Instituto Tecnológico y de Estudios Superiores de Occidente

Reconocimiento de validez oficial de estudios de nivel superior según acuerdo secretarial 15018, publicado en el Diario Oficial de la Federación del 29 de noviembre de 1976.

Departamento de Electrónica, Sistemas e Informática

Maestría en Sistemas Computacionales

Analítica Académica: Análisis de la Graduación de Estudiantes de un Posgrado con PNPC

TRABAJO RECEPCIONAL que para obtener el **GRADO** de
MAESTRO EN SISTEMAS COMPUTACIONALES

Presenta: **GUILLERMO CORONADO GONZÁLEZ**

Directora **DRA. MILDRETH ISADORA ALCARAZ MEJÍA**

Tlaquepaque, Jalisco. 20 de julio de 2020.

AGRADECIMIENTOS

El autor desea dar las gracias al CONACYT, institución que hizo posible este trabajo y la finalización del programa, por medio de la beca 498422; al ITESO, que apoyó con otra beca; a los profesores que guiaron en el aprendizaje de todos los temas de la maestría; a los compañeros del programa con los cuales trabajé en equipo en múltiples materias; y a la directora del trabajo, por sus incontables consejos, sugerencias, mejoras y dirección en la realización del presente trabajo.

DEDICATORIA

Dedico este trabajo a mis padres, que desde que era niño buscaron lo mejor para mí y me enseñaron que el camino a seguir en la vida era el del estudio y de los conocimientos. Por siempre creer en mí y por enseñarme que en la vida lo más importante es ponernos retos a nosotros mismos, a través de los cuales nos podemos superar personal y profesionalmente.

RESUMEN

El objetivo principal de este trabajo es comprender la deserción de estudiantes de un posgrado que cuenta con las características particulares de ser un posgrado PNPC en la modalidad de programa con la industria. De tal forma que, si se logran identificar posibles eventos o circunstancias particulares, se pueda trabajar anticipadamente con los alumnos para evitar que esto suceda.

Se realizó un enfoque orientado a datos con una aproximación en aprendizaje máquina, para encontrar modelos de predicción que tuvieran métricas definidas, además, de encontrar el mejor modelo posible para determinar con la entrada de ciertas variables la probabilidad de un alumno de concluir el programa con un grado aceptable de certeza.

Se realizaron diferentes aproximaciones, todas basadas en diferentes algoritmos y paradigmas de aprendizaje máquina, se encontraron resultados prometedores de acuerdo con las métricas definidas, que ayudaron a entender mejor la deserción del programa de posgrado en estudio.

Este trabajo está dividido en cinco secciones: en la primera se exploran trabajos previos similares al que se presenta; en la segunda se presenta el sustento teórico con el que se trabajó; en la tercera se desglosa el desarrollo del trabajo; en la cuarta se presentan los resultados obtenidos a partir de distintas aproximaciones; y en el quinto se detallan las conclusiones, trabajo futuro y recomendaciones a quienes deseen retomar este trabajo o un trabajo similar.

TABLA DE CONTENIDO

LISTA DE FIGURAS

LISTA DE TABLAS

LISTA DE ACRÓNIMOS Y ABREVIATURAS

SSE	*Error Sum of Squares*
TP	*True positive*
FP	*False positive*
FN	*False negative*
IDI	Investigación, Desarrollo e Innovación
RELU	*Rectified Linear Unit*

1. INTRODUCCIÓN

Resumen: *En este capítulo se presenta brevemente la justificación del objeto de estudio, la definición del problema, objetivo general y específico, y un resumen de los trabajos relacionados con la deserción escolar, se verán trabajos realizados en universidades de otros países, para programas de licenciatura y de posgrado, los modelos que propusieron, las métricas que utilizaron para medir sus trabajos y los resultados que obtuvieron.*

1.1. Justificación

Este trabajo busca encontrar las razones que llevan a un estudiante de posgrado a egresar definitivamente de sus estudios, basándonos en variables previamente recolectadas.

Se busca encontrar las razones de deserción para encontrar formas de apoyar al estudiante a terminar sus estudios en tiempo y forma. De esta manera, se busca incrementar el porcentaje de egreso de estudios en el programa de Maestría en Sistemas Computacionales del ITESO.

De acuerdo con el estudio realizado por [1] el abandono en estudios de posgrado puede ser visto de dos formas: el abandono temprano, es decir en las primeras semanas de iniciado el programa de estudios; suele tener una relación con el hecho de que el estudiante tenga una beca o no y también la demanda de tiempo para estudiar o elaborar trabajos que puede tener un impacto negativo en las exigencias laborales o de atención familiar. La segunda forma de ver esto es el abandono tardío o el hecho de no tener una graduación, este problema se asocia frecuentemente con la elaboración de la tesis y el rol que juega el asesor.

De acuerdo con el estudio realizado por [2], durante el ciclo escolar 2015-2016, la tasa de abandono escolar en educación media superior fue de 15.5% en términos absolutos. Según el mismo estudio [2], las principales causas asociadas a la deserción o abandono escolar son las siguientes:

- Económicas: en relación con la falta de dinero en el hogar.
- Institucionales o escolares: por razones Inter sistémicas, por ejemplo, la oferta educativa, la desigualdad en la calidad de los servicios educativos y los mecanismos de acceso; y por razones intra sistémicas, en la cual existen “prácticas pedagógicas inadecuadas, formación docente limitada y condiciones laborales precarias, infraestructura y equipamiento insuficiente”.
- Individuales: La desmotivación y el desinterés por la escuela y el aprendizaje en general.

Más de la tercera parte de los jóvenes (35.4%) abandonan la escuela por causas económicas, mientras que otra tercera parte (32.3%), la abandona por causas escolares-individuales [2].

En el trabajo realizado por [3] se menciona que realizar predicciones acertadas del futuro rendimiento de los estudiantes, basado en su historial académico es crucial para efectuar intervenciones pedagógicas para asegurar que los estudiantes se graduarán en tiempo.

De acuerdo con lo realizado en [4] a pesar de que en el programa de posgrado bajo estudio existen métodos para encontrar a los mejores estudiantes, como el examen de admisión o el promedio del programa de estudios anterior, en general, sería de gran apoyo encontrar métricas que ayuden a predecir con alta precisión si existirá una graduación y una eventual obtención de grado.

1.2. Problema

Según los estudios previos realizados por los autores en [1-4], existen diversos factores que pueden ser una o más causas fundamentales del porque un alumno de posgrado deserta de su programa de estudios.

En el presente trabajo se busca encontrar específicamente qué variables medibles llevan a un alumno a terminar o no el programa de la Maestría en Sistemas Computacionales, a través de diversas técnicas enfocadas en predicción/clasificación/segmentación de los datos que se recolectan antes de iniciar el programa y mientras el mismo programa transcurre.

1.3 Objetivos

Se busca encontrar un modelo de clasificación que nos ayude a determinar si un estudiante completará el programa de estudios, y que tenga las mejores métricas de certeza posibles.

La intención es tener la información sobre esta predicción antes de comenzar el primer periodo o al finalizar el mismo.

Se busca utilizar diversos modelos de algoritmos de aprendizaje, de Machine Learning, comparar entre los modelos obtenidos e identificar cuál de ellos satisface mejor las necesidades del presente problema.

1.4 Trabajos relacionados

La analítica académica tiene diferentes aproximaciones. Se encontraron múltiples trabajos relacionados, los cuales se centran en buscar el mejor rendimiento de los estudiantes o evitar la deserción de estos.

En [4] se presenta un trabajo de estudiantes de Doctorado en Ciencias de la Computación, se generan múltiples hipótesis sobre el por qué los alumnos terminan o no el programa y se trabaja con variables de las siguientes categorías:

- Registro académico
- Rendimiento en los exámenes de admisión
- Factores individuales

El análisis se realizó utilizando un modelo de regresión lineal y se utilizó el coeficiente de correlación de Pearson para medir la precisión de dicho modelo. Realizaron dos modelos diferentes, en donde el segundo demostró algunas de sus hipótesis: la primera, que una selección de cursos fundamentales tiene relación con la terminación del programa; la segunda, que el hecho de haber escrito una tesis de Maestría también tenía relación con el éxito de terminar el programa.

En los resultados obtenidos, después de experimentos utilizando únicamente una variable y aproximaciones multivariables, se encontró que solo tres variables eran suficientes para encontrar un modelo satisfactorio: la calificación obtenida en el curso de algoritmos, el promedio obtenido de un conjunto de cursos denominados "núcleo" y si el estudiante había escrito una tesis de maestría.

En [3] se presenta un estudio basado en la información de dos carreras: Ingeniería Mecánica e Ingeniería Aeroespacial. Se desarrolló un sistema de predicción de doble capa, el cual consistió en una capa de predictores bases y una capa de conjunto. En la capa de conjunto, se sintetiza la información obtenida de los predictores bases, así como de los resultados de las capas de conjunto de los predictores de ensamble de periodos académicos pasados. Para generar los predictores, se utilizaron diferentes algoritmos de aprendizaje supervisado y no supervisado como, regresión lineal, regresión logística, redes neuronales y k-means.

En sus resultados, encontraron que los errores de predicción decrecen conforme avanzan los periodos de estudio. Se aplicaron diferentes tamaños de clústeres: 5, 10, 20. También, encontraron que sus algoritmos basados en segmentación de cursos fueron los que mejores resultados dieron. Entre las segmentaciones realizadas, se encontraron: mismo departamento, prerrequisito directo únicamente y serie de prerrequisitos.

En [5] se presenta una comparativa entre diferentes sistemas de analítica académica de diferentes universidades en Estados Unidos, cada sistema tiene entradas de datos distintas, como, por ejemplo, las personalidades de los estudiantes, hábitos sociales y comportamientos en un sistema; En otro, se mencionan datos como asistencias y calificaciones de materias. Después de resaltar una comparativa entre las entradas de dichos sistemas, el trabajo presenta los tipos de datos con los que recomiendan trabajar:

- Demográficos
- Habilidad académica
- Rendimiento académico
- Historial académico
- Información académica
- Información de participaciones
- Información institucional
- Soporte financiero

En [5] se propone investigar y diseñar un sistema inteligente de predicción y recomendación, que determine cuál es la variable que más afecta el éxito de los estudiantes en educación superior. Se propone utilizar Redes Neuronales y Árboles de Decisión para realizar predicciones con base en datos previos, con la siguiente metodología:

1. Paso uno: Los datos recolectados son alimentados para una etapa inicial de aprendizaje y para la creación de las reglas de los árboles de decisión.
2. Paso dos: Se hace un sistema que acepte retroalimentación para ajustar las reglas.
3. Paso tres: Los datos de los estudiantes actuales se alimentan al sistema para predecir y otorgar información para el paso 2.

En [6] se desarrolla un modelo basado en árboles de decisión para realizar predicciones sobre el rendimiento académico de los estudiantes, únicamente para un curso (Estructura de datos) de segundo año de Ciencias de la Computación de múltiples Universidades Nigerianas. Las variables con las que se trabajó fueron las siguientes:

- Calificación obtenida
- Género
- Fortaleza financiera
- Motivación

En los resultados, se encuentra que la variable más importante, es decir el primer nodo, corresponde a las calificaciones obtenidas, mientras que el siguiente nodo en importancia, pasa a ser el género, donde las mujeres obtuvieron mejores resultados que los varones, al tener que 59.03% de las mujeres aprueban y 55.10% de los hombres también lo hacen.

En [7] se evalúa el rendimiento de estudiantes que buscan el grado de Maestría en Aplicaciones de Computación, de la Universidad de Pune, de India, se utiliza el modelo de Redes Neuronales para seleccionar atributos de un conjunto de datos, dichos atributos seleccionados fueron:

- Educación previa
- Conocimientos previos de computación
- Si los padres del alumno obtuvieron educación
- Porcentaje de graduación
- Porcentaje de Asistencias
- Calificaciones

En [7], los autores realizaron una comparación de los resultados obtenidos entre un modelo que utiliza Árboles de Decisión para selección de atributos y un modelo de Redes Neuronales que mide la precisión únicamente con dichos atributos, contra un modelo de Redes Neuronales que mida la precisión con todos los atributos, los pasos que siguieron fueron los siguientes:

1. Recolección de datos
2. Preprocesamiento
3. Árbol de decisión
    a. Creación de reglas
    b. Selección de atributos
    c. Red Neuronal
    d. Medición de la Precisión
4. Red Neuronal con todos los atributos
    a. Medición de la Precisión

En los resultados obtenidos, se encontró que las variables de principal importancia fueron tres. En primer lugar, la educación previa, en segundo lugar, conocimientos previos de computación y, por último, si alguno de los progenitores había tenido acceso a la educación.

En [8] se presenta una predicción de rendimiento de estudiantes de preparatoria de la Universidad Autónoma de Zacatecas, entre los atributos con los que se trabajó se encentran:

- Información personal y familiar para identificar factores que afectan rendimiento escolar
- Un estudio socioeconómico que es realizado al ingresar al programa
- Las calificaciones obtenidas en cursos al finalizar cada semestre

En [8] se trabajó con un modelo de clasificación basado en Árboles decisión, se utilizaron cinco algoritmos distintos con los cuales se realizaron múltiples experimentos. Se midió la precisión obtenida para cada experimento. Los resultados se representan en verdaderos positivos (TP), verdaderos negativos (TN) y como precisión.

Entre los resultados obtenidos, se encontró que las variables más importantes, fueron: las calificaciones obtenidas en algunos cursos de preparatoria, el nivel de motivación y el promedio obtenido en la preparatoria.

2 CONCEPTOS BÁSICOS

Resumen: *En este capítulo se presentan las bases teóricas y conceptuales sobre la deserción escolar y su impacto. Se detalla lo que es machine learning y los paradigmas de aprendizaje que se utilizaron para el presente trabajo clasificados como (aprendizaje supervisado y no supervisado).*

2.1 Aprendizaje máquina

El aprendizaje máquina es un conjunto de métodos que las computadoras usan para realizar y mejorar predicciones o comportamientos basados en datos [9].

Un algoritmo de aprendizaje máquina aprende un modelo estimando parámetros o aprendiendo estructuras. El algoritmo es guiado por una función de costo, la cual debe de ser minimizada, comúnmente apoyada por un algoritmo de optimización para encontrar el mejor modelo posible. [9]

Las máquinas son mejores que los humanos en múltiples tareas, como en juegos o en predicción del clima, y tienen la gran ventaja en términos de velocidad, escalabilidad y reproducibilidad de los problemas.

Un ejemplo en el que nos puede ayudar el aprendizaje máquina es en la predicción del precio de una casa. Si se cuenta con datos previos como la cantidad de metros cuadrados, cuartos, ubicación geográfica, entre otros y se conocen los precios finales, entonces se podrá obtener una predicción de precio de otra casa con valores diferentes para metros cuadrados, cuartos y ubicación geográfica.

Otra aplicación en la cual nos puede ayudar el aprendizaje máquina es en la segmentación de los clientes que tiene un corporativo. Quizás dicha empresa quiere tener a sus clientes en diferentes segmentos para aplicar distintas estrategias de mercadotecnia a cada grupo, o por el tipo de producto y servicio que ofrece a cada uno de los grupos, y así ayudar a reducir costos y aumentar la productividad.

Machine learning, su término en inglés, se divide en dos tipos de aprendizaje, supervisado y no supervisado. Los algoritmos que se observarán a continuación pertenecen a la rama de aprendizaje supervisado: Árboles de decisión binaria, regresión logística, bosques aleatorios y redes neuronales.

Los algoritmos de aprendizaje supervisado nos permiten generar modelos de regresión o clasificación basado en aprendizaje de datos en los cuales se tiene claro cuáles son las clasificaciones con las que se quiere trabajar.

2.1.1 Aprendizaje supervisado

El objetivo del aprendizaje supervisado es utilizado para aprender un modelo de predicción, que mapee múltiples entradas, a una salida. Si dicha salida es categórica, se llama clasificación, si es numérica se llama regresión [9].

El aprendizaje supervisado nos es útil cuando se conoce el resultado al que se desea llegar. Por ejemplo, si se conoce la cantidad de ventas realizadas por una empresa en un periodo determinado de tiempo y se tiene clasificado como bueno, malo o regular, se puede entrenar un algoritmo de aprendizaje supervisado. Posteriormente, se podrá predecir lo que se espera vender en el futuro y el tipo de ventas que obtendrá la empresa.

2.1.1.1 Regresión logística

La regresión logística modela las probabilidades para dos posibles resultados. Su uso principal es el de la clasificación.

Los modelos lineales no entregan probabilidades. Simplemente interpolan entre puntos, lo cual no se puede interpretar como probabilidades. Los modelos lineales no pueden extenderse a clasificación con múltiples clases tampoco.

En lugar de buscar ajustar una línea recta o un hiperplano, la regresión logística ajusta la salida de una ecuación lineal entre 0 y 1 [10].

Para encontrar el mejor modelo en regresión logística con base en un conjunto de datos, se utiliza, un algoritmo de optimización, a partir del cual, se busca el mejor modelo posible. Para determinar si un modelo es el mejor, el algoritmo de optimización se apoya de una función de "costo", la cual nos ayuda a determinar qué tan bueno es el modelo actual con el conjunto de datos utilizado para entrenar.

El modelo de Regresión logística falla cuando las variables interactúan entre ellas o cuando su relación con el resultado esperado es lineal.

2.1.1.2 Árboles de decisión

Los modelos basados en árboles de decisión realizan particiones del conjunto de datos de acuerdo con ciertas variables. Las particiones finales son conocidas como terminales u hojas. Estos árboles pueden ser utilizados para clasificación y para regresión [9].

Existen múltiples fórmulas para la creación de un árbol de decisión. La utilizada en este trabajo es la fórmula desarrollada por Shannon [11].

La fórmula de Shannon es la siguiente:

$$H(S) = -p_{(+)}log_2 p_{(+)} - p_{(-)}log_2 p_{(-)}$$

Donde S representa al subconjunto de datos que se utilizará como "partición", $p_{(+)}$ representa al porcentaje de positivos utilizados (es decir, clasificados positivamente) y $p_{(-)}$ representa al porcentaje de negativos utilizados (es decir, clasificados negativamente).

Posteriormente se aplica la fórmula de la ganancia de la información, que representa la caída de la entropía.

La fórmula de la ganancia es la siguiente:

$$Gain(S, A) = H(S) - \sum_{V \in Values(A)} \frac{|S_v|}{|s|} H(S_V)$$

Esta fórmula se aplica para cada una de las columnas del conjunto de datos y a partir del resultado se escoge la que tenga el número mayor. Dicha selección corresponde a la primer "hoja" del árbol de decisión y representa a la columna de datos que tiene mejor correlación con el resultado. Este proceso se repite con las demás variables para ir determinando las hojas posteriores del árbol.

2.1.1.3 Bosques Aleatorios

Los bosques aleatorios funcionan a partir de múltiples árboles de decisión binaria, se crean múltiples árboles sin alguna correlación que funcionan como si fueran un "comité" para escoger la mejor opción. Cada uno de estos árboles en el bosque consideran un subconjunto aleatorio de variables. Lo cual incrementa la diversidad y robustece el resultado a obtener [9].

2.1.1.4 Redes Neuronales

Las redes neuronales consisten de un conjunto de unidades de procesamiento simples que se comunican enviándose señales una a otra a través de múltiples conexiones con determinadas ponderaciones. [12]

Cada unidad de procesamiento realiza un trabajo sencillo:

1. Recibe una entrada de sus vecinos o fuentes externas.
2. Utiliza esa entrada para propagar una señal a otras unidades.

Existen tres tipos de unidades:

- Unidades de entrada, reciben datos por fuera de la red neuronal
- Unidades de salida, envían datos fuera de la red neuronal
- Unidades ocultas, cuyas señales se quedan dentro de la red neuronal

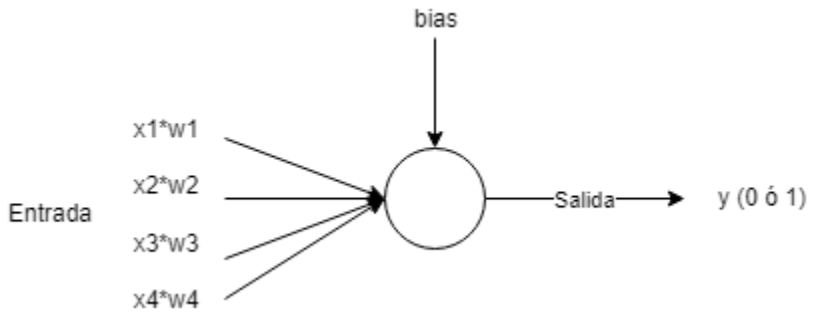

Figura 1 Unidad de procesamiento de una red neuronal

En la figura 1, se muestra una unidad sencilla, siendo x, el vector de entrada y w, la ponderación por la cual se multiplica esta entrada.

La entrada de una unidad es simplemente la suma de la multiplicación de los pesos por las entradas, sumado al bias. La salida z, es un vector del mismo tamaño que la entrada x:

$$z = \sum_{j}^{J} w_j * x_j + b$$

Debido a que las salidas de las unidades suelen ser lineales, se utilizan funciones de activación para alterar su comportamiento, de manera que sea no-lineal.

El vector y es del mismo tamaño que la entrada. Esta salida se utiliza como entrada para la siguiente capa de neuronas.

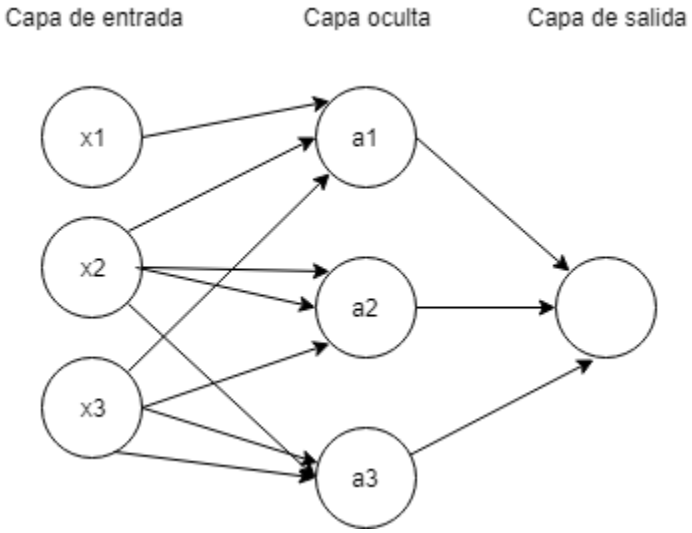

Figura 2 Red neuronal simple

En la figura 2 podemos observar un ejemplo sencillo, con una capa oculta. Si deseáramos que esta red calcule una predicción, tendríamos que hacer lo siguiente:

Suponiendo que tenemos una entrada $[x_1, x_2, x_3]$ esta debe ser multiplicada por una matriz de pesos para calcular su salida $[[w_{11}, w_{12}, w_{13}], [w_{21}, w_{22}, w_{23}], [w_{31}, w_{32}, w_{33}]\ ]$, de tal forma que

$$z_1 = x_1 * w_{11} + x_2 * w_{12} + x_3 * w_{13} + b_{11}$$

$$z_2 = x_1 * w_{11} + x_2 * w_{22} + x_3 * w_{23} + b_{12}$$

$$y\ z_3 = x_1 * w_{31} + x_2 * w_{32} + x_3 * w_{33} + b_{13}\ .$$

Posteriormente, se activan dichas salidas con la función sigmoide:

$$a_1 = sig(z_1)$$

$$a_2 = sig(z_2)$$

$$a_3 = sig(z_3)$$

Para calcular la predicción final se hace un cálculo similar a:

Tenemos una entrada $[a_1, a_2, a_3]$, y tenemos una nueva matriz $[w_{11}, w_{12}, w_{13}]$

$$a = sig(a_1 * w_{11} + a_2 * w_{12} + a_3 * w_{13} + b_{11})$$

Las redes neuronales aprenden por medio del cálculo del error para el modelo actual (es decir, la ponderación de todos los pesos) y la reducción (o aumento) de dichos pesos.

2.1.2 Aprendizaje no supervisado

¿Qué sucede cuando no sabemos cómo clasificar los datos? Es decir, tenemos datos, pero no sabemos cómo separarlos o segmentarlos. Aquí es donde entran los algoritmos de aprendizaje no supervisado, los cuales nos ayuda a realizar una segmentación de la información que tenemos.

Un ejemplo muy claro de esto es el *clustering*, que nos ayuda a segmentar la información en clases diferentes. Existen múltiples algoritmos para trabajar el *clustering*, pero el que tiene suma importancia para el presente trabajo, es el algoritmo de *K-Means* [13].

2.1.2.1 K-Means

Como se mencionó previamente, K-Means es un algoritmo de aprendizaje no supervisado. Su objetivo es generar clasificación a partir de datos que no se encuentran propiamente segmentados.

El algoritmo intenta agrupar un conjunto de datos en *k* grupos, en donde cada grupo tiene una varianza similar. El número de grupos es especificado como hiper parámetro [13]:

1. Se inicializan *k* cantidad de centroides en posiciones aleatorias.
2. Para cada renglón del conjunto de datos:
    a. Se calcula la distancia entre el renglón y todos los centroides.
    b. Se asigna el centroide más cercano a dicho renglón.
3. Los centroides se mueven al promedio de sus respectivos clústeres.
4. Los puntos dos y tres se repiten hasta que no se observe un cambio de asignación de centroide en los datos.

K-Means asume que todos los clústeres tienen forma convexa y que todas las variables se encuentran dentro de la misma escala [13].

Para determinar la cantidad de clústeres a elegir existen múltiples métodos, uno de ellos es el conocido como "Elbow Method*"*. Este método se basa en una métrica llamada inercia, la cual representa la suma de los cuadrados de las distancias de cada uno de renglones de datos a su centroide más cercano, Para determinar el número óptimo de clústeres, se tiene que seleccionar el valor de *k* en el cual se hace un "codo" y la distorsión comienza a decrecer de forma lineal [13].

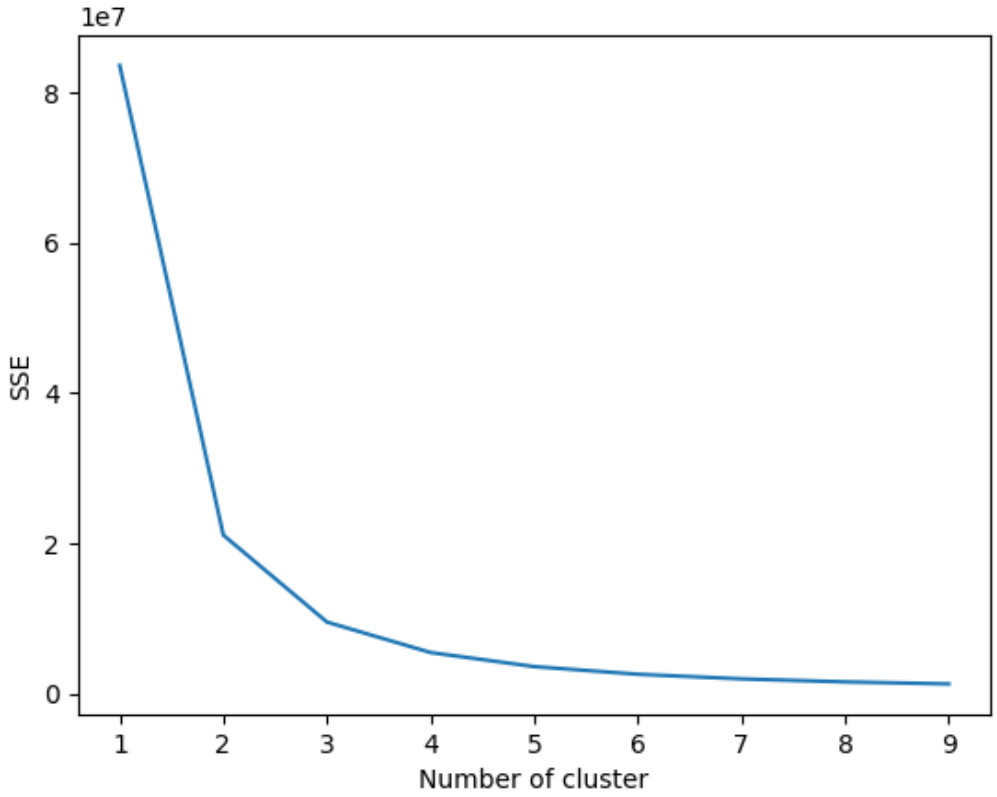

Figura 3 Gráfica método de codo

En la figura 3 se puede observar un ejemplo donde “SSE” representa la inercia. Para este ejemplo, la cantidad de clústeres que se escogería sería 2.

3 DESARROLLO DEL PROYECTO

Resumen: *En este capítulo se presenta en detalle el desarrollo metodológico que incluye los pasos que se siguieron para llegar al resultado final y las métricas con las que se midió el resultado obtenido.*

3.1 Introducción

Para encontrar el modelo adecuado para resolver el presente problema, se realizaron diferentes aproximaciones utilizando diversas técnicas, con diferentes variables y con distintos algoritmos de aprendizaje máquina.

Antes de aplicar alguna técnica, es importante conocer los datos seleccionados para este estudio.

3.2 Conjunto de datos

Se trabajó en un conjunto de datos que consistió en 78 estudiantes del programa, que corresponde a 4 generaciones completas. Dicho conjunto se clasificó en dos diferentes clases:

1. Bajas, es decir, aquellos que no terminaron el programa.
2. Egresados, es decir, aquellos que terminaron el programa y se titularon o se encuentran en proceso de titularse.

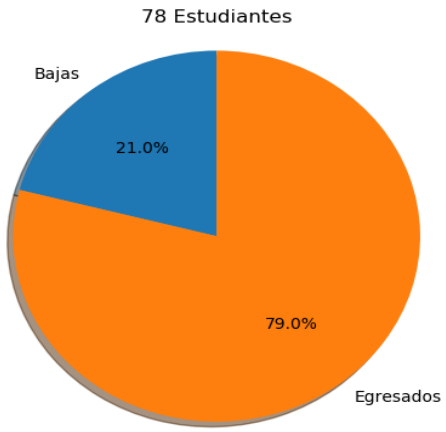

Figura 4 Porcentaje de Egreso del programa

En la figura 4 se puede observar el porcentaje de estudiantes que se dieron de baja del programa y el porcentaje de egresados. Se puede observar que las bajas son menores que la cantidad de egresados, lo cual marca un desbalance en la información.

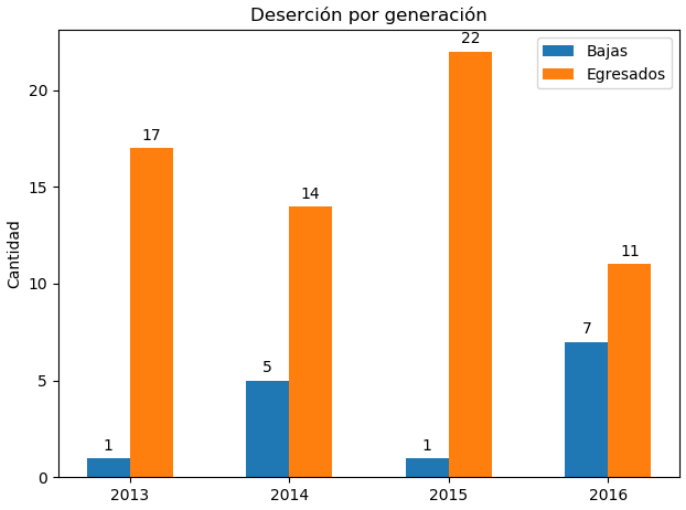

Figura 5 Egreso por generación

Los 78 estudiantes se encuentran distribuidos en diferentes generaciones, desde la primera generación del programa hasta la última de la cual se tienen datos, como se muestra en la figura 5.

Las variables con las que se trabajó para cada estudiante fueron las siguientes:

1. Calificación Kardex, es decir, la categoría que se obtuvo a partir del promedio obtenido en licenciatura.
2. Calificación propedéutica, es decir, la categoría que se obtuvo a partir del promedio obtenido en el curso propedéutico obligatorio que funciona como filtro de admisión al programa.
3. Misma empresa, es decir, si durante todo el programa trabajaron para la misma empresa.
4. Grupo edad, es decir el grupo de edad al que pertenece el estudiante al entrar al programa.
5. Licenciatura, que representa la categoría a la que pertenece el título de licenciatura del estudiante.
6. Calificación Matemáticas, que representa la calificación obtenida en la asignatura “Matemáticas Avanzadas para Computación” cursada en el primer semestre del programa.
7. Calificación Algoritmos, que representa la calificación obtenida en la asignatura “Análisis y diseño de algoritmos” cursada en el primer semestre del programa.
8. Calificación IDI, que representa la calificación obtenida en la asignatura “Innovación, Diseño e Investigación I” cursada en el primer semestre del programa.

Para convertir cada una de las variables en las categorías previamente mencionadas se realizó un análisis de la información basado en diagramas de caja y en datos estadísticos de cada variable.

3.2.1 Calificación Kardex

Se trabajó en diferentes estadísticas para cada una de las clasificaciones trabajadas, como se puede observar en la tabla 1, comenzando por el promedio, la desviación estándar, el mínimo y el máximo obtenidos, lo cual nos permite entender la información obtenida. Es interesante observar que no existe tanta diferencia entre los promedios y desviaciones estándar de cada clase y que, por ejemplo, la diferencia entre mínimo y máximo es más alto para los egresados (21.3) que para las bajas (14.04).

Clase	Promedio	Desviación Estándar	Mínimo	Máximo
Baja	89.24	3.72	82.9	96.94
Egresado	89.74	4.42	76.7	98.0

Tabla 1 Métricas Calificación Kardex

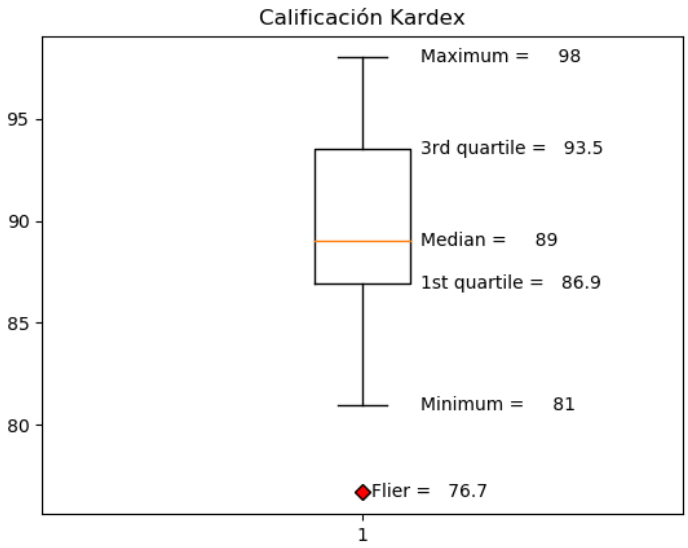

Figura 6 Diagrama de caja de Calificación Kardex

En la figura 8 se puede observar un diagrama de caja, en el cual se pueden observar los mismos datos previamente mostrados, a partir de los cuales se categorizan las variables:

1. Los menores o iguales a 86.9, es decir el primer cuartil, pertenecen a la primera categoría.
2. Los mayores a 86.9 y menores o iguales a 89.0 pertenecen a la segunda categoría.
3. Los mayores a 89.0 y menores o iguales a 93.5 pertenecen a la tercera categoría.
4. Los mayores a 93.5 pertenecen a una cuarta categoría.

3.2.2 Calificación Propedéutico

Para la calificación del propedéutico, se trabajó con las mismas estadísticas que para la calificación del Kardex, como se observa en la Tabla 2, la diferencia entre clasificaciones es mínima. Siendo que únicamente el promedio de todos los egresados es ligeramente superior a los que se dieron de baja, por 0.21.

Clase	Promedio	Desviación Estándar	Mínimo	Máximo
Baja	8.0	1.32	6	10
Egresado	8.21	1.19	6	10

Tabla 2 Métricas Calificación Propedéutico

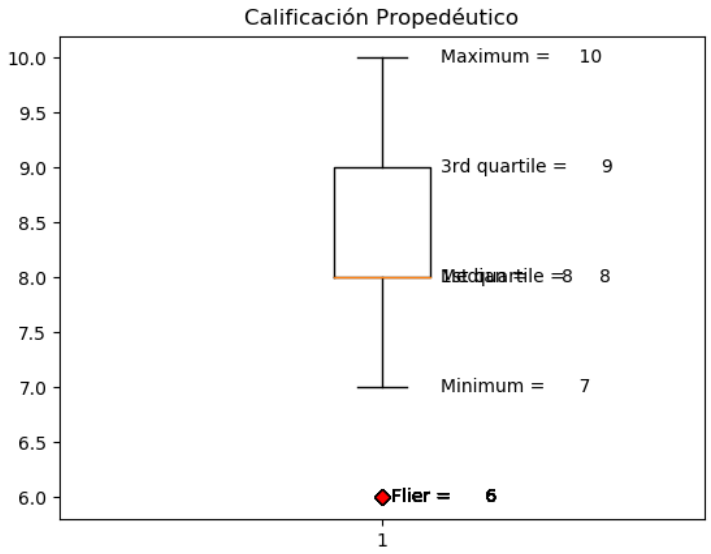

Figura 7 Diagrama de caja de Calificación Propedéutico

En la figura 9 se puede observar un diagrama de caja, en el cual se observan los datos obtenidos de la calificación de propedéutico para todo el conjunto de datos. A partir del diagrama obtenido se seleccionaron las siguientes categorías:

1. Los menores o iguales a 8 pertenecen a la primera categoría.
2. Los que fueron iguales a 9 pertenecen a la segunda categoría.
3. El resto (es decir los que obtuvieron 10), pertenecen a la tercera categoría.

3.2.3 Misma Empresa

Esta variable se refiere a si el estudiante al momento en el que terminó el programa seguía trabajando en la misma empresa que cuando comenzó. Al ser una variable binaria, no es necesario realizar una separación en múltiples categorías.

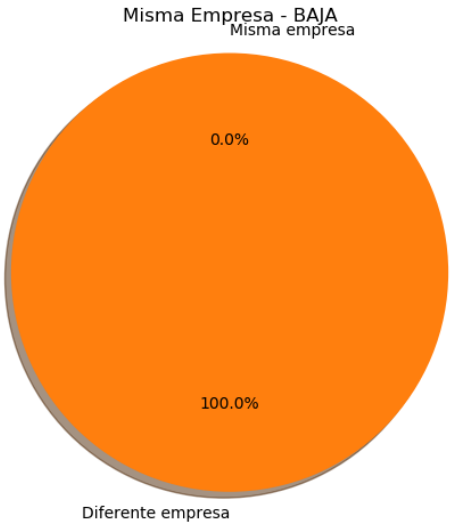

Figura 8 Porcentaje de cambio de empresa

En la figura 10 se puede observar que todos los que se dieron de baja del programa se cambiaron de empresa, en contraste con la figura 11, que muestra un porcentaje bastante diferente, se observa que el 70% de los estudiantes tuvieron un cambio de empleo también.

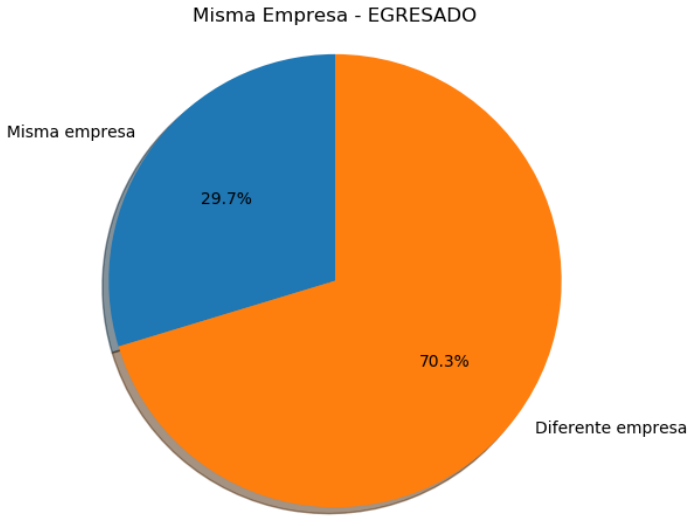

Figura 9 Porcentaje de cambio de empresa

3.2.4 Edad

Para esta variable se calculó la edad que tenían los estudiantes al momento de comenzar el programa. En la figura 12 se muestra el diagrama de caja con la distribución obtenida en la variable.

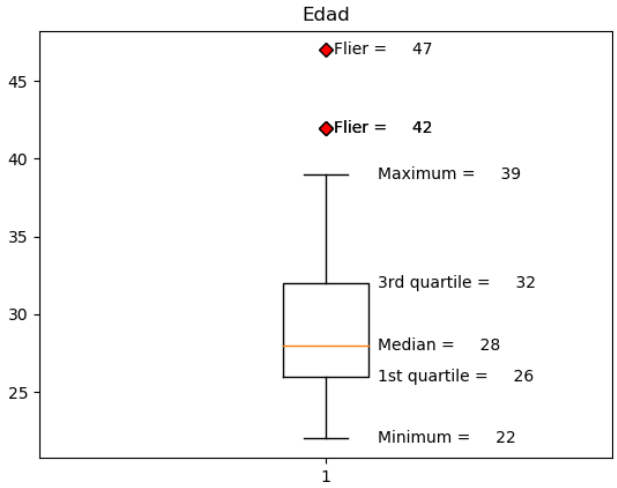

Figura 10 Diagrama de caja de Edad

Se determinó que las categorías serían de la siguiente manera:

1. Menores de 26 años.
2. Mayores de 26 años y de menor o igual edad a 28.
3. Mayores de 28 años y de menor o igual edad a 32.
4. Mayores de 32 años.

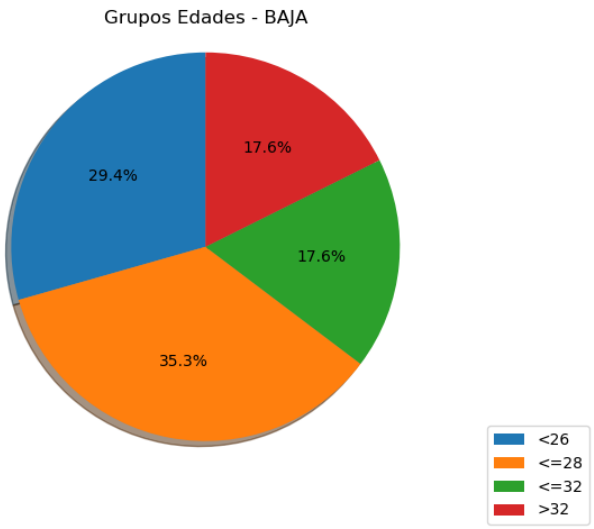

Figura 11 Porcentaje de grupos de edad

En las figuras 13 y 14 se pueden observan gráficas de la distribución de edades, se observa la distribución de los datos en las categorías previamente seleccionadas a partir del diagrama de caja. Es interesante observar que el grupo con mayor incidencia en ambos es el segundo grupo, es decir, los que se encuentran en el rango de edad de 26 a 28 años.

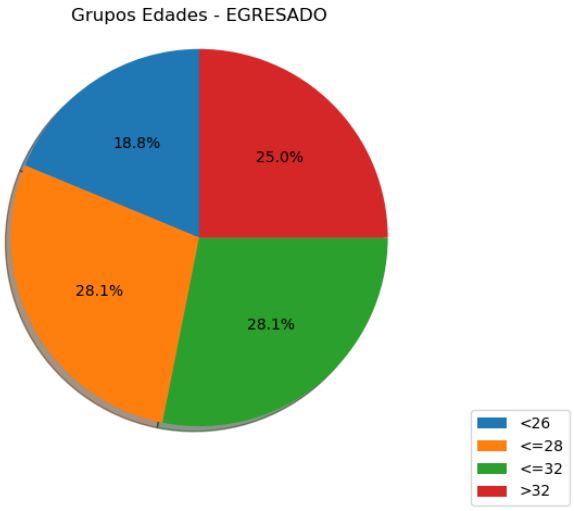

Figura 12 Porcentaje de grupos de edad

3.2.5 Licenciatura

Para la licenciatura se dividió en categorías de acuerdo con el tipo de licenciatura de la que el estudiante obtuvo título, dichas categorías fueron las siguientes:

1. Computación
2. Electrónica
3. Software
4. Mecatrónica
5. Informática
6. Finanzas
7. Biomédica

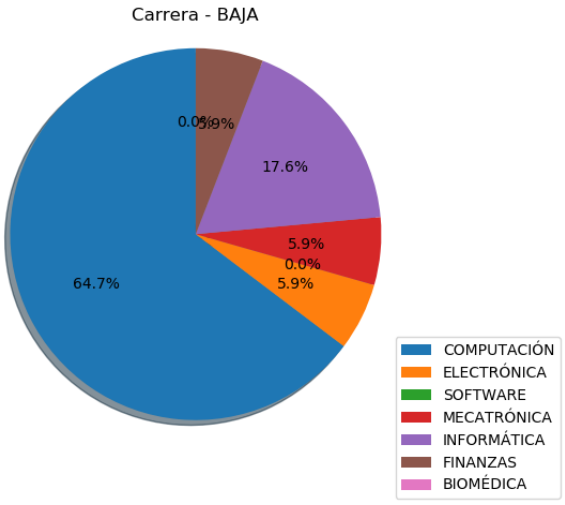

Figura 13 Porcentaje de carreras

En la figura 15 se puede observar los porcentajes de estudiantes que se dieron de baja del programa para cada una de las categorías previamente mencionadas.

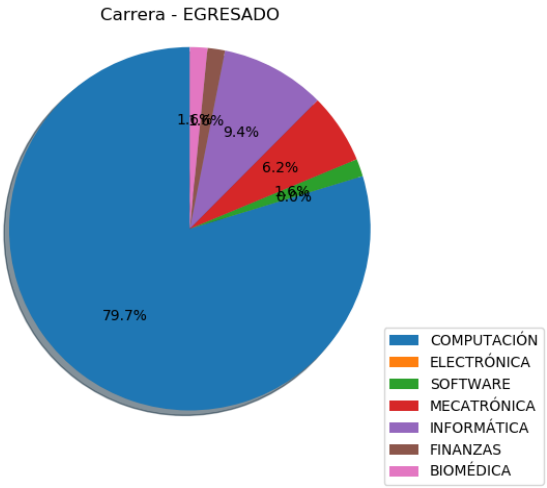

Figura 14 Porcentaje de carreras

En la figura 16 se observan los porcentajes de estudiantes que terminaron el programa para las mismas categorías. Dado el alto porcentaje de estudiantes de Computación, el mayor porcentaje de Bajas y Egresados es de Computación, como es de esperarse.

3.3 Limpieza de datos

En la limpieza de datos se desecharon aquellos registros incompletos o que pudieran causar confusión a los algoritmos de aprendizaje. Para cumplir con este objetivo, se realizó lo siguiente:

1. Eliminar registros que no tuvieran datos.
2. Unir las diferentes fuentes de datos en una.

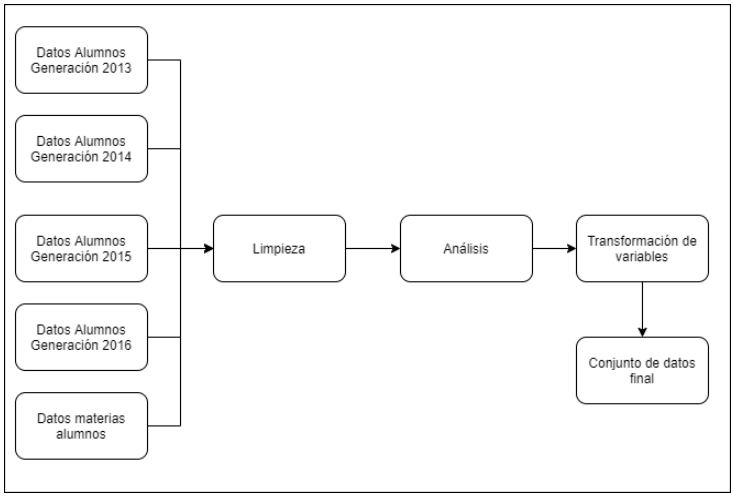

Figura 15 Proceso de Limpieza de datos

En la figura 6 se observan las diversas fuentes de datos que se procesaron para obtener un solo conjunto de datos. Estas fuentes se encontraban en archivos con formato xlsx donde había información de los alumnos de cada una de las generaciones, incluyendo información de las calificaciones por materia por alumno.

3.4 Análisis y Transformación de Variables

El primer problema que se encontró fue que algunas de las variables se encontraban dentro de un dominio continuo y otras variables se encontraban dentro de un dominio discreto. Esto podría causar resultados inciertos para los modelos de predicción, por lo cual se crearon categorías para cada una de las variables.

Para hacer esta conversión de variables continuas a categóricas, se realizó un análisis basado en las siguientes estadísticas:

1. El promedio de dicha variable.
2. Su desviación estándar.
3. Su mínimo y máximo.
4. Una representación visual por medio de un diagrama de caja, que permitiera observar las categorías por cuartiles.

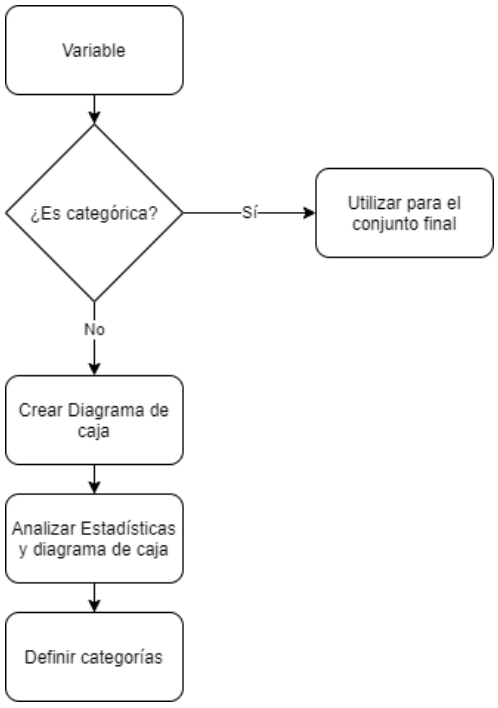

Figura 16 Proceso de Análisis y Transformación de Variables

En la figura 7 se observa un diagrama en el cual se explica el flujo de análisis aplicado a las variables con las que se trabajó.

De acuerdo con las estadísticas previamente generadas y a los diagramas obtenidos, se hizo una transformación de variables. Aquellas variables que no se podían trabajar directamente en el modelo matemático se transformaron a variables binarias, numéricas o categóricas.

Se realizó un análisis de correlación para determinar el impacto que cada una de las variables analizadas tienen con el resultado final, de tal forma que se identifiquen aquellas variables que no aporten al modelo final y puedan ser descartadas.

3.5 Entrenamiento de modelos de predicción

Esta etapa consistió en entrenar un modelo de aprendizaje, en el cual nos apoyaremos para determinar futuras predicciones.

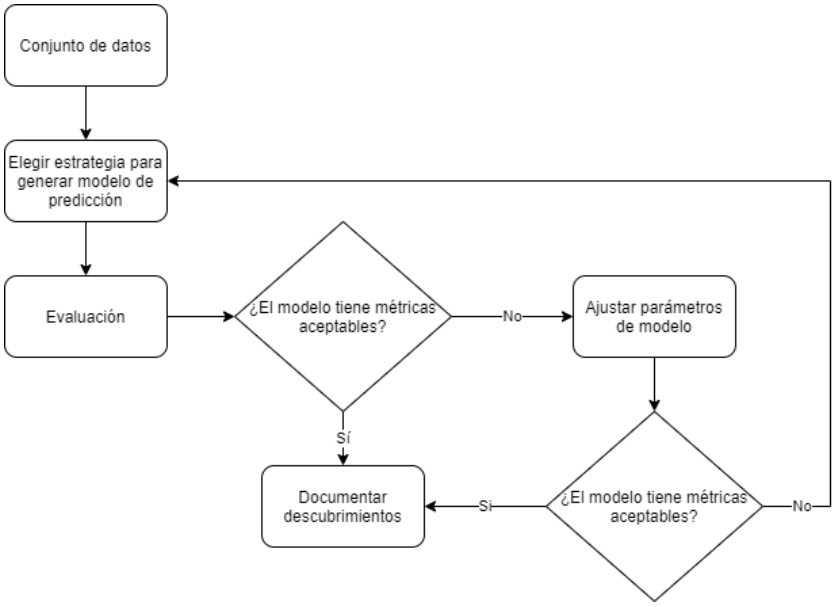

Figura 17 Entrenamiento de un modelo de predicción

En la figura 17 se observan los pasos que se siguieron para llegar al modelo adecuado. En la sección de elección de estrategia para generar el modelo de predicción, se iteró con diferentes hipótesis para finalmente comparar resultados.

3.5.1 Primer Enfoque

El primer enfoque fue utilizado para realizar una primera aproximación exclusivamente con redes neuronales.

Para desarrollar este modelo, se utilizó el lenguaje de programación Python3 con las librerías *Pandas, Matplotlib, Numpy, y Scikit Learn*, principalmente.

Las variables con las que se trabajó fueron seleccionadas porque ambas son numéricas y se encuentran dentro del mismo rango, lo cual nos podría dar un primer acercamiento al modelo final y partir de los resultados obtenidos como comparativa:

- Calificación de propedéutico
- Calificación de Kardex

Se desarrolló una red neuronal basada en el modelo de perceptrón multicapa, utilizando la librería *Scikit Learn*, Se seleccionó un modelo sencillo de una capa oculta y tres neuronas en dicha capa. Esta arquitectura es la que presenta la librería por defecto.

3.5.2 Segundo Enfoque

En este enfoque se buscó realizar una comparativa entre diferentes modelos de aprendizaje y utilizando diferentes variables. El objetivo fue ejecutar cada uno de los algoritmos de aprendizaje con todos los subconjuntos posibles de variables y documentar dichos resultados, de tal forma que se pudiera encontrar aquel que tuviera las mejores métricas.

Se utilizaron cuatro diferentes algoritmos de aprendizaje: *Regresión Logística; Redes Neuronales; Árboles de decisión binaria; y Bosques aleatorios.* Se buscó que se hicieran múltiples combinaciones de dichos algoritmos con las diferentes variables existentes en el conjunto de datos:

- Calificación Kardex
- Calificación del curso propedéutico
- Misma empresa
- Grupo edad
- Grupo licenciatura
- Calificación de la materia *Matemáticas Avanzadas para computación*
- Calificación de la materia *Análisis y diseño de algoritmos*
- Calificación de la materia *Innovación, Diseño e Investigación I*

Para realizar dicho análisis, se utilizó el lenguaje de programación Python3 con las librerías *Pandas, Matplotlib, Numpy, y Scikit Learn*, principalmente.

En la figura 18 podemos observar el flujo que se siguió en el algoritmo desarrollado, de tal forma que al final se generó un reporte detallando los resultados de cada una de las ejecuciones generadas.

3.5.3 Tercer Enfoque

Este enfoque consistió en la creación de un árbol de decisión binaria, en dicho árbol, el primer nodo representa la variable con mayor importancia, es decir, la variable que más impacto tiene en el resultado final para el conjunto de datos. De tal forma que, el último nodo representa la variable menos importante o la que menos impacto tiene para el resultado final.

Para generar este enfoque se utilizó el lenguaje de programación Python3 y las librerías *Numpy y Matplotlib.*

Se trabajó con las siguientes variables para generar el modelo:

- Calificación Kardex
- Calificación del curso propedéutico
- Misma empresa
- Grupo edad
- Grupo licenciatura
- Calificación de la materia *Matemáticas Avanzadas para computación*
- Calificación de la materia *Análisis y diseño de algoritmos*
- Calificación de la materia *Innovación, Diseño e Investigación I*

Figura 18 Ejecución Segundo Modelo

En la figura 19, podemos observar el algoritmo que se siguió para armar el árbol de decisión binaria, se va armando uno a uno los nodos de acuerdo con la entropía de cada una de las variables, la variable que se selecciona va de acuerdo con la entropía y la certeza generada, lo cual nos dice si una variable por sí sola es capaz determinar el resultado final (egreso o baja).

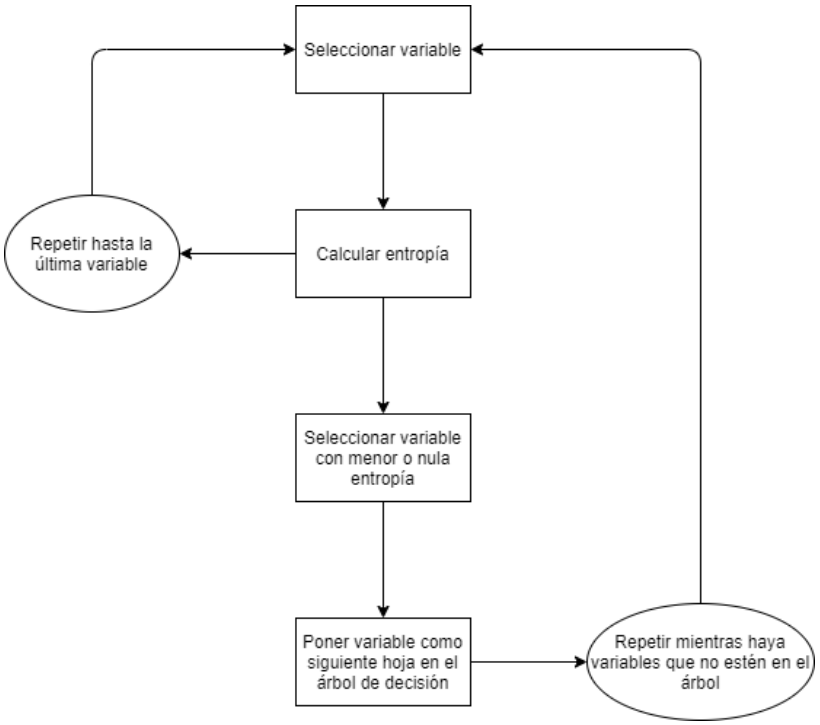

Figura 19 Ejecución Tercer Modelo

3.5.4 Cuarto Enfoque

Este enfoque consistió en generar clasificaciones distintas a las originales (Baja, Egresado) con base en el conjunto de datos original, y generar agrupaciones nuevas, en las cuales se determinen "grupos de riesgo".

Las variables utilizadas para realizar este análisis fueron:

1. Calificación Kardex.
2. Calificación Propedéutico.
3. Misma Empresa.
4. Grupo de edad.
5. Licenciatura.
6. Calificación *Matemáticas Avanzadas para Computación.*
7. Calificación *Análisis y Diseño de Algoritmos.*
8. Calificación *Innovación, Diseño e Investigación I.*

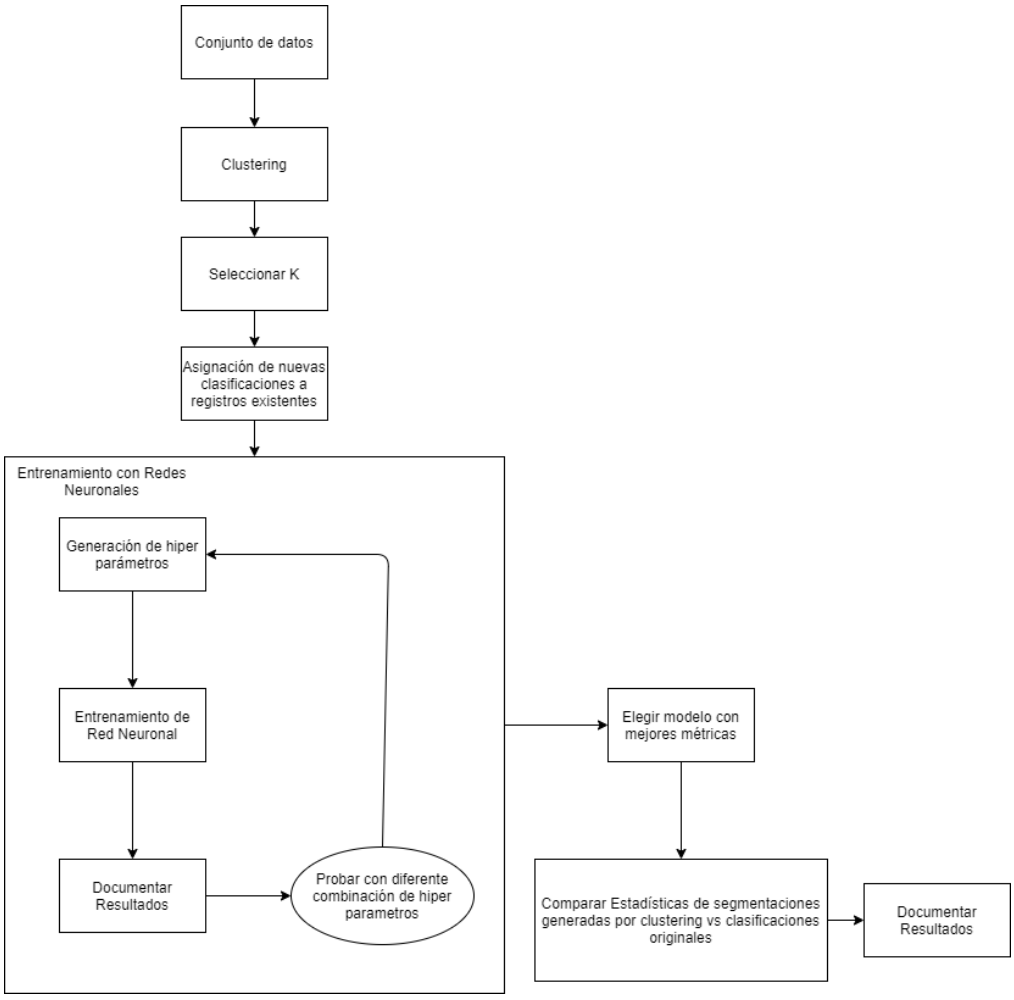

Figura 20 Ejecución Cuarto Modelo

Para generar este modelo se utilizó el lenguaje de programación Python y las librerías *matplotlib, numpy, pandas y scikit-learn.*

En la figura 20 observamos los pasos que se siguieron para generar este modelo. Primero se utilizó el algoritmo de segmentación *k-means,* el cual nos ayudó a generar clasificaciones diferentes a las originales, de tal forma que pudiéramos generar diferentes grupos a los originales e intentar obtener resultados diferentes.

De acuerdo con la gráfica de "codo" que se observa en la figura 21, se seleccionaron tres tipos de clasificación para agrupar a los alumnos.

Posteriormente y siguiendo el flujo de la figura 20, una vez obtenidas las agrupaciones, se procedió a generar un modelo de clasificación basado en redes neuronales. Se realizaron múltiples iteraciones, en las cuales se utilizaron diferentes hiper parámetros para comparar diferentes modelos. Se hicieron las siguientes combinaciones:

1. Arquitectura de la red neuronal, con las siguientes variables (el primer valor representa la cantidad de capas ocultas, el segundo valor la cantidad de neuronas por capa oculta):
    a. (5, 2), (5, 3), (5, 4), (5,5), (5,6)
    b. (6, 2), (6, 3), (6, 4), (6,5), (6,6)
    c. (7, 2), (7, 3), (7, 4), (7,5), (7,6)
    d. (8, 2), (8, 3), (8, 4), (8,5), (8,6)
    e. (9, 2), (9, 3), (9, 4), (9,5), (9,6)
2. Alpha, es decir, la taza de aprendizaje tuvo los siguientes valores:
    a. 0.001
    b. 0.003
    c. 0.01
    d. 0.03
    e. 0.1
    f. 0.3
3. Se utilizó RELU para la activación.
4. Se utilizó Adam como algoritmo de optimización.

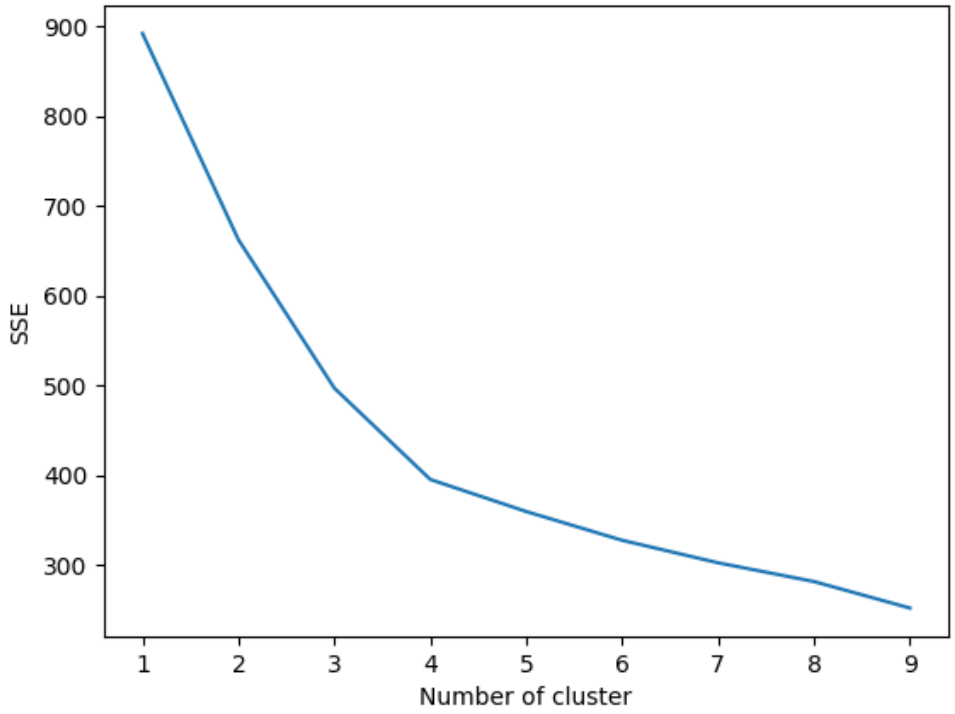

Figura 21 Gráfica de codo del cuarto modelo

3.5.5 Quinto Enfoque

Como quinto enfoque, se propuso trabajar con las clasificaciones originales y con el algoritmo de aprendizaje supervisado de Redes Neuronales. Para sacar el potencial de este modelo, se decidió realizar

múltiples iteraciones con diferentes hiper parámetros para comparar los resultados obtenidos con datos de pruebas.

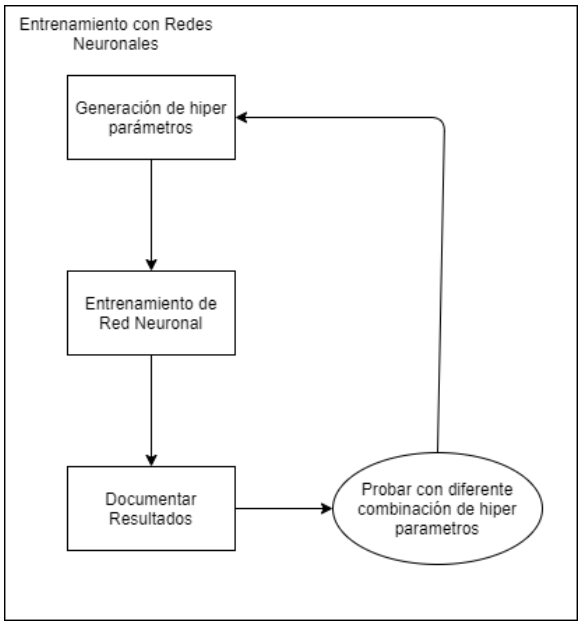

Figura 22 Ejecución Quinto Modelo

En la figura 22, podemos observar la ejecución del quinto modelo. Este modelo se centró únicamente en utilizar el algoritmo de redes neuronales con diferentes hiper parámetros, documentando todos los resultados para encontrar aquel que tuviera las mejores métricas.

Los hiper parámetros con los que se realizaron las iteraciones fueron las siguientes:

1. La arquitectura de la red neuronal, es decir, la cantidad de capas ocultas y la cantidad de neuronas contenidas en cada capa, tuvo las siguientes variantes (el primer valor representa las capas ocultas, el segundo la cantidad de neuronas): [(5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6), (7, 2), (7, 3), (7, 4), (7, 5), (7, 6),(8, 2), (8, 3), (8, 4), (8, 5), (8, 6), (9, 2), (9, 3), (9, 4), (9, 5), (9, 6)].
2. Alpha, es decir, la taza de aprendizaje, tuvo los siguientes valores: [0.001, 0.003, 0.01, 0.03, 0.1, 0.3].
3. Activación, es decir qué función de activación se utilizó, tuvo los siguientes valores: [Tangente hiperbólica, *Relu*].
4. Tipo de taza de aprendizaje, tuvo los siguientes valores: [adaptativa, constante].

La combinación de las variables dio un resultado de 600 iteraciones en total, las variables con las que se ejecutó este modelo, fueron las siguientes:

1. Calificación Kardex.
2. Calificación propedéutica.
3. Misma empresa.
4. Grupo edad.
5. Grupo licenciatura.
6. Calificaciones matemáticas.

7. Calificación algoritmos.
8. Calificación IDI I.
9. Beca asignada.

3.6 Métricas y Evaluación de modelo

Para el presente trabajo, se dividió el conjunto de datos de la siguiente forma:

1. 80% de los datos escogidos de manera aleatoria para entrenar los modelos seleccionados
2. 20% de los datos escogidos de manera aleatoria para validar el modelo seleccionado y determinar si la generalización que se hacía funcionaba de manera correcta.

Estos datos se utilizaron para el entrenamiento de los 5 enfoques diferentes de aprendizaje presentados previamente.

Para medir qué tan buenos son dichos modelos se utilizaron las métricas de precisión y *recall*. La precisión determina que proporción de las identificaciones positivas fue correcta. *Recall* nos ayuda a determinar qué proporción de los positivos actuales fueron identificados de manera correcta. Se definen de la siguiente manera:

$$Precisión = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

TP es el valor de los elementos clasificados positivamente y que fueron clasificados correctamente, FP significa aquellos elementos clasificados positivamente, pero clasificados de manera incorrecta y FN representa a los clasificados de manera negativa, pero de manera incorrecta.

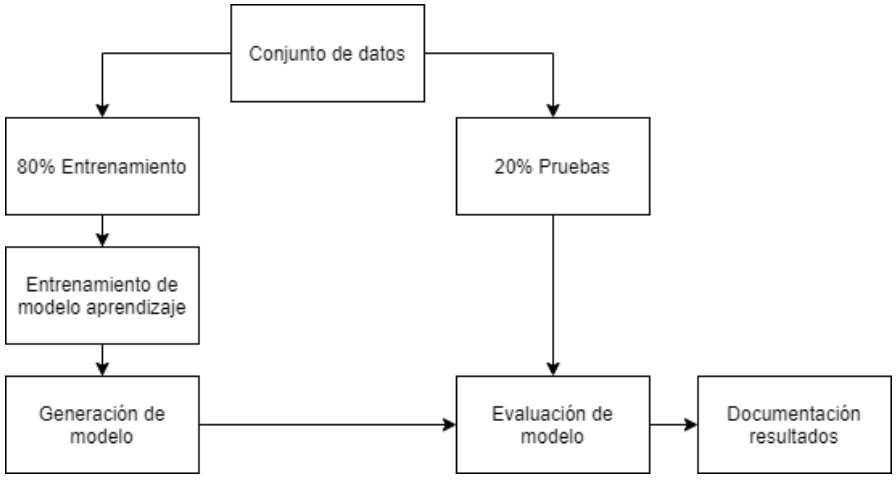

Figura 23 Evaluación de modelos de predicción

En la figura 23, podemos observar los pasos que se siguieron para realizar la evaluación de los modelos generados, se documentaron los resultados obtenidos con base en las métricas de precisión y exhaustividad.

4 RESULTADOS

Resumen: *En este capítulo se presentan los resultados obtenidos del desarrollo de este trabajo. Se detallan los resultados que generó cada modelo descrito en el capítulo anterior, así como una comparación entre los resultados de los modelos.*

4.1 Primer Enfoque

Los resultados que se obtuvieron a partir del primer enfoque generado los podemos observar en la tabla 3. Tanto la precisión, como la exhaustividad son muy bajas. Esto debido a que no se utilizaron todas las variables existentes y porque el modelo que se utilizó fue muy sencillo.

Las variables con las que se trabajó fueron seleccionadas porque ambas son variables que se tiene antes del inicio formal del programa. Si estas dos variables nos dieran un buen resultado, tendríamos un modelo confiable de predicción antes de la inscripción formal al posgrado en estudio.

Clase	Precisión	Exhaustividad (*Recall)*
Baja	0.34	0
Egresado	0.56	0.93

Tabla 3 Métricas obtenidas en una corrida aleatoria

4.2 Segundo Enfoque

En total se crearon 511 combinaciones posibles de las variables seleccionadas para este enfoque y los algoritmos. Se realizó un entrenamiento para cada una de estas combinaciones. Se generó un reporte contra los datos de entrenamiento y otro reporte más contra los datos de pruebas.

Como ejemplo, para una corrida aleatoria, se obtuvieron los siguientes resultados. Cabe recordar que los valores de precisión y exhaustividad están en un rango entre 0 y 1, donde 0 es un muy mal resultado y 1 es un resultado ideal.

Clase	Precisión	Exhaustividad (*Recall)*
Baja	0	0
Egresado	0.86	0.93

Tabla 4 Métricas obtenidas en una corrida aleatoria

Clase	Precisión	Exhaustividad (*Recall)*
Baja	0	0
Egresado	0.50	1.0

Tabla 5 Métricas obtenidas sin datos de pruebas

En la tabla 4 se puede observar un ejemplo de los resultados obtenidos por el algoritmo Regresión logística haciendo pruebas con los mismos datos de entrenamiento. Mientras que en la tabla 5 se observan los resultados obtenidos con los datos de pruebas. Se puede observar que existe un claro sesgo en cuanto a la clasificación de “egresados”.

Clase	Precisión	Exhaustividad (*Recall)*	Tipo
Baja	0.9	1	Entrenamiento
Egresado	1	0.98	Entrenamiento
Baja	0.6	0.375	Pruebas
Egresado	0.54	0.75	Pruebas

Tabla 6 Métricas obtenidas con datos de pruebas

El mejor resultado obtenido fue por parte del algoritmo de árboles de decisión binaria, utilizando las variables:

- Calificación Kardex
- Calificación del curso propedéutico
- Misma empresa
- Grupo edad
- Grupo licenciatura
- Calificación de la materia *Matemáticas Avanzadas para computación*
- Calificación de la materia *Análisis y diseño de algoritmos*
- Calificación de la materia *Innovación, Diseño e Investigación I*

En la tabla 5 podemos observar las métricas obtenidas por el algoritmo, tanto en pruebas como en el mismo entrenamiento. Si hacemos una comparación con los resultados obtenidos en las tablas 5 y 6 por el algoritmo de regresión logística, podemos ver grandes diferencias. La desventaja de este modelo es que falla en hacer una generalización, dado que es muy baja tanto su precisión como su exhaustividad en las métricas de pruebas.

4.3 Tercer Enfoque

El tercer acercamiento fue realizando un árbol de decisión binaria para determinar la correlación entre las variables y las clasificaciones. De acuerdo con las fórmulas de entropía y ganancia revisadas se generó un árbol de decisión binaria de manera gráfica, en el cual se va desglosando una a una las variables, demostrando cuáles tienen mayor o menor impacto sobre el conjunto de datos trabajado.

En la figura 24 se puede observar el árbol generado, de arriba hacia abajo se desglosan las variables que implica de mayor a menor importancia, siendo “Calificación algoritmos” la más importante y “Calificación propedéutico” la menos importante. Sin embargo, lo que este árbol nos demuestra también es que no hay certeza con el conjunto de datos trabajado para generar un modelo adecuado que nos ayude a realizar predicciones más precisas.

Este modelo nos ayuda a determinar que las variables que tienen mayor importancia o “peso” para el resultado final son:

- Calificación Análisis y Diseño de Algoritmos
- Misma Empresa
- Licenciatura

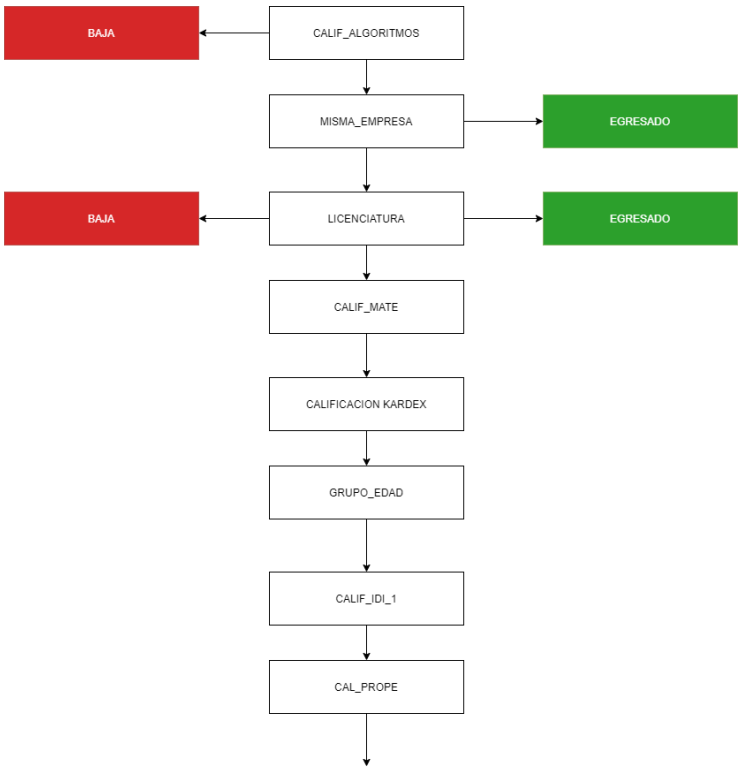

Figura 24 Árbol de decisión

4.4 Cuarto Enfoque

De esta aproximación, se tuvo un resultado interesante. En la tabla 7 se tiene un reporte de clasificación con datos de prueba, con los cuales se puede observar un modelo prácticamente perfecto.

Para generarlo se utilizaron como hiper parámetros, una arquitectura de 9 capas ocultas y 5 neuronas en cada capa, con una taza de aprendizaje de 0.3. Se realizaron múltiples ejecuciones, con diferentes arquitecturas, en las cuales realizaron variaciones en el número de capas ocultas y la cantidad de neuronas en ellas. Esta arquitectura fue seleccionada porque fue con la que se obtuvieron mejores resultados.

Clasificación	Precisión	*Recall*
0	1.0	1.0
1	1.0	1.0
2	1.0	1.0

Tabla 7 Métricas obtenidas en el cuarto modelo

De acuerdo con los datos obtenidos por el modelo de *clustering*, se realizó un análisis de las segmentaciones generadas. En la tabla 8 tenemos las estadísticas de la primera clasificación generada por el modelo, en la cual, se tiene una cantidad de **14** estudiantes.

Clase 0	Calificación Kardex	Calificación Propedéutico	Calificación Matemáticas	Calificación Algoritmos	Calificación IDI I
Promedio	9.0	8.21	9.21	8.9	9.64
Desviación Estándar	0.53	1.25	0.8	1.5	0.49
Mínimo	8.1	6	8	5	9
Máximo	9.66	10	10	10	10

Tabla 8 Estadísticas del primer grupo segmentado

En la tabla 9 se puede observar el siguiente grupo de la segmentación, en la cual se obtuvieron un total de **64** estudiantes, lo que quiere decir, que es el grupo más grande.

Clase 1	Calificación Kardex	Calificación Propedéutico	Calificación Matemáticas	Calificación Algoritmos	Calificación IDI I
Promedio	8.96	8.23	9.1	8.9	9.4
Desviación Estándar	0.41	1.2	0.85	1.23	0.81
Mínimo	7.67	6	7	6	7
Máximo	9.8	10	10	10	10

Tabla 9 Estadísticas del segundo grupo segmentado

En la tabla 10 tenemos al último grupo de la segmentación, el cual cuenta con únicamente **3** estudiantes, es decir, el grupo más pequeño.

Clase 2	Calificación Kardex	Calificación Propedéutico	Calificación Matemáticas	Calificación Algoritmos	Calificación IDI I
Promedio	8.8	6.6	5.6	3.3	5
Desviación Estándar	0.22	1.15	1.15	2.88	0
Mínimo	8.56	6	5	0	5
Máximo	9	8	7	5	5

Tabla 10 Estadísticas del tercer grupo segmentado

En la tabla 11 se puede observar una comparativa entre las segmentaciones obtenidas con el modelo de *clustering* y las clasificaciones reales, es decir, se muestra en cada una de las segmentaciones, la cantidad de bajas y de egresados que hubo en realidad.

Clase	Bajas	Egresados
0	2	12
1	12	52
2	3	0

Tabla 11 Comparativa de clasificaciones reales vs segmentaciones

De acuerdo con las segmentaciones y a las estadísticas previamente observadas, se puede concluir lo siguiente:

1. El grupo dos (cuyas estadísticas se observan en la tabla 9) representa un claro riesgo, debido a los siguientes datos:
    a. El promedio de las calificaciones de matemáticas, algoritmos e IDI no es aprobatorio, y los máximos en dos de tres de estas calificaciones tampoco lo es.

b. El promedio de calificación de propedéutico es el mínimo aprobatorio para ingresar al programa.

2. Sin embargo, a pesar de lo anterior, al ser un grupo de únicamente tres personas es difícil considerar que el grupo tres representa una generalización adecuada.
3. Los grupos cero y uno representan una mejoría sustancial en cuestión de calificaciones en comparación con el grupo dos, sin embargo, como podemos observar en la tabla 11, en ambos grupos existe un porcentaje de estudiantes que se dieron de baja del programa y de acuerdo con los porcentajes, que es 16% de bajas en la clase uno y 23% para la clase dos.

4.5 Quinto Enfoque

Los mejores resultados se obtuvieron a partir de una combinación de una arquitectura de seis capas ocultas y tres neuronas con cualquier taza de aprendizaje, cualquier tipo de taza de aprendizaje y con función de activación *Relu*.

En la tabla 12 se pueden observar los resultados obtenidos, no otorga métricas precisas para determinar una clasificación de bajas, pero en contraste, da excelentes resultados en el caso de predecir egresados.

Clase	Precisión	Recall
Baja	1.0	0.6
Egresado	0.85	1.0

Tabla 12 Métricas obtenidas en el quinto modelo

5. CONCLUSIONES

Resumen: *En este capítulo se presentan las conclusiones y trabajo futuro. Se explica qué se puede inferir a partir de los resultados obtenidos, y qué es lo que se puede realizar como trabajo futuro a partir de lo encontrado.*

5.1 Conclusiones

De acuerdo con los enfoques generados, se puede determinar que los mejores, de acuerdo con las métricas, fueron los últimos dos enfoques generados, con las clasificaciones originales (quinto enfoque) y con las obtenidas a partir de la segmentación realizada en el algoritmo de *clustering* (cuarto enfoque).

Las diferencias observadas con los tres primeros modelos son las siguientes:

1. Si bien en el primer enfoque se utilizó el mismo algoritmo de aprendizaje (Redes Neuronales) se utilizaron únicamente dos variables y no se compararon diferentes arquitecturas, es decir, se utilizó la arquitectura *default*, de la librería con la que se trabajó y no se realizaron más comparaciones.
2. El segundo enfoque se centró en hacer fuerza bruta para comparar los resultados obtenidos con diferentes algoritmos (Regresión Logística, Árboles de Decisión Binaria, Bosques Aleatorios, y Redes Neuronales), Sin embargo, estos modelos fallaron en tener buenos resultados y a pesar de utilizar Redes Neuronales, se cometió el mismo error que en el primer modelo, es decir, no alterar la arquitectura utilizada.
3. El tercer enfoque se basó en generar un árbol de decisión en el cual se muestre cuáles eran las variables más importantes para realizar predicciones, sin embargo, como este modelo se basa en encontrar certezas al 100% para realizar sus nodos de decisión se falló en encontrar un modelo perfecto.

El cuarto y quinto enfoque son similares porque ambos se centraron en las siguientes ideas:

1. Se utiliza el algoritmo de Redes Neuronales como algoritmo de aprendizaje supervisado.
2. Se realizan múltiples ejecuciones de entrenamiento con diferentes arquitecturas.

Su principal diferencia se encuentra en que el cuarto enfoque primero ejecuta un algoritmo no supervisado para realizar segmentaciones, y el quinto enfoque ya tiene determinadas las clasificaciones del conjunto de datos originales. Es decir, en el cuarto enfoque se aplica primero el algoritmo de k-means para generar las clasificaciones, mientras que en el quinto enfoque las clasificaciones son las originales.

El cuarto enfoque al parecer es perfecto, dado que sus métricas muestran los mejores resultados. Sin embargo, el 79% de los estudiantes están concentrados en una clase. Si observamos la figura 11 y las estadísticas generadas en las tablas 8, 9 y 10, podemos observar que los grupos generados por el algoritmo de segmentación son parecidos a las clasificaciones reales. Lo que significa, que las clasificaciones obtenidas por el algoritmo k-means, son muy parecidas a las clasificaciones originales.

En el quinto enfoque se encontró una arquitectura que da métricas aceptables utilizando las clasificaciones originales. Sin embargo, falla un poco al generalizar las bajas, lo cual se puede observar en la tabla 12, donde el *recall* obtenido es de 0.6.

Un dato importante para resaltar en el conjunto de datos trabajado es que se encuentra desbalanceado, es decir, se tienen cuatro veces más datos de egresados, que, de bajas, es por esto que los primeros tres enfoques fallan al generalizar las bajas.

También es importante observar que el conjunto de datos es pequeño, con únicamente 81 registros, por lo cual, es difícil realizar generalizaciones. De todos los algoritmos utilizados, el que mejores resultados otorgó, incluso con un conjunto de datos pequeño, fue el de redes neuronales con variaciones en las arquitecturas utilizadas.

En conclusión, el quinto enfoque fue el que mejores resultados otorgó para predecir si un alumno completará el programa de posgrado, como observamos en la tabla 12, tiene una precisión de 0.85 y un recall de 1.0, lo cual nos ayuda para determinar si futuros estudiantes del posgrado serán posibles egresados.

5.2 Trabajo Futuro

Como se mencionó previamente, es importante resaltar que los datos están desbalanceados. Para trabajar esta situación, hay que buscar algunas técnicas para balancear datos.

Un punto importante a considerar es el contar con más registros que con los que se trabajó originalmente, dado que son muy pocos, y se puede llegar a fallar en generalizar.

También, hay que considerar la recopilación de datos adicionales para cada estudiante, como:

- Género
- Fortaleza financiera
- Datos demográficos

Un siguiente estudio puede partir del cuarto o quinto enfoque, trabajar con las mismas métricas o agregar alguna que no exista.

Considerar que el algoritmo de aprendizaje que mejor soluciona nuestro problema inicial fue el de Redes Neuronales con sus debidas experimentaciones.

Es posible también aplicar una generalización de lo que se realizó en este trabajo para otros posgrados dentro de la universidad. Sin embargo, es importante revisar las particularidades de cada uno de los posgrados y transformarlas en variables categóricas o numéricas. Una de las particularidades del programa que se analiza en este trabajo es que cuenta con un examen propedéutico, una característica que no es común a todos los posgrados de esta universidad.

BIBLIOGRAFÍA

[1] M. Fresán, “Factores que propician el abandono y obstaculizan la culminación de los estudios de posgrado”, [En línea], [13 Nov 2013] Disponible en http://revistas.utp.ac.pa/index.php/clabes/article/view/877/904

[2] F. Miranda, “Abandono escolar en educación media superior: conocimiento y aportaciones de política pública”, [En línea], [28 Feb 2018] Disponible en https://sinectica.iteso.mx/index.php/SINECTICA/article/view/863

[3] J. Xu, K. H. Moon and M. van der Schaar, "A Machine Learning Approach for Tracking a Predicting Student Performance in Degree Programs," en IEEE Journal of Selected Topics Signal Processing, vol. 11, no. 5, pág. 742-753, Aug. 2017. doi: 10.1109/JSTSP.2017.2692560

[4] G. W. Cox, W. E. Hughes Jr., L. H. Etzkorn and M. E. Weisskopf, "Predicting Computer Science Ph.D. Completion: A Case Study," en IEEE Transactions on Education, vol. 52, no. 1, pág. 137-143, Feb. 2009. doi: 10.1109/TE.2008.921458.

[5] U. B. Mat, N. Buniyamin, P. M. Arsad and R. Kassim, "An Overview of Using Academic Analytics to Predict and Improve Students’ Achievement: A Proposed Proactive Intelligent Intervention" en IEEE 5th Conference on Engineering Education (ICEED)

[6] K. D. Kolo, S. A. Adepoju and J. K. Alhassan, "A Decision Tree Approach for Predicting Students Academic Performance" en I.J. Education and Management Engineering, 2015, 5, pág. 12-19

[7] S. Borkar and K. Rajeswari, "Attributes Selection for Predicting Students’ Academic Performance using Education Data Mining and Artificial Neural Network" en International Journal of Computer Applications, Volumen 86 – No 10, Enero 2014

[8] C. Márquez-Vera, C. Romero and S. Ventura, "Predicting School Failure Using Data Mining"

[9] C. Molnar, Interpretable Machine Learning, “A Guide for Making Black Box Models Explainable”, [En línea], [17 Dic 2019] Disponible en https://christophm.github.io/interpretable-ml-book/

[10] G. James, et. All, “An Introduction to Statistical Learning”, Springer, 2013.

[11] C.E. Shannon, “A Mathematical Theory of Communication”, 1949, pág. 10-14.

[12] B. Kröse, P. van der Smagt, An Introduction to neural networks. Amsterdam: The University of Amsterdam, 1996, pág. 15-46.

[13] C. Albon, *Machine Learning with Python Cookbook*, O’Reilly Media Inc, [En línea] [2018], Disponible en https://learning.oreilly.com/library/view/machine-learning-with/9781491989371/