el 65% de los estudiantes con rendimiento académico satisfactorio pertenecían a familias funcionales, sugiriendo que la disfunción familiar podría afectar la sensación de seguridad, protección y bienestar del estudiante. Además, en el estudio llevado por Tamayo, Hernández y Díaz (2022) en estudiantes de odontología, encontró que del 69,1% de los estudiantes con rendimiento académico destacado, el 55,1% describen un entorno familiar funcional, documentando una correlación estadísticamente significativa entre bajo rendimiento académico y disfunción familiar.

Entre los factores más frecuentemente mencionados en la literatura se incluyen la autoconfianza, la satisfacción, las prácticas educativas y el trasfondo cultural (Brown et al., 2016). No obstante, se destaca que la autoconfianza figura como uno de los elementos más determinantes para la persistencia en carreras universitarias (Brown et al., 2016). De lo antes planteado según Oermann (2017), la autoconfianza y la autoeficacia son conceptos estrechamente vinculados. Por

un lado, la autoeficacia establece un fundamento para motivar a los individuos hacia el bienestar y el logro de metas; por otro lado, la autoconfianza fomenta las habilidades en ciencias y la toma de decisiones de los estudiantes, aspectos que pueden incidir directamente en el desarrollo profesional (Tuttle, 2009). Por lo cual, en su estudio, Lin (2016) indicó que la educación basada en el aprendizaje activo se revela como un mecanismo efectivo para potenciar las habilidades de los estudiantes. Destacó que el aprendizaje activo contribuye a fortalecer la confianza y el éxito en las competencias de los estudiantes. Además, el aprendizaje activo en las ciencias cultiva el conocimiento mediante la participación en trabajos prácticos, generando así autoconfianza y el desarrollo del pensamiento crítico en los estudiantes (Acosta, 2019). Asimismo, la relación entre la satisfacción de la vida y el rendimiento académico ha sido objeto de estudio en psicología y educación (Martínez et al., 2010) sugiriendo que existe una conexión significativa entre una mayor satisfacción de la vida y un rendimiento académico más sólido. De la misma forma, la satisfacción de la vida hace referencia al bienestar del estudiante en relación con sus estudios, lo que implica una actitud positiva hacia la universidad y su programa académico. Por otra parte, el abandono hace referencia a las posibilidades que el estudiante considera de retirarse de la universidad. “La superación de retos y la consecución de objetivos aumenta la autoestima, la autoeficacia y en general produce satisfacción”. (Salonava et al., 2005). Haciendo énfasis en que la satisfacción personal tiene un papel importante en el rendimiento académico, como se muestra en el estudio realizado por Vélez y Roa (2005) con estudiantes universitarios en la Universidad del Rosario, donde la mayoría de los estudiantes que estaban satisfechos con su carrera y universidad, y alejados de presiones por parte de sus padres presentaron resultados positivos.

Por otro lado, se observó que los estudiantes que ingresan al programa académico que seleccionaron como segunda opción presentan un mayor riesgo de deserción temprana, debido a

que al acceder a un programa afín pero que se encuentra alineado en menor medida con sus intereses, puede asociarse a una reducción de la motivación intrínseca. El autor Salonava (2005), se refieren a este campo como el “engagement”, que definen como “un estado psicológico relacionado con los estudios que es positivo y significativo”. El “engagement” se caracteriza por: vigor, dedicación y absorción. Estas características se observan en aquellos estudiantes para quienes el estudio va más allá de una simple tarea; y lo perciben como una oportunidad para disfrutar del conocimiento. Dedicar extensas horas a sus actividades académicas con un entusiasmo evidente, ya que sienten que el tiempo transcurre rápidamente. Además, demuestran un elevado nivel de compromiso y capacidad de concentración en sus estudios. Por lo general, experimentan una sensación de gratificación al realizar estas tareas, ya que, para ellos, el estudio es una fuente de verdadero disfrute y satisfacción. A favor de esta premisa, se observa que los estudiantes muestran mayor motivación intrínseca cuando tienen la posibilidad de ubicarse en la carrera universitaria escogida como primera opción, lo que se correlaciona con mayor persistencia en las acciones que conllevan altos logros educativos (Song et al., 2012). En síntesis, los estudiantes que cursan el programa académico de interés experimentan altos niveles de satisfacción, lo que se correlaciona positivamente con el rendimiento académico.

Finalmente, el beneficio de recibir las tres comidas por la Dirección de Bienestar Estudiantil se comportó como un factor protector para la deserción temprana, esta asociación es consistente con los resultados obtenidos en un metaanálisis que sugieren que la ingesta de una dieta favorable se asocia con un mayor rendimiento académico en estudiantes universitarios (Burrows et al., 2017). Por otra parte, también se ha observado que los estudiantes universitarios con malnutrición ya sea por carencias como la desnutrición o por excesos como el sobrepeso o la obesidad presentan un bajo rendimiento académico (Salazar-Rendón et al., 2018). Ahora bien, los

estudiantes universitarios tienen altas demandas de energía debido a que la mayoría se encuentran en la transición de la adolescencia a la vida adulta, tiene alta movilidad, participan en diversas actividades y tiene una alta carga académicamente, por lo que requieren aportes nutritivos suficientes que permita el normal funcionamiento físico y cognitivo. Sin embargo, los estudiantes universitarios no practican hábitos alimentarios saludables, ya que consumen cantidades insuficientes de frutas y verduras, consumen altas cantidades de azúcares refinados y de grasas, por lo que su alimentación es poco variada y de baja calidad nutricional (Avila et al., 2022).

12. Conclusiones y Recomendaciones

A partir del modelo de regresión logística construido con los coeficientes de cada variable se calculó la incidencia para la muestra de 4273 estudiantes estableciendo un umbral de probabilidad de deserción del 0,5 dando como resultado una incidencia del 34,2%, que se obtuvo dividiendo el número de estudiantes que cumplen el criterio entre el número de estudiantes de la muestra, y luego multiplicamos por 100 para dar el resultado como porcentaje. Además, se calcula la prevalencia de los problemas de salud mental en los estudiantes con el mismo proceso dando como resultado 32,5%.

Con base en los resultados obtenidos es importante lograr una mayor participación de las familias en la educación de sus hijos, así como la comunicación con las instituciones de educación familiar para facilitar el afrontamiento de las dificultades que pueda conllevar la vida universitaria. Además, teniendo en cuenta que un porcentaje significativo de los estudiantes que provienen fuera

del AMB por diversos motivos (económicos, tiempo y conocimiento) están en riesgo de desarrollar patrones inadecuados de alimentación es necesario generar estrategias que promuevan hábitos nutricionales saludables. En ese orden de ideas, es necesario contar con un profesional en Nutrición y Dietética en la Dirección de Bienestar Estudiantil que cuide la distribución, combinación y proporciones de los diferentes grupos alimentarios, constituye una ventaja para garantizar una nutrición balanceada que potencie el rendimiento cognitivo. Adicionalmente, debido a que a medida que el puesto de ingreso del estudiante es mayor corre un mayor riesgo de desertar debido a que pueden no contar con los conocimientos previos necesarios para las asignaturas de primer nivel, por lo cual, es importante que los profesores diseñen evaluación diagnósticas en la primera semana para identificar los vacíos académicos que puedan interferir en el desarrollo de las competencias de la asignatura, para generar estrategias de nivelación dado que los estudiantes provienen de diferentes entornos académicos.

Por otra parte, mediante el modelo de redes neuronales artificiales se encontró una precisión del 87% que puede orientar la toma de decisiones en la vicerrectoría académica y la Dirección de Bienestar Estudiantil para conducir al fortalecimiento de las acciones para favorecer la retención universitaria, y la Universidad pueda continuar formando los profesionales que se requieren para solucionar los problemas que aquejan nuestra sociedad y contribuyan al desarrollo sostenible de la región de influencia. igualmente, mediante la combinación de la red neuronal artificial y la regresión logística se hizo una simplificación en las variables de entrada que inicialmente eran 26 dejando estas en solo nueve, facilitando la recolección de información que realizan las IES, observándose solo una reducción de la precisión del modelo en 3%, manteniéndose en niveles óptimos para su implementación.

Referencias Bibliográficas

- Acosta Ruiz de Mesones, R. L. (2019). *La formación investigativa de los estudiantes a través de prácticas de laboratorio en el área de ciencia y tecnología.*
- Alarcón, R. D. (2019). La Salud Mental de los estudiantes universitarios. *Revista Médica Herediana, 30*(4), 219-221.
- Aragón Borja, L. E. (2011). Perfil de personalidad de estudiantes universitarios de la carrera de Psicología: El caso de la Facultad de Estudios Superiores Iztacala. *Perfiles educativos, 33*(133), 68-87.
- Avila, F. J. M., Caneda-Bermejo, M. C., & Vivas-Castillo, A. C. (2022). Hábitos alimenticios y sus efectos en la salud de los estudiantes universitarios. Una revisión sistemática de la literatura: Dietary habits and health effects among university students. A systematic review. *Psicogente, 25*(47), 1-31.
- Ayçaguer, L. C. S., & Utra, I. M. B. (2004). *Regresión logística*. La Muralla, Hespérides.
- Bean, J. (1985). Interaction effects based on class level in an explanatory model of college student dropout syndrome. *American Educational Research Journal 22* (1): 35-64.
- Bean, J., & Eaton, S. B. (2001). The psychology underlying successful retention practices. *Journal of College Student Retention: Research, Theory & Practice, 3*(1), 73-89.
- Brown, P. L., Concannon, J. P., Marx, D., Donaldson, C. W y Black, A. (2016). An examination of middle school students' STEM self-efficacy with relation to interest and perceptions of STEM. *Journal of STEM Education, 17*(3), 27-37.

- Bruffaerts, R., Mortier, P., Kiekens, G., Auerbach, R. P., Cuijpers, P., Demyttenaere, K., ... & Kessler, R. C. (2018). Mental health problems in college freshmen: Prevalence and academic functioning. *Journal of affective disorders*, 225, 97-103.
- Burrows, T. L., Whatnall, M. C., Patterson, A. J., & Hutchesson, M. J. (2017, September). Associations between dietary intake and academic achievement in college students: A systematic review. *Healthcare* 5 (4), 60. MDPI.
- Cardona-Arias, J. A., Perez-Restrepo, D., Rivera-Ocampo, S., Gómez-Martínez, J., & Reyes, Á. (2015). Prevalencia de ansiedad en estudiantes universitarios. *Diversitas: perspectivas en psicología*, 11(1), 79-89.
- Ceballos, G. A., Charris, K. R., Gutiérrez, S. M. C., Acevedo, H. C. O., Herazo, E., & Arias, A. C. (2013). Asociación entre religiosidad y síntomas ansiosos y depresivos en estudiantes universitarios. *Medunab*, 16(1), 19-24.
- Cornfield, J., Gordon, T., & Smith, W. W. (1961). Quantal response curves for experimentally uncontrolled variables. *Bull Int Stat Inst*, 38(3), 97-115.
- Cristina, D., Guevara, D., Rodrigo, G. C., Ricardo, C., Aldana, B., de Paula, F., ... & Gabriel, V. C. (2017). *Modelado y simulación de sistemas naturales*. Editorial Tadeo Lozano.
- Cuadras, C. M. (2007). *Nuevos métodos de análisis multivariante* (Vol. 20). Barcelona: CMC editions.
- Cuamba Osorio, N., & Zazueta Sánchez, N. A. (2020). Salud mental, habilidades de afrontamiento y rendimiento académico en estudiantes universitarios. *Psicumex*, 10(2), 71-94.
- De la Horra Navarro, J. (2018). *Modelos matemáticos para ciencias experimentales*. Ediciones Diaz de Santos SA.

- Duval, F., González, F., & Rabia, H. (2010). Neurobiología del estrés. *Revista chilena de neuro-psiquiatría*, 48(4), 307-318.
- Fullér Robert. (1995). *Neural Fuzzy Systems*. Abo Akademi University. Turku.
- Galicia, L. E. V. (2009). ¿Estudias y trabajas?. Los estudiantes trabajadores de la Universidad Autónoma Metropolitana, Unidad Azcapotzalco. *Revista Latinoamericana de Estudios Educativos (México)*, 39(3-4), 121-149.
- García, E. A., Castillo-Jimenez, D. A., Cepeda, I., Pacheco, J. L., & López, R. P. (2019). Ansiedad y depresión en estudiantes universitarios: relación con rendimiento académico. *Interdisciplinary Journal of Epidemiology and Public Health*, 2(1).
- Gómez, M. V., Simich, L., Strike, C., Brands, B., Giesbrecht, N., & Khenti, A. (2012) *Medio social y uso simultáneo de alcohol y tabaco en estudiantes universitarios de pregrado de carreras de ciencias de la salud de una universidad, Cundinamarca–Colombia*.
- González Manteiga, M. T. A. (2003). *Modelos matemáticos discretos en las ciencias de la naturaleza*.
- González, J. R. H., & Hernando, V. J. M. (1995). *Redes neuronales artificiales: fundamentos, modelos y aplicaciones*.
- González-Ramírez, Teresa & Pedraza-Navarro, Inmaculada. (2016). The role of family in university drop out cases facing social and family related difficulties. *Journal of Global Research in Education and Social Science*. 8, 138-143. ISSN 2454-1834
- Gutiérrez Rodas, J. A., Montoya Vélez, L. P., Toro Isaza, B. E., Briñón Zapata, M. A., Rosas Restrepo, E., & Salazar Quintero, L. E. (2010). Depresión en estudiantes universitarios y su asociación con el estrés académico. *CES Medicina*, 24(1), 7-17.

- Hjorth, C. F., Bilgrav, L., Frandsen, L. S., Overgaard, C., Torp-Pedersen, C., Nielsen, B., & Bøggild, H. (2016). Mental health and school dropout across educational levels and genders: a 4.8-year follow-up study. *BMC public health*, 16, 1-12.
- Ibarra Mora, J., Hernández Mosqueira, C. M., & Ventura-Vall-Llovera, C. (2019). Hábitos alimentarios y rendimiento académico en escolares adolescentes de Chile. *Revista española de nutrición humana y dietética*, 23(4), 292-301.
- Insúa, S. R., Caballero, A. M., Lozoya, M. C. B., & Martín, A. J. (2004). *Investigación operativa: modelos determinísticos y estocásticos*. Editorial universitaria Ramón Areces.
- Jadue, G. (2001). Algunos efectos de la ansiedad en el rendimiento escolar. *Estudios pedagógicos (Valdivia)*, (27), 111-118.
- Kelsey, J. L. (1996). *Methods in observational epidemiology* (Vol. 10). Monographs in Epidemiology and.
- Lin, H. (2016). Effectiveness of simulation-based learning on student nurse's self-efficacy and performance while learning fundamental learning skills. *Technology and health Care*, 24, 369-375.
- Martin, B., & Sanz Molina, A. (2002). *Redes neuronales y sistemas difusos*.
- Martínez, J. A. M., León, A. G., Landa, J. M. A., & Zafra, E. L. (2010). La enseñanza de la Psicología a través de actividades desde una perspectiva multidisciplinar: aumento de la motivación y rendimiento. *European journal of education and psychology*, 3(1), 33-44.
- Mateos-Aparicio Morales, G., & Hernández Estrada, A. (2021). Análisis multivariante de datos: cómo buscar patrones de comportamiento en Big Data. *Análisis multivariante de datos: cómo buscar patrones de comportamiento en Big Data*.

McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5, 115-133.

MEN. (2009). *Deserción estudiantil en la educación superior colombiana*. Metodología de seguimiento, diagnóstico y elementos para su prevención.

Ministerio de Justicia y del Derecho – Observatorio de Drogas de Colombia (2019), *Estudio Nacional de Consumo de Sustancias Psicoactivas Colombia*. Bogotá DC.: ODC. <https://www.minjusticia.gov.co/programas-co/ODC/Documents/Publicaciones/Consumo/Estudios/estudio%20Nacional%20de%20consumo%202019v2.pdf?csf=1&e=iV5lh3>

Ministerio de Salud y Protección Social (2015) *Encuesta Nacional de Salud Mental, Colombia, 2015* <https://www.minsalud.gov.co/salud/publica/epidemiologia/Paginas/Estudios-y-encuestas.aspx>

Morales, V. H., Méndez, E. R., & Ibarrola, R. V. (2011). *Modelos probabilísticos y optimización*. Ediciones Académicas.

Navalón Mira, A., & Ruiz-Callado, R. (2017). Consumo de sustancias psicoactivas y rendimiento académico. una investigación en estudiantes de educación secundaria obligatoria. *Salud y drogas*, 17(1), 45-52.

Oermann, M. H., De Gagne, J. C., NPD-BC, C. N. E., & Phillips, B. C. (Eds.). (2017). *Teaching in nursing and role of the educator: The complete guide to best practice in teaching, evaluation, and curriculum development*. Springer Publishing Company.

Organización Mundial de la Salud (1997), *Life Skills Education in Schools*, Ginebra, Suiza.

Osada, J., Rojas, M., Rosales, C., & Vega-Dienstmaier, J. (2010). Sintomatología ansiosa y depresiva en estudiantes de medicina. *Revista de Neuro-psiquiatría*, 73(1), 15-19.

- Palacio Sprockel, L. E., Vargas Babilonia, J. D., & Monroy Toro, S. L. (2020). Bibliometric Analysis of Studies on Socioeconomic Factors in University Students. *Educación y Educadores*, 23(3), 355-375.
- Paula, I. (2016). *Problemas de salud mental en la adolescencia: Vulnerabilidad, riesgo y oportunidad*. https://www.researchgate.net/publication/255978892_Problemas_de_salud_mental_en_la_adolescencia_Vulnerabilidad_riesgo_y_oportunidad.
- Peña Axt, J. C., Soto Figueroa, V. E., & Calderón Aliante, U. A. (2016). La influencia de la familia en la deserción escolar: estudio de caso en estudiantes de secundaria de dos instituciones de las comunas de Padre las Casas y Villarrica, Región de la Araucanía, Chile. *Revista mexicana de investigación educativa*, 21(70), 881-899.
- Peña, D. (2014). *Fundamentos de estadística*. Alianza editorial.
- Pérez, A. B. D., Quispe, F. M. P., Aguilar, O. A. G., & Cortez, L. C. C. (2020). Transición secundaria-universidad y la adaptación a la vida universitaria. *Revista de Ciencias Sociales*, 26(3), 244-258.
- Pinzón, N. (2020). *Diseño de un sistema de apoyo para estudiantes foráneos en la Universidad Javeriana*. [Tesis de Licenciatura. Pontificia Universidad Javeriana. Colombia].
- Piper C. (2019). System for the Unified Management, Assessment, and Review of Information (SUMARI). *Journal of the Medical Library Association JMLA*, 107(4), 634–636. <https://doi.org/10.5195/jmla.2019.790>
- Prieto Meléndez, R. & Herrera, A & Pérez, J & Padrón-Godínez, A. (2000). *El modelo neuronal de McCulloch y Pitts Interpretación Comparativa del Modelo*.

- Pusztai, G., Fényes, H., & Kovács, K. (2022). Factors influencing the chance of dropout or being at risk of dropout in higher education. *Education Sciences*, 12(11), 804.
- Quesada, F. J. G., Graciani, M. A. F., Bonal, M. T. L., & Díaz-Mata, M. A. (1994). Aprendizaje con redes neuronales artificiales. *Ensayos: Revista de la Facultad de Educación de Albacete*, (9), 169-180.
- Ramírez, P. L., Cruz, Z. G., Rentería, M. D. J. R., Barrera, G. P., & Núñez-Hernández, A. (2019). Índice de masa corporal y rendimiento académico en estudiantes universitarios. *Revista Salud Pública y Nutrición*, 17(4), 34-40.
- Reyes, N. S., Morales, J. B., Moya, J. G., Teran, C. E., Rodriguez, D. N., & Altamirano, G. C. (2019). Modelo para predecir el rendimiento académico basado en redes neuronales y analítica de aprendizaje. *Revista Ibérica de Sistemas e Tecnologías de Informação*, (E17), 258-266.
- Ríos-Guzmán, R., Guzmán-Urizar, C., Pérez-Mazariegos, V., & Valdez-de León, H. (2020). Rendimiento académico y su relación con la funcionalidad familiar en estudiantes de Medicina. *Ciencias Sociales y Humanidades*, 7(2), 1-9.
- Ruiz-Casanova, S.M.R., Pérez-Coello, E.A., Canto-Esquivel, J.C., Ortega-Santana, G. y Canto-Esquivel, J.A. (2021). Aspectos relevantes en la adaptación de estudiantes foráneos para determinar estrategias académicas. *Revista del Centro de Graduados e Investigación. Instituto Tecnológico de Mérida*, 36(87), 56-61
- Salazar-Rendón, J. C., Méndez, N., & Azcorra, H. (2018). Asociación entre el sobrepeso y la obesidad con el rendimiento académico en estudiantes de secundaria de la ciudad de Mérida, México. *Boletín médico del Hospital Infantil de México*, 75(2), 94-103.

- Salonava, M., Cifre, E, Grau, R., Martínez Martínez, I. (2005). Antecedentes de la autoeficacia en profesores y estudiantes universitarios: un modelo causal. *Revista de Psicología del Trabajo y de las Organizaciones*. 1-2(21), 159-176.
- Silva-Ramos, M. F., López-Cocotle, J. J., & Columba Meza-Zamora, M. E. (2020). Estrés académico en estudiantes universitarios. *Investigación y Ciencia*, 28(79), 75-83.
- Song, C., Benin, M., & Glick, J. (2012). Dropping out of high school: The effects of family structure and family transitions. *Journal of Divorce & Remarriage*, 53(1), 18-33.
- Sosu, E. M., & Pheunpha, P. (2019, February). Trajectory of university dropout: Investigating the cumulative effect of academic vulnerability and proximity to family support. In *Frontiers in Education* (Vol. 4, p. 6). Frontiers Media SA.
- Spady, W. G. (1970). Dropouts from higher education: An interdisciplinary review and synthesis. *Interchange* 1 (1): 64-85.
- Tamayo-Cabeza, G., Hernández-Torres, A., & Díaz-Cárdenas, S. (2022). Funcionalidad familiar, soporte de amigos y rendimiento académico en estudiantes de odontología. *Universidad Y Salud*, 24(1), 18-28.
- Tinto, V. (1998). Colleges as communities: taking research on student persistence seriously. The educational character of student experience. *Review of Higher Education* 21 (2): 167-177.
- Torres, C. M., Vélez, P. J., & Altamar, B. F. (2015). La calidad de la educación superior en Colombia. Una aproximación econométrica (2007 – 2012). *Clío América*. 9(18), 143 – 156
- Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., ... & Straus, S. E. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, 169(7), 467-473.

Tuttle, R. E. (2009). *The impact of* simulation in nursing education on the self-efficacy and learner satisfaction of nursing students* [Doctoral dissertation, Capella University].

Universidad de La Salle. (2003). *Fortalecimiento del acompañamiento, seguimiento y control de los estudiantes en sus procesos académicos*. Bogotá.

Universidad de los Andes. (2006). *Investigación sobre deserción en las instituciones de educación superior en Colombia*. Informe final.

Universidad Industrial de Santander (2021). *Modelo Pedagógico*. <https://uis.edu.co/wp-content/uploads/2022/06/Acuerto-233-de-agosto-10-de-2021-Modelo-Pedagogico-UIS21.pdf>

Universidad Nacional de Colombia. (2002). *Estudio de la deserción estudiantil en la educación superior en Colombia*. Bogotá: Instituto Colombiano para el Fomento de la Educación Superior – ICFES.

Universidad Pontificia Bolivariana, seccional Bucaramanga. (2005). *Programa de acompañamiento académico: una experiencia exitosa para disminuir la deserción*. Ponencia en el Encuentro Internacional sobre Deserción en Educación Superior: Experiencias Significativas, Bogotá.

Vargas, G. M. G. (2007). Factores asociados al rendimiento académico en estudiantes universitarios, una reflexión desde la calidad de la educación superior pública. *Revista educación*, 31(1), 43-63.

Vélez Van, M. A., Roa, N. C. (2005). Factors associated with academic performance in medical students. *PSIC. Educación Médica*. 2(8), 1-10.

Walter S, Duncan D. (1967) Estimation of the probability of an event as a function of several variables. *Biometrika* 35(5), 79-167.

Yen J & Langari R. (2000). *Fuzzy Logic: Intelligence, Control and Information*. Prenhall. EUA.

Zárate-Rueda, R., & Socha, C. M. (2009). Estudio sobre las motivaciones de deserción estudiantil en la Universidad Industrial de Santander. *PROSPECTIVA. Revista de Trabajo Social e Intervención Social*, 411-430.

Apéndices

Apéndice A. Esquema de extracción de datos

ID	Autores	Título	Año	Tipo de estudio	Participantes	Concepto (método)	Contexto	VARIABLES explicativas (forma de medición)	Precisión del modelo	Principales hallazgos
1	Mayra Albán, David Mauricio	Decision Trees for the Early Identification of University Students at Risk of Desertion	2019	Estudio de Cohorte	3162 estudiantes de pregrado de las carreras de ingeniería de la universidad pública de Ecuador	Árbol de decisiones	Ecuador	Adicción a las redes sociales, edad, adicción al alcohol, dependencia emocional, adicción a las drogas, adicción al celular, adicción a los videojuegos, adicción a los juegos, adicción a las compras y deserción	97,95%	Las variables adicción a las redes sociales, dependencia emocional, estado marital y la edad, fueron considerados influenciadores del proceso de deserción
2	Dursun Delena, Kazim Topuz b, Enes Eryarsoyc	Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen	2019	Estudio experimental	Base de datos de 36,461 estudiantes de primer ingreso de una universidad pública	Modelo probabilístico de redes bayesianas	No reportado	Tipo de admisión (Nominal), edad (Numérica), Nivel de examen para el programa (Numérica), Escuela (Nominal), Nivel de estudio (Nominal), Etnia (Nominal), GPA	84%	Las variables con mayor índice de predicción son GPA de otoño, GPA acumulativo de otoño y Horas ganadas registras en otoño, con un

ID	Autores	Título	Año	Tipo de estudio	Participantes	Concepto (método)	Contexto	Variables explicativas (forma de medición)	Precisión del modelo	Principales hallazgos
		student attrition						acumulativo (Numérica), Horas gastadas (Numérica), Trabajo de otoño (Binaria), GPA (Numérica), Beca (Binaria), Horas registradas (Numérica), Préstamo estudiantil (Binaria), GPA ponderado (Numérica), Género (Binaria), Puntuacion SAT combinada (Numérica), Puntuación SAT ingles (Numerica), Puntuación SAT matemáticas (Numérica), Puntuación SAT lectura (Numerica), Puntuacion SAT ciencia (Numérica), GPA colegio (Numérica), Estado marital (Binaria), Meses luego del colegio (Numérica), Dirección estatal (Nominal), Ganancia de registro (Numérica), Financiamiento de otoño (Binaria), Financiamiento verano (Binaria), Residencia (Binaria), Registro segundo semestre (Binaria), Horas	índice mayor al 20%	

ID	Autores	Título	Año	Tipo de estudio	Participantes	Concepto (método)	Contexto	Variables explicativas (forma de medición)	Precisión del modelo	Principales hallazgos
								registradas en verano (Numérica), Beca de verano (Binaria), Trabajo de verano (Binaria), Préstamo de verano (Binaria) y Horas de transferencia (Numeric)		
3	Sergi Rovira, Eloi Puertas, Laura Igual	Data-driven system to predict academic grades and dropout	2017	Estudio experimental	Una base de datos de 4,434 estudiantes de derecho, matemáticas y ciencias de la computación en la universidad de Barcelona entre los años 2009-2014	Regresión logística, Adda-Boost, Naive Bayes, Bosque aleatorio y soporte maquina vector	España	No reportado	Naive bayes 75 % Regresión logística 92% Bosque aleatorio 76% Adda-Boost 73% SVM no reportado	Por medio de los diferentes modelos matemáticos se puede obtener una alta precisión en la deserción escolar
4	Dayana Vila, Saúl Cisneros, Pedro Granda, Cosme Ortega, Miguel Posso-Yépez, Iván García-Santillán	Detection of Desertion Patterns in University Students Using Data Mining Techniques: A Case Study	2018	Estudio experimental	Base de datos de 17,882 estudiantes de la Universidad Técnica del Norte	Bosque aleatorio y Naive Bayes	Ecuador	Edad, Etnia, Nacionalidad, Lugar de origen, Lugar de residencia, Grado de discapacidad, Género, Estado civil, Cargas familiares, Modalidad de estudio, Facultad, Notas promedio y Estado actual (Activo si o no)	Bosque aleatorio 97,6% Naive Bayes 97,5%	El método de bosque aleatorio proporciona ligeramente un mejor rendimiento que Naive Bayes, lo que nos permite identificar patrones relevantes para la deserción estudiantil.

ID	Autores	Título	Año	Tipo de estudio	Participantes	Concepto (método)	Contexto	Variables explicativas (forma de medición)	Precisión del modelo	Principales hallazgos
5	Y. Amaya, E. Barrientos, D. Heredia	Student Dropout Predictive Model Using Data Mining Techniques	2015	Estudio experimental	Estudiantes de la Universidad Simón Bolívar pertenecientes al programa de ingeniería de sistemas entre los periodos académicos 2009 y 2013	Árboles de decisiones	Colombia	Edad, Ciudad de procedencia, estrato, sexo, ocupacion, estado civil, nivel de estudios del padre, nivel de estudios de la madre, valor de matrícula, ingresos económicos, semestre cursado, jornada, materias cursadas, materias perdidas y promedio	No reportado	En el análisis de diferentes técnicas que se utilizan en la minería de datos específicamente para la predicción, se observa que los árboles de decisión resultan ser buenos clasificadores, según los resultados obtenidos y con base a estos algoritmos es mucho más sensible el IDE3
6	N. Lam-On, T. Boongoen	Using Cluster Ensemble to Improve Classification of Student Dropout in Thai University	2014	Estudio experimental	Una base de datos de 811 estudiantes de 26 departamentos académicos de la universidad de Mae Fah Luang	Cluster ensembles	Tailandia	Sexo, Ciudad de procedencia, Tipo de admision, Programa academico, GPAX (capacidad academica), GPA1, GPA2, GPA3 y GPA4	91%	El uso de Cluster ensembles para este propósito es novedoso que se puede mejorar con otros tipos de información como el carácter familiar y el carácter de los estudiantes.
7	Aulck Lovenoor, Velagapudi Nishant,	Predicting Student Dropout in	2016	Estudio experimental	Una base de datos de 69,111 estudiantes de	Regresion logistica Bosque aleatorio	USA	Raza, Género, Fecha de nacimiento, Estado de residencia, Identificación de hispano, Puntaje SAT,	Regresion logistica: 66,59%	Los resultados en matemáticas, inglés, química y las clases de

ID	Autores	Título	Año	Tipo de estudio	Participantes	Concepto (método)	Contexto	Variables explicativas (forma de medición)	Precisión del modelo	Principales hallazgos
	Blumenstock Joshua, West Jevin	Higher Education			la universidad de Washington's	k vecinos más próximos		Puntaje ACT, Clases tomadas, Hora en que fueron tomadas, Calificaciones recibidas y Especialidades declaradas	Bosque aleatorio: 62,24% k vecinos más próximos: 64,6%	psicología están entre los predictores más fuertes para la retención estudiantil, además de que el año de nacimiento también es un predictor fuerte para la deserción.
8	Abu-Oda, Ghadeer El-Halees, Alaa	DATA MINING IN HIGHER EDUCATION : UNIVERSITY STUDENT DROPOUT CASE STUDY	2015	Estudio experimental	Una base de datos de 1290 estudiantes de informática de la Universidad de ALAQSA matriculados entre 2005 y 2011	Árboles de decisiones Naive Bayes	Israel	Sexo (Binomial), GPA (Numerica), Promedio (Numerica), Nota del curso (Nominal), Nota programacion I (Numerica), Nota introducción a la computación (Numérica), Nota análisis de algoritmos (Numérica), Nota diseño lógico (Numérica), Nota estructura de datos I (Numérica), Nota estructura de datos II (Numérica) y región del estudiante (Binomial)	Árboles de decisiones: 98,14% Naive Bayes: 96,86%	El estudio muestra que los cursos de diseño lógico y análisis de algoritmos tienen un gran efecto en la predicción de la deserción estudiantil.

Apéndice B. Factores de la deserción

Determinantes de la Deserción			
Individuales	Académicos	Institucionales	Socioeconómicos
Edad	Nivel examen para el programa	Beca	Trabajo
Adicción a las redes sociales	Escuela (Departamento de ciencias)	Préstamo estudiantil	Estado marital
Adicción al alcohol	Tipos de admisión	Financiamiento	Ganancia de registro
Dependencia emocional	Nivel de escolaridad previo	Residencia	Cargas familiares
Adicción a las drogas	GPA* acumulativo		Escolaridad del padre
Adicción al celular	GPA* ponderado		Escolaridad de la madre
Adicción a los videojuegos	GPA* colegio		Valor matrícula
Adicción a las compras	tiempo indirecto de estudio		Ingresos económicos
Etnia	tiempo directo de estudio		
Género	Puntuación SAT*		
Meses luego del colegio	Semestre cursado		
Dirección estatal	Materias cursadas		
Nacionalidad	Materias perdidas		
Lugar de origen	Horario de clase		
Grado de discapacidad	Especialidades declaradas		
Fecha de nacimiento			
El promedio de calificaciones (GPA) es un sistema de calificación utilizado para medir el rendimiento académico de un estudiante.			
El puntaje (SAT) mide un rango de destrezas en lectura, análisis y escritura			

Apéndice C. Tabla de programas académicos

Programa académico	Total	Mujeres	Hombres
INGENIERIA DE PETROLEOS	155	45	110
LICENCIATURA EN INGLÉS	22	11	11
MATEMÁTICAS	106	30	76
INGENIERÍA ELÉCTRICA	202	36	166
GEOLOGÍA	196	75	121
LIC.LENGUAS EXTRANJERAS CON ÉNFASIS EN INGLÉS	77	39	38
INGENIERÍA INDUSTRIAL	299	152	147
FILOSOFÍA	145	69	76
INGENIERÍA ELECTRÓNICA	229	46	183
INGENIERÍA CIVIL	280	88	192
DISEÑO INDUSTRIAL	144	78	66
INGENIERÍA METALÚRGICA	183	65	118
INGENIERÍA QUÍMICA	223	106	117
FISIOTERAPIA	44	30	14
DERECHO	194	105	89
INGENIERÍA DE SISTEMAS	254	48	206
ECONOMÍA	141	53	88
LICENCIATURA EN ESPAÑOL Y LITERATURA	70	26	44
LICENCIATURA EN LITERATURA Y LENGUA CASTELLANA	60	32	28
INGENIERÍA MECÁNICA	274	33	241
QUÍMICA	136	70	66
TRABAJO SOCIAL	148	118	30
LICENCIATURA EN MATEMÁTICAS	137	71	66
LICENCIATURA EN MÚSICA	105	28	77
FISICA	48	8	40
MEDICINA	111	62	49
MICROBIOLOGIA Y BIOANALISIS	72	43	29
BIOLOGÍA	47	28	19
LIC EN EDUC BASICA CON ENFASIS CIENCIAS NAT EDU AMBI	18	15	3
HISTORIA Y ARCHIVÍSTICA	52	25	27
NUTRICION Y DIETETICA	32	26	6
ENFERMERÍA	47	34	13
LIC.EN EDUC.BÁSICA ÉNFASIS.EN LENGUA CASTELLANA	21	17	4

Apéndice D. Tabla de puesto de ingreso

Puesto de ingreso	Cantidad	Porcentaje
1	65	1,52
2	69	1,62
3	68	1,59
4	62	1,45
5	72	1,69
6	64	1,50
7	51	1,19
8	64	1,50
9	68	1,59
10	75	1,76
11	69	1,62
12	68	1,59
13	53	1,24
14	57	1,33
15	73	1,71
16	66	1,54
17	72	1,69
18	73	1,71
19	68	1,59
20	71	1,66
21	60	1,40
22	67	1,57
23	59	1,38
24	69	1,62
25	65	1,52
26	64	1,50
27	68	1,59
28	68	1,59
29	66	1,54
30	60	1,40
31	61	1,43
32	69	1,62
33	65	1,52
34	78	1,83
35	61	1,43
36	61	1,43
37	70	1,64
38	54	1,26
39	62	1,45
40	64	1,50

Puesto de ingreso	Cantidad	Porcentaje
41	54	1,26
42	58	1,36
43	49	1,15
44	61	1,43
45	60	1,40
46	56	1,31
47	44	1,03
48	44	1,03
49	46	1,08
50	47	1,10
51	34	0,80
52	40	0,94
53	38	0,89
54	38	0,89
55	42	0,98
56	36	0,84
57	43	1,01
58	36	0,84
59	35	0,82
60	36	0,84
61	33	0,77
62	33	0,77
63	29	0,68
64	26	0,61
65	24	0,56
66	23	0,54
67	23	0,54
68	24	0,56
69	25	0,59
70	26	0,61
71	25	0,59
72	26	0,61
73	24	0,56
74	26	0,61
75	27	0,63
76	22	0,51
77	21	0,49
78	27	0,63
79	14	0,33
80	23	0,54
81	15	0,35
82	10	0,23
83	18	0,42
84	12	0,28

Puesto de ingreso	Cantidad	Porcentaje
85	13	0,30
86	13	0,30
87	14	0,33
88	13	0,30
89	16	0,37
90	11	0,26
91	14	0,33
92	14	0,33
93	7	0,16
94	10	0,23
95	10	0,23
96	7	0,16
97	9	0,21
98	6	0,14
99	8	0,19
100	8	0,19
101	4	0,09
102	3	0,07
103	5	0,12
104	4	0,09
106	1	0,02
107	3	0,07
108	1	0,02
109	1	0,02
111	1	0,02
112	2	0,05
113	1	0,02
115	1	0,02

Apéndice E. Tabla de variables en el análisis univariado

Variables	Todos		Mujeres		Hombres	
	Población	IC95%	Población	IC95%	Población	IC95%
Sin riesgo socioeconómico	4115	0,95 - 0,96	1665	0,94 - 0,96	2450	0,95 - 0,97
Con riesgo socioeconómico	157	0,03 - 0,04	73	0,03 - 0,05	84	0,02 - 0,04
Satisfecho con la vida	3847	0,89 - 0,9	1547	0,87 - 0,9	2300	0,89 - ,91
Bajo riesgo de insatisfacción en la vida	378	0,08 - 0,09	169	0,08 - 0,11	209	0,07 - 0,09
Alto riesgo de insatisfacción en la vida	47	0,008 - 0,014	22	0,007 - 0,012	25	0,006 - 0,014
Buenas competencias lectoras	2036	0,46 - 0,49	802	0,43 - 0,48	1234	0,46 - 0,50
Bajas competencias lectoras	1953	0,44 - 0,47	862	0,47 - 0,51	1091	0,41 - 0,45
Pésimas competencias lectoras	283	0,05 - 0,07	74	0,03 - 0,05	209	0,07 - 0,09
Buenas capacidades de consulta	2365	0,53 - 0,56	1013	0,04 - 0,07	1352	0,51 - 0,55
Bajas capacidades de consulta	1459	0,32 - 0,35	495	0,26 - 0,30	964	0,36 - 0,39
Pésimas capacidades de consulta	448	0,09 - 0,11	230	0,11 - 0,14	218	0,07 - 0,09
Buena motivación para el aprendizaje	2376	0,54 - 0,57	1056	0,58 - 0,63	1320	0,50 - 0,54
Baja motivación para el aprendizaje	97	0,01 - 0,02	46	0,01 - 0,03	51	0,01 - 0,02
Pésima motivación para el aprendizaje	1799	0,40 - 0,43	636	0,34 - 0,38	2534	0,43 - 0,47
Buena confianza para el aprendizaje	1557	0,35 - 0,37	702	0,38 - 0,42	855	0,31 - 0,35
Baja confianza para el aprendizaje	2574	0,58 - 0,61	976	0,53 - 0,58	1598	0,61 - 0,64
Pésima confianza para el aprendizaje	141	0,02 - 0,03	60	0,02 - 0,04	81	0,02 - 0,03
Buenas estrategias de aprendizaje	2230	0,50 - 0,53	910	0,49 - 0,54	1320	0,50 - 0,54
Malas estrategias de aprendizaje	886	0,19 - 0,21	534	0,28 - 0,32	352	0,12 - 0,15
Pésimas estrategias de aprendizaje	1156	0,25 - 0,28	294	0,15 - 0,18	862	0,32 - 0,35

Variables	Todos		Mujeres		Hombres	
	Población	IC95%	Población	IC95%	Población	IC95%
Buen cuidado de sí mismo	1676	0,37 - 0,40	716	0,38 - 0,43	960	0,35 - 0,39
Bajo cuidado de sí mismo	921	0,20 - 0,22	386	0,20 - 0,24	535	0,19 - 0,22
Pésimo cuidado de sí mismo	1675	0,37 - 0,40	636	0,34 - 0,38	1039	0,39 - 0,42
Buenas relaciones interpersonales	2334	0,53 - 0,56	950	0,52 - 0,57	1384	0,52 - 0,56
Malas relaciones interpersonales	623	0,13 - 0,15	244	0,12 - 0,15	379	0,13 - 0,16
Pésimas relaciones interpersonales	1315	0,29 - 0,32	544	0,29 - 0,33	771	0,28 - 0,32
Buena eficiencia personal	2828	0,64 - 0,67	1151	0,63 - 0,68	1677	0,64 - 0,68
Mala eficiencia personal	347	0,07 - 0,08	143	0,06 - 0,09	204	0,07 - 0,09
Pésima eficiencia personal	1097	0,24 - 0,27	444	0,23 - 0,27	653	0,24 - 0,27
Buenos hábitos del sueño	2345	0,53 - 0,56	954	0,52 - 0,57	1391	0,52 - 0,56
Malos hábitos del sueño	832	0,18 - 0,20	301	0,15 - 0,19	531	0,19 - 0,22
Pésimos hábitos del sueño	1095	0,24 - 0,26	483	0,25 - 0,29	612	0,22 - 0,25
Funcionalidad familiar	3402	0,78 - 0,80	1392	0,78 - 0,81	2010	0,77 - 0,80
Disfuncionalidad familiar	807	0,19 - 0,21	346	0,18 - 0,21	524	0,19 - 0,22
Sin riesgo de ansiedad	2530	0,57 - 0,60	935	0,51 - 0,56	1595	0,61 - 0,64
Bajo riesgo de ansiedad	530	0,11 - 0,13	160	0,07 - 0,10	370	0,13 - 0,16
Alto riesgo de ansiedad	1212	0,27 - 0,29	643	0,34 - 0,39	569	0,20 - 0,24
Sin riesgo de depresión	2880	0,65 - 0,68	1073	0,59 - 0,64	1807	0,69 - 0,73
Bajo riesgo de depresión	322	0,06 - 0,08	108	0,05 - 0,07	214	0,07 - 0,09
Alto riesgo de depresión	1070	0,23 - 0,26	557	0,29 - 0,34	513	0,18 - 0,21
Sin riesgo de alcoholismo	4039	0,83 - 0,95	1675	0,95 - 0,97	2364	0,92 - 0,94
Con riesgo de alcoholismo	233	0,04 - 0,06	63	0,02 - 0,04	170	0,05 - 0,07
Colegio público	2865	0,65 - 0,68	1176	0,65 - 0,69	1689	0,64 - 0,68
Colegio privado	1407	0,31 - 0,34	562	0,30 - 0,34	845	0,31 - 0,35

Variables	Todos		Mujeres		Hombres	
	Población	IC95%	Población	IC95%	Población	IC95%
Primera opción	2373	0,54 - 0,57	944	0,51 - 0,56	1429	0,54 - 0,58
Segunda opción	1899	0,42 - 0,45	794	0,43 - 0,48	1105	0,41 - 0,45
Sin descuento	4060	0,94 - 0,95	1675	0,95 - 0,97	2385	0,93 - 0,95
Con descuento	212	0,02 - 0,04	63	0,01 - 0,03	149	0,03 - 0,05
Sin auxiliaturas	4042	0,93 - 0,95	1631	0,92 - 0,94	2411	0,94 - 0,95
Con auxiliaturas	230	0,04 - 0,06	107	0,05 - 0,07	123	0,04 - 0,057
Sin servicio de comedores	3563	0,82 - 0,84	1451	0,81 - 0,85	2112	0,81 - 0,84
Una comida de comedores	378	0,08 - 0,09	171	0,08 - 0,11	207	0,07 - 0,09
Dos comidas de comedores	168	0,03 - 0,04	64	0,02 - 0,04	104	0,03 - 0,04
Tres comidas de comedores	163	0,03 - 0,04	52	0,02 - 0,03	111	0,03 - 0,05