



FACULTAD DE POSTGRADO

TESIS DE POSTGRADO

**IMPLEMENTACIÓN DE MODELO PREDICTIVO PARA
IDENTIFICAR DESERCIÓN ESTUDIANTIL EN UNITEC**

SUSTENTADO POR:

**WILLIAM JOSUÉ CABALLERO CÁCERES
MELISSA ALEJANDRA CABRERA BURGOS**

PREVIA INVESTIDURA AL TÍTULO DE

**MÁSTER EN
ANALÍTICA DE NEGOCIOS**

TEGUCIGALPA, F. M., HONDURAS, C.A.

NOVIEMBRE 2023

UNIVERSIDAD TECNOLÓGICA CENTROAMERICANA

UNITEC

FACULTAD DE POSTGRADO

AUTORIDADES UNIVERSITARIAS

RECTORA

ROSALPINA RODRÍGUEZ GUEVARA

PRORECTOR/ SECRETARIO GENERAL

ROGER MARTÍNEZ MIRALDA

VICERRECTOR ACADÉMICO

JAVIER ABRAHAM SALGADO LEZAMA

DIRECTORA NACIONAL DE POSTGRADO

ANA DEL CARMEN RETTALLY

**IMPLEMENTACIÓN DE MODELO PREDICTIVO PARA
IDENTIFICAR DESERCIÓN ESTUDIANTIL EN UNITEC**

**TRABAJO PRESENTADO EN CUMPLIMIENTO DE LOS
REQUISITOS EXIGIDOS PARA OPTAR AL TÍTULO DE**

**MÁSTER EN
ANALÍTICA DE NEGOCIOS**

**ASESOR METODOLÓGICO
HENRY OSORTO**

MIEMBROS DE LA TERNA:

**KEVIN FÚNEZ
DANIEL LUNA
JULISSA CORTÉS**

DERECHOS DE AUTOR

© Copyright 2023
William Josué Caballero Cáceres
Melissa Alejandra Cabrera Burgos

Todos los derechos son reservados.



FACULTAD DE POSTGRADO

**IMPLEMENTACIÓN DE MODELO PREDICTIVO PARA IDENTIFICAR DESERCIÓN
ESTUDIANTIL EN UNITEC**

WILLIAM CABALLERO

MELISSA CABRERA

RESUMEN

Esta investigación se realizó con el objetivo de implementar un modelo predictivo para identificar los estudiantes con riesgo de deserción de un periodo académico a otro en la Universidad Tecnológica Centroamericana (UNITEC) en estudiantes de pregrado de las ciudades de Tegucigalpa y San Pedro Sula. En el estudio se realizaron procesos de análisis de datos, procesamiento de datos como limpiezas y transformaciones, se entrenó el algoritmo de aprendizaje y se evaluaron los resultados en función de cada uno de los indicadores de eficiencia, una vez realizado el entrenamiento se llevó a cabo el proceso de implementación, todo ello con el fin de que el modelo predictivo sea una alerta temprana para la universidad para mitigar la deserción, mejore el enfoque de esfuerzos y recursos y aumente los índices de retención.

Palabras claves: Algoritmo de aprendizaje / Aprendizaje Automático / Deserción/ Inteligencia Artificial / Modelos Predictivos



GRADUATE SCHOOL

**IMPLEMENTATION OF A PREDICTIVE MODEL TO IDENTIFY STUDENT
DROPOUT AT UNITEC**

WILLIAM CABALLERO

MELISSA CABRERA

ABSTRACT

This research was conducted with the aim of implementing a predictive model to identify students at risk of dropping out from one academic period to another at the Universidad Tecnológica Centroamericana (UNITEC) among undergraduate students in the cities of Tegucigalpa and San Pedro Sula. The study involved data analysis processes, data preprocessing such as cleaning and transformations, training of the learning algorithm, and evaluation of results based on each of the efficiency indicators. Once the training was completed, the implementation process was carried out, all with the goal of making the predictive model an early alert for the university to mitigate dropout, improve the focus of efforts and resources, and increase retention rates.

Key words: Learning Algorithm / Machine Learning / Dropout / Artificial Intelligence / Predictive Models

DEDICATORIAS

Dedico con mucho aprecio este trabajo a mi madre Marcela y mi padre Francisco quienes han sido los pilares fundamentales de mi vida al brindarme su apoyo en todo momento para cumplir cada una de mis metas. William Caballero

Dedico este trabajo a mi esposo Miguel Cáliz por su paciencia, apoyo y motivación a lo largo de toda la maestría. Melissa Cabrera

AGRADECIMIENTOS

Agradezco primeramente a Dios y la virgen Maria por el conocimiento, salud y sabiduría que siempre me proporcionan en mi vida para lograr mis metas, de igual manera un agradecimiento especial para mi amada esposa Floribeth que ha sido mi apoyo incondicional a lo largo de esta carrera para la obtener este nuevo logro. William Caballero

Quiero expresar un profundo agradecimiento en primer lugar a Dios por guiarme y darme salud y fuerzas para culminar esta maestría, a mi madre por inspirarme a ser mejor cada día y porque le debo todos mis logros y a mi hermana por sus palabras de ánimo y siempre estar pendiente de mí. Melissa Cabrera

De igual manera esta investigación no hubiese sido posible sin el apoyo de Unitec, agradecemos su contribución y disposición para proporcionarnos los datos necesarios para llevar a cabo este estudio.

ÍNDICE DE CONTENIDO

CAPITULO I. PLANTEAMIENTO DEL PROBLEMA.....	12
1.1 INTRODUCCIÓN	12
1.2 ANTECEDENTES.....	13
1.3 DEFINICIÓN DEL PROBLEMA	14
1.4 OBJETIVOS.....	15
1.4.1 OBJETIVO GENERAL	15
1.4.2 OBJETIVO ESPECÍFICOS	15
1.5 JUSTIFICACIÓN.....	15
CAPITULO II. MARCO TEÓRICO.....	16
2.1 ANÁLISIS DE SITUACIÓN ACTUAL.....	16
2.1.1 LA EDUCACIÓN SUPERIOR EN AMÉRICA LATINA	16
2.1.2 IMPACTO DE COVID-19 EN DESERCIÓN ESTUDIANTIL.....	16
2.1.3 EDUCACIÓN SUPERIOR EN HONDURAS.....	17
2.1.4 DESERCIÓN UNIVERSITARIA EN HONDURAS	19
2.2 CONCEPTUALIZACIÓN.....	20
2.2.1 DESERCIÓN ESTUDIANTIL.....	20
2.2.2 FACTORES DE DESERCIÓN	21
2.2.3 MODELOS PREDICTIVOS.....	22
2.3 TEORÍAS DE SUSTENTO	22
2.3.1 BASES TEÓRICAS	22
2.3.2 METODOLOGÍAS DESARROLLADAS POR OTROS INVESTIGADORES.....	29
2.3.3 INSTRUMENTOS UTILIZADOS.....	31
2.4 MARCO LEGAL	32
CAPITULO III. MARCO METODOLÓGICO.....	33
3.1 CONGRUENCIA METODOLÓGICA	33
3.1.1 MATRIZ METODOLÓGICA.....	33
3.2.3 ESQUEMA DE VARIABLES DE ESTUDIO.....	35
3.1.3 OPERACIONALIZACIÓN DE LAS VARIABLES	36
3.1.4 HIPÓTESIS	38

3.2 ENFOQUE Y MÉTODOS	39
3.3 DISEÑO DE LA INVESTIGACIÓN	39
3.3.1 POBLACIÓN	39
3.3.2 MUESTRA	39
3.3.3 TÉCNICAS DE MUESTREO.....	40
3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS	40
3.5 FUENTES DE INFORMACIÓN.....	43
3.5.1 FUENTES PRIMARIAS	43
CAPITULO IV. RESULTADOS Y ANÁLISIS.....	43
4.1 ANÁLISIS EXPLORATORIO DE DATOS (EDA)	43
4.2 RESULTADOS DE MODELOS	52
CAPITULO V. CONCLUSIONES	54
5.1 CONCLUSIONES	54
5.2 RECOMENDACIONES	55
CAPITULO VI. APLICABILIDAD.....	57
6.1 NOMBRE DE LA PROPUESTA	57
6.2 JUSTIFICACIÓN DE LA PROPUESTA.....	57
6.3 ALCANCE DE LA PROPUESTA	57
6.4 DESARROLLO DE LA PROPUESTA.....	57
6.4.1 DESCRIPCIÓN DE LA PROPUESTA	57
6.4.2 DESARROLLO DE LA PROPUESTA	58
6.5 MEDIDAS DE CONTROL	63
6.6 CRONOGRAMA Y PRESUPUESTO DE IMPLEMENTACIÓN	64
6.7 CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA.....	65
BIBLIOGRAFÍA.....	68
ANEXOS.....	73
ANEXO I.....	73
ANEXO II	74
ANEXO III.....	74

ÍNDICE DE TABLAS

TABLA 1. MATRIZ METODOLÓGICA.....	33
TABLA 2. ESQUEMA DE VARIABLES DE ESTUDIO.....	35
TABLA 3. OPERACIONALIZACIÓN DE LAS VARIABLES	36
TABLA 4. ASOCIACIÓN DE VARIABLES CATEGÓRICAS CONTRA LA VARIABLE DE DESERCIÓN ESTUDIANTIL.....	50
TABLA 5. ASOCIACIÓN DE VARIABLES NUMÉRICAS CONTRA LA VARIABLE DE DESERCIÓN ESTUDIANTIL.....	51
TABLA 6. CORRELACIÓN DE VARIABLES NUMÉRICAS RELEVANTES PARA EL ESTUDIO	52
TABLA 7. INDICADORES DE EFICIENCIA DE MODELOS PREVIOS A IMPLEMENTAR	52
TABLA 8. VARIABLES DE ENTRADA DEL MODELO.	59
TABLA 9. INDICADORES DE RESULTADOS DEL MODELO.....	63
TABLA 10. CRONOGRAMA DE ACTIVIDADES PARA IMPLEMENTACIÓN DE MODELO PREDICTIVO.....	64
TABLA 11. CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA.....	65

ÍNDICE DE FIGURAS

FIGURA 1. DIAGRAMA DE ENFOQUE Y MÉTODOS	39
FIGURA 2. POBLACIÓN Y DESERTORES DE PREGRADO.....	44
FIGURA 3. GÉNERO DE ESTUDIANTES NO DESERTORES Y DESERTORES DE PREGRADO	45
FIGURA 4. GRÁFICO DE EDAD DE ESTUDIANTES NO DESERTORES Y DESERTORES DE PREGRADO.....	45
FIGURA 5. TIPO DE ESTUDIANTE DE PREGRADO.....	46
FIGURA 6. GRÁFICO DE CAJAS DE PROMEDIO DE MATERIAS MATRICULADAS Y MATERIAS APROBADAS DE ESTUDIANTES DESERTORES DE PREGRADO	46
FIGURA 7. RELACIÓN ENTRE ÍNDICE ACADÉMICO DE PERÍODO E INASISTENCIAS DE ESTUDIANTES PREGRADO.....	47
FIGURA 8. PORCENTAJE DE ESTUDIANTES DE PREGRADO CON DESCUENTO	47
FIGURA 9. COMPARATIVA DE DESCUENTO DE ESTUDIANTES DESERTORES Y NO DESERTORES	47
FIGURA 10. PORCENTAJE DE PERMISOS DE MATRICULA DE ESTUDIANTES DE PREGRADO.....	48
FIGURA 11. PORCENTAJE DE PERMISOS DE MATRICULA DE ESTUDIANTES DESERTORES DE PREGRADO	49
FIGURA 12. PROMEDIO DE CUOTAS PAGADAS Y CUMPLIMIENTO DE CUOTAS DE DESERTORES Y NO DESERTORES	50
FIGURA 13. MATRIZ DE CONFUSIÓN DE ALGORITMOS ENTRENADOS.....	53
FIGURA 14. CURVAS ROC DE ALGORITMOS ENTRENADOS	53
FIGURA 15. ENTRENAMIENTO CON ÁRBOL DE DECISIÓN	55
FIGURA 16. DIAGRAMA DE COMPORTAMIENTO.....	56
FIGURA 17. CÁLCULO DE VARIABLE DE RESPUESTA	59
FIGURA 18. PREPROCESAMIENTO DE VARIABLE NUMÉRICAS.....	61
FIGURA 19. PREPROCESAMIENTO DE VARIABLE CATEGÓRICAS	61
FIGURA 20. ENTRENAMIENTO DE MODELO	62

CAPITULO I. PLANTEAMIENTO DEL PROBLEMA

1.1 INTRODUCCIÓN

La deserción es una condición en la que un estudiante deja de tener actividad académica ya sea parcial o definitiva a lo largo de su carrera estudiantil, esto genera impacto negativo al estudiante en diferentes ámbitos a largo plazo como ser primordialmente la dificultad para la inserción laboral y crecimiento académico, por otro lado la deserción también afecta a las instituciones académicas y al país en general, en donde para las universidades tanto públicas como privadas se ha convertido en uno de los principales problemas ya que estas dejan de percibir ingresos por matrícula y/o mensualidades además de que sus indicadores de matrículas se ven afectados constantemente y a nivel de país la deserción produce un impacto a largo plazo de manera socioeconómica ya que incrementa el desempleo y los índices de pobreza.

Existen varias razones que causan la deserción estudiantil, en algunos casos puede darse por razones económicas, psicológicas como la falta de motivación, problemas emocionales, situaciones familiares o salud, por otro lado, la calidad de educación, docentes y facilidades de las instituciones universitarias para promover una buena metodología de enseñanza – aprendizaje.

Por ello, ante esta problemática el objetivo de este estudio es aplicar técnicas de inteligencia artificial mediante modelos predictivos para la Universidad Tecnológica Centroamericana (UNITEC) durante un periodo académico para la identificación temprana de estudiantes con riesgo de abandonar sus estudios universitarios, utilizando información histórica de estudiantes recolectada de la universidad de tal manera que se convierta en una herramienta de alerta, para que la institución tenga la capacidad de prevenir la deserción estudiantil, minimizar pérdidas y sea capaz de aplicar estrategias de retención de estudiantes de manera anticipada.

1.2 ANTECEDENTES

La deserción es un indicador que ayuda a medir la eficiencia y calidad en una institución educativa (Pineda, Pedraza y Moreno, 2011) y es entendida como el abandono del sistema de manera permanente o transitorio, ya sea por características personales o de desempeño académico (Zavala et al. 2018). A su vez también se puede definir como una variedad de situaciones identificadas en el proceso educativo del estudiantado con un denominador común, detención o interrupción de estudios iniciados antes de finalizarlos (Cabrera et al. 2006)

Así mismo, la deserción estudiantil tiene consecuencias sociales, emocionales y económicas, no solo en el propio estudiante sino en su entorno más cercano (González y Girón, 2005). Adicionalmente, quienes no concluyen sus estudios son subempleados y no obtienen los ingresos económicos deseados (Díaz, 2008). De igual manera, es un tema que tiene trascendencia no solo en los individuos, sino también en las instituciones y en las políticas nacionales (Barrero, 2015).

En Honduras durante el 2020, la deserción aumentó en las principales universidades tanto públicas como privadas, se estimó que el 20% a 25% de los estudiantes abandonaron sus estudios lo que representa cerca de 50 mil estudiantes (Diario Tiempo, 2020). Según el informe del SITEAL (2016) en el 2013 los estudiantes con un nivel económico bajo presentan mayor porcentaje de deserción (65.6%) en comparación con aquellos que provienen de estratos altos (56.2%) y en ambos casos los índices de deserción supera el 50%, los estudiantes que provienen de zonas urbana desertan el 62.9%, mientras que la rural en un 72.7%, a nivel de género los hombres abandonan sus estudios en un 63.3% y las mujeres en 36.3%, en términos generales en promedio el 63.7% de los estudiantes desertan de sus estudios universitarios (Educación superior en Iberoamérica. Informe 2016. Informe nacional: Honduras - CINDA, s. f.)

A la fecha la deserción se ha tratado de comprender tanto de manera cualitativa como cuantitativa, en donde de manera cuantitativa, se ha llevado a cabo mediante modelos estadísticos buscando la probabilidad de deserción, como presenta Barrero (2015), los principales modelos estadísticos que explican la deserción son; de elección discreta y los de supervivencia. Asimismo, con el apoyo de la inteligencia artificial y a través de minería de datos se pueden encontrar patrones de los estudiantes para entender mejor las variables que pueden ser determinantes de la deserción y poder aplicar un modelo de aprendizaje automático predictivo tal como el ejemplo que menciona Sifuentes (2018) mediante arboles de decisión. De manera similar Jiménez y Tamiran (2015)

realizaron estudios acerca de la detección de patrones a partir de datos socioeconómicos, académicos, disciplinares e institucionales de los estudiantes para obtener conocimiento que ayude en la toma de decisiones enfocadas en las políticas y estrategias relacionadas con retención estudiantil, este estudio utilizó los algoritmos J48, Apriori y K-means y los resultados obtenidos determinaron un factor común para los estudiantes que desertan, el cual está relacionando con bajos promedios y materias perdidas en los primeros semestres de la carrera.

La manera cualitativa comprende que el problema de deserción está basado en la retención de estudiantes y se enfoca en tutorías, asesoría interdisciplinaria, cursos preparatorios, atención psicológica, seguimiento a docente, entre otros (Barrero, 2015).

1.3 DEFINICIÓN DEL PROBLEMA

Los elevados índices de deserción estudiantil tanto en las universidades públicas y privadas generan pérdidas económicas, así mismo, tiene impactos negativos en la calidad educativa, posicionamiento o prestigio de la universidad, enfoque de recursos y en el seguimiento estudiantil.

La identificación de la deserción estudiantil representa una limitación por el gran volumen de estudiantes que dificulta el seguimiento oportuno de los mismos, esto representa una oportunidad para la identificación temprana de estudiantes con riesgo de deserción y que mediante la aplicación de herramientas predictivas se busca determinar la causas para la aplicación de medidas de retención.

1. ¿Se reducirían las pérdidas económicas de la universidad al poder identificar de manera temprana la deserción estudiantil?
2. ¿Cuáles son los factores más relevantes que ocasionan la deserción estudiantil?
3. ¿Cuál es el comportamiento de los estudiantes desertores en UNITEC?
4. ¿Al implementar un modelo predictivo de retención se reducirán los índices de deserción estudiantil en la Universidad Tecnológica Centroamericana (UNITEC)?

1.4 OBJETIVOS

1.4.1 OBJETIVO GENERAL

Identificar de manera temprana la deserción estudiantil en un periodo académico en la Universidad Tecnológica Centroamericana (UNITEC) en los centros de Tegucigalpa y San Pedro Sula, mejorando los índices de retención y mitigando pérdidas económicas para la institución, mediante la aplicación de modelos predictivos.

1.4.2 OBJETIVO ESPECÍFICOS

- Identificar los principales factores que causan la deserción de un estudiante.
- Conocer el comportamiento de los estudiantes desertores en UNITEC
- Desarrollar una propuesta de diseño e implementación de un modelo predictivo de inteligencia artificial mediante algoritmos de aprendizaje basados en la deserción académica para mejorar la retención estudiantil.

1.5 JUSTIFICACIÓN

Se realizará el estudio para mitigar la deserción estudiantil en UNITEC, con el fin de tener una visión temprana para aplicación de estrategias de retención en las instituciones educativas, para disminuir los índices de deserción, reducir pérdidas económicas, mejorar el enfoque de esfuerzos y recursos, tanto financieros como administrativos para la retención de estudiantes.

El estudio se llevará a cabo mediante el diseño e implementación de un modelo predictivo en busca de predecir la deserción de los estudiantes en UNITEC para estudiantes de Pregrado de las ciudades de Tegucigalpa y San Pedro Sula durante un periodo académico.

Por otro lado, el estudio mediante la implementación de técnicas de inteligencia artificial busca comprender la eficiencia de los modelos predictivos para la estimación de la deserción de estudiantes y los factores que causan la misma.

CAPITULO II. MARCO TEÓRICO

2.1 ANÁLISIS DE SITUACIÓN ACTUAL

2.1.1 LA EDUCACIÓN SUPERIOR EN AMÉRICA LATINA

La Educación Superior Universitaria (ESUP) en América Latina y el Caribe, está en un momento determinante respecto a la retención y deserción de estudiantes, aunque la tasa bruta promedio de matrícula casi se ha duplicado entre el año 2000 al 2013 de 21% a 43% respectivamente, (Ferreira, Avitabile, Botero, Haimovich, & Urzúa, 2017, citado de Munizaga, Cifuentes & Beltrán, 2017), el índice de retención es de 46%, por lo que más de la mitad de la población entre las edades de 25 a 29 años que estudia a nivel de educación superior aún no han finalizado su plan de estudio, ya sea porque aún están estudiando (32%) o porque han desertado de manera permanente de las instituciones educativas (22%) (Banco Mundial, 2017, citado de Munizaga, Cifuentes & Beltrán, 2017).

La educación superior conlleva enormes beneficios tanto para las personas que pueden acceder a ella como para el conjunto de la sociedad. No sólo implica el acceso a mejores empleos y remuneraciones, sino también permite contar con mejores condiciones de salud, participar más activamente en los desafíos ambientales y otros que compartimos como sociedad. Además, este nivel educativo se vincula estrechamente con la construcción de un proyecto individual de desarrollo, facilita la participación en la sociedad del conocimiento y facilita un proceso de aprendizaje continuo y una mayor adaptación a la vorágine de cambios que afectan nuestra sociedad. (Valenzuela & Yañez, 2022).

2.1.2 IMPACTO DE COVID-19 EN DESERCIÓN ESTUDIANTIL

Con la pandemia del año 2020 para evitar la propagación masiva del virus del COVID-19, en América Latina, el Instituto Internacional de la UNESCO para la Educación Superior en América Latina y el Caribe (IESALC) sugirió el cierre de las Instituciones de Educación Superior y la preparación de acciones de contingencia para modificar la enseñanza presencial a espacios virtuales (Jaramillo, 2021, citado en Flores, et al. 2023).

No obstante, los países de dicha región experimentaron problemas tales como; la baja conectividad (el 52% de los hogares cuenta con equipamiento tecnológico y conectividad), plataformas obsoletas, deserción educativa, problemas económicos, falta de contenido en línea,

baja preparación de los profesores para la enseñanza en línea, donde las poblaciones con mayores ingresos lograron acceder a la tecnología y garantizaron su formación, mientras que las poblaciones con menores ingresos no lo lograron al carecer de las condiciones requeridas para la continuidad de su formación. (Balderas et al., 2021, citado en Flores, et al. 2023).

El mayor impacto que el COVID-19 ha tenido en las Instituciones de Educación Superior (IES) de América Latina, ha estado en la transición súbita de la docencia presencial a la docencia no presencial, mediada por las tecnologías. Esta situación ha repercutido negativamente en todos los factores que hacen posible el funcionamiento de las instituciones, tales como estudiantes, profesores, gerentes institucionales y trabajadores, así como para las instituciones y el sistema en su conjunto. Ruiz Bolívar, C. (2022)

Desde el inicio de la pandemia por covid-19 la deserción estudiantil ha tenido un crecimiento exponencial y se ha dado en todas las instituciones educativas, a nivel socioeconómico ha tenido un impacto significativo en el entorno de los estudiantes, por otro lado las instituciones educativas se enfrentaron a una nueva modalidad virtual, empleando distintas plataformas educativas y diversos recursos tecnológicos, ya que se limitó la educación presencial, esto tuvo un efecto negativo provocando un alto índice de deserción estudiantil. (Vásquez, 2023).

2.1.3 EDUCACIÓN SUPERIOR EN HONDURAS

La institución rectora de la educación superior en Honduras es la Universidad Nacional Autónoma de Honduras (UNAH) y por medio del Consejo de Educación Superior (CES) es responsable de dirigir y desarrollar el sistema de educación superior.

El Sistema de educación superior es regido y orientado por la Ley de Educación Superior y su Reglamento. De igual manera, se cuenta con las Normas Académicas de Educación Superior que son de aplicación y observancia obligatoria en todas las instituciones y centros del Nivel, públicos y privados y que fueron emitidas por el Consejo de Educación Superior en 1992 (Calderón, 2011).

En Honduras la demanda inmediata o población meta para el nivel de educación superior es la población entre los 18 a los 24 años, para el año 2012 se estimaba en 1,130,578 y la demanda potencial que es la población entre 0 a 17 años estimada en el mismo año 2012 en 3,693,052 habitantes y la demanda potencial total de estudiantes para educación superior era de 304,164 entre

las edades de 18-24, alrededor del 60% fue la población atendida/matriculada en educación superior y un 40% fue demanda insatisfecha (Plan Estratégico SES 2014-2023).

La cobertura brindada por las 20 Instituciones de Educación Superior (IES) es de 14.3% de la población estimada que tiene entre 18 y 24 años, aunque se han creado nuevas instituciones esta cifra no ha variado en la última década, su oferta educativa está concentrada en las ciudades de Tegucigalpa y San Pedro Sula, donde se ubica cerca del 40% de los campus o sedes, siendo que para el año 2008 ninguna universidad tenía cobertura de los departamentos de Valle, La Paz, Lempira, Gracias a Dios e Islas de la Bahía, teniendo en cuenta que Honduras está conformada por 18 Departamentos (Calderón, 2011).

Existe un grupo de la población entre las edades de 25 a 64 años que no podrán realizar sus estudios de educación superior de manera inmediata, algunos de los motivos son por no tener su título de educación media o porque trabajan y no continúan con los estudios de educación superior (Plan Estratégico SES 2014-2023).

Como mencionan López & Moncada, G. (2016). El sistema de educación superior en Honduras está cambiando, tal como lo muestra la transformación de las escuelas normales, la regulación de la educación a distancia y la virtualidad. El Plan Estratégico de Desarrollo de la Educación Superior 2014 – 2023 es una muestra del interés que se está brindando a este rubro, con objetivos estratégicos claros que finalmente impactarán tanto en la cobertura, como en la calidad de la educación superior entre ellos se encuentran:

- Promover e impulsar el crecimiento armonioso e integral del Sistema de Educación Superior, para fortalecer el desarrollo del país.
- Aumentar la cobertura de Educación Superior al promedio latinoamericano, con equidad poblacional, regional, étnica y con atención a la población con necesidades especiales.
- Ampliar y diversificar la oferta académica a nivel de grado y postgrado, en carreras orientadas fundamentalmente al desarrollo de la ciencia y la tecnología.
- Certificar la gestión y calidad del sistema de Educación Superior.
- Actualizar las Normas académicas de la Educación Superior de Honduras para armonizarlas con las tendencias de homologación/armonización de estándares, títulos, currículos, en la región Centroamericana.

(Cristal, 2021; Guerrero y Soto, 2019; citado en Castillo, et al. 2022) menciona que, en Honduras se ha ampliado la oferta académica de carreras de grado, y se han creado nuevas universidades en el sector privado para atender la demanda nacional. De las cuales la universidad pública constituye la mayor parte de la población de estudiantes universitarios, produciendo consigo fenómenos como la reducción de la calidad educativa, limitaciones para la introducción de innovaciones pedagógicas, uso de modelos clásicos de enseñanza para grupos amplios, dificultad para responder a las necesidades de cada alumno durante su ingreso y permanencia.

En el 2010 el 61% de la matrícula se registra en las universidades públicas, la UNAH es responsable del 72% de ese porcentaje; las IES privadas registran el 39% de la matrícula total en el Sistema de Educación Superior de Honduras. El 57% del total de estudiantes en las IES del país son mujeres, también se observa que las universidades públicas, con excepción de la UNAH y la UPNFM, no han tenido un aumento estadísticamente significativo en el número de estudiantes matriculados. En el caso de las IES privadas, la UNITEC, UTH y UNICAH son las que más han aumentado el número de estudiantes matriculados en los últimos cinco años, otro dato relevante es que a partir del año 2008 las universidades privadas son las que cuentan con un mayor número de carreras a nivel de licenciatura. Sin embargo, las universidades públicas ofertan el mayor número de posgrados (Calderón, 2011).

2.1.4 DESERCIÓN UNIVERSITARIA EN HONDURAS

En el 2013 en Honduras el 36.3% culminó los estudios universitarios, sin embargo, el 63.7% deserto, los estudiantes con bajos ingresos económicos tienen mayor tasa de deserción (65.6%) en comparación a los que provienen de estratos altos (56.2%), cabe mencionar que en ambos casos la deserción representa más del 50%. A nivel demográfico, la población urbana deserta el 62.9%, mientras que de zonas rurales es el 72.7%, en cuanto a género, el 64.1% de los hombres no logran culminar los estudios superiores y el 63.3% de las mujeres se retiran. Respecto a la tasa de graduación, no hay diferencia significativa entre las universidades privadas y públicas de Honduras, en el 2014 la tasa de las universidades públicas y privadas fue de 12.6% y 13% respectivamente, cabe mencionar que a nivel de postgrado las tasas son más altas que en pregrado (Duriez M et al., 2016).

Para 2020 según diario el tiempo se reportó cifras deserción de estudiantes tan solo para la Universidad Nacional de Honduras aproximadas del 20 %, de igual manera el rector de la UTH

Javier Mejía informo que presentaron deserción estudiantil del 30% y así mismo el rector de UNITEC Marlon Brevé comunico que la deserción en total para la institución fue de un 20%.

2.2 CONCEPTUALIZACIÓN

2.2.1 DESERCIÓN ESTUDIANTIL

No existe una única definición para el concepto de deserción estudiantil universitaria. Según el Diccionario Enciclopédico de Ciencias de la Educación (Picardo, O., Escobar, J., & Balmore, R., 2004, citado en Viale, H., 2014), la deserción es el acto deliberado o forzado mediante el cual el estudiante deja su aula o centro educativo.

La deserción estudiantil es el abandono prematuro de las obligaciones académicas que puede tener un estudiante en una institución en particular o totalmente del sistema educativo, la deserción se puede dividir de tal manera como menciona Sifuentes (2018), en deserción interna, ocurre cuando el estudiante decide cambiarse de carrera en la misma institución de educación superior; deserción institucional, se produce cuando el estudiante abandona la universidad para matricularse en otra; y deserción definitiva del sistema educativo, se da cuando el estudiante no se matricula en ninguna universidad.

La deserción estudiantil hoy en día se ha convertido en un problema no solo para las instituciones académicas sino en un problema general para la sociedad, economía y para el individuo, en donde un individuo sin estudios universitarios o estudios incompletos conlleva un mayor riesgo para adquirir un trabajo de acuerdo con sus destrezas, preferencias o expectativas salariales afectándolo directamente de manera social, psicológica y económica.

Cada estudiante que abandona crea un lugar vacante en el conjunto estudiantil que pudo ser ocupado por otro alumno que persistiera en los estudios. Por consiguiente, la pérdida de estudiantes causa serios problemas financieros a las instituciones al producir inestabilidad en la fuente de sus ingresos (Portillo, 2023). Es común ver repercusiones económicas por lo general en universidades privadas no obstante las universidades públicas también se ven afectadas no solo por lo ingresos que pueden generar, sino que también por la reputación de la universidad frente a la matrícula de estudiantes ingresados contra estudiantes egresados.

La deserción en su mayor parte es causada por el comportamiento de los estudiantes ya sea a través de un factor social, académico o financiero, sin embargo, la deserción también puede ser ocasionada por el ambiente de la institución, calidad de educación o por accesibilidad.

2.2.2 FACTORES DE DESERCIÓN

La deserción puede ser ocasionada por múltiples causas que tienen una alta variabilidad de acuerdo con la situación de cada individuo por lo cual estas causas se categorizan según los factores de deserción estudiantil detallados a continuación:

- Factores sociales, los cuales abarcan el comportamiento del estudiante desde la motivación por el aprendizaje y el deseo por tener un cierto grado académico.
- Factores académicos, que incluyen el rendimiento académico del estudiante a lo largo de su vida estudiantil en función de sus calificaciones promedio, carga de cursos aprobados, abandonados o reprobados.
- Factores económicos, considerados como de los más relevantes dado que no todos los estudiantes tienen las mismas capacidades financieras ya que depende de un ingreso ya sea familiar o individual, así mismo dependerá la capacidad de pago de un estudiante.
- Factores institucionales, los cuales se tratan de la calidad de educación proporcionada al estudiante, el proveer recursos necesarios para el proceso de aprendizaje, la comunicación y seguimiento continuo con los estudiantes.

Según Portillo (2023), algunas de las causas más comunes son los factores económicos, la escasa información acerca de la carrera, el proceso de adaptación a la vida universitaria, la falta de metodologías de estudio, y de una acción de retención estudiantil o preventiva a la deserción por parte de la institución de educación superior. La mayoría de las causas mencionadas se encuentran dentro del alcance de cada institución universitaria de tal manera que pueden mejorarse estos puntos una vez identificados para poder reducir la deserción continua o prevenirla, sin embargo, la deserción abarca más causas que están relacionadas directamente con los estudiantes en donde las instituciones deben de implementar medidas de retención para una mejor comprensión del problema.

De igual manera Canales & De los Rios (2007) mencionan que los factores socioeconómicos son especialmente relevantes. Dado que el nacimiento de hijos, matrimonio, muerte o enfermedad de la cabeza de hogar son factores que podrían desencadenar la deserción definitiva del sistema educativo.

2.2.3 MODELOS PREDICTIVOS

Los modelos predictivos tienen como objetivo predecir valores futuros de los datos en busca de un valor concreto o una respuesta como afirma García et al. (2018) que los datos van acompañados de una salida (clase, categoría o valor numérico) en donde la regresión y la clasificación son técnicas comúnmente usadas en este tipo de modelos

Como define (¿Qué es la inteligencia artificial? | Microsoft Azure, s. f.), los modelos de aprendizaje automático se entrenan con subconjuntos de datos. Cuando los datos que se usan para entrenar el modelo representan con similitud el conjunto de datos real al cual se implementará el modelo, el algoritmo provee una mayor precisión.

Siendo lo anterior se debe de tomar en cuenta que para el entrenamiento de soluciones predictivas se debe de contar con suficiente información histórica a manera de obtener una mejor respuesta a la variable que se desea predecir o un mejor poder predictivo, siendo así que a mayor cantidad de datos es posible que el algoritmo de aprendizaje pueda ajustar mejor los resultados en base a los comportamientos de las variables proporcionadas.

Los modelos predictivos orientados a dar una respuesta a la deserción estudiantil son soluciones de clasificación siendo como tal, identificar de una observación la probabilidad de que ocurra un suceso y categorizando en base a una respuesta esperada. Siendo en el caso de la deserción estudiante se busca conocer si un estudiante puede ser o no desertor.

2.3 TEORÍAS DE SUSTENTO

2.3.1 BASES TEÓRICAS

En este apartado se establecen los conceptos teóricos que dan soporte a la investigación, al hablar de la implementación de modelos predictivos en busca de conocer el riesgo de deserción estudiantil, se consideran varias teorías relevantes que influyen en el desarrollo de una solución predictiva en el ámbito académico.

Determinar las causas o factores del abandono es una tarea compleja, puesto que en su predicción inciden diferentes variables, modelos teóricos indican que la intención de abandono antecede la concretización del abandono y se denota como uno de los principales predictores de deserción universitaria, de igual forma, se evidencia como uno de los elementos claves implicados en el proceso psicológico de este fenómeno educacional (Mashburn, 2000, citado en Castillo, M., 2022).

Las investigaciones de Bethencourt et al., 2008, citado en Viale, H., 2014), han permitido demostrar la importancia de ciertas características psicológicas del estudiantado universitario en la terminación de sus estudios, parece ser que la persistencia para acabar la titulación a pesar de los obstáculos, la motivación hacia la titulación cursada, la capacidad de esfuerzo en favor de logros futuros, el ajuste entre las capacidades del alumnado y las exigencias de la titulación, y la satisfacción con la titulación cursada son garantías de éxito académico.

Por su parte, (Bean y Metzner, 1985, citado en Castillo, M., 2022) proponen un modelo teórico en el que definen la intención de abandono como la probabilidad estimada de un estudiante de suspender sus estudios, la cual incluye la intención informada de los discentes de cambiar de carrera o abandonar la universidad, este modelo plantea que la intención es vista como una señal de alerta temprana de deserción real y es explicada por factores que incluyen resultados académicos (rendimiento), antecedentes previos a la universidad, variables académicas y ambientales.

Otros factores relevantes en cuánto al abandono estudiantil son los factores personales y socioeconómicos especialmente aquellos estudiantes con dificultades de ajuste financiero a nivel familiar y con bajos niveles de escolaridad parental (García et al., 2016). Asimismo, estudiantes con bajo rendimiento académico en áreas como las ciencias exactas, físico y matemáticas presentan mayor probabilidad de abandono de estudios (Gairín et al., 2014, citado en Castillo, M., 2022).

Las experiencias académicas de adaptación temprana, en lo vinculado con las vivencias de carrera y orientación vocacional, aspectos personales como el bienestar físico y psicológico, hábitos de estudio, aspectos sociales/interpersonales (tener una red de apoyo en la universidad) y la percepción de la calidad de los servicios de la institución educativa representan factores protectores ante la intención de abandono universitario (Casanova et al., 2021, citado en Castillo, M., 2022).

Tomando en consideración el proceso de transición vital y académica que experimentan los estudiantes de primer ingreso, este segmento estudiantil se ha identificado como uno de los principales grupos de riesgo de abandono en educación superior (Blair, 2017, citado en Castillo, M., 2022), ya que los estudiantes de primer año deben adaptarse a nuevas dinámicas educacionales y personales que demandan un fuerte proceso de reajustes de autonomía, autorregulación y

madurez que no todos los estudiantes afrontan con suceso (Holdsworth et al., 2018, citado en Castillo, M., 2022).

Gran parte de la literatura referida a la permanencia de los estudiantes en la universidad se desarrolla sobre la base de dos principales teorías sociológicas: el modelo de integración del estudiante Student Integration Model Spady de 1971 y de Tinto de 1975, explica que dadas todas las demás variables constantes, un mayor grado de integración del estudiante en el ambiente académico y social contribuye a un mayor grado de compromiso institucional y esto afecta directamente a la decisión del alumno de permanecer o desertar. El segundo modelo acerca de desgaste del estudiante Student Attrition Model de Bean y Metzner de 1985, atribuye mayor importancia relativa a los factores externos a la institución (Viale, H., 2014).

(Cabrera, Bethencourt, Álvarez y González 2006, citado en Viale, H., 2014) consideran que, si bien estos modelos parecieran ser opuestos, en realidad existe complementariedad entre ellos y a su vez superposición en algunos aspectos.

2.3.1.1 RETENCIÓN ESTUDIANTIL EN EDUCACIÓN SUPERIOR

2.3.1.1.1 MEDIDAS DE RETENCIÓN

Menciona Fernández, Corengia, & Durand (2014), que el nivel de retención hace referencia a la tasa de permanencia de los estudiantes en una institución en particular.

Según Arguelles & Steffanell (2023) los modelos de retención y permanencia estudiantil se deben desarrollar teniendo en cuenta las variables institucionales y las características de la población estudiantil, de esta forma se determinan las estrategias que se deben implementar y los objetivos que se desean alcanzar.

Con el aumento de la deserción estudiantil en las instituciones universitarias se ha identificado que se deben de implementar estrategias de retención estudiantil las cuales deben de estar basada en los factores que causan la deserción con el objeto de aumentar los niveles de retención, pero en la mayoría de los casos se desconoce cuáles son esas causas, sin embargo a través de los modelos predictivos se pueden establecer posibles factores y probabilidades de deserción de los estudiantes, los cuales pueden contribuir a la institución a la identificación de medidas a implementar para la retención de estudiantes ya sea a través del acompañamiento estudiantil para un seguimiento cercano con el estudiante hasta un apoyo financiero mediante becas o financiamientos.

A manera de mitigar la deserción y poder mejorar los índices de retención estudiantil o niveles de retención hoy en día se pueden hacer uso de modelos predictivos que sean capaces de poder determinar la deserción de un estudiante a lo largo de su vida académica.

2.3.1.2 LA CIENCIA DE DATOS EN LA EDUCACIÓN SUPERIOR

2.3.1.2.1 CIENCIA DE DATOS

La ciencia de datos es el estudio de datos con el fin de extraer información significativa para empresas. Es un enfoque multidisciplinario que combina principios y prácticas del campo de las matemáticas (¿Qué es la ciencia de datos? - Explicación de la ciencia de datos - AWS, s. f.)

La ciencia de datos es aplicable a múltiples áreas de estudio siempre y cuando existan datos que se puedan recopilar, procesar y analizar para poder realizar soluciones analíticas capaces de dar respuestas a comportamientos desconocidos, la mejora de procesos y predicciones futuras. No obstante, la ciencia de datos no solo es aplicable para el desarrollo de soluciones analíticas sino para la búsqueda de conocimiento en los datos dando respuesta al pasado, presente y el futuro a través del análisis de estos.

2.3.1.2.2 ANÁLISIS DE DATOS

El análisis de datos es un proceso que involucra técnicas de consumo y preprocesamiento de datos a través de la minería de datos y el análisis exploratorio mediante el cual intenta conocer el comportamiento de cada variable en un conjunto de datos interpretándolo tanto de manera técnica como orientado al negocio en busca de generar información valiosa para dar respuesta a un problema.

El análisis de datos es primordial en la construcción de una solución predictiva, ya que consiste en mantener los datos lo más entendibles en función de contexto de cada variable a medida que la información resultante sea consistente y que el proceso para aplicar técnicas de aprendizaje al momento de realizar entrenamientos con algoritmos de aprendizaje automático.

2.3.1.3 LA INTERPRETACIÓN DE LOS DATOS

2.3.1.3.1 ANÁLISIS EXPLORATORIO DE DATOS

El análisis exploratorio de datos forma parte fundamental del análisis de datos al proveer técnicas para mantener consistente la información de tal manera que involucra técnicas de visualización, estadística descriptiva, manipulación y limpieza de datos con el fin de comprender el contexto de cada variable en busca de la coherencia de la información. Dentro de las técnicas

de visualizaciones de datos más utilizadas en el análisis de variables numéricas se encuentran los box plot también conocidos como gráficos de cajas o bigotes ayudan a identificar de manera rápida valores atípicos y medias de las variables de estudio.

Es importante realizar un análisis exploratorio de datos ya que implica en la limpieza y transformaciones de los datos importante convirtiéndose en un paso previo para la aplicación de técnicas como minería de datos tal como lo afirma Llaugel & González (2016) la recolección, preparación de datos, y la interpretación de los resultados son la etapa preliminar de la minería de datos, y pertenecen a todo el proceso de descubrimiento de patrones como pasos importantes.

Por otro lado, el análisis exploratorio de datos nos ayuda a comprender cada variable de una manera más específica en función de todo el conjunto de datos por lo cual es el paso en donde se decide el tratamiento de cada variable siendo que se pueden considerar para ser normalizadas, imputadas o transformadas para que puedan otorgar mayor valor al orientarlos a modelos predictivos con inteligencia artificial.

La aplicación del análisis exploratorio de datos menciona Buzai & Baxendale (2009) que permite descubrir errores en la codificación de los datos, determinar los casos anómalos (outliers) y la posibilidad de comprobar supuestos necesarios para la aplicación de la mayoría de las pruebas estadísticas. Además de ello permite identificar variables potenciales o relevantes en un estudio al conocer cómo se relaciona cada variable con el resto de las variables del conjunto de datos por medio de las correlaciones.

2.3.1.3.2 MINERÍA DE DATOS

La minería de datos al formar parte del análisis de datos encargándose de la búsqueda de patrones existentes entre los datos, relaciones que podrían existir entre los mismos datos de tal manera que se pueda clasificar mejor la información esto mediante la implementación de algoritmos en grandes volúmenes de datos (García et al., 2018) afirma que la minería de datos consiste en la búsqueda de los patrones de interés en una determinada forma de representación o sobre un conjunto de representaciones, utilizando para ello métodos de clasificación, reglas o árboles, regresión, agrupación, etc.

De igual manera Llaugel & González (2016) menciona que algunas de las aplicaciones de la minería se centran en detectar patrones de comportamiento de individuos o entidades de datos, como son los agrupamientos de registros de datos por ciertas similitudes de sus atributos (análisis

de clúster). De lo cual al aplicar algoritmos de agrupamiento provee una visión más clara de los datos y como se puede obtener una solución especializada orientada a las necesidades de acuerdo con los datos sin generalizar.

2.3.1.4 LA INTELIGENCIA ARTIFICIAL

La inteligencia artificial (IA) se refiere a sistemas o máquinas que imitan la inteligencia humana para realizar tareas y pueden mejorar iterativamente a partir de la información que recopilan. ¿Qué es la inteligencia artificial? (s. f.).

La inteligencia artificial es aplicable en múltiples áreas como en los negocios, salud, educación, finanzas entre otras de tal manera que ayudan a mejorar procesos o sirven de apoyo para tomar de decisiones como por ejemplo hoy en día se encuentran los chat bots que son capaces de dar respuestas personalizadas a un usuario en función de cada negocio, por otro lado aplicaciones con inteligencia artificial capaces de generar imágenes en base a texto o hasta inteligencia artificial aplicada para la autonomía de un vehículo.

2.3.1.4.1 APRENDIZAJE AUTOMÁTICO

El aprendizaje automático o también conocido como machine learning es uno de los enfoques principales de la inteligencia artificial, tratándose de un aspecto de la informática en el que las computadoras o las máquinas tienen la capacidad de aprender sin estar programados para ello (Rouhiainen, 2018). Siendo así el aprendizaje automático se basa en la aplicación de algoritmos encargados de realizar ese aprendizaje simulando el pensamiento analítico humano y haciendo eficiente el procesamiento de datos y la facilidad para la toma de decisiones.

El aprendizaje automático se basa en el entrenamiento con datos históricos o bajo demanda de manera que busca que las máquinas entiendan como realizar un proceso tal como lo hacen los humanos. El aprendizaje automático utiliza algoritmos para identificar patrones en los datos, y esos patrones luego se usan para crear un modelo de datos que puede hacer predicciones. (¿Qué es la inteligencia artificial? | Microsoft Azure, s. f.)

2.3.1.4.1.1 TIPOS DE APRENDIZAJE AUTOMÁTICO

2.3.1.4.1.1.1 ALGORITMOS SUPERVISADOS

Predicen un dato o un conjunto completo desconocido a priori, a partir de otros conocidos y de la historia de una entidad registrada en sus atributos explicativos y de la denominada variable respuesta (Llaugel & González, 2016). En el estudio de la deserción

estudiantil al aplicar algoritmos supervisado se tiene como objetivo dar respuesta a la variable de deserción especificando si un estudiante puede ser o no un desertor.

2.3.1.4.1.1.2 ALGORITMOS NO SUPERVISADOS

Este tipo de algoritmos se encargan de identificar patrones y comportamientos específicos en los datos por lo cual son muy útiles al aplicar técnicas de minería de datos para ayudar a comprender la existencia de grupos característicos en un conjunto de datos, en función del comportamiento o patrones y similitudes de cada uno de los campos en un conjunto de datos, esto mediante técnicas de agrupaciones, por ejemplo, algoritmos como K-Means. Dando la posibilidad de clasificar en grupos en base a el patrón de los datos.

2.3.1.4.2 TÉCNICAS DE APRENDIZAJE

Hoy en día, hay una gran variedad de algoritmos que han sido aplicados en varios problemas del mundo real con alta precisión, estos algoritmos incluyen árboles de decisión, basados en regresiones, máquinas de soporte vectorial, redes neuronales artificiales, aprendizaje bayesiano, métodos basados en instancias, métodos de kernel, entre otros (Urbina et al., 2020). Sin embargo, se cuentan con una variedad más amplia de algoritmos para poder identificar patrones, comportamientos, grupos y predicciones los cuales están categorizados según el enfoque del problema a solucionar ya sean por clasificación, regresión o agrupación.

2.3.1.4.2.1 CLASIFICACIÓN

Los algoritmos de clasificación buscan predecir el valor de una clase o variable de respuesta en los nuevos registros evaluados en base un entrenamiento con datos históricos, por ejemplo, clasificar si un estudiante en particular puede ser desertor o no en función de datos históricos de estudiantes desertores.

Dentro de los algoritmos de clasificación más utilizados se encuentran las regresiones logísticas, arboles de decisión, máquinas de soporte vectorial entre otros.

2.3.1.4.2.2 REGRESIÓN

Los algoritmos de regresión buscan predecir valores numéricos calculándolo en función de la relación de las variables y su comportamiento histórico como por ejemplo el cálculo de ventas, utilidades o el valor de vida de un cliente para una empresa en un periodo desconocido futuro.

Dentro de los algoritmos de regresión se encuentra la regresión lineal, regresión múltiple, regresión polinomial, arboles de decisión.

2.3.1.4.2.3 AGRUPACIÓN

Los algoritmos de agrupación tienen como objetivo la identificación de grupos, subgrupos, comportamientos y patrones dentro de los conjuntos de datos de tal manera que puedan dar una respuesta clasificando los datos en función de las categorías establecidas.

Dentro de los algoritmos de agrupación se encuentran la agrupación por clúster con K-Means y K-Medoids

2.3.1.5 INTERPRETACIÓN DE UN MODELO ESTADÍSTICO

Dentro de los indicadores más utilizados según Palacios (2023) se encuentra la exactitud, la precisión, la sensibilidad, especificidad, tasa de error, puntuación F1-score. En problemas de clasificación encontramos estos indicadores de eficiencia del modelo antes mencionados derivados mediante la denominada matriz de confusión que tiene como objetivo representar la salida proporcionada por un modelo en función de true positives, false positives, true negatives y false negatives.

El medir la eficiencia de un modelo dependerá del contexto que se desea conocer de la variable de respuesta sin embargo gracias a los indicadores mencionado anteriormente se puede determinar si los resultados de un modelo predictivo son favorables, siendo como tal los indicadores más útiles para la aprobación de un modelo predictivo es la exactitud y la precisión, no obstante, siempre se debe realizar una revisión general del resto de los indicadores.

2.3.2 METODOLOGÍAS DESARROLLADAS POR OTROS INVESTIGADORES

En general, los estudios descriptivos o de tipo correlacional causal en los que se utilizan modelos predictivos, utilizan técnicas de análisis de datos y aprendizaje automático, con el propósito de identificar patrones, factores y variables que influyen en la deserción y que ayuden a predecir la probabilidad que un estudiante abandone sus estudios, asimismo, utilizan enfoques cuantitativos o mixtos, todo depende de la orientación o el objetivo de la investigación, se basan en procedimientos similares para la recolección de datos, aplicación de minería de datos e inteligencia artificial (modelos de árbol de decisión, regresión logística, KNN y redes neurales), así como en las pruebas y verificación e interpretación de resultados.

Un estudio de Calderón (2020), utilizó un modelo predictivo para determinar la deserción de estudiantes en la Universidad Tecnológica del Perú, en el que se utilizó una metodología descriptiva y correlacional para conocer la relación entre las variables relevantes, con el objetivo de describir el compartimento, características y rasgos importantes relacionados a la deserción estudiantil. Hay investigaciones que se basan en características y grupos específicos tal es el caso de Canales & De los Ríos (2009), en donde exploran la retención y permanencia de estudiantes vulnerables, desde un enfoque cuantitativo de tipo descriptivo, para la muestra de su estudio seleccionaron 58 estudiantes universitarios con características específicas catalogadas como vulnerables como; pertenecer a grupos étnicos minoritarios, ser jefe de hogar, tener hijos, laborar más de 20 horas a la semana y pertenecer a los dos primeros quintiles de ingreso.

Por su parte, Vásconez Altamirano (2023), realizó un estudio de implementación de un modelo de predicción de deserción escolar en los estudiantes de la unidad educativa Los Andes a consecuencia del Covid 19 y es un estudio en que se utilizó la metodología cuantitativa, cuyo objetivo es conocer y estimar la cantidad de estudiantes que desertaron de institución educativa Los Andes a consecuencia de la pandemia.

Otros estudios enfocados en la minería de datos y aprendizaje automático orientados a la creación de modelo predictivos y búsqueda de patrones, como el caso del estudio de Solis et al. (2018), en que analizó el comportamiento de cuatro algoritmos de aprendizaje automático que fueron; bosque aleatorio, redes neuronales, máquinas de vectores de soporte y regresión logística y utilizó diferentes enfoques y perspectivas para la predicción de la deserción de estudiantes universitarios.

Los autores Valero, Vargas, Salvador, et al. (2014), quienes utilizaron algoritmos como árboles de decisión y KNN vecinos, lograron identificar que la deserción en la Universidad Tecnológica de Izúcar de Matamoros se debe a tres factores principales; en cuanto a la variable edad se determinó como un factor importante ya que se relaciona la madurez y perspectiva de los estudiantes, a nivel financiero los ingresos familiares fueron otra variable relevante, puesto que los estudiantes menores a 18 años no tienen independencia financiera ya que dependen de sus padres para el pago de mensualidades, a nivel académico, el nivel de inglés es muy bajo y afecta el rendimiento de los estudiantes, especialmente entre los mayores de 18 años, el proceso para realizar el estudio fue recopilar los datos, selección, limpieza y transformación de los mismos, aplicación de minería de datos, pruebas y verificación e interpretación de resultados.

Un estudio similar es el “modelo predictivo de deserción estudiantil basado en arboles de decisión”, se basa en la creación de un modelo para predecir la probabilidad de deserción en el período académico, la metodología empleada fue Knowledge Discovery in Database, que consta de 5 pasos: selección, procesamiento, transformación, minería de datos y evaluación del modelo. (Cuji, Gavilanes & Sánchez, 2017)

Barramuño, Meza & Gálvez (2021) utiliza un método basado en aprendizaje automático supervisado para predecir la deserción de los estudiantes de la carrera de fisioterapia y se utilizó una muestra de 336 estudiantes universitarios, en donde se tomaron en cuenta cinco variables cuantitativas y 11 cualitativas, relacionadas a la deserción de estudiantes universitarios, con esta base de datos histórica, se probaron 23 clasificadores basados en aprendizaje automático supervisado.

2.3.3 INSTRUMENTOS UTILIZADOS

Los investigadores que realizan estudios desde la perspectiva cuantitativa, en donde se emplean diferentes tipos de modelos predictivos basados en aprendizaje automático, entrenan su modelo utilizando datos de primera mano como bases de datos históricas de estudiantes y seleccionan diferentes tipos de variables numéricas y categóricas, otros estudios se apoyan de diferentes recursos e instrumentos apoyándose de metodologías mixtas en las que se aplican diferentes fuentes y técnicas de selección de datos.

En la investigación de Sifuentes (2018) se utilizó un modelo para predecir la deserción estudiantil en una universidad privada peruana, es un estudio de tipo descriptivo y corte transversal y como instrumento utilizó datos académicos históricos con una muestra de 4,478 estudiantes matriculados en el periodo 2016-I (2,731 hombre y el resto mujeres con una edad promedio de 23 años y para que la muestra fuera significativa se incluyen estudiantes de distintas carreras, se elaboraron siete modelos de predicción y se consideraron como críticos los estudiantes con altos índices de reprobación, al finalizar el semestre se logró reducir los índices de reprobación en los estudiantes que fueron parte del estudio en un 25% a 40%, en comparación con los años anteriores en los que no se utilizaron modelos predictivos, este estudio tuvo un alto impacto en el nivel académico de los estudiantes y consecutivamente un efecto positivo en la mejora de calidad educativa.

Por su parte, un estudio en el que se empleó un modelo de clasificación basado en árboles de decisión, el cuál constaba de varios parámetros optimizados para la predicción de la deserción de estudiantes universitarios, utilizó una base de datos históricos de estudiantes y en la muestra se analizaron 5,288 casos de estudiantes de la universidad pública chilena, usando la técnica CBAD para mejorar la predicción y se utilizó el software RapidMiner. (Ramírez & Grandón, 2018).

Ahora bien, el estudio de Portillo (2023) el cual tiene un alcance descriptivo y de diseño no experimental de corte transversal, con el objetivo de identificar factores de deserción, utilizó otros instrumentos para el análisis de información, basados en el levantamiento de datos por medio de encuestas semiestructura y el análisis documental.

El estudio de Castillo (2022) basado en perfiles estudiantiles socioeconómico y académico de ingreso y vivencias académicas implicadas en el abandono estudiantil en Honduras el cual tiene como objetivo analizar las variables predictoras de deserción estudiantil universitaria en Honduras, por medio de un cuestionario que indaga variables socioeconómicas y académicas se tomó una muestra de 653 estudiantes de la Universidad Nacional Autónoma de Honduras (UNAH) matriculados en el primer período académico del año 2022., os participantes fueron conformados en su mayoría por estudiantes mujeres (n= 431, 66.6%) y se evidenció una edad promedio de 19 años, con una desviación estándar de 2.58 años y se contó con estudiantes universitarios de todas las áreas del conocimiento que oferta la UNAH.

2.4 MARCO LEGAL

Dentro de las regulaciones en Honduras aplicadas a la educación superior se encuentran;

- Ley de Educación Superior, 14 de septiembre de 1989
- Normas Académicas de la Educación Superior, se emiten en cumplimiento del Artículo 7 de la Ley de Educación Superior, 6 de noviembre de 1992

CAPITULO III. MARCO METODOLÓGICO

Para llevar a cabo la investigación se implementarán modelos predictivos mediante algoritmos de aprendizaje en la Universidad Tecnológica Centroamericana (UNITEC) en busca de dar respuesta a la deserción estudiantil en donde se realizará la investigación con enfoque cuantitativo de tipo correlacional.

3.1 CONGRUENCIA METODOLÓGICA

3.1.1 MATRIZ METODOLÓGICA

Tabla 1. Matriz Metodológica

Implementación de modelo predictivo para identificar deserción estudiantil en Unitec					
Formulación del Problema	Objetivo General	Objetivos específicos	Preguntas de Investigación	Variable Dependiente	Variabes Independientes

<p>Los elevados índices de deserción estudiantil tanto en las universidades públicas y privadas generan pérdidas económicas, así mismo, tiene impactos negativos en la calidad educativa, posicionamiento o prestigio de la universidad, enfoque de recursos y en el seguimiento estudiantil.</p>	<p>Identificar de manera temprana la deserción estudiantil en un periodo académico en la</p>	<p>Identificar los principales factores que causan la deserción de un estudiante.</p>	<p>¿Cuáles son los factores más relevantes que ocasionan la deserción estudiantil?</p>	<p>Deserción</p>	<p>Datos demográficos: Género, edad, ciudad y estado civil</p>
	<p>Universidad Tecnológica Centroamericana (UNITEC) en Tegucigalpa y San Pedro Sula,</p>	<p>Conocer el comportamiento de los estudiantes desertores en UNITEC.</p>	<p>¿Cuál es el comportamiento de los estudiantes desertores en UNITEC?</p>		<p>Datos financieros: Cuotas pendientes, número de cuotas pagadas, descuento, beca, permiso de matrícula.</p>
	<p>mejorando los índices de retención y mitigando las pérdidas económicas para la institución, mediante la aplicación de modelos predictivos.</p>	<p>Desarrollar una propuesta de diseño e implementación de un modelo predictivo de inteligencia artificial mediante algoritmos de aprendizaje basados en la deserción académica para mejorar la retención estudiantil.</p>	<p>¿Al implementar un modelo predictivo de retención se reducirán los índices de deserción estudiantil en la Universidad Tecnológica Centroamericana (UNITEC)?</p>		<p>Datos académicos: tipo de estudiante, facultad, jornada, notas promedio, uvs matriculadas, inasistencia, índice académico, carrera, uvs aprobadas, uvs reprobadas, uvs retiradas,</p>

					materias matriculadas, materias retiradas, materias reprobadas, materias aprobadas.
--	--	--	--	--	--

Fuente: (Elaboración Propia, 2023)

3.2.3 ESQUEMA DE VARIABLES DE ESTUDIO

Tabla 2. Esquema de Variables de Estudio

Variable	Tipo de Variable	Indicadores
Deserción Estudiantil	Dependiente	Índice de deserción
Demográficas	Independiente	Género, edad, ciudad y estado civil

Académicas	Independiente	Tipo de estudiante, facultad, jornada, notas promedio, uvs matriculadas, inasistencia, índice académico, carrera, uvs aprobadas, uvs reprobadas, uvs retiradas, materias matriculadas, materias retiradas, materias reprobadas, materias aprobadas.
Financieras	Independiente	Cuotas pendientes, número de cuotas pagadas, descuento, beca, permiso de matrícula.

Fuente: (Elaboración Propia, 2023)

3.1.3 OPERACIONALIZACIÓN DE LAS VARIABLES

Tabla 3. Operacionalización de las variables

Variable	Definición		Dimensiones	Indicador
	Conceptual	Operacional		
Deserción	Es el abandono prematuro de las obligaciones académicas que puede tener un estudiante en una institución en particular o totalmente del sistema educativo.	Es un indicador que mide el porcentaje de estudiantes que abandonan sus estudios	Estudiantes no matriculados	Índice de deserción

Datos Demográficos	Son datos que a partir de la observación dan cuenta de características específicas de una persona como ser biológicas, sociales y culturales (CEPAL, 2014)	Ayudan en conocer características de género, edad, ciudad y estado civil	Género, edad, ciudad y estado civil	Rango de edades con mayor deserción
				% Hombres o mujeres que desertan
				Ciudad con más deserción
				% Estado civil
Datos Financieros	Es el conjunto de datos cuya magnitud es medida y expresada en términos de dinero que describen la situación financiera y otros aspectos relacionados con la obtención y el uso del dinero (Gómez, C et al., 2017)	Información financiera del estudiante de saldos y beneficios económicos	Cuotas pagadas, cuotas pendientes, descuento, beca y permiso de matrícula	Cantidad de cuotas pagadas del estudiante
				Cantidad de Cuotas pendientes de pago
				% becas
				% de descuento
				% permisos de matrícula con saldo
Datos Académicos	Son medidas de las capacidades del alumno, que expresa lo que ha aprendido a lo largo del proceso formativo (Pérez Porto, J., Gardey, 2008)	Desempeño del estudiante en sus clases	Tipo de estudiante, facultad, jornada, notas promedio, uvs matriculadas, inasistencia, índice académico, carrera, uvs aprobadas, uvs reprobadas, uvs retiradas, materias matriculadas,	Notas promedio
				Tipo de estudiante
				Porcentaje de inasistencias
				Índice del período
				Número de clases matriculadas
				Número de clases reprobadas

			materias retiradas, materias reprobadas, materias aprobadas.	Número de clases retiradas
				Número de clases aprobadas
				UVS matriculadas, reprobados y retirados
				% Facultad con mayor deserción
				% Jornada con más deserción

Fuente: (Elaboración Propia, 2023)

3.1.4 HIPÓTESIS

H1: La deserción en estudiantes de pregrado de Unitec está asociada a factores demográficos como género, edad y estado civil.

H2: La falta de beneficios económicos como becas, descuentos junto a las dificultades por el pago de mensualidades y matrícula puede llevar a que los estudiantes de pregrado de Unitec desertar.

H3: El rendimiento académico y las inasistencias de los estudiantes de pregrado de Unitec están correlacionadas con los índices de deserción.

3.2 ENFOQUE Y MÉTODOS

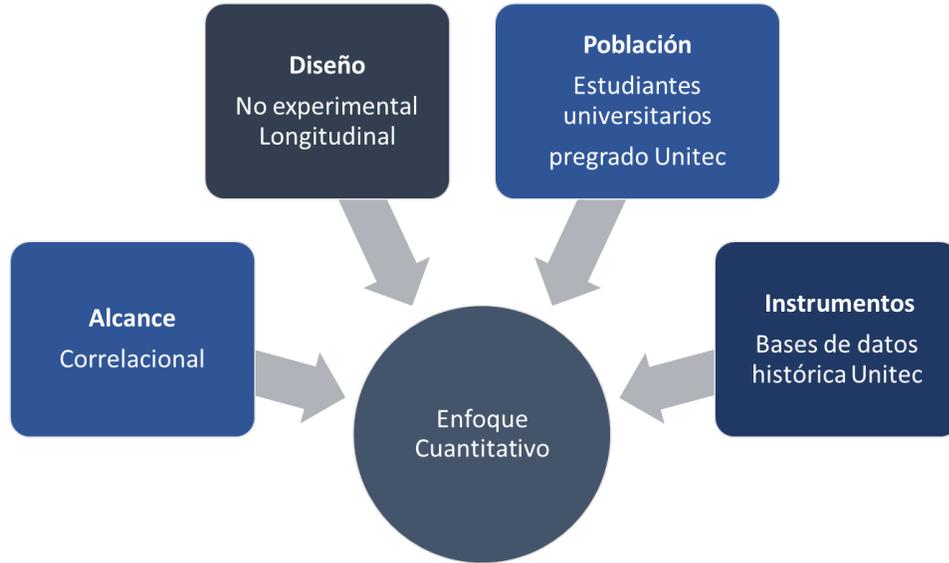


Figura 1. Diagrama de Enfoque y Métodos.

Fuente: (Elaboración Propia, 2023)

3.3 DISEÑO DE LA INVESTIGACIÓN

El diseño la investigación es de tipo no experimental de corte longitudinal, ya que el estudio cuenta con un enfoque cuantitativo de tipo correlacional y se basa en la implementación de la solución predictiva en un periodo o ciclo académico en específico, en función del comportamiento a lo largo de ciclos académicos pasados.

3.3.1 POBLACIÓN

Para este estudio se tomará en cuenta datos históricos de todos los estudiantes de Pregrado de las ciudades de Tegucigalpa y San Pedro Sula proporcionados por Unitec y contiene datos demográficos, financieros y académicos con los que se deberá realizar análisis previos al entrenamiento de los modelos predictivos.

3.3.2 MUESTRA

El estudio no cuenta con una muestra específica, por lo tanto, se considera el universo completo de la población de estudiantes de pregrado de Unitec, sin embargo, para el entrenamiento de los modelos predictivos es necesario contar con un conjunto de entrenamiento y pruebas, por

lo cual la población se debe de dividir en una proporción aproximada del setenta y cinco por ciento (75%) o más para datos de entrenamiento y el resto para realizar pruebas de implementación para la evaluación de indicadores de eficiencia de los modelos predictivos.

3.3.3 TÉCNICAS DE MUESTREO

Para el estudio no es necesario aplicar técnicas de muestreo de datos dado que se considera toda la población.

3.4 TÉCNICAS, INSTRUMENTOS Y PROCEDIMIENTOS APLICADOS

3.4.1 TÉCNICA

En este caso el estudio se realizará mediante la observación de una base de datos proporcionada por que contienen datos históricos desde el primer periodo del 2020 al primer periodo del 2023 de los estudiantes de pregrado de Unitec, para poder generar indicadores para una mejor comprensión del comportamiento de las variables frente a la deserción estudiantil.

3.4.2 INSTRUMENTOS

Como instrumento se hará uso de una base de datos así mismo para la identificación de variables relevantes para el estudio y aplicación en los modelos predictivos para la deserción estudiantil.

3.4.3 PROCEDIMIENTOS APLICADOS

Para llevar a cabo el proyecto se definen los procedimientos aplicados en función de las siguientes fases conforme a la metodología CRISP-DM:

3.4.3.1 COMPRENSIÓN DEL NEGOCIO

Este estudio se centra en identificar de manera temprana los estudiantes de pregrado con probabilidad de deserción en Unitec, mediante la implementación de un modelo predictivo de inteligencia artificial mediante algoritmos de aprendizaje y que ayude a reducir la tasa de deserción, mejorar la retención estudiantil, apoyo en la toma de decisiones y estrategias en materia de retención, enfoque de recursos financieros y esfuerzos administrativos en los estudiantes con mayor riesgo de deserción.

Unitec proporcionó las bases de datos históricas de los periodos académicos comprendidos entre el 2020-01 al 2023-01, la cual está conformada con datos demográficos, académicos y

financieros de los estudiantes de pregrado de las ciudades de Tegucigalpa y San Pedro Sul, los cuales servirán para el entrenamiento e implementación del modelo.

3.4.3.2 COMPRESIÓN DE LOS DATOS

Los datos serán analizados con el apoyo de herramientas de visualización como Power BI y Microsoft Excel, el cual será de apoyo para el análisis exploratorio de datos para entender el comportamiento en función de la relevancia de cada una de las variables conforme a la deserción estudiantil y poder determinar la relevancia de cada variable.

Para describir el comportamiento de las variables categóricas se basarán en el análisis de frecuencias mediante distribuciones de gráficos de barra y para las variables numéricas se utilizará elementos visuales tipo box plot, gráficos de líneas a través del tiempo y gráficos de dispersión entre variables de lo cual se mantendrá un tablero de hallazgos del análisis exploratorio de datos.

Por otro lado, en función de comprender mejor los datos se hará uso de tablas de correlación para las variables continuas y tablas de frecuencia para variables categóricas para poder identificar variables relevantes para el estudio.

3.4.3.3 PREPARACIÓN DE LOS DATOS

Posterior a un análisis exploratorio de datos se descubren anomalías en los datos, por lo cual es necesario llevar a cabo un proceso de limpieza y transformación de datos que tiene como objetivo dar congruencia y validez a cada variable del conjunto de datos. De igual manera la transformación de datos da lugar a la creación de nuevas variables en función de las existentes para que puedan proporcionar valor al estudio.

En el caso de encontrar valores atípicos, valores nulos o vacíos será necesario realizar imputaciones de las variables de tal manera que el tratamiento de imputación para variables numéricas se determina en función del valor promedio o mediana de los valores de la variable a imputar, por otro lado, para las variables categóricas el valor a imputar se determinará en función del valor con mayor frecuencia de dicha variable.

Previo al modelado de los datos es necesario preprocesar las variables categóricas para obtener las observaciones de cada dimensión de tal manera que sea comprensible para entrenar un

modelo, así mismo aplicar técnicas de normalización de datos para las variables numéricas para obtener un mejor resultado en el entrenamiento de un modelo predictivo.

3.4.3.4 MODELADO

Luego de los procesos de limpieza y transformaciones necesarias en los datos, es posible realizar entrenamientos de modelos estadísticos para poder predecir la deserción estudiantil en base a comportamientos de datos históricos, de lo cual se hará uso librerías especializadas como scikit-learn a través del lenguaje de programación Python para poder entrenar un modelo capaz de proporcionar una respuesta al problema de la deserción estudiantil, en donde para ello se hará uso de algoritmos de clasificación capaces de determinar si un estudiante puede desertar o no de un periodo académico a otro.

3.4.3.5 EVALUACIÓN

Se llevará a cabo procesos de evaluación en función de los modelos entrenados en busca de identificar que modelo se ajusta mejor para proporcionar una respuesta a la deserción estudiantil, de tal manera que se evaluarán indicadores importantes en función de matrices de confusión principalmente como la precisión y exactitud, así mismo a través de curvas ROC.

3.4.3.6 DESPLIEGUE

Por último, se llevará a cabo la implementación de la solución predictiva con el apoyo del modelo de entrenamiento que mejor se ajusta a dar una respuesta y que tenga indicadores de eficiencia aceptables, del cual se generan resultados para identificar la deserción en los estudiantes a través de variables de respuesta como el valor de la predicción, la probabilidad de deserción y los factores más relevantes causantes de la deserción en cada estudiante, además de crear otras variables que puedan generar valor a la toma de decisiones como una clasificación basada en la probabilidad de deserción.

Como resultado de implementar la solución predictiva se genera conocimiento para la institución con respecto a identificar la deserción estudiantil de manera temprana al conocer el grupo de estudiantes clasificados según su riesgo de desertar y los factores que influyen en la deserción de tal manera que mejora el enfoque de recursos.

3.5 FUENTES DE INFORMACIÓN

3.5.1 FUENTES PRIMARIAS

Como fuente primaria para el estudio se considera principalmente las bases de datos de estudiantes universitarios de pregrado de Unitec, ya que los mismos contienen las variables relevantes y precisas para el entrenamiento del modelo predictivo, además se tendrá mayor control sobre la calidad y la confiabilidad de los datos.

3.5.2 FUENTES SECUNDARIAS

Por la naturaleza y problema de estudio no se incorporarán datos externos de instituciones o personas, ya que en el estudio se necesitan datos específicos que no están disponibles en fuentes secundarias, por lo que la recopilación de datos será por medio de fuentes primarias proporcionados directamente de Unitec.

CAPITULO IV. RESULTADOS Y ANÁLISIS

4.1 ANÁLISIS EXPLORATORIO DE DATOS (EDA)

El objetivo del análisis exploratorio de datos es poder conocer comportamientos de variables demográficas, académicas y financieras de los estudiantes en busca de entender la deserción estudiantil que se presenta cuando un estudiante se ausenta de un periodo académico a otro. Por lo cual se utiliza para el análisis la base de datos proporcionada por la Universidad Tecnológica Centroamericana (UNITEC), la cual está conformada de 96,103 observaciones de estudiantes de pregrado en periodos académicos desde el año 2020 al primer periodo del año 2023 con un total de 13 periodos académicos.

A lo largo de los periodos académicos UNITEC percibe una matrícula promedio de 7,392 estudiantes de pregrado para los campus de San Pedro Sula y Tegucigalpa de los cuales se presenta un índice de deserción estudiantil promedio del 12% de los estudiantes en donde estos se encuentran caracterizados como aquellos estudiantes que no se encuentran presentes de un periodo académico a otro.



Figura 2. Población y Desertores de Pregrado.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Del grafico anterior se puede notar un comportamiento atípico en los periodos académicos del año 2020, el cual presento el mayor índice de deserción estudiantil entre el primer y segundo periodo académico del 2020, la razón de este comportamiento fue por el inicio de la pandemia del Covid-19 y la adaptación de la educación mediante tele docencia. Dado este comportamiento en los datos todos los periodos académicos del año 2020 no se tomarán en cuenta para realizar soluciones predictoras, ya que pueden conllevar a comportamientos o patrones inusuales.

Al analizar la deserción estudiantil desde otra perspectiva se analizan variables categóricas como el género el cual se observa con una distribución uniforme entre la población que son mujeres contra hombres, además de ello mantiene un comportamiento uniforme en relación con la deserción por género en donde al menos el 49 % de estudiantes desertores son mujeres y el resto son hombres.

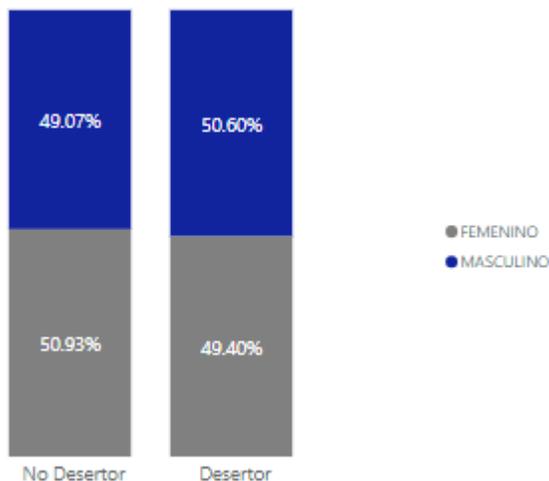


Figura 3. Género de Estudiantes No Desertores y Desertores de Pregrado.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Con respecto a la edad de los estudiantes al observar el comportamiento a través de gráficos de caja se puede notar que los estudiantes desertores presentan una edad promedio de 29 años aproximadamente, cabe resaltar que la fuente de datos en bruto presenta valores atípicos como es notables en el gráfico los cuales serán tratados mediante imputaciones de datos por medio de medianas, el 96% de los estudiantes de pregrado que desertan de sus estudios son solteros, por su parte un 4% de las observaciones aparecen como vacías, por lo que se imputará la variable y se reemplazarán los valores vacíos por la moda del estado civil que en este caso es “soltero”, ya que es el valor categórico con mayor frecuencia en la base de datos. Asimismo, el 99% de los estudiantes desertores no trabajan.

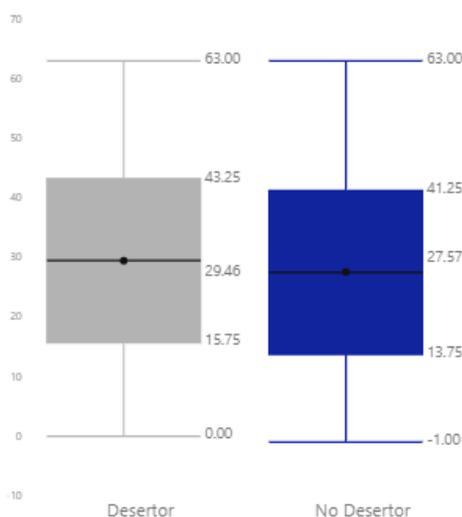


Figura 4. Gráfico de Edad de Estudiantes No Desertores y Desertores de Pregrado.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Como se refleja en el gráfico de dona el estudiante de reingreso representa la mayor parte del conjunto de datos proporcionado con un total de 93.3 %, el restante corresponde a los estudiantes de primer ingreso, siendo como tal que la deserción estudiantil se encuentra más enfocada en estudiantes de reingreso debido a la poca población categorizada como tipo de estudiantes de primer ingreso. Sin embargo, cabe resaltar que el comportamiento de los índices de deserción se mantiene entre tipos de estudiantes en un aproximado de 12%.

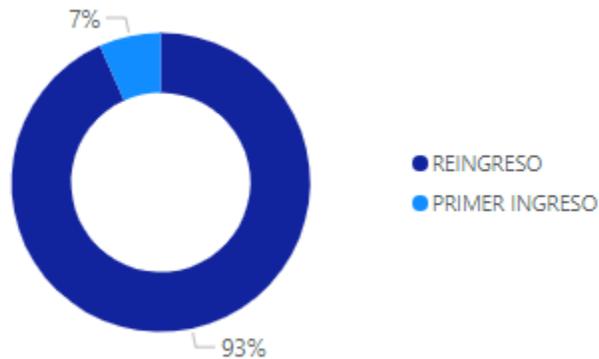


Figura 5. Tipo de Estudiante de Pregrado.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Con respecto a la parte académica los estudiantes en promedio matriculan aproximadamente 5 materias por periodo de los cuales en relación con el promedio se aprueban 4 materias, sin embargo, el comportamiento de los estudiantes clasificados como desertores se refleja de la siguiente manera, en donde se matricula en promedio 4 clases de la cuales en promedio se aprueban 3 como se puede ver en los siguientes gráficos en donde se muestra el comportamiento de estas variables.

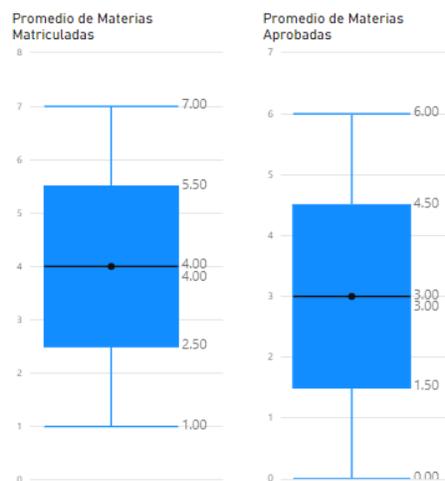


Figura 6. Gráfico de Cajas de Promedio de Materias Matriculadas y Materias Aprobadas de Estudiantes Desertores de Pregrado.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

En base al comportamiento de los datos académicos se ha notado cierta correlación entre las inasistencias que posee un estudiante en un periodo y el índice académicos resulta de ello una

correlación negativa, que se interpreta de tal manera que a mayor cantidad de inasistencias que tenga un estudiante menor será su índice académico y viceversa a menor inasistencia mejor rendimiento tendrá un estudiante.

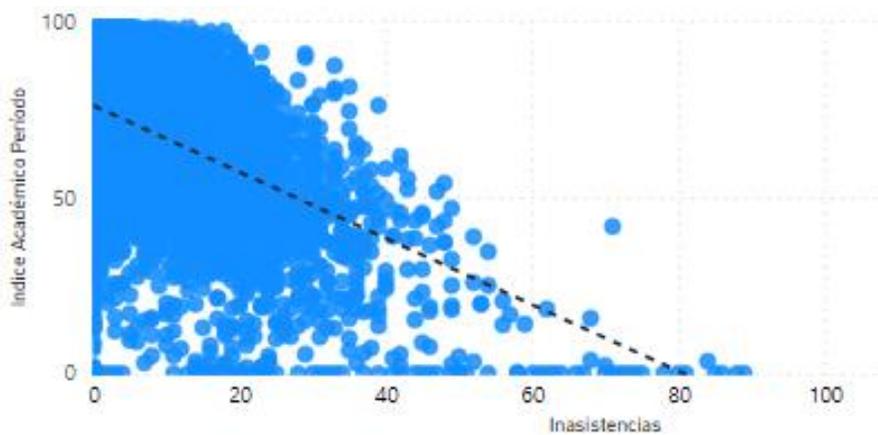


Figura 7. Relación entre Índice Académico de Período e Inasistencias de Estudiantes Pregrado.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Dentro de los aspectos financieros se resalta que del total de la población tan solo el 3.43% son estudiante becados, no obstante, al hablar de beneficios financieros la universidad proporciona descuentos en cuotas a más del 69% de los estudiantes de la población. Sin embargo, de dichos estudiantes que son desertores más del 66% poseen el beneficio financiero de descuentos.

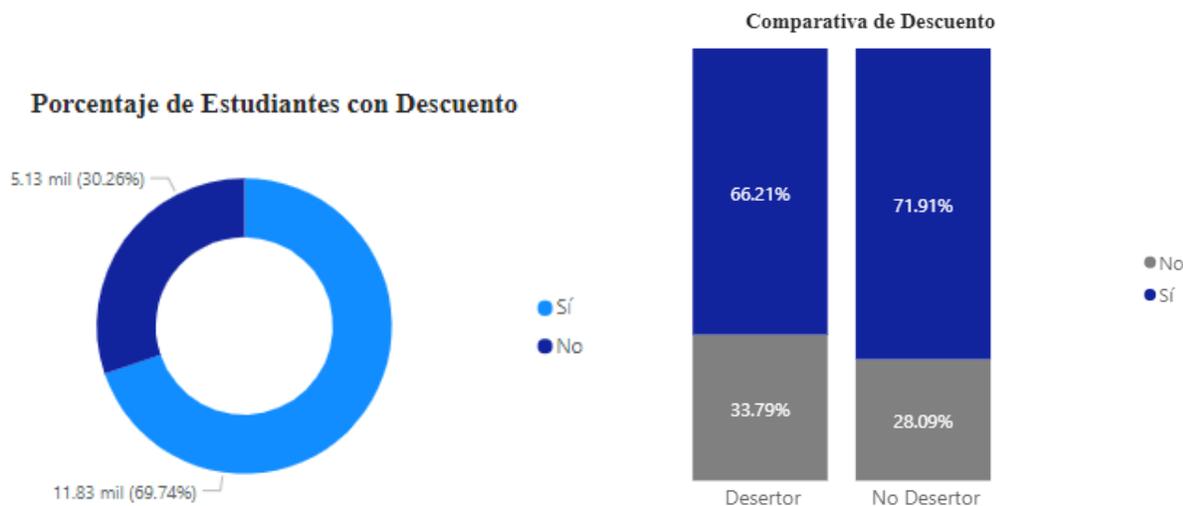


Figura 8. Porcentaje de Estudiantes de Pregrado con descuento.

Figura 9. Comparativa de Descuento de Estudiantes Desertores y No desertores.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

En promedio 2,910 estudiantes de pregrado utilizan permiso de matrícula, lo que equivale al 39% de la población de pregrado, el permiso de matrícula es un mecanismo de apoyo financiero por parte de Unitec, el cual permite que los estudiantes puedan matricularse teniendo un cierto número de cuotas pendientes de pago o realizando un abono al saldo pendiente, Unitec revisa cada caso y determina el porcentaje de abono adecuado para cada estudiante.

A partir del segundo periodo del 2020 aumentó el porcentaje de estudiantes de pregrado que solicitaron permiso de matrícula, siendo el último período del 2020 el de mayor porcentaje alcanzado el 64%, lo que equivale a que 4,501 estudiantes de pregrado necesitaron un permiso para realizar su matrícula.

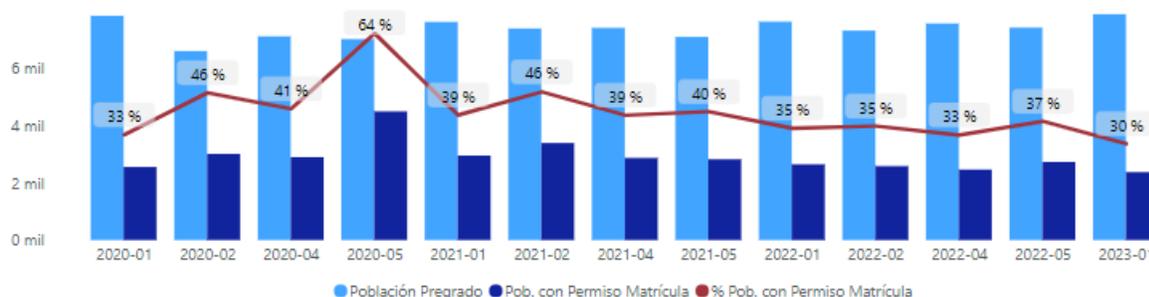


Figura 10. Porcentaje de Permisos de Matrícula de Estudiantes de Pregrado.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Ahora bien, observando el comportamiento de permisos de matrícula específicamente de los estudiantes desertores, se puede ver que el porcentaje es más alto comparado con el promedio de los estudiantes de pregrado con permiso de matrícula, ya que en promedio del 53% de los estudiantes desertores solicitan permiso para realizar su matrícula, esta tendencia se ha mantenido parte del 2021 y 2022, en los últimos tres períodos del 2020 los permisos de matrícula en estudiantes desertores aumentaron alrededor del 60%, esto asociado a la pandemia.



Figura 11. Porcentaje de Permisos de Matrícula de Estudiantes Desertores de Pregrado.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

En cuanto al comportamiento de pago de los estudiantes de pregrado, los estudiantes no desertores tienen en promedio 14 cuotas pagadas, que también es la media de cuotas pagadas de los estudiantes de pregrado, respecto al cumplimiento de pago que es un indicador que muestra la relación entre las cuotas pagadas y las cuotas del periodo, los estudiantes no desertores tienen un promedio de cumplimiento de cuotas del 54%.

Los estudiantes desertores tienen en promedio 9 cuotas pagadas, que es un promedio más bajo a la media de cuotas pagadas de los estudiantes de pregrado (14 cuotas), en cuanto al cumplimiento de cuotas tienen un 51%. Asimismo, se puede observar en el gráfico de caja que la variable cuotas pagadas presenta valores atípicos, por lo que antes del modelado se ajustará los valores mediante la normalización de datos.

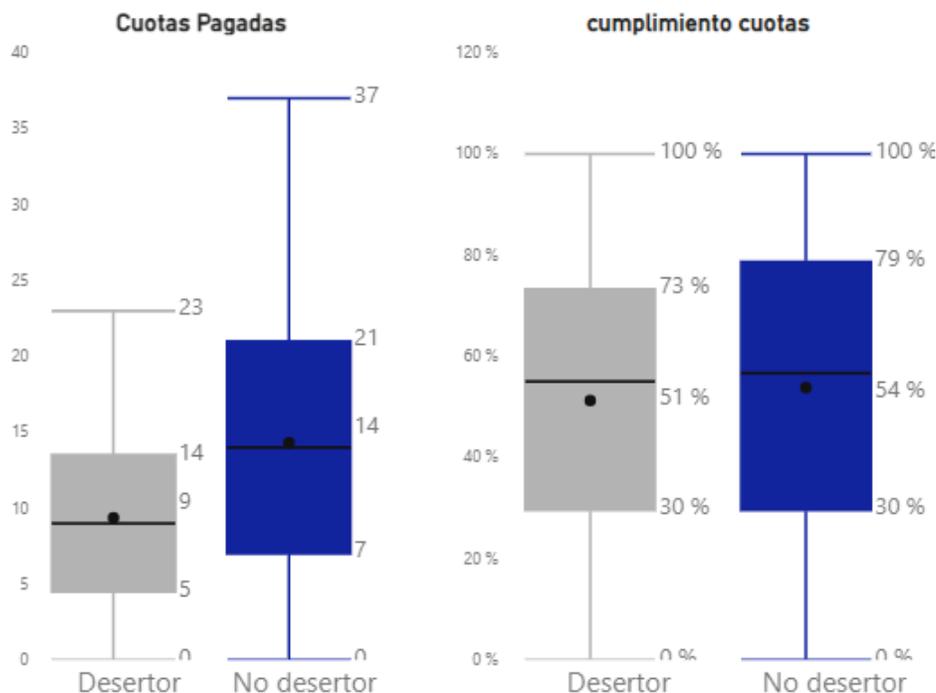


Figura 12. Promedio de Cuotas Pagadas y Cumplimiento de Cuotas de Desertores y No Desertores.

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Tabla 4. Asociación de variables categóricas contra la variable de deserción estudiantil

Variables	X ²	P
Genero	14.7169	0.0001
Trabajo	100.1777	0.0000
Estado Civil	17.6666	0.0001
Forma de Pago	2650.1381	0.0000
Beca	217.6783	0.0000
Descuento	1157.4886	0.0000
Permiso de Matricula	958.6216	0.0000
Jornada	11466.4141	0.0000
Tipo de Estudiante	0.4345	0.5098
Facultad	78.5058	0.0000

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Al implementar el estadístico de Chi cuadrado en busca de conocer la asociación que se tiene de cada variable categórica en función de la variable de deserción en donde se observa según este estadístico que dentro variables con mayor asociación contra la variable de respuesta se encuentra la jornada en la que se encuentra los estudiantes, forma de pago y descuento, por otro lado la que tiene la menor asociación es la variable de tipo de estudiante y esto se debe a como anteriormente se reflejó las distribuciones entre estudiantes de primer ingreso que son un porcentaje muy pequeño en comparación a los estudiantes de reingreso.

Tabla 5. Asociación de variables numéricas contra la variable de deserción estudiantil

VARIABLES	T	P
Cuotas del Periodo	-16.2939	0.0000
Cuotas Pendientes del Periodo	45.9944	0.0000
Cuotas Pagadas del Periodo	-33.9812	0.0000
Notas Promedio	-59.0176	0.0000
Unidades Valorativas Matriculadas	-73.9339	0.0000
Unidades Valorativas Reprobadas	19.9142	0.0000
Unidades Valorativas Retiradas	24.5998	0.0000
Cumplimiento de Cuotas	-53.5727	0.0000
Indice de Reprobación	34.7389	0.0000
Inasistencias	9.5793	0.0000
Materias Matriculadas	-111.4558	0.0000
Materias Reprobadas	22.5541	0.0000
Materias Retiradas	26.2428	0.0000

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Al implementar pruebas T de student en busca de conocer analizar la significancia estadística de cada variable numérica en función de la variable de deserción en donde se observa según este estadístico que dentro variables con mayor significancia contra la variable de respuesta se encuentran las cuotas pendientes del periodo, el índice de reprobación, unidades valorativa retiradas entre otra que pueden conllevar a dar una buena respuesta al implementar en una solución predictiva en busca de la deserción estudiantil.

Tabla 6. Correlación de variables numéricas relevantes para el estudio

	Cuotas Pendientes	Notas Promedio	Inasistencias	Materias Matriculadas	Materias Retiradas	Materias Reprobadas	Materias Aprobadas	Cumplimiento de Cuotas	Índice Aprobación	Índice Reprobación
Cuotas Pendientes	1.00	-0.14	0.10	-0.08	0.06	0.06	-0.13	-0.91	-0.14	0.08
Notas Promedio	-0.14	1.00	-0.43	0.19	-0.52	-0.40	0.62	0.16	0.87	-0.43
Inasistencias	0.10	-0.43	1.00	0.20	0.50	0.16	-0.13	-0.10	-0.39	0.10
Materias Matriculadas	-0.08	0.19	0.20	1.00	0.05	0.07	0.80	0.11	0.15	-0.07
Materias Retiradas	0.06	-0.52	0.50	0.05	1.00	0.03	-0.30	-0.06	-0.49	0.01
Materias Reprobadas	0.06	-0.40	0.16	0.07	0.03	1.00	-0.38	-0.06	-0.62	0.91
Materias Aprobadas	-0.13	0.62	-0.13	0.80	-0.30	-0.38	1.00	0.16	0.66	-0.44
Cumplimiento de Cuotas	-0.91	0.16	-0.10	0.11	-0.06	-0.06	0.16	1.00	0.16	-0.09
Índice Aprobación	-0.14	0.87	-0.39	0.15	-0.49	-0.62	0.66	0.16	1.00	-0.67
Índice Reprobación	0.08	-0.43	0.10	-0.07	0.01	0.91	-0.44	-0.09	-0.67	1.00

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

En base al cuadro de correlación anterior se puede notar que resaltar correlaciones más fuertes de variables como ser el cumplimiento de cuotas contra cuotas pendientes; Notas promedio contra índice de aprobación; Materias reprobadas contra índice de reprobación, estas correlaciones se deben dado que las variables son variables creadas a partir de otra variable base para obtener una respuesta desde otra perspectiva. Por otro lado, correlaciones como la que tiene la variable Inasistencia contra las variables notas promedio y materias retiradas resultan en dar una respuesta favorable para determinar o comprender de mejor manera el comportamiento de los datos y el análisis de la deserción estudiantil.

4.2 RESULTADOS DE MODELOS

En esta sección se muestra el resultado de cada modelo estadístico elaborado con el fin de determinar el mejor modelo que se ajuste para proveer una respuesta a la deserción estudiantil, el cual se conoce con mayor detalle en el capítulo Aplicabilidad.

Para poder conocer el comportamiento de cada modelo estadístico se tuvo que entrenar y realizar pruebas con el fin de conocer los indicadores de eficiencia del modelo ejecutado.

Tabla 7. Indicadores de eficiencia de modelos previos a implementar

Algoritmo	Precisión	Exactitud	Especificidad	F1 Score	ROC Auc
Regresión Logística	88.41%	90.00%	97.89%	88.36%	86.80%
Árbol de Decisión	87.55%	87.30%	92.51%	87.42%	70.30%
Bosques Aleatorios	90.72%	91.61%	98.16%	90.55%	89.10%

Fuente: Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Además de obtener los indicadores de eficiencia se realizó la evaluación de cada matriz de confusión con relación al resultado de cada entrenamiento, de lo cual se obtiene lo siguiente.

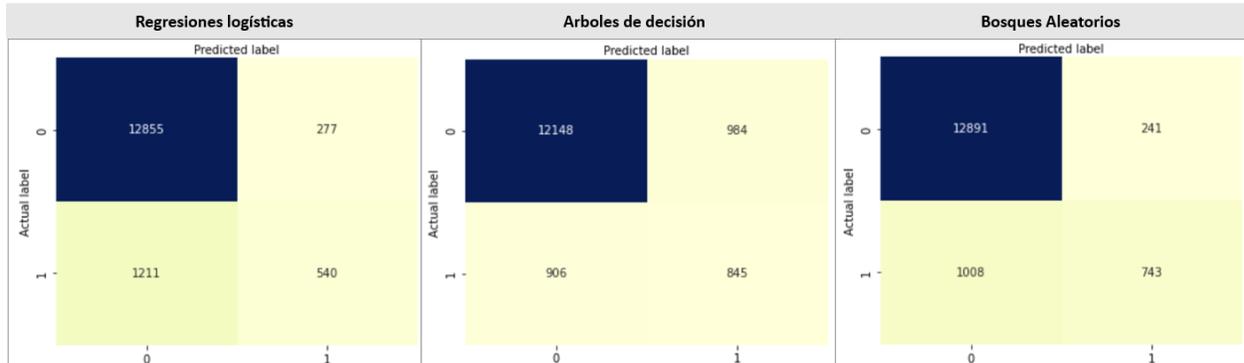


Figura 13. Matriz de confusión de algoritmos entrenados

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Por otro lado, en función de conocer la eficiencia de cada algoritmo se evalúa gráficamente a través de curvas ROC.

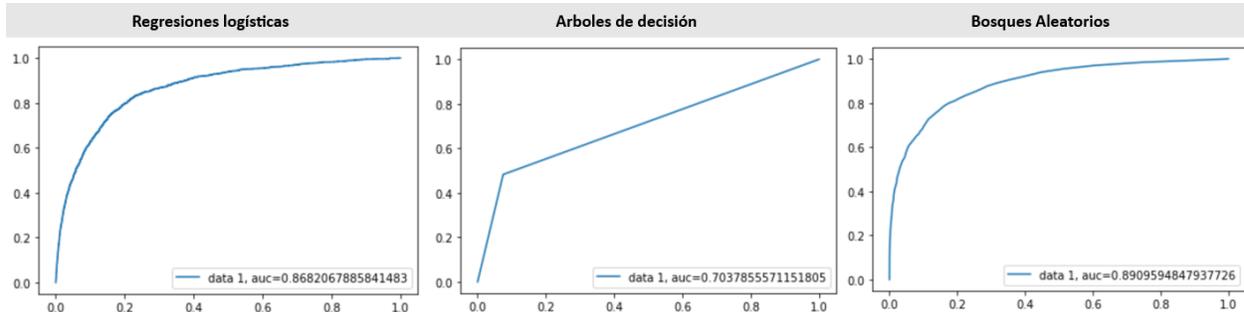


Figura 14. Curvas ROC de algoritmos entrenados

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

Los modelos seleccionados fueron los que mejor se ajustaban a dar una respuesta al problema de la deserción estudiantil en el cual los algoritmos utilizados son orientados a dar respuesta a problemas de clasificación, en donde con relación a los resultados presentados los árboles de decisión son los que menos se ajustan a dar una respuesta predictiva favorable, sin embargo, en el capítulo de aplicabilidad se dará a conocer más a detalle del modelo implementado y la razón de estas.

CAPITULO V. CONCLUSIONES

5.1 CONCLUSIONES

- Los factores que influyen en la deserción de los estudiantes en la Universidad Tecnológica Centroamericana se identificó mediante el análisis exploratorio de datos y entrenamiento de soluciones predictoras para dar una respuesta a la deserción, que dentro los aspectos más influyentes se encuentra variables académicas relacionadas con el rendimiento del estudiante, como los índices de notas bajas en un estudiantes, el aumento en la cantidad de materias reprobadas y la cantidad de materias retiradas, sin embargo no solamente el aspecto académico tiene relevancia también las variables financieras como las relacionadas con el pago de cuotas las cuales contribuyen a la deserción estudiantil.
- Conforme al comportamiento de los estudiantes que abandonan sus estudios de un periodo académico a otro, resaltan comportamientos relacionados con el rendimiento académico en donde se ha notado que los estudiantes que tienen a desertar presentan índice académico, materias aprobadas más bajas que el promedio además que se presentan mayor cantidad de inasistencias. Por otro lado, dentro de la parte financiera los estudiantes desertores presentan mayor deuda implicando un mayor número de cuotas adeudas, por otro lado, independiente del alto porcentaje de la población que poseen beneficios financieros siempre está presente el fenómeno de la deserción de un periodo académico a otro.
- En el capítulo de aplicabilidad se define la propuesta de implementación de un modelo predictivo para dar respuesta a la deserción estudiantil para estudiantes de pregrado en la Universidad Tecnológica Centroamericana. con el apoyo de información histórica proporcionada, mediante procesos de datos y algoritmos de aprendizaje entrenados en busca de dar respuesta al problema se concluye que la solución predictiva sirve como una herramienta de apoyo en donde se podrá identificar de mejor manera los estudiantes con riesgo de deserción, mejorando el enfoque de esfuerzos para la retención de estudiantes.

5.2 RECOMENDACIONES

- Siendo las variables académicas los factores más relevantes asociados a la deserción, Unitec puede aplicar este modelo predictivo que se basa en factores de rendimiento académico, inasistencias, calificaciones y otros datos relevantes, para identificar de manera temprana los estudiantes con mayor riesgo de deserción, ya que el modelo asigna una puntuación de riesgo a cada estudiante y de esta manera podrá enfocar y priorizar los recursos en los estudiantes con mayor riesgo de deserción, que son los que necesitan más apoyo, los acercamientos pueden ser personalizados desde la academia hasta el personal administrativo y se puede comenzar días antes del primer examen o en el transcurso del periodo académico. Asimismo, se muestra como referencia un diagrama de comportamientos de los estudiantes desertores basados en arboles de decisión solamente de referencia para entender los factores o variables relevantes causantes de la deserción a través del siguiente algoritmo de entrenamiento.

```
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn import tree

X_train, X_test, y_train, y_test = train_test_split(df_x,df_y, train_size=0.75,random_state=42,stratify=df_y)

DT = tree.DecisionTreeClassifier(max_leaf_nodes=7).fit(X_train, y_train)
y_pred = DT.predict(X_test)

cn=["Deserter", "Non Deserter"]
plt.figure(figsize=(40,15))
plot_tree(DT, filled=True, feature_names=df_x.columns,class_names=cn)
plt.show()
```

Figura 15. Entrenamiento con árbol de decisión

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

El diagrama de comportamiento resalta las variables más importantes causantes de la deserción para la comprensión de los estudiantes desertores en donde resaltan aspectos académicos y financieros.

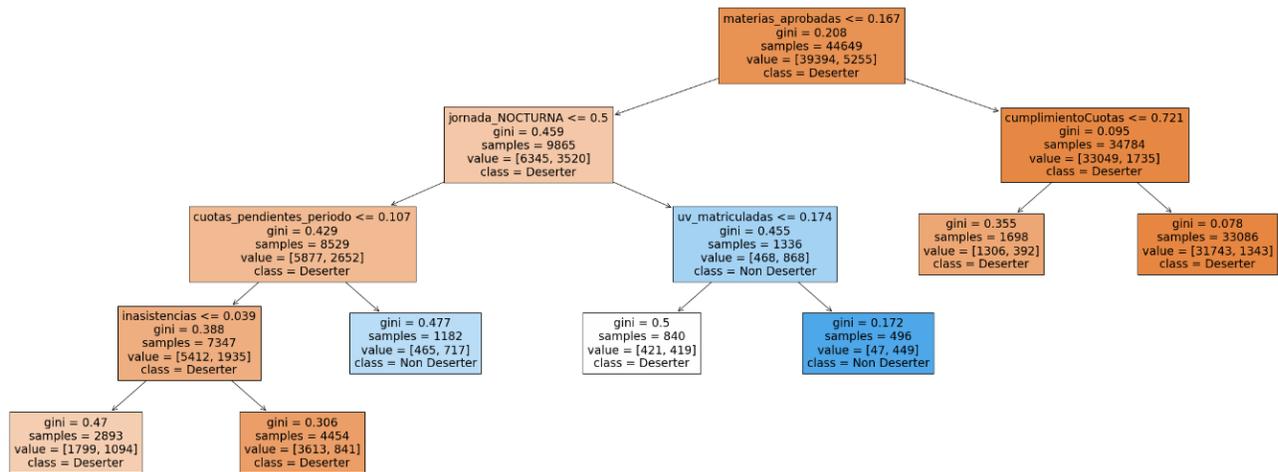


Figura 16. Diagrama de comportamiento

Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

- Monitorear constantemente el progreso de los estudiantes con riesgo de deserción, ya que puede ayudar a saber qué tipo de apoyo puede necesitar el estudiante y también se tendrá una mejor visibilidad del comportamiento y evolución de los estudiantes que desertan.
- Capacitación continua al personal involucrado en las estrategias de retención, informar acerca de la relevancia del uso de modelos predictivos orientados a identificar estudiantes con riesgo de deserción, desde el personal académico hasta el administrativo como docentes, tutores, call center, psicólogos y servicio al cliente.
- Junto con la aplicación del modelo predictivo otro método que puede ayudar a la retención estudiantil es la retroalimentación constante de los estudiantes ya sea por encuestas, comentarios en redes sociales, solicitudes, correos, es decir tener una vista 360 de las opiniones de los estudiantes, esto ayudará a la universidad a enfocar sus estrategias de retención y a la mejora continua.

CAPITULO VI. APLICABILIDAD

6.1 NOMBRE DE LA PROPUESTA

Implementación de modelo predictivo para identificar deserción estudiantil en Unitec.

6.2 JUSTIFICACIÓN DE LA PROPUESTA

Con respecto al análisis exploratorio de datos los factores más influyentes para que un estudiante deserte de un periodo a otro son las variables académicas relacionadas con el rendimiento del estudiante, índices de notas bajos en estudiantes, el aumento en la cantidad de materias reprobadas y la cantidad de materias retiradas. Sin embargo, no solamente el aspecto académico tiene relevancia, también las variables financieras como las relacionadas con el pago de cuotas las cuales contribuyen a la deserción estudiantil, esto proporcionará una visión temprana a la universidad de cuáles y cuántos son los estudiantes con mayor riesgo de deserción, apoyo en la toma de decisiones en materia de retención, optimización de recursos y satisfacción estudiantil.

6.3 ALCANCE DE LA PROPUESTA

- Desarrollar un modelo predictivo de inteligencia artificial mediante algoritmos de aprendizaje en busca de dar respuesta a la deserción estudiantil llevando a cabo una implementación local de la solución predictiva, con el objeto de que sirva de apoyo para la identificación temprana de la deserción estudiantil y así enfocar esfuerzos administrativos y mejorar los índices de retención.

6.4 DESARROLLO DE LA PROPUESTA

6.4.1 DESCRIPCIÓN DE LA PROPUESTA

La propuesta consta del desarrollo e implementación de un modelo predictivo mediante algoritmos de aprendizaje capaces de dar respuesta de manera temprana a la deserción estudiantil, la cual es presente cuando un estudiante se ausenta de un periodo académico a otro. De tal manera el modelo predictivo servirá de apoyo como herramienta para mitigar la deserción y aumentar los índices de retención en la universidad.

Para la creación e implementación del modelo predictivo es necesario realizar procesos de análisis de datos, procesamiento de datos como limpiezas y transformaciones necesarias previos

al entrenamiento de un algoritmo de aprendizaje, posterior al proceso de datos es posible entrenar un algoritmo de aprendizaje y evaluar los resultados de este en función de cada uno de sus indicadores de eficiencia, una vez realizado el entrenamiento se lleva a cabo el proceso de implementación.

6.4.2 DESARROLLO DE LA PROPUESTA

Para el análisis, desarrollo e implementación de la solución se hizo uso de diversas tecnologías y herramientas en donde se utilizó como lenguaje de programación Python en un ambiente de Anaconda con Jupiter Notebooks y con apoyo de herramientas de visualización de datos como Power BI.

A continuación, se define las diversas etapas que fueron desarrolladas para lograr la obtención de los resultados e implementación de la solución predictiva.

6.4.2.1 OBTENCIÓN DE LA VARIABLE DE RESPUESTA

Se entiende como la variable de respuesta a el valor desconocido o el valor que se desea que calcule o puede predecir un modelo, sin embargo, un modelo estadístico debe de aprender en función de comportamientos de datos históricos, por lo cual el cálculo de la variable de respuesta en esta investigación está centrado en conocer la deserción estudiantil que se produce con la ausencia de un estudiante en el transcurso de un periodo académico a otro subsiguiente. De lo cual de los datos históricos proporcionados se realizó el cálculo de la variable de respuesta denominada como 'Salida' en cual indica si un estudiante es desertor dado que no se encuentra en el siguiente periodo académico.

Para realizar este cálculo se crearon subconjuntos de datos para poder comprobar la información en un Periodo académico [X] definido como lo actual contra un Periodo académico [X+1] definido como lo futuro, en donde mediante iteraciones de todos los ciclos académicos presentes en el conjunto de datos completo se determinó si un estudiante se encontraba o no de lo 'actual' en lo 'futuro' por medio de uniones de los conjuntos de datos determinando como tal si la ausencia del mismo implica la deserción.

```

for i in range(0, len(Ciclos)-1) :
    Actual = df_Resumen[df_Resumen['CicloAcademico'] == Ciclos[i]]
    Futura = df_Resumen[df_Resumen['CicloAcademico'] == Ciclos[i+1]]

    Futura =Futura.assign(Salida=0)

    Futura = Actual.join( Futura, how='left', lsuffix="left", rsuffix="right")
    Futura = Futura.filter(['cuentaleft', 'CicloAcademicoleft', 'Salida'])
    Futura= Futura.rename(columns={'cuentaleft': 'cuenta', 'CicloAcademicoleft': 'CicloAcademico'})

df_VR = pd.concat([df_VR, Futura])

```

Figura 17. Cálculo de variable de respuesta

Fuente: (Elaboración Propia, 2023)

6.4.2.2 TRANSFORMACIONES Y LIMPIEZA DE DATOS

En relación con los valores atípicos y valores nulos encontrados en los datos históricos proporcionado se realizó una serie de pasos para la limpieza y transformaciones de variables, en fin, de garantizar la congruencia de los datos previo al entrenamiento de un modelo estadístico.

Las variables numéricas que presentaron valores atípicos se hizo uso de imputaciones de esos valores atípicos con el valor de la mediana de dicha variable, por otro lado, para las variables categóricas los valores nulos se imputaron mediante el valor más frecuente de la variable.

Dentro de la transformación conlleva la modificación de variables con el objetivo de reducir el número de observaciones o categorías posibles, de igual manera se crearon nuevas variables con el fin de proporcionar mayor valor al análisis y al entrenamiento del modelo, a continuación, se definen las variables utilizadas.

Tabla 8. Variables de entrada del modelo.

Variable	Tipo	Descripción
Edad	Numérica	Representa la edad en años del estudiante
Genero	Categórica	Representa el género del estudiante como Masculino o Femenino
Trabajo	Categórica	Indica si el estudiante trabaja
Estado Civil	Categórica	Representa el estado civil del estudiante como Soltero, Casado o Unión libre. Esta variable ha sido tratada en donde los valores nulos se reemplazan con el valor más frecuente.
Campus	Categórica	Representa el campus al que pertenece el estudiante sea Tegucigalpa o San Pedro Sula
Cuotas del periodo	Numérica	Indica la cantidad de cuotas del periodo
Cuotas pagadas del periodo	Numérica	Indica la cantidad de cuotas pagadas en el periodo

Cuotas pendientes del periodo	Numérica	Indica la cantidad de cuotas pendientes de pagar en el periodo
Forma de pago	Categórica	Representa la forma de pago que utiliza el estudiante. Esta variable se ha transformado para simplificar el número de observaciones.
Tiene Beca	Categórica	Indica si el estudiante es becado
Tiene Descuento	Categórica	Indica si el estudiante cuenta con descuentos aplicados
Permiso de matricula	Categórica	Indica si el estudiante posee permiso para matricular condicionado.
Jornada	Categórica	Representa la jornada a la que pertenece el estudiante. Esta variable se ha transformado para simplificar el número de observaciones
Tipo de Estudiante	Categórica	Representa el tipo de estudiante en donde se clasifican como primer ingreso y reingreso
Facultad	Categórica	Representa el nombre de la facultad a la que pertenece la carrera de un estudiante.
Notas promedio	Numérica	Indica las notas promedio del estudiante en el periodo
Índice Académico	Numérica	Indica el índice académico que tiene el estudiante
UV Matriculadas	Numérica	Indica la cantidad de unidades valorativas matriculadas por el estudiante
UV Reprobadas	Numérica	Indica la cantidad de unidades valorativas reprobadas por el estudiante
UV Retiradas	Numérica	Indica la cantidad de unidades valorativas retiradas por el estudiante
Inasistencias	Numérica	Indica la cantidad de inasistencias que tiene un estudiante
Materias matriculadas	Numérica	Indica la cantidad de materias matriculadas por el estudiante
Materias reprobadas	Numérica	Indica la cantidad de materias reprobadas por el estudiante
Materias retiradas	Numérica	Indica la cantidad de materias retiradas por el estudiante
Materias aprobadas	Numérica	Indica la cantidad de materias aprobadas por el estudiante
Cumplimiento de Cuotas	Numérica	Indicador que muestra el cumplimiento de cuotas pagadas contra las cuotas del periodo
Cumplimiento de Aprobadas	Numérica	Indicador que muestra el cumplimiento de materias aprobadas contra las materias matriculadas
Índice de reprobación	Numérica	Indicador que muestra el índice de reprobación en relación con las materias reprobadas contra las matriculadas
Días	Numérica	Indicador de diferencia entre fechas de pago de cuotas
Salida	Categórica	Variable de respuesta que representa si un estudiante es desertor con un 1 o no desertor con 0.

Fuente: (Elaboración Propia, 2023)

Del conjunto de datos una vez con las variables transformadas y limpias se procede al preprocesamiento previo al entrenamiento del modelo estadístico con el apoyo de librerías en Python como pandas, numpy, sklearn entre otras. De lo cual se realiza la separación de variables numéricas y de variables categorías esto con el objeto de tratarlas de acuerdo con el tipo de datos en donde las variables numéricas se normalizan mediante la técnica de mínimos y máximos con la función `MinMaxScaler()` de sklearn.

```
from sklearn.preprocessing import MinMaxScaler

df_categoricalvar = df_fulldata.loc[:, df_fulldata.dtypes == object]
df_numericalvar = df_fulldata.loc[:, df_fulldata.dtypes != object]
df_numericalvar = df_numericalvar.fillna(0)
df_numericalvar_norm = df_numericalvar.loc[:, df_numericalvar.columns != 'Salida']
df_numericalvar_salida = df_numericalvar.loc[:, df_numericalvar.columns == 'Salida']

cols = df_numericalvar_norm.columns
index = df_numericalvar_norm.index
x = df_numericalvar_norm.values #returns a numpy array
min_max_scaler = MinMaxScaler()
x_scaled = min_max_scaler.fit_transform(x)
df_numericalvar = pd.DataFrame(x_scaled, columns=cols, index=index)
df_numericalvar = pd.concat([df_numericalvar, df_numericalvar_salida], axis=1)
```

Figura 18. Preprocesamiento de variable numéricas.

Fuente: (Elaboración Propia, 2023)

Por otro lado, las variables categóricas se convierte a valores ficticios o valores dummies que son entendibles para entrenar el modelo estadístico al mantener todas las observaciones en columnas con indicadores booleanos, esto con el apoyo de la función `get_dummies()` de pandas.

```
df_categoricalvar = pd.get_dummies(df_categoricalvar)
```

Figura 19. Preprocesamiento de variable categóricas.

Fuente: (Elaboración Propia, 2023)

6.4.2.3 EL ENTRENAMIENTO

Para el entrenamiento del modelo es necesario separar los conjuntos de datos en un conjunto de entrenamiento y un conjunto de pruebas por lo cual se utiliza la siguiente distribución 75% para entrenamiento y 25% para pruebas del modelo entrenado. La separación del conjunto de datos se realiza mediante la función `train_test_split()` de sklearn. Posteriormente a ello se realiza el entrenamiento del mediante el algoritmo de aprendizaje seleccionado, en este caso en base a los

resultados presentados los capítulos anteriores se decidió realizar la solución predictiva con el apoyo del algoritmo de aprendizaje de regresión logística dado que es el que mejor se apega a los datos reales de acuerdo con la predicción que realiza.

```
from sklearn.preprocessing import StandardScaler
X_train, X_test, y_train, y_test = train_test_split(df_x,df_y, train_size=0.75, random_state=42,stratify=df_y)
LR = LogisticRegression(max_iter=1200,random_state=42).fit(X_train, y_train)
y_pred = LR.predict(X_test)
LR.score(X_test, y_test)
```

Figura 20. Entrenamiento de modelo.

Fuente: (Elaboración Propia, 2023)

Una vez entrenado el modelo se procede a verificar los indicadores de eficiencia anteriormente presentados en el capítulo de resultados y se guarda el modelo entrenado para poder utilizarlo en las implementaciones a través de un archivo pickle, además de guardar el modelo en el archivo también se deben de guardar las variables de entrada del modelo en otro archivo para poder utilizarlos correctamente las variables en la implementación.

6.3 LA IMPLEMENTACIÓN

En la implementación se utilizan observaciones o registros nuevos a los cuales se desea conocer la variable de respuesta en donde todo el proceso de la etapa I y II se debe llevar a cabo para garantizar la ejecución del modelo entrenado con los nuevos datos de entrada.

Se realiza el llamado de las variables de entrada del modelo desde el archivo almacenado en la etapa anterior al igual que el modelo entrenado desde el archivo pickle. Al cargar lo archivos el modelo está disponible para poder realizar las predicciones de las nuevas observaciones, estas predicciones son almacenadas en una nueva columna en el conjunto de datos a manera de respuesta como Predicción y para obtener la probabilidad de clasificación del modelo de igual manera se almacena en una variable determinada como Score, la cual sirve para categorizar los resultados de la deserción mediante la variable Clasificación.

Los resultados del modelo son almacenados en un archivo de resultados en formato csv el cual está disponible para utilizar en la toma de decisiones a través del consumo de datos en bruto o mediante herramientas de visualización véase el anexo II. Para implementar la solución predictiva consultar el manual de uso e implementación del modelo predictivo véase el anexo III.

6.5 MEDIDAS DE CONTROL

Una vez implementado el modelo es de utilidad la respuesta proporcionada en donde se obtienen tres variables importantes que sirven como indicadores para determinar la deserción siendo el primero de conjunto de resultados la variable “Predicción” que clasifica como verdadero o falso si el modelo determina al estudiante como desertor o no, de igual manera esta variable se argumenta con la variable “Score”, que define la probabilidad de que ocurra el suceso y por último en función de esta probabilidad y para proporcionar una perspectiva más entendible se encuentra la variable “Clasificación” el cual categoriza a los estudiantes desde un riesgo bajo a un riesgo muy alto basado en el valor del score.

La variable de clasificación ayuda a enfocar los esfuerzos al conocer los estudiantes clasificados con riesgo “MUY ALTO” e implica que deben de ser atendidos con prioridad ya que poseen una mayor probabilidad de que sean estudiantes desertores. La clasificación se categoriza de la siguiente manera:

Tabla 9. Indicadores de resultados del modelo.

Predicción	Score	Criterio	Clasificación
1	0.9	Score \geq 0.9	MUY ALTO
1	0.75	Score \geq 0.75 y Score $<$ 0.90	ALTO
1	0.5	Score \geq 0.50 y Score $<$ 0.75	MEDIO
0	0.3	Score $<$ 0.50	BAJO

Fuente: (Elaboración Propia, 2023)

6.6 CRONOGRAMA Y PRESUPUESTO DE IMPLEMENTACIÓN

Tabla 10. Cronograma de actividades para implementación de modelo predictivo.

Actividad	Costo USD	Semana						
		1	2	3	4	5	6	7
Levantamiento de requerimientos y datos históricos	250	■						
Análisis exploratorio de datos	500	■	■					
Limpieza y transformación de variables	350		■	■				
Entrenamiento de modelo	500			■	■	■		
Evaluación de modelo	500				■	■	■	
Implementación	500						■	■
Total	2600							

Fuente: (Elaboración Propia, 2023)

Valor estimado diario de desarrollo e implementación 50 USD el cual asciende a un total de 2600 USD en las siete semanas, adicionalmente para llevar a cabo la implementación es necesario contar con el ambiente y equipo necesario por lo cual el cliente debe de contar con un equipo de cómputo con las siguientes capacidades mínimas:

Sistema operativo: Windows 10

Memoria RAM: 8GB

Disco duro: 256GB

Costo aproximado 600 USD

6.7 CONCORDANCIA DE LOS SEGMENTOS DE LA TESIS CON LA PROPUESTA

Tabla 11. Concordancia de los segmentos de la tesis con la propuesta

Capítulo I			Capítulo II	Capítulo III			Capítulo V	Capítulo VI	
Título Investigación	Objetivo General	Objetivos específicos	Teorías/Metodologías de sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la Propuesta	Objetivos de la Propuesta
Implementación de modelo predictivo para identificar deserción estudiantil en Unitec	Identificar de manera temprana la deserción estudiantil en un periodo académico en UNITEC en los centros de Tegucigalpa y San Pedro Sula, mejorando los índices de retención y mitigando pérdidas económicas para la institución, mediante la aplicación de modelos predictivos.	Identificar los principales factores que causan la deserción de un estudiante.	Teorías de factores de deserción	Datos demográficos, académicos y financieros	96,103 estudiantes pregrado	Observación de base de datos de toda la población de estudiantes de pregrado del 2020-1 al 2023-1	Los factores que influyen en la deserción de los estudiantes en Unitec se identificó mediante el análisis exploratorio de datos y entrenamiento de soluciones predictoras para dar una respuesta a la deserción, que dentro los aspectos más influyentes se encuentra variables académicas; rendimiento del estudiante, índices de notas bajos, cantidad de materias reprob. y ret .	Implementación de modelo predictivo para identificar deserción estudiantil en Unitec	Desarrollar un modelo predictivo de inteligencia artificial mediante algoritmos de aprendizaje en busca de dar respuesta a la deserción estudiantil llevando a cabo una implementación local de la solución predictiva, con el objeto de que sirva de apoyo para la identificación temprana de la deserción estudiantil y así enfocar esfuerzos administrativos y mejorar los índices de retención.

Continuación tabla 11

Capítulo I			Capítulo II	Capítulo III			Capítulo V	Capítulo VI	
Título Investigación	Objetivo General	Objetivos específicos	Teorías/ Metodologías de sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la Propuesta	Objetivos de la Propuesta
		Conocer el comportamiento de los estudiantes desertores en UNITEC.	Teorías de deserción	Datos demográficos, académicos y financieros			Conforme al comportamiento de los estudiantes que abandonan sus estudios de un periodo académico a otro resaltan comportamientos relacionados con el rendimiento académico en donde se ha notado que los estudiantes que tienen a desertar presentan índice académico, materias aprobadas más bajas que el promedio además que se presentan mayor cantidad de inasistencias.		

Continuación tabla 11

Capítulo I			Capítulo II	Capítulo III			Capítulo V	Capítulo VI	
Título Investigación	Objetivo General	Objetivos específicos	Teorías/Metodologías de sustento	Variables	Poblaciones	Técnicas	Conclusiones	Nombre de la Propuesta	Objetivos de la Propuesta
		Desarrollar una propuesta de diseño e implementación de un modelo predictivo de inteligencia artificial mediante algoritmos de aprendizaje basados en la deserción académica para mejorar la retención estudiantil.	Teorías de modelos predictivos/inteligencia artificial	Datos demográficos, académicos y financieros			En el capítulo de aplicabilidad se define la propuesta de implementación de un modelo predictivo para dar respuesta a la deserción estudiantil para estudiantes de pregrado en la Unitec, con el apoyo de información histórica proporcionada, mediante procesos de datos y algoritmos de aprendizaje entrenados en busca de dar respuesta a la deserción		

Fuente: (Elaboración Propia, 2023)

BIBLIOGRAFÍA

- Arguelles Ballen, Y y Steffanell Piñeres, M. (2023). Modelo basado en el uso de la gamificación para la retención de los estudiantes de pregrado de la Universidad Nacional Abierta y a Distancia UNAD. Corporación Universidad de la Costa.
- Barramuño, M., Meza, C., & G. Gálvez, G. (2021). Prediction of student attrition risk using machine learning,” J. Appl. Res. High. Educ. Vol. 14 No. 3, pp. 974-986. <https://doi.org/10.1108/JARHE-02-2021-0073>
- Barrero Rivera, F. (2015). Investigación en deserción estudiantil universitaria: educación, cultura y significados. Revista Educación Y Desarrollo Social, 9(2), 86–101. <https://doi.org/10.18359/reds.948>
- Buzai, G. D., & Baxendale, C. A. (2009). Análisis exploratorio de datos espaciales. Geografía y Sistemas de Información Geográfica, N° 1,(2009). <http://ri.unlu.edu.ar/xmlui/handle/rediunlu/702>
- Cabrera, L., Bethencourt, J. T., Álvarez Pérez, P., & González Alfonso, M. (2006). El problema del abandono de los estudios universitarios. *Relieve*, 12(2), 171-203.
- Calderón, L. R. A. (2020, 8 julio). Modelo predictivo para determinar deserción de estudiantes en la Universidad Tecnológica del Perú. <https://repositorio.utp.edu.pe:443/handle/20.500.12867/2924>
- Calderón, R. (2011). El crecimiento y desarrollo de la educación superior en Honduras, una perspectiva desde la UNAH. <https://www.redalyc.org/pdf/1794/179422350010.pdf>
- Canales, A., & De los Ríos, D. (2007). Factores explicativos de la deserción universitaria. *Calidad en la Educación*, (26), 173-201. <https://doi.org/10.31619/caledu.n26.239>
- Canales, A., & De los Ríos, D. (2009). Retención de estudiantes vulnerables en la educación universitaria. *Calidad en la Educación*, (30), 50-83. <https://doi.org/10.31619/caledu.n30.173>

- Castillo Díaz, M.A., Mendoza-Aly, J.L., Martínez-Martínez, C.V. y Martínez-Gonzales, L.E. (2022). Alteraciones de salud mental y adaptación universitaria en estudiantes hondureños: un análisis de grupos con rendimiento académico diferenciado. *Revista Española de Orientación y Psicopedagogía*, 33(3), 144-161. <https://doi.org/10.5944/reop.vol.33.num.3.2022.36466>
- Castillo, M., (2022). Perfil de Ingreso y Vivencias Académicas Tempranas Implicadas en la Intención de Abandono en Educación Superior. Análisis En Contexto Hondureño. https://www.researchgate.net/profile/Marcio-Alexander-Castillo-Diaz/publication/365872579_Perfil_de_ingreso_y_vivencias_academicas_tempranas_implicadas_en_la_intencion_de_abandono_en_educacion_superior_Analisis_en_contexto_hondureno/links/63878d0178f94b73a0ba5e65/Perfil-de-ingreso-y-vivencias-academicas-tempranas-implicadas-en-la-intencion-de-abandono-en-educacion-superior-Analisis-en-contexto-hondureno.pdf
- Cuji, B., Gavilanes W. , & Sanchez R., Modelo predictivo de deserción estudiantil basado en arboles de decisión, *Espacios*, vol. 38, no. 55, p. 17, 2017. <http://ww.revistaespacios.com/a17v38n55/a17v38n55p17.pdf>
- Díaz-Peralta, C. (2008). Modelo conceptual para la deserción estudiantil universitaria chilena. *Estudios Pedagógicos*, 34(2), 65-86. Donoso,S., & Schiefelbein,E.(2007). Análisis de los modelos explicativos de retención de estudiantes en la universidad: una visión desde la desigualdad social. *Estudios Pedagógicos*, 33(1), 7-27.
- Duriez, M., López, V., & Moncada, G. (2016). INFORME NACIONAL: HONDURAS. Recuperado de <https://cinda.cl/wp-content/uploads/2019/01/educacion-superior-en-iberoamerica-informe-2016-informe-nacional-honduras.pdf>
- Educación superior en Iberoamérica. Informe 2016. Informe nacional: Honduras - CINDA. (s. f.). Centro Interuniversitario de Desarrollo. <https://cinda.cl/publicacion/educacion-superior-en-iberoamerica-informe-2016-informe-nacional-honduras/>
- Fernández Hileman, M. D. R., Corengia, Ángela, & Durand, J. (2014). Deserción y retención universitaria: una discusión bibliográfica. *Pensando Psicología*, 10(17), 85–96. <https://doi.org/10.16925/pe.v10i17.787>

- Flores Juanez, K. M., & Acevedo Amaya, M. R. (2023). Disrupción en las condiciones socioeconómicas y educativas derivadas del COVID19 en Honduras: caso MAE-UNAH. *Universidad Y Sociedad*, 15(4), 673-684. Recuperado a partir de <https://rus.ucf.edu.cu/index.php/rus/article/view/4024>
- García, M. E., Gutiérrez, A. B. B., Herrero, E. T., Menéndez, R. C., & Pérez, J. C. N. (2016). El contexto sí importa: Identificación de relaciones entre el abandono de titulación y variables contextuales. *European Journal of Education and Psychology*, 9(2), 79-88. <https://doi.org/10.1016/j.ejeps.2015.06.001>
- García, J., Molina, J., Berlanga, A., Patricio, M., Bustamante, A., & Padilla, W. (2018). *Ciencia de datos. Técnicas Analíticas y Aprendizaje Estadístico*. Bogotá, Colombia. Publicaciones Altaria, SL.
- González, D. & Girón, L. (2005). Determinantes del rendimiento académico y la deserción estudiantil en el programa de Economía de la Pontificia Universidad Javeriana de Cali. *Economía, Gestión y Desarrollo*, 3, 173-201.
- Jiménez, A. J. y Tamiran, R. S. (2015). Caracterización de la deserción estudiantil en educación superior con minería de datos. *Revista Tecnológica ESPOL*, 28(5), 447-463.
- J. P. Valenzuela y N. Yañez, “Trayectoria y políticas de inclusión en educación superior en América Latina y el Caribe en el contexto de la pandemia: dos décadas de avances y desafíos”, *Documentos de Proyectos (LC/TS.2022/50)*, Santiago, Comisión Económica para América Latina y el Caribe (CEPAL), 2022.
- Ley de Educación Superior. (n.d.) Recuperado de <https://des.unah.edu.hn/repositorio/normativa/>
- Llaugel, Felipe & González-Disla, Renato. (2016). Modelo Predictivo de Desercion Escolar para la Republica Dominicana. *Revista de Investigación y Evaluación Escolar*. <http://dx.doi.org/10.13140/RG.2.2.18485.88806>
- López-Dr, M. D. D. V., & Moncada, G. (2016). *INFORME NACIONAL: HONDURAS*. <https://cinda.cl/wp-content/uploads/2019/01/educacion-superior-en-iberoamerica-informe-2016-informe-nacional-honduras.pdf>

- Munizaga Mellado, F., Cifuentes Orellana, M., & Beltrán Gabrie, A. (2017). Variables y factores asociados al fenómeno de la retención y abandono estudiantil universitario en investigaciones de Latinoamérica y el Caribe. Congresos CLABES. Recuperado a partir de <https://revistas.utp.ac.pa/index.php/clabes/article/view/1671>
- Palacios, B. M. E. (2023, 7 febrero). Repositorio Universidad Técnica de Ambato: Modelo matemático para predecir el grado de deserción de los estudiantes en el Instituto Superior Tecnológico Bolívar. <https://repositorio.uta.edu.ec/jspui/handle/123456789/37204>
- Pineda, C., Pedraza, A. y Moreno, I. (2011). Efectividad de las estrategias de retención universitaria: la función del docente. *Educación y Educadores*, 14(1). Recuperado de <https://bit.ly/2ziibm>
- Plan Estratégico del SES 2014 - 2023. (n.d.). Edu.hn. Recuperado de <https://des.unah.edu.hn/repositorio/plan-estrategico-del-ses-2014-2023/>
- Portillo Vera., E. (2023). Factores asociados a la deserción universitaria en la Licenciatura en Administración de Empresas y Contaduría Pública, en el Paraguay: Factors associated with university desertion in the Bachelor of Business Administration and Public Accounting, in Paraguay. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades*, 4(1), 1312–1319. <https://doi.org/10.56712/latam.v4i1.340>
- ¿Qué es la ciencia de datos? - Explicación de la ciencia de datos - AWS. (s. f.). Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/data-science/>
- ¿Qué es la inteligencia artificial? | Microsoft Azure. (s. f.). Microsoft Azure. <https://azure.microsoft.com/es-es/resources/cloud-computing-dictionary/what-is-artificial-intelligence/>
- ¿Qué es la inteligencia artificial? (s. f.). ORACLE. <https://www.oracle.com/mx/artificial-intelligence/what-is-ai/>
- Ramírez, P., & Grandón, E. Predicción de la Deserción Académica en una Universidad Pública Chilena a través de la Clasificación basada en Árboles de Decisión con Parámetros Optimizados. *Form. Univ.*, vol. 11, no. 3, pp. 3–10, 2018. <http://dx.doi.org/10.4067/S0718-50062018000300003>

- Rouhiainen, L. (2018). *Inteligencia artificial*. Madrid: Alienta Editorial.
- Ruiz Bolívar, C. (2022). Enfoque tecno-pedagógico post COVID-19: una propuesta para las instituciones de Educación Superior de América Latina. *INVESTIGACIÓN Y POSTGRADO*, 36(2), 9-23. Recuperado de <https://revistas-historico.upel.edu.ve/index.php/revinpost/article/view/9677>
- Sifuentes Bitocchi, O. (2018). Modelos predictivos de la deserción estudiantil en una universidad privada peruana. *Industrial Data*, 21(2), 47–52. <https://doi.org/10.15381/idata.v21i2.15602>
- Solis, M., Moreira, T., Gonzalez, R., Fernandez, T., & Hernandez, M., “Perspectives to predict dropout in university students with machine learning,” in 2018 IEEE International Work Conference on Bioinspired Intelligence (IWOBI), 2018, pp. 1–6. <https://doi.org/10.1109/IWOBI.2018.8464191>
- Urbina-Nájera, A. B., Camino-Hampshire, J., & Raúl, C. B. (2020). Deserción escolar universitaria: Patrones para prevenirla aplicando minería de datos educativa. *Revista Electrónica De Investigación y Evaluación Educativa*, 26(1)<https://doi.org/10.7203/relieve.26.1.16061>
- Valero S., Oreal, Vargas, Salvador A., et.al (2014). *Minería de datos*. México: Universidad de Puebla. 2014. 10 p. <https://www.ties.unam.mx/num02/pdf/mineria.pdf>
- Vásconez Altamirano, G. E. (2023). Modelo de predicción de deserción escolar en los estudiantes de la unidad educativa Los Andes por impacto de la pandemia. *Ciencia Latina Revista Científica Multidisciplinar*, 7(1), 3038-3052. https://doi.org/10.37811/cl_rcm.v7i1.4640
- Viale Tudela H., (2014). Una Aproximación Teórica a la Deserción Estudiantil Universitaria. *Revista Digital de Investigación en Docencia Universitaria*, vol. 8, no. 1, pp. 59–75, 2014. <https://www.redalyc.org/pdf/4985/498573047004.pdf>
- Web, R. (2020, 26 diciembre). Honduras: Por la pandemia, unos 50 mil alumnos abandonaron la universidad. *Tiempo.hn*. <https://tiempo.hn/honduras-50-mil-abandonaron-universidad-2020/>

Zavala, M., Álvarez, M., Vázquez, M., González, I. y Bazán, A. (2018). Factores Internos, Externos y Bilaterales asociados con la deserción en estudiantes universitarios. *Interacciones*, 4(1), 59-69. doi: <https://doi.org/10.24016/2018.v4n1.103>

ANEXOS

ANEXO I

San Pedro Sula, Cortés 09 de junio del 2023

Srs. UNITEC

Estimado Señor(a):

Reciba un cordial y atento saludo. Por medio de la presente solicitamos su apoyo, dado que somos estudiantes de UNITEC y nos encontramos desarrollando el Proyecto Final de Graduación, previo a obtener nuestro título de maestría en Analítica de Negocios.

Hemos seleccionado como tema Modelo de Predicción de Deserción Estudiantil para Estudiantes de Pregrado, por lo que estaríamos muy agradecidos de contar con su apoyo para poder desarrollar nuestra investigación.

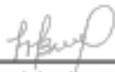
Los datos que necesitaríamos son de los últimos 3 años, y serían las variables abajo descritas:

- Datos demográficos
- Datos académicos
- Datos financieros

Nos comprometemos a no suministrarle a ninguna persona o empresa de manera verbal, escrita o de cualquier otra manera ninguna clase de información o datos sobre la información brindada.

Por lo antes descrito, y en conocimiento pleno, firmamos la misma.

Atentamente,



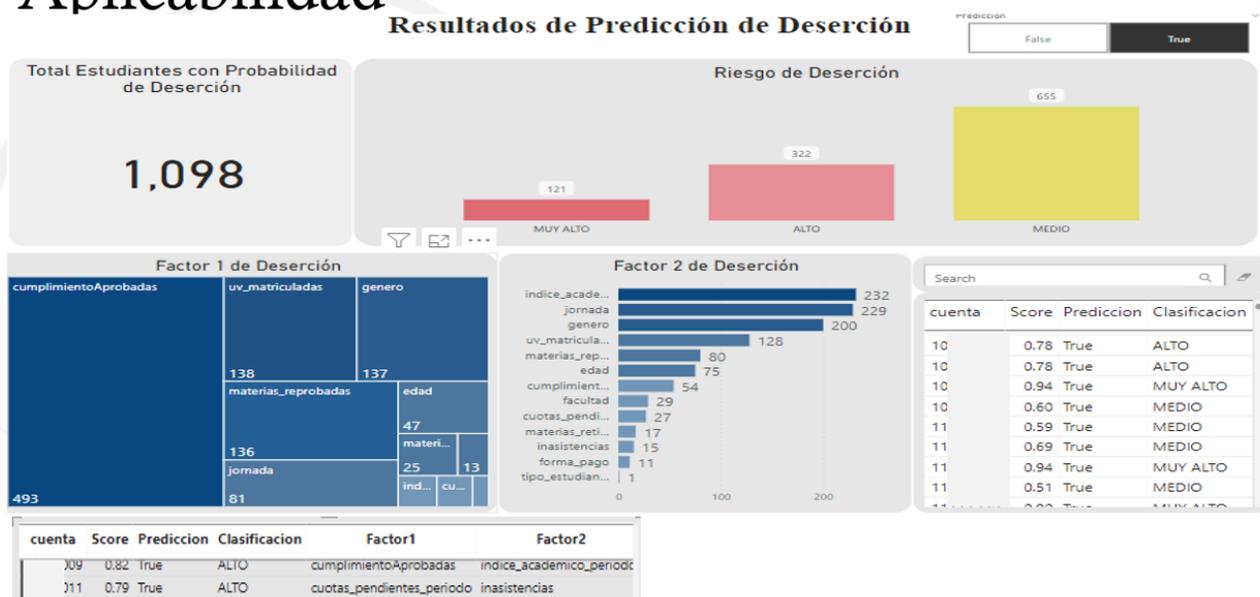
Melissa Alejandra Cabrera Burgos
12213193



William Josué Caballero Cáceres
12213036

ANEXO II

Aplicabilidad



Fuente: Elaboración propia con base de datos de UNITEC 2020-2023

ANEXO III

MANUAL DE USO E IMPLEMENTACIÓN DEL MODELO

En el presente manual se describe una serie de pasos necesarios para llevar a cabo la implementación y uso correcto de la solución predictiva, desde la preparación del ambiente necesario hasta el manejo de los resultados.

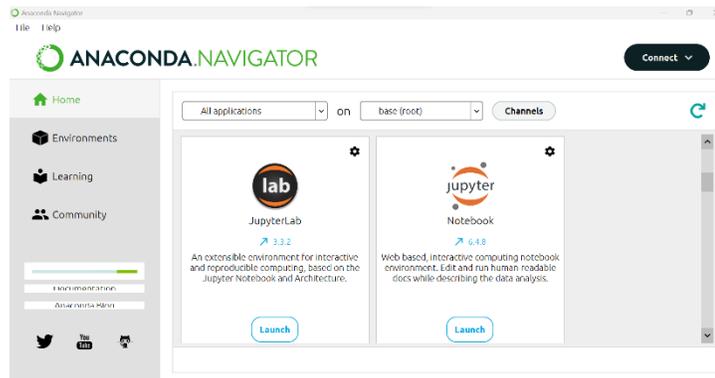
CONFIGURACIÓN DEL AMBIENTE

El modelo predictivo esta implementado en un ambiente Anaconda en Jupyter Notebooks con Python como lenguaje de programación por lo cual se debe de seguir los siguientes pasos:

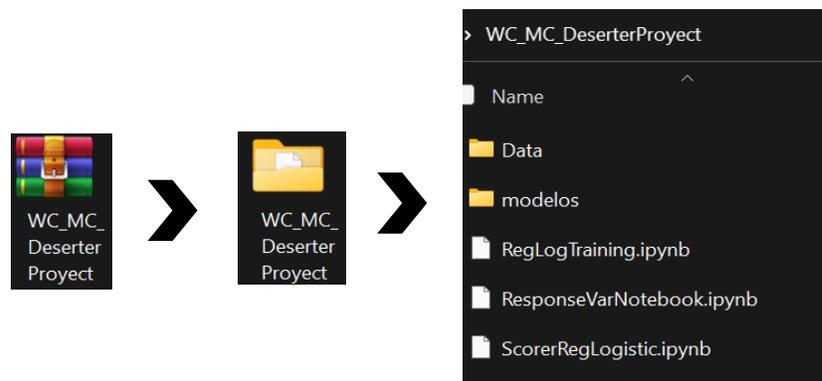
1. Descargar Anaconda desde <https://www.anaconda.com/download>.



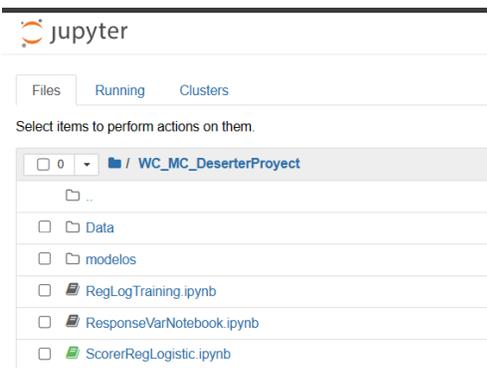
2. Ejecutar el instalador de Anaconda dando doble clic al archivo descargado.
3. Seleccionar la ruta de instalación y seguir los pasos del asistente de instalación.
4. Ejecutar Anaconda Navigator, buscar el apartado de Jupyter Notebook y dar clic en “Launch”



5. En la ruta de instalación o la de su preferencia crear la copia del directorio que contiene el modelo y los datos, el cual se debe de descomprimir y mantener la estructura proporcionada.



6. Leer el directorio desde Jupyter Notebook



PREPARACIÓN DE LOS DATOS

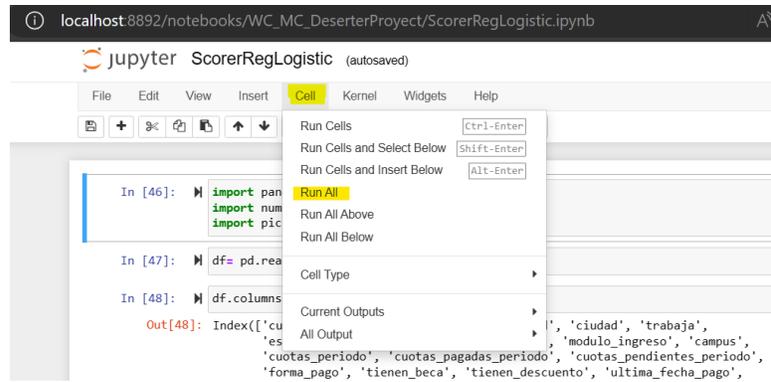
En la carpeta “Data” se encuentra un archivo csv nombrado DatasetEntrante sirve de referencia para corroborar el formato de los datos que deben de ingresarse los datos y las variables a utilizar en el modelo para que pueda realizar la predicción.

Name	Date modified	Type	Size
Resultado	20/9/2023 23:08	Microsoft Excel Com...	672 KB
DatasetCompletoVR	20/9/2023 22:54	Microsoft Excel Com...	31,182 KB
DatasetEntrante	20/9/2023 22:54	Microsoft Excel Com...	2,376 KB
Financieros	20/8/2023 10:06	Microsoft Excel Com...	8,970 KB
Academicos	20/8/2023 10:06	Microsoft Excel Com...	15,808 KB
Demograficos	20/8/2023 10:04	Microsoft Excel Com...	1,249 KB

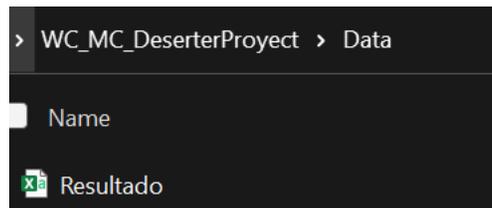
El “DatasetEntrante” debe contener las variables en bruto para realizar la predicción, en donde se encuentra información demográfica, académica y financiera a nivel de ciclo académico en general. Estos datos se procesarán posteriormente previo a implementar el modelo.

EJECUCIÓN DEL MODELO PREDICTIVO

1. Abrir el archivo “ScorerRegLogistic.ipynb” desde Jupyter notebooks.
2. En el menú “celdas” seleccionar la opción “Ejecutar todo”.



3. Al ejecutar el modelo de implementación automáticamente se crea un archivo de “Resultados.csv” que se encuentra almacenado en la carpeta de data del mismo directorio.



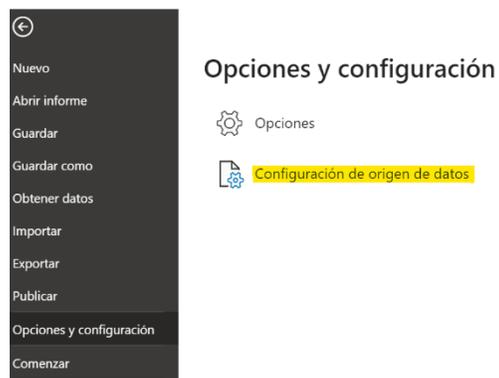
MANEJO DE RESULTADOS.

Una vez implementado el modelo se encuentran disponibles los datos en el archivo de resultados que podrá ser utilizado mediante cualquier herramienta de lectura o visualización como por ejemplo Excel o PowerBI.

1. Si se desea manipular los datos en Excel solo será necesario abrir el archivo Resultados.csv.
2. Si se desea visualizar los datos mediante Power BI es necesario configurar la ruta del tablero elaborado de manera de actualizar los datos con los valores de ejecución de modelo, para ello se debe los siguientes pasos:
 - a. Abrir el tablero de power BI de resultados de implementación.
 - b. En el menú “archivo”, dar clic en la sección de “opciones y configuración”



c. Dar clic en configuración de origen de datos.



d. Dar clic en “Cambiar Origen”, en el cual se debe de establecer la ruta del archivo de resultados que se encuentra en el directorio proporcionado.

Configuración de origen de datos

