

**PROYECTO INSTITUCIONAL VICERRECTORIA ACADÉMICA, FACULTAD DE
INGENIERÍA Y TECNOLÓGICAS, PROGRAMA DE INGENIERÍA ELECTRÓNICA**

**TRABAJO DE TESIS PARA OPTAR AL TÍTULO DE INGENIERO
ELECTRÓNICO**

**MODELO DE DETECCIÓN DE LA DESERCIÓN UNIVERSITARIA COMO
OBSTÁCULO PARA LA CALIDAD DE LA EDUCACIÓN Y EL DESARROLLO
SOCIOECONÓMICO APLICADO AL PROGRAMA DE INGENIERÍA
ELECTRÓNICA DE LA UNIVERSIDAD POPULAR DEL CESAR**

Yorman Amílkar Quintana Pabón
yaquintana@unicesar.edu.co

DIRECTOR: José Ramón Iglesias Gamarra
ASESOR: Omaira Luz Tapias Díaz PhD (C)

DSP-ASIC BUILDER GROUP
Semillero de investigación
Línea de investigación: PROCESAMIENTO DE SEÑALES

UNIVERSIDAD POPULAR DEL CESAR
FACULTAD DE INGENIERIA Y TECNOLÓGICAS
PROGRAMA DE INGENIERÍA ELECTRÓNICA
VALLEDUPAR
2024

DEDICATORIAS

La culminación de este proceso se lo quiero dedicar llenamente a mi familia que pese a que algunos no creían en mí con el tiempo y por distintas situaciones fueron entregándome apoyo, pero quisiera hacer una especial dedicatoria a mi madre Edilma Pabón que pese a que ya no me acompaña en esta experiencia que llamamos vida, fue la persona que nunca dejo de creer en mí y que me apoyo en todo, quiero expresarle mis más grandes disculpas por toda la demora que pase para llegar a este punto y aunque no pude darte la dicha que me vieras tomando el título, este logro es por y para ti, perdóname siempre te amare.

AGRADECIMIENTOS

Expreso mi más sincero reconocimiento a todos aquellos que, directa o indirectamente, contribuyeron al éxito de este proyecto de grado.

A mis padres, Edilma Pabón y Alfonso Quintana principalmente a mi madre Edilma Pabón que me acompañó y animó, su amor incondicional, apoyo constante y especialmente por creer en mí, en cada paso de mi formación académica. Su sacrificio y dedicación han sido fundamentales para alcanzar este logro.

A mi novia Yuranis Duran que ha estado conmigo durante todo este extenso proceso, apoyándome y animándome ahí a mi lado demostrándome que si se puede.

A mi tutor de tesis, José Ramón por su guía experta, sus valiosas sugerencias y su paciencia durante todo el arduo y extenso proceso de investigación. Su conocimiento y experiencia han sido cruciales para el desarrollo de este trabajo.

A mis compañeros de estudio y amigos, por su apoyo, camaradería y por hacer de esta experiencia algo inolvidable. Gracias por estar siempre ahí, especialmente en los momentos más difíciles.

A todas las personas que participaron en las encuestas y entrevistas, y a aquellos que de alguna manera contribuyeron con su tiempo y conocimientos a la recopilación de datos para esta investigación.

CONTENIDO

RESUMEN	6
INTRODUCCIÓN	8
ANTECEDENTES	9
LÍNEA Y TEMA DE INVESTIGACIÓN	11
DESCRIPCIÓN Y FORMULACIÓN DEL PROBLEMA	12
JUSTIFICACIÓN	14
OBJETIVO GENERAL	15
OBJETIVOS ESPECÍFICOS	15
MARCOS REFERENCIALES	16
ANTECEDENTES	16
MARCO TEÓRICO	18
MARCO HISTÓRICO	22
MARCO CONCEPTUAL	23
MARCO LEGAL	28
DISEÑO METODOLÓGICO	30
MÉTODO DE INVESTIGACIÓN	34
METODOLOGÍA	52
RESULTADOS Y ANÁLISIS	53
CONCLUSIONES	59
RECOMENDACIONES	60
REFERENCIAS BIBLIOGRÁFICAS	61

ÍNDICE DE TABLAS Y FIGURAS

<i>Tabla 1. Línea de investigación.</i>	<i>¡Error! Marcador no definido.</i>
<i>Figura 1. Deserción en programas de formación superior en Colombia (2016).</i>	20
<i>Figura 2. Correlación minería de datos con la estadística, IA y Machine Learning.</i>	25
<i>Figura 3. Registros y variables totales obtenidas a través de la encuesta realizada.</i>	31
<i>Figura 4. Correlación de variables.</i>	32
<i>Tabla 2. Variables con base a la información personal</i>	35
<i>Tabla 3. Variables categóricas con base a la información personal y sus valores.</i>	37
<i>Tabla 4. Variables con base a la información socioeconómica</i>	37
<i>Tabla 5. Variables categóricas con base a la información socioeconómica y sus valores</i>	39
<i>Tabla 6. Variables con base a la información académica</i>	39
<i>Tabla 7. Variables categóricas con base a la información académica y sus valores</i>	41
<i>Figura 5. Clasificación de tablas en la base de datos creada en PostgreSQL.</i>	42
<i>Figura 6. Base de datos creada en PostgreSQL.</i>	42
<i>Figura 7. Estructura jerárquica del aplicativo.</i>	44
<i>Figura 8. Construcción páginas del aplicativo.</i>	46
<i>Figura 9. Cuadro de operaciones.</i>	48
<i>Figura 10. Tabla de consulta de datos del aplicativo.</i>	48
<i>Figura 11. Componente de edición del aplicativo.</i>	49
<i>Figura 12. 1er gráfico informativo del aplicativo (Cantidad de estudiantes según su género).</i>	50
<i>Figura 13. Visualización página principal (home) aplicativo web.</i>	53
<i>Figura 14. Formulario de recopilación de la información personal.</i>	54
<i>Figura 15. Formulario de recopilación de la información académica.</i>	55
<i>Figura 16. Formulario de recopilación de la información socioeconómica.</i>	56
<i>Figura 17. Gestor de archivos – parte 1.</i>	57
<i>Figura 18. Gestor de archivos – parte 2.</i>	58

RESUMEN

La deserción universitaria, la falta de cualificación docente y el panorama de inserción laboral de los egresados son problemáticas que enfrentan la mayoría de las instituciones de educación superior en Colombia, trayendo efectos negativos a todo el sistema educativo, pero especialmente a los estudiantes debido a que truncan el ciclo de formación educativo y sus posibilidades de ocupación laboral, lo cual trae consecuencias a todo nivel. Igualmente, para las universidades porque afecta los indicadores de calidad educativa; y para el caso de las instituciones públicas los recursos que recibe por cada estudiante; ya que, al no cumplirse el objeto de dichos recursos, estos se desaprovechan afectando las finanzas de las instituciones. Este escenario, junto a la falta de cualificación docente y la situación de egresados en el espectro laboral colombiano, podría desmejorar la calidad de la educación lo mismo que a las actividades de gestión docencia y administrativas en los claustros de educación superior.

Actualmente en Colombia las universidades comienzan a utilizar las técnicas de alta computación enfocadas a sus sistemas de información tratando de mejorar los procesos de gestión, aprendizaje y desarrollo cualitativo de sus profesores. Es por eso que, la Universidad Popular del Cesar viene desarrollando un trabajo de investigación con el fin de desarrollar modelos de IA (Inteligencia Artificial) y de reconocimiento de patrones capaces de brindar información para trazar políticas públicas relacionadas con el bienestar estudiantil, profesoral y de los egresados, a fin de evitar o reducir la deserción estudiantil a partir de sus causas, el retiro de profesores por falta de motivación, entre otras problemáticas. Tomando el programa de Ingeniería Electrónica como piloto para desarrollar esta propuesta.

Tanto la deserción estudiantil como profesoral tiene diferentes causas y está supeditada en un gran porcentaje a las condiciones tanto de los estudiantes como de los docentes y del entorno en que estos se desenvuelven. Lo primero que este trabajo propone es poder incluir la mayor cantidad de variables posibles (características individuales, socio-económicas, académicas e institucionales) con las cuales podamos dar explicación al problema de la deserción estudiantil y de docentes y el impacto de nuestros egresados en la sociedad.

El objetivo de este trabajo es proponer algunos modelos cuyo núcleo principal sea el *learning analytics* a partir de la información recogida en una encuesta realizada a los estudiantes, profesores y egresados del programa de Ingeniería Electrónica de la Universidad Popular del Cesar, lo que se busca con este trabajo es que la

información obtenida sirva de insumo para poder identificar a los estudiantes que puedan estar en riesgo de abandonar la universidad desde el momento en que ingresan a la universidad y de los profesores que se encuentran desmotivados por las condiciones que ofrece la institución para ellos. Con el fin de poder cumplir con los propósitos investigativos se planea implementar una metodología con modelos sucesivos, a través de los que se crearán modelos predictivos en tres momentos diferentes, los cuales son normales en nuestra institución de acuerdo con los estudios iniciales.

INTRODUCCIÓN

Teniendo en cuenta la problemática que representa la deserción académica, y considerando para este caso el ámbito universitario, esta afecta a muchas instituciones educativas, tanto a nivel internacional, como nacional, por lo tanto, no exceptúa a nuestro claustro, la Universidad Popular del Cesar. Como consecuencia de esto se genera un impacto negativo en el desarrollo socioeconómico de la región.

Con este enfoque, se plantea la creación de un modelo de detección temprana de la deserción, utilizando métodos de análisis y ciencia de datos, así como técnicas de *Learning Analytics*; el cual permita identificar qué estudiantes tienen un mayor riesgo de desertar del programa de su escogencia.

Para el desarrollo del mismo se usará como piloto el programa de Ingeniería Electrónica, puesto que es bien conocido de que es uno de los programas ofertados por la Universidad Popular del Cesar con mayor exigencia, debido a que requiere altos niveles de dedicación y habilidades técnicas especializadas; y, por ende, tiene al mismo tiempo, altos niveles de deserción.

Además, se estudiará el impacto de la deserción en el desarrollo socioeconómico de la región, especialmente en lo que se refiere a la formación de profesionales especializados en áreas de alta demanda.

La meta es desarrollar un modelo de detección preciso y confiable, que pueda ser utilizado por la universidad para mejorar la retención de estudiantes y en el que se desarrollen estrategias de apoyo hacia dichos estudiantes, para lograr una disminución en la deserción y abandono de los estudios del estudiantado; y al mismo tiempo contribuir al desarrollo socioeconómico de la región.

ANTECEDENTES

La deserción universitaria es un problema común en todo el mundo, que afecta a estudiantes de todos los niveles de educación superior y esto se ve evidenciado en varios textos y estudios. La tasa de deserción universitaria en programas de ingeniería electrónica es particularmente alta, debido a la naturaleza rigurosa del programa y la complejidad de los cursos. A continuación, se presentan algunos de los estudios más relevantes sobre la detección de la deserción universitaria:

El análisis de datos académicos es uno de los métodos más utilizados para detectar la deserción universitaria. Se ha demostrado que el bajo rendimiento académico, las calificaciones insuficientes y la asistencia deficiente son algunos de los factores más importantes asociados con la deserción universitaria. [1]

En [2] se detalla que la deserción universitaria es un obstáculo importante para la calidad de la educación superior, ya que representa una pérdida de recursos y esfuerzos para las instituciones educativas y para la sociedad en general. La deserción puede ser causada por una variedad de factores, como la falta de apoyo financiero, la falta de preparación académica, la falta de motivación y la falta de orientación vocacional.

En el estudio de [3] se realizó un análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos. Aquí se identificaron los factores que influyeron sobre la deserción de los estudiantes de la carrera de Ingeniería en Informática de la Universidad Gastón Dachary en Argentina.

Y así en [4] se afronta el impacto negativo en la calidad de la educación a nivel macro, ya que puede afectar la capacidad del país para competir en el mercado global y para desarrollar su economía. En un estudio realizado en México, se encontró que la deserción universitaria puede afectar negativamente el desarrollo económico del país, ya que los estudiantes que abandonan sus estudios universitarios tienen menos posibilidades de encontrar empleos bien remunerados y de contribuir al crecimiento económico.

Además, la realización de encuestas y entrevistas puede proporcionar información valiosa sobre las razones detrás de la deserción universitaria. Estos métodos permiten a los estudiantes expresar sus inquietudes y preocupaciones sobre el proceso de aprendizaje y la universidad en general detallados en [5].

En [6] se habla de cómo el uso de modelos predictivos también ha resultado efectivo en la detección temprana de los estudiantes en riesgo de desertar. Estos modelos

utilizan una combinación de datos académicos y personales para identificar patrones y factores de riesgo de deserción.

En general, estos estudios sugieren que las técnicas de minería de datos y aprendizaje automático pueden ser útiles para predecir la deserción universitaria en programas académicos. Además, el rendimiento académico y la tutoría como herramienta de seguimiento son factores importantes que pueden influir en la retención de estudiantes en estos programas.

LÍNEA Y TEMA DE INVESTIGACIÓN

Tabla 1.

Línea de investigación

Línea 3 Bioingeniería	Procesamiento de señales	<ul style="list-style-type: none">• Procesamiento digital de señales e Imágenes• Inteligencia artificial y analítica de datos
	Robótica medica aplicada	<ul style="list-style-type: none">• Desarrollo de equipos biomédicos• Desarrollo de Instrumentación aplicada a la medicina

DESCRIPCIÓN Y FORMULACIÓN DEL PROBLEMA

Teniendo en cuenta el hecho de que todo el sistema de Educación Superior a nivel mundial, y por ende a nivel nacional tiene una problemática bastante grande en consecuencia a la deserción estudiantil; y también considerando que en nuestra universidad los cupos para admisión de nuevos estudiantes son limitados, al momento de un estudiante desertar se le estaría negando la posibilidad de estudio a otras personas, y dichos desertores muy seguramente ejerzan otro tipo de actividades económicas y laborales poco productivas, generando esto un aumento en el círculo de pobreza; además, se produce un estropeo y desaprovechamiento de recursos por parte del gobierno, las instituciones, el estudiante y sus familias.

Puede resultar aún más preocupante, cuando al analizar cifras vemos que solamente uno de cada dos estudiantes que finaliza su bachillerato (educación media) puede acceder e ingresar a la educación superior, y aun ingresando, se dan unos índices de deserción altos. Las deserciones en la educación universitaria y en la técnica y la tecnológica son del 50 % y del 67 %, respectivamente; uno de cada dos estudiantes que ingresa a carreras universitarias deserta, y dos de cada tres que ingresan a carreras técnicas y tecnológicas lo hacen. Esto nos genera a nivel nacional una pérdida de talento considerable dejando en evidencia el hecho de que tenemos un sistema de educación superior altamente ineficiente. Además, también es importante mencionar que muchas personas incurren en obligaciones financieras para buscar un mejor futuro, y la deserción hace muy difícilmente cumplir con los compromisos [7].

La disminución y detección de la deserción, que generaría enormes beneficios al potenciar el talento, requiere acciones efectivas. Es por esto que las IES (Instituciones de Educación Superior) y el Gobierno deben considerar muy relevante esta reducción y adoptar las medidas correctivas necesarias. ¿Hay algún tipo de acción por parte de las IES y el Gobierno para la detección y reducción de la misma? ¿En nuestro claustro universitario se es consciente de esta problemática?, y si lo es, ¿qué tipo de acciones se están llevando a cabo para intentar dar solución?

Por tal motivo, actualmente, dar respuesta y solución a esa problemática, es un motivo de interés tanto para los entes gubernamentales como para las directivas de las Universidades, incluyendo la Universidad Popular del Cesar. Considerando lo anterior, La Vicerrectoría Académica, el Programa de Ingeniería Electrónica de la Facultad de Ingeniería y Tecnológicas plantearon este proyecto de investigación, cuyo objetivo es que a través de un modelo se pueda detectar y predecir las causas que generan y llevan a la deserción estudiantil en los programas de pregrado de la Universidad Popular del Cesar Sede Valledupar, empleando técnicas de *Learning Analytics* y de minería de datos para investigar este fenómeno, poder categorizar a

los estudiantes desertores tomando en consideración aspectos como los socioeconómicos, académicos, institucionales y de admisión. Y tomando como prueba piloto de dicho modelo el programa de Ingeniería Electrónica de la Universidad Popular del Cesar.

JUSTIFICACIÓN

La deserción universitaria en la Universidad Popular del Cesar es un problema que afecta tanto a los estudiantes como a la institución educativa en sí, esta puede tener un alto impacto en la calidad de la educación que ofrece la universidad, así como en su reputación y financiamiento.

Además, la deserción puede afectar la formación de profesionales especializados en áreas de alta demanda, lo que genera un impacto en el desarrollo socioeconómico de la región. Es importante señalar que la deserción es un problema con un impacto negativo en los estudiantes, ya que abandonar los estudios puede resultar en una pérdida de tiempo, esfuerzo y recursos invertidos en su educación. Además, la deserción puede tener consecuencias negativas en su futuro académico y profesional, limitando sus oportunidades de desarrollo socioeconómico y crecimiento personal.

En el contexto específico del programa de ingeniería electrónica de la Universidad Popular del Cesar, la deserción puede llegar a ser especialmente crítica debido a la complejidad y especialización requeridas en esta área. Por lo tanto, la creación de un modelo de detección que permita identificar a los estudiantes en riesgo de deserción y ofrecerles apoyo puede tener un impacto significativo en la retención de estudiantes y en la calidad de la educación ofrecida por la universidad. Además, este modelo puede ser un recurso valioso para replicarlo en otros programas académicos e incluso en otras instituciones educativas que enfrenten problemas similares en cuanto a la deserción universitaria.

OBJETIVO GENERAL

Proponer un modelo que pueda predecir la deserción temprana en el programa de Ingeniería Electrónica de la Universidad Popular del Cesar utilizando métodos de minería de datos y técnica de *Learning Analytics*.

OBJETIVOS ESPECÍFICOS

- ✓ Definir las categorías de variables y las variables asociadas a la deserción.
- ✓ Definir herramientas de minería de datos, descubrir patrones y relaciones validas en el conjunto de datos.
- ✓ Diseñar el prototipo del software con base al modelo de datos.

MARCOS REFERENCIALES

Poder definir y conceptualizar de forma estandarizada la deserción es una tarea ardua y compleja, debido a que hay múltiples perspectivas en las que se puede centrar. Con base en esto hay varios estudios que analizan la deserción en diferentes etapas de la educación, para nuestro caso, vamos enfocados en la superior. En uno de los tantos estudios (Elder, 1985) se contempla la deserción usando el enfoque de historia de vida, el cual se centra en la secuencia de eventos y transiciones. En este planteamiento se tiene en cuenta el historial educativo del individuo, y otros factores desde los individuales, académicos, socioeconómicos e institucionales desde el momento en que ingresa al nivel superior.

Es importante notar el hecho de que, por medio de diversos estudios se ha podido evidenciar que, para una misma persona, el suceso o acto abandonar sus estudios no es constante a través del tiempo ni la edad, y a su vez, los efectos de los diferentes factores tampoco son constantes. Hay una variación que permite trabajar desde dicho enfoque, considerando no solo el tiempo en la ocurrencia del evento mismo (deserción) sino también cómo cambia en el tiempo el efecto de cada factor determinante. [8].

ANTECEDENTES

Siguiendo la idea del hecho de que definir de manera única la deserción, de tal modo que dicha definición abarque en su totalidad la complejidad de este fenómeno. Tinto (1989) nos muestra que es posible analizar la definición de la deserción estudiantil desde varias perspectivas y de acuerdo con los diferentes tipos de abandono. Estos planteamientos están relacionados de manera directa con las partes que intervienen en el proceso educativo, como son los estudiantes, los funcionarios de las instituciones de educación superior y los responsables de la política nacional de educación. Partiendo de ello, la deserción se puede observar, analizar e investigar desde tres puntos de vista: individual, institucional y estatal o nacional.

Desde el punto de vista individual, la deserción implica que hay que reconocer que la forma o el sentido que un estudiante asigna al acto de abandono son completamente diferentes a los que alguien más (un observador) atribuye a ese mismo comportamiento. Se podría decir entonces que desertar denota un fallo que lleva a un sentir de frustración en querer completar determinado rumbo de acciones

para alcanzar una meta anhelada; razón principal por la que ingresó a una institución de educación superior. [9].

El término deserción puede estar mejor utilizado cuando integra una comunidad de intereses (Tinto, 1989). En esta se abarca tanto al estudiante como a los observadores externos que tienen como objetivo la disminución de la deserción estudiantil, siendo estos los mismos funcionarios de la institución (docentes, directivos...). Analizando desde el punto de vista de éstos, la deserción del individuo es vista como un fracaso de la institución debido a que no ayudó al estudiante a cumplir lo que inicialmente se había propuesto al ingresar en la institución de educación superior. Una forma de poder integrar una definición de la deserción que incluya todos los intereses (o la mayoría) es realizar el estudio desde una perspectiva institucional, la cual comprende la comunidad de intereses involucrados en un proceso de deserción. Se puede emplear como escala de la deserción la cantidad de estudiantes que abandonan en un período determinado una institución de educación superior, antes de haber obtenido el título correspondiente (ICFES, 2002). Este panorama permite la facultad de tener un marco conceptual donde todos los individuos que se retiran de una institución de educación superior pueden, considerando las razones del abandono, ser catalogados como desertores (Tinto, 1989; Castaño et. al., 2004). [10]

El enfoque estatal o nacional de la definición de deserción tiene como fundamento la estructuración educativa del país (ICFES, 2002). En este caso, el fenómeno de deserción contempla el abandono del estudiante de sistema educativo en general. Por ende, no se puede considerar que todos los abandonos de instituciones corresponden a deserciones del sistema, sino que algunos pueden definirse como transferencias entre instituciones educativas o cambios al interior del sistema. [11]. Por consiguiente, solamente las veces en que el abandono estudiantil denotan un abandono total del sistema de educación superior son tenidas en cuenta como deserción, y aquellas que son transferencias entre instituciones corresponderían a movilidad dentro del sistema. Esta definición sitúa las medidas de políticas públicas en la vigilancia y armonización de los flujos internos de alumnos, así como en la reducción, si es posible, de la pérdida de estudiantes de las instituciones educativas del sistema nacional de educación superior. [12].

MARCO TEÓRICO

La deserción estudiantil representa en una gran medida una falla e inconveniente estructural del sistema educativo en general. Las estadísticas evidencian cómo este problema afecta de manera directa la permanencia y la graduación de los estudiantes. En un ámbito internacional, enfocados en la educación universitaria, que es el tópico de este estudio, según los datos presentados en el informe *Education at the glance* (2016), la deserción universitaria en los países pertenecientes a la OCDE alcanza un 31 %. Para el caso del Espacio Europeo de Educación Superior (EEES), conformado por 47 países, la tasa de deserción varía de un 20 % a un 55 %. En el ámbito latinoamericano las mediciones de deserción van de un 8 % a un 48 %. En el ámbito colombiano, la deserción se ubica en un 48.8 %. [2]. Estos datos nos muestran un claro y evidente panorama en el que la deserción es un problema que afecta a todos los sistemas educativos a escala global.

En Colombia tenemos un Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior – SPADIES, este se encarga de recopilar, organizar y clasificar los datos suministrados por las instituciones, los cuales permiten hacer un seguimiento de la deserción, los diferentes factores que generan el fenómeno y de las estrategias implementadas para intervenirlo. Dicho sistema de información, el cual se encuentra adscrito al Ministerio de Educación Nacional de Colombia, define el fenómeno de la deserción como:

“Estado de un estudiante que de manera voluntaria o forzosa no registra matrícula por dos o más períodos académicos consecutivos del programa en el que se matriculó; y no se encuentra como graduado, o retirado por motivos disciplinarios. La deserción es el resultado del efecto de diferentes factores como individuales, académicos, institucionales, y socioeconómicos. (SPADIES, Glosario, 2020).” [2]

Algo a tener en cuenta, es que, al ser un tema tan amplio tanto a nivel nacional como global, existen varias definiciones para el concepto de deserción, según Tinto [13] define la deserción como una situación a la que se enfrenta un estudiante cuando aspira y no logra concluir su proyecto educativo.

Muchas discusiones se han dado alrededor de la deserción estudiantil y a pesar de que no se tiene una definición consensuada y global sobre este fenómeno [14] existe una coincidencia en cuanto a que este debe ser explicado por diferentes grupos de variables las cuales en términos generales se han agrupado en 4 categorías principales: socioeconómicas, individuales, institucionales y académicas

[15]. Dentro de cada una de estas categorías existen una serie de elementos que interaccionan en el estudiante y que, para cada región, momento económico y político del país son cambiantes [16]; por este motivo se hace complejo el estudio de los factores que influyen en la deserción [17] ya que se debe abarcar no sólo una variedad de perspectivas, sino que también hay que tener en cuenta que existen diferentes tipos de abandono, los cuales se ubican dentro de dos categorías.

La primera categoría establece el abandono en el tiempo, el cual a su vez puede ser precoz, que se presenta cuando el estudiante no alcanza a iniciar su proceso de formación; temprana que se presenta cuando el estudiante deserta en los primeros años de estudio y tardía que indica que el abandono se presenta cuando el estudiante llevaba un avance considerable en su proceso formativo.

La segunda categoría es el tipo de deserción que se considera en el espacio; en esta clasificación se presentan la deserción interna o del programa, la deserción institucional que es cuando el estudiante abandona la institución educativa y finalmente la deserción del sistema en la cual se considera como desertor a aquel que abandona definitivamente el proceso de formación [18],[19] Hay que tener en cuenta que el estudio de la deserción debe realizarse desde diferentes acepciones de acuerdo con el investigador y la situación del ambiente en el que está el estudiante, es por esto por lo que las variables para incluir en el estudio dependen del punto de vista desde el cual se haga el análisis; esto es, individual, institucional y estatal o nacional [18].

A rasgos generales, y haciendo una recopilación de algunas de las definiciones y causas de la deserción, podemos establecer que los alumnos dimiten o abandonan el sistema educativo por múltiples razones o circunstancias, que son consecuentes de diversos factores, dentro de los que podemos encontrar:

Factores socioculturales: se encuentran relacionados a determinadas limitaciones familiares, culturales o sociales, que generan una narrativa en la que, dependiendo la procedencia étnica, vuelven inaccesible la educación superior.

Factores estructurales: las condiciones económicas prolongas las deficiencias, fallas y carencias, que hacen tomar más fuerza a la idea de que la educación es un beneficio solo para la élite.

Factores políticos: la falta de interés político para un correcto financiamiento de una educación a nivel de calidad, acceso, y cobertura, como un eje principal en los planes de desarrollo y las políticas vinculadas con los sectores productivo y educativo de un país.

Factores institucionales: están basados a la disposición de la institución educativa y en general al sistema educativo, en cuanto a modelos de gestión y pedagógicos

que vayan al pie de las nuevas prácticas de enseñanza y aprendizaje, así mismo como la oferta de programas pertinentes a las necesidades del mundo global, etc.

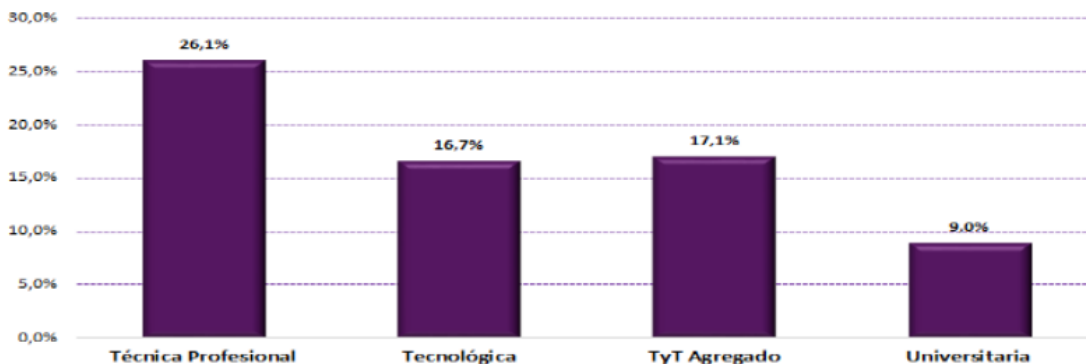
Factores personales: están fuertemente relacionados con aspectos emocionales, afectivos, fuerza mayor, salud, entre otros; que pueden inducir y llevar de manera directa a un abandono del proceso académico.

Factores de aprendizaje: planeamiento y uso de sistemas inadecuados, que resultan obsoletos en cuanto a métodos de enseñanza y aprendizaje, los cuales no tienen en cuenta las investigaciones y estudios más recientes en cuanto a estilos y ritmos de aprendizaje, necesidades educativas especiales, inteligencias múltiples; y cada una de las condiciones necesarias que requieren innovación, transformación y cambios en función de las necesidades de formación del estudiante.

Unos datos que evidencian dicha problemática, tomados de la estadística reportada por el Sistema para la Prevención de la Deserción en las Instituciones de Educación Superior -SPADIES-, Figura 1, para el año 2016, indican que en Colombia se han logrado avances en reducir de manera significativa los índices de deserción universitaria, respecto a los informes entregados por entes internacionales en 2015. Según estos datos la formación técnica profesional que representa el índice más alto de deserción en la formación superior en Colombia con un 26.1%, y la disminución más significativa con un 9.0% corresponde a la educación universitaria.[2]

Figura 1.

Tasa de deserción en niveles de formación superior en Colombia (2016)



*Nota. Deserción en programas de formación superior en Colombia (2016).
Fuente: Ministerio de Educación Nacional de Colombia – SPADIES.*

La tecnología ha transformado fundamentalmente el mundo en el que vivimos, y el sector educativo no ha sido ajeno a esta transformación. Hoy en día, los educadores y estudiantes tienen acceso a una amplia variedad de herramientas tecnológicas que pueden mejorar significativamente el proceso de aprendizaje. Por ejemplo, los programas de simulación y las aplicaciones interactivas permiten a los estudiantes experimentar con conceptos complejos de una manera segura y sin riesgos. Las plataformas de aprendizaje en línea y los recursos digitales proporcionan acceso instantáneo a una gran cantidad de información, lo que puede facilitar el aprendizaje autodirigido y el trabajo colaborativo. [20]

Sin embargo, aunque la tecnología puede tener muchos beneficios para la educación, también puede presentar desafíos significativos. El uso excesivo de dispositivos electrónicos puede distraer a los estudiantes y disminuir su capacidad de atención. Además, algunos estudiantes pueden tener dificultades para adaptarse a las nuevas tecnologías o pueden carecer de acceso a ellas, lo que puede generar desigualdades en el aprendizaje. [21]

Es importante que los educadores reconozcan tanto los beneficios como los desafíos de la tecnología y busquen formas efectivas de integrarla en su enseñanza. Esto puede requerir un cambio en la forma en que se aborda la educación, con un mayor énfasis en la enseñanza de habilidades digitales y la incorporación de nuevas herramientas tecnológicas en el aula. Al hacerlo, podemos aprovechar al máximo el potencial de la tecnología para mejorar la educación y preparar a los estudiantes para el mundo en constante evolución de la tecnología. [22]

MARCO HISTÓRICO

La problemática de la deserción estudiantil no es nueva y ha sido objeto de estudio durante décadas. En los últimos años, la deserción universitaria se ha convertido en un problema global, que afecta tanto a países desarrollados como a países en desarrollo. En Latinoamérica, la deserción universitaria ha sido un problema crónico que afecta significativamente a la calidad de la educación y al desarrollo socioeconómico de la región. [23]

La preocupación por la deserción universitaria y su impacto en la calidad de la educación y el desarrollo socioeconómico se remonta a la década de 1960, cuando se comenzaron a desarrollar estrategias para mejorar la retención de los estudiantes en las universidades.

En los años 70, se realizaron estudios sobre la deserción universitaria que concluyeron que el problema era multifactorial y que involucraba no solo aspectos académicos, sino también sociales, económicos y psicológicos. A partir de entonces, se comenzaron a desarrollar estrategias para mejorar la retención de los estudiantes, incluyendo programas de apoyo académico, asesoramiento psicológico y financiero, y programas de orientación y tutoría.

Con la llegada del siglo XXI, el avance de la tecnología ha permitido la creación de herramientas y técnicas más sofisticadas para abordar el problema de la deserción universitaria. La aplicación de métodos de análisis de datos y técnicas de *Learning Analytics*, por ejemplo, ha permitido a las instituciones educativas recopilar y analizar grandes cantidades de información sobre los estudiantes, lo que a su vez ha permitido la identificación de patrones y tendencias que pueden ayudar a predecir el éxito o fracaso académico de los estudiantes.

Según estadísticas del Ministerio de Educación Nacional, la tasa de deserción en la educación superior es del 19,8%. Esto significa que casi 2 de cada 10 estudiantes que inician una carrera universitaria no la terminan. La deserción universitaria tiene un impacto negativo en la economía y en la sociedad, ya que reduce la cantidad de profesionales calificados y limita las oportunidades de desarrollo para los jóvenes. [23]

MARCO CONCEPTUAL

Educación: Se puede denotar como el conjunto de estrategias destinadas a desarrollar la capacidad cognitiva, ética, moral y habilidades en un individuo. Predeterminadas por el entorno social, económico, cultural y ambiental que lo rodean. Donde se involucran directamente una serie de aspectos que van en pro de la transmisión de conocimientos hacia una persona. [24]

Sistema educativo: Grupo de elementos relacionados entre sí para brindar educación a través de diferentes protocolos que llevan a la transmisión de conocimientos, hacia los individuos. Es importante destacar que, en el sistema educativo, encontramos entidades públicas y privadas, que prestan el acceso a la educación a las personas, siendo este un derecho fundamental de todos los seres humanos. [24]

Deserción: Es la interrupción voluntaria o involuntaria de un proceso. Principalmente esta definición es implementada en el ámbito educativo, cuando un alumno se retira de sus estudios sin haberlos finalizado. [24]

Deserción estudiantil: La deserción estudiantil son eventos que se presentan en todos los escenarios educativos cuando el estudiante o alumno se retira completamente de su proceso de aprendizaje; esto aplica en los diferentes niveles educativos, desde la primaria pasando por la secundaria básica, en la secundaria media o bachillerato, y en el ciclo profesional (técnico, tecnólogo y profesional universitario). Esta decisión puede ser personal o forzosa, en la que influyen diversos aspectos y motivos. [24]

Deserción universitaria: Suceso en el que un estudiante universitario abandona totalmente su carrera.

Bajo rendimiento académico: Hace referencia a los resultados poco favorables y negativos que obtiene el estudiante en las diversas áreas de conocimiento que está recibiendo, incluso en las temáticas más básicas.

Semestre inactivo: Período en que el estudiante interrumpe sus estudios de forma parcial.

Semestre de deserción: Período académico en el cual el estudiante abandona el proceso de aprendizaje dentro del aula.

Aprendizaje analítico: Es un enfoque para el análisis de datos que utiliza técnicas estadísticas y de aprendizaje automático para identificar patrones y tendencias en grandes conjuntos de datos. Se aplica en diversos campos, como el marketing, la educación y la salud, para comprender mejor el comportamiento humano y mejorar la toma de decisiones.

Ciencia de datos: Es una disciplina que se enfoca en la extracción de información útil y conocimiento a partir de grandes conjuntos de datos. La ciencia de datos utiliza técnicas estadísticas, matemáticas, de visualización y de modelado para analizar los datos y descubrir patrones, tendencias y relaciones ocultas en ellos. En el contexto del modelo de detección de la deserción universitaria, la ciencia de datos se utilizaría para analizar los datos de los estudiantes y desarrollar un modelo de detección temprana de la deserción.

Learning Analytics: Es un conjunto de técnicas y herramientas que se utilizan para analizar el rendimiento de los estudiantes y mejorar su aprendizaje. Esta utiliza datos de diferentes fuentes, como los sistemas de gestión del aprendizaje y los sistemas de evaluación para identificar patrones de comportamiento de los estudiantes y determinar qué factores pueden afectar su éxito académico. En el contexto del modelo de detección de la deserción universitaria, se puede usar para analizar el rendimiento de los estudiantes y poder desarrollar estrategias de apoyo personalizadas para aquellos que tienen un mayor riesgo de desertar.

Detección temprana de la deserción: Es el proceso de identificar a los estudiantes que tienen un mayor riesgo de abandonar sus estudios antes de que ocurra la deserción. En el contexto del modelo de detección de la deserción universitaria, la detección temprana de la deserción se lograría mediante el uso de métodos de análisis de datos y *Learning Analytics* para identificar patrones de comportamiento de los estudiantes que indican un mayor riesgo de deserción. Con esta información, se pueden desarrollar estrategias de apoyo personalizadas para mejorar la retención de los estudiantes y reducir la tasa de deserción.

Minería de datos: Consiste en la recopilación de técnicas asistidas por computadora con el fin de detectar la información procesable de enormes conjuntos de datos, y además explorarla. Emplea el análisis matemático por medio del cual se pueden deducir patrones, relaciones ocultas y tendencias existentes en los datos. Generalmente dichos patrones no es posible detectarlos a través de la exploración tradicional de los datos debido la complejidad de los mismos, o por la cantidad. Podemos decir que la minería de datos transforma datos en bruto en conocimiento práctico, el cual puede ser utilizado para múltiples propósitos. [26]

La minería de datos, Figura 2, se encuentra cimentada en tres disciplinas científicas entrelazadas: estadística (el estudio numérico de relaciones de datos), inteligencia artificial (inteligencia similar a la humana exhibida por software y/o máquinas) y *Machine Learning* (algoritmos que pueden aprender de datos para hacer predicciones). [27]

Figura 2.

Bases de la minería de datos



Nota. Correlación minería de datos con la estadística, IA y Machine Learning.

Dichos patrones y tendencias se pueden seleccionar y definir como un *modelo de minería de datos*. Los modelos de minería de datos se pueden aplicar en escenarios como los siguientes desde un punto de vista empresarial:

- **Pronóstico:** cálculo de las ventas y predicción de las cargas del servidor o del tiempo de inactividad del servidor.
- **Riesgo y probabilidad:** elección de los mejores clientes para la distribución de correo directo, determinación del punto de equilibrio probable para los escenarios de riesgo, y asignación de probabilidades a diagnósticos y otros resultados.
- **Recomendaciones:** determinación de los productos que se pueden vender juntos y generación de recomendaciones.

- **Búsqueda de secuencias:** análisis de los artículos que los clientes han introducido en el carrito de la compra y predicción de posibles eventos.
- **Agrupación:** distribución de clientes o eventos en grupos de elementos relacionados, y análisis y predicción de afinidades.

En rasgos generales podemos decir que la minería de datos permite:

- Filtrar todo el ruido caótico y repetitivo en los datos.
- Entender qué es relevante y luego hacer un buen uso de esa información para evaluar resultados probables.
- Acelerar el ritmo de la toma de decisiones informadas.

Para el campo en el que estamos llevando la investigación el cual es la Educación, la minería de datos se puede emplear de diversas formas, por ejemplo, las instituciones educativas pueden emplear algoritmos de minería de datos para evaluar a los estudiantes, personalizar las lecciones y hacer que el aprendizaje sea más lúdico. Por medio de vistas unificadas basadas en datos del progreso de los estudiantes, los docentes pueden predecir el desempeño de sus alumnos antes de que pongan un pie en el salón de clases – y desarrollar estrategias de intervención para mantenerlos en curso. La minería de datos ayuda a los docentes a través de una visión unificada y basada en los datos de los estudiantes, a predecir niveles de logro y a detectar estudiantes o grupos de estudiantes que necesitan atención extra y poder apoyarlos de una mejor manera. Gracias a estas ventajas y aplicabilidad de la minería de datos sustentamos la idea de la creación e implementación del modelo de detección temprana de la deserción estudiantil en la Universidad Popular del Cesar, teniendo como fin y objetivo principal ser una herramienta en pro del beneficio de los estudiantes, y por ende para la universidad. [28]

PostgreSQL: o también conocido como Postgres, es un sistema de gestión de bases de datos relacionales (RDBMS) libre y de código abierto (*Open Source*) que hace énfasis en la extensibilidad y el cumplimiento de SQL.

Es gratuito y libre, además de que hoy nos ofrece una gran cantidad de opciones avanzadas. De hecho, es considerado el motor de base de datos más avanzado en la actualidad. [29]

Postgres se basa en el clásico **modelo cliente-servidor**: el componente central que hace de servidor, con el nombre de “*postmaster*”, administra todos los archivos de la BD, así como todas las conexiones que se establecen con el servidor de la BD para las funciones de comunicación (input/output). Para establecer estas

conexiones solo se necesita un programa-cliente, si bien el paquete de software de PostgreSQL ya integra de forma nativa el cliente `psql` para operarlo desde la terminal de líneas de comando. También podría optarse por una aplicación con **interfaz gráfica de usuario** como `pgAdmin` o `phpPgAdmin`, fáciles de instalar y manejar. En las páginas web interactivas es el servidor web el que se encarga de hacer de cliente. [30]

Dentro de las ventajas de PostgreSQL podemos mencionar las siguientes:

- ✓ *Open source*
- ✓ Altamente ampliable
- ✓ En gran parte conforme con el estándar SQL
- ✓ Permite procesar tipos complejos de datos (p. ej., datos geográficos)
- ✓ Búsqueda de texto completo flexible
- ✓ Pueden crearse funciones propias, *triggers*, tipos de datos, etc.
- ✓ Amplia compatibilidad con varios lenguajes (Python, Java, Perl, PHP, C, C++, etc.)
- ✓ Soporta JSON
- ✓ Multiplataforma

Como desventajas o inconvenientes podemos mencionar:

- No está disponible por defecto en todos los *hosters*
- La documentación es mejorable y está solo disponible en inglés
- La velocidad de lectura es menor que en otros gestores

MARCO LEGAL

Sistema de Educación en Colombia

La **Ley 115 de 1994** (Ley General de Educación), de conformidad con el artículo 67 de la Constitución Política, define y desarrolla la organización y la prestación de la educación formal en sus niveles de preescolar, básica (primaria y secundaria) y media, no formal e informal.

La educación superior, por su parte, es reglamentada por la **Ley 30 de 1992** que define el carácter y autonomía de las Instituciones de Educación Superior - IES-, el objeto de los programas académicos y los procedimientos de fomento, inspección y vigilancia de la enseñanza. [25]

Estas dos leyes indican los principios constitucionales sobre el derecho a la educación que tiene toda persona, por su parte, las condiciones de calidad que debe tener la educación se establecen mediante el **Decreto 2566 de 2003** y la **Ley 1188 de 2008**.

El **Decreto 2566 de 2003** reglamentó las condiciones de calidad y demás requisitos para el ofrecimiento y desarrollo de programas académicos de educación superior, norma que fue derogada con la **Ley 1188 de 2008** que estableció de forma obligatoria las condiciones de calidad para obtener el registro calificado de un programa académico, para lo cual las Instituciones de Educación Superior, además de demostrar el cumplimiento de condiciones de calidad de los programas, deben demostrar ciertas condiciones de calidad de carácter institucional. [25]

Esta normatividad se complementa con la **Ley 749 de 2002** que organiza el servicio público de la educación superior en las modalidades de formación técnica profesional y tecnológica, amplía la definición de las instituciones técnicas y tecnológicas, hace énfasis en lo que respecta a los ciclos 38 propedéuticos de formación, establece la posibilidad de transferencia de los estudiantes y de articulación con la media técnica. [25]

Ley 30 de 1992- Servicio Público de Educación Superior. Expresa normas por medio de las cuales se reglamenta la organización del servicio público de la Educación Superior. (El Congreso de Colombia, 28 de diciembre de 1992).

Decreto 1403 de 1993- Reglamentación de Ley 30 de 1992. Establece que mientras se dictaminan los requisitos para la creación y funcionamiento de los 21 programas académicos de pregrado que puedan ofrecer las instituciones de educación superior, estas deberán presentar al Ministerio de Educación Nacional por conducto del Instituto Colombiano para el Fomento de la Educación Superior (ICFES), con el fin de garantizar el cumplimiento de los fines de la educación, la información referida al correspondiente programa. Así también regula lo referido a la autorización de la oferta de programas de Maestría, Doctorado y Postdoctorado, de conformidad con la referida Ley. (Dado el 21 de Julio de 1993 y Publicado en el Diario Oficial N° 41.476 del 5 de agosto de 1994. [25]

Ley 115 de 1994- Ley General de Educación. Ordena la organización del Sistema Educativo General Colombiano. Esto, establece normas generales para regular el Servicio Público de la Educación que cumple una función social acorde con las necesidades e intereses de las personas, de la familia y de la sociedad. Respecto a la Educación Superior, señala que ésta es regulada por ley especial, excepto lo dispuesto en la presente Ley. "Excepto en lo dispuesto en la Ley 115 de 1994, sobre Educación Tecnológica que había sido omitida en la Ley 30 de 1992. Ver Artículo 213 de la Ley 115. [25]

DISEÑO METODOLÓGICO

El interés en este caso está centrado en el rendimiento académico de los estudiantes de pregrado al término de su primer semestre (cohorte 2022-2), por lo importante que resulta para la Universidad contar con información para: la identificación temprana de factores que incidan en la permanencia, la definición de acciones que conduzcan al mejoramiento del rendimiento académico, así como también para desarrollar políticas públicas y curriculares.

Los datos estadísticos que se presentan son en realidad los valores de los correspondientes parámetros, dado que están calculados sobre los registros de la muestra obtenida de los estudiantes de pregrado de la Facultad de Ingeniería en el programa de Ingeniería Electrónica en el semestre 2022-2. Por tal razón, no se hace inferencia estadística y, de la diferencia entre parámetros, se dice si es significativa o no en términos prácticos. El procesamiento se realiza sobre los registros de 318 estudiantes con información acerca de: sexo, programa académico, edad, modalidad (presencial, virtual), estrato socioeconómico, promedio semestre, condición académica, modo de admisión y naturaleza del colegio, entre otras. La estadística descriptiva, la clasificación cruzada y la comparación entre niveles de las variables, son las técnicas estadísticas empleadas para la generación de resultados, con los propósitos de avanzar en la descripción de las variables asociadas a la vida académica y, de explorar relaciones de asociación con la variable rendimiento y condición académicos.

Etapas de selección

El objetivo de esta etapa es obtener las fuentes de datos internas y externas que sirven de base para el proceso de minería de datos. Como fuente interna se seleccionó la encuesta realizada por los estudiantes del programa de ingeniería electrónica de la Universidad Popular del Cesar, compuesta por información personal y académica de 318 estudiantes activos, nos está faltando la de egresados y retirados, acumulada en un periodo de un semestre. Como fuente externa se seleccionó el archivo con la información de los colegios de educación secundaria del país, que se obtuvo con el Ministerio de Educación Nacional de Colombia, con el fin de obtener algunas características de los colegios entre las que se cuentan: si el colegio es público o privado, el tipo de calendario A o B, la jornada de estudio, si el colegio es urbano o rural. Estas características son importantes para incluirlas

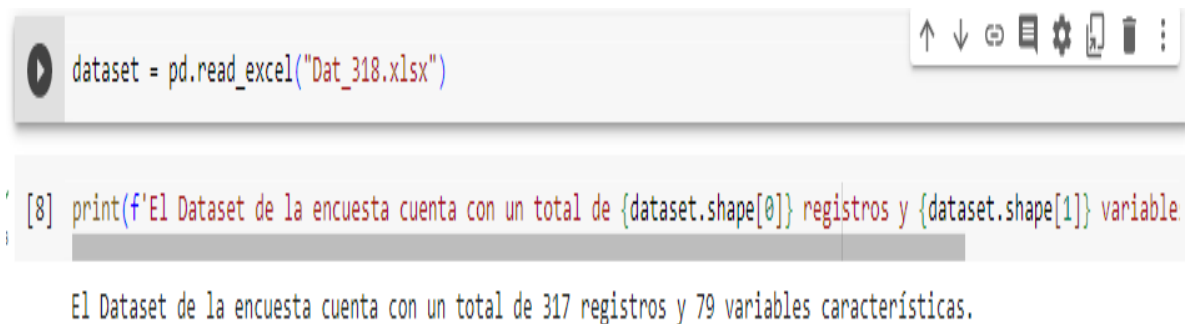
dentro de los atributos relevantes que permitan determinar el perfil de los estudiantes que abandonan sus estudios o tienen bajo rendimiento académico.

Etapa de pre procesamiento de datos

Para la ejecución de todas las fases, es necesario tener en cuenta la consideración de un conjunto específico de variables que influyen en mayor medida al comportamiento desertor, al resultado final que se obtendrá en la predicción. Asimismo, en esta fase se realiza la extracción de datos desde la fuente original y se realiza un proceso de limpieza y transformación de los datos. Finalmente, se cargan los datos finales para su análisis. El objetivo de esta etapa es obtener datos limpios, datos sin valores nulos o anómalos que permitan obtener patrones de calidad. Por medio de consultas sobre la información recogida de la encuesta se analizó minuciosamente la calidad de los datos contenidos en cada uno de los atributos de las columnas. Figura 3, los atributos seleccionados de las diferentes columnas en su gran mayoría no contenían valores nulos ni anómalos (*outliers*), pero en aquellos casos que se presentaban, estos fueron reemplazados utilizando técnicas estadísticas tales como la media y la moda o derivando sus valores a través de otros de acuerdo con el dominio de donde provienen los datos.

Figura 3.

Proceso de carga y análisis de datos en Python



```
dataset = pd.read_excel("Dat_318.xlsx")

[8] print(f'El Dataset de la encuesta cuenta con un total de {dataset.shape[0]} registros y {dataset.shape[1]} variables características.')

El Dataset de la encuesta cuenta con un total de 317 registros y 79 variables características.
```

*Nota. Registros y variables totales obtenidas a través de la encuesta realizada.
(Fuente: Propia, derivada de la investigación).*

Etapa de depuración de variables

A partir de la identificación y análisis de cada una de las variables en la presente y pasadas iteraciones, se procede a depurar a aquellas que no se requieren para el modelo, Figura 4, dado a que son redundantes, (están altamente correlacionadas con otras variables del Datasheet que se mantiene) su nivel de importancia es muy bajo y/o contienen valores errados. En la etapa de depuración se buscan características útiles para representar los datos según la meta del proceso de minería de datos. Se utilizan métodos de reducción de dimensiones o de transformación para disminuir el número efectivo de variables en consideración o para encontrar representaciones invariantes de los datos “según (Fayyad et al., 1996)”. En esta etapa se construyó el conjunto de datos Dat_318.xlsx. Se discretizaron los atributos con valores continuos, es decir, se transformaron los valores numéricos en valores discretos o nominales, con el fin de disminuir el número de valores distintos de estos atributos. La discretización de estos atributos se realizó teniendo en cuenta el hecho de que las diferencias en ciertas zonas del rango de valores eran más importantes que otras, lo que resultaba en intervalos más significativos que otros dentro del conjunto de valores de cada atributo. La variable a predecir en la cual se tiene como propósito clasificar a aquellos posibles estudiantes desertores y no desertores, se encuentra implícita en la variable 'ESTADO_DEFINITIVO' del conjunto de datos en la cual, los registros iguales a 1 corresponden a los no desertores y los registros iguales a 2 corresponden a los desertores a final de cada semestre. Por lo anterior, se procede a crear una nueva columna llamada 'DESERTOR' en el conjunto de datos a partir de la columna 'ESTADO_DEFINITIVO'

Figura 4.

Depuración de variables

	MID_NUM_CONTACTO	MID_CREDITOS_APROBADOS	MID_CREDITOS_MATRIC	MID_DOCUMENTO	MID_EDAD	MID_ESTRATO
MID_NUM_CONTACTO	1.000000	0.096432	-0.096021	-0.280259	0.158322	-0.182762
MID_CREDITOS_APROBADOS	0.096432	1.000000	0.078450	0.018216	0.477364	-0.055085
MID_CREDITOS_MATRIC	-0.096021	0.078450	1.000000	-0.044322	0.228231	0.001588
MID_DOCUMENTO	-0.280259	0.018216	-0.044322	1.000000	0.090373	0.055692
MID_EDAD	0.158322	0.477364	0.228231	0.090373	1.000000	-0.122875
MID_ESTRATO	-0.182762	-0.055085	0.001588	0.055692	-0.122875	1.000000
MID_MATERIAS_CANCELADAS	-0.064589	0.200854	0.105313	-0.028275	0.143458	0.093210
MID_PUESTO_ICFES	0.044898	0.143847	0.241313	0.104177	0.473437	-0.123806
MID_PUNTAJE_ICFES	0.074410	0.057221	-0.040399	-0.005848	0.075316	0.064369
Semestres_pagos_del_estudiante	0.096028	0.611777	0.116167	0.090342	0.717534	-0.058898
MID_CREDITOS_APROBADOS2	0.150736	0.149778	0.260729	-0.432904	0.283899	-0.089170
creditos_Aprobados_Segundo_Semestre_Estudiante	0.072602	0.788927	0.052973	0.087179	-0.022742	0.088304
MID_CREDITOS_MATRIC2	-0.236376	-0.004524	0.491042	-0.146680	0.333939	0.014065
creditos_Matriculados_Segundo_Semestre_Estudiante	-0.144480	0.392289	0.316054	-0.265755	0.376920	-0.229423
MID_CREDITOS_NO_APROBADOS	0.012948	0.460351	0.169697	0.102373	0.575821	-0.090671

Nota. Correlación de variables. (Fuente: Propia, derivada de la investigación).

Etapas de minería de datos

El objetivo de esta etapa es la búsqueda y descubrimiento de patrones insospechados y de interés utilizando diferentes técnicas de descubrimiento tales como clasificación, *clustering*, patrones secuenciales, asociación, entre otras. Para el descubrimiento de patrones de bajo rendimiento académico y deserción estudiantil se desarrollarán tareas de clasificación y asociación.

Etapas de Interpretación y evaluación de resultados

En esta etapa se interpretan los patrones descubiertos y posiblemente se retorna a los anteriores pasos o etapas para posteriores iteraciones. Esta etapa puede incluir la visualización de los patrones extraídos, la remoción de los patrones redundantes o irrelevantes y la traducción de los patrones útiles en términos que sean entendibles para el usuario.

MÉTODO DE INVESTIGACIÓN

Para el desarrollo del modelo de la detección temprana de la deserción en la Universidad Popular del Cesar, en la que se empleará como plan piloto el programa de ingeniería electrónica. Se desarrolló a través de métodos lógicos y empíricos, debido a que se tuvo que tomar como guías y modelos a seguir los procesos empleados por otras universidades, así como los dictaminados por el Ministerio de Educación Nacional (*MEN*); y al mismo tiempo, por medio del conocimiento directo y propio obtenido a través de los mecanismos de recolección de datos, que en nuestro caso fue la encuesta realizada (<https://script.google.com/a/macros/unicesar.edu.co/s/AKfycbxZlqGuOQeK2EGc4FZ6xVL21Uc9HC9SveIF0gZn8V8/dev>).

Teniendo en cuenta nuestros objetivos propuestos, se procedió de la siguiente manera:

- ✓ Definir las categorías de variables y las variables asociadas a la deserción.

Con base a lo aprendido previamente (marco teórico) tenemos que se puede dividir en 3 tipos de factores principales los que influyen en la deserción (personales, socioeconómicos y académicos), con base a estos y por medio de la encuesta pudimos obtener e identificar todas y cada una de las variables que están asociadas a la deserción, y posteriormente se hizo la respectiva caracterización de las mismas.

TABLAS CON LOS DATOS:

Tabla 2.

Caracterización estudiante

VARIABLE	TIPO
ID_NOMBRE	CHAR (35)
ID_TIPO_DOC	CATEGÓRICA
ID_DOCUMENTO	INT → CHAR (12)
ID_EDAD	INT → CHAR (2)
ID_NCONTACTO	INT
ID_GENERO	CATEGÓRICA
ID_VICTIMA_CONFLICTO	CATEGÓRICA
ID_ESTADO_CIVIL	CATEGÓRICA
ID_DPTO_NACIMIENTO	CHAR (50)
ID_MUN_NACIMIENTO	CHAR (50)
ID_TIPO_DISCAPACIDAD	CATEGÓRICA
ID_ETNIA	CATEGÓRICA
ID_CORREO_INST	CHAR (30)
ID_CABEZA_HOGAR	CATEGÓRICA
¿TIENE HIJOS?	
ID_CANTIDAD_HIJOS	→ INT
ID_EDAD_MAYOR	→ INT
ID_EDAD_MENOR	→ INT
ID_OCUPACION_MAMÁ	CATEGÓRICA
ID_OCUPACION_PAPÁ	CATEGÓRICA
ID_CANT_HERMANOS	→ INT
ID_POSICION_HERMANO	→ INT
ID_INTEGRANTES	→ INT
ID_TALENTO	CATEGÓRICA
ID_EPS	CATEGÓRICA
ID_SISBEN	CHAR (3)
ID_LENTES	CATEGÓRICA

*Nota. Variables con base a la información personal.
 (Fuente: Propia, derivada de la investigación).*

Tabla 3.

Categorías estudiante

VARIABLE	CATEGORÍAS	VALOR
ID_TIPO_DOC	T. IDENTIDAD → CHAR (10)	0
	C.C → CHAR (10)	1
	PASAPORTE → CHAR (10)	2
	C. EXTRANJERIA → CHAR (10)	3

ID_GENERO	MASCULINO → INT	0
	FEMENINO → INT	1
	OTRO → INT	2
ID_VICTIMA_CONFLICTO	SI → INT	0
	NO → INT	1
ID_ESTADO_CIVIL	SOLTERO → INT	0
	CASADO → INT	1
	SEPARADO → INT	2
	UNION LIBRE → INT	3
	VIUDO → INT	4
ID_TIPO_DISCAPACIDAD	VISUAL → INT	0
	AUDITIVA → INT	1
	MOVILIDAD REDUCIDA → INT	2
	OTRA → INT	3
ID_ETNIA	SIN PERTENECER GRUPO ETNICO → INT	0
	NEGRO → INT	1
	MULATO → INT	2
	AFRODESCENDIENTE → INT	3
	AFROCOLOMBIANO → INT	4
	INDIGENA → INT	5
	RAIZAL → INT	6
	PALENQUERO → INT	7
ID_CABEZA_HOGAR	SI_MADRE_CAB_HOGAR → INT	0
	SI_PADRE_CAB_HOGAR → INT	1
	NO → INT	2
	NO REGISTRA → INT	3
ID_OCUPACION_MAMA	LABORA → INT	0
	CASAS DE FAMILIA → INT	1
	PENSIONADA → INT	2
	INDEPENDIENTE → INT	3
	FALLECIDA → INT	4
	ESTUDIANTE → INT	5
	CAB HOGAR → INT	6
	NO REGISTRA → INT	7
ID_OCUPACION_PAPÁ	LABORA → INT	0
	CASAS DE FAMILIA → INT	1
	PENSIONADA → INT	2

	INDEPENDIENTE → INT	3
	FALLECIDA → INT	4
	ESTUDIANTE → INT	5
	CAB HOGAR → INT	6
	NO REGISTRA → INT	7
ID_TALENTO	SI → INT	0
	NO → INT	1
	NO REGISTRA → INT	2
ID_EPS	SI → INT	0
	NO → INT	1
	NO REGISTRA → INT	2
ID_LENTES	SI	0
	NO	1
	NO REGISTRA	2

*Nota. Variables categóricas con base a la información personal y sus valores.
 (Fuente: Propia, derivada de la investigación).*

Tabla 4.

Datos socioeconómicos.

VARIABLE	TIPO
ID_OCUPACION	CATEGÓRICA
ID_ESTRATO	CATEGÓRICA
ID_TIPO_VIVIENDA	CATEGÓRICA
ID_INGRESOS	CATEGÓRICA
ID_ACUDIENTE	CATEGÓRICA
ID_NIVEL_ACADEMICO	CATEGÓRICA
ID_INGRESO FAMILIAR	INT
ID_TIENE PC	CATEGÓRICA
ID_TIENE INTERNET	CATEGÓRICA
ID_TIENE SMARTPHONE	CATEGÓRICA
ID_PLAN DATOS	CATEGÓRICA
ID_INGRESOS TRABAJO	CATEGÓRICA

*Tabla 4. Variables con base a la información socioeconómica.
 (Fuente: Propia, derivada de la investigación).*

Tabla 5.

Categorías socioeconómicas

VARIABLE	CATEGORÍAS	VALOR
ID_OCUPACION	SOLO ESTUDIANTE → INT	0
	EMPLEADO → INT	1
	INDEPENDIENTE → INT	2
	PENSIONADO → INT	3
ID_ESTRATO	1 → INT	0
	2 → INT	1
	3 → INT	2
	4 → INT	3
	5 → INT	4
ID_TIPO_VIVIENDA	PROPIA → INT	0
	ARRENDADA → INT	1
	FAMILIAR → INT	2
	OTRA → INT	3
ID_INGRESOS	SI → INT	0
	NO → INT	1
ID_ACUDIENTE	SI → INT	0
	NO → INT	1
ID_NIVEL_ACADEMICO	BACHILLER → INT	0
	BACHILLER TECNICO → INT	1
	UNIVERSITARIO → INT	2
	TECNICO → INT	3
	TECNOLOGO → INT	4
ID_TIENE PC	SI → INT	0
	NO → INT	1
ID_TIENE INTERNET	SI → INT	0
	NO → INT	1
ID_TIENE SMARTPHONE	SI → INT	0
	NO → INT	1
ID_PLAN DATOS	SI → INT	0
	NO → INT	1
ID_INGRESOS TRABAJO	SI → INT	0
	NO → INT	1

*Nota. Variables categóricas con base a la información socioeconómica y sus valores.
 (Fuente: Propia, derivada de la investigación).*

Tabla 6.

Datos academicos

VARIABLE	TIPO
ID_CARRERA ACTUAL	→ INT
ID_HERMANOS UPC	→ INT
ID_INGRESO UPC	CATEGÓRICA
ID_VALIDACION BACHILLER	CATEGÓRICA
ID_TIPO_COLEGIO	CATEGÓRICA
ID_INGRESO(ADMITIDO)	CATEGÓRICA
ID_MOTIVACION (CARRERA)	CATEGÓRICA
ID_INICIO CARRERA (FECHA)	CHAR (6)
ID_REPITIO ICFES	CATEGÓRICA
ID_PUNTAJE ICFES	→ INT
ID_PUESTO ICFES	→ INT
ID_LECTURA CRITICA	→ INT
ID_MATEMATICAS	→ INT
ID_COMP CIUDADANAS	→ INT
ID_CIENCIAS	→ INT
ID_INGLES	→ INT
ID_SEM CURSADOS	→ INT
ID_UBICACION SEM	→ INT
ID_CREDITOS APROBADOS	→ INT
ID_PROM PONDERADO	FLOAT
ID_PROM SEM ANTERIOR	FLOAT
ID_CREDITOS_NO_APROBADOS	→INT
ID_CREDITOS_MATRIC	→ INT
ID_MATERIAS_CANCELADAS	→ INT
ID_ASIGNATURA_ENCUESTA	CHAR (6)
ID_JORNADA ACADEMICA	CATEGÓRICA
ID_NOTA_ACTUAL	FLOAT
ID_RENDIMIENTO_ACTUAL	CATEGÓRICA
ID_NIVEL SATISFACION_ASIG	CATEGÓRICA

*Nota. Variables con base a la información académica.
 (Fuente: Propia, derivada de la investigación).*

Tabla 7.

Categorías académicas

VARIABLE	CATEGORÍA	VALOR
ID_INGRESO UPC	ASPIRANTE REGULAR → INT	0
	PREUNIVERSITARIO → INT	1
	DEPORTISTA DESTACADO → INT	2
	CABILDO INDIGENA → INT	3
	TRANSF INTERNA → INT	4
	TRANSF EXTERNA → INT	5
	REINTEGRO → INT	6
	RESERVA CUPO → INT	7
	OTRO → INT	8
	MEJOR BACHILLER → INT	9
ID_VALIDACION BACHILLER	SI → INT	0
	NO → INT	1
ID_TIPO_COLEGIO	PUBLICO → INT	0
	PRIVADO → INT	1
	NO REGISTRA → INT	2
ID_INGRESO(ADMITIDO)	SI → INT	0
	NO → INT	1
ID_MOTIVACION (CARRERA)	BAJO → INT	0
	MEDIO → INT	1
	ALTO → INT	2
	EXCELENTE → INT	3
ID_REPITIO ICFES	SI → INT	0
	NO → INT	1
ID_JORNADA ACADEMICA	DIURNA → INT	0
	NOCTURNA → INT	1
	ESPECIAL → INT	2
	SABATINA → INT	3
ID_RENDIMIENTO_ACTUAL	BAJO → INT	0
	MEDIO → INT	1
	ALTO → INT	2
	EXCELENTE → INT	3
ID_NIVEL_SATISFACION_ASIG	BAJO → INT	0
	MEDIO → INT	1
	ALTO → INT	2

	EXCELENTE →INT	3
--	----------------	---

*Nota. Variables categóricas con base a la información académica y sus valores.
(Fuente: Propia, derivada de la investigación).*

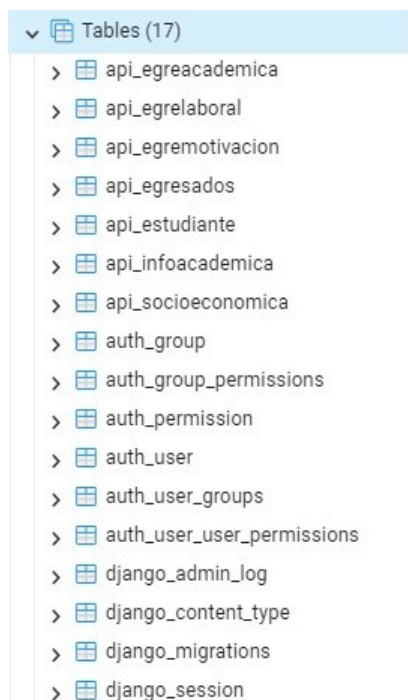
- ✓ Definir herramientas de minería de datos, descubrir patrones y relaciones validas en el conjunto de datos.

Después de haber completado la organización y traducción de todos los datos en el documento de Excel, se procedió a la creación de tablas de datos en la herramienta de software PostgreSQL. En esta etapa, se construyeron todas las tablas requeridas junto con sus datos correspondientes para continuar con el proceso de análisis del proyecto.

La razón detrás de este paso es esencial. Al organizar y traducir los datos en un formato tabular en PostgreSQL, se estableció una base sólida para llevar a cabo análisis más avanzados y efectivos. Las bases de datos relacionales como PostgreSQL ofrecen una estructura de almacenamiento eficiente y permiten realizar consultas complejas, lo que facilita la identificación de patrones, tendencias y relaciones en los datos. Además, proporcionan una plataforma segura y escalable para gestionar grandes conjuntos de información. Cada tabla que se creó en PostgreSQL. Figura 5, representa una entidad específica o una relación entre entidades en nuestros datos.

Figura 5.

Tablas de datos en PostgreSQL



*Nota. Clasificación de tablas en la base de datos creada en PostgreSQL.
 (Fuente: Propia, derivada de la investigación).*

Al construir la base de datos de esta manera, se están sentando las bases para el análisis de datos que ayudará a cumplir los objetivos del proyecto, que es entender y predecir los factores que influyen en la deserción universitaria. Esta base de datos. Figura 6, será la columna vertebral del análisis, permitiendo explorar relaciones entre variables, aplicar algoritmos de predicción y, en última instancia, extraer conocimientos valiosos de los datos encuestados.

Figura 6.

Base de datos para el análisis de deserción

	ID_DOCUMENTO [PK] integer	ID_TIPO_DOCUMENTO character varying (35)	ID_NOMBRE character varying (50)	ID_EDAD integer	ID_NUM_CONTACTO character varying (20)	ID_GENERO integer	ID_VIC_CONFLICTO integer
1	1000241868	cedulaCiudadania	luis daniel medina campo	19	3188393811	1	1
2	1001852955	cedulaCiudadania	Edgar Daniel Torres Padilla	22	3233291955	1	1
3	1001919993	cedulaCiudadania	saith galindo	19	3046231989	1	2
4	1002132476	cedulaCiudadania	breider farid bovea lara	22	0	1	2
5	1002187929	cedulaCiudadania	carlos miguel martes martes	22	2269984396	1	2
6	1002188303	cedulaCiudadania	alvaro javier julio aragon	24	0	1	1
7	1003038093	cedulaCiudadania	jesus andres amaya rangel	21	3136598426	1	1
8	1003121874	cedulaCiudadania	adelson javier ochoa carpio	19	3183690393	1	2
9	1003123708	cedulaCiudadania	diego andres cassiani melendres	20	3126896762	1	1
10	1003173210	cedulaCiudadania	rafael ramiro ramirez matoma	19	0	1	1

*Nota. Base de datos creada en PostgreSQL.
 (Fuente: Propia, derivada de la investigación).*

Es fundamental resaltar que se ha garantizado la precisión y la calidad en la traducción de datos, ya que se llevó a cabo una revisión detallada de cada dato, y cualquier corrección necesaria se realizó de forma manual.

Este enfoque meticuloso se implementó con el propósito de asegurar la confiabilidad de la conversión de datos y garantizar que los resultados del análisis posterior sean precisos y significativos.

```
/*CREACIÓN DE LA BASE DE DATOS 'DATOS_ENCUESTA' */
```

A continuación, se presenta el script para la creación de la base de datos, que se incluye en el **Anexo A**

- ✓ Diseñar el prototipo del software con base al modelo de datos.

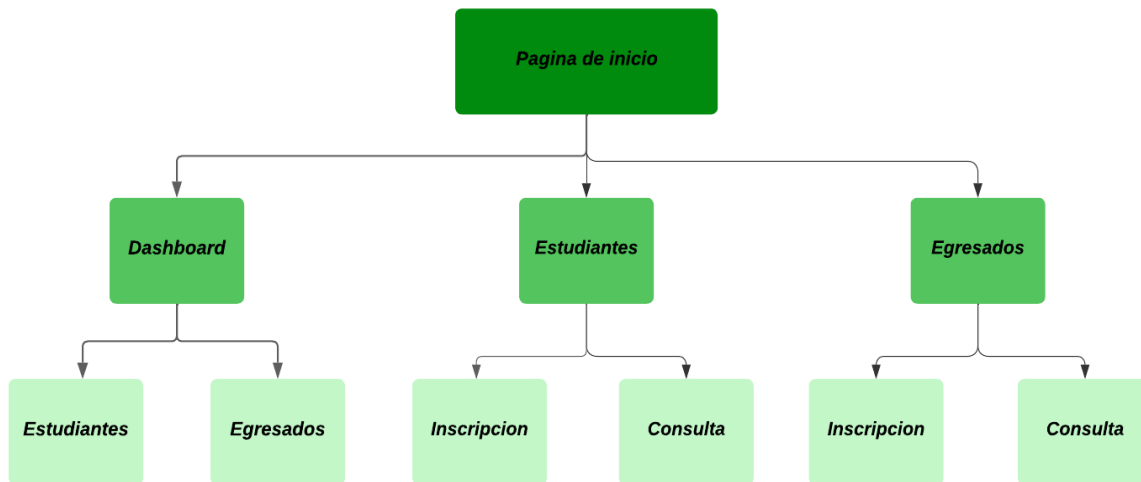
Previo a la implementación y realización del aplicativo, es necesario tener claro el comportamiento y funcionamiento del mismo, razón por la cual se crea un modelo en el que se definen y crean las distintas metodologías a seguir; en este caso siguiendo los parámetros y variables que se tienen se establecen la forma en que se realizarán los gráficos, las predicciones a través de la red neuronal realizando múltiples pruebas para lograr un entrenamiento de la misma y poder ajustarla al objetivo mismo. A continuación, el código implementado en el lenguaje Python del modelo:

A continuación, se presenta el código implementado en el lenguaje Python del modelo, que se incluye en el **Anexo B**

Es importante crear una estructura sólida para guiar el desarrollo del aplicativo. Esto garantizará que la aplicación sea intuitiva y fácil de usar, manteniendo al mismo tiempo secciones exclusivas para estudiantes y exalumnos que contendrán información. Imaginemos esta estructura como un diagrama de jerarquía observado en la figura 7, con un nivel principal llamado página de inicio y tres subniveles básicos: *dashboards*, estudiantes y egresados.

Figura 7.

Diagrama de Jerarquía del aplicativo



*Nota. Estructura jerárquica del aplicativo.
(Fuente: Propia, derivada de la investigación).*

Nivel principal: página de inicio

En la parte superior de la estructura se encuentra la página de inicio, que es la puerta de entrada de la aplicación. La página debe ser atractiva y debe proporcionar una descripción general de la funcionalidad de la aplicación. Su propósito es ayudar a los usuarios a obtener los recursos y funciones que necesitan de manera eficiente.

Subnivel 1: Dashboards

Debajo de la página de inicio, el primer subnivel es el *dashboard*. Es un panel de control que recopila información clave y proporciona una descripción general de los datos relevantes a todos los usuarios. Aquí, los usuarios pueden acceder a estadísticas generales y cualquier información general que deseen observar. Es importante destacar que estos paneles son personalizables para que los usuarios puedan adaptar la información que ven a sus necesidades específicas.

Subnivel 2: Estudiantes

El segundo subnivel es para estudiantes. Esta sección es fundamental para mantener la integridad de los datos de sus estudiantes, ya que contendrá información personal de cada estudiante registrado en las encuestas. Aquí, se puede consultar, editar y eliminar los datos de cada alumno por separado todo esto se hace a través de una consulta mediante el documento de identidad.

Subnivel 3: Egresados

Finalmente, el tercer subnivel está reservado a los egresados. Al igual que con los estudiantes, esta sección contendrá información sobre los exalumnos. Aquí, usuarios podrán consultar, editar y eliminar información mediante el uso del documento de identidad del egresado.

Para garantizar la usabilidad de la aplicación, las pruebas de usuario son esenciales durante todo el proceso de diseño. Esto le permite identificar posibles puntos de fricción y realizar los ajustes necesarios en la interfaz. Además, se debe prestar especial atención al diseño visual, que garantiza que los colores, las fuentes y los elementos visuales sean coherentes con la identidad del proyecto.

Construcción de las páginas

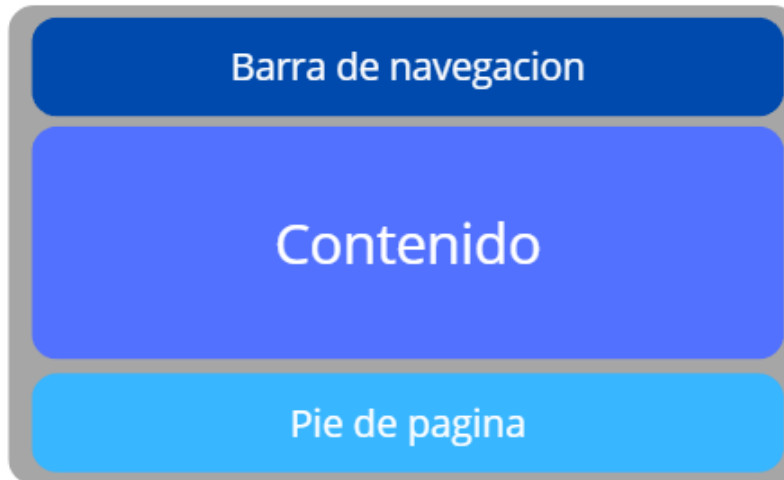
La construcción de las páginas como se ha mencionado anteriormente que es modularizado, quiere decir, que no es un solo archivo HTML que contiene todas las partes de la página, sino que son varios archivos, por separado lo cuales podemos modificar e importar hacia otros con el fin de simplificar el tamaño del código y la implementación de nuevos apartados.

Todo esto sucede dentro un el archivo que se llama “**Main layout**” el cual es la configuración principal de cada página.

En este caso primero se trae la barra de navegación, luego el **main** donde va el contenido de la página, este también es importado a través de un archivo de rutas invocado por “**router-view**” y por último el pie de página. Figura 8

Figura 8.

Estructura modular de la página web



*Nota. Construcción páginas del aplicativo.
(Fuente: Propia, derivada de la investigación).*

De manera que tiene mucha similitud a un rompecabezas donde se unen las partes para ir formando la figura que se desea, esta misma analogía es la que se implementa para cambiar el contenido de la página dependiendo del apartado en el que se encuentre el usuario.

Principalmente la barra de navegación y el pie de página no van a sufrir modificaciones, pero si el contenido y esto se hace mediante un archivo de rutas llamado **routes.js**, usando JavaScript, se crean las rutas necesarias para navegar por cada página de la plataforma.

Lo primero que se debe hacer es cargar la ruta principal ("/"), esta ruta está configurada para la página de inicio de la aplicación, importando el componente que se debe inicializar cuando se accede a esta ruta. En este caso, se carga el componente "**Main.layout**" ubicado en la carpeta (**layouts**).

A continuación, dentro de la ruta principal, se definen las rutas hijas (**children**) que se utilizarán para mostrar contenido específico en áreas de la página principal:

Ruta " / ": En el contexto de un sitio web, una ruta es un camino o dirección que se utiliza para acceder a una página específica. Por ende, es la ruta por defecto de la página de inicio, es decir, la página principal cuando se accede a la raíz del sitio.

Ruta consulta: Esta ruta está configurada para mostrar la página de "Consulta del estudiante".

Ruta edition: Esta ruta está configurada para mostrar la página de "Edición del estudiante".

Ruta consulta_egresados: Esta ruta está configurada para mostrar la página de "Consulta de Egresados".

Ruta inscripción: Esta ruta está configurada para mostrar la página de "Inscripción".

Ruta inscripcion_egresados: Esta ruta está configurada para mostrar la página de "Inscripción de Egresados".

Ruta dashStu: Esta ruta está configurada para mostrar un *dashboard* para estudiantes.

Para cada una de estas rutas hijas, se especifica el nombre de la ruta y el componente asociado este componente es un archivo (**.vue**) el cual contiene código HTML, JavaScript y CSS necesario para el contenido que se desea mostrar, que se cargará cuando el usuario acceda a esa ruta en particular.

Creación de los componentes

Los componentes son el contenido de cada apartado siendo este el que cambia según la página donde se encuentre el usuario, estas son inscripción, consulta, edición y el panel de visualizaciones (*dashboard*), tanto para estudiantes y egresados.

- **INSCRIPCIÓN:** Este componente es donde se registra a los estudiantes y egresados que no estén en las encuestas, teniendo en un formato de registro mucho más limpio y por secciones que el de la encuesta, el código de este apartado es:

Dentro de este componente se pueden agregar más componentes; estos son *PersonalForm*, *EconomicForm*, *AcademicForm*. Éstos tienen las preguntas de las encuestas distribuidas para cada apartado.

Desde esta función los datos son enviados como Json al back-end donde son desmostado e ingresados a la base de datos.

- **CONSULTA:** Este componente es el que contiene más utilidades, ya que permite consultar, entrar al componente de edición y eliminar; esto se hace mediante un cuadro con una entrada numérica que permite digitar el documento de identidad y 3 botones para consultar, editar y eliminar cada uno de estos, programados con una función en el apartado de *JavaScript* del componente que ejecutan dichas acciones, además cuenta con una tabla para mostrar los datos obtenidos. Figura 9 - Figura 10. separadas por un guía de navegación que

independiza las tablas una de la otra como se puede apreciar en las siguientes figuras.

Figura 9.

Componente de consulta de estudiantes

Consulta de estudiantes

Documento
1000241868

Buscar Editar Eliminar

*Nota. Cuadro de operaciones.
 (Fuente: Propia, derivada de la investigación).*

Figura 10.

Tabla de datos personales

Datos personales		Datos socioeconómicos		Datos académicos	
Programa actual	Ing. Electronica	Ingreso a la upc	Aspirante regular	Hermanos IES	1
Motivación	Excelente	Validó bachiderato	No	Tipo de colegio	Público
Repitió icfes	No	Puntaje icfes	266	Puesto icfes	65
Matemática	64	Ciudadanos	44	Ciencias	50
Ubicación semestre	4	Créditos aprobados	78	Promedio acumulado	4
Créditos no aprobados	Valor no definido	Créditos matriculados	20	Materias canceladas	0
Jornada académica	Diurno	Nota actual	0	Rendimiento actual	Alto
				Ingreso oferta	No
				Inicio semestre	2021-1
				Lectura crítica	55
				Inglés	53
				Promedio semestre anterior	3.6
				Asignatura Encuesta	Circuitos III, código EL411
				Nivel satisfacción asignatura	Excelente

*Nota. Tabla de consulta de datos del aplicativo.
 (Fuente: Propia, derivada de la investigación).*

- **EDICIÓN:** En el componente de edición tenemos prácticamente el mismo código que en la inscripción puesto que maneja el mismo principio, pero con dos diferencias muy relevantes; como es un apartado de edición de datos, se despliega desde el apartado de consultas, ya que ahí es donde se solicita la información de algún estudiante o egresado, y al ser lanzado redirige hacia otra página donde estará la información actual de dicha persona. Figura 11.

Figura 11.

Componente de edición de datos

Tipo de documento Cédula	Documento 1234567	Nombre Andres Reyes	Edad 23
Telefono 310999999	Email acamiloreyes@unicesar.edu.co	Genero Masculino	Estado Civil Soltero
Departamento cesar	Municipio valledupar	¿Tiene alguna discapacidad? Ninguna	Etnia Sin distincion racial
¿Es victima del conflicto? No	¿Es Madre o Padre Cabeza de hogar? No	Edad hijo Mayor 0	Edad hijo Menor 0
Cantidad de hijos 0	Cantidad de hermanos 1	Posición hermanos 1	Integrantes 6
Ocupacion de la madre Labora	Ocupacion del padre Independiente	Sisben A12	Talento Fútbol
EPS Sí	¿Usa lentes? No		

*Nota. Componente de edición del aplicativo.
(Fuente: Propia, derivada de la investigación).*

Se cambian los campos que se desean modificar y continuar con el flujo de la encuesta hasta finalizar el componente de edición, esto creará el nuevo objeto que será enviado al back-end ingresado a la base de datos para hacer dicha actualización.

- **DASHBOARD:** Este componente muestra la parte de visualizaciones después del análisis de los datos, esto se consigue gracias a la librería “**Apache-Echarts**”, que facilita el uso de herramientas para graficar en el entorno HTML con gráficos dinámicos e interactivos, lo primero que se debe realizar para el correcto funcionamiento de esto es la importación de los datos que vienen de back-end, para ello se crea un archivo de configuración para cada gráfico, esto es necesario para mantener la integridad del apartado principal; estos archivos

van dentro de una carpeta llamada “**config**” en el lenguaje JavaScript, cada archivo contiene una petición **GET** a la API.

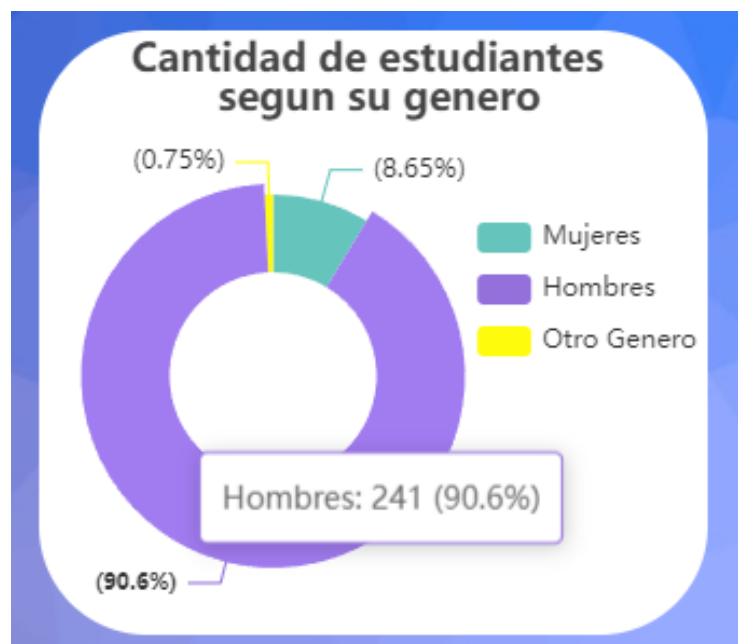
Al obtener los datos, es necesario la construcción del gráfico, esto usando las herramientas suministradas por “*Apache-Echarts*”, donde establece parámetros tanto de los axis, *legend*, series, títulos, *tooltip* dentro de una variable exigida por la librería para su funcionamiento, cada grafico es diferente, pero para mostrar el desarrollo de la construcción del grafico usaremos el primer gráfico. Este grafico tiene como objetivo conocer la cantidad de estudiantes según su género y para ello se usó un **gráfico tipo pie**, la configuración es la siguiente:

“*Apache-Echarts*”, contiene la propiedad (**V-chart**), esta propiedad exige una clase “*Chart*” que inicializa la configuración de los gráficos mostrados anteriormente, mediante la variable opción 1,2,3,4,5 respectivamente para cada gráfico.

Con esto se despliega el gráfico mostrado en la siguiente figura, la cual muestra un gráfico de pie; este gráfico contiene animación de inicio, filtro de información a través de los títulos de las variables mostradas y *tooltips* con la información cuando el mouse se posiciona encima de cada variable.

Figura 12.

Gráfico de estudiantes por genero



Nota. 1er gráfico informativo del aplicativo (Cantidad de estudiantes según su género). (Fuente: Propia, derivada de la investigación).

Conexión entre el *front-end* y el *back-end*

El intercambio de información entre estas dos capas se ha mencionado mucho durante el transcurso de este proyecto, es esencial una buena conexión entre el *Front-end* y al *API (back-end)* y para eso se usa la librería conocida como “**AXIOS**”, esta librería permite el uso de peticiones *HTTP* al *API* a través de sus herramientas.

Estas peticiones se realizan hacia las rutas del *back-end*, establecidas en el apartado de “URL” del *back-end*, estas extraen la información obtenida mediante un formato *Json* el cual es usado en el *front-end*, para su visualización en los diferentes apartados.

Dependiendo del tipo de función el método *HTTP* cambia, si es una consulta se usa *GET* para pedir datos, si es una eliminación se usa *DELETE* para borrar el dato en la base de datos, si es una petición de agregar o editar información se usa una petición *POST* que envía datos hacia el servidor.

Estas almacenan la información generalmente en la variable “**data**” que es enviada al *front-end* mediante el llamado de cada función **async**.

METODOLOGÍA

Teniendo en cuenta todos los factores socioculturales, estructurales, políticos, institucionales, personales, de aprendizaje, entre otros; los cuales influyen de manera directa e indirecta en la deserción universitaria; podemos establecer qué variables y patrones inciden, así como la relación que tienen por medio del trabajo investigativo, a través del análisis de los mismos. Y siguiendo el enfoque en nuestro claustro universitario; usando como muestra y objeto de estudio el programa de Ingeniería Electrónica poder tener una mayor viabilidad de ejecución, esto con el fin de ampliar todas las posibles soluciones a la problemática presentada.

A través de encuestas realizadas a los estudiantes usando diferentes medios tecnológicos entre los que destacan los Formularios de Google que ayudan a la recopilación de los datos, y con los que se busca poder categorizar todas las variables que abordan la problemática, y de ahí empezar a tener datos reales que por medio de un análisis cuantitativo y documental describan y permitan tener una visualización más específica de todas estas variables, ayudados nuevamente de las herramientas tecnológicas a disposición, mencionando algunas como *Python* y varios de sus módulos especializados en bases de datos (*Peewee*, *MySQLdb*, *Gadfly*), ciencia de datos (*Pandas*, *Numpy*, *Matplotlib*), desarrollo web (*Django Rest Framework*, *Beautiful Soup*), entre otros. Los cuales, al mismo tiempo, ayudan al diseño y estructuración del prototipo de software que se desea realizar e implementar.

Considerando los factores socioculturales, estructurales, políticos, institucionales, personales y de aprendizaje que afectan directamente e indirectamente la deserción universitaria, la aplicación de métodos de predicción adecuados y eficientes en los modelos de encuestas puede contribuir significativamente a la intervención temprana y prevención de este fenómeno. Además, estas herramientas proporcionan información valiosa para ajustar el programa académico, específicamente en el contexto del programa de Ingeniería Electrónica, al comprender mejor las variables y patrones que influyen en la deserción estudiantil. A través de encuestas tecnológicas, como Formularios de Google, y el uso de herramientas de análisis de datos como *Python* y sus módulos especializados, se puede categorizar y analizar estas variables para desarrollar soluciones más efectivas.

RESULTADOS Y ANÁLISIS

Figura 13.

Página principal del aplicativo web



*Nota. Visualización página principal (home) aplicativo web.
(Fuente: Propia, derivada de la investigación).*

A continuación, se muestra la visualización de los distintos formularios para recopilar la información necesaria que posteriormente será empleada por los modelos de inteligencia artificial para analizar y prever los factores que influyen en la deserción estudiantil. Figura 14 – Figura 15 – Figura 16, estos forman parte de la interfaz de usuario de la aplicación web en proceso de desarrollo.

Figura 14.

Formulario de recopilación de información personal

Información personal

Complete todos los datos.

Tipo de documento	Documento	Nombre	Edad
Seleccionar ▼	<input type="text"/>	<input type="text"/>	<input type="text"/>
Telefono	Email	Genero	Estado Civil
<input type="text"/>	<input type="text"/>	Seleccionar ▼	Seleccionar ▼
Departamento	Municipio	¿Tiene alguna discapacidad?	Etnia
<input type="text"/>	<input type="text"/>	Seleccionar ▼	Seleccionar ▼
¿Es victima del conflicto?	¿Es Madre o Padre Cabeza de hogar?	Edad hijo Mayor	Edad hijo Menor
Seleccionar ▼	Seleccionar ▼	<input type="text"/>	<input type="text"/>
Cantidad de hijos	Cantidad de hermanos	Posición hermanos	Integrantes
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Ocupacion de la madre	Ocupacion del padre	Sisben	Talento
Seleccionar ▼	Seleccionar ▼	<input type="text"/>	Seleccionar ▼
EPS	¿Usa lentes?		
Seleccionar ▼	Seleccionar ▼		

Siguiente

Nota. (Fuente: Propia, derivada de la investigación).

Figura 15.

Formulario de recopilación de información académica

Información Académica

Complete todos los datos.

Seleccione la carrera universitaria que está estudiando actualmente <input type="text" value="Seleccionar"/>	Su nota obtenida actualmente en la asignatura donde está siendo encuestado <input type="text"/>	Usted ingresó a la Universidad Popular del Cesar por medio de según su oferta académica <input type="text" value="Seleccionar"/>
Ingresó a la universidad por (Mejor bachiller, preuniversitario o deportista) <input type="text" value="Seleccionar"/>	Qué tan motivado te sentías al momento de ingresar a estudiar su carrera profesional <input type="text" value="Seleccionar"/>	En qué semestre iniciaste tu carrera profesional en la Universidad Popular del Cesar <input type="text" value="Seleccionar"/>
Tipo de colegio en el que estudio <input type="text" value="Seleccionar"/>	Validó el bachillerato <input type="text" value="Seleccionar"/>	Repitió usted las pruebas saber 11° ICSES <input type="text" value="Seleccionar"/>
PUNTAJE de las pruebas saber 11° ICSES <input type="text"/>	PERCENTIL de las pruebas saber 11° ICSES <input type="text"/>	Puntaje de lectura crítica <input type="text"/>
Puntaje de matemáticas <input type="text"/>	Puntaje de sociales y ciudadanas <input type="text"/>	Puntaje de ciencias naturales <input type="text"/>
Puntaje de inglés <input type="text"/>	¿Cuántos semestres ha cursado usted actualmente? <input type="text"/>	Su ubicación semestral según el vortal <input type="text"/>

Créditos aprobados hasta el 2023-1 <input type="text"/>	¿Cual es su promedio acumulado actual? <input type="text"/>	¿Cual es su promedio del semestre anterior? <input type="text"/>
Cuantos créditos no aprobados? <input type="text"/>	Créditos matriculados este semestre 2023-1 <input type="text"/>	Desde que asignatura esta haciendo la encuesta <input type="text" value="Seleccionar"/>
Jornada académica <input type="text" value="Seleccionar"/>	Actual rendimiento académico en la asignatura <input type="text" value="Seleccionar"/>	Nivel de satisfacción con la asignatura <input type="text" value="Seleccionar"/>
¿Tiene hermanos estudiando actualmente en alguna IES? <input type="text"/>	¿Ha cancelado alguna asignatura este semestre 2023-1? <input type="text"/>	

Nota. (Fuente: Propia, derivada de la investigación).

Este formulario está diseñado para recoger datos sobre el progreso académico de los alumnos. Se abordan diversos temas, como los créditos obtenidos hasta el tercer semestre de 2021, el promedio acumulativo actual, el promedio del semestre anterior, los créditos inscritos este semestre, el horario académico, el rendimiento actual en la materia, el nivel de satisfacción con la materia, si el alumno está teniendo dificultades en alguna Institución de Educación Superior (IES) en este momento, y si ha retirado alguna materia este semestre.

Este formulario es una herramienta esencial para recoger información sobre los alumnos. Los datos recogidos se utilizan para alimentar los modelos de inteligencia

artificial y análisis de aprendizaje que están creados para poder identificar a los alumnos que corren el riesgo de abandonar la universidad.

Figura 16.

Formulario de recopilación de información socioeconómica

Información Socioeconómica
Complete todos los datos.

¿Tiene usted alguna ocupación? Seleccionar	¿Estrato socioeconómico? Seleccionar	¿Tipo de vivienda? Seleccionar
¿Recibe usted ingresos mensualmente? Seleccionar	¿Alguien lo representa legalmente? Seleccionar	Nivel académico Seleccionar
¿Cuánto es el ingreso familiar? <input type="text"/>	¿Posee usted computador? Seleccionar	¿Tiene acceso a Internet? Seleccionar
¿Posee usted un celular inteligente? Seleccionar	¿Posee usted un plan de datos? Seleccionar	¿Al momento de entrar a la universidad usted trabajaba? Seleccionar

Siguiente

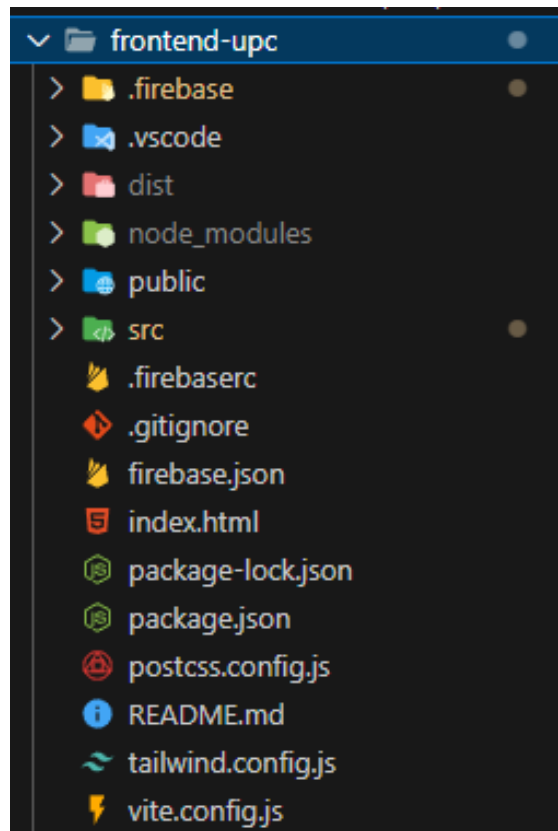
Nota. (Fuente: Propia, derivada de la investigación).

Este formulario recoge datos sobre la condición socioeconómica de los alumnos, dicha información es fundamental para comprender algunas de las razones que llevan a la deserción universitaria. Se abordan diversos temas, como el empleo del alumno, los ingresos mensuales, el nivel de estudios, el acceso a Internet y el uso de un *Smartphone*. Las respuestas a estas preguntas ofrecen una visión valiosa de las razones que llevan a la deserción universitaria, permitiendo así la formulación de políticas y estrategias para enfrentar este problema.

A continuación. Figura 17 – Figura 18, se mostrará la estructura organizacional del proyecto en el entorno de desarrollo. En este caso, se ha utilizado Visual Studio Code para gestionar y desarrollar el proyecto. Se presentan capturas de la estructura del proyecto en Visual Studio Code, mostrando cómo se han organizado los archivos y carpetas para facilitar el desarrollo y mantenimiento del aplicativo.

Figura 17.

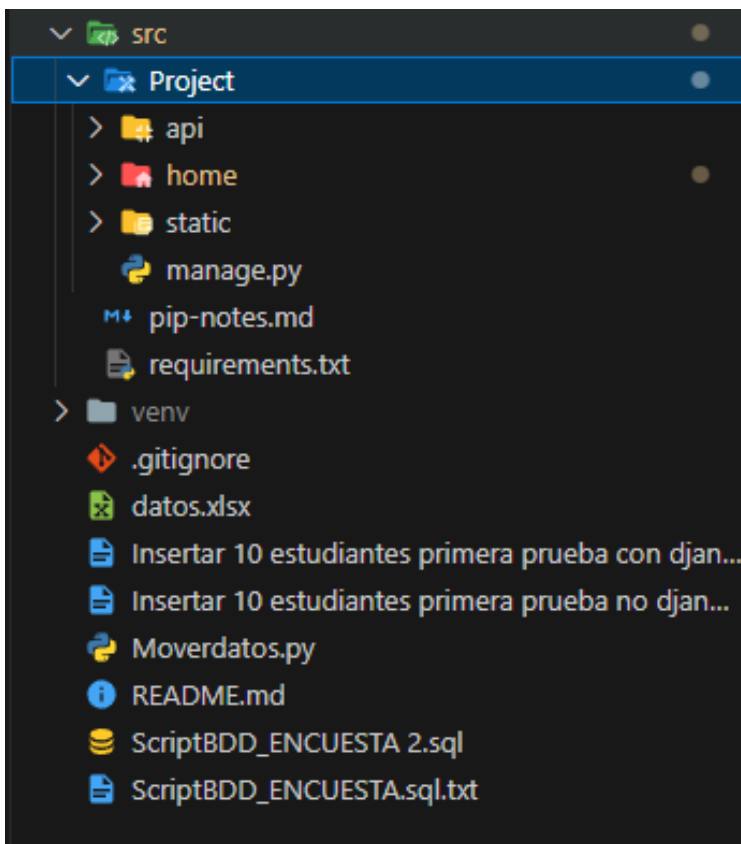
Gestor de archivos parte 1



Nota. (Fuente: Propia, derivada de la investigación).

Figura 18.

Gestor de archivos parte 2



Nota. (Fuente: Propia, derivada de la investigación).

El gestor de archivos muestra la organización del proyecto de la aplicación web en desarrollo. Esta organización es el producto de la aplicación de diversas técnicas informáticas al sistema de información que se está creando para la universidad.

Este gestor de archivos presenta la estructura de los archivos de la sección visual del proyecto, denominada “*frontend-upc*”. Los archivos y carpetas están dispuestos en una estructura de árbol y se distinguen por colores según su tipo: naranja para los archivos relacionados con *git*, verde para los módulos de nodo y azul para los archivos fuente. Los archivos también se diferencian por colores: rojo para los archivos *HTML*, amarillo para los archivos *JSON* y morado para los archivos *JavaScript*.

CONCLUSIONES

Durante la ejecución del proyecto, se investigaron diversas formas de análisis de datos. La que arrojó los mejores resultados fueron las Redes Neuronales Recurrentes, por lo que se eligió esta técnica como base para lograr un mejor desempeño en el objetivo de predecir la deserción universitaria. En este contexto, se desarrolló un modelo predictivo utilizando una base de datos que contenía información académica y sociodemográfica de los estudiantes del programa de Ingeniería Electrónica. Para validar la aplicabilidad del modelo en distintos contextos, se realizaron pruebas adicionales utilizando datos de estudiantes de otras disciplinas (Psicología y Microbiología). Además, se llevó a cabo un trabajo integral en todas las fases del proyecto para diseñar una plataforma de visualización de datos, que cuenta con una interfaz gráfica de usuario (GUI) con opciones visuales adecuadas, modelos definidos, conexión con bases de datos específicas, y capacidades de consulta y filtrado de datos para fines estadísticos y analíticos.

Este prototipo de software ha evidenciado claramente que la minería de datos y el análisis de aprendizaje son herramientas efectivas para prevenir la deserción universitaria en el programa de Ingeniería Electrónica de la universidad popular del cesar. Al examinar datos académicos y sociodemográficos, el sistema es capaz de identificar señales y proyecciones que facilitan la predicción y comprensión más profunda de este fenómeno tan preocupante como lo es la deserción estudiantil universitaria nuestra alma mater.

Aunque este prototipo ha mostrado un gran potencial, es esencial superar algunas limitaciones para mejorar su validez. Ampliar el tamaño de la muestra, incorporando estudiantes de diversos programas académicos y cohortes, contribuiría a obtener resultados más sólidos y representativos. Asimismo, la introducción de mecanismos de control de calidad durante la recolección de datos, como la validación de la información y el fomento de una mayor participación, podría elevar la precisión de los análisis. Implementar estas medidas ayudaría a garantizar que el prototipo no solo sea más confiable, sino también más eficaz en la identificación de factores críticos relacionados con la deserción universitaria y el rendimiento académico.

RECOMENDACIONES

Una estrategia clave para enriquecer el análisis y mejorar la precisión del modelo predictivo es ampliar las fuentes de datos. Esto se puede lograr integrando nuevos conjuntos de datos tanto cuantitativos como cualitativos. Por ejemplo, se podrían llevar a cabo encuestas y entrevistas exhaustivas con estudiantes actuales y desertores. Estas encuestas y entrevistas permitirían recopilar información detallada sobre sus experiencias educativas, motivaciones personales, desafíos y dificultades que enfrentan en su trayectoria académica.

Al obtener datos cualitativos a través de entrevistas, se puede capturar un espectro más amplio de factores que afectan la deserción, proporcionando una comprensión más profunda y matizada de los problemas subyacentes. Las entrevistas pueden revelar aspectos personales y contextuales que no son fácilmente capturados por los datos cuantitativos, como las dinámicas familiares, la salud mental, las expectativas personales y las influencias sociales.

Para maximizar el impacto y la utilidad de este prototipo de software, es crucial extender su implementación a toda la universidad. Al aplicar la solución de analítica predictiva a una gama más amplia de programas académicos, se pueden construir modelos más robustos y generalizables a nivel institucional. Esta expansión permitirá identificar factores transversales que afectan la deserción y el rendimiento académico en diferentes disciplinas y contextos socioeconómicos. En consecuencia, se obtendrá una comprensión más completa y holística de los factores que influyen en la deserción, permitiendo desarrollar estrategias de intervención más efectivas y personalizadas para cada programa de estudios.

Es fundamental proporcionar una formación continua y especializada al personal académico y administrativo en el uso de herramientas de minería de datos y análisis avanzado. Esta capacitación debe incluir no solo el manejo técnico del software, sino también una comprensión profunda de las metodologías y técnicas analíticas aplicadas. Al fortalecer las habilidades del personal, se asegura que puedan interpretar los datos correctamente, identificar patrones y tendencias significativas, y utilizar esta información para tomar decisiones estratégicas informadas.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Fernández, E. (2016). Deserción universitaria: causas y consecuencias. *Revista Científica de Administración, Finanzas e Investigación*, 2(2), 73-85.
- [2] Indicadores de deserción universitaria y factores asociados; Escuela de Ciencias de la Educación (ECEDU), Universidad Nacional Abierta y a Distancia (UNAD), Bogotá, Colombia; Deyser Gutiérrez A. John Fredy Vélez Díaz, Juan López M., *Revista EducaT: Educación virtual, Innovación y Tecnologías*. Vol. 2 No. 1, 2021.
- [3] K. B. Eckert and R. Suénaga, “Análisis de deserción-permanencia de estudiantes universitarios utilizando técnica de clasificación en minería de datos,” *Form. Univ.*, vol. 8, no. 5, pp. 3–12, 2015.
- [4] García-Sandoval, J. P., Muñoz-Organero, M., & Sandoval, R. H. (2021). Deserción universitaria y su impacto en el desarrollo económico de México. *Revista Científica de Administración, Finanzas e Informática*
- [5] Castellanos, J. (2015). La deserción universitaria en América Latina y el Caribe: situación, causas y estrategias de intervención. CEPAL.
- [6] Muñoz, R. (2019). Aplicación de modelos de regresión logística para la detección de estudiantes en riesgo de deserción en la educación.
- [7] Efectos de la deserción, Universidad del Valle, Carlos Angulo Galvis, 2017; tomado de: <http://uvsalud.univalle.edu.co/comunicandosalud/wp-content/uploads/2017/10/06.10.17-Efectos-de-la-deserci%C3%B3n1.pdf>
- [8] Elder, Glen, (1985). “Perspectives on the life course”, en: Elder (ed.): *Life course dynamics*. Cornell University Press.
- [9] Tinto, V. (1989): “Definir la deserción una cuestión de perspectiva”, *Revista Educación Superior*, 71, México.
- [10] Castaño et al. (2004). “Deserción estudiantil universitaria: una aplicación de modelos de duración.” *Lecturas de Economía*, No. 60 Enero – Junio.
- [11] ICFES (2002). “Estudio de la deserción estudiantil en la educación superior en Colombia.” Documento convenio UN – ICFES.
- [12] Informe Determinantes de la deserción “Informe mensual sobre el soporte técnico y avance del contrato para garantizar la alimentación, consolidación, validación y uso de la información del SPADIES”. Bogotá 2014. https://www.mineducacion.gov.co/sistemasdeinformacion/1735/articles-254702_Informe_determinantes_desercion.pdf
- [13] V. Tinto, “Completing College: Rethinking Institutional Action.,” *Univ. Chicago*

Press, p. 228, Apr. 2012.

- [14] J. Areth, E. Jaime, C.-M. Henry, and R. Granobles, “La educación virtual en Colombia: exposición de modelos de deserción.,” *Rev. Innovación Educ.*, vol. 7, no. 1, p. 10, 2015.
- [15] C. Guzmán Ruiz *et al.*, “Deserción estudiantil en la educación superior colombiana. Metodología de seguimiento, diagnóstico y elementos de prevención,” Bogotá- Colombia, 2009. Accessed: Aug. 27, 2021. [Online]. Available: www.mineducacion.gov.co.
- [16] I. Velasco Quintero, “ANÁLISIS DE LAS CAUSAS DE DESERCIÓN UNIVERSITARIA,” Universidad Nacional Abierta y a Distancia UNAD, 2016.
- [17] E. E. Rocío Balmori Méndez, M. DE Teresa La Garza Carranza, and E. Reyes Varela, “EL MODELO DE DESERCIÓN DE TINTO COMO BASE PARA LA PLANEACIÓN INSTITUCIONAL: EL CASO DE DOS INSTITUCIONES DE EDUCACIÓN SUPERIOR TECNOLÓGICA,” Accessed: Nov. 23, 2021. [Online]. Available: www.dgest.gob.mx.
- [18] V. D. Vera Gil, “Learning Analytics and Scholar Dropout: A Predictive Model Learning Analytics and Scholar Dropout : A Predictive Model,” vol. 25, no. January, pp. 1414–1419, 2018, doi: 10.5829/idosi.mejsr.2017.1414.1419.
- [19] I. Velasco Quintero, “ANÁLISIS DE LAS CAUSAS DE DESERCIÓN UNIVERSITARIA,” Universidad Nacional Abierta y a Distancia UNAD, 2016.
- [20] Formación basada en las Tecnologías de la Información y comunicación: Análisis didáctico del proceso de enseñanza-aprendizaje available: https://www.tdx.cat/bitstream/handle/10803/8909/Etesis_1.pdf
- [21] Las tecnologías digitales frente a los desafíos de una educación. Available: https://repositorio.cepal.org/bitstream/handle/11362/35386/S2012809_es.pdf
- [22] Importancia del manejo de competencias tecnológicas en las prácticas docentes de la Universidad Nacional Experimental de la Seguridad (UNES). Available: <https://www.redalyc.org/journal/440/44057415013/html/>
- [23] DESERCIÓN ESCOLAR EN COLOMBIA: ANÁLISIS, DETERMINANTES Y POLÍTICA DE ACOGIDA, BIENESTAR Y PERMANENCIA. Available: https://www.mineducacion.gov.co/1780/articles-363488_recurso_34.pdf
- [24] Montaña Omar (2015). Factores que inciden en la deserción estudiantil en el programa académico tecnología en gestión portuaria de la universidad del valle sede pacífico en el periodo 2009-2014. Universidad del Valle sede Pacífico. Buenaventura, Colombia.
- [25] Normas Generales de la Educación Superior, Marco Legal - Sistema de Educación en Colombia. Tomado de:

[https://www.mineducacion.gov.co/1621/w3-printer-184681.html#:~:text=Ley%2030%20de%201992%2D%20Servicio,28%20de%20Diciembre%20de%201992\).](https://www.mineducacion.gov.co/1621/w3-printer-184681.html#:~:text=Ley%2030%20de%201992%2D%20Servicio,28%20de%20Diciembre%20de%201992).)

- [26] Conceptos de minería de datos. Septiembre 2022. Tomado de: <https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>
- [27] Minería de datos, Qué es y por qué es importante?. Tomado de: https://www.sas.com/es_co/insights/analytics/data-mining.html
- [28] ¿Qué es la minería de datos?. Aws.amazon. Tomado de: <https://aws.amazon.com/es/what-is/data-mining/>
- [29] PostgreSQL: qué es, cómo funciona y cuáles son sus ventajas. 2015. Tomado de: <https://platzi.com/blog/que-es-postgresql/>
- [30] PostgreSQL: el gestor de bases de datos a fondo. Digital Guide IONOS. 2022. Tomado de: <https://www.ionos.es/digitalguide/servidores/know-how/postgresql/>

ANEXOS

Anexo A: Script de Creación de la Base de Datos

Este anexo contiene el script utilizado para la creación de la base de datos que respalda el modelo de detección temprana de la deserción en la universidad popular del Cesar.

```
--Script para la creación de la base de datos  
drop database if exists DATOS_ENCUESTA;  
create database DATOS_ENCUESTA;  
CREATE TABLE ESTUDIANTES(  
ID_DOCUMENTO int NOT NULL,  
ID_TIPO_DOCUMENTO char(1) NOT NULL,  
ID_NOMBRE char(35) NOT NULL,  
ID_EDAD INT NOT NULL,  
ID_NUM_CONTACTO CHAR(10) NOT NULL,  
ID_GENERO int NULL,  
ID_VIC_CONFLICTO int NOT NULL,  
ID_ESTADO_CIVIL INT NOT NULL,  
ID_DPTO_NAC CHAR(30) NOT NULL,  
ID_MUN_NAC CHAR(30) NOT NULL,  
ID_TIPO_DISC INT NOT NULL,  
ID_ETNIA INT NOT NULL,  
ID_EMAIL CHAR(50) NOT NULL,  
ID_MAD_CAB_HOGAR INT NOT NULL,  
ID_EDAD_MAYOR INT NOT NULL,  
ID_EDAD_MENOR INT NOT NULL,  
ID_CANT_HIJOS INT NOT NULL,  
ID_OCUP_MADRE INT NOT NULL,  
ID_OCUP_PADRE INT NOT NULL,
```

```
ID_CANT_HERMANOS INT NOT NULL,  
ID_POS_HERMANO INT NOT NULL,  
ID_INTEGRANTES INT NOT NULL,  
ID_TALENTO INT NOT NULL,  
ID_EPS INT NOT NULL,  
ID_SISBEN CHAR(3) NOT NULL,  
ID_LENTES INT NOT NULL,  
CONSTRAINT PK_ESTUDIANTES PRIMARY KEY (ID_DOCUMENTO));
```

```
CREATE TABLE SOCIO_ECONOMICA(  
ID_ESTUDIANTE int NOT NULL,  
ID_OCUPACION int NOT NULL,  
ID_ESTRATO INT NOT NULL,  
ID_TIPO_VIVIENDA int NULL,  
ID_RECIBE_INGRESOS INT NOT NULL,  
ID_ACUDIENTE INT NOT NULL,  
ID_NIVEL_ACADEMICO INT NOT NULL,  
ID_INGRESO_FAMILIAR NUMERIC NOT NULL,  
ID_TIENE_PC INT NOT NULL,  
ID_TIENE_INTERNET INT NOT NULL,  
ID_CEL_SMART INT NOT NULL,  
ID_PLAN_DATOS INT NOT NULL,  
ID_INGRESO_TRABAJA INT NOT NULL,  
CONSTRAINT PK_SOCIO_ECONOMICA PRIMARY KEY  
(ID_ESTUDIANTE));
```

```
CREATE TABLE INFO_ACADEMICA(  
ID_ESTUDIANTE int NOT NULL,  
ID_PROGR_ACTUAL int NOT NULL,  
ID_HERMANOS_IES int NOT NULL,
```

ID_INGRESO_UPC int NOT NULL,
ID_INGRESO_OFERTA INT NOT NULL,
ID_MOTIVACION INT NOT NULL,
ID_VALIDO_BACH INT NOT NULL,
ID_TIPO_COLEGIO INT NOT NULL,
ID_SEMESTRE_INICIO CHAR(6) NOT NULL,
ID_REPITIO_ICFES INT NOT NULL,
ID_PUNTAJE_ICFES INT NOT NULL,
ID_PUESTO_ICFES INT NOT NULL,
ID_LECTURA_CRITICA INT NOT NULL,
ID_MATEMATICA INT NOT NULL,
ID_CIUDADANOS INT NOT NULL,
ID_CIENCIAS INT NOT NULL,
ID_INGLES INT NOT NULL,
ID_SEMESTRES_CURSADOS INT NOT NULL,
ID_UBICACION_SEMESTRE INT NOT NULL,
ID_CREDITOS_APROBADOS INT NOT NULL,
ID_PROMEDIO_ACUMULADO INT NOT NULL,
ID_PROMEDIO_SEMESTRE_ANT INT NOT NULL,
ID_CREDITOS_NO_APROBADOS INT NOT NULL,
ID_CREDITOS_MATRIC INT NOT NULL,
ID_MATERIAS_CANCELADAS INT NOT NULL,
ID_ASIGNATURA_ENCUESTA CHAR(6) NOT NULL,
ID_JORNADA_ACAD INT NOT NULL,
ID_NOTA_ACTUAL INT NOT NULL,
ID_RENDIMIENTO_ACTUAL INT NOT NULL,
ID_NIVEL_SATISFACION_ASIG INT NOT NULL,
CONSTRAINT PK_INFO_ACADEMICA PRIMARY KEY
(ID_ESTUDIANTE));

ANEXO B: código del modelo en Python

Este anexo contiene el código implementado en el lenguaje Python del modelo para la predicción.

```
# Métricas de desempeño
def model_metrics(model, X_test, y_test):
    y_test_pred = model.predict(X_test)
    print(f"Accuracy = {accuracy_score(y_test,y_test_pred)}")
    print(f"Recall = {recall_score(y_test,y_test_pred)}")

#Las métricas F1, precision and recall requieren que se establezca la convención de cuál es la clase positiva
(1)
print(f"F1 score = {f1_score(y_test,y_test_pred)}")

disp = plot_confusion_matrix(model, X_test, y_test, display_labels=model.classes_,
                             cmap=plt.cm.Blues,
                             normalize='true')
disp.ax_.set_title('Matriz de confusión')
plt.show()

def confusion_matrix_labels(model, X_test, y_test):
    """
    Esta función permite graficar la matriz de confusión con los labels en valor predicho correspondientes a:
    - Verdaderos positivos
    - Falsos positivos
    - Verdaderos negativos
    - Falsos negativos
    """
    predicciones = model.predict(X_test)

    df_predicciones = pd.DataFrame({"True": y_test, "Pred": predicciones})
    df_predicciones.head()

    cf_matrix = confusion_matrix(y_test, predicciones)

    group_names = ["True Neg", "False Pos", "False Neg", "True Pos"]
    group_counts = ["{0:0.0f}".format(value) for value in
                    cf_matrix.flatten()]
    group_percentages = ["{0:.2%}".format(value) for value in
                          cf_matrix.flatten()/np.sum(cf_matrix)]
    labels = [f"{v1}\n{v2}\n{v3}" for v1, v2, v3 in
              zip(group_names,group_counts,group_percentages)]
    labels = np.asarray(labels).reshape(2,2)
    sns.heatmap(cf_matrix, annot=labels, fmt="", cmap='Blues')

# Función para obtener la métrica de validación cruzada del modelo

def cross_val(model, X, y):
```

```
kf = KFold(n_splits=10)
scores = cross_val_score(model, X, y, cv=kf, scoring="f1")
print(f"Metricas cross_validation \n{scores.round(2)}")
print("Media de cross_validation", scores.mean().round(2))
```

Función para obtener la métrica del AUC; área bajo la curva ROC

```
def roc_auc_metrics(model, name_model, X_test, y_test):
```

```
    ns_probs = [0 for _ in range(len(y_test))]
    lr_probs = model.predict_proba(X_test)[:, 1]
    # Calculamos el AUC
    ns_auc = roc_auc_score(y_test, ns_probs)
    lr_auc = roc_auc_score(y_test, lr_probs)
    # Imprimimos en pantalla
    print('Sin entrenar: ROC AUC=%.3f % (ns_auc)')
    print(name_model, 'ROC AUC=%.3f % (lr_auc)')
    # Calculamos las curvas ROC
    ns_fpr, ns_tpr, _ = roc_curve(y_test, ns_probs)
    lr_fpr, lr_tpr, _ = roc_curve(y_test, lr_probs)
    # Pintamos las curvas ROC
    plt.plot(ns_fpr, ns_tpr, linestyle='--', label='Sin entrenar')
    plt.plot(lr_fpr, lr_tpr, marker='.', label=name_model)
    # Etiquetas de los ejes
    plt.xlabel('Tasa de Falsos Positivos')
    plt.ylabel('Tasa de Verdaderos Positivos')
    plt.legend()
    plt.show()
```

Dataset Entrenamiento y prueba

```
X_train, X_test, y_train, y_test = train_test_split(X_res, y_res, test_size=0.3, random_state=12)
print (X_train.shape, y_train.shape, X_test.shape, y_test.shape)
```

```
y_train.value_counts()
y_test.value_counts()
```

Modelos

Modelo Random Forest Classifier

```
model_randomforest = RandomForestClassifier(max_depth=2, random_state=0, class_weight = 'balanced')
model_randomforest.fit(X_train, y_train)
```

```
parameteres = {'n_estimators':[40,50,80], 'max_depth':[4,5,6], 'class_weight':['None','balanced']}
```

```
grid = GridSearchCV(model_randomforest, param_grid=parameteres, cv=10)
kf_hiper = KFold(n_splits=10)
model_randomforest = grid.fit(X_train, y_train).best_estimator_
```

```
model_randomforest
```

```
model_metrics(model=model_randomforest, X_test=X_test, y_test=y_test)
```

```
Accuracy = 0.8993516659606211  
Recall = 0.8608809339667808  
F1 score = 0.8953471670289473
```

```
confusion_matrix_labels(model_randomforest, X_test, y_test)
```

```
cross_val(model_randomforest, X_test, y_test)  
Metricas cross_validation  
[0.88 0.88 0.88 0.88 0.88 0.88 0.89 0.9 0.89 0.88]  
Media de cross_validation 0.88
```

```
roc_auc_metrics(model_randomforest,"Random Forest Classifier",X_test, y_test)  
Sin entrenar: ROC AUC=0.500  
Random Forest Classifier ROC AUC=0.966
```

```
importances = model_randomforest.feature_importances_  
features = pd.Series(importances, index=X_train.columns)  
plt.figure(figsize=(10, 25))  
features.plot(kind="barh")  
plt.show()
```

```
# Modelo de Red Neuronal con Autoencoder
```

```
# Para un mejor ajuste en este modelo, se escalan los datos  
t = MinMaxScaler()  
t.fit(X_train)  
X_train_RN = t.transform(X_train)  
X_test_RN = t.transform(X_test)
```

```
# Se define el encoder  
n_inputs = X_train.shape[1]  
visible = Input(shape=(n_inputs,))  
# encoder level 1  
e = Dense(n_inputs*2)(visible)  
e = BatchNormalization()(e)  
e = LeakyReLU()(e)  
# encoder level 2  
e = Dense(n_inputs)(e)  
e = BatchNormalization()(e)  
e = LeakyReLU()(e)  
# bottleneck  
n_bottleneck = n_inputs  
bottleneck = Dense(n_bottleneck)(e)
```

```
# Se define decoder, level 1
```

```
d = Dense(n_inputs)(bottleneck)
d = BatchNormalization()(d)
d = LeakyReLU()(d)
# decoder level 2
d = Dense(n_inputs*2)(d)
d = BatchNormalization()(d)
d = LeakyReLU()(d)
# Capa de salida
output = Dense(n_inputs, activation='linear')(d)
# Se define el modelo del autoencoder
model_RN = Model(inputs=visible, outputs=output)
```

```
# Se realiza la compilación del modelo de autoencoder
model_RN.compile(optimizer='adam', loss='mse')
```

```
model_RN.summary()
Model: "model"
```

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 86)]	0
dense (Dense)	(None, 172)	14964
batch_normalization (BatchNo	(None, 172)	688
leaky_re_lu (LeakyReLU)	(None, 172)	0
dense_1 (Dense)	(None, 86)	14878
batch_normalization_1 (Batch	(None, 86)	344
leaky_re_lu_1 (LeakyReLU)	(None, 86)	0
dense_2 (Dense)	(None, 86)	7482
dense_3 (Dense)	(None, 86)	7482
batch_normalization_2 (Batch	(None, 86)	344
leaky_re_lu_2 (LeakyReLU)	(None, 86)	0
dense_4 (Dense)	(None, 172)	14964
batch_normalization_3 (Batch	(None, 172)	688
leaky_re_lu_3 (LeakyReLU)	(None, 172)	0
dense_5 (Dense)	(None, 86)	14878

Total params: 76,712

Trainable params: 75,680
Non-trainable params: 1,032

```
# Se entrena el modelo de autoencoder para reconstruir la entrada
history = model_RN.fit(X_train_RN, X_train_RN, epochs=10, batch_size=16, verbose=2,
validation_data=(X_test_RN,X_test_RN))
```

```
Epoch 1/10
15836/15836 - 41s - loss: 0.0135 - val_loss: 0.0040
Epoch 2/10
15836/15836 - 39s - loss: 0.0070 - val_loss: 0.0023
Epoch 3/10
15836/15836 - 39s - loss: 0.0049 - val_loss: 0.0021
Epoch 4/10
15836/15836 - 40s - loss: 0.0034 - val_loss: 0.0015
Epoch 5/10
15836/15836 - 40s - loss: 0.0026 - val_loss: 0.0011
Epoch 6/10
15836/15836 - 39s - loss: 0.0022 - val_loss: 0.0011
Epoch 7/10
15836/15836 - 41s - loss: 0.0020 - val_loss: 0.0013
Epoch 8/10
15836/15836 - 39s - loss: 0.0018 - val_loss: 0.0011
Epoch 9/10
15836/15836 - 40s - loss: 0.0016 - val_loss: 7.6036e-04
Epoch 10/10
15836/15836 - 44s - loss: 0.0015 - val_loss: 7.6939e-04
```

```
# Se grafica la función de pérdida
plt.plot(history.history['loss'], label='train')
plt.plot(history.history['val_loss'], label='test')
plt.legend()
plt.show()
```