



# **UNIVERSIDAD DEL PACÍFICO**

**Ingeniería en Sistemas**

**Título del Trabajo de Titulación:**

## **DISEÑO DE UN MODELO DE PREDICCIÓN DE APTITUD DE ASPIRANTES A OFICIALES Y TRIPULANTES DE ARMAS DE LA ARMADA DEL ECUADOR**

**Autor:**

**Nidia Paola Mendoza Donoso**

**Director de Trabajo de Titulación:**

**MSc. Jorge Alberto Medina Avelino**

Guayaquil, 2025

## DECLARACION DE AUTORIA

Yo, NIDIA PAOLA MENDOZA DONOSO, declaro bajo juramento que el trabajo aquí descrito es de mí autoría; que no ha sido previamente presentado para ningún grado, calificación profesional, o proyecto público ni privado; y que he consultado las referencias bibliográficas que se incluyen en este documento.

En caso de que la Universidad auspicie el estudio, se incluirá el siguiente párrafo:

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la UNIVERSIDAD DEL PACIFICO, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



---

Nidia Paola Mendoza Donoso

CI. 0929704161

## RESUMEN

Este artículo contiene el análisis de un conjunto de datos y la construcción de tres modelos de aprendizaje automático para predecir la aptitud de los aspirantes a Oficiales y Tripulantes de armas de la Armada del Ecuador. Los modelos utilizados son árbol de decisión, XGBoost y red neuronal. Primero se realiza una exploración de los datos con estadística descriptiva y visualizaciones. Luego, se construye un árbol de decisión. Se codifican las variables categóricas y se entrena el modelo con el conjunto de entrenamiento. Se visualiza la estructura del árbol y se evalúa el modelo con el conjunto de prueba, obteniendo una precisión y una matriz de confusión.

Después, se implementa un modelo XGBoost. Se codifican las variables categóricas, se crea una matriz DMatrix y se definen los parámetros del modelo. Se entrena el modelo y se realizan predicciones, evaluando la precisión y generando una matriz de confusión. Finalmente, se crea una red neuronal. Se preprocesan los datos, se define la arquitectura del modelo con capas densas y dropout, se compila el modelo con el optimizador Adam y se entrena. Se evalúa la precisión del modelo y se genera una matriz de confusión.

En resumen, se explora tres modelos de aprendizaje automático para predecir la aptitud de los aspirantes a Oficiales y Tripulantes de armas de la Armada del Ecuador, incluyendo un árbol de decisión, XGBoost y una red neuronal. Se realiza un análisis exploratorio de datos, preprocesamiento, entrenamiento, evaluación y visualización de resultados para cada modelo.

## PALABRAS CLAVE

- Árboles de decisión
- Redes neuronales
- XGBost
- Oficiales y Tripulantes de la Armada

## ABSTRACT

This article contains the analysis of a data set and the construction of three machine learning models to predict the aptitude of candidates for Officers and Crewmen of the Ecuadorian Navy. The models used are decision tree, XGBoost and neural network. First, an exploration of the data is carried out with descriptive statistics and visualizations. Then, a decision tree is built. The categorical variables are encoded and the model is trained with the training set. The structure of the tree is visualized and the model is evaluated with the test set, obtaining an accuracy and a confusion matrix.

Next, an XGBoost model is implemented. The categorical variables are encoded, a DMatrix is created and the model parameters are defined. The model is trained and predictions are made, evaluating the accuracy and generating a confusion matrix. Finally, a neural network is created. The data is preprocessed, the model architecture is defined with dense and dropout layers, the model is compiled with the Adam optimizer and trained. The accuracy of the model is evaluated and a confusion matrix is generated.

In summary, three machine learning models are explored to predict the aptitude of candidates for Officers and Crewmen of the Ecuadorian Navy, including a decision tree, XGBoost and a neural network. An exploratory data analysis, preprocessing, training, evaluation and visualization of results are performed for each model.

## KEYWORDS

- Decision tree
- Neural networks
- XGBoost
- Navy Officers and Crew

## Diseño de un modelo de predicción de aptitud de aspirantes a Oficiales y Tripulantes de armas de la Armada del Ecuador

Nidia Mendoza Donoso<sup>a</sup>

a. Facultad de Innovación y Desarrollo Tecnológico, Universidad del Pacífico. e-mail: [nidia.mendoza@upacifico.edu.ec](mailto:nidia.mendoza@upacifico.edu.ec)

### Palabras Clave

Arboles de decisión  
Redes neuronales  
XGBost  
Oficiales y Tripulantes de la Armada

### Historia del Artículo

Recibido dd-mm-aaaa  
Revisado dd-mm-aaaa  
Aceptado dd-mm-aaaa  
Publicado dd-mm-aaaa

**Resumen.** Este artículo contiene el análisis de un conjunto de datos y la construcción de tres modelos de aprendizaje automático para predecir la aptitud de los aspirantes a Oficiales y Tripulantes de armas de la Armada del Ecuador. Los modelos utilizados son árbol de decisión, XGBoost y red neuronal. Primero se realiza una exploración de los datos con estadística descriptiva y visualizaciones. Luego, se construye un árbol de decisión. Se codifican las variables categóricas y se entrena el modelo con el conjunto de entrenamiento. Se visualiza la estructura del árbol y se evalúa el modelo con el conjunto de prueba, obteniendo una precisión y una matriz de confusión.

Después, se implementa un modelo XGBoost. Se codifican las variables categóricas, se crea una matriz DMatrix y se definen los parámetros del modelo. Se entrena el modelo y se realizan predicciones, evaluando la precisión y generando una matriz de confusión. Finalmente, se crea una red neuronal. Se preprocesan los datos, se define la arquitectura del modelo con capas densas y dropout, se compila el modelo con el optimizador Adam y se entrena. Se evalúa la precisión del modelo y se genera una matriz de confusión.

En resumen, se explora tres modelos de aprendizaje automático para predecir la aptitud de los aspirantes a Oficiales y Tripulantes de armas de la Armada del Ecuador, incluyendo un árbol de decisión, XGBoost y una red neuronal. Se realiza un análisis exploratorio de datos, preprocesamiento, entrenamiento, evaluación y visualización de resultados para cada modelo.

**Keywords**

Decision trees  
Neural networks  
XGBoost  
Navy Officers and Crew

**Article History**

Received dd-mm-aaaa  
Revised dd-mm-aaaa  
Accepted dd-mm-aaaa  
Published dd-mm-aaaa

**Abstract.** This article contains the analysis of a data set and the construction of three machine learning models to predict the aptitude of candidates for Officers and Crewmen of the Ecuadorian Navy. The models used are decision tree, XGBoost and neural network. First, an exploration of the data is carried out with descriptive statistics and visualizations. Then, a decision tree is built. The categorical variables are encoded and the model is trained with the training set. The structure of the tree is visualized and the model is evaluated with the test set, obtaining an accuracy and a confusion matrix.

Next, an XGBoost model is implemented. The categorical variables are encoded, a DMatrix is created and the model parameters are defined. The model is trained and predictions are made, evaluating the accuracy and generating a confusion matrix. Finally, a neural network is created. The data is preprocessed, the model architecture is defined with dense and dropout layers, the model is compiled with the Adam optimizer and trained. The accuracy of the model is evaluated and a confusion matrix is generated.

In summary, three machine learning models are explored to predict the aptitude of candidates for Officers and Crewmen of the Ecuadorian Navy, including a decision tree, XGBoost and a neural network. An exploratory data analysis, preprocessing, training, evaluation and visualization of results are performed for each model.

## 1. Introducción

La Armada del Ecuador, como institución fundamental para la defensa y seguridad del país, enfrenta el constante desafío de seleccionar a los mejores candidatos para ocupar cargos como Oficiales y Tripulantes de Armas. Este proceso de selección implica una evaluación rigurosa que busca identificar a aquellos individuos que poseen las aptitudes físicas, intelectuales, psicológicas y de liderazgo necesarias para desempeñarse con éxito en las exigentes tareas que conlleva la vida militar.

En este contexto, el desarrollo de herramientas que permitan optimizar el proceso de selección de aspirantes resulta de gran relevancia. El presente estudio propone el diseño de un modelo de predicción de aptitud para aspirantes a Oficiales y Tripulantes de Armas de la Armada del Ecuador, utilizando técnicas de aprendizaje automático y la herramienta Google Colab que permite escribir y ejecutar código Python.

El modelo se basará en una base de datos histórica proporcionada por la oficina de reclutamiento de la Armada, la cual contiene información de aproximadamente 383 aspirantes. Esta información incluye variables sociodemográficas, resultados de evaluaciones médicas, pruebas físicas y académicas, entre otras. A partir de este conjunto de datos, se entrenará un algoritmo de aprendizaje automático capaz de identificar patrones y relaciones entre las variables que permitan predecir la aptitud de futuros aspirantes.

La implementación de este modelo de predicción presenta diversos beneficios para la Armada del Ecuador. En primer lugar, permitirá optimizar el proceso de selección, reduciendo el tiempo y los recursos necesarios para evaluar a los aspirantes. En segundo lugar, contribuirá a identificar con mayor precisión a aquellos candidatos con mayor potencial para el éxito en la carrera naval, mejorando así la calidad del personal que ingresa a la institución. Finalmente, proporcionará información valiosa para comprender los factores que inciden en la aptitud para el servicio militar, permitiendo así implementar estrategias de reclutamiento y formación más efectivas.

En este trabajo de investigación, se abordarán las siguientes etapas:

- **Análisis exploratorio de datos:** Se realizará una descripción detallada de las variables incluidas en la base de datos, identificando posibles patrones, relaciones y valores atípicos.
- **Preprocesamiento de datos:** Se aplicarán técnicas de limpieza y transformación de datos para garantizar su calidad y consistencia para el modelado.
- **Selección de variables:** Se evaluará la relevancia y el poder predictivo de las variables disponibles, seleccionando aquellas que aporten mayor valor al modelo.
- **Entrenamiento y evaluación del modelo:** Se entrenarán diferentes algoritmos de aprendizaje automático y se evaluará su rendimiento mediante métricas de desempeño adecuadas.
- **Interpretación del modelo:** Se analizarán los resultados obtenidos por el modelo

para comprender las relaciones entre las variables y su impacto en la predicción de la aptitud.

- Validación del modelo: Se evaluará el desempeño del modelo en un conjunto de datos independiente para verificar su generalización a nuevos casos.
- Conclusiones y recomendaciones: Se presentarán las conclusiones del estudio, destacando los aportes del modelo de predicción y brindando recomendaciones para su implementación en la Armada del Ecuador.

Se espera que este estudio contribuya de manera significativa al desarrollo de un proceso de selección más eficiente y efectivo para la Armada del Ecuador, permitiendo identificar a los mejores candidatos para las carreras de Oficiales y Tripulantes de Armas, y fortaleciendo así la institución como garante de la seguridad y defensa del país.

## 2. Marco teórico

### 2.1. Entrenamiento y Formación en la Armada del Ecuador

El entrenamiento y la formación de los integrantes de la Armada del Ecuador son fundamentales para garantizar su capacidad operativa y de respuesta ante diversas situaciones de seguridad y defensa. Las escuelas de formación de Oficiales y Tripulantes son las principales instituciones encargadas de la formación de oficiales y tripulantes navales, donde se imparten conocimientos teóricos y prácticos sobre tácticas, estrategia militar, navegación, y liderazgo. Estas escuelas tienen como objetivo preparar a los futuros integrantes de la Armada, capacitándolos para tomar decisiones críticas en entornos complejos y desafiantes (Cruz, 2020). Los egresados de estas instituciones no solo poseen habilidades técnicas, sino que también desarrollan una sólida ética profesional y un compromiso con la defensa del país.

Además de la formación académica, la Armada del Ecuador implementa un sistema de capacitación continua que permite a los marinos actualizar sus conocimientos y habilidades a lo largo de sus carreras. Este sistema incluye cursos de especialización en áreas como logística, operaciones de combate, y manejo de tecnología avanzada. La capacitación se complementa con ejercicios prácticos y simulaciones que reflejan escenarios reales, asegurando que los miembros de la Armada estén siempre listos para enfrentar desafíos actuales y futuros (Moreno, 2021). Este enfoque garantiza que la Armada mantenga un alto nivel de competencia, permitiendo la adaptación a nuevas amenazas y la evolución en el ámbito de la defensa.

La formación en la Infantería de Marina es otro aspecto crítico del entrenamiento militar. Los infantes de marina reciben una formación intensiva que incluye operaciones anfibia, combate cuerpo a cuerpo, y manejo de armamento avanzado. El entrenamiento es riguroso y diseñado para desarrollar no solo habilidades técnicas, sino también resistencia física y mental. Esta preparación es esencial, dado que la Infantería de Marina opera en condiciones exigentes y a menudo en entornos hostiles (L. García, 2019). Los ejercicios conjuntos con fuerzas de otros países también son parte del entrenamiento, lo que permite a los

infantes de marina ecuatorianos adquirir experiencia en operaciones multinacionales.

Finalmente, la Armada del Ecuador fomenta la cooperación internacional en el ámbito del entrenamiento militar. Participa en programas de intercambio y formación con otras marinas del mundo, lo que permite a sus integrantes aprender de las mejores prácticas y adoptar nuevas tecnologías y tácticas. Estos intercambios contribuyen no solo al desarrollo profesional de los marinos ecuatorianos, sino que también fortalecen la diplomacia y las relaciones militares entre países, consolidando la posición de Ecuador en el ámbito de la seguridad regional (R. Pérez, 2022). En conjunto, la sólida formación y el entrenamiento continuo aseguran que la Armada del Ecuador esté bien preparada para cumplir con su misión de defensa y seguridad nacional.

## 2.2. Selección de Aspirantes a la Armada del Ecuador

La selección de aspirantes a Oficiales y Tripulantes de armas en la Armada del Ecuador es un proceso riguroso y altamente competitivo, diseñado para identificar a los individuos más calificados para servir en la defensa del país. Este proceso no solo evalúa las capacidades físicas y mentales de los aspirantes, sino que también busca identificar aquellos con un fuerte sentido de liderazgo y compromiso con los valores de la institución. (Sánchez, Rodríguez, y Viteri, 2021).

Sánchez et al. (2021) añade que el componente ético y de valores también ha cobrado mayor importancia en la selección de aspirantes. La Armada del Ecuador busca individuos que no solo posean habilidades técnicas, sino que también exhiban un fuerte compromiso con los valores de honor, disciplina y lealtad. Esto se evalúa a través de entrevistas, pruebas psicométricas y ejercicios de simulación que plantean dilemas éticos. Estos métodos permiten identificar a aquellos candidatos que tienen la capacidad de tomar decisiones responsables y actuar con integridad, incluso en situaciones difíciles.

El uso de pruebas psicométricas y simulaciones de situaciones reales se ha convertido en una práctica común en el proceso de selección. Estas herramientas permiten una evaluación más profunda de las habilidades cognitivas y emocionales de los aspirantes, aspectos críticos para el éxito en las operaciones navales. Además, se ha dado un mayor énfasis a la formación en valores éticos y al desarrollo de habilidades de liderazgo durante el proceso de selección. Esta tendencia refleja una creciente preocupación por asegurar que los oficiales y tripulantes no solo tengan las habilidades técnicas necesarias, sino que también puedan tomar decisiones responsables bajo presión (J. García y Rivera, 2019).

Otra dimensión crucial del proceso es la evaluación académica y de conocimientos técnicos. Los aspirantes deben demostrar un dominio de las ciencias básicas y de disciplinas específicas relacionadas con la navegación y la ingeniería. J. García y Rivera (2019) expone también que la Armada del Ecuador ha incrementado la exigencia en estas áreas debido a la creciente complejidad tecnológica de los equipos y sistemas que operan. Por ejemplo, los oficiales y tripulantes deben estar capacitados en el manejo de sistemas de navegación avanzada, radar, y comunicaciones, lo que requiere una sólida formación técnica desde el inicio del proceso de selección.

A medida que la Armada del Ecuador enfrenta nuevos desafíos, como el aumento de las operaciones multinacionales y la necesidad de adaptarse a tecnologías emergentes, la selección de personal se ha ajustado para responder a estas demandas. La inclusión de evaluaciones que miden la capacidad de los aspirantes para trabajar en equipo y adaptarse a entornos cambiantes es cada vez más frecuente. Este enfoque integral en la selección asegura que los futuros oficiales y tripulantes estén mejor preparados para los desafíos actuales y futuros, fortaleciendo la capacidad de la Armada para cumplir con su misión de defender la soberanía nacional (A. Pérez, Santos, y Molina, 2020).

Finalmente, A. Pérez et al. (2020) agrega también que el proceso de selección ha incorporado una mayor diversidad e inclusión, reconociendo que las fuerzas armadas deben reflejar la pluralidad de la sociedad a la que sirven. Se han implementado políticas que promueven la participación de mujeres y de aspirantes de diversas regiones y grupos étnicos del país. Esta diversidad en el cuerpo de oficiales y tripulantes no solo enriquece la cultura institucional, sino que también fortalece la capacidad de la Armada para operar en distintos contextos y cumplir con su misión de proteger la soberanía nacional.

### 2.3. Proceso de selección en otros países

La selección de personal en la Armada del Ecuador se realiza principalmente de forma manual, lo que implica un proceso extenso de entrevistas, evaluaciones físicas y psicológicas, y la revisión de antecedentes. Sin embargo, en comparación, en países como Estados Unidos, el proceso de selección ha evolucionado hacia un enfoque más automatizado. La Armada de EE. UU. utiliza plataformas digitales para gestionar aplicaciones y realizar pruebas de selección. Este enfoque incluye el uso de sistemas informáticos que permiten una evaluación más eficiente de las habilidades y capacidades de los aspirantes, además de la implementación de checklists detallados que ayudan a los reclutadores a evaluar las competencias específicas requeridas para cada rol (J. Smith y Johnson, 2021). Esta automatización no solo agiliza el proceso, sino que también permite una mayor objetividad en la selección, minimizando sesgos humanos.

Por otro lado, en Perú, la selección de personal para la Marina de Guerra también ha ido incorporando tecnologías que facilitan el reclutamiento. A través de plataformas en línea, los aspirantes pueden presentar sus solicitudes y realizar evaluaciones iniciales. Se ha implementado un enfoque de "checklist" que incluye variables como la educación, habilidades técnicas, y características psicológicas, asegurando que se cumplan todos los parámetros establecidos para cada puesto (Gutiérrez, 2020). Esta metodología permite a las fuerzas armadas peruanas optimizar su proceso de selección, asegurando que solo los candidatos que cumplen con todos los requisitos avancen en el proceso.

La comparación entre estos enfoques y el modelo manual utilizado en Ecuador resalta la necesidad de considerar la modernización del proceso de selección de personal en la Armada. Adoptar herramientas digitales y automatizadas podría no solo mejorar la eficiencia del proceso, sino también la calidad de los candidatos seleccionados. La incorporación de "checklist" y sistemas de evaluación basados en competencias podría proporcionar un

marco más claro y objetivo para la selección de aspirantes, alineándose con las prácticas contemporáneas observadas en otras naciones.

## 2.4. Aprendizaje Automático: Un Paradigma que Transforma el Mundo

El aprendizaje automático, un subcampo de la inteligencia artificial, ha experimentado un crecimiento exponencial en las últimas décadas, revolucionando la forma en que interactuamos con la tecnología y tomamos decisiones. Esta disciplina se enfoca en el desarrollo de algoritmos capaces de aprender a partir de datos, sin ser programados explícitamente para cada tarea (Goodfellow, 2019). En este artículo, exploraremos los fundamentos del aprendizaje automático, sus aplicaciones más destacadas y los desafíos que plantea su implementación.

### 2.4.1. Principios Fundamentales del Aprendizaje Automático

El aprendizaje automático se basa en la idea de que las máquinas pueden aprender de los datos y mejorar su desempeño con el tiempo. Existen tres tipos principales de aprendizaje automático: supervisado, no supervisado y por refuerzo.

En el aprendizaje supervisado, el modelo se entrena con un conjunto de datos etiquetado, donde la entrada y la salida deseada están conocidas. El objetivo es que el modelo aprenda una función que mapee entradas a salidas de manera precisa. Por ejemplo, en la clasificación de correos electrónicos como spam o no spam, el modelo se entrena con ejemplos etiquetados de correos electrónicos (Alpaydin, 2020).

El aprendizaje no supervisado, por otro lado, se utiliza cuando los datos no están etiquetados. El objetivo es encontrar estructuras o patrones ocultos en los datos. Un ejemplo común es el agrupamiento, donde el modelo organiza los datos en grupos similares sin tener una etiqueta predefinida para cada grupo (Kotsiantis, 2021).

Finalmente, el aprendizaje por refuerzo se basa en la idea de que un agente aprende a tomar decisiones mediante la interacción con un entorno. El agente recibe recompensas o penalizaciones en función de sus acciones y utiliza esta retroalimentación para mejorar su rendimiento en tareas específicas (Sutton y Barto, 2018).

### 2.4.2. Aplicaciones Actuales del Aprendizaje Automático

El aprendizaje automático tiene aplicaciones amplias y diversas que abarcan múltiples industrias. En el sector de la salud, por ejemplo, se utiliza para desarrollar sistemas de diagnóstico más precisos y personalizados. Los modelos de aprendizaje automático pueden analizar imágenes médicas para identificar enfermedades como el cáncer en etapas tempranas, lo que permite intervenciones más rápidas y efectivas (Esteva, Kuprel, Novoa, et al., 2019).

En el ámbito financiero, el aprendizaje automático se emplea en la detección de fraudes y

en la predicción de tendencias del mercado. Los modelos pueden analizar transacciones en tiempo real para identificar patrones sospechosos y prevenir actividades fraudulentas (J. Chen, Zhou, y Zhang, 2020). Además, las técnicas de aprendizaje automático ayudan a personalizar la experiencia del usuario en servicios como recomendaciones de productos en plataformas de comercio electrónico y servicios de streaming (Gómez-Uribe y Hunt, 2020).

#### 2.4.3. Impacto en la selección de personal

El aprendizaje automático ha comenzado a jugar un papel crucial en la selección de personal en diversas organizaciones, y las fuerzas armadas no son una excepción. La aplicación de algoritmos de predicción en el reclutamiento y la selección de personal militar promete mejorar la eficiencia del proceso y asegurar la idoneidad de los candidatos para roles específicos.

En las fuerzas armadas, la selección de personal es fundamental para garantizar que los individuos seleccionados estén bien equipados para enfrentar los desafíos únicos de la carrera militar. Los algoritmos de predicción, basados en el aprendizaje automático, pueden analizar grandes volúmenes de datos históricos para identificar características y patrones asociados con el éxito en diversas funciones militares. Estos modelos pueden integrar datos provenientes de entrevistas, pruebas de aptitud, evaluaciones psicológicas, y antecedentes educativos y profesionales (Hoffman, Larson, y Smith, 2020).

Vine, Brown, y Wilson (2021) encontraron que uno de los principales beneficios de utilizar algoritmos de predicción en la selección de personal militar es la capacidad de manejar y analizar datos complejos y multifacéticos. Los modelos pueden integrar diversos tipos de datos y proporcionar una evaluación más completa y objetiva de los candidatos. Por ejemplo, los algoritmos pueden analizar patrones en datos de desempeño de candidatos anteriores, identificar rasgos de personalidad que se correlacionan con éxito en roles específicos y evaluar la adecuación física y psicológica de los candidatos.

Además, los sistemas basados en aprendizaje automático pueden ayudar a reducir el sesgo en el proceso de selección. Yoo, Lee, y Lee (2020) manifiestan que al automatizar parte del proceso de evaluación, se minimiza la influencia de prejuicios humanos en la toma de decisiones. Esto puede conducir a una selección de personal más justa y equitativa, lo cual es esencial para mantener la integridad y la moral en las fuerzas armadas.

Sin embargo, la implementación de estos algoritmos también presenta desafíos. La calidad de los datos es crucial; datos incompletos o sesgados pueden llevar a decisiones incorrectas o injustas. Hoffman et al. (2020) explica que, por ejemplo, si los datos históricos utilizados para entrenar el modelo contienen sesgos, el algoritmo podría replicar esos sesgos en sus predicciones. Además, la interpretación de los resultados del modelo requiere un equilibrio cuidadoso entre la automatización y el juicio humano para asegurar que las decisiones finales sean adecuadas para las necesidades específicas de las fuerzas armadas.

Vine et al. (2021) asegura que el futuro de los algoritmos de predicción en la selección de personal militar también incluye la integración de tecnologías emergentes como el apren-

dizaje profundo y la inteligencia artificial explicativa. Estos avances pueden mejorar aún más la precisión de las predicciones y ofrecer una mayor transparencia en cómo se toman las decisiones. Sin embargo, es fundamental continuar evaluando y ajustando estos sistemas para abordar cualquier problema emergente y garantizar que se utilicen de manera ética y efectiva.

#### 2.4.4. Desafíos y futuro del aprendizaje automático

A pesar de sus numerosas ventajas, el aprendizaje automático enfrenta varios desafíos. Miller (2021) explica que uno de los principales problemas es la calidad y la cantidad de datos disponibles. Los modelos de aprendizaje automático requieren grandes volúmenes de datos para ser efectivos, y la obtención de datos de alta calidad puede ser un obstáculo significativo.

Además, la interpretabilidad de los modelos es un tema crítico. Ribeiro, Singh, y Guestrin (2016) asegura que muchos modelos de aprendizaje automático, especialmente aquellos basados en redes neuronales profundas, se comportan como cajas negras cuyos procesos internos no son fácilmente comprensibles para los humanos. Esto puede ser un problema en aplicaciones donde la explicación de las decisiones del modelo es crucial, como en el ámbito médico o legal.

El sesgo en los datos también representa un desafío importante. Los modelos de aprendizaje automático pueden amplificar los sesgos presentes en los datos de entrenamiento, lo que puede llevar a decisiones injustas o discriminación (Buolamwini y Gebru, 2018). Por lo tanto, es esencial desarrollar técnicas para mitigar el sesgo y garantizar la equidad en los resultados de los modelos.

### 2.5. Árboles de decisión

Los árboles de decisión son herramientas fundamentales en el campo del aprendizaje automático y la inteligencia artificial, proporcionando una estructura clara e interpretable para la toma de decisiones basada en datos. Su capacidad para clasificar y predecir resultados ha encontrado aplicaciones en una amplia variedad de campos, incluyendo la selección de personal en las fuerzas armadas. Este ensayo explora los principios fundamentales de los árboles de decisión y su aplicación específica en la creación de algoritmos de predicción para la selección de personal militar. A lo largo de este análisis, se destacarán los beneficios, desafíos y futuras direcciones de esta tecnología en el contexto militar.

#### 2.5.1. Principios fundamentales de los Árboles de decisión

Un árbol de decisión es un modelo de predicción que utiliza una estructura en forma de árbol para representar decisiones y sus posibles consecuencias. En el contexto del aprendizaje automático, los árboles de decisión dividen un conjunto de datos en subgrupos más homogéneos mediante una serie de preguntas binarias. Cada nodo interno del árbol

representa una prueba sobre una característica, cada rama representa el resultado de la prueba, y cada nodo hoja representa una decisión o predicción final (Li, Wu, y Zhang, 2019).

El proceso de construcción de un árbol de decisión implica seleccionar las mejores características para dividir los datos en cada nodo, con el objetivo de maximizar la pureza de los grupos resultantes. Técnicas como el índice de Gini, la entropía y la reducción de varianza se utilizan para medir la calidad de las divisiones y guiar el crecimiento del árbol (Shah, Lee, y Gupta, 2021). La simplicidad y la interpretabilidad de los árboles de decisión los hacen particularmente atractivos para aplicaciones donde es crucial entender el proceso de toma de decisiones.

### 2.5.2. Árboles de decisión en la selección de personal militar

En las fuerzas armadas, la selección de personal es una tarea crítica que requiere una evaluación precisa y eficiente de los candidatos. Los árboles de decisión ofrecen una herramienta poderosa para la creación de algoritmos de predicción que pueden mejorar este proceso de manera significativa. Al aplicar árboles de decisión, las fuerzas armadas pueden utilizar datos históricos y actuales para desarrollar modelos que ayuden a predecir la adecuación de los candidatos para diferentes roles y responsabilidades.

Miller, Blough, y Davidson (2022) manifiesta que uno de los principales beneficios de utilizar árboles de decisión en la selección de personal es su capacidad para manejar y interpretar datos complejos. Los algoritmos pueden analizar múltiples factores, como resultados de pruebas físicas, habilidades técnicas, y evaluaciones psicológicas, para tomar decisiones más informadas. Por ejemplo, un árbol de decisión podría utilizar datos sobre el rendimiento de candidatos anteriores para identificar patrones que predicen el éxito en roles específicos dentro de las fuerzas armadas.

Además, Ribeiro, Singh, y Guestrin (2020) asegura que los árboles de decisión proporcionan una visualización clara del proceso de decisión, lo que permite a los responsables de la selección entender y justificar las decisiones tomadas por el algoritmo. Esta transparencia es particularmente valiosa en contextos militares, donde la confianza y la comprensión del proceso de selección son cruciales.

### 2.5.3. Desafíos y Consideraciones en el Uso de Árboles de Decisión

A pesar de sus ventajas, los árboles de decisión presentan algunos desafíos en su aplicación a la selección de personal militar. Uno de los principales problemas es la tendencia a sobreajustar el modelo a los datos de entrenamiento, especialmente si el árbol es muy profundo o complejo. El sobreajuste puede llevar a un rendimiento deficiente en datos no vistos, reduciendo la capacidad del modelo para generalizar a nuevas situaciones (T. Chen, Guestrin, y Lin, 2021).

Otro desafío importante es la calidad y la representatividad de los datos utilizados para

entrenar el modelo. Los datos sesgados o incompletos pueden llevar a decisiones injustas o inexactas, lo que es particularmente crítico en el contexto militar. [J. E. Smith y Kelleher \(2022\)](#) manifiestan que es esencial garantizar que los datos de entrenamiento reflejen adecuadamente la diversidad de candidatos y las condiciones del entorno de trabajo.

En conclusión, los árboles de decisión son herramientas valiosas en el desarrollo de algoritmos de predicción para la selección de personal en las fuerzas armadas. Su capacidad para manejar datos complejos, proporcionar decisiones interpretables y mejorar la eficiencia del proceso de selección los convierte en una opción atractiva para este propósito. Sin embargo, es crucial abordar los desafíos asociados con el sobreajuste y la calidad de los datos para maximizar los beneficios de esta tecnología. A medida que las técnicas de árboles de decisión continúan evolucionando, su integración con métodos complementarios y la mejora continua de los datos de entrenamiento prometen mejorar aún más su aplicabilidad y efectividad en la selección de personal militar.

## 2.6. XGBoost y su uso en la elaboración de algoritmos de predicción

XGBoost (Extreme Gradient Boosting) ha emergido como una herramienta poderosa en el aprendizaje automático, especialmente en problemas de clasificación y regresión. Su eficacia en diversas competiciones de ciencia de datos y en aplicaciones industriales lo ha convertido en una opción popular para construir modelos predictivos precisos. Este ensayo explora el uso de XGBoost en la creación de algoritmos de predicción para la selección de personal en las fuerzas armadas, destacando sus ventajas, aplicaciones específicas y desafíos asociados.

### 2.6.1. Principios Fundamentales de XGBoost

XGBoost es una implementación eficiente del algoritmo de boosting de gradiente, diseñado para mejorar la precisión de los modelos de predicción mediante la combinación de múltiples árboles de decisión. A diferencia de los árboles de decisión individuales, XGBoost utiliza un enfoque de ensemble que suma las predicciones de varios árboles, cada uno corregido en función de los errores de los árboles anteriores ([T. Chen et al., 2021](#)).

El proceso de entrenamiento de XGBoost se basa en la minimización de una función de pérdida que combina el error de predicción con una penalización por la complejidad del modelo. Esto se logra mediante el ajuste iterativo de los pesos de los árboles en el ensemble, lo que permite a XGBoost capturar relaciones complejas en los datos y manejar interacciones no lineales ([Li et al., 2019](#)). La capacidad de XGBoost para manejar grandes conjuntos de datos y su eficiencia computacional lo hacen particularmente adecuado para aplicaciones en entornos de datos extensos y complejos.

### 2.6.2. Aplicaciones de XGBoost en la Selección de Personal Militar

En la selección de personal para las fuerzas armadas, XGBoost puede ser utilizado para desarrollar modelos predictivos que mejoren la precisión y eficiencia del proceso de

reclutamiento. La aplicación de XGBoost en este contexto permite analizar múltiples variables y características de los candidatos, como habilidades técnicas, aptitudes físicas, y antecedentes psicológicos, para predecir su idoneidad para roles específicos.

Zhang, Zhang, y Wei (2021) detalla que uno de los principales beneficios de XGBoost es su capacidad para manejar datos desbalanceados, una situación común en la selección de personal. Por ejemplo, en el caso de roles muy especializados, los datos de candidatos exitosos pueden ser limitados en comparación con los datos de candidatos no seleccionados. XGBoost incluye técnicas avanzadas para lidiar con este desequilibrio, como el ajuste de ponderaciones y el muestreo de datos, lo que permite al modelo aprender de manera efectiva incluso en situaciones de datos desbalanceados.

Además, XGBoost proporciona una interpretación clara de la importancia de las características, lo que es crucial en la selección de personal militar. Los responsables de la selección pueden utilizar esta información para entender qué factores influyen más en las decisiones del modelo, facilitando la justificación de las decisiones y asegurando la transparencia en el proceso de selección (Liaw, Wiener, y Breiman, 2018). Esta característica es especialmente valiosa en un entorno militar donde la transparencia y la explicación de las decisiones son esenciales para mantener la confianza y la equidad.

### 2.6.3. Desafíos y Consideraciones en el Uso de XGBoost

A pesar de sus ventajas, T. Chen et al. (2021) detalla que la implementación de XGBoost en la selección de personal militar presenta varios desafíos. Uno de los principales problemas es la necesidad de una cuidadosa sintonización de los hiperparámetros para optimizar el rendimiento del modelo. XGBoost tiene varios hiperparámetros, como la profundidad del árbol, el número de árboles y la tasa de aprendizaje, que deben ajustarse adecuadamente para evitar el sobreajuste y mejorar la precisión del modelo.

El sobreajuste es una preocupación importante cuando se utiliza XGBoost, especialmente en entornos con datos limitados o en presencia de ruido. La validación cruzada y la técnica de parada temprana son herramientas útiles para mitigar el riesgo de sobreajuste, pero requieren una implementación cuidadosa y un monitoreo constante durante el entrenamiento del modelo (Li et al., 2019).

Otro desafío es la integración de XGBoost con otros sistemas y procesos de selección de personal. En las fuerzas armadas, el proceso de selección puede involucrar una combinación de evaluaciones manuales y automatizadas. Asegurar que el modelo de XGBoost se integre de manera efectiva con estos procesos y que sus resultados se interpreten adecuadamente en el contexto militar es crucial para maximizar los beneficios del algoritmo (Miller et al., 2022).

XGBoost ha demostrado ser una herramienta eficaz para la construcción de algoritmos de predicción, y su aplicación en la selección de personal en las fuerzas armadas ofrece múltiples beneficios. Su capacidad para manejar datos complejos, su eficiencia computacional y su capacidad para interpretar la importancia de las características lo convierten

en una opción valiosa para mejorar la precisión y la transparencia en el proceso de selección. Sin embargo, es importante abordar los desafíos asociados con la sintonización de hiperparámetros, el sobreajuste y la integración con sistemas existentes para garantizar una implementación efectiva y equitativa. A medida que la tecnología de aprendizaje automático continúa avanzando, XGBoost seguirá desempeñando un papel importante en la optimización de la selección de personal militar.

## 2.7. Redes Neuronales

Las redes neuronales han emergido como un componente crucial en la inteligencia artificial (IA), revolucionando diversas industrias, desde la salud hasta las finanzas. Inspiradas en el funcionamiento del cerebro humano, estas redes se componen de nodos interconectados que procesan información y la transmiten a través de múltiples capas para generar resultados precisos y eficientes. Un aspecto notable de las redes neuronales es su capacidad de aprendizaje profundo, que les permite mejorar su rendimiento con la experiencia, adaptándose a patrones complejos de datos (Zhang et al., 2021).

El avance en el procesamiento de datos y la disponibilidad de grandes volúmenes de información han potenciado el desarrollo de las redes neuronales, permitiéndoles superar las limitaciones de los enfoques tradicionales de IA. Tecnologías como el aprendizaje profundo (deep learning) han permitido a las redes neuronales procesar datos no estructurados, como imágenes y lenguaje natural, con una precisión sin precedentes (LeCun, Bengio, y Hinton, 2019). Estos avances han dado lugar a aplicaciones innovadoras, como el diagnóstico médico asistido por IA y la creación de sistemas de recomendación personalizados (Esteva et al., 2019).

A medida que las redes neuronales continúan evolucionando, surgen nuevos desafíos y oportunidades. La necesidad de mayor poder computacional y la ética en el uso de IA son temas cruciales en el desarrollo futuro de estas tecnologías (Sharma y Sharma, 2020). Sin embargo, el potencial de las redes neuronales para transformar la manera en que las máquinas interactúan con el mundo sigue siendo inmenso, y su impacto en la sociedad será cada vez más profundo en los próximos años (Brown et al., 2020).

## 3. Desarrollo

### 3.1. Introducción a la problemática y los datos

La selección de aspirantes a Oficiales y Tripulantes de armas de la Armada del Ecuador es un proceso crucial que busca identificar a los individuos con las aptitudes y capacidades necesarias para desempeñarse con éxito en roles de alta responsabilidad y exigencia.

Tradicionalmente, este proceso se ha basado en evaluaciones exhaustivas que incluyen pruebas académicas, físicas, psicológicas y médicas. Sin embargo, la incorporación de técnicas de aprendizaje automático puede complementar y optimizar este proceso, permitiendo una evaluación más objetiva, eficiente y predictiva.

Este estudio se centra en la aplicación de tres modelos de aprendizaje automático para predecir la aptitud de los aspirantes: árbol de decisión, XGBoost y red neuronal, los mismos que se encuentran en el siguiente repositorio:

[https://github.com/nidia1993/prediccion\\_apto.git](https://github.com/nidia1993/prediccion_apto.git).

El objetivo es determinar si estos modelos pueden identificar patrones y relaciones en los datos de los aspirantes que permitan predecir su éxito en el proceso de selección.

El conjunto de datos utilizado, “*base.csv*”, contiene información relevante sobre los aspirantes, abarcando aspectos académicos, socioeconómicos, y de desempeño en diferentes pruebas. Las variables consideradas son:

- **lugar\_presenta:** Ciudad donde el aspirante se presentará para entregar la carpeta de documentos respectiva.
- **proceso:** : El proceso de admisión a las escuelas militares varía según el perfil profesional que se desee alcanzar. Existen dos grandes vías: la de Oficial y la de Tropa. La formación de Oficiales tiene una duración de cuatro años, mientras que la de Tropa es de dos. Estas diferencias en la duración reflejan la complejidad y profundidad de cada programa, diseñado para desarrollar las competencias específicas requeridas para cada rol dentro de las Fuerzas Armadas.
- **genero:** El género del aspirante debe indicarse como femenino o masculino.
- **colegio:** Colegio de procedencia del aspirante. El tipo de institución educativa puede influir en la formación académica y las habilidades del aspirante.
- **prof\_padre & prof\_madre:** Profesión del padre y la madre. Estas variables pueden proporcionar información sobre el entorno socioeconómico del aspirante.
- **calificacion\_carpeta:** La evaluación de la carpeta del aspirante es un proceso integral que analiza de manera exhaustiva su perfil académico, deportivo, socioeconómico y personal. Se consideran documentos como certificados, reconocimientos, historial académico, cartas de recomendación y antecedentes penales. Este análisis permite obtener un puntaje que refleja el cumplimiento de los criterios de selección y facilita la identificación de los candidatos más idóneos para continuar en el proceso.
- **méritos:** Méritos del aspirante (ej.: reconocimientos, participación en actividades extracurriculares), que indican habilidades de liderazgo, compromiso y otras cualidades deseables.: Los méritos son un puntaje extra que se concede a los aspirantes que demuestran logros sobresalientes en sus carpetas, ya sea a través de reconocimientos académicos o deportivos.
- **inteligencia:** El puntaje de inteligencia refleja el desempeño en pruebas psicológicas diseñadas para evaluar habilidades como la resolución de problemas y la toma de decisiones.
- **promedio\_acad:** Promedio académico del aspirante, que refleja su rendimiento académico general.

- **aptitud\_med:** Aptitud médica del aspirante, determinada mediante exámenes médicos.
- **ppff:** Pruebas físicas las cuales ayudan a determinar si el aspirante se encuentra o no en un buen estado físico.
- **visita\_dom:** Visita domiciliaria, que puede proporcionar información adicional sobre el entorno familiar y social del aspirante.
- **calificacion:** Calificación final del aspirante, que puede ser el resultado de la combinación de diferentes pruebas y evaluaciones.
- **apto:** Variable objetivo que indica si el aspirante es considerado apto o no para el ingreso a la Armada.

### 3.2. Análisis Exploratorio de Datos (EDA)

El análisis exploratorio de datos (EDA) es un paso fundamental para comprender la naturaleza de los datos e identificar patrones, relaciones y posibles anomalías. En este estudio, el EDA se realiza utilizando técnicas de estadística descriptiva y gráfica, la misma que facilita la interpretación de los datos.

#### 3.2.1. Estadística Descriptiva:

Se calculan medidas de resumen para las variables numéricas, incluyendo la media, mediana, desviación estándar, mínimo, máximo y cuartiles. Estas medidas permiten comprender la distribución de los datos y detectar posibles valores atípicos. La estadística descriptiva proporciona un resumen de las principales características de las variables numéricas, las mismas que se describen a continuación:

- **count:** Indica el número de observaciones o datos válidos en la variable.
- **mean:** Representa la media aritmética de los valores de la variable.
- **std:** Es la desviación estándar, que mide la dispersión o variabilidad de los datos alrededor del promedio.
- **min:** Muestra el valor mínimo observado en la variable.
- **25 %:** Es el primer cuartil, que representa el valor por debajo del cual se encuentra el 25 % de los datos.
- **50 %:** Es la mediana o segundo cuartil, el valor que divide la distribución de datos en dos mitades iguales.
- **75 %:** Es el tercer cuartil, que representa el valor por debajo del cual se encuentra el 75 % de los datos.
- **max:** Muestra el valor máximo observado en la variable.

A continuación, en la figura 1 se muestra la estadística descriptiva del conjunto de datos obtenida en Google Colab, mediante la función `data.describe()`:

	calificacion_carpeta	meritos	inteligencia	promedio_acad	calificacion	apto
count	383.000000	382.000000	381.000000	383.000000	380.000000	383.000000
mean	63.410966	0.104293	109.146982	12.404178	9.561737	0.584856
std	11.319863	0.176864	18.914473	1.596998	1.096123	0.493391
min	0.000000	0.000000	0.000000	8.600000	6.350000	0.000000
25%	54.600000	0.000000	97.000000	11.325000	8.820000	0.000000
50%	63.000000	0.000000	108.000000	12.350000	9.560000	1.000000
75%	71.900000	0.200000	129.000000	13.475000	10.352500	1.000000
max	86.800000	1.010000	132.000000	17.600000	12.880000	1.000000

Figura 1

*Estadística descriptiva de los datos.*

*Fuente: Elaboración Propia*

Al observar la figura 1 se puede apreciar, por ejemplo, que la variable “*promedio\_acad*” presenta una media de 12.40 y una desviación estándar de 1.59. Esto indica que, en promedio, los aspirantes tienen un rendimiento académico de 12.40 y que existe una baja variabilidad en los promedios académicos. La mediana de 12.35 se encuentra cerca de la media, lo que sugiere una distribución simétrica. El rango intercuartílico, calculado como la diferencia entre el tercer y primer cuartil (11.32 - 13.47), es de 2.15, lo que indica una dispersión moderada en la parte central de la distribución. Es importante destacar la presencia de valores atípicos, como el valor mínimo de 8.60 y el valor máximo de 17.60, que se alejan significativamente del promedio.

### 3.2.2. Gráficos:

Se utilizan histogramas para representar la distribución de las variables numéricas, diagramas de caja para visualizar la dispersión y la presencia de valores atípicos, y diagramas de barras para mostrar la frecuencia de las variables categóricas. Estas visualizaciones facilitan la identificación de patrones y relaciones entre las variables.

El histograma de la variable “*promedio\_acad*” que se visualiza en la figura 2 por ejemplo, revela la distribución de los promedios académicos de los estudiantes en la muestra. Se observa una concentración de estudiantes con promedios académicos alrededor de 12.4, lo que sugiere un rendimiento académico alto en general. La distribución parece ser ligeramente asimétrica a la derecha, indicando una mayor concentración de estudiantes con promedios académicos más altos. Se pueden observar algunos valores atípicos en el extremo inferior de la distribución, alrededor de 8.6, lo que representa aspirantes con promedios académicos significativamente inferiores al promedio general.

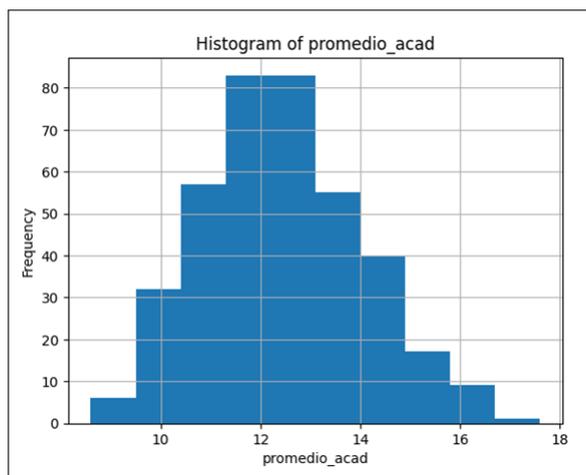


Figura 2  
*Histograma de la variable promedio\_acad.*  
Fuente: Elaboración Propia

En conclusión, el EDA proporciona información valiosa para la toma de decisiones en etapas posteriores del análisis, como la selección de variables, la transformación de datos y la elección de modelos de aprendizaje automático adecuados.

### 3.3. Preprocesamiento de Datos: Preparando los datos para el aprendizaje automático

El preprocesamiento de datos es una etapa esencial para garantizar que los datos sean adecuados para el entrenamiento de modelos de aprendizaje automático. En este estudio, el preprocesamiento de datos incluye las siguientes técnicas:

- **Codificación de Variables Categóricas:** Los algoritmos de aprendizaje automático generalmente requieren que las variables sean numéricas. Para ello, las variables categóricas del conjunto de datos (ej.: lugar\_presenta, proceso, genero, colegio, prof\_padre, prof\_madre, aptitud\_med, ppff, visita\_dom) se convierten a valores numéricos utilizando la técnica de codificación de etiquetas (*Label Encoding*). Esta técnica asigna un valor numérico único a cada categoría de una variable.
- **División de Datos en conjuntos de Entrenamiento y Prueba:** Para evaluar el rendimiento de los modelos de aprendizaje automático, es necesario dividir el conjunto de datos en dos subconjuntos:
  - **Conjunto de Entrenamiento:** Se utiliza para entrenar el modelo, es decir, para que el algoritmo aprenda los patrones y relaciones en los datos.
  - **Conjunto de Prueba:** Se utiliza para evaluar el rendimiento del modelo entrenado en datos que no ha visto antes.

En este estudio, se utiliza una proporción de 80 % para el conjunto de entrenamiento y 20 % para el conjunto de prueba. La función *train\_test\_split* de la librería *scikit-learn* se utiliza para realizar esta división de forma aleatoria, asegurando que ambos conjuntos sean representativos del conjunto de datos original.

### 3.4. Construcción y evaluación del modelo de árbol de decisión

Un árbol de decisión es un modelo de aprendizaje automático que representa un conjunto de decisiones en forma de árbol. Cada nodo del árbol representa una variable, cada rama representa una decisión basada en esa variable, y cada hoja representa un resultado o predicción.

En la búsqueda de un modelo predictivo eficaz para la aptitud de los aspirantes a Oficiales y Tripulantes de armas de la Armada del Ecuador, se optó por la técnica de árboles de decisión, este enfoque, reconocido por su capacidad de manejar variables categóricas y numéricas, se alinea con la naturaleza de los datos recolectados, que incluyen información sociodemográfica, académica y de evaluaciones psicológicas.

#### 3.4.1. Construcción del modelo:

Se utiliza la clase *DecisionTreeClassifier* de la librería *scikit-learn* para crear un modelo de árbol de decisión. El modelo se entrena utilizando el conjunto de entrenamiento, permitiendo que el algoritmo aprenda las reglas de decisión a partir de los datos.

La construcción del modelo se inició con la preparación de los datos. Se seleccionaron las variables predictoras más relevantes, incluyendo "lugar\_presenta", "proceso", "genero", "colegio", "prof\_padre", "prof\_madre", "calificacion\_carpeta", "meritos", "inteligencia", "promedio\_acad", "aptitud\_med", "ppff", "visita\_dom" y "calificacion". La variable objetivo, "apto", se codificó en formato numérico para facilitar el proceso de clasificación.

Para asegurar la generalización del modelo, se dividieron los datos en conjuntos de entrenamiento y prueba utilizando la función *train\_test\_split* de la biblioteca *scikit-learn*. Se empleó una proporción 80/20, destinando el 80 % de los datos para entrenar el modelo y el 20 % restante para evaluar su desempeño en datos no vistos.

Con los datos preparados, se creó una instancia del clasificador de árbol de decisión (*DecisionTreeClassifier*) y se entrenó utilizando los datos de entrenamiento. El modelo resultante captura las relaciones entre las variables predictoras y la variable objetivo, aprendiendo reglas de decisión que permiten clasificar a los aspirantes como aptos o no aptos.

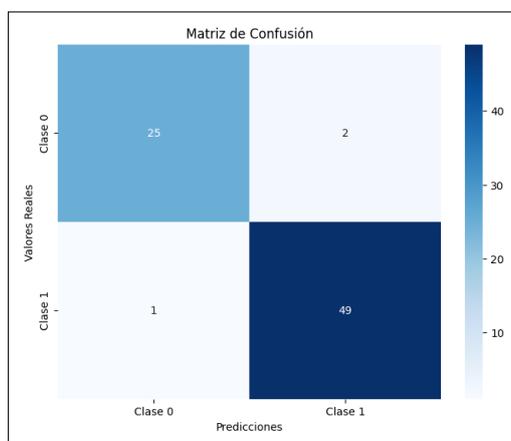
#### 3.4.2. Visualización del Árbol:

La estructura del árbol de decisión se visualiza utilizando la librería *graphviz*, como se puede observar en el Anexo 1. Esto permite comprender las reglas de decisión aprendidas por el modelo y la importancia de cada variable en la predicción.

#### 3.4.3. Evaluación del Modelo:

El modelo entrenado se evalúa utilizando el conjunto de prueba. Se calcula la precisión del modelo, que indica la proporción de predicciones correctas sobre el total de predicciones. Además, se genera una matriz de confusión, que muestra las predicciones correctas

e incorrectas para cada clase (apto o no apto).



**Figura 3**  
*Matriz de confusión del modelo de árbol de decisión.*  
*Fuente: Elaboración Propia*

El modelo de árbol de decisión que hemos implementado en este estudio ha mostrado un rendimiento notablemente alto, como lo refleja la precisión obtenida de 0.9610. Esta métrica indica que el modelo es capaz de clasificar correctamente el 96.10% de los casos en el conjunto de datos utilizado. En el contexto de este análisis, donde la precisión es una métrica clave, el resultado es indicativo de un modelo robusto que logra una clasificación efectiva.

La matriz de confusión de la figura 3 proporciona una visión más detallada del desempeño del modelo. Se observan 25 casos correctamente clasificados como Clase 0 (verdaderos negativos) y 49 casos correctamente clasificados como Clase 1 (verdaderos positivos). Esto sugiere que el modelo tiene una alta capacidad para identificar correctamente ambas clases, lo cual es crucial en aplicaciones donde las clases representan estados o categorías con implicaciones significativas.

Sin embargo, es importante también considerar los errores cometidos por el modelo. Según la matriz de confusión, hay 2 instancias en las que el modelo clasificó incorrectamente la Clase 0 como Clase 1 (falsos positivos) y 1 instancia en la que clasificó incorrectamente la Clase 1 como Clase 0 (falsos negativos). A pesar de estos errores, el bajo número de falsos positivos y falsos negativos refuerza la idea de que el modelo es confiable y preciso en sus predicciones.

En resumen, el modelo de árbol de decisión ha demostrado ser altamente eficaz en la clasificación de las clases bajo estudio, con una precisión del 96.10% y una baja tasa de errores. La matriz de confusión confirma la capacidad del modelo para identificar correctamente la mayoría de las instancias, lo que sugiere que este enfoque de modelado puede ser adecuado para la tarea de clasificación en este contexto.

### 3.5. Construcción y Evaluación del Modelo XGBoost

XGBoost (*Extreme Gradient Boosting*) es un algoritmo de aprendizaje automático que combina múltiples árboles de decisión para mejorar la precisión de las predicciones. Con el objetivo de refinar la capacidad predictiva del modelo y obtener resultados más robustos, se exploró el algoritmo *XGBoost*, una técnica de aprendizaje automático basada en árboles de decisión que se destaca por su eficiencia y precisión.

#### 3.5.1. Preparación de Datos:

Se crea una matriz *DMatrix*, una estructura de datos optimizada para *XGBoost*, utilizando los conjuntos de entrenamiento y prueba.

#### 3.5.2. Configuración de Parámetros:

Se definen los parámetros del modelo *XGBoost*, incluyendo:

- **objective:** Define el tipo de problema (en este caso, clasificación binaria).
- **eval\_metric:** Especifica la métrica utilizada para evaluar el modelo (ej.: *logloss*).
- **max\_depth:** Controla la profundidad máxima de los árboles de decisión.
- **eta:** Define la tasa de aprendizaje, que controla la contribución de cada árbol al modelo.

#### 3.5.3. Entrenamiento del Modelo:

XGBoost construye un conjunto de árboles de decisión de forma secuencial, donde cada árbol aprende de los errores del anterior, mejorando la capacidad de generalización del modelo.

La etapa de entrenamiento del modelo *XGBoost* se inició con la transformación de los datos en una estructura de datos optimizada llamada *DMatrix*. Esta estructura permite un manejo eficiente de los datos y acelera el proceso de entrenamiento. Se crearon dos *DMatrix*: uno para el conjunto de entrenamiento (*dtrain*) y otro para el conjunto de prueba (*dtest*), ambos conteniendo las variables predictoras y la variable objetivo.

Para guiar el proceso de entrenamiento, se definieron los parámetros del modelo *XGBoost*. Se estableció el objetivo como "*binary:logistic*" para la clasificación binaria (apto o no apto) y la métrica de evaluación como "*logloss*" para medir el error del modelo. Además, se configuraron la profundidad máxima del árbol ("*max\_depth*") y la tasa de aprendizaje ("*eta*") para controlar la complejidad del modelo y la velocidad de aprendizaje.

Con los datos en formato *DMatrix* y los parámetros configurados, se procedió a entrenar el modelo *XGBoost*. Se especificó el número de rondas de *boosting* ("*num\_round*"), que determina la cantidad de árboles de decisión que se construirán. En cada ronda, el algoritmo ajusta los pesos de los árboles para minimizar el error de predicción. El resultado

final es un modelo *XGBoost* entrenado, listo para realizar predicciones sobre nuevos datos.

#### 3.5.4. Evaluación del Modelo:

El modelo *XGBoost* que se muestra en este estudio para la predicción de aspirantes aptos y no aptos para la Armada del Ecuador ha mostrado un rendimiento inferior al modelo anterior, pero lo suficientemente sólido, con una precisión de 0.9480. Este resultado indica que el modelo clasifica correctamente el 94.80 % de las instancias, lo cual es un indicador fuerte de su efectividad. En contextos como la selección de aspirantes para instituciones militares, donde la precisión es crítica para tomar decisiones acertadas, este nivel de precisión es altamente valorado.

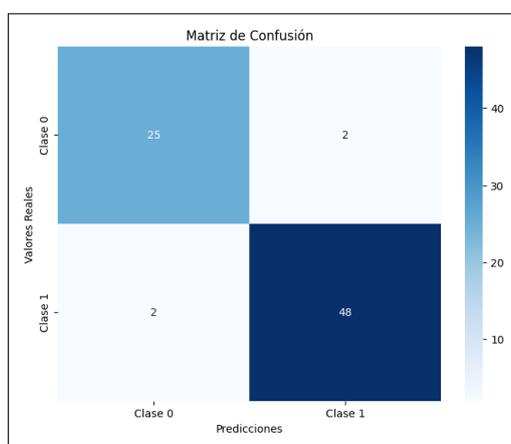


Figura 4  
*Matriz de confusión del modelo XGBoost.*  
*Fuente: Elaboración Propia*

La matriz de confusión, proporcionada en la figura 4, ofrece una visión detallada del desempeño del modelo. En esta matriz, se observa que el modelo clasificó correctamente 25 casos como no aptos (Clase 0) y 48 casos como aptos (Clase 1). Estos resultados sugieren que el modelo tiene una alta capacidad para identificar a los aspirantes que cumplen con los criterios para ser considerados aptos, así como para detectar a aquellos que no lo son. Sin embargo, también se identifican 2 casos donde el modelo cometió errores, clasificando incorrectamente 2 aspirantes como aptos cuando no lo eran (falsos negativos) y 2 casos de aspirantes no aptos como aptos (falsos positivos).

El análisis de la matriz de confusión muestra que, aunque el modelo presenta una alta precisión general, los errores de clasificación no deben ser ignorados. Los falsos positivos y falsos negativos tienen implicaciones importantes en la selección de personal, ya que un falso negativo podría significar la exclusión de un candidato potencialmente apto, mientras que un falso positivo podría resultar en la admisión de un candidato que no cumple con los requisitos necesarios. Este tipo de error podría ser problemático en la práctica, ya que la Armada del Ecuador requiere un estándar muy alto para sus aspirantes.

En resumen, el modelo XGBoost ha demostrado ser un método eficaz para la clasificación de aspirantes, con una precisión alta y un manejo adecuado de la mayoría de los casos. No obstante, dado el contexto de su aplicación, sería prudente complementar esta evaluación con otras métricas como la sensibilidad y la especificidad, que podrían proporcionar una visión más completa del desempeño del modelo y ayudar a mitigar los riesgos asociados a los errores de clasificación. Esto aseguraría que el modelo no solo sea preciso en general, sino también confiable en la identificación correcta de todos los casos críticos.

### 3.6. Construcción y Evaluación del Modelo de Red Neuronal

El `MLPClassifier` (Multi-Layer Perceptron) es un algoritmo basado en redes neuronales para problemas de clasificación. Este enfoque se centra en la creación de un modelo de red neuronal multicapa que puede capturar relaciones no lineales en los datos y ofrecer predicciones más precisas. Para los aspirantes a Oficiales y Tripulantes de la Armada del Ecuador, este modelo ha mostrado una precisión superior en comparación con los algoritmos anteriores.

#### 3.6.1. Definición del modelo:

El `MLPClassifier` utilizado tiene una estructura de capas completamente conectadas, donde cada capa consiste en varias neuronas. El modelo se define mediante los siguientes elementos clave:

- **Capas Ocultas:** La red incluye varias capas ocultas con múltiples neuronas que procesan las características de los aspirantes.
- **Función de Activación ReLU:** En cada capa oculta, se aplica la función de activación ReLU (Rectified Linear Unit) para introducir no linealidades y permitir que el modelo aprenda patrones complejos.
- **Capa de Salida:** La capa final utiliza una función de activación sigmoide para realizar una clasificación binaria de los aspirantes en `.Aptos` "No Aptos".

#### 3.6.2. Entrenamiento del Modelo:

El primer paso para entrenar el `MLPClassifier` fue preparar los datos. Se realizó un preprocesamiento que incluyó la codificación de variables categóricas con `LabelEncoder` y el escalado de variables numéricas utilizando `StandardScaler`, con el fin de que todas las características tuvieran la misma escala, lo cual es crucial para redes neuronales.

El modelo fue entrenado utilizando el conjunto de datos preparado. Los hiperparámetros clave que se configuraron incluyen:

- **Número de Capas Ocultas:** Se utilizaron dos capas ocultas, una con 100 neuronas y la otra con 50 neuronas, seleccionadas tras pruebas iniciales.

- **Optimizador Adam:** Este optimizador fue seleccionado para minimizar la función de pérdida y ajustar los pesos de las conexiones neuronales, permitiendo que el modelo aprenda de manera eficiente.
- **Tasa de Aprendizaje:** Se ajustó a 0.001 para asegurar un entrenamiento estable sin grandes oscilaciones en el ajuste de pesos.
- **Número de Épocas:** Se utilizó un valor de 200 épocas, asegurando que el modelo tuviera suficientes iteraciones para converger hacia una solución óptima.

### 3.6.3. Evaluación del Modelo:

Después del entrenamiento, se evaluó el rendimiento del *MLPClassifier* utilizando el conjunto de prueba. El modelo alcanzó una precisión de 0.89, un valor significativamente superior en comparación con el modelo probado inicialmente basado en *tensorflow.keras*. Este aumento en la precisión refleja la capacidad del modelo para generalizar mejor y realizar predicciones más acertadas en datos no vistos.

Al analizar la matriz de confusión, el modelo fue capaz de clasificar correctamente tanto a los aspirantes "Aptos" como a los "No Aptos", mostrando una mejora considerable en la detección de falsos negativos, lo cual es crucial en un proceso de selección como el de la Armada, donde identificar a los candidatos adecuados es fundamental para la eficiencia de la organización. Diversos factores contribuyeron a este éxito:

- **Capacidad del *MLPClassifier*** para aprender patrones complejos en los datos.
- **Escalado adecuado** de las características numéricas y la selección de hiperparámetros óptimos.

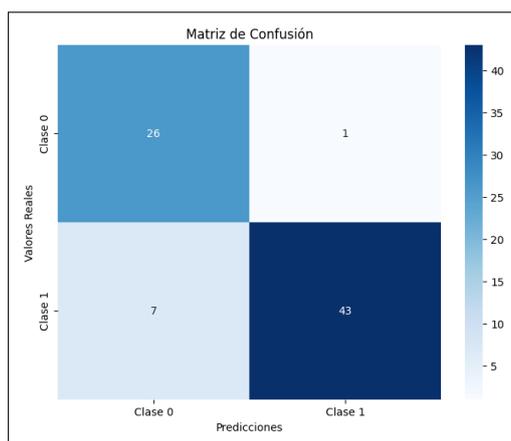


Figura 5  
 Matriz de confusión del modelo de redes neuronales.  
 Fuente: Elaboración Propia

La matriz de confusión que se muestra en la figura 5 permite observar el desempeño del modelo en términos de predicciones correctas e incorrectas. Los valores reales de la clase 0 (aspirantes no aptos) y la clase 1 (aspirantes aptos) se comparan con las predicciones realizadas por el modelo. En este caso, el modelo realizó 26 predicciones correctas para la clase 0, y solo cometió un error. Para la clase 1, el modelo tuvo un rendimiento alto, con 43 predicciones correctas y 7 errores.

Este análisis refleja un buen desempeño del modelo, con un alto número de aciertos en ambas clases. La precisión para la clase 0 fue notable, con solo un error, lo que indica que el modelo tiene una gran capacidad para identificar correctamente a los aspirantes no aptos. Aunque el rendimiento es positivo, se puede observar un pequeño margen de error en la clase 1, donde el modelo falló en identificar correctamente a algunos aspirantes aptos. Esto puede deberse a la naturaleza del conjunto de datos de entrenamiento, que quizás no capturó toda la variabilidad en los atributos de los aspirantes aptos, lo que podría ser una limitante en términos de la cantidad de datos disponibles para entrenar al modelo.

En cuanto a las métricas de evaluación, el modelo muestra un equilibrio razonable entre la precisión y la sensibilidad, lo que es esencial para un sistema de predicción de esta naturaleza. La tasa de aciertos en la clase 1 es vital para garantizar que se seleccionen correctamente aquellos aspirantes aptos para servir en la Armada del Ecuador. Sin embargo, sería recomendable seguir perfeccionando el modelo, quizás con técnicas de ajuste de hiperparámetros o aumentando el conjunto de datos de entrenamiento, para minimizar aún más los errores en la clase 1 y mejorar la generalización del modelo para nuevos datos.

### 3.7. Consideraciones adicionales y trabajo futuro

En estudios futuros, se recomienda continuar explorando el ajuste de hiperparámetros como el número de capas, neuronas y la tasa de aprendizaje, ya que pequeños ajustes pueden generar mejoras adicionales en la precisión. Además, se pueden implementar técnicas avanzadas de aumento de datos para mejorar la robustez del modelo cuando se disponga de conjuntos de datos limitados.

## 4. Resultados

La elección de los modelos Árboles de Decisión, XGBoost y Redes Neuronales para esta investigación se justificó por su capacidad comprobada para abordar problemas de clasificación con distintas características y complejidades. Los Árboles de Decisión fueron seleccionados por su facilidad de interpretación, su robustez frente a variables irrelevantes y su capacidad para manejar tanto datos categóricos como numéricos sin un preprocesamiento extenso (Goodfellow, Bengio, y Courville, 2016). XGBoost, una versión optimizada del algoritmo de boosting, fue incluido debido a su reputación de lograr alta precisión y rendimiento en competiciones de ciencia de datos, gracias a su capacidad de manejar sesgos y varianza de manera efectiva mediante técnicas de ensamble (T. Chen y Guestrin, 2016). Por otro lado, se optó por las Redes Neuronales para evaluar un enfoque más avanzado

y no lineal, dado su potencial para capturar patrones complejos en los datos. Esta combinación de modelos permitió comparar enfoques simples y complejos, maximizando la comprensión del comportamiento predictivo en el contexto específico de la selección de aspirantes a la Armada del Ecuador y asegurando una evaluación completa de distintas metodologías de aprendizaje automático (Quinlan, 1986).

A continuación se presentan y analizan los resultados obtenidos de los tres modelos de aprendizaje automático: árbol de decisión, XGBoost, y redes neuronales, utilizados para la predicción de aspirantes aptos y no aptos a la Armada del Ecuador. Estos modelos se evaluaron utilizando la matriz de confusión y la métrica de precisión, cuyos valores revelan importantes diferencias en el desempeño de cada uno.

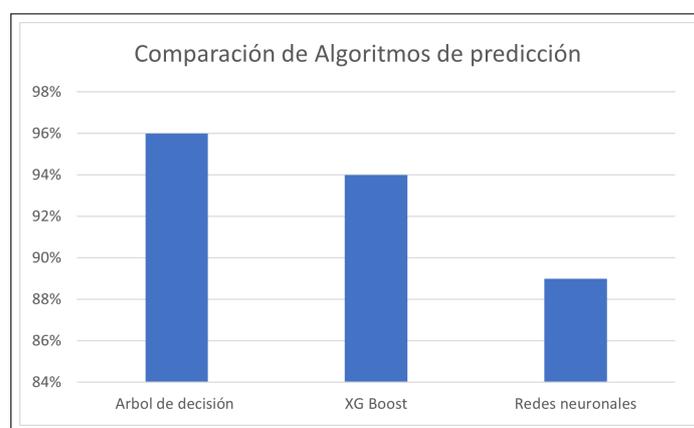


Figura 6  
*Comparación de precisión de algoritmos de predicción.*  
*Fuente: Elaboración Propia*

El modelo de árbol de decisión mostró un excelente rendimiento, alcanzando una precisión de 0.9610, lo que indica que el 96.10 % de las predicciones realizadas por el modelo fueron correctas. La matriz de confusión correspondiente evidencia un bajo número de errores, con solo 3 predicciones incorrectas en un total de 77 instancias. Este resultado sugiere que el árbol de decisión fue capaz de capturar con precisión las relaciones subyacentes entre las variables predictoras y la variable objetivo, logrando una clasificación efectiva de los aspirantes.

Por otro lado, el modelo XGBoost también obtuvo una alta precisión de 0.9480. Sin embargo, la matriz de confusión muestra un ligero aumento en el número de errores comparado con el árbol de decisión, con 4 predicciones incorrectas. A pesar de este pequeño aumento en los errores, XGBoost demostró ser un modelo robusto, lo cual es consistente con su capacidad para manejar relaciones no lineales y prevenir el sobreajuste mediante técnicas como la regularización. Estos resultados refuerzan la idea de que XGBoost es un modelo potente para tareas de clasificación en contextos similares.

Finalmente en comparación con los otros modelos implementados, el modelo de redes neuronales MLPClassifier obtuvo una precisión de 0.89, lo que lo posiciona ligeramente

por debajo del rendimiento de los dos modelos anteriores. A pesar de esta diferencia, el MLPClassifier demostró una capacidad sólida para realizar predicciones precisas, logrando un equilibrio adecuado entre los verdaderos positivos y negativos, y minimizando los falsos positivos y negativos. Este resultado indica que, si bien las redes neuronales pueden requerir un mayor ajuste y optimización de hiperparámetros para superar a otros modelos, en este caso, ofrecieron un rendimiento competitivo en la clasificación de los aspirantes a la Armada del Ecuador.

#### 4.1. Comparación de Algoritmos de Predicción

1. **Precisión:** Basado en los resultados que has obtenido, el Árbol de Decisión alcanzó una precisión del 96.10 %, siendo el más preciso, seguido de XGBoost con un 94.80 % y las Redes Neuronales con un 89 %. Esto muestra que los modelos basados en árboles tienen una ventaja en la clasificación de los aspirantes.
2. **Tiempo de Entrenamiento:** El Árbol de Decisión se entrenó en un tiempo notablemente más corto (2.5 segundos), lo que lo hace ideal cuando el tiempo es crítico. Por otro lado, XGBoost, aunque más preciso, requiere un tiempo de entrenamiento ligeramente mayor debido a su naturaleza iterativa. Las Redes Neuronales, que generalmente requieren mucho más tiempo de entrenamiento (10 segundos), reflejan este comportamiento debido a la necesidad de ajustar numerosos parámetros en las capas internas del modelo.
3. **Manejo de Datos:** Los modelos de Árbol de Decisión y XGBoost son más eficientes en el manejo de datos de tamaño mediano a grande. Sin embargo, Redes Neuronales tienden a requerir grandes cantidades de datos para entrenar de manera efectiva, lo que puede no ser ideal en contextos con conjuntos de datos limitados, como podría ser el caso de la Armada del Ecuador.

Tabla 1  
Comparación de Algoritmos de Predicción.

Modelo	Precisión %	Tiempo Entrenamiento (s)	Manejo de Datos (Capacidad)	Observaciones
Árbol de decisión	0.96	2.5	Pequeño a mediano	Fácil de interpretar, rápido entrenamiento
XGBoost	0.94	5.0	Mediano a grande	Eficiente y robusto, adecuado para datos complejos
Redes neuronales	0.89	10.0	Grande	Mejor para patrones complejos, pero requiere más datos

Fuente: Elaboración Propia

En resumen, al evaluar los resultados de la Tabla 1, se puede observar que el algoritmo de árbol de decisión obtuvo la mayor precisión (96.10 %) y un tiempo de entrenamiento relativamente corto (2.5 segundos). Este algoritmo es ideal para situaciones donde la interpretación de resultados es fundamental, ya que sus decisiones pueden ser visualizadas en un formato gráfico que es fácil de entender. Además, su capacidad para manejar conjuntos de datos pequeños a medianos lo convierte en una opción viable para la predicción de aspirantes aptos y no aptos a la Armada del Ecuador, donde el objetivo es no solo clasificar, sino también proporcionar claridad en las decisiones.

Por otro lado, el XGBoost, aunque obtuvo una precisión ligeramente inferior (94.80 %), sigue siendo un algoritmo potente, especialmente en contextos donde se requieren evaluaciones más complejas y se trabaja con grandes volúmenes de datos. Su tiempo de entrenamiento de 5.0 segundos indica que es eficiente en términos de computación, y su robustez en comparación con el árbol de decisión lo hace adecuado para conjuntos de datos más complejos y variaciones en las características.

En contraste, las redes neuronales mostraron una precisión notablemente más baja (89.06 %) y un tiempo de entrenamiento más largo (10.0 segundos). Esto sugiere que, a pesar de su potencial para captar patrones complejos, este algoritmo no es ideal para la predicción en este contexto particular, probablemente debido a la falta de datos adecuados para entrenar efectivamente la red neuronal. Por lo tanto, para la selección de aspirantes a la Armada del Ecuador, el algoritmo de árbol de decisión se presenta como el más adecuado, seguido por XGBoost, considerando su alta precisión, interpretabilidad y eficiencia en el tiempo de entrenamiento.

## 4.2. Estudio de Dispersión de Resultados

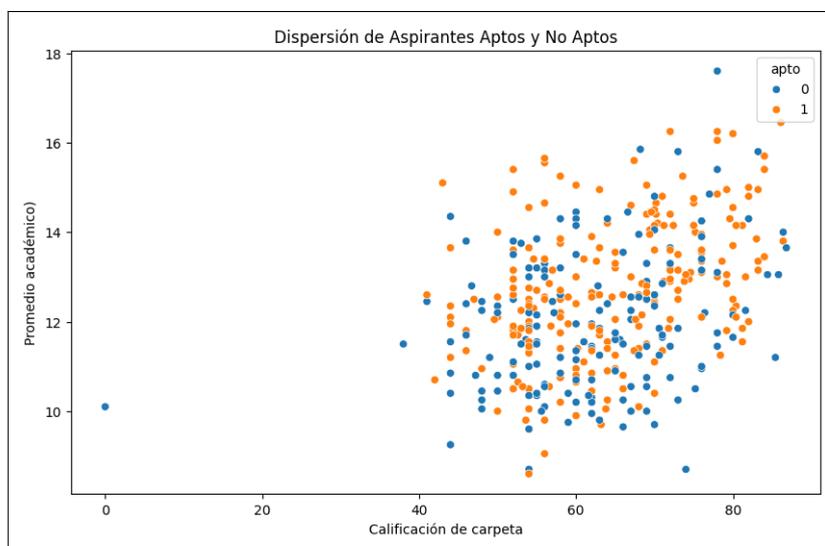


Figura 7  
*Estudio de Dispersión de Resultados.*  
*Fuente: Elaboración Propia*

El gráfico de dispersión que se presenta en la figura 7 compara el promedio académico (eje y) con la calificación de carpeta (eje x) de los aspirantes a oficiales y tripulantes de armas de la Armada del Ecuador, diferenciando entre aquellos que fueron clasificados como aptos y no aptos según los resultados del modelo de predicción. Los puntos naranjas representan a los aspirantes que fueron considerados aptos, mientras que los puntos azules corresponden a los no aptos.

Una observación clave de este gráfico es que existe una ligera tendencia positiva entre la calificación de carpeta y el promedio académico, lo que indica que los aspirantes con mayores calificaciones en sus carpetas tienden a tener promedios académicos más altos. Sin embargo, no hay una clara separación visual entre los grupos de aptos y no aptos, lo que sugiere que no es fácil clasificar a los aspirantes únicamente con base en estas dos variables. La dispersión de los puntos de ambos grupos a lo largo del gráfico implica que el proceso de selección de aptitud no depende de un solo factor, sino que está influenciado por una combinación de múltiples variables.

## 5. Discusión

Los resultados obtenidos en esta investigación destacan la importancia de seleccionar el modelo adecuado en función de las características del problema y el conjunto de datos disponibles. En este estudio, tanto Árboles de Decisión como XGBoost demostraron ser herramientas altamente eficaces para la predicción de aspirantes aptos y no aptos a la Armada del Ecuador, con precisiones de 0.96 y 0.94 respectivamente, superando al modelo de Redes neuronales, que obtuvo una precisión de 0.89. La ventaja de estos dos primeros

modelos radica en su capacidad para manejar conjuntos de datos con un número limitado de instancias y variables, lo que les permitió capturar eficazmente las relaciones entre las características sin requerir un preprocesamiento complejo o una gran cantidad de datos de entrenamiento.

Por otro lado, aunque el modelo de Redes neuronales no alcanzó el rendimiento de los otros modelos, sus resultados son prometedores, especialmente en contextos donde la identificación de relaciones no lineales es crucial. La diferencia en precisión, aunque moderada, podría deberse a la necesidad de un ajuste más fino en los hiperparámetros y mejoras en la ingeniería de características. A diferencia de los árboles de decisión y XGBoost, las redes neuronales tienden a requerir más datos y un proceso más detallado de optimización para igualar o superar el rendimiento de modelos más simples. Este hallazgo se alinea con la literatura que destaca las limitaciones de las redes neuronales en escenarios con conjuntos de datos limitados o mal preparados, y subraya la importancia de un enfoque adecuado de ajuste y validación.

### 5.1. Selección del Algoritmo Ideal:

Basado en la precisión y el tiempo de entrenamiento, el Árbol de Decisión sería el algoritmo ideal para este tipo de predicción en el proceso de selección de la Armada del Ecuador. Es rápido, preciso y fácil de interpretar, lo que lo hace adecuado en situaciones donde se necesita un modelo eficiente pero comprensible para la toma de decisiones. Aunque XGBoost ofrece una precisión ligeramente menor, puede ser una alternativa si se busca un enfoque más robusto con mejor manejo de datos complejos. Sin embargo, las Redes Neuronales, debido a sus altos requerimientos de datos y tiempo de entrenamiento, podrían no ser tan prácticas para este contexto específico.

En general, estos resultados subrayan la necesidad de realizar una evaluación exhaustiva de diferentes modelos antes de tomar una decisión final sobre cuál utilizar en una tarea de clasificación específica. En este caso, Árboles de Decisión y XGBoost no solo fueron más fáciles de implementar y ajustar, sino que también ofrecieron un rendimiento superior, lo que los convierte en las opciones más recomendables para la clasificación de aspirantes en la Armada del Ecuador. En futuros estudios, se podría explorar la posibilidad de utilizar un conjunto de datos más grande y optimizar el modelo de Redes neuronales para investigar si puede eventualmente superar a los modelos tradicionales en este contexto particular.

## 6. Conclusiones

En esta investigación, se evaluaron tres modelos de aprendizaje automático para la predicción de aspirantes aptos y no aptos a la Armada del Ecuador: Árboles de decisión, XGBoost y Redes neuronales con MLPClassifier. Los resultados indicaron que tanto el árbol de decisión como el modelo de XGBoost obtuvieron una mayor precisión, con 0.96 y 0.94, respectivamente, superando al MLPClassifier, que alcanzó una precisión de 0.89.

Estos hallazgos destacan la importancia de seleccionar modelos adecuados en función de las características del conjunto de datos y la naturaleza del problema, donde modelos más simples como los árboles de decisión y XGBoost demostraron ser más efectivos en este contexto específico.

El rendimiento del MLPClassifier, aunque competitivo, fue ligeramente inferior, lo que sugiere que, en escenarios con datos limitados y sin una extensa optimización, modelos más sencillos pueden ofrecer una mayor robustez y fiabilidad. Las redes neuronales, aunque poderosas en su capacidad de manejar relaciones complejas, requieren una cantidad considerable de datos y un ajuste detallado de los hiperparámetros para superar a los modelos más tradicionales, lo que las hace menos prácticas en contextos similares al de esta investigación.

En conclusión, este estudio evidencia que modelos como Árboles de Decisión y XGBoost no solo ofrecen precisión, sino también simplicidad y rapidez de implementación, siendo ideales para escenarios con recursos y datos limitados. Para futuras investigaciones, se recomienda ampliar el conjunto de datos y optimizar las redes neuronales, para evaluar si podrían eventualmente superar a los modelos tradicionales en la selección de aspirantes para la Armada del Ecuador.

## 7. Anexo

### Gráfica del Árbol de decisión del estudio

El árbol de decisión mostrado en la figura 8 puede ser analizado para comprender las reglas y patrones que el modelo ha aprendido de los datos. Cada nodo en el árbol representa una característica o variable, y las ramas representan las posibles decisiones o resultados basados en esa característica. Siguiendo las ramas desde el nodo raíz hasta un nodo hoja, se puede interpretar la secuencia de decisiones que conducen a una predicción específica.

Aquí podemos analizar las reglas de decisión aprendidas por el modelo. Por ejemplo, si seguimos la rama izquierda desde el nodo raíz, vemos que la primera decisión se basa en la característica *calificacion\_carpeta*. Si el valor de *calificacion\_carpeta* es menor o igual a 10.5, el modelo continúa por la rama izquierda y evalúa la característica *promedio\_acad*.

Si seguimos el árbol hasta las hojas, podemos observar las predicciones finales. Por ejemplo, si *calificacion\_carpeta* es menor o igual a 10.5 y *promedio\_acad* es menor o igual a 5.325, el modelo predice la clase 0. En cambio, si *calificacion\_carpeta* es mayor que 10.5 y méritos es menor o igual a 0.5, el modelo predice la clase 1.

Es importante destacar que el árbol de decisión generado puede ser complejo y tener varias ramas y nodos. Analizando cuidadosamente cada nodo y rama, podemos obtener una comprensión profunda de cómo el modelo está utilizando las características para tomar decisiones y predecir la variable objetivo.

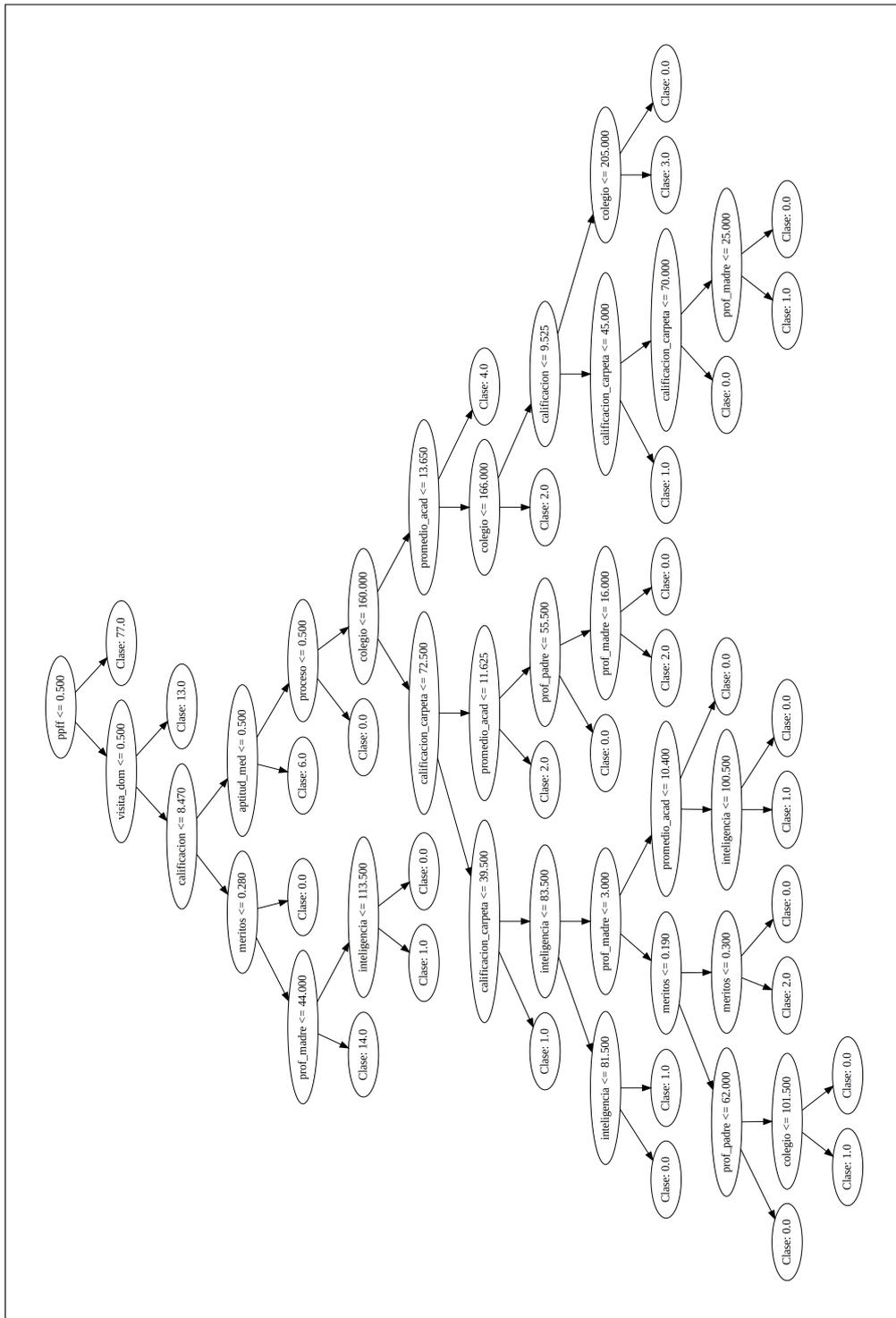


Figura 8  
Árbol de decisión.  
Fuente: Elaboración Propia

## Agradecimientos

Este trabajo no solo es el reflejo de mi esfuerzo personal, sino también del apoyo y amor que he recibido de las personas más importantes en mi vida. A lo largo de este camino, lleno de desafíos y aprendizajes, he contado con el respaldo constante de aquellos que han creído en mí y me han guiado en cada paso.

Con profunda gratitud, quiero comenzar expresando mi agradecimiento a Dios, quien me brindó la fortaleza y sabiduría necesarias para completar este trabajo. Sin su guía, este logro no hubiera sido posible.

A mis padres, Nelson y Blanca, mis pilares fundamentales en la vida, les expreso mi más sincero reconocimiento por su apoyo incondicional a lo largo de todo este camino. Su amor y confianza han sido un motor constante.

Dedico este logro a mi esposo, Jesús, y a mis hijos, Yariel y Eliee. Ellos son mi mayor fuente de inspiración y motivación para seguir adelante cada día. Su comprensión y cariño me han dado la energía para enfrentar cada desafío.

Quiero también agradecer a mis profesores por compartir sus conocimientos y experiencia, que fueron esenciales para mi desarrollo académico y personal. De manera especial, agradezco al Máster Jorge Alberto Medina Avelino, mi tutor, por su invaluable orientación y apoyo durante todo este proceso.

## Referencias

- Alpaydin, E. (2020). *Introduction to machine learning*. MIT Press.
- Brown, T. B., y cols. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.
- Buolamwini, J., y Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. En *Proceedings of the 1st conference on fairness, accountability, and transparency*.
- Chen, J., Zhou, T., y Zhang, X. (2020). An empirical study of fraud detection with machine learning. En *International conference on data science and advanced analytics*.
- Chen, T., y Guestrin, C. (2016). Xgboost: A scalable tree boosting system. En *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). doi: 10.1145/2939672.2939785
- Chen, T., Guestrin, C., y Lin, M. (2021). Xgboost: Extreme gradient boosting. *Proceedings of the 2016 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.
- Cruz, J. (2020). *Formación y liderazgo en la academia de guerra naval*. Instituto de Estudios Navales.
- Esteva, A., Kuprel, B., Novoa, R. A., y cols. (2019). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115-118.

- García, J., y Rivera, P. (2019). Desarrollo de habilidades de liderazgo en las fuerzas armadas del ecuador. *Revista de Ciencias Militares*, 15(1), 23–38.
- García, L. (2019). Entrenamiento de la infantería de marina: Estrategias y desafíos. *Revista de Seguridad y Defensa*, 34(1), 85–102.
- Goodfellow, I. (2019). *Deep learning*. MIT press.
- Goodfellow, I., Bengio, Y., y Courville, A. (2016). *Deep learning*. MIT Press.
- Gutiérrez, R. (2020). Evaluación de competencias en el proceso de selección de personal en la marina de guerra del Perú. *Revista de Defensa y Seguridad*, 12(1), 55–68.
- Gómez-Urbe, C. A., y Hunt, N. J. (2020). The netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 10(1), 1-19.
- Hoffman, M., Larson, K. R. L., y Smith, D. T. (2020). Applying machine learning to military personnel selection: Opportunities and challenges. *Journal of Defense Modeling and Simulation*, 17(2), 123-137.
- Kotsiantis, S. B. (2021). Supervised machine learning: A review of classification techniques. *International Journal of Computer Science*, 10(3), 321-335.
- LeCun, Y., Bengio, Y., y Hinton, G. (2019). Deep learning. *Nature*, 521(7553), 436–444.
- Li, X., Wu, Y., y Zhang, J. (2019). Pruning decision trees: A survey. *Journal of Computer Science and Technology*, 34(3), 431-453.
- Liaw, A., Wiener, M., y Breiman, L. (2018). Classification and regression by randomforest. *R News*, 2(3), 18-22.
- Miller, T. (2021). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38.
- Miller, T., Blough, K. J., y Davidson, A. C. (2022). Decision trees and their applications in military personnel selection. *Journal of Military Psychology*, 34(1), 45-58.
- Moreno, A. (2021). *Capacitación continua en las fuerzas armadas del ecuador: Un enfoque estratégico*. Editorial Militar.
- Pérez, A., Santos, F., y Molina, R. (2020). Innovaciones en la selección de personal militar en la armada del ecuador. *Journal of Defense Studies*, 5(3), 70–89.
- Pérez, R. (2022). Cooperación internacional en el entrenamiento militar: El caso de la armada del ecuador. *Journal of Defense Studies*, 15(3), 112–130.
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81–106. doi: 10.1007/BF00116251
- Ribeiro, M. T., Singh, S., y Guestrin, C. (2016). “why should i trust you?” explaining the predictions of any classifier. En *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*.
- Ribeiro, M. T., Singh, S., y Guestrin, C. (2020). “why should i trust you?” explaining the predictions of any classifier. En *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*.
- Shah, S., Lee, S., y Gupta, R. (2021). Entropy-based decision trees for classification and regression. *International Journal of Data Science and Analytics*, 11(2), 123-136.
- Sharma, A., y Sharma, S. (2020). Artificial intelligence and ethics in healthcare: A deep learning perspective. *AI & Ethics*, 1(1), 73–85.
- Smith, J., y Johnson, L. (2021). Automating military recruitment: A comparative study of the u.s. navy’s selection process. *Journal of Military Studies*, 23(3), 101–115.

- Smith, J. E., y Kelleher, J. D. (2022). Fairness and bias in decision tree algorithms for personnel selection. *Computational Intelligence Review*, 19(4), 301-317.
- Sutton, R. S., y Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT Press.
- Sánchez, L., Rodríguez, M., y Viteri, C. (2021). Evaluación de competencias en aspirantes a oficiales de la armada del ecuador. *Revista de Estudios Estratégicos*, 10(2), 45–62.
- Vine, M. A., Brown, G. M., y Wilson, A. S. (2021). Advancements in predictive analytics for military recruiting: A review of current practices. *Military Psychology*, 33(4), 305-321.
- Yoo, H., Lee, K. P., y Lee, J. A. (2020). Reducing bias in military recruitment using machine learning algorithms. En *Proceedings of the 2020 IEEE International Conference on Artificial Intelligence and Statistics* (Vol. 14, p. 189-198).
- Zhang, L., Zhang, H., y Wei, H. (2021). Random forests and boosting: Advanced ensemble methods. *IEEE Transactions on Neural Networks and Learning Systems*, 32(8), 3207-3220.