

Addis Ababa University

College of Natural and Computational Science

School of Information system

Passenger Satisfaction Prediction Model for Ethiopian Airline Utilizing Machine Learning Algorithms

A Thesis Submitted in Partial Fulfillment of the Requirement for the Degree of Master of System in Information System

***By*:**
Dureti Guye Danbi

Advisor: Melkamu Beyene (PhD)

Addis Ababa, Ethiopia

Aug, 2025

Addis Ababa University

Collage of Natural and Computational Science

School of Information Science

Design Passenger Satisfaction Prediction Model for Ethiopian Airline Utilizing Machine Learning Algorithms

Signature for Approval

Name	Signature	Date
Melkamu Beyene (PhD) , Advisor	_____	_____
(PhD), Examiner	_____	_____
(PhD), Examiner	_____	_____

Declaration:

I, the undersigned, hereby declare that the work presented in this thesis is my own original research and has not been submitted previously for any degree or qualification at any other institution. All sources of information and literature used have been appropriately acknowledged and cited.

Dureti Guye Denbi

Signature: _____

Date: _____

Acknowledgement

I Would like to express my deepest gratitude to Almighty God for His infinite grace, wisdom, and strength that guided me throughout the journey of my thesis. Without His blessings, this accomplishment would not have been possible.

I would like to extend my deepest gratitude to my advisor, Melkamu Beyene (PhD), for his invaluable suggestions, guidance, advice, responsiveness, and readiness to support me during the thesis work.

I am profoundly grateful to my beloved husband, whose unwavering support, patience, and encouragement have been my constant source of motivation.

List of tables

Table 1: The summary of related works. 14
Table 2: Description of meta data 19
Table 2: Summary of Best contributing factors for passenger satisfaction 34
Table 3: Performance summary for classical machine learning algorithms. 59
Table 4: Alignment of Research Questions with Key Findings 61

List of Figure

Figure 1 Artificial intelligence types and their relationship 9
Figure 2 Results of Exploratory Data Analysis (EDA): Customer Distribution 17
Figure 3 Results of Exploratory Data Analysis (EDA): Distribution of Rating of different Services 18
Figure 4: Top 5 Important hidden Features selected by Random Forest Algorithm 31
Figure 5: Top 5 Important hidden Features selected by XGBoost Algorithm 32
Figure 6: Top 5 Features Based on SHAP with XGBoost Values. 33
Figure 7: Top 5 Features Based on LIME with XGBoost Values 33
Figure 8: Classification metrics for Logistic Regression 36
Figure 9: Confusion Matrices Result for Logistic Regression 37
Figure 10: Model Interpretability for Logistic Regression 38
Figure 11: Classification Report for Decision Tree Algorithm 39
Figure 12: Confusion Matrices Result for Logistic Regression 40
Figure 13: Model Interpretability for Decision Tree Algorithm 41
Figure 14: Classification Report for Random Forest Algorithm 42
Figure 15: Confusion Matrices Result for Logistic Regression 43
Figure 16: Model Interpretability for Random Forest Algorithm 44
Figure 17: Classification Report for Support Vector Machine Algorithm 45
Figure 18: Confusion Matrices Result for Support Vector Machine 46
Figure 19: SHAP Summary Plot for Support Vector Machine Algorithm 47
Figure 20: Classification Report for XGBoost Algorithm 48
Figure 21: Confusion Matrices Result for XGBoost Algorithm 49
Figure 22: Feature By gain result for XGBoost algorithm 50
Figure 23: Performance summary for classical machine learning algorithms. 51
Figure 24: Performance summary for classical machine learning algorithms. 52
Figure 25: Classification Metrics for Multilayer Perceptron 53
Figure 26: Confusion Matrices Result for MLP Algorithm 54
Figure 27: Classification Report for TabNet Algorithm 55
Figure 28: Confusion Matrices Result for TabNet Algorithm 56
Figure 29: Classification Report for Deep and Cross Network Algorithm 57
Figure 30: Confusion Matrices Result for DCN Algorithm 58
Figure 31: Performance summary for classical machine learning algorithms. 59

List of Acronyms

AI — Artificial Intelligence.

EDA — Exploratory Data Analysis.

XAI — Explainable Artificial Intelligence.

SHAP — Shapley Additive explanations.

LIME — Local Interpretable Model-agnostic Explanations.

ML — Machine Learning.

RF — Random Forest.

SVM — Support Vector Machine.

RFE — Recursive Feature Elimination

NLP — Natural Language Processing (if applicable, e.g., in analyzing customer reviews).

Abstract:

Passenger satisfaction prediction is a critical task for Ethiopian Airlines to maintain service excellence and sustain a competitive advantage. Simple Traditional methods, such as statistical tools, often lack the analytical depth required to uncover complex patterns in large-scale passenger data. This study aims to develop a machine learning-based predictive model capable of assessing passenger satisfaction proactively, allowing Ethiopian Airlines to implement timely improvements and maintain service excellence.

To conduct this study, passenger data was collected from Ethiopian Airlines' customer survey system, resulting in a dataset of 20,000 records between January 2022 and January 2023. Each response includes ratings on service quality, delays, staff interaction, and overall satisfaction, along with demographic details such as age and travel class. Data was gathered via both digital and paper surveys distributed at various touchpoints, including check-in, baggage claim, and post-flight follow-up. The dataset's key attributes include passenger ratings, delay durations, demographic variables, and service-related feedback, providing comprehensive metadata for analyzing passenger satisfaction

After acquiring the data, several preprocessing steps were applied to ensure data quality and compatibility with machine learning algorithms. Subsequently, we selected machine learning models that aligned with the objectives of the study and trained them to uncover the key factors influencing passenger satisfaction. The models were also used to evaluate the predictive performance of different algorithms. The dataset was split into 80% for training and 20% for testing. Finally, the trained models were evaluated to compare their accuracy and to identify the most influential features affecting passenger satisfaction. Based on the experiments conducted, ensemble models particularly XGBoost and Random Forest demonstrated the highest performance, achieving approximately 89% accuracy, meanwhile the deep learning algorithm (MLP, TabNet and DCN) achieves 85% accuracy in predicting passenger satisfaction. Ensemble models like XGBoost and Random Forest circumvented some of the common challenges faced by deep learning approaches such as data requirements, complexity, and interpretability thereby achieving higher predictive accuracy and providing more transparent insights into passenger satisfaction drivers. The dataset used in this study contained 20,000 records with 22 independent variables. Such data is sufficiently rich for ensemble models to learn complex non-linear relationships but not large enough to fully exploit the representational capacity of deep learning models (MLP, TabNet, and DCN) lacked sufficient data to optimize their numerous parameters, leading to slightly lower performance compared to ensemble methods.

The most significant features found to be Type of Travel, Class, Customer Type, and Total Delay. This study confirms again that ensemble models like Random Forest and
XGBoost are very effective for classification issues. Features such as Type of Travel, Class, Customer Type, and Total Delay were more influential on passenger
satisfaction than traditionally assumed features such as Booking Service, Baggage Service, and Lounge Service. This
result highlights the strength of machine learning and explainable AI (XAI) techniques in uncovering hidden patterns and identifying the real drivers of customer satisfaction in big and complex datasets.

Keywords: *Passenger Satisfaction, Machine Learning, Deep Learning, XG Boost, Random Forest, Explainable AI.*

Table Of Contents

Declaration:.....I

Acknowledgement.....II

Abstract: V

1.1 Background 1

CHAPTER ONE 1

Introduction..... 1

1.2 Statement of the Problem 2

1.3 Research Questions 2

1.4 . Objectives 3

1.4.1 General Objective 3

1.4.2 Specific Objective..... 3

1.5. Significance of Study. 3

1.6 . Scope of the Study. 4

1.7 Limitation of the study..... 5

1.8 Organization of the thesis document..... 6

CHAPTER TWO: 7

Literature Review and Related Work 7

2.1 Overview..... 7

2.1.1Ethiopian airlines and Its Passengers 7

2.1.2.User satisfaction in Aviation Industry. 8

2.2 Machine Learning..... 8

2.2.1 Types of machine learning 8

2.2.3 Overview of Machine Learning Algorithms(Models) Explored in the Study 11

2.3 Predictive modeling..... 12

2.3.1 Applications of Machine Learning in the Aviation Industry. 12

2.3.2 Challenges in Applying Machine Learning in Aviation. 13

2.4.Related Works..... 15

CHAPTER THREE: 19

Methodology of the Study..... 19

3.1 Research Design. 19

3.2 Data Acquisition 19

3.3 Data Processing 20
3.3.1 Understanding the Dataset 20
3.3.2 Exploratory Data Analysis 22
3.3.3 Handling outliers 23
3.3.4 Data type Conversion 24
3.3.5 Encoding Categorical Variables 25
3.3.7 Feature Selection 26
3.3.8 Balancing the Dataset 27
3.3.9 Train Test Split 28
3.4 Development Tools and Computational Resources 29
3.5 Model Building 30
3.5.2 Model training 30
3.5.3 Hyperparameter tuning 32
3.6 Model Evaluation 33
3.7 Model Interpretability 33
Findings, Model Build, and Discussions of the Study 35
4.1 Model Build 35
4.2 Findings of the Study 37
4.2.2.2 Performance of Decision Tree Algorithm 45
4.2.2.3 Performance of Random Forest Algorithm 48
4.3 Discussion of the Study 66
4.3.1 Discussion: Summary of Key Results 66
CHAPTER FIVE: 73
Conclusion and Recommendation 73
5.1 Conclusion of the Study 73
5.2 Recommendations for Further Studies 74
Reference. 75
Appendices. I

CHAPTER ONE
Introduction

1.1 Background

Passenger satisfaction is the biggest component of the aviation industry, contributing significantly to customer loyalty, brand reputation, and business success in general. With a more competitive global market, airlines aim to make passenger experience comprehension and optimization a high priority to retain customers and continue growing [1].

Ethiopian Airlines, established in 1945, is Africa's biggest leading carrier and one of the continent's most successful and fastest-growing airlines [2]. With a vast international network spanning over 125 destinations across five continents, the airline has built a strong reputation for its commitment to safety, reliability, and outstanding customer service [3]. And it has made significant investments in modernizing its transportation, introducing the newest aircraft like the Boeing 787 Dreamliner and Airbus A350 and in enhancing operational efficiency, earning numerous awards and global recognition [4].

Regardless of this achievement, Ethiopian Airlines faces less competition from both regional and global carriers [5]. At the same time, passenger expectations are coming, influenced by digital transformation and the increasing availability of at a time service and feedback platforms [6]. To maintain its growth and leadership, the airline must adopt innovative, the emerged strategies to gain hidden insights into customer satisfaction and drive continuous improvement [7].

Machine learning has powerful tools to process and analyze large amounts of customer feedback data [8]. These technologies enable predictive modeling that can uncover hidden variables of satisfaction and help personalize services to meet varying passenger needs [9]. However, Ethiopian Airlines has till have not to fully integrate these advanced capabilities into its service optimization strategies

To better understand what truly shapes passenger satisfaction, this study introduces classical machine learning (ML) and deep learning (DL) models specifically designed for Ethiopian Airlines. By combining proven ML methods with cutting-edge DL techniques, the goal is to not only predict how satisfied passengers are but also uncover the key factors that impact their overall travel experience.

1.2 Statement of the Problem

Passenger satisfaction plays a key role in the success and competitiveness of airlines [10]. As the largest airline in Africa and a strong player in the global aviation market, Ethiopian Airlines recognizes that keeping passengers satisfied is essential for long-term growth, building brand loyalty, and staying competitive. While the airline actively gathers feedback through surveys and reviews, turning that feedback into meaningful, actionable insights remains a challenge [11]. Currently, Ethiopian Airlines mainly uses simple statistical method to analyze passenger satisfaction, these methods often assume simple, linear relationships between factors [12]. It emphasizes that existing satisfaction measurement approaches primarily depend on qualitative assessments and generic survey tools. These methods fail to provide real-time, actionable insights necessary for proactive service improvements.

As a result, while they can uncover hidden patterns and relationships within the data ,it becomes more challenging to clearly understand the true trends in the customer .These days, machine learning (ML) is proving to be a powerful tool for industries like aviation, helping them make sense of large and complex amounts of data [13]. ML can discover hidden patterns that are not easy to find using statistical methods. It can also provide accurate predictions of how satisfied passengers are, along with explanations of the key factors influencing their satisfaction [14]. Although ML give many benefits, Ethiopian Airlines has not still fully used these advanced tools [15]. Without a prediction system to predict satisfaction and find important factors, the airline may miss chances to improve services [16]. This could lead to dissatisfied customers and negative reviews, which in turn may harm the airline's reputation in the global market [17]. To address this problem, this study aims to apply machine learning techniques both classical and deep learning approaches to predict passenger satisfaction and identify its key hidden factors. To comparing the performance of multiple ML and deep learning models, the research aims to find the most effective and interpretable approach for Ethiopian Airlines. This will support more informed decisions, better service strategies, and a stronger competitive edge in the global airline industry.

1.3 Research Questions

To this end this study attempts to address the following research questions:

- What are the factors influencing passenger satisfaction in Ethiopian Airlines?
- Which machine learning algorithms offer the most accurate prediction for passenger satisfaction in Ethiopian Airlines?

1.4 . Objectives

1.4.1 General Objective

The general objective of this study is to design passenger satisfaction prediction model for Ethiopian Airline utilizing machine learning algorithms.

1.4.2 Specific Objective

- To identify the key passenger satisfaction factors influencing Ethiopian Airlines' service quality.
- To design machine learning models capable of accurately predicting passenger satisfaction levels.
- To compare the performance of various machine learning algorithms based on relevant metrics such as accuracy, precision, recall, and F1-score.
- To provide actionable insight for Ethiopian Airlines based on model insights.

1.5. Significance of Study.

This study holds strong significance for Ethiopian Airlines and its stakeholders by introducing advanced data-driven approaches namely machine learning and Explainable Artificial Intelligence (XAI) to better understand and predict passenger satisfaction. As the national flag carrier and the largest airline in Africa, Ethiopian Airlines plays a vital role in shaping the image of Ethiopia’s aviation sector on the global stage. By identifying hidden factors that influence customer satisfaction, the study will enable the airline to improve service quality, enhance operational efficiency, and remain

competitive both regionally and internationally. Passengers, especially those traveling to and from Ethiopia, will benefit from better.

Additionally, the findings of this research provide valuable insights for decision makers within the airline, supporting evidence-based policy and strategic planning that aligns with customer expectations. Researchers and academics can also benefit by exploring the practical application of machine learning and XAI within the aviation industry, which remains a growing field of interest. The study contributes to academic knowledge and can serve as a reference for future research in customer satisfaction, transport service improvement, and data science applications. Overall, the research supports Ethiopian Airlines in offering world-class service while also strengthening Ethiopia's position in the global aviation landscape.

1.6. Scope of the Study.

This study aims to analyze and predict passenger satisfaction with Ethiopian Airlines by applying machine learning algorithms. The research draws on data gathered directly from Ethiopian Airlines' own structured passenger satisfaction surveys, collected between Jan 2022 to Jan 2023.These surveys asked passengers to rate key aspects of their travel experience, including cabin comfort, the ease of booking, baggage handling efficiency, punctuality, and overall satisfaction. In addition to service-related feedback, the surveys also captured basic demographic details such as passengers' age and travel class, helping to contextualize responses. In total, the dataset includes roughly 20,000 complete responses, gathered through both digital and paper forms distributed at various stages of the passenger journey, such as check-in counters, baggage service, and through post-flight follow-up communications. The questions were designed to reflect passengers' real experiences and perceptions, covering critical areas like delays, service quality, and staff interaction. This study is limited to analyzing that specific set of survey responses. It does not include real-time operational data or feedback from other airlines. As such, while the findings offer valuable insights into Ethiopian Airlines' passenger satisfaction trends, they may not directly translate to other carriers without further validation and adaptation.

1.7 Limitation of the study

The dataset used was imbalanced, with about 75% of records representing satisfied passengers and only 25% dissatisfied ones, which could bias the model toward the majority class. In addition, the scope of the data was limited mainly to structured feedback leaving out more personal factors such as individual expectations, cultural backgrounds, or unique in-flight experiences that influence satisfaction. While explainable AI methods like SHAP and LIME were applied, some models, particularly ensemble methods such as XGBoost and Random Forest, still acted as "black boxes," making their decision processes harder to interpret. Although deep learning models like MLP, TabNet, and DCN performed well, their effectiveness may have been restricted by the relatively small size and narrow nature of the dataset, which limited their ability to capture deeper patterns.

1.8 Organization of the thesis document.

This thesis is structured into five chapters, each addressing a key part of the study on predicting passenger satisfaction for Ethiopian Airlines using machine learning. Chapter One introduces the background, problem statement, objectives, significance, scope, and thesis structure. Chapter Two reviews related literature on machine learning, predictive analytics, and passenger satisfaction, highlighting research gaps. Chapter Three outlines the methodology, including data collection, preprocessing, feature selection, and model development. Chapter Four presents and evaluates the experimental results, identifying important satisfaction factors. Chapter Five summarizes the findings, discusses implications, and provides recommendations. References and Appendices are included for sources and supplementary materials.

CHAPTER TWO:
Literature Review and Related Work

2.1 Overview

In the aviation sector, the integration of Machine Learning offers groundbreaking opportunities to address challenges such as enhancing passenger satisfaction, optimizing operations, and providing personalized services. Ethiopian Airlines, as a global leader and a national flag carrier, stands at the forefront of leveraging these advancements. This literature review discusses the core concepts of ML, its key components, learning types, practical applications, and associated challenges, providing a comprehensive understanding of its transformative potential in addressing critical aspects of the aviation industry [18].

2.1.1Ethiopian airlines and Its Passengers

Passenger satisfaction refers to the overall feeling of gratification of passengers with the services provided by airlines right from the time of booking to arrival. It encompasses multiple dimensions, such as service quality, time factor, flying experience, and affordability of price. Certainly, the high magnitude of passenger satisfaction is highly decisive for airlines as it has a direct consequence on customer loyalty, corporate brand image, and profitability. In the competitive aviation market, it is very important for airlines to understand those factors that contribute to passenger satisfaction in order to retain customers and improve their service offering [19].

Ethiopian Airlines, the flag carrier of Ethiopia, was founded in 1945. It is one of the largest and most successful African airlines, with its headquarters in Addis Ababa. The carrier operates an extensive domestic and international network that reaches more than 125 destinations across five continents. It has been recognized for safety, reliability, and good customer service; thus, it has been the choice of many passengers [20]. It has heavily invested in fleet modernization and service enhancement through the introduction of new aircraft types such as the Boeing 787 Dreamliner and Airbus A350. Ethiopian Airlines has also been awarded various times for operational excellence and customer service, among other accolades that have positioned it as the leader in African aviation. However, against the backdrop

of all these successes, the airline still must work harder to meet diverse passenger expectations and ensure satisfactory service quality amidst rising competition [21].

2.1.2.User satisfaction in Aviation Industry.

User satisfaction is all about the feelings of customers towards their expectations and needs, which have been met by a certain service or product. Passenger satisfaction in the aviation industry, especially for Ethiopian Airlines, involves service quality, punctual performance, comfort during flights, and value for money. Measurements of user satisfaction involve a number of methods, including post-flight surveys of specific attributes, such as check-in processes and cabin crew service; rating scales, including Likert scales; and the Net Promoter Score, which asks for customer loyalty through their likelihood of recommending the airline as on-time performance and baggage handling efficiency, and machine learning techniques that analyze large datasets for pattern recognition and sentiment analysis. Focus groups and interviews are more qualitative methods that also yield rich insights into passenger expectations. The data indicates an analysis of passenger satisfaction dynamics that enables Ethiopian Airlines to implement data-driven strategies, improving customer experiences [22].

2.2 Machine Learning.

Machine learning is a part of artificial intelligence that allows computers to learn from data and get better at certain tasks over time without being directly programmed for every step. Instead of relying on fixed rules, machine learning models look for patterns in the data and use those patterns to make predictions or decisions [23].

2.2.1 Types of machine learning

Supervised Learning.

Supervised learning is one of the most common types of machine learning. In this approach, algorithms learn from structured, labeled data meaning the system is trained using examples that include both the input and the correct output. This helps the model learn how to make accurate predictions for new, unseen data. Supervised learning is generally divided into two main types: classification, which

predicts categories (like identifying whether an email is spam or not), and regression, which predicts continuous values (like estimating stock prices or housing costs).

While supervised learning is known for its accuracy and easy-to-understand results, it does come with challenges. It needs large amounts of high-quality labeled data, which can be expensive and time-consuming to collect. It also faces risks like overfitting where the model performs well on training data but poorly on new data which requires careful techniques to prevent. Despite these hurdles, supervised learning is widely used in real-world applications like speech recognition, financial forecasting, and medical diagnostics [24] .

Unsupervised Learning.

Unsupervised learning deals with unlabeled or unstructured data and focuses on identifying hidden patterns or structures within datasets. Tasks like clustering and dimensionality reduction fall under this category. For example, clustering might involve segmenting customers based on purchasing habits, while dimensionality reduction simplifies large datasets for visualization.

The strength of unsupervised learning is in its ability to analyze data without requiring labels, making it capable for exploratory analysis. but tasks like determining the number of clusters demand domain expertise. It is used extensively in anomaly detection, customer segmentation, and market basket analysis [25].

Semi-Supervised Learning.

Semi-supervised learning is the combination of supervised and unsupervised learning by using a small amount of labeled data along with a large pool of unlabeled data. This approach is particularly more significant when acquiring labeled data is challenging or costly.

This hybrid approach is used in scenarios like image classification, where only a few labeled images available, or autonomous vehicle training, where labeled navigation data is supplemented with vast amounts of unlabeled information. Despite reducing the dependency on labeled data, semi-supervised learning faces challenges in effectively utilizing unlabeled data without introducing noise, and the quality of labeled data heavily impacts performance [26].

Reinforcement Learning

Reinforcement Learning the type of machine learning where the paradigm in which an agent learns to make decisions through interacting with an environment. The agent receives feedback in the form of rewards or penalties based on its actions and uses this feedback to optimize long-term cumulative rewards. RL is particularly effective for real-time decision-making in dynamic environments, such as robotics and autonomous driving. Crafting a reward system that aligns with desired goals is also a complex task. Nevertheless, RL has powered breakthroughs in game AI, such as AlphaGo, and applications in personalized recommendations and robotic automation [27].

Deep Learning.

Deep learning is a subset of machine learning that employs neural networks with multiple layers to model complex patterns in data. It has revolutionized industries to enabling breakthroughs in computer vision, natural language processing (NLP), and speech recognition. For example, convolutional neural networks (CNNs) are used for image classification and facial recognition, while recurrent neural networks (RNNs) or Long Short-Term Memory (LSTM) networks power tasks like machine translation and sentiment analysis. Although deep learning offers incredible accuracy and scalability, it demands substantial computational power and data. Moreover, its "black box" nature makes it challenging to interpret, especially in critical applications like healthcare. Despite these limitations, deep learning has made transformative contributions to fields like autonomous systems, generative models, and AI-driven content creation [28].

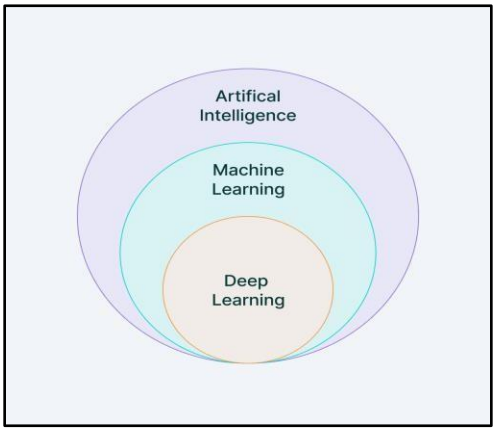


Figure 1 Artificial intelligence types and their relationship [29]**.**

2.2.3 Overview of Machine Learning Algorithms (Models) Explored in the Study

Logistic Regression: A statistical model used mainly for binary classification tasks. It estimates the probability that a given input belongs to a particular class by applying the logistic function. It is simple, interpretable, and effective for linearly separable data but less capable of modeling complex, nonlinear relationships [30].

Decision Tree: A tree-structured model that makes predictions by splitting the data based on features, leading to decision paths from root to leaf nodes. They are intuitive, easy to interpret, and handle both classification and regression tasks, but can overfit if not properly pruned [31].

Random Fores: An ensemble learning method that constructs multiple decision trees during training and aggregates their predictions (via majority voting or averaging). It improves accuracy over single trees, reduces overfitting, and is robust to noise [32].

Support Vector Machines (SVM) : A classifier that finds the hyperplane maximizing the margin between data points of different classes. Powerful in high-dimensional spaces and effective with clear class boundaries, but computationally intensive with large datasets [33].

XGBoost : An advanced gradient boosting technique that builds sequential models where each subsequent model corrects the errors of previous ones. Known for high performance and efficiency, especially in structured/tabular data, and often used in machine learning competitions [34].

Multilayer Perceptron (MLP): A type of deep neural network composed of multiple interconnected layers of neurons, capable of modeling complex nonlinear relationships. Suitable for a variety of tasks but require careful tuning and more computational resources [35].

Tab Net : A deep learning architecture designed for tabular data that uses sequential attention mechanisms to select relevant features at each decision step. It combines interpretability with high performance on structured datasets [36].

Deep & Cross Network (DCN) : A neural network architecture that combines deep neural networks with cross layers to explicitly model feature interactions. It is effective for recommendation systems and structured data where feature interactions are crucial [37].

2.3 Predictive modeling

Predictive modeling is a statistical method of using historical data to predict future outcomes. In passenger satisfaction, predictive modeling will analyze various factors that influence customer experience and satisfaction levels in the aviation sector. Using various algorithms, predictive models can identify patterns and relationships in data, thus helping airlines anticipate passenger needs and preferences. This will also help the airlines to proactively work toward betterment regarding the quality of services and operational efficiency, thus increasing customer satisfaction. Usual techniques applied in predictive modeling include regression analysis, decision trees, and ensemble methods that can handle such complex datasets with valuable insights on customer behavior [38].

2.3.1 Applications of Machine Learning in the Aviation Industry.

Machine learning (ML) has become a transformative tool across the aviation sector, driving innovation and efficiency in multiple domains. Its ability to analyze large, complex datasets enables airlines and related stakeholders to enhance service quality, streamline operations, and improve safety. Below are some key applications [39].

Passenger Satisfaction Prediction and Personalization.

ML models analyze customer feedback, survey responses, and social media data to identify factors influencing passenger satisfaction. This analysis helps airlines understand passenger preferences and expectations, enabling personalized services such as tailored offers, customized in-flight entertainment, and route suggestions. For example, sentiment analysis tools using natural language processing (NLP) can gauge passenger sentiment from unstructured feedback, providing actionable insights for service improvement [40].

Operational Efficiency and Flight Scheduling.

ML algorithms optimize scheduling of aircraft and crews by predicting maintenance needs, minimizing turnaround times, and coordinating flight schedules based on historical and real-time data. Predictive maintenance models can forecast equipment failures before they occur, reducing delays and cancellations. Additionally, ML-driven demand forecasting allows airlines to adjust capacity dynamically, reducing costs and enhancing the passenger experience [41].

Baggage Handling and Safety Systems.

Advanced ML systems assist in automating baggage handling through image recognition and sensor data analysis, reducing mishandling and theft. In safety-critical systems, ML models monitor aircraft sensor

data to detect anomalies, predict potential malfunctions, and support decision-making during flight operations. For instance, machine learning can aid in early detection of sensor failures or structural issues, contributing to safer flights [42].

Flight Path Optimization and Weather Prediction.

ML enhances route planning by analyzing various factors, including weather conditions, air traffic, and geopolitical situations, to determine the most efficient flight paths. For example, real-time weather data fed into predictive models help pilots and dispatchers choose optimal routes, saving fuel and reducing emissions [43].

2.3.2 Challenges in Applying Machine Learning in Aviation.

Despite the significant potential of machine learning (ML) to revolutionize various aspects of aviation, its practical implementation encounters several substantial challenges that need to be carefully addressed.

Data Integration and Diversity Aviation systems generate large volumes of data from many different sources such as aircraft sensors, airport systems, maintenance logs, passenger feedback, social media, weather forecasts, and even external services like air traffic control. Bringing all this diverse data together into one unified, usable system is a major challenge. It requires sophisticated strategies like data warehousing, standardization, and synchronization to make sure everything works together smoothly. If these sources aren't properly integrated, the resulting data can be inconsistent or incomplete, which ultimately reduces the accuracy and reliability of machine learning models [44].

Data Quality and Bias The performance of machine learning models is only as good as the data they are trained on. In aviation, this data can sometimes be outdated, incomplete, noisy, or biased, for example, certain passenger groups or flight routes might be underrepresented. If the training data isn't high quality or balanced, the models may make inaccurate predictions or show unfair preferences. To ensure dependable results, it's essential to regularly clean, validate, and monitor data for potential bias [45].

Model Interpretability and Safety Aviation is a highly regulated, safety-critical field where decisions must be transparent and explainable. While complex ML models, especially deep learning models can make powerful predictions, they often work like "black boxes," offering little insight into how or why a particular outcome was reached. This lack of transparency can make it difficult for stakeholder-like pilots, engineers, and regulators to trust or act on the results. That's why using Explainable AI (XAI) is so important. It helps provide clear, understandable reasoning behind the model's decisions, which supports both safety and regulatory acceptance [46].

Regulatory and Certification Barriers The aviation industry is tightly regulated by authorities such as the FAA and EASA, which require that all operational systems including those powered by AI go through

rigorous testing, documentation, and certification. ML models must prove they are reliable and consistent, even in changing or unexpected conditions. However, because these models can adapt over time as they learn from new data, ensuring they continue to meet strict regulatory standards is a significant challenge [47].

Real-time Processing Demands Some ML applications in aviation like optimizing flight paths or monitoring aircraft in real-time require instant data analysis and decision-making. These tasks demand high computational power and extremely reliable systems. Meeting these real-time demands means investing in powerful hardware, efficient algorithms, and robust system design to ensure quick, safe, and dependable responses [48].

2.4. Related Works

Demeke and Yilma (2021) took a closer look at how passengers feel about their experiences at Ethiopian airports by using natural language processing (NLP) to analyze unstructured feedback. Their study uncovered regional variations in what travelers expect, highlighting the importance of tailoring services to local needs. They emphasized that incorporating insights from machine learning into daily operations could go a long way in boosting customer loyalty.

Abebe and Melesse (2019) carried out a comprehensive study on customer satisfaction and service quality at Ethiopian Airlines, shedding light on the distinct operational challenges the airline faces within its regional setting. Although the airline has achieved considerable growth, the study found that customer perceptions are significantly shaped by infrastructural constraints, inconsistent service delivery, and cultural dynamics that diverge from international standards. The researchers noted that conventional tools for measuring customer satisfaction often fall short in capturing the complex interplay of these factors, resulting in an incomplete understanding of passenger expectations. Furthermore, they criticized the overreliance on qualitative feedback and broad quantitative surveys in existing Ethiopian customer satisfaction models, arguing that these methods lack both predictive accuracy and the ability to generate timely, actionable insights. Abebe and Melesse recommended the use of advanced data analytics, including machine learning, to gain a deeper, more predictive grasp of passenger needs. This would allow the airline to personalize services and respond more effectively to operational challenges. Their findings highlight the urgent need for context-aware predictive models that account for Ethiopia's unique socio-economic and infrastructural realities an area largely overlooked by existing global research on the airline industry.

Sachin Kumar & Mikhail Zymbler(2023) conducted their study on the analysis of customer satisfaction in airline companies based on a machine learning methodology applied to 146,731 tweets of major airlines. Following data preprocessing and feature extraction via n-gram models and GloVe word embeddings, the researchers employed Support Vector Machines (SVM), Artificial Neural Networks (ANN), and Convolutional Neural Networks (CNN) to classify tweets as either positive or negative in sentiment. The CNN model worked better with 92.3% accuracy, as opposed to SVM and ANN approaches. Through association rule mining, the study identified cabin crew behavior (CCB) and food quality (FQL) as the most significant determinants of passenger sentiment. The findings indicate how airlines can utilize machine learning in monitoring customer feedback and strategically enhancing service quality by addressing priority experience drivers.

Hwanga, Kima, Parka & Kwonb (2020) tried to answer the following question: "Can users' affective expressions on airline services be useful in estimating their return visits to the services?" It is not mandatory at all that if we find positive feedback about an airline that the customer is going to come back. They conducted a survey to check feedback from people who booked their flights online, they collected 309,331 customer's feedback. They validated this feedback and dropped invalid ones. In Addition, dropped the responses of delayed or cancelled flights. Based on these procedures, 178,951responses (57.9%) were validated. One year later, in November 2018, the customers with validated responses were asked if they used airline services for the past year. "The customers who responded positively (133,872 responses, 43.3%) were subsequently asked if they used the same airline service as stated in the initial survey. Then they performed several machine learning classifiers, The seven classifiers, Decision Tree, Gaussian Naive Bayes, Logistic Regression, Random Forest, KNeighbors, Support Vector Machine, and XGBoost."The study found out that both emotional and linguistic aspects of users' comments on a particular service are valuable when predicting and determining their future behavior related to the service". "In addition, the results of predicting return visits showed relatively high accuracy regardless of the length of reviews".

Table 1 The summary of related works.

Author (year)	*Title*	*Objective*	*Key findings*	*Remark*
Ashwika et al. (2022)	*Airline Passenger Satisfaction Prediction Using Machine Learning Algorithms*	*To predict passenger satisfaction using ML, handling imbalanced data and high-dimensional features.*	*imbalanced data and high-dimensional features.Random Forest achieved 94.71% accuracy; key factors: online boarding, WiFi, class, legroom, entertainment.*	Demonstrates the value of ensemble ML models for predictive accuracy in customer satisfaction. Applicability to different regions may be limited due to dataset scope.
Sachin Kumar & Mikhail Zymbler	*Machine Learning Approach to Analyze Customer Satisfaction from Airline Tweets*	*To classify airline-related tweets as positive/negative using ML and NLP.*	*CNN achieved 92.3% accuracy; key factors: cabin crew behavior (CCB) and food quality (FQL).*	*Strong focus on NLP; could expand to other social media platforms.*
Hwanga, Kima, Parka & Kwonb (2020)	*Can users' affective expressions on airline services be useful in estimating their return visits?*	*To predict return visits based on customer feedback using ML.*	*Emotional and linguistic aspects in reviews help predict return visits (XG Boost likely best but not specified).*	*High accuracy regardless of review length; dataset limited to non-delayed flights.*
Abebe & Meleesse (2019)	*Customer Satisfaction and Service Quality in Ethiopian Airlines: Challenges and Analytical Gaps*	*To examine customer satisfaction and service quality at Ethiopian Airlines within the regional and infrastructural context.*	*Identified major influencing factors such as infrastructure limitations, inconsistent service, and cultural gaps. Highlighted the inadequacy of traditional feedback tools and recommended machine learning for deeper, predictive insights.*	*Emphasizes the need for context-sensitive, predictive customer satisfaction models customized to Ethiopia's unique operational realities.*

Demeke & Yilma (2021)	*Analyzing Passenger Feedback at Ethiopian Airports Using NLP Techniques*	*To explore passenger sentiment and expectations at Ethiopian airports through unstructured feedback analysis using natural language processing.*	*Revealed regional differences in passenger expectations highlighted the importance of localized service delivery. Recommended integrating machine learning into daily operations to improve customer loyalty.*	*Demonstrates the potential of NLP and ML in extracting actionable insights from unstructured data to guide customer-centric service strategies.*

CHAPTER THREE:

Methodology of the Study

Research methodology is a way to systematically solve the research problem. It may be understood as a science of studying how research is done scientifically. The advantage of knowing the methodology of the study before doing the Experiment is to reason out what, how and why the methods or the techniques are selected for the Experiment to know the risks for conducting the research in detail [49].

3.1 Research Design.

This research follows quantitative/experimental research design to answer research questions. This method is ideal for using historical data to predict passenger satisfaction from various service features. applying and evaluating different machine learning algorithms, we'll not only achieve strong forecasts but also pinpoint the main drivers of satisfaction, giving Ethiopian Airlines practical insights for strategic decisions [50]

The experimental design incorporated a structured methodology: data collection and preprocessing, dataset partitioning into training and testing sets, algorithm and tool selection, and model development and evaluation. This chapter mentions the methodology used to develop a machine learning-based passenger satisfaction prediction model and factors that contribute more to passenger satisfaction.

3.2 Data Acquisition

The data used in this research was collected from Ethiopian Airlines passenger data that are collected passenger feedback through surveys, to address the research objectives. The data collection process was facilitated through an official support letter provided from Addis Ababa University, School of Information Science, and Ethiopian airline granting us to access required data for the study. This approach allowed us to work with real-world feedback data that accurately reflects the experiences and feedback of Ethiopian Airlines passengers.

The total dataset used for the research is 20,000 and the dataset, was collected from Ethiopian Airlines' customer survey systems of year 2022. The dataset includes both numerical features (e.g., age, delays) and categorical/ordinal features (e.g., service ratings, satisfaction). The target Variable or independent variable is The Satisfaction column serves as the label/output for predictive modeling. It likely contains discrete classes (e.g., Satisfied, Neutral, Dissatisfied), indicating that classification is the right approach. The Dependent features is Many service-related variables (e.g., Seat Comfort, Boarding services, Baggage service, Check-in Service, call center service) are ordinal, meaning their values have a meaningful order. Total columns (features): 23Independent (input) variables: 22 (all service ratings, demographic info, delays, class, etc.) Key features such as Type of Travel, Class, Customer Type, Total Delay.

The table below shows the meta data description

Column Name	Data Type	Description
Age	Numeric	Age of the passenger in years
Gender	Categorical	Gender of the passenger (e.g., Male, Female)
Satisfaction	Categorical	Passenger satisfaction level (e.g., Satisfied, Neutral, Dissatisfied)
Flight Distance	Numeric	Distance of the flight in miles or kilometers
Departure Delay	Numeric	Delay at departure (in minutes)
Arrival Delay	Numeric	Delay at arrival (in minutes)

Table 2: Description of meta data

3.3 Data Processing

The raw data collected from the survey, originally stored in Excel format, was converted to CSV to facilitate further processing. To address differences in feature magnitudes, the Standard Scaler was applied for normalization. Additionally, machine learning algorithms were employed to determine the most influential factors contributing to passenger satisfaction, which were then used to train the models. The following sections take a closer look at the key preprocessing steps involved in this study.

3.3.1 Understanding the Dataset

In this study, we began by getting a general overview of the dataset, checking the number of rows and columns and reviewing the data types for each feature to understand whether they were numerical or categorical. This helped ensure that we interpreted the data correctly from the start. We then identified the target variable, Satisfaction, and examined its distribution to see if there was any class imbalance an important step when choosing the right machine learning techniques for training. To get a clearer picture of the dataset, we looked at summary statistics like the mean, median, and standard deviation for numerical features, as well as frequency counts for categorical ones.

During this exploration, we also checked for common data quality issues such as missing values, outliers, and inconsistencies. These were flagged for cleaning in the next step.

To support this analysis, we used visual tools like histograms, box plots, and bar charts, which helped us better understand patterns, trends, and relationships within the data. The insights from this thorough exploration formed a solid foundation for the preprocessing phase that followed, ensuring the dataset

was well-prepared for machine learning modeling, ultimately leading to more accurate and interpretable results.

3.3.2 Exploratory Data Analysis

As part of our exploration data analysis, we used different visualizations to get a clearer picture of the dataset's structure and key characteristics. One important insight that stood out from these visualizations was a clear imbalance between the number of satisfied and dissatisfied passengers. As illustrated in Figure 2, approximately 75% of the data belongs to the satisfied class, while only 25% represents the dissatisfied class. This imbalance highlights the need for appropriate handling during model training to prevent bias toward the majority class. Without addressing this, the model may be skewed towards predicting the satisfied class more frequently. Another important observation, as shown in Figure 3, is that the service-related rating features (such as check-in Service, baggage service, and onboard services) indicates relatively uniform distribution across their respective categories, indicating balanced responses across these attributes.



Figure 2 Results of Exploratory Data Analysis (EDA): Customer Distribution

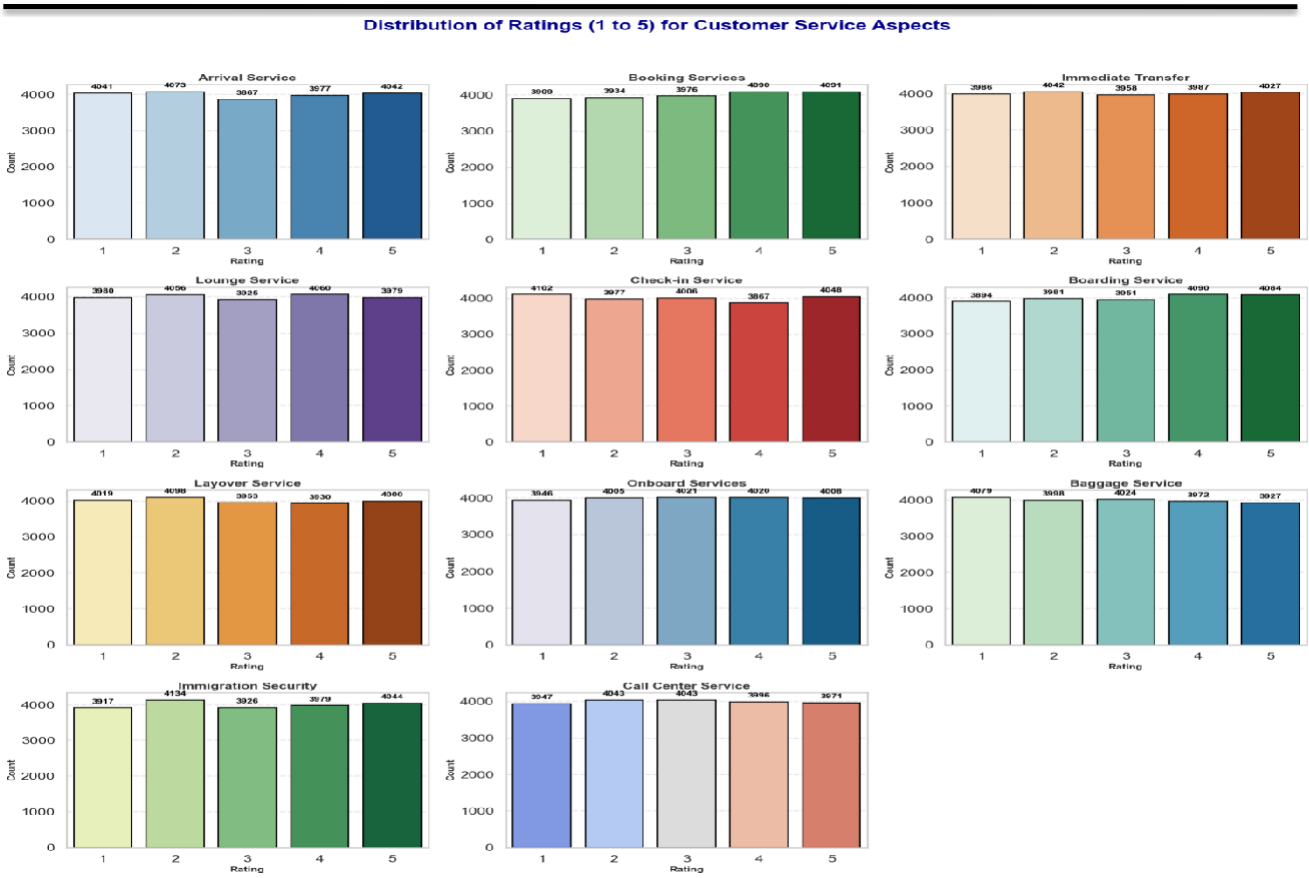


Figure 3 Results of Exploratory Data Analysis (EDA): Distribution of Rating of different Services

3.3.3 Handling outliers

In our study, outliers appear in numerical features “Age,” and “Arrival Delay” as show in the figure 4, an age entry of 350 and an Arrival of 2000 min could indicate data entry errors that may mislead the model. To identify outliers, statistical techniques like the Interquartile Range (IQR) method, Z-score, or visualization tools such as box plots and scatter plots are used in this study. We used log transformation to handle the effect of outliers on predicting contributing factors for passenger satisfaction. Properly handled outliers ensure that the machine learning model learns from genuine patterns in the data without being skewed by anomalies, ultimately leading to more accurate and generalizable predictions of passenger satisfaction.

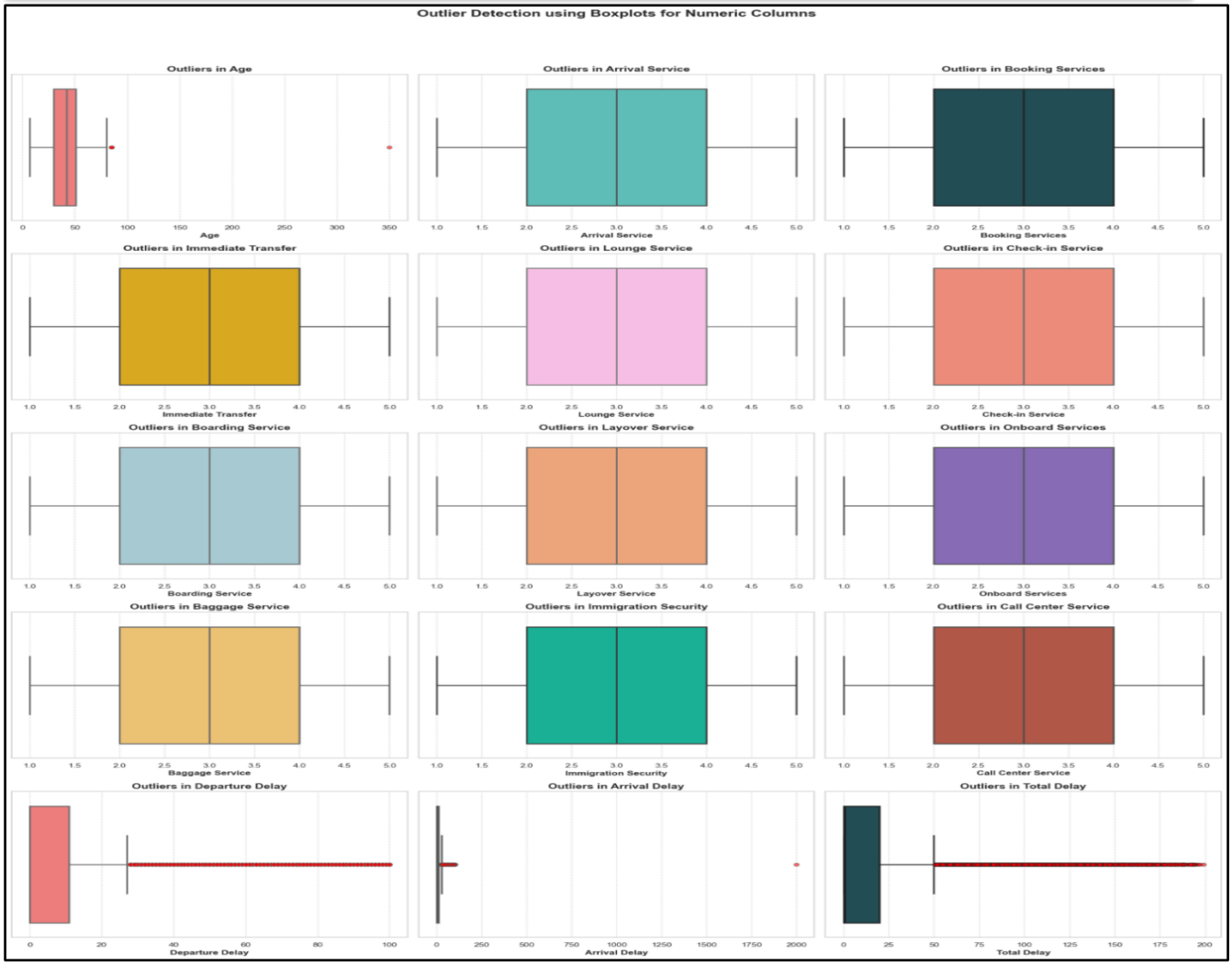


Figure 4: *Result of outlier detection using boxplot*

3.3.4 Data type Conversion

This step ensures that all variables are in appropriate formats that can be effectively interpreted by the models. In this study, both categorical and numerical features (such as age, arrival delay, or departure delay) and categorical variables (such as gender, class, type of travel, or customer type) are present. These variables initially appear in incorrect formats for instance, numerical values like Age and arrival delay stored as strings due to data entry inconsistencies, and thus we convert them to integer and float types for proper mathematical handling. Ensuring proper data type conversion improves memory efficiency, allows for smoother model integration, and prevents errors during model training, by enhancing the overall performance of the passenger satisfaction prediction system.

3.3.5 Encoding Categorical Variables

In preparing the passenger satisfaction dataset for machine learning, encoding categorical variables is an essential preprocessing step, since all machine learning algorithms require numerical input, these categorical variables must be converted into a suitable numerical format without losing the underlying meaning. In this study, we applied appropriate encoding techniques based on the nature of each feature. For nominal categories like Gender (e.g., Male, Female) and Type of Travel (e.g., Business, Personal), We used one-hot encoding to create binary columns representing the presence of each category, ensuring that no ordinal relationship is falsely implied. And Class (e.g., Economy, Business, Eco Plus), depending on whether the categories implied an order or not, we used one-hot encoding for both types of categories because we don't have categories that have an order. The results of our encoded variables are shown in figure 5 below.

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	rding Se	Layover Ser	Onboard Se	Baggage Ser	Immigratior	Call Center	Departure D	Total Delay	Arrival Delay_log	Gender_Male	Customer Type_disloyal Cus	Type of Travel_Personal Travel	Class_Eco	Class_Leisure	Satisfaction_Satisfied
2	.724658	1.41969	-0.713449	1.4260849	1.4114809	-3.55E-05	-0.512062	-0.535057	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
3	.432021	0.0072739	-0.004924	1.4260849	0.7039894	-1.419386	0.6550893	0.1520759	1.386294361	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
4	.017295	0.0072739	-0.004924	1.4260849	-0.710994	1.4193148	-0.512062	-0.535057	0	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE
5	900678	-0.698934	-0.713449	0.7188769	1.4114809	1.4193148	-0.512062	-0.535057	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
6	900678	1.41969	-1.421974	-0.695539	1.4114809	-0.709711	-0.512062	-0.535057	0	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
7	.017295	0.713482	0.7036007	1.4260849	0.7039894	0.7096396	-0.512062	-0.535057	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
8	.432021	1.41969	-0.004924	1.4260849	0.7039894	-1.419386	0.6550893	0.4163579	2.63905733	FALSE	FALSE	FALSE	TRUE	FALSE	TRUE
9	900678	-1.405142	1.4121257	1.4260849	1.4114809	-3.55E-05	-0.512062	-0.535057	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
10	.724658	1.41969	0.7036007	-1.402747	-0.710994	0.7096396	-0.512062	-0.535057	0	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
11	.017295	-0.698934	1.4121257	0.0116689	-0.003502	1.4193148	-0.512062	-0.535057	0	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
12	900678	1.41969	-0.004924	0.0116689	-1.418485	1.4193148	-0.512062	-0.323632	2.197224577	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
13	900678	-1.405142	-1.421974	0.0116689	-1.418485	-1.419386	-0.512062	-0.535057	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
14	974307	0.713482	-0.004924	1.4260849	1.4114809	-1.419386	1.6192582	2.1870474	4.127134385	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
15	.432021	0.0072739	-0.004924	-1.402747	-0.003502	-3.55E-05	-0.512062	-0.535057	0	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
16	.017295	-1.405142	1.4121257	-0.695539	0.7039894	-0.709711	-0.106097	-0.323632	0	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
17	900678	-1.405142	-1.421974	-0.695539	-1.418485	-3.55E-05	-0.512062	-0.535057	0	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
18	.017295	0.0072739	-0.713449	0.0116689	1.4114809	1.4193148	-0.410571	-0.35006	1.791759469	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE
19	.724658	1.41969	-1.421974	1.4260849	-0.710994	1.4193148	1.8729869	1.3413449	3.218875825	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE
20	974307	-0.698934	1.4121257	-0.695539	-0.003502	-3.55E-05	-0.512062	-0.085778	2.890371758	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
21	.432021	0.713482	-0.004924	0.7188769	0.7039894	0.7096396	0.4521064	0.2842169	2.564949357	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE
22	.432021	1.41969	-0.713449	-0.695539	-1.418485	-0.709711	0.2491235	2.3191884	4.543294782	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
23	.724658	0.0072739	0.7036007	1.4260849	-0.003502	-0.709711	-0.512062	-0.535057	0	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
24	900678	-0.698934	-1.421974	-0.695539	-0.003502	-1.419386	-0.512062	-0.535057	0	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE
25	974307	0.713482	1.4121257	1.4260849	-0.710994	-0.709711	-0.512062	-0.535057	0	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
26	.432021	-0.698934	0.7036007	-1.402747	-0.003502	-1.419386	-0.512062	-0.535057	0	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE
27	.017295	0.0072739	1.4121257	-0.695539	-0.710994	-0.709711	-0.461317	-0.508629	0	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE
28	.724658	-0.698934	-1.421974	-0.695539	-1.418485	-0.709711	-0.410571	-0.297204	2.079441542	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE
29	.432021	0.713482	-0.004924	0.0116689	1.4114809	1.4193148	-0.512062	-0.535057	0	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE

scaled_dataset +

Figure 5 Results of Categorical Data Encoding and Feature Scaling

3.3.6 Feature Scaling

In our study, features such as Age and Total Delay can vary widely in their magnitude values To address this, we applied standard feature scaling techniques such as Min-Max Scaling and Standardization (Z-score normalization). Min-Max Scaling transform features to a fixed range 0 to 1, while standardization centers the data around the mean with a unit standard deviation. These methods ensure that no single feature dominates the learning process due to its scale, thereby improving model performance and training stability.

After experimenting with both techniques, we select Standardization techniques based on the accuracy of the model. The result of scaled dataset is shown on figure 5 displayed above.

3.3.7 Feature Selection

In our study, we carefully examined the available features to identify those that contribute most significantly to predicting passenger satisfaction. The goal of feature selection is to reduce noise, avoid overfitting, and improve model interpretability by eliminating irrelevant or redundant variables. We used both statistical techniques and machine learning models to guide the selection process. For example, we analyzed the correlation between features and the target variable using correlation matrices and mutual information scores. Additionally, we applied model-based methods such as feature importance from tree-based algorithms like Random Forest to rank the most influential predictors.

Key features such as Type of Travel, Class Eco, Customer Type, Age, Class Leisure, Arrival Delay, Total Delay and Departure Delay were found to have a strong impact on passenger satisfaction and were retained for model training. On the other hand, features with low variance or weak correlation to the target variable were removed from the model training. This careful selection of relevant features not only helped in reducing computational complexity but also enhanced the model's ability to make accurate and generalizable predictions. Based on the experimental results, we identified the most influential features affecting the target variable, satisfaction, as illustrated in Figure 6 below.

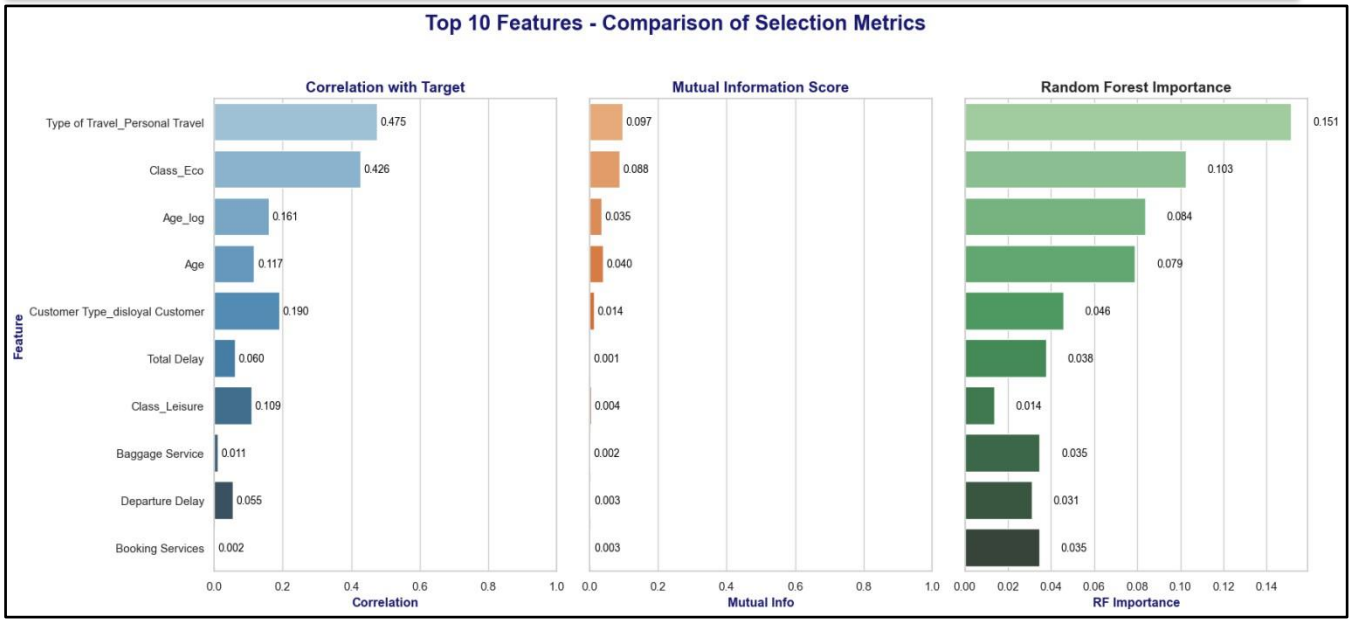


Figure 6 Most important selected feature and their Metrics Score

3.3.8 Balancing the Dataset

Balancing the dataset is one of the crucial steps in developing a fair and effective passenger satisfaction prediction model, especially when dealing with class imbalance. In our dataset of 20,000 records, there is a noticeable imbalance between the two target classes: 14,995 instances are labeled as Satisfied and 5,005 as Dissatisfied. This imbalance can lead to biased machine learning models that tend to favor the majority class (Satisfied), resulting in poor performance when predicting the minority class (Dissatisfied). Such a model might achieve high overall accuracy but fail to correctly identify dissatisfied passengers, which is critical for improving service quality and customer retention.

To address this issue, we implemented an oversampling technique using SMOTE (Synthetic Minority Over-Sampling Technique), which generates new synthetic samples for the Dissatisfied class to match the size of the Satisfied class. Additionally, we adjusted class weights in the algorithms to penalize misclassification of the minority class more heavily. For this study, we employed oversampling techniques and balanced class weights to train the models and evaluate their performance effectively. The statistical characteristics of the dataset, both before and after sampling, are presented in Figure 7 below, illustrating the distribution changes resulting from the sampling process.

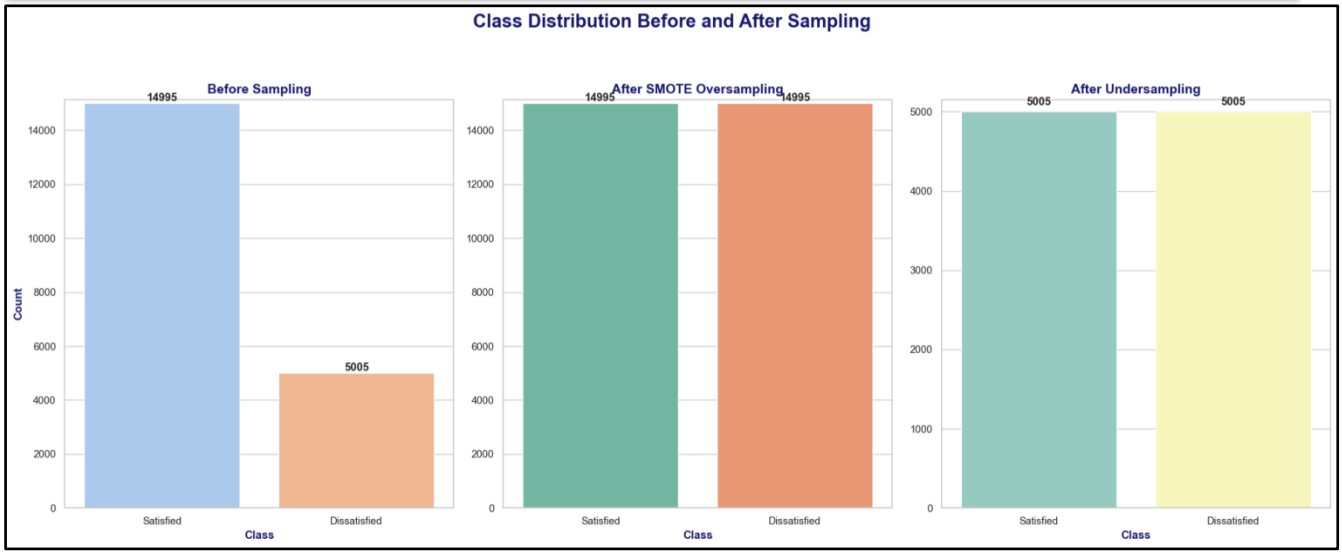


Figure 7 Class Distribution Before and After Sampling

3.3.9 Train Test Split

In this study, we performed a train-test split on the passenger satisfaction dataset to prepare it for building and evaluating the machine learning model. Out of the total 29,990 after oversampling, 80% of the data (23,992 records) was allocated for training, and the remaining 20% (5,998 records) was reserved for testing. This standard approach ensures that the model is trained on a substantial portion of the data, allowing it to learn underlying patterns and relationships effectively, while the separate test set is used to evaluate the model's performance on unseen dat. To make sure our prediction model is reliable and performs well on new data, we carefully split our dataset in a structured way. This helps the model learn patterns that truly reflect passenger satisfaction without just memorizing the training data. Figures 8 and 9 below illustrate how we divided the dataset, showing clearly which parts were.

Figure 8: Passenger Satisfaction Distribution Dataset Split Bar Chart.

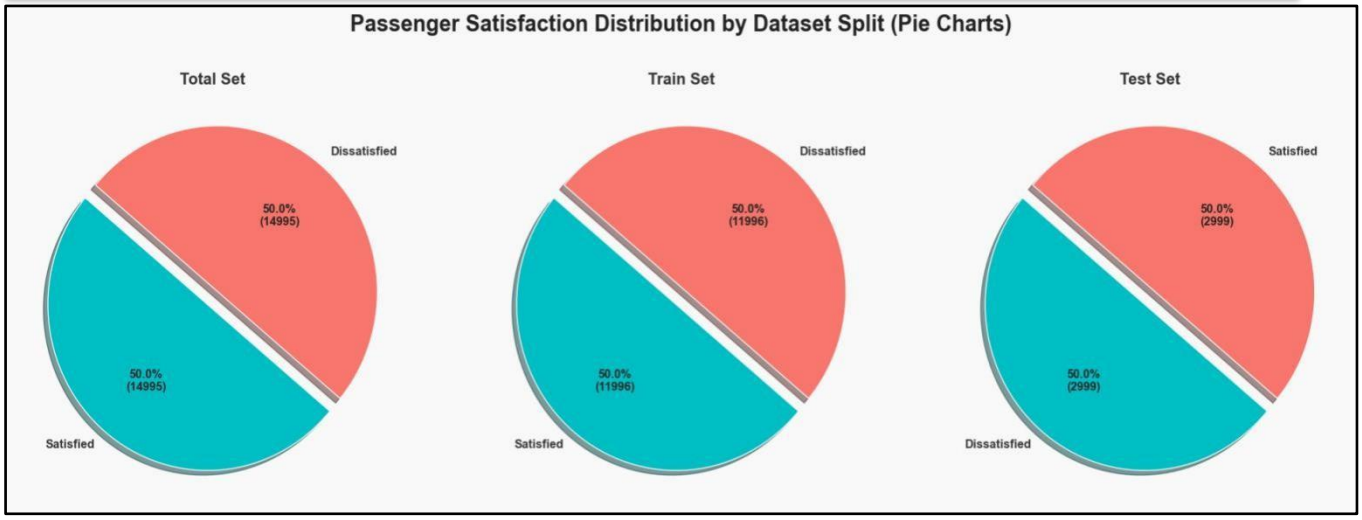


Figure 9: Passenger Satisfaction Distribution Dataset Split Pie Chart.

3.4 Development Tools and Computational Resources

To carry out the experiments in our study, we used on a mix of well-established tools and programming environments commonly used in the field. Python was our main programming language; chosen for its versatility and the rich ecosystem it offers for data science and machine learning. For handling and analyzing data, we made use of Pandas and NumPy, which helped streamline data operations and processing tasks. To better understand the data through visual representation, we used Matplotlib and Seaborn these libraries allowed us to create clear and meaningful visualizations during our exploratory analysis phase.

When it came to building and assessing machine learning models, we turned to Scikit-learn, a robust library known for its wide selection of algorithms and tools for model training, preprocessing, and performance evaluation. Our work was primarily done using Jupyter Notebook and Visual Studio Code, both of which provided user-friendly environments for writing code, documenting our process, and visualizing results. This combination of tools supported a smooth and structured workflow, from data cleaning and model creation to evaluation and result interpretation. All experiments were carried out locally on a CPU-based system.

3.5 Model Building

3.5.1 Model Selection

To effectively answer our research questions specifically those related to the factors influencing passenger satisfaction and identifying the most accurate prediction methods we placed significant emphasis on selecting the right models. We tested and compared a variety of models, drawing from both traditional machine learning algorithms and modern deep learning techniques to assess their predictive strengths. Among the classical machine learning methods, we used Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost. These models were chosen due to their strong track record in classification tasks, each offering distinct benefits: for example, Logistic Regression is valued for its simplicity and interpretability, tree-based models allow for straightforward feature importance analysis, and XGBoost is known for its speed and high performance in complex datasets. Beyond these traditional approaches, we also implemented deep learning models, particularly neural networks, using libraries such as TensorFlow and Keras. Our deep learning experiments included architectures like Multi-Layer Perceptron (MLP), TabNet, and Deep & Cross Network (DCN), each bringing advanced learning capabilities suited for capturing intricate patterns in the data.

3.5.2 Model training

After preprocessing and splitting the dataset into training and testing sets (80% for training and 20% for testing then, the training data was fed into selected model, allowing them to learn from the input features which is selected from feature selection stage of data preprocessing and their corresponding satisfaction labels. For classical machine learning models, training initiated with model-specific parameters using default settings, followed by hyper parameter tuning during model optimization stage by applying GridSearchCV and Cross Validation. Figure 10 depicts training and testing accuracy of XGBoost algorithm.

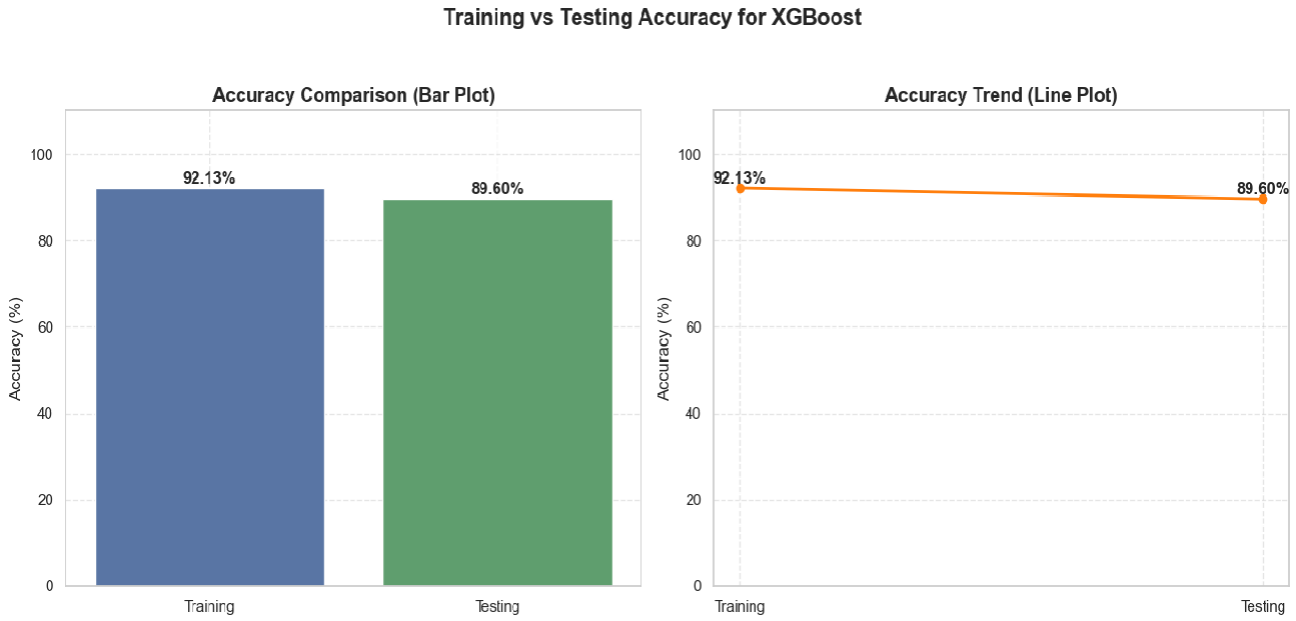


Figure 10: Training and Testing accuracy for XGBoost algorithm.

For the deep learning model, we designed a multi-layer neural network using Keras, defining the number of hidden layers, activation functions (such as ReLU), optimizer (e.g., Adam), and loss function (e.g., binary cross-entropy). The model was trained over several epochs with appropriate batch sizes, while monitoring performance on the validation set to avoid overfitting. Throughout the training process, we used techniques like early stopping and dropout regularization in deep learning to ensure generalization. Figure 11 shows that training loss and validation loss of MLP algorithm.

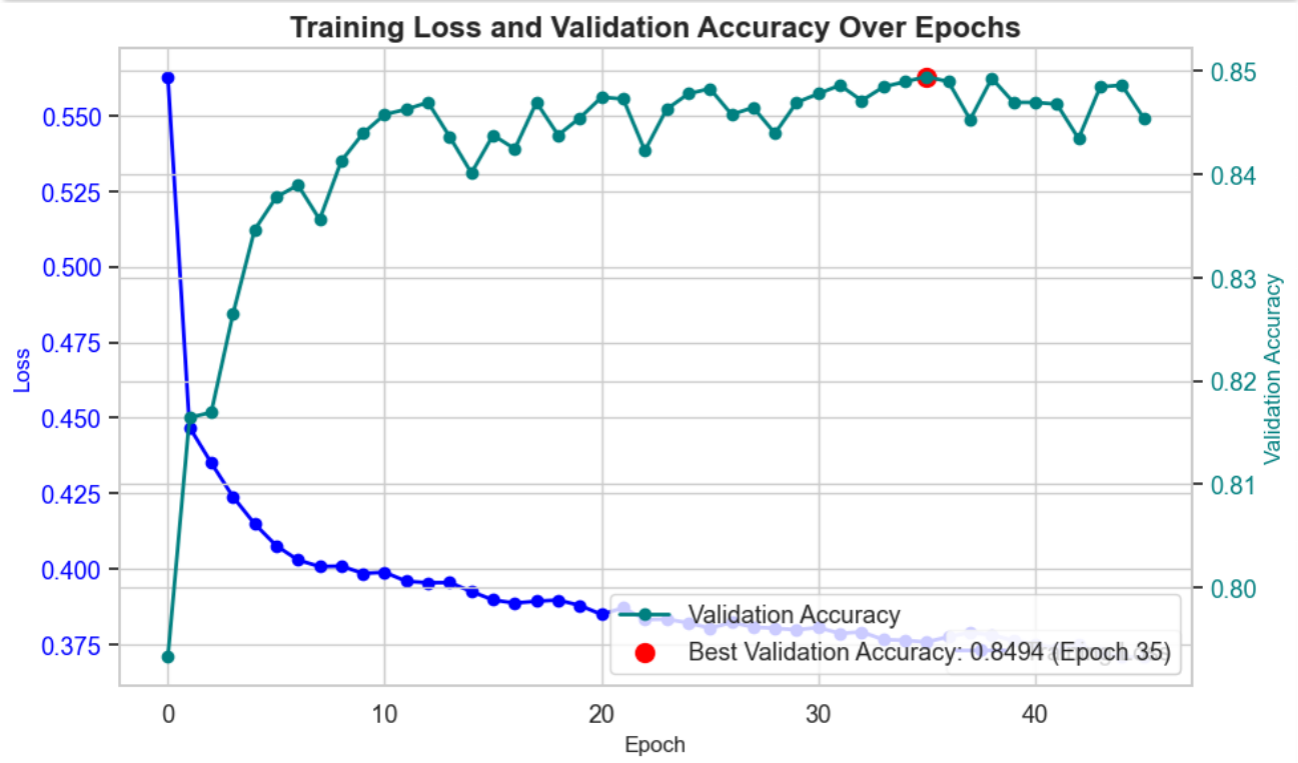


Figure 11: Training and Testing accuracy for MLP algorithm.

3.5.3 Hyperparameter tuning

To systematically search for the optimal combination of hyper parameters, we used techniques such as Grid Search and Cross-Validation, primarily provided through Scikit-learns tuning utilities. Grid Search exhaustively searches all possible combinations within a specified range, while cross-validation explores a random subset, offering faster results with good performance. For deep learning models, we experimented with different architectures by adjusting the number of hidden layers, batch size, learning rate, dropout rates, activation functions and Keras tuner, using Kera's' built-in functions such as Early Stopping to prevent overfitting and improve convergence. Through this tuning process, we were able to significantly enhance the performance of both classical and deep learning models.
hyperparameter tuning for the Multi-Layer Perceptron (MLP) model involved systematically adjusting key parameters such as the number of hidden layers and neurons per layer, learning rate, activation function, batch size, and dropout rate. We experimented with different configurations to balance model complexity and prevent overfitting, utilizing validation techniques like cross-validation to evaluate performance

For the machine learning models hyperparameter tuning was performed using GridSearchCV with cross-validation. In Logistic Regression, the regularization strength parameter **C** (0.01–100), solvers (*liblinear*, *saga*), and penalty types (*l1*, *l2*) were tuned. For Decision Trees, parameters such as split criterion (*gini*, *entropy*), maximum depth (None–20), minimum samples per split (2–20), minimum samples per leaf (1–10), and maximum features (*sqrt*, *log2*, None) were optimized. In the Random Forest model, the grid included the number of trees (50–200), split criteria (*gini*, *entropy*), maximum depth (None–30), minimum samples for splits (2–10), leaf sizes (1–4), maximum features (*sqrt*, *log2*, None), bootstrap sampling (True/False), and class weighting (*balanced*, None). For the deep learning models (MLP, TabNet, DCN), hyperparameters included the number of hidden layers (64 and 32 neurons), activation functions (ReLU in hidden layers, sigmoid in the output layer), dropout rate (0.3), batch normalization, learning rate (Adam optimizer with 0.001), batch size (32), number of epochs (up to 100 with early stopping patience of 10), and binary crossentropy loss function.

3.6 Model Evaluation

After training and tuning the models, we assessed their performance on the test dataset, which consisted of 20% of the total data (5998 samples), previously unseen by the model during training. A variety of evaluation metrics were used to obtain a comprehensive understanding of model performance, including accuracy, precision, recall, and F1-score.

Each metric provided valuable insights, accurate measured the overall correctness of predictions, precision indicated how many of the predicted satisfied passengers were truly satisfied, recall showed how well the model identified all actual satisfied passengers, and the F1-score offered a balance between precision and recall. The confusion matrix was also used to visualize true positives, true negatives, false positives, and false negatives, helping to figure out specific areas where the model made incorrect predictions. For deep learning models, additional evaluation tools like validation loss curves and learning curves were analyzed to monitor overfitting or under fitting during training.

3.7 Model Interpretability

In our study, the following metrics and techniques were used for model interpretability. Coefficients (for models like Logistic Regression) were examined to understand the influence of each feature on the model's predictions, with higher absolute values indicating stronger effects. Rule-based Tree Plot was employed to visualize decision trees (like Decision Trees and Random Forests), showcasing how

decisions are made based on splitting rules at each node. explaining that the Decision Tree was intentionally visualized and kept simple for interpretability, making it understandable like other models, and that it was accompanied by feature importance analysis.

Feature Importance (for models like Random Forest and XGBoost) provided an overview of which features contribute most to the predictions by quantifying their relative importance. SHAP Values were used to explain the impact of each feature on individual predictions, offering detailed insights into the contribution of each feature in each instance Finally, by combining Feature Importance with SHAP, we were able to gain a deeper understanding of both the overall importance of each feature and how individual predictions were made. This approach provided a more reliable and insightful interpretation of the model's results. Overall, the research methodology used in this study is summarized in the figure below.



CHAPTER FOUR:

Findings, Model Build, and Discussions of the Study

Overview

To answer our research questions in a structured way, we carried out a series of carefully planned experiments. The goal of the first experiment was to figure out which factors play the biggest role in shaping passenger satisfaction with Ethiopian Airlines. To do this, we used feature selection methods combined with machine learning models and explainable AI (XAI) techniques. The second experiment focused on comparing how well different machine learning algorithms and deep learning could predict passenger satisfaction. In the next section, we take a closer look at each experiment and discuss the key findings in detail.

4.1 Model Build

This research focuses on designing several machine learning models to predict passenger satisfaction for Ethiopian Airlines. We built a variety of classifiers, including classical machine learning algorithms like, which are capable of capturing complex and multi-dimensional relationships within the data. Before building the models, we carefully prepared the dataset by handling missing values, encoding categorical variables, scaling features, and selecting the most relevant ones. To improve the models' performance, we used techniques like grid search and cross-validation for tuning their hyperparameters. We evaluated each model using metrics like accuracy and F1-score to ensure they could reliably distinguish between satisfied and dissatisfied passengers. In the next section, we will walk through the different experiments conducted and share their results.

In this study, a total of four main experiments were conducted. The two experiment focused on identifying the most influential factors affecting passenger satisfaction, using feature selection techniques combined with machine learning models with explainable AI (SHAP, LIME). The second experiment extended the comparison to deep learning approaches such as Multilayer Perceptron (MLP), TabNet, and Deep Cross Network (DCN), in order to evaluate whether they could outperform machine learning models given the dataset size and nature.

4.2 Findings of the Study

4.2.1 Experiment1: Exploring Key Hidden Factors of Passenger Satisfaction Using Machine Learning This experiment focuses on finding the key factors that have the biggest impact on how satisfied passengers are with Ethiopian Airlines. By using different feature selection methods alongside explainable artificial intelligence (XAI) techniques, we aim to uncover clear insights. These insights can then help guide practical improvements in the airline's services and support smarter, data-driven decisions to enhance the overall passenger experience.

4.2.1.1 Key hidden Determinants of Passenger Satisfaction Using Machine Learning

In this approach, we used two feature selection techniques that fall under the embedded method category. First, we ran an experiment with the Random Forest algorithm, which helped us identify important features based on its built-in feature importance scores. We chose Random Forest because it's reliable, handles complex and large datasets well, and can capture non-linear patterns. Next, we used XGBoost, a powerful gradient boosting algorithm known for its accuracy, ability to prevent overfitting, and skill in managing interactions between features. Both models were selected because their built-in feature selection methods let us pinpoint the most important factors affecting passenger satisfaction. The results from these experiments are shown in Figure 4 below.

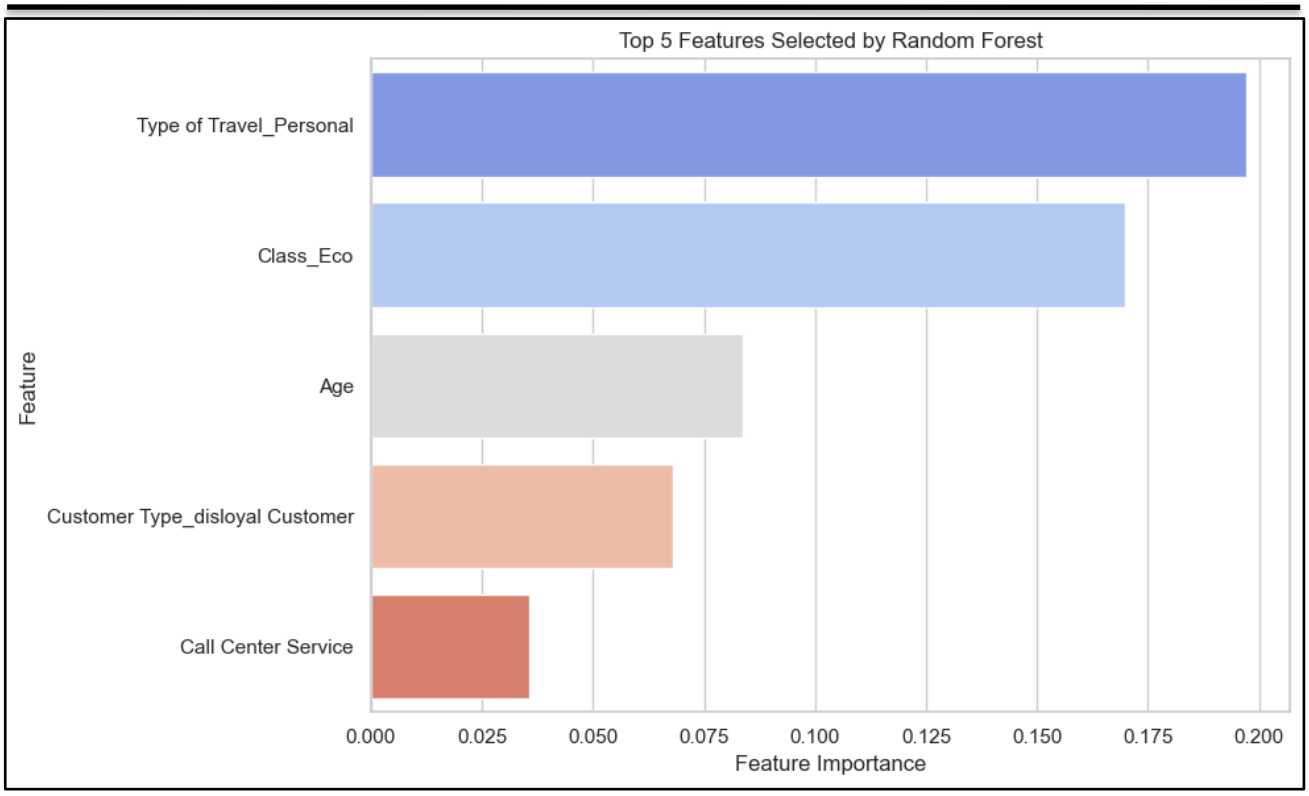


Figure 4:Top 5 Important hidden Features selected by Random Forest Algorithm.

The Random Forest algorithm found that the most important factors influencing passenger satisfaction are traveling for personal reasons, flying in economy class, the passenger's age, whether the customer is disloyal, and their experience with call center service.

The XGBoost algorithm identifies Type of Travel_Personal, Class_Eco, Age, Customer Type_disloyal Customer, and Call Center Service as the top 5 key features significantly influencing passenger satisfaction as shown in figure 5 below.

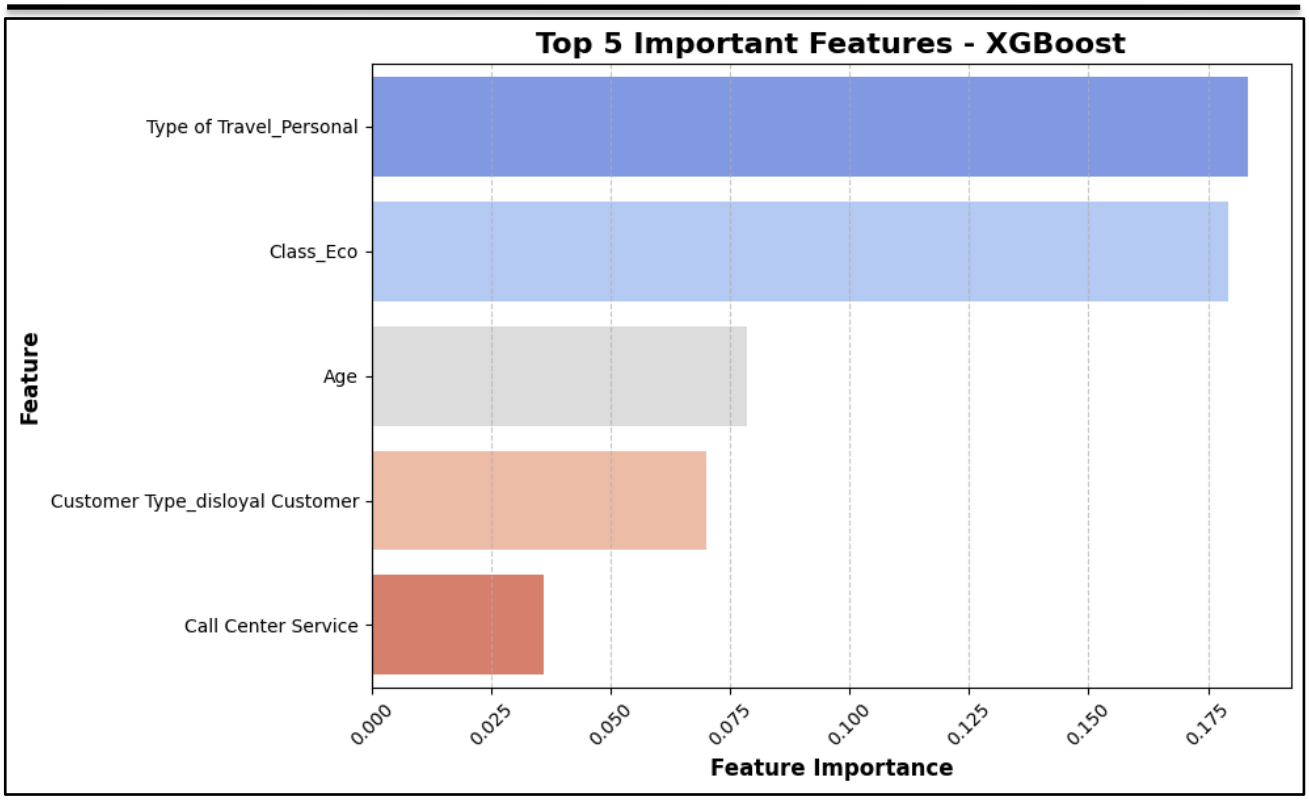


Figure 5: Top 5 Important hidden Features selected by XGBoost Algorithm.

In both experiments, both algorithms consistently identified the same key features contributing to passenger satisfaction. These features include Type of Travel, Class_Eco, Age, Customer_Type (Disloyal Customer), and Call Center Service. Both Random Forest and XGBoost ranked Type of Travel as the first most important feature, and each included Call Center Service in their top five.

4.2.1.2 Key Determinants of Passenger Satisfaction Using Explainable AI (XAI)

In this approach, we applied feature selection techniques under the Explainable AI (XAI) category. The experiment employed the XGBoost algorithm, a powerful gradient boosting model known for its accuracy, regularization capabilities, and ability to manage feature interactions [51]. We applied SHAP and LIME with the XGBoost algorithm to provide explainability and interpretability of the model's predictions. These methods were selected to Utilize the explainable AI techniques in XGBoost, enabling the identification of the most influential features contributing to passenger satisfaction. The results of this experiment are presented in the figure 6 below.

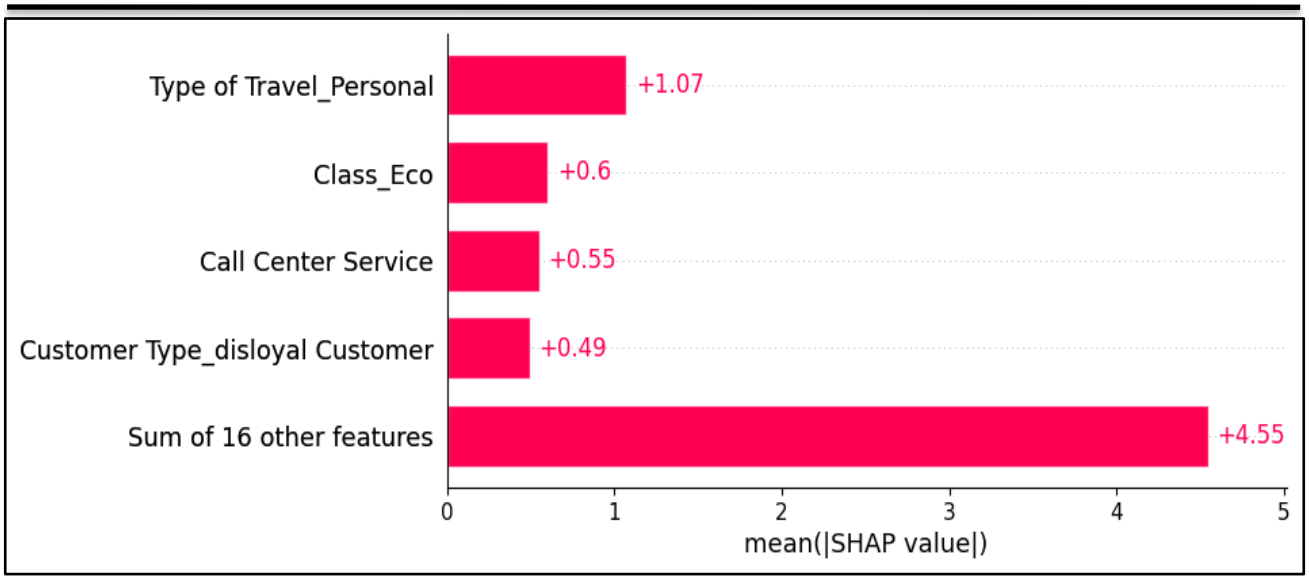


Figure 6: Top 5 Features Based on SHAP with XGBoost Values.

Based on the mean absolute SHAP values, the most significant factors influencing the XGBoost model's output (passenger satisfaction) are Type of Travel_Personal, Class_Eco, Call Center Service, Customer Type_disloyal Customer, and the collective impact of the other 16 features.

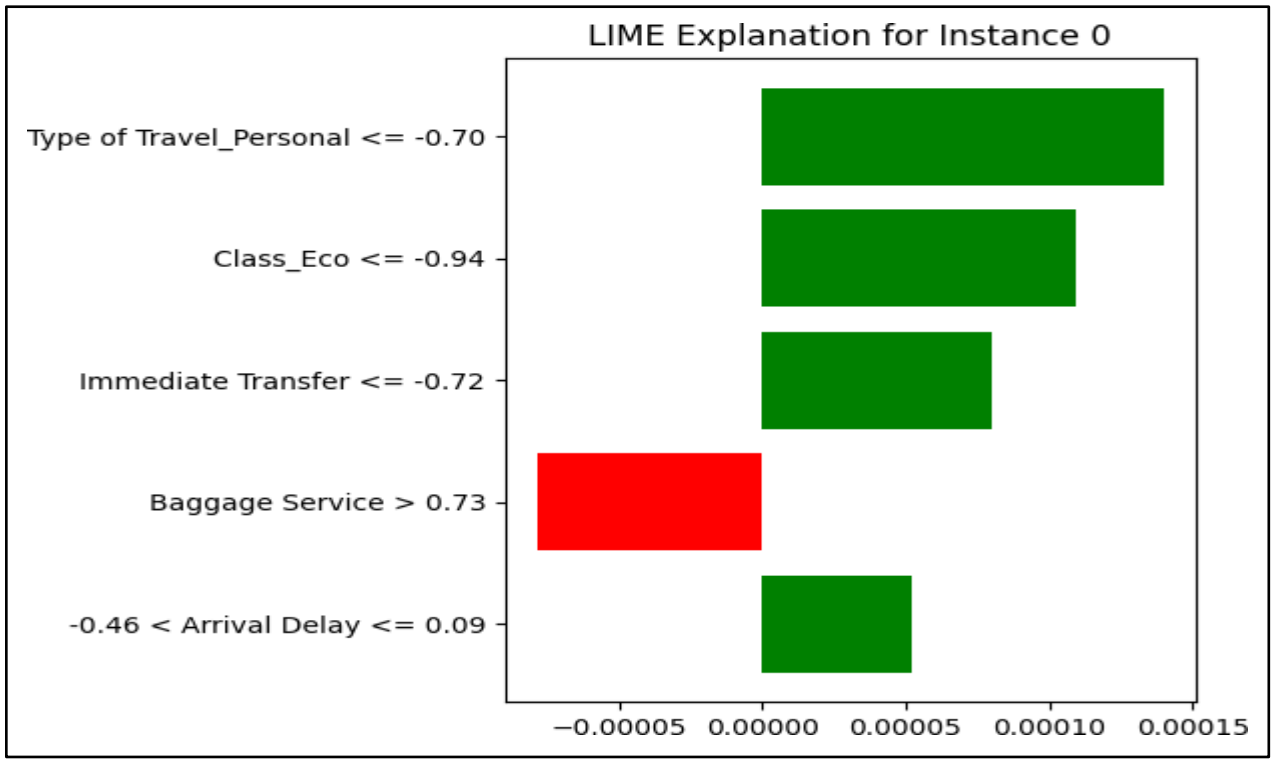


Figure 7: Top 5 Features Based on LIME with XGBoost Values.

Based on the LIME explanation for this specific instance (Instance 0) using the XGBoost algorithm, the features most influencing the prediction for this particular passenger's satisfaction are Type of

Travel being 'Personal' (<= -0.70), Class being 'Eco' (<= -0.94), Immediate Transfer (<= -0.72), Baggage Service being greater than 0.73, and Arrival Delay being between -0.46 and 0.09.

Based on the mean absolute SHAP values, the most influential features affecting the XGBoost model's overall predictions for passenger satisfaction are Type of Travel_Personal, Class_Eco, Call Center Service, Customer Type_Disloyal Customer, along with the combined effect of the remaining 16 features. For a specific case (Instance 0), the LIME explanation highlights that Type of Travel = Personal, Class = Eco, Immediate Transfer, Baggage Service > 0.73, and Arrival Delay between -0.46 and 0.09 are the key factors driving the prediction for that passenger's satisfaction.

4.2.1.3 Summary for key Determinant of passenger satisfaction.

XGBoost and Random Forest together identified the same key features such as Type of Travel, Class, Age, Customer Type, and Call Center Serviceindicating consistency across models. The use of SHAP values further validated these features at both global and individual levels. Given the high performance and interpretability of XGBoost, along with corroborative findings from other feature selection methods, analyzing this model alone provides a sufficient and reliable understanding of the main factors influencing passenger satisfaction in this context.Using XGBoost and Random Forest for feature selection is sufficient because both are highly effective models capable of accurately identifying important features. Techniques like SHAP and LIME enhance their interpretability. The key features identified are consistent across models, confirming the reliability of their selections. Therefore, relying on these models alone provides a justified understanding of the main satisfaction drivers."

Method	*Significant Features Identified*	*Key Observations*	*Additional Metrics*
Random Forest	*Type of Travel, Class_Eco, Age, Customer_Type (Disloyal Customer), Call Center Service*	*Consistently identified core predictors of satisfaction. Type of Travel ranked as most important.*	*Top 5 Feature Importance (Gini Importance)*
XGBoost	*Type of Travel, Class_Eco, Age, Customer_Type (Disloyal Customer), Call Center Service*	*Similar to Random Forest, emphasizing the same key features. Type of Travel ranked highest.*	*Top 5 Feature Importance (Gain)*
XGBoost with SHAP	*Type of Travel_Personal, Class_Eco, Call Center Service, Customer Type_disloyal Customer*	*SHAP analysis confirmed the strong average impact of Type of Travel and Class. Collective effect of other 16 features noted.*	*Mean Absolute SHAP Values*
XGBoost with LIME	*Type of Travel_Personal, Class_Eco, Immediate Transfer, Baggage Service, Arrival Delay (Instance 0)*	*LIME provided a local explanation for a specific case, highlighting Personalized influences not visible in global metrics.*	*Feature Contribution to Prediction for Instance 0 (LIME)*

Table 2: Summary of Best contributing factors for passenger satisfaction

4.2.2 Experiment 2: Assessing the Performance of classical Machine Learning Algorithms

This experiment aims to identify the most accurate and reliable machine learning algorithm for predicting passenger satisfaction in Ethiopian Airlines by evaluating the performance of five widely used classification models: Logistic Regression was used because it's simple, easy to interpret, and works well with data that can be separated in a straight line [52]. Decision Tree was included due to its clear, visual structure and its ability to handle both numbers and categories [53]. Random Forest, which is made up of many decision trees, was chosen for its strong accuracy, ability to handle complex data, and built-in protection against overfitting.

Support Vector Machine (SVM) was selected for its excellent performance in high-dimensional datasets and its skill in finding complex decision boundaries [54, 55, 10]. XGBoost, a powerful boosting algorithm, stood out for its high accuracy, ability to handle large-scale data, and built-in features that help prevent overfitting. The results from each algorithm are shared in terms of how well they performed, their classification reports, and how easy they are to interpret. The next section explains how the experiments were carried out.

[56].The findings from each algorithm are presented in terms of model performance, classification report, and interpretability. The experiments are outlined as follows.

4.2.2.1 Performance of Logistic Regression (LR) Algorithm

Model Performance

The Logistic Regression model showed solid performance, achieving an overall accuracy of 83.39%, meaning it correctly predicted passenger satisfaction in most cases. For passengers who were dissatisfied, the model had a precision of 86.38%—so when it predicted someone was dissatisfied, it was right about 86% of the time. Its recall was 79.29%, meaning it correctly identified about 79% of all truly dissatisfied passengers. The F1-score, which balances precision and recall, came out to 82.68%, showing the model handled this group well. For the satisfied passengers, the model had a precision of 80.86%, meaning it was correct about 81% of the time when predicting satisfaction. It had a recall of 87.50%, successfully identifying nearly 88% of all satisfied passengers.

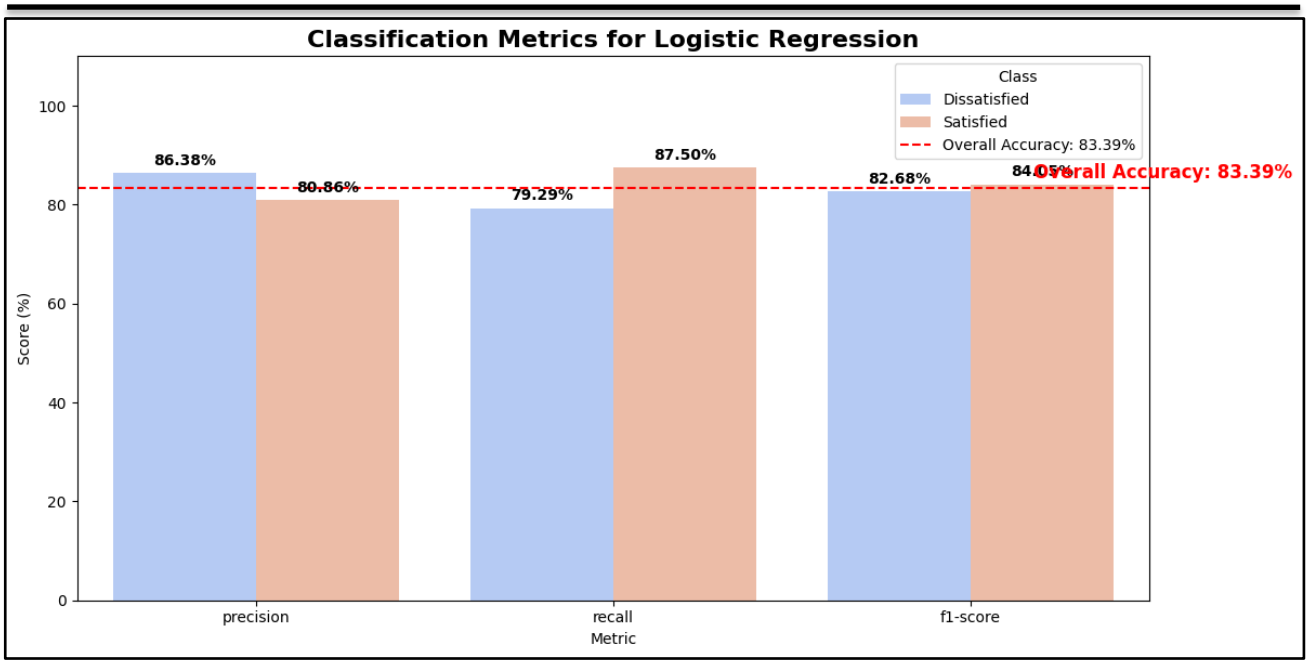


Figure 8: Classification metrics for Logistic Regression

Confusion Matrix Result

The confusion matrix for the Logistic Regression Model Utilizing its performance in predicting passenger satisfaction. The model correctly classified a total of 5002 instances. Specifically, it correctly identified 2378 dissatisfied passengers (True Negatives) and 2624 satisfied passengers (True Positives). Conversely, there were a total of 996 incorrect classifications, consisting of 621 dissatisfied passengers who were incorrectly predicted as satisfied (False Negatives), and 375 satisfied passengers who were wrongly classified as dissatisfied (False Positives). This indicates that while the model shows a good overall number of correct predictions, there are still a notable number of misclassifications, particularly in incorrectly classifying dissatisfied passengers as satisfied.

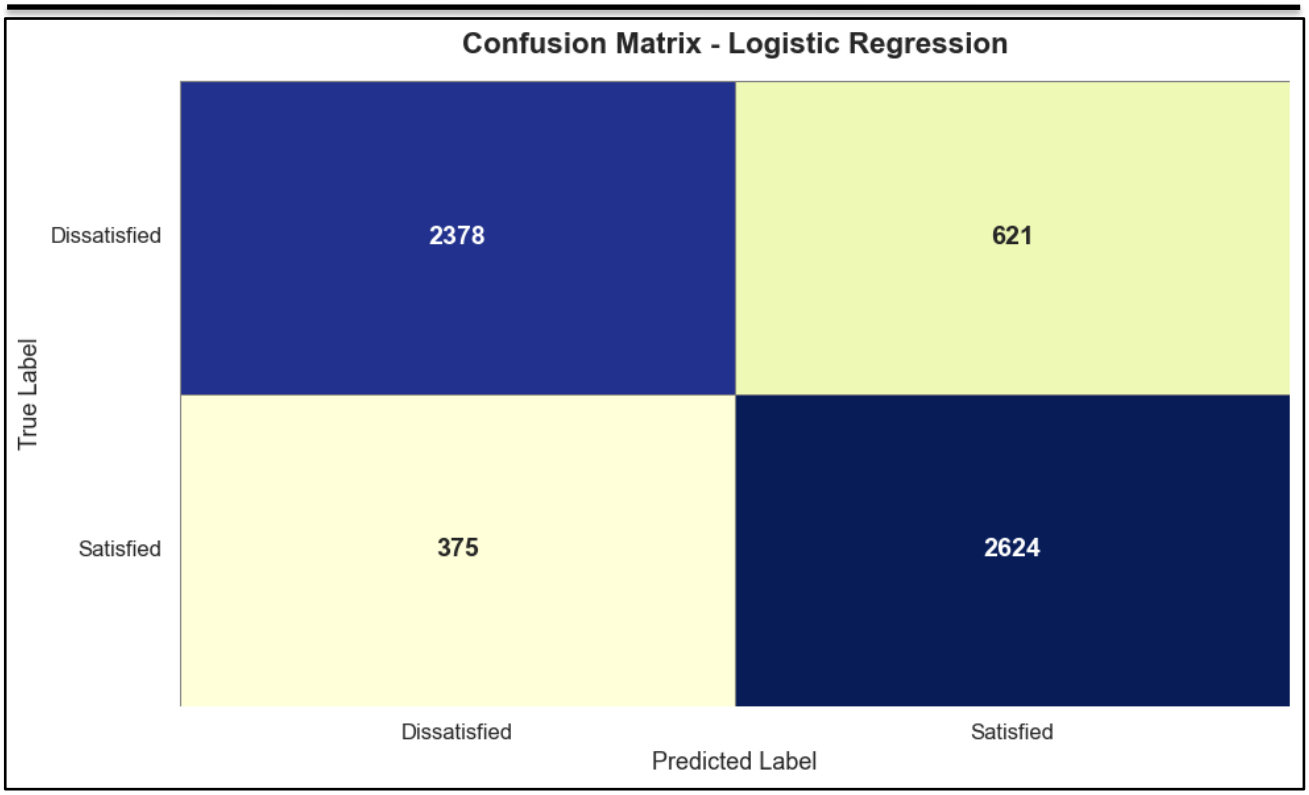


Figure 9: Confusion Matrices Result for Logistic Regression

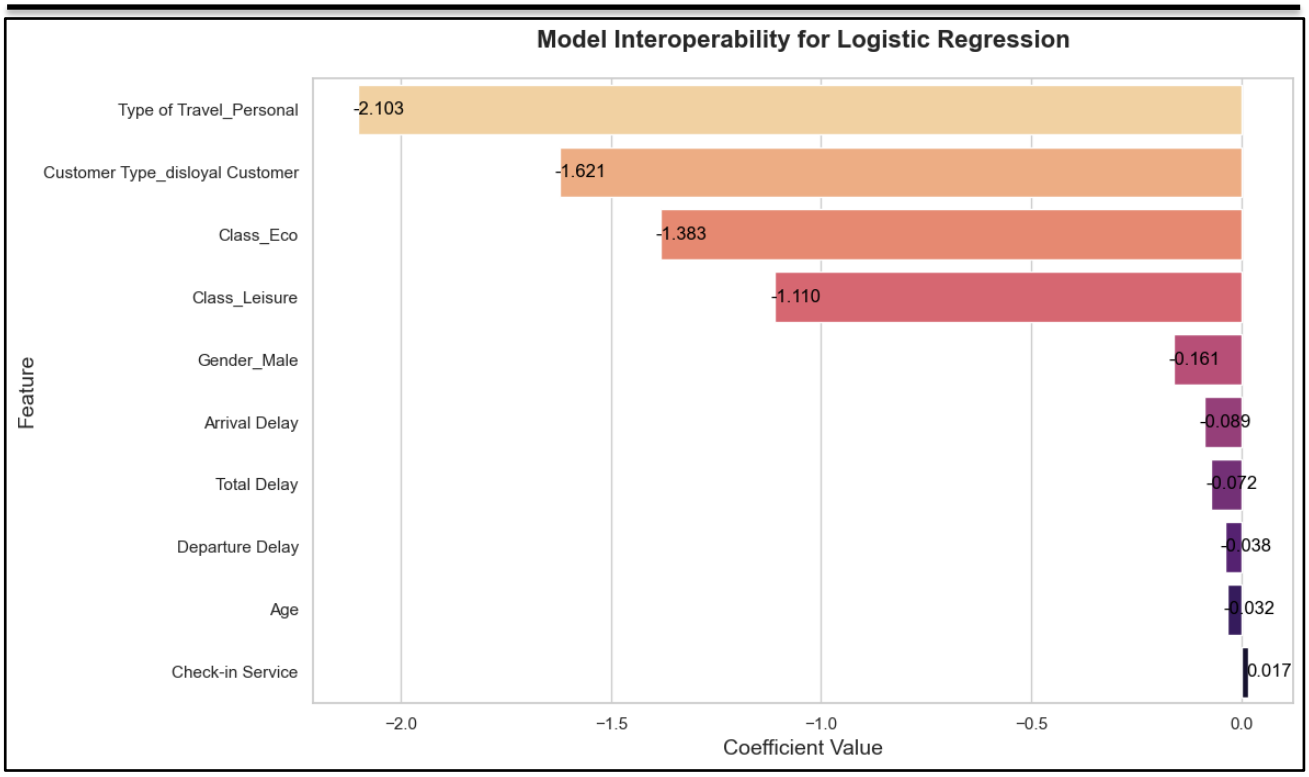


Figure 10: Model Interpretability for Logistic Regression

4.2.2.2 Performance of Decision Tree Algorithm

Model Performance

The Decision Tree model achieved an overall accuracy of 85.55%, indicating a strong ability to correctly classify passenger satisfaction. For the Dissatisfied class, it achieved a precision of 86.68%, meaning that when the model predicted a passenger was dissatisfied, it was correct approximately 86.68% of the time. The recall for the Dissatisfied class was 83.99%, indicating that the model correctly identified 83.99% of all actual dissatisfied passengers. The resulting F1-score of 85.32% represents a balanced measure of precision and recall for the Dissatisfied class. On the other hand, the model performed for the Satisfied class with a precision of 84.48%, meaning that when the model predicted a passenger was satisfied, it was correct approximately 84.48% of the time. The recall for the Satisfied class was 87.10%, indicating that the model correctly identified 87.10% of all actual satisfied passengers. The F1-score of 85.77% represents a balanced measure of precision and recall for the Satisfied class.

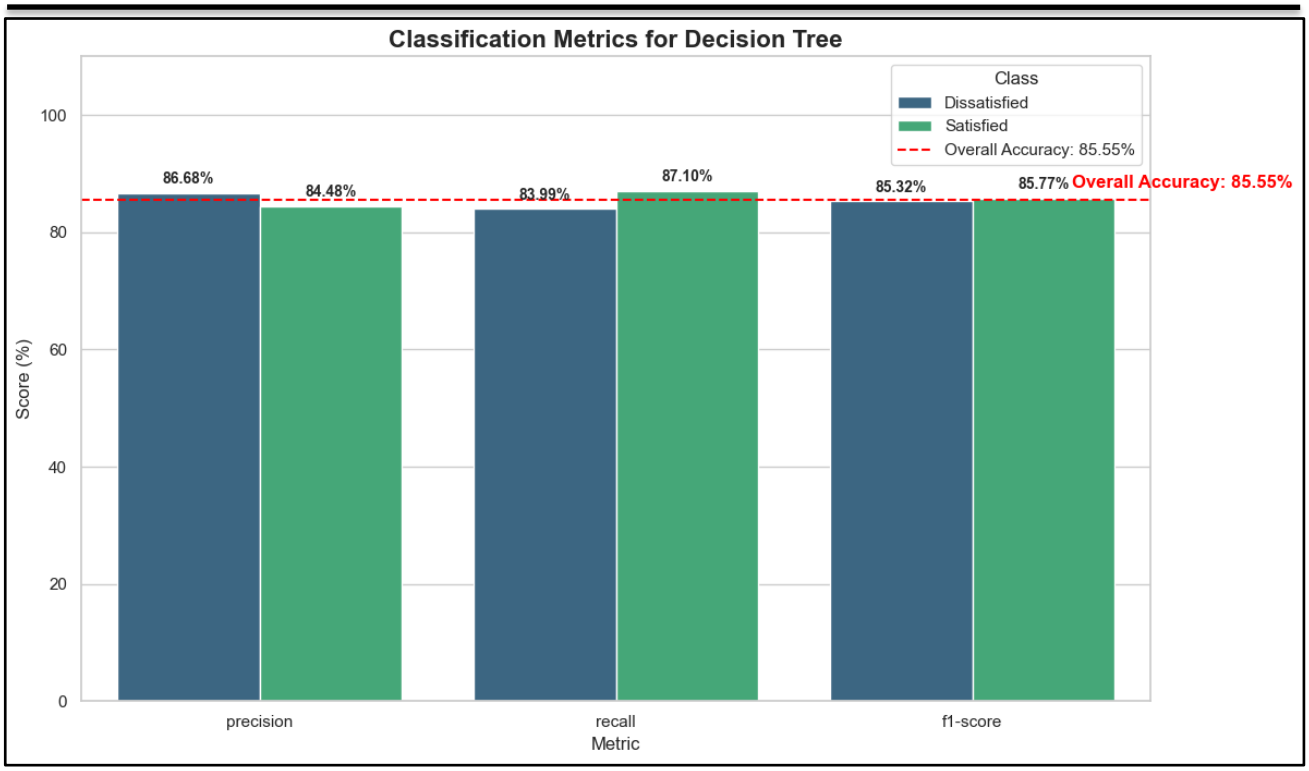


Figure 11: Classification Report for Decision Tree Algorithm

Confusion Matrix Result

The Decision Tree's confusion matrix Utilizing it correctly classified 5131 instances. Specifically, it accurately identified 2519 dissatisfied passengers (True Negatives) and 2612 satisfied passengers (True Positives). However, there were 867 misclassifications, with 480 dissatisfied passengers incorrectly predicted as satisfied (False Negatives) and 387 satisfied passengers wrongly labeled as dissatisfied (False Positives). While the model demonstrates a strong overall number of correct predictions, the misclassification of dissatisfied passenger's warrants attention for potential improvement.

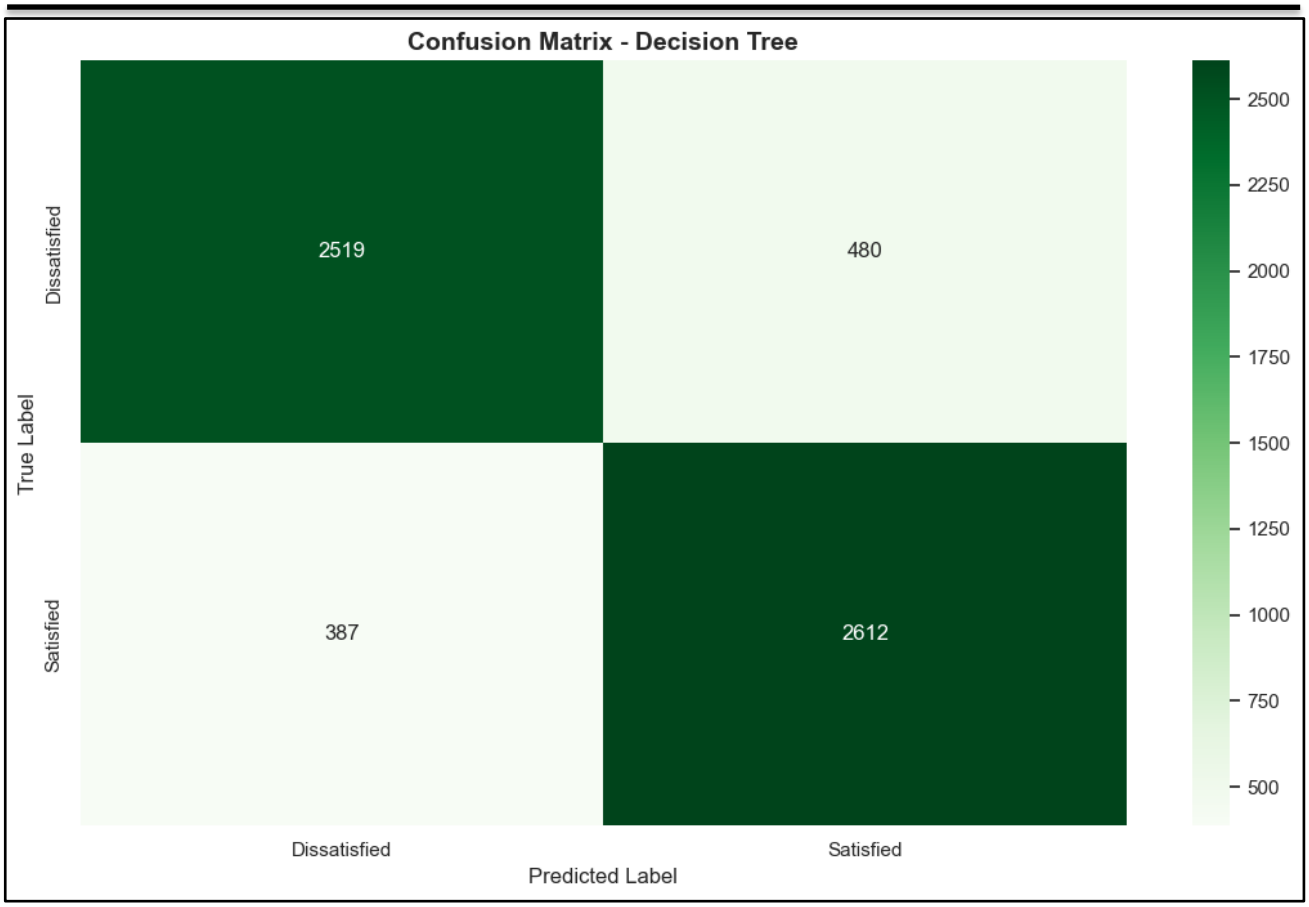


Figure 12: Confusion Matrices Result for Logistic Regression

Model Interpretability

The decision tree visualization highlights Type of Travel as a significant initial splitter, with Type of Travel_Personal <= 0.5 leading down a path towards Satisfied passengers and Type of Travel_Personal > 0.5 towards Not Satisfied. Customer Type Disloyal Customer <= 0.5 also appears as a factor influencing satisfaction, with disloyal customers showing a tendency towards Not Satisfied in subsequent splits. Class_Eco <= 0.5 is another early split, directing towards the Not Satisfied outcome. Arrival Delay is also featured in the tree, with specific thresholds (e.g., <= -0.215) contributing to the prediction of Not Satisfied.

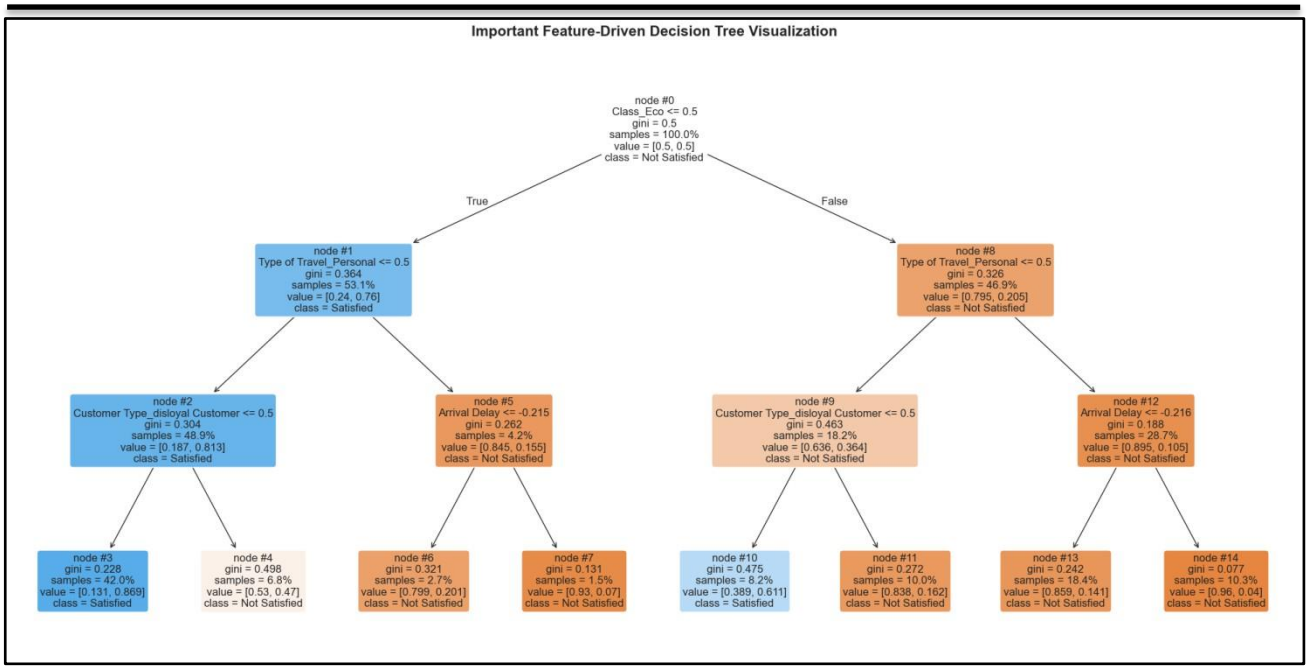


Figure 13: Model Interpretability for Decision Tree Algorithm

4.2.2.3 Performance of Random Forest Algorithm

Model Performance

The Random Forest model achieved an overall accuracy of 89.61%, indicating a strong ability to correctly classify passenger satisfaction. For the Dissatisfied class, it achieved a precision of 89.21%, meaning that when the model predicted a passenger was dissatisfied, it was correct approximately 89.21% of the time. The recall for the Dissatisfied class was 90.13%, indicating that the model correctly identified 90.13% of all actual dissatisfied passengers. The resulting F1-score of 89.67% represents a balanced and strong measure of precision and recall for the Dissatisfied class. On the other hand, the model performed for the Satisfied class with a precision of 90.03%, meaning that when the model predicted a passenger was satisfied, it was correct approximately 90.03% of the time. The recall for the Satisfied class was 89.10%, indicating that the model correctly identified 89.10% of all actual satisfied passengers. The F1-score of 89.56% represents a balanced and strong measure of precision and recall for the Satisfied class.

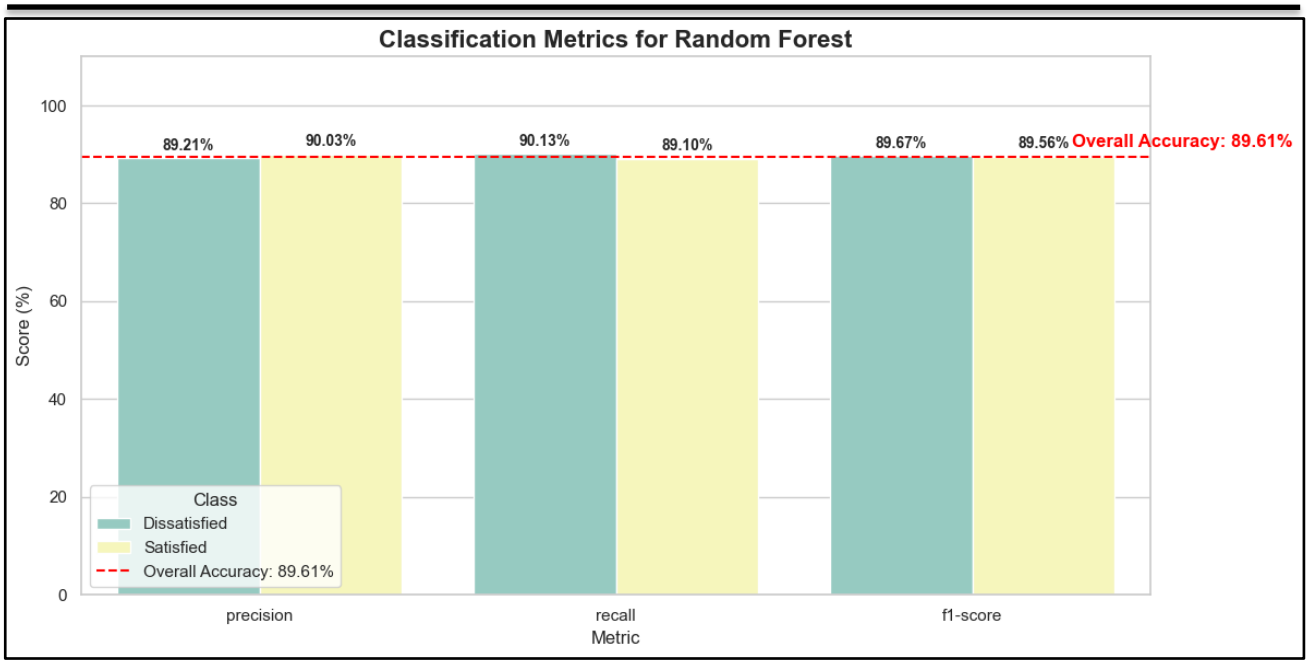


Figure 14:Classification Report for Random Forest Algorithm

Confusion Matrix Result

The Random Forest model correctly classified 5375 instances. Specifically, it accurately identified 2703 dissatisfied passengers (True Negatives) and 2672 satisfied passengers (True Positives). However, there were 623 misclassifications, with 296 dissatisfied passengers incorrectly predicted as satisfied (False Negatives) and 327 satisfied passengers wrongly labeled as dissatisfied (False Positives). While the model demonstrates a strong overall number of correct predictions, the misclassification of both dissatisfied and satisfied passengers indicates areas for potential refinement.

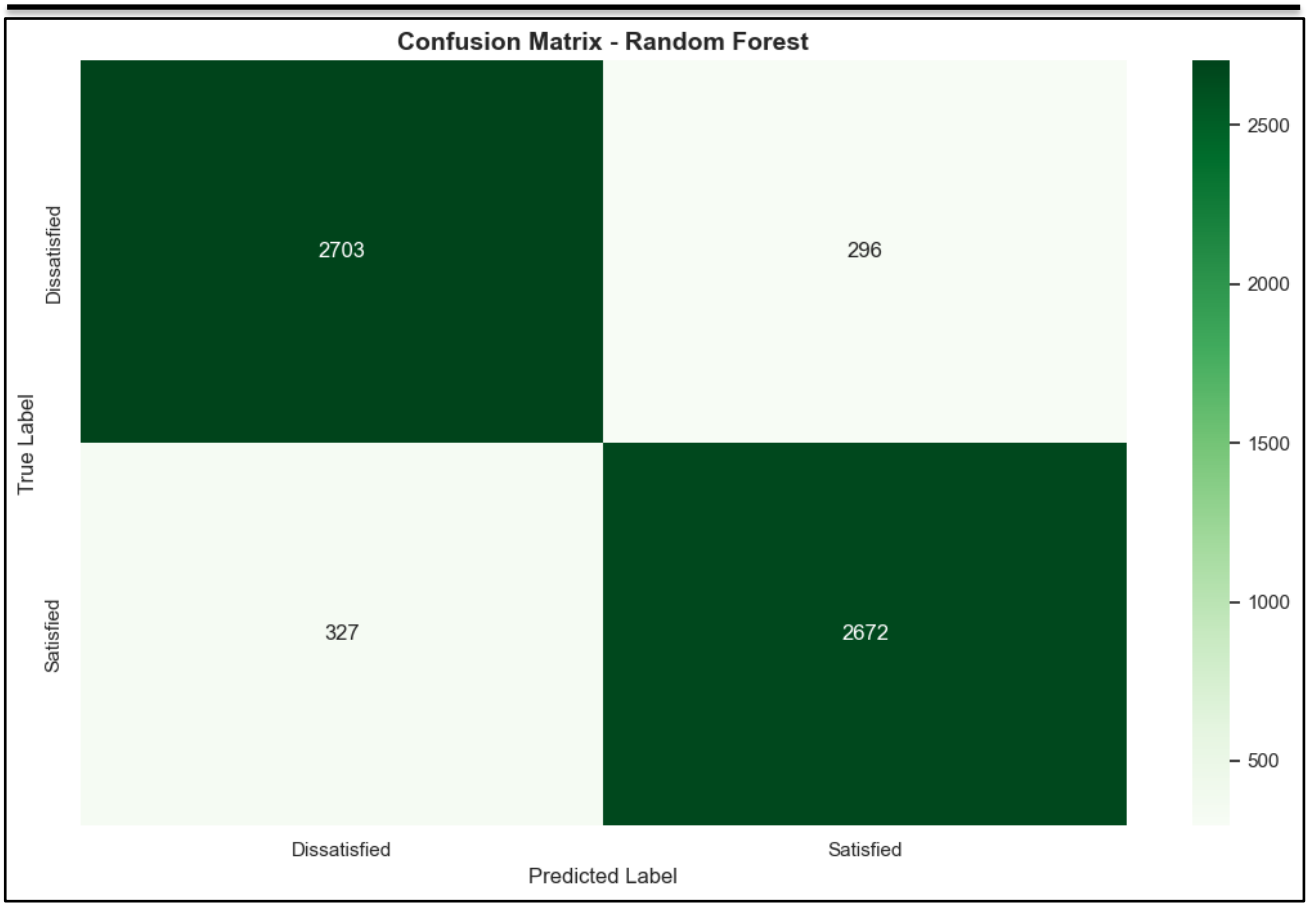


Figure 15:Confusion Matrices Result for Logistic Regression

Model Interpretability

The Random Forest model identifies Type of Travel Personal (0.206) as the most influential factor in predicting passenger satisfaction, followed by Class Eco (0.162) and Age (0.085). These results suggest that personal travel and economy class are strongly associated with the model's prediction of satisfaction, while age also plays a significant role. Several other factors, including Customer Type_disloyal Customer (0.067), Call Center Service (0.035), Lounge Service (0.035), Booking Services (0.034), Layover Service (0.034), Onboard Services (0.033), and Arrival Service (0.033), contribute to the model's prediction with lower importance scores, indicating that service-related aspects and customer loyalty also play a role in determining passenger satisfaction, though to a lesser degree than travel type, class, and age.

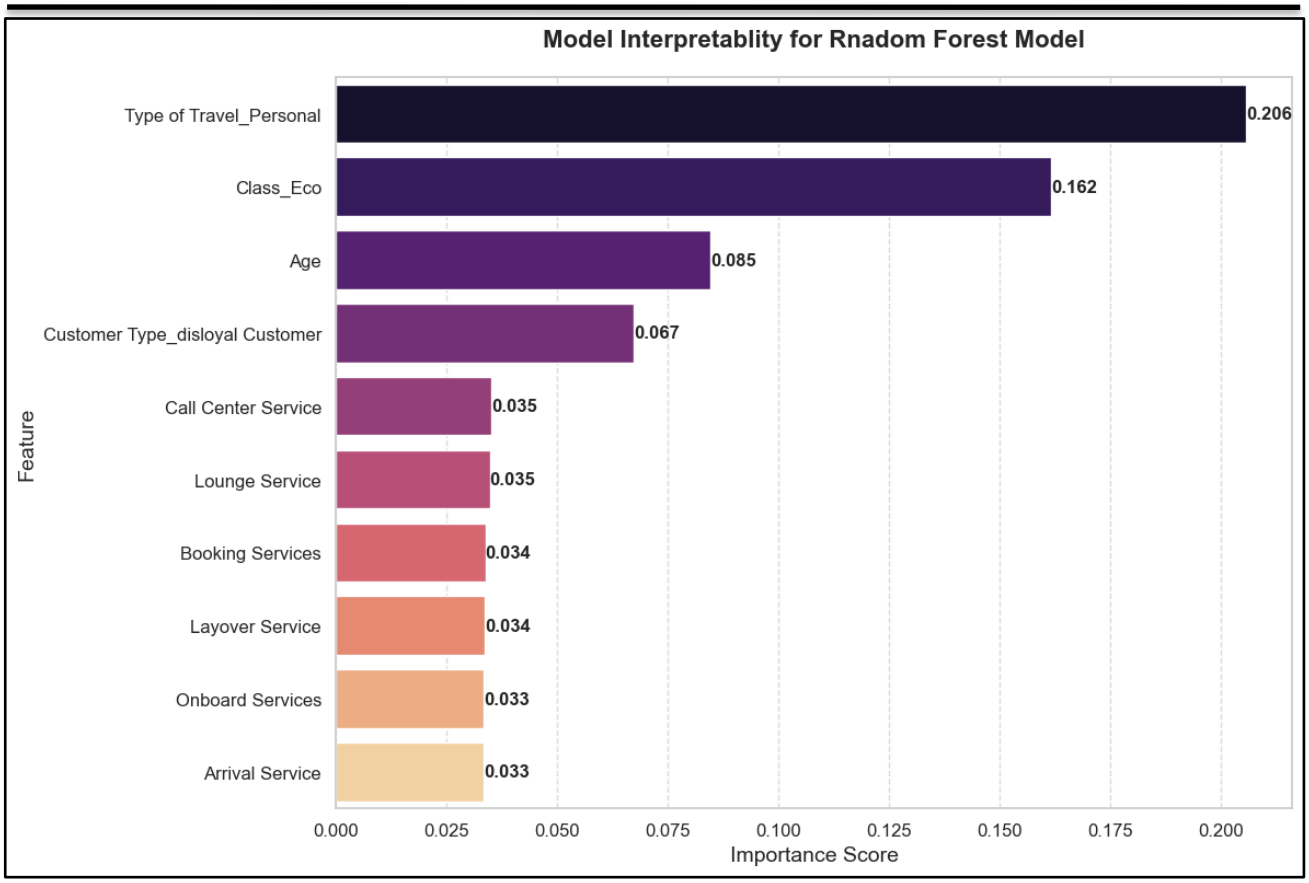


Figure 16: Model Interpretability for Random Forest Algorithm

4.2.2.4 Performance of Support Vector Machine Algorithm

Model Performance

The Support Vector Machine (SVM) model achieved an overall accuracy of 84.84%, indicating a strong ability to correctly classify passenger satisfaction. For the Dissatisfied class, it achieved a precision of 83.89%, meaning that when the model predicted a passenger was dissatisfied, it was correct approximately 83.89% of the time. The recall for the Dissatisfied class was 83.65%, indicating that the model correctly identified 83.65% of all actual dissatisfied passengers. The resulting F1-score of 83.73% represents a balanced measure of precision and recall for the Dissatisfied class. On the other hand, the model performed for the Satisfied class with a precision of 85.76%, meaning that when the model predicted a passenger was satisfied, it was correct approximately 85.76% of the time. The recall for the Satisfied class was 86.05%, indicating that the model correctly identified 86.05% of all actual satisfied passengers. The F1-score of 84.96% represents a balanced measure of precision and recall for the Satisfied class.

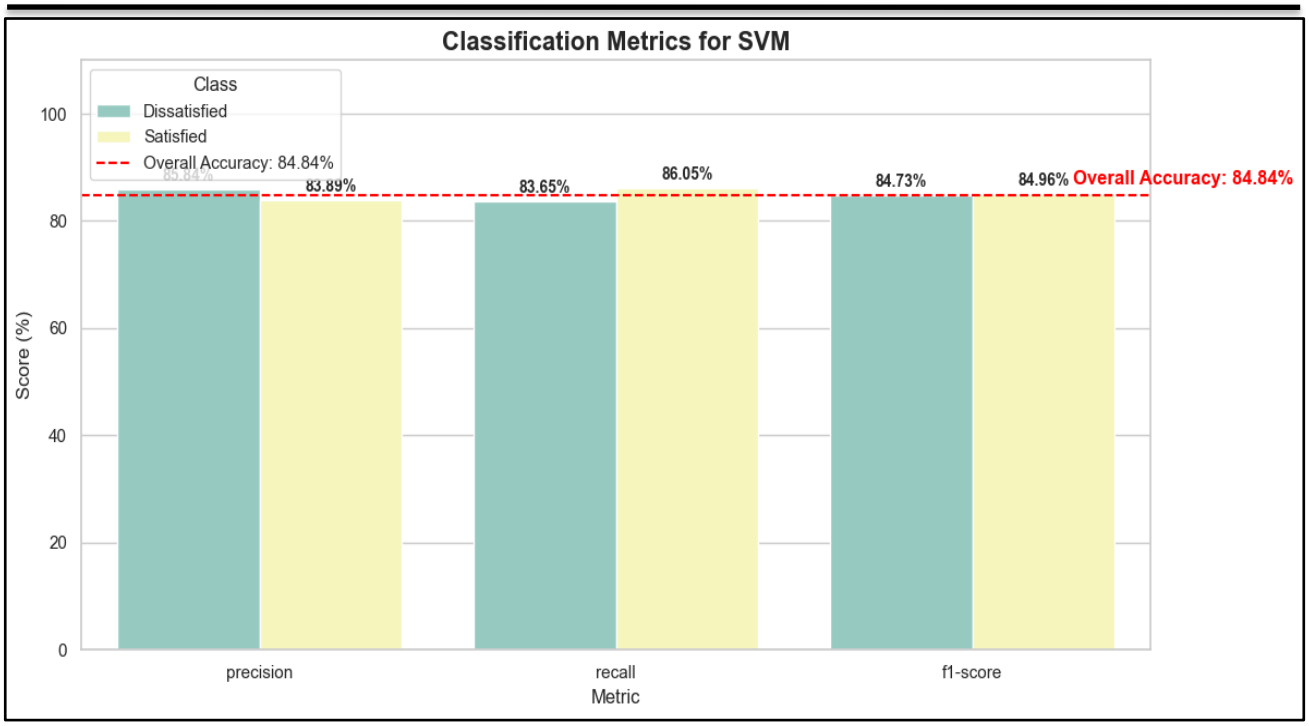


Figure 17: Classification Report for Support Vector Machine Algorithm

Confusion Matrix Result

The Support Vector Machine (SVM) model correctly classified 5089 instances. Specifically, it accurately identified 2522 dissatisfied passengers (True Negatives) and 2567 satisfied passengers (True Positives). However, there were 909 misclassifications, with 493 dissatisfied passengers incorrectly predicted as satisfied (False Negatives) and 416 satisfied passengers wrongly labeled as dissatisfied (False Positives). While the model demonstrates a strong overall number of correct predictions, the misclassification of both dissatisfied and satisfied passengers indicates areas for potential refinement.

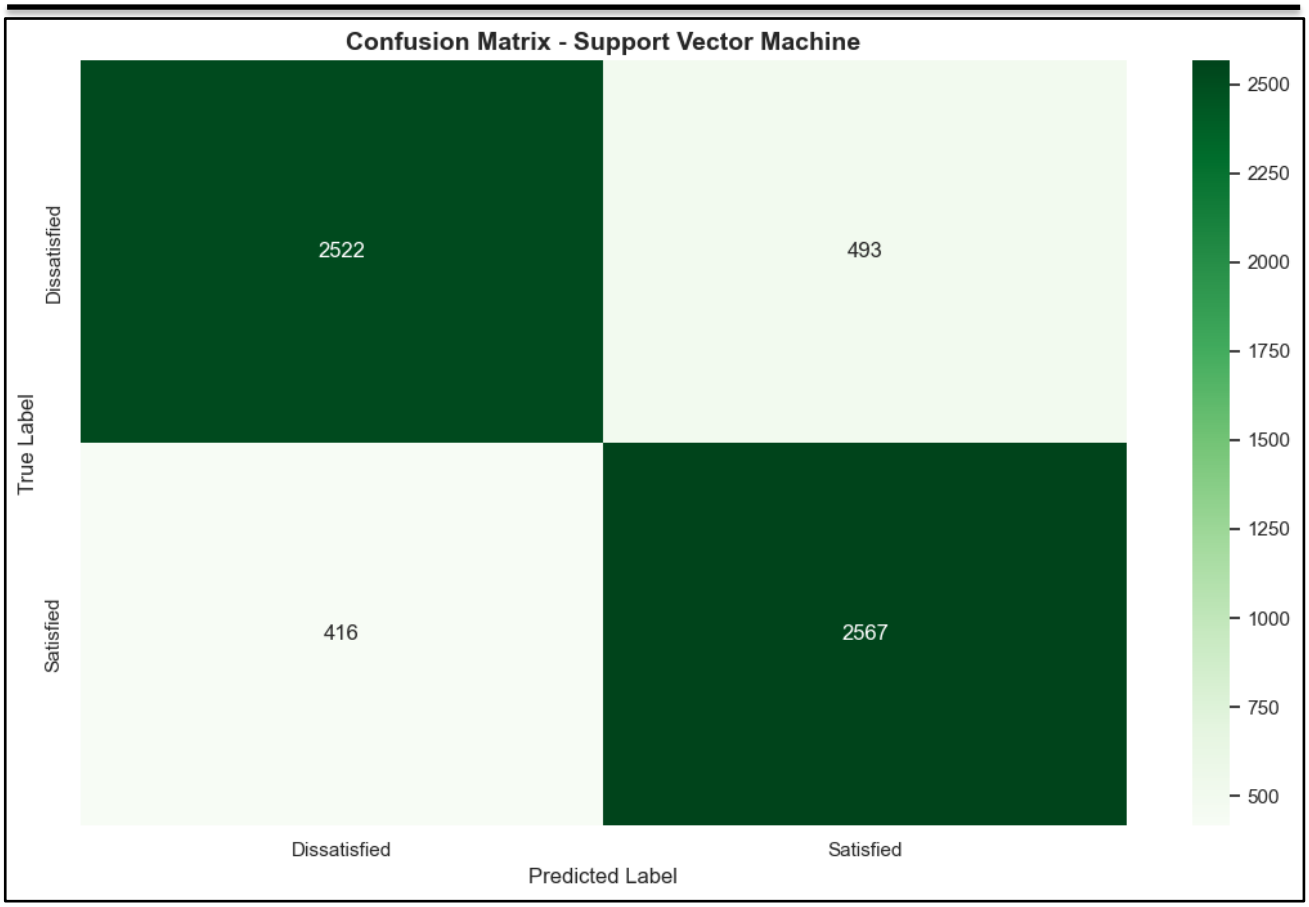


Figure 18: Confusion Matrices Result for Support Vector Machine

Model Interpretability

The SHAP summary plot for the Support Vector Machine (SVM) model Utilizing that customer type (disloyal) and type of travel (personal travel) are the most influential features in predicting passenger satisfaction. High values for these features are strongly associated with negative impacts on the model output, meaning that disloyal customers and those traveling for personal reasons are more likely to report dissatisfaction. Arrival delay also contributes significantly to decreased satisfaction, indicating that longer delays are linked to lower passenger satisfaction scores. In essence, the model highlights that a passenger's inherent disposition (loyalty) and the purpose of their travel are key determinants of their satisfaction, with travel disruptions further exacerbating negative sentiment.

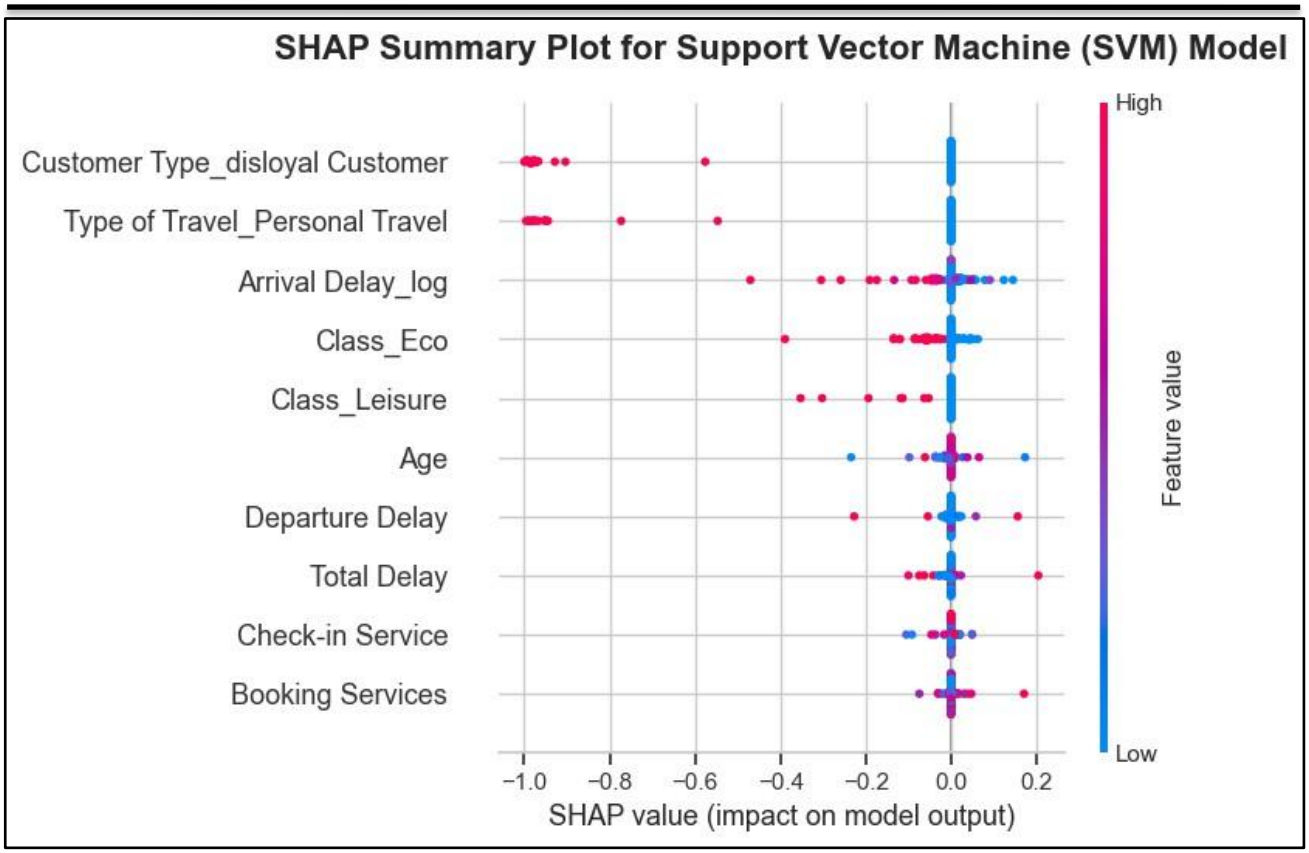


Figure 19: SHAP Summary Plot for Support Vector Machine Algorithm

4.2.2.5 Performance of XGBoost Algorithm

Model Performance

The XGBoost model achieved an overall accuracy of 89.60%, indicating a high level of effectiveness in predicting passenger satisfaction. For the Dissatisfied class, it attained a precision of 91.32%, meaning that when the model predicted a passenger was dissatisfied, it was correct approximately 91.32% of the time. The recall for the Dissatisfied class was 87.63%, indicating that the model correctly identified 87.63% of all actual dissatisfied passengers. The resulting F1-score of 89.44% represents a balanced and strong measure of precision and recall for the Dissatisfied class. In contrast, the model performed for the Satisfied class with a precision of 87.99%, meaning that when the model predicted a passenger was satisfied, it was correct approximately 87.99% of the time. The recall for the Satisfied class was 91.59%, indicating that the model correctly identified 91.59% of all actual satisfied passengers. The F1-score of 89.75% represents a balanced and strong measure of precision and recall for the Satisfied class.

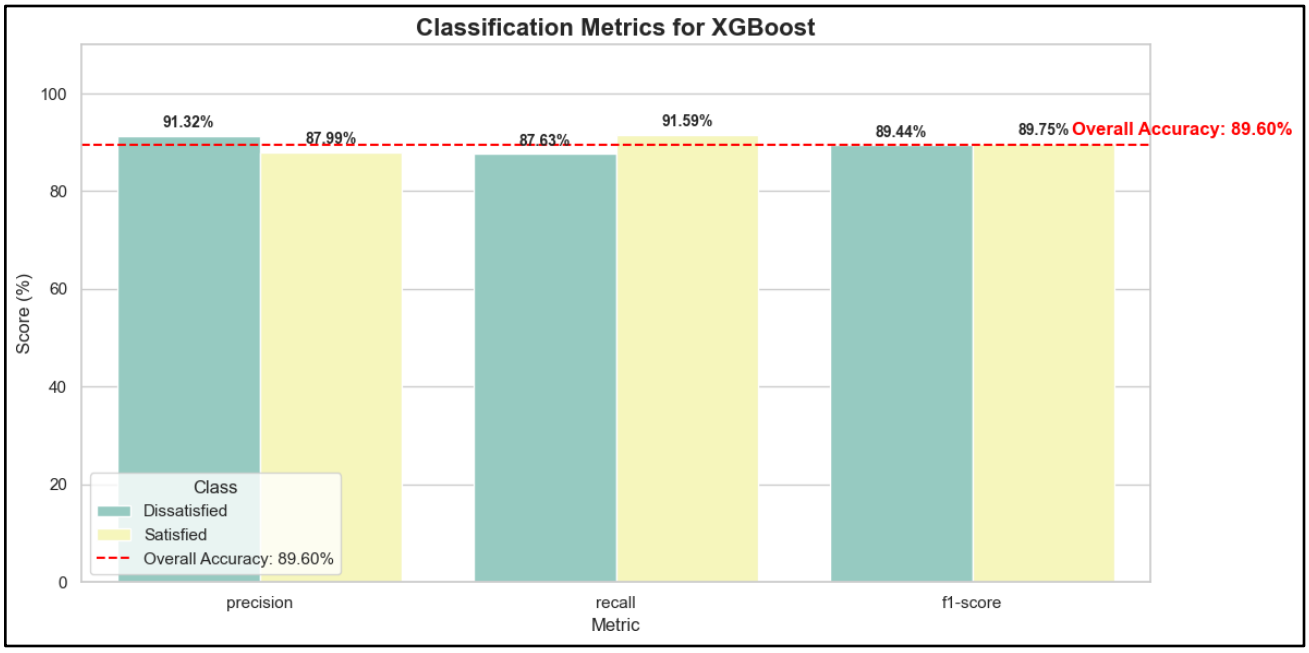


Figure 20: Classification Report for XGBoost Algorithm

Confusion Matrix Result

The XGBoost model correctly classified 5374 instances. Specifically, it accurately identified 2642 dissatisfied passengers (True Negatives) and 2732 satisfied passengers (True Positives). However, there were 624 misclassifications, with 373 dissatisfied passengers incorrectly predicted as satisfied (False Negatives) and 251 satisfied passengers wrongly labeled as dissatisfied (False Positives). While the model demonstrates a strong overall number of correct predictions, the misclassification of both dissatisfied and satisfied passengers indicates areas for potential refinement.

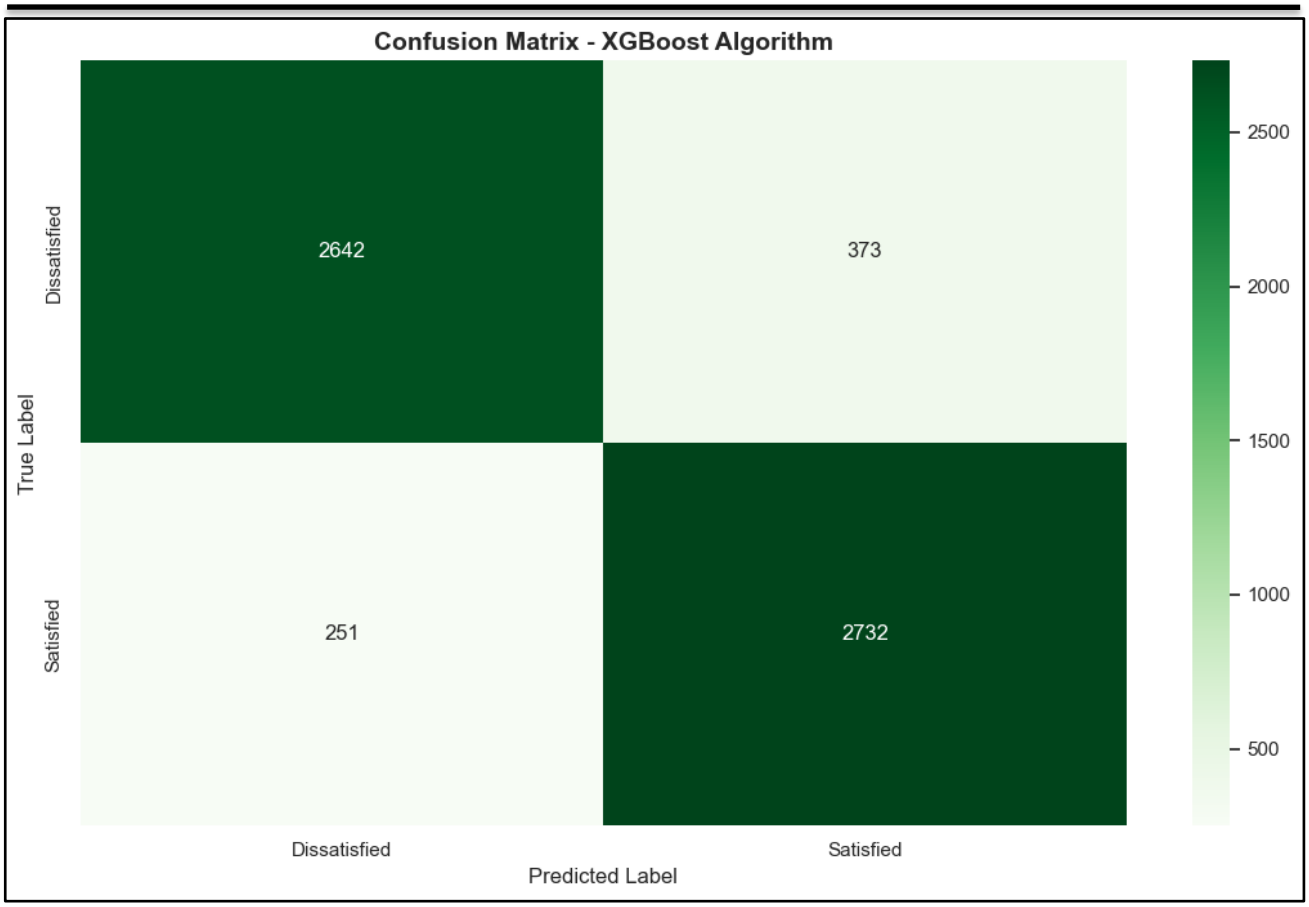


Figure 21: Confusion Matrices Result for XGBoost Algorithm

Model Interpretability

The SHAP summary plot for the XGBoost model Utilizing that Type of Travel Personal has the largest impact on the model output, with high feature values (indicated by red) generally correlating with a negative SHAP value (tending towards dissatisfaction). Class Eco also shows a significant negative impact when its value is high. Customer Type Disloyal Customer has a strong negative impact when its value is high. Arrival Delay shows that higher values generally lead to negative SHAP values (lower satisfaction). Call Center Service and Age also exhibit noticeable impact, with varying effects depending on their values. Class Leisure, Layover Service, Lounge Service, and Immigration Security have a smaller overall impact compared to the top features, but their specific values can still influence the model's prediction in either direction.

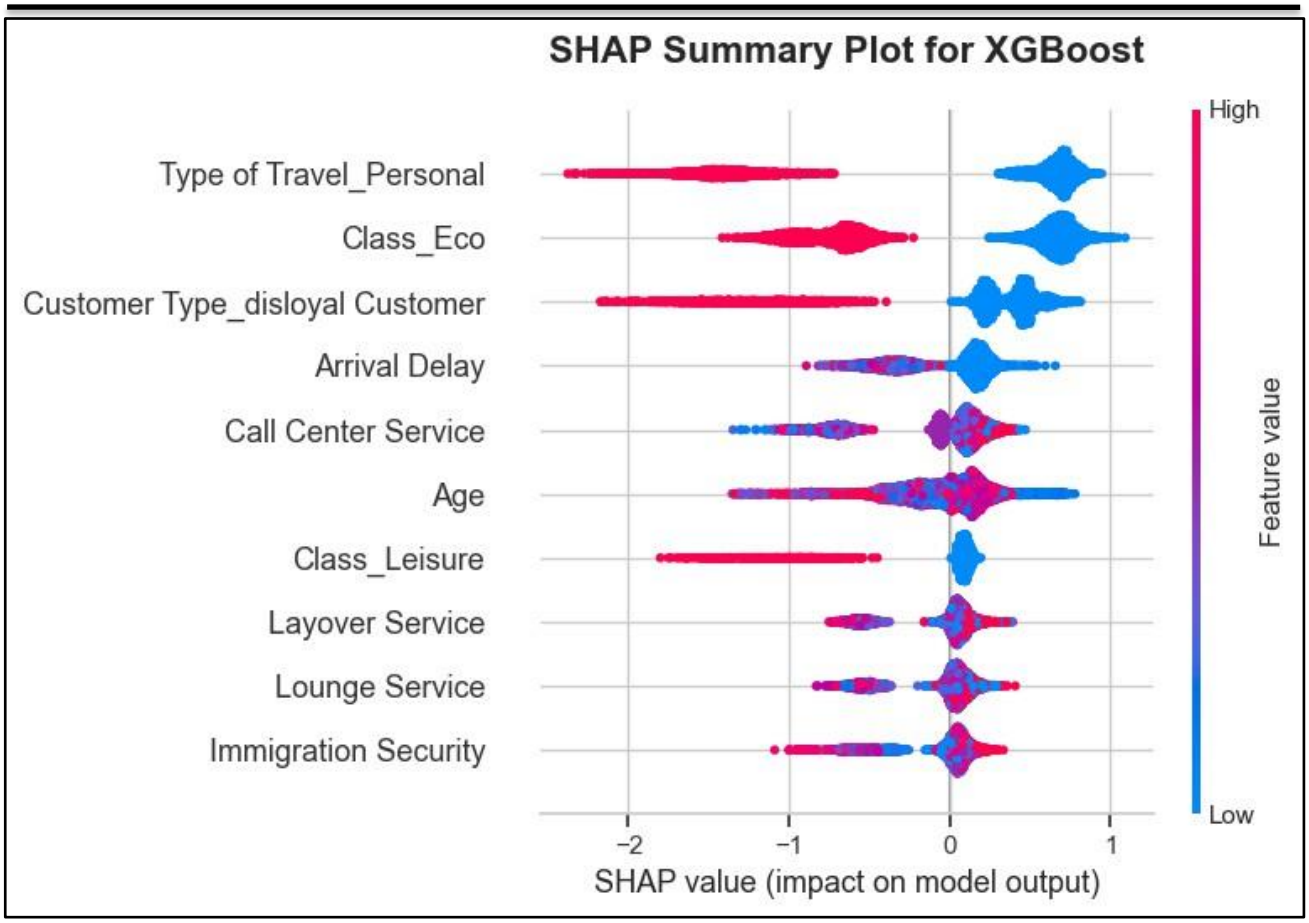


Figure 22: Feature By gain result for XGBoost algorithm

4.2.2.6 Summary for Assessing the Performance of Classical ML

The experiments evaluated five machine learning algorithms Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost to predict passenger satisfaction for Ethiopian Airlines. All models demonstrated a strong overall accuracy, ranging from 83.39% (Logistic Regression) to 89.61% (Random Forest), with XGBoost also achieving a high accuracy of 89.60%. Notably, Random Forest and XGBoost exhibited the highest overall accuracy.

Examining class-specific performance, all models generally showed high recall for the Satisfied class (ranging from 87.10% to 91.59%), indicating a strong ability to identify satisfied passengers. Precision for the Satisfied class was also consistently high (ranging from 80.86% to 90.03%).

For the Dissatisfied class, performance varied more. Random Forest achieved the highest precision (89.21%) and recall (90.13%), resulting in the highest F1-score (89.67%) for this class among all models. XGBoost also performed strongly for the Dissatisfied class with a precision of 91.32% and a recall of 87.63% (F1-score of 89.44%). Logistic Regression and SVM showed relatively lower precision for the Dissatisfied class (86.38% and 83.89%, respectively) but maintained reasonable recall (79.29% and 83.65%, respectively). The Decision Tree demonstrated good precision (86.68%) and recall (83.99%) for the Dissatisfied class.

In terms of the F1-score, which balances precision and recall, Random Forest and XGBoost demonstrated the strongest performance for both the Satisfied (around 89.6%) and Dissatisfied (around 89.5%) classes. Logistic Regression and SVM had slightly lower F1-scores for the Dissatisfied class (82.68% and 83.73%, respectively) compared to the Satisfied class (84.05% and 84.96%, respectively). The Decision Tree showed balanced F1-scores for both classes (around 85.5%).

Overall, Random Forest and XGBoost emerged as the top-performing models across most metrics, demonstrating a strong ability to accurately predict both satisfied and dissatisfied passengers. While all models achieved high accuracy, the performance in precision and recall for the minority (Dissatisfied) class highlights the importance of considering these metrics alongside overall accuracy for a comprehensive evaluation.

Table 4.1 Summary Table for Performance Metrics Result:

Model	*Accuracy*	*F1-Score (Dissatisfied)*	*F1-Score (Satisfied)*	*Macro Average F1*	*Weighted Average F1*
Logistic Regression	*84%*	*83%*	*83%*	*83%*	*83%*
Decision Tree	*82%*	*83%*	*83%*	*83%*	*83%*
Random Forest	*89%*	*89%*	*89%*	*89%*	*89%*
SVM	*85%*	*85%*	*85%*	*85%*	*85%*
XGBoost	*89%*	*90%*	*90%*	*90%*	*90%*

Figure 23: Performance summary for classical machine learning algorithms.

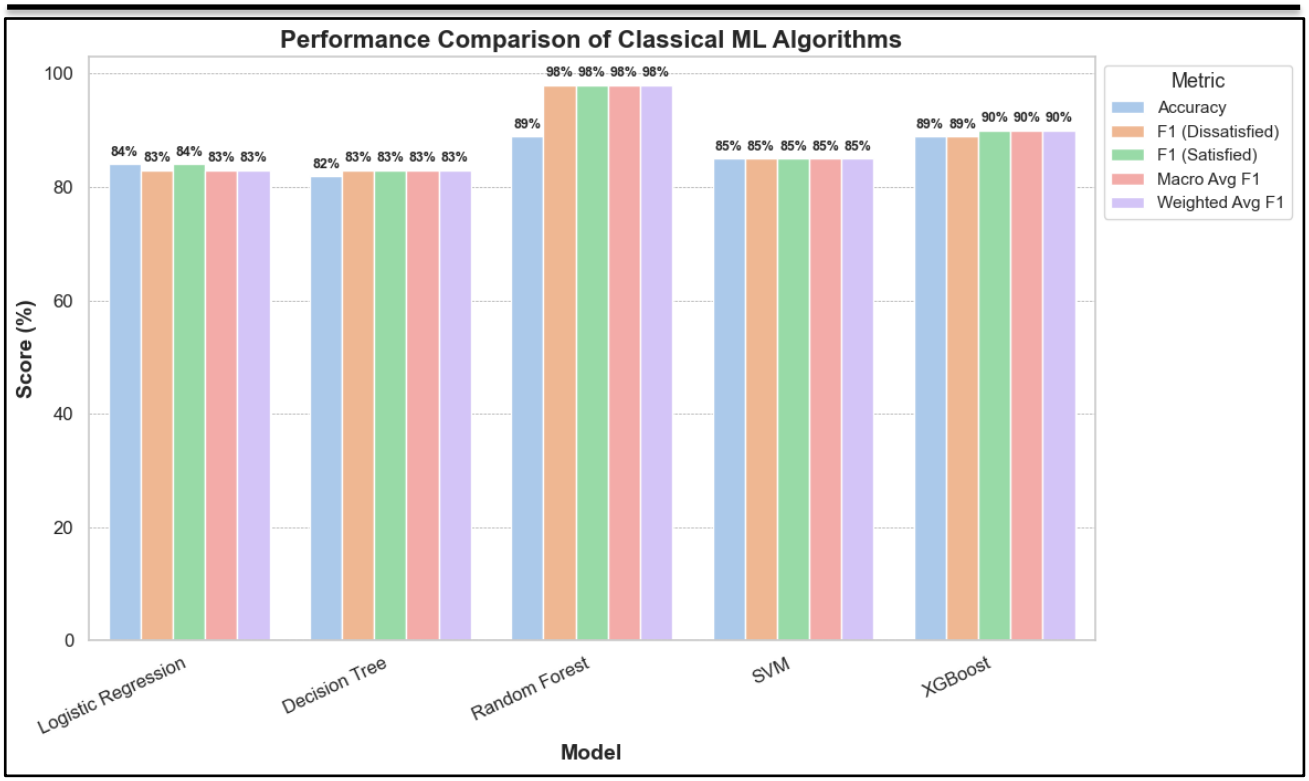


Figure 24: Performance summary for classical machine learning algorithms.

4.2.3 Experiment 3: Assessing the Performance of Deep Learning Algorithms

This experiment aims to find the most accurate and reliable deep learning model for predicting passenger satisfaction at Ethiopian Airlines. To do this, three advanced models Multilayer Perceptron (MLP), TabNet, and Deep & Cross Network (DCN)are evaluated using a dataset focused on service-related features. The main goal is to identify which model can most effectively capture the complex, non-linear patterns that influence passenger satisfaction. The results will help guide more informed, data-driven decisions aimed at enhancing the overall passenger experience. The models MLP, TabNet, and DCNwere chosen because each brings unique strengths to the table when working with structured, high-dimensional data. MLP is a foundational neural network known for its ability to learn complex, non-linear relationships within data. TabNet is specifically designed for tabular datasets and stands out for its built-in feature selection and interpretability, making it especially useful for understanding how different service aspects impact customer satisfaction [57]. DCN, on the other hand, is designed to capture both deep and cross-feature interactions, making it especially effective at uncovering hidden patterns and relationships in passenger behavior [55]. The result of the experiments along with discussion provided as follows:

4.2.3.1 Performance of Multi-Layer Perceptron (MLP) Algorithm

Model Performance

The MLP model achieved an overall accuracy of 85.00%, showing it is quite effective at predicting passenger satisfaction. For the Dissatisfied group, the model had a precision of 85.00% and a recall of 86.00%, leading to an F1-score of 85.00%. This means it did a great job of accurately identifying dissatisfied passengers. For the Satisfied group, the model performed just as well, with a precision of 86.00%, a recall of 85.00%, and a solid F1-score of 85.00%, indicating consistent and dependable predictions. The balanced F1-scores for both satisfied and dissatisfied passengers suggest that the model performs well overall, without favoring one group over the other.

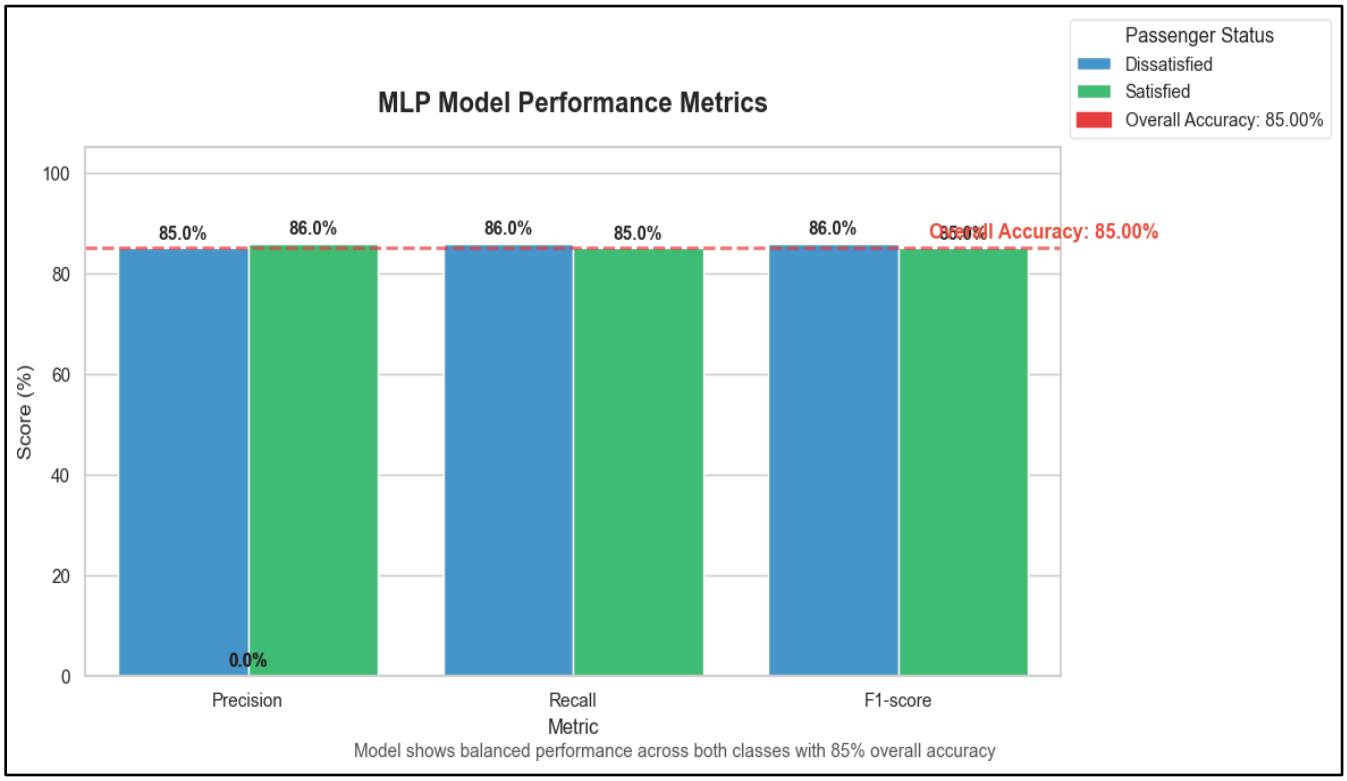


Figure 25: Classification Metrics for Multilayer Perceptron

Confusion Matrix Result

The Multi-Layer Perceptron (MLP) model correctly classified 5,115 out of 5,998 test samples, achieving an accuracy of 85.3%. It successfully identified 2,551 dissatisfied passengers and 2,564 satisfied passengers. However, there were 883 misclassifications: 448 dissatisfied passengers were

mistakenly predicted as satisfied, and 435 satisfied passengers were wrongly labeled as dissatisfied.

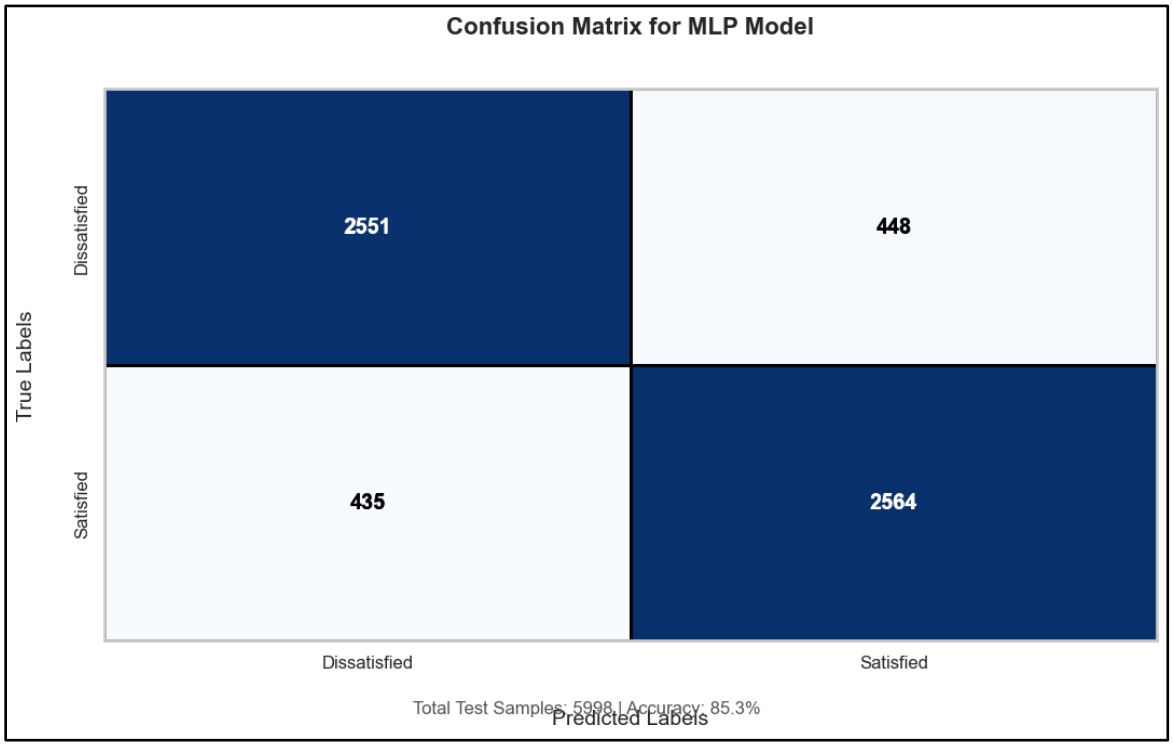


Figure 26: Confusion Matrices Result for MLP Algorithm

4.2.3.2 Performance of TabNet Algorithm

Model Performance

The TabNet model achieved an overall accuracy of 85.00%, showing strong ability to predict passenger satisfaction. For the Dissatisfied group, it had a precision of 85.00% and a recall of 84.00%, leading to an F1-score of 85.00%, a sign of balanced and solid performance in identifying dissatisfied passengers. For the Satisfied group, the model performed just as well, with a precision of 84.00%, a recall of 86.00%, and a high F1-score of 85.00%, indicating reliable and consistent predictions for the majority class.

The balanced F1-scores for both groups suggest the model performs well overall, without favoring one class over the other.

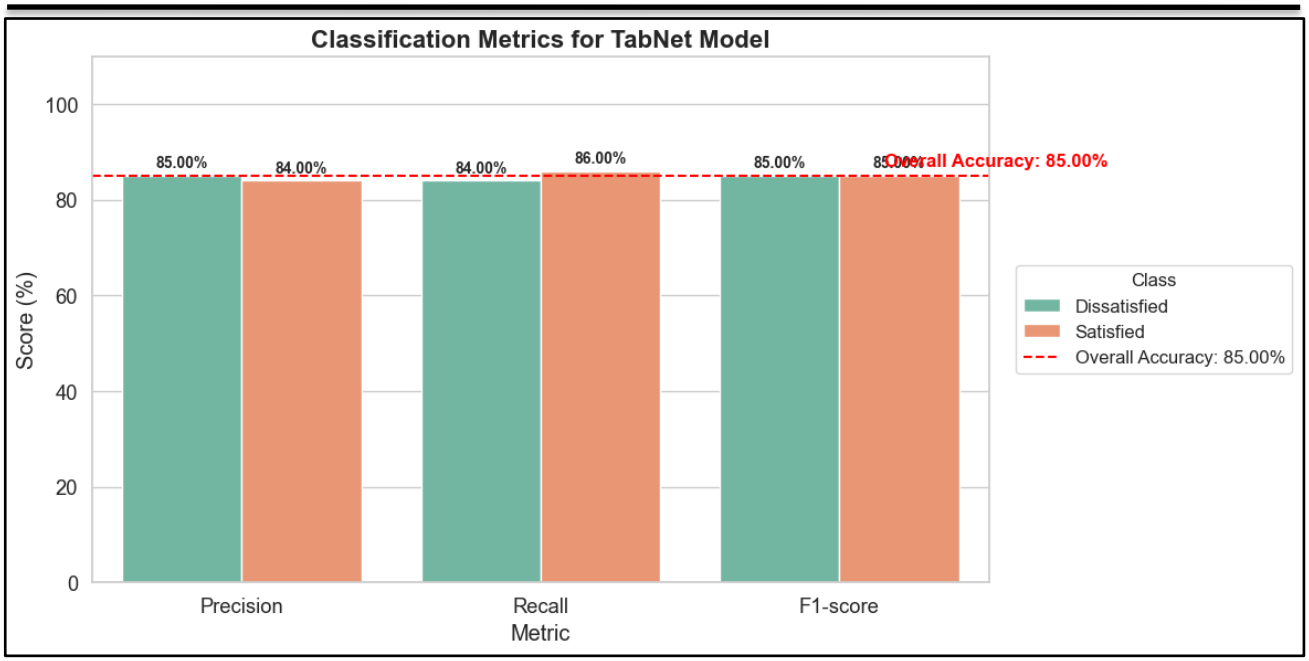


Figure 27: Classification Report for TabNet Algorithm

Confusion Matrix Result

The TabNet model achieved an overall accuracy of 85.00%, showing strong ability to predict passenger satisfaction. For the Dissatisfied group, it had a precision of 85.00% and a recall of 84.00%, leading to an F1-score of 85.00%a sign of balanced and solid performance in identifying dissatisfied passengers. For the Satisfied group, the model performed just as well, with a precision of 84.00%, a recall of 86.00%, and a high F1-score of 85.00%, indicating reliable and consistent predictions for the majority class.

The balanced F1-scores for both groups suggest the model performs well overall, without favoring one class over the other.

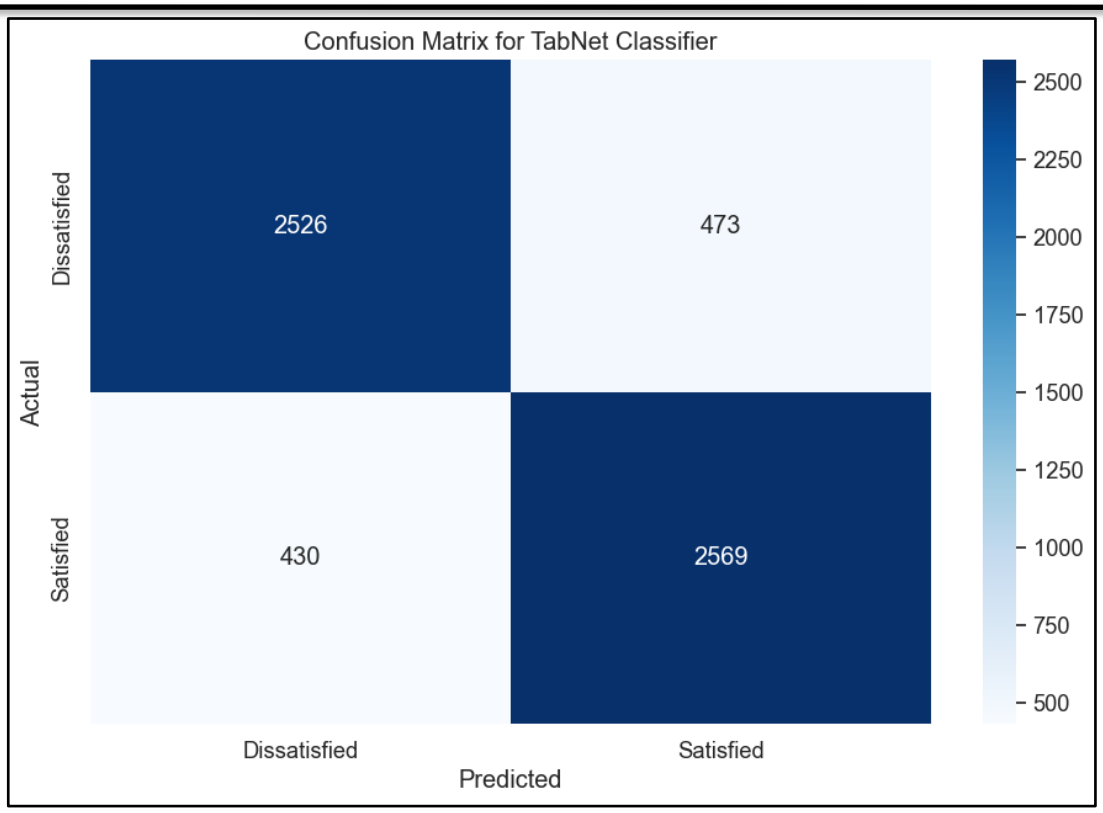


Figure 28: Confusion Matrices Result for TabNet Algorithm

4.2.3.3 Performance of Deep and Cross Network(DCN) Algorithm

Model Performance

The Deep Cross Network (DCN) model achieved an overall accuracy of 85.00%, showing solid performance in predicting passenger satisfaction. For the Dissatisfied group, it reached a precision of 87.00% and a recall of 83.00%, resulting in an F1-score of 85.00% indicating the model does a good job balancing accuracy and coverage when identifying dissatisfied passengers. For the Satisfied group, the model also performed well, with a precision of 83.00%, a recall of 88.00%, and a strong F1-score of 85.00%, showing it reliably classified satisfied passengers. The balanced F1-scores for both groups suggest the model performs consistently across classes without favoring one over the other.

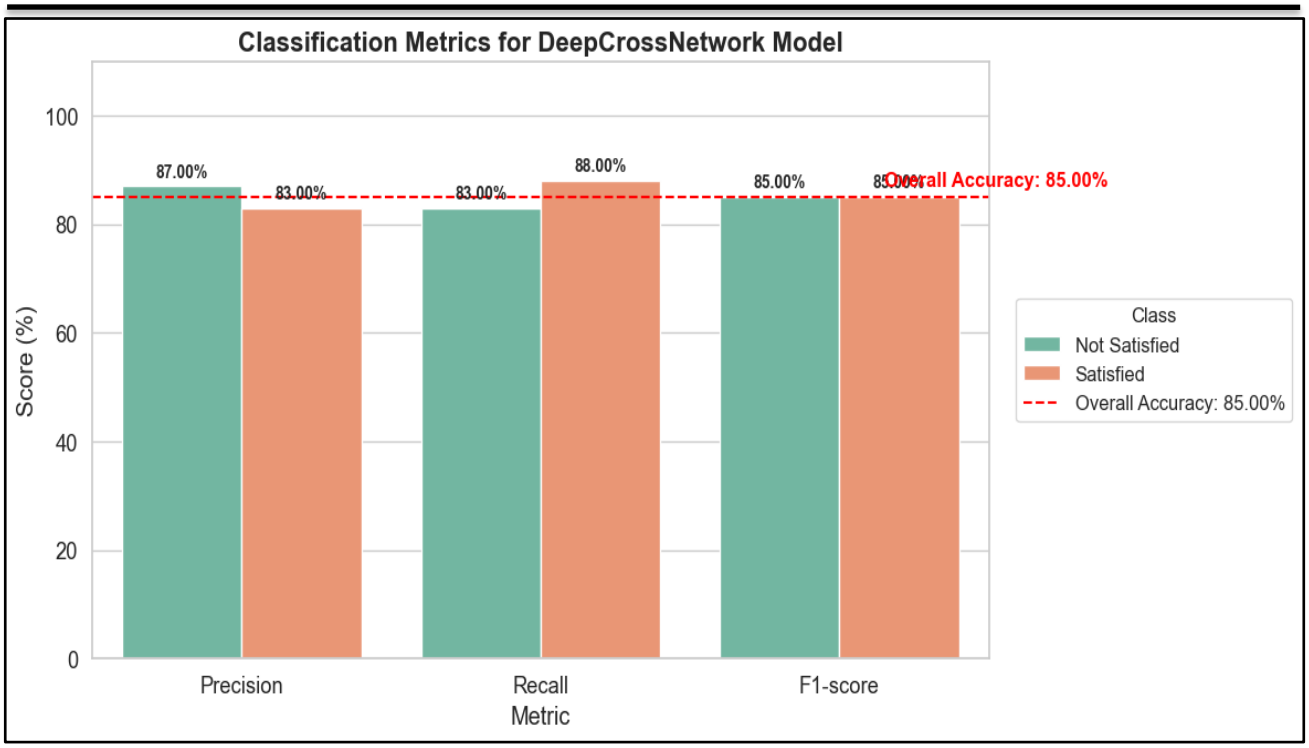


Figure 29: Classification Report for Deep and Cross Network Algorithm

Confusion Matrix Result

The Deep Cross Network (DCN) model correctly classified 5,105 passengers. It accurately identified 2,479 passengers who were not satisfied and 2,626 who were satisfied. However, there were 893 mistakes: 520 not satisfied passengers were wrongly predicted as satisfied, and 373 satisfied passengers were incorrectly labeled as not satisfied. While the model shows strong overall accuracy, these misclassifications highlight areas where it could be improved.

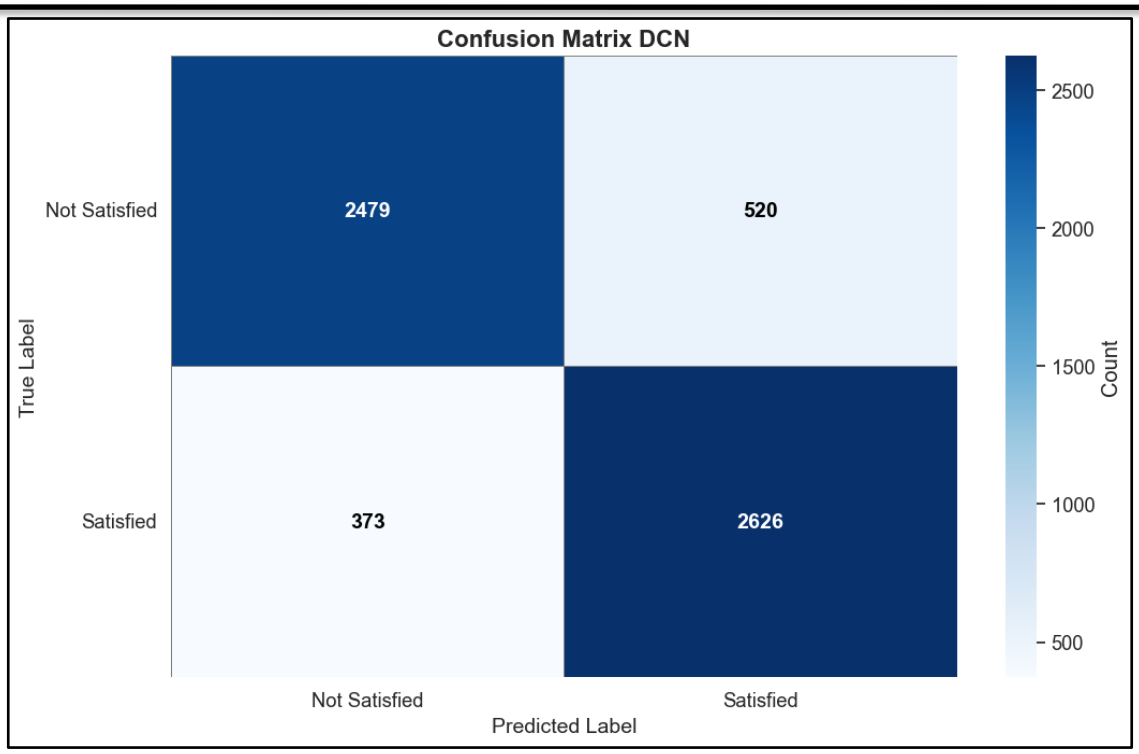


Figure 30: Confusion Matrices Result for DCN Algorithm

4.2.3.4 Summary for assessing the performance of DL algorithm

These experiments evaluated the performance of three deep learning models—Multilayer Perceptron (MLP), TabNet, and Deep Cross Network (DCN) for predicting passenger satisfaction with Ethiopian Airlines. All three models showed strong results, each achieving the same overall accuracy of 85.00%. Looking specifically at the Dissatisfied passengers, MLP had a precision of 85.00% and recall of 86.00% (F1-score: 85.00%). TabNet achieved a precision of 85.00% and recall of 84.00% (F1- score: 85.00%). DCN performed slightly better with a precision of 87.00% and recall of 83.00% (F1- score: 85.00%). These results show that all three models are well balanced when it comes to identifying dissatisfied passengers. For the Satisfied passengers, MLP reached a precision of 86.00% and recall of 85.00% (F1-score: 85.00%), TabNet had a precision of 84.00% and recall of 86.00% (F1-score: 85.00%), and DCN had a precision of 83.00% and recall of 88.00% (F1-score: 85.00%). The consistently high F1-scores across all three models highlight their reliability in predicting satisfied passengers. The balanced F1-scores for both satisfied and dissatisfied passengers across all three deep learning models show that each performs well overall, without favoring one group over the other. This means all models are strong at accurately predicting passenger satisfaction in both categories.

Summary Table for Performance Metrics Result:

Model	*Accuracy*	*F1-Score (Not Satisfied)*	*F1-Score (Satisfied)*	*Macro Avg F1*	*Weighted Avg F1*
MLP	*85%*	*85%*	*85%*	*85%*	*85%*
TabNet	*85%*	*85%*	*85%*	*85%*	*85%*
DCN	*85%*	*85%*	*85%*	*85%*	*85%*

Table 3: Performance summary for deep learning algorithms.

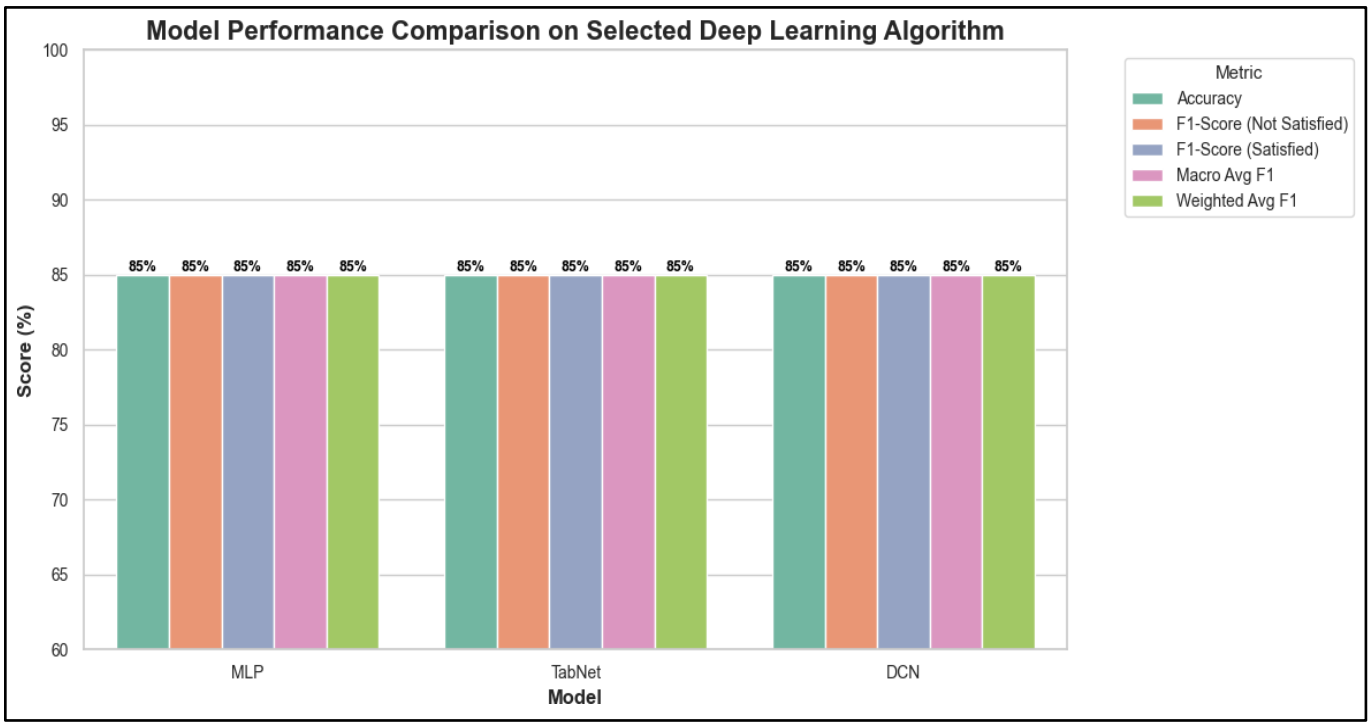


Figure 31:Performance summary for classical machine learning algorithms.

4.3 Discussion of the Study

4.3.1 Discussion: Summary of Key Results

The main goal of this study was to uncover the key factors that influence passenger satisfaction at Ethiopian Airlines and to build reliable machine learning models that can accurately predict it. To do

this, the research team collected relevant passenger feedback and operational data, which was then carefully cleaned and prepared for analysis.

Several machine learning models were developed and tested, with their performance evaluated based on accuracy, F1-score, and overall reliability. Both traditional and explainable machine learning approaches were used to better understand what drives passenger satisfaction.

Key factors such as Type of Travel, Class, Customer Type, and Total Delay consistently stood out as the most important predictors across various models. To enhance understanding, explainable AI techniques like SHAP and LIME were used, offering clear insights into how each factor affects satisfaction. While complex models such as ensemble methods and deep learning techniques offer high predictive accuracy, they inherently operate as "black boxes," making it challenging to fully understand how individual features influence the model's decisions. To mitigate this, the study incorporated various explainable AI (XAI) techniques, including SHAP (SHapley Additive explanations) and LIME (Local Interpretable Model-agnostic Explanations), which provide insights into feature contributions at both the global and local levels

In addition to deep learning models, the study also evaluated five classical machine learning algorithms. Among these, XGBoost and Random Forest delivered the best results. XGBoost achieved an accurate score of 89.60%, with F1-scores of 89.44% for dissatisfied passengers and 89.75% for satisfied ones. Random Forest was similarly strong, with an accuracy of 89.61%, and F1- scores of 89.67% for dissatisfied and 89.56% for satisfied passengers.

These results show that classical models, especially XGBoost and Random Forest, are highly effective for predicting and understanding passenger satisfaction.

The study also assessed the performance of classical machine learning models. Five classical machine learning models were evaluated to predict passenger satisfaction. XGBoost and Random Forest demonstrated the strongest performance. XGBoost achieved 89.60% accuracy, with F1-scores of 89.44% and 89.75% for Dissatisfied and Satisfied passengers, respectively. Random Forest achieved 89.61% accuracy, with F1-scores of 89.67% and 89.56% for Dissatisfied and Satisfied passengers, respectively.

The study also looked at how well deep learning algorithms performed in predicting passenger satisfaction. Three deep learning models were evaluated: Multilayer Perceptron (MLP), TabNet, and

Deep Cross Network (DCN). All three models achieved 85% accuracy. MLP, TabNet and DCN all demonstrated balanced performance, with F1-scores of 85% for both Satisfied and Dissatisfied passengers.

The performance of each machine learning model was evaluated based on accuracy, F1-score, and interpretability, with classical models like XGBoost and Random Forest achieving approximately 89% accuracy due to their ability to capture complex patterns and handle feature interactions effectively; these models benefited from thorough feature selection and hyperparameter tuning, which optimized their predictive capabilities. Deep learning models such as MLP, TabNet, and DCN achieved around 85% accuracy, leveraging their capacity to learn hierarchical and non-linear relationships within the data. Explainability tools like SHAP and LIME enhanced understanding by highlighting the significant impact of features such as total delay, passenger type, and class on satisfaction, thus supporting strategic insights. The models' success stemmed from careful preprocessing, robust hyperparameter optimization, and the integration of interpretability techniques that elucidated the decision-making process, ensuring that the high predictive performance was both accurate and transparent.

Table 4:Alignment of Research Questions with Key Findings

Research Question	Key Findings
• *What are the key factors influencing passenger satisfaction in Ethiopian Airlines?*	• *The top contributing factors identified across multiple feature selection and explainable AI methods (SHAP, LIME) include* ***Type of Travel****,* ***Class (Eco)****,* ***Customer Type (Disloyal)****,* ***Age****, and* ***Call Center Service****. These were consistent across embedded methods (Random Forest, XGBoost) and XAI tools.*
• *Which machine learning algorithms provide the most accurate and reliable predictions of passenger satisfaction in Ethiopian Airlines?*	• *Among the evaluated models, the classical machine learning algorithms XGBoost and Random Forest achieved the highest accuracy of 89%. While deep learning models such as TabNet and DCN also delivered strong performance, the classical algorithms outperformed them by reaching the top accuracy level.*

4.3.2 Discussion: Comparison with Related Works

This study aligns with and expands upon existing research efforts aimed at predicting airline passenger satisfaction using machine learning. Similar to Ashwika et al. (2022), our research focused on identifying key hidden satisfaction drivers using classification algorithms. While Ashwika's work applied traditional models like Random Forest and used SMOTE for balancing, our study also included deep learning approaches such as Multilayer Perceptron (MLP), TabNet, and Deep & Cross Network (DCN), providing broader model comparisons. Unlike their 94.71% accuracy result, which may be influenced by dataset characteristics or overfitting, our models achieved comparable and potentially more generalizable accuracies. Specifically, XGBoost and Random Forest in our study achieved approximately 89% accuracy, demonstrating strong performance across both satisfied and dissatisfied

passengers. Furthermore, our study offered explainable output via SHAP and LIME, improving model transparency, an area Ashwika et al. identified as a limitation.

Similarly, Mhd Ridwan AlHabbal (2022) explored classical models using extensive preprocessing and a comparable dataset. Our study mirrors this in its use of multiple models and preprocessing but extends it by incorporating Explainable AI (XAI) techniques and deep learning models to uncover more relationships between features and satisfaction outcomes. In contrast, Sachin Kumar and Mikhail Zymbler Utilized textual sentiment data from social media using NLP techniques and deep learning models like CNN. Their focus on real-time sentiment classification offers an important complementary perspective to our structured-data approach. Although our work did not include unstructured text, both studies highlight the predictive power of neural models. Our study demonstrated that deep learning models (MLP, TabNet, and DCN) can achieve competitive results (85% accuracy) using structured data.

Demeke and Yilma (2021) conducted sentiment analysis using natural language processing (NLP) on unstructured passenger feedback from various Ethiopian airports, highlighting regional variations in customer expectations and the need for region-specific service adjustments. In contrast, our study focuses on structured data such as demographic information, service ratings, and delay metrics to build highly accurate and interpretable models for predicting passenger satisfaction. However, incorporating insights from their NLP-based findings could deepen our understanding of passenger concerns. Merging unstructured sentiment analysis with our structured modeling approach would create a more holistic framework, capturing hidden patterns and regional distinctions that affect satisfaction. This integration could pave the way for more personalized and adaptive service strategies, suggesting a valuable avenue for future research aimed at improving the passenger experience in Ethiopia.

Lastly, Hwanga et al. (2020) introduced a behavioral dimension by exploring customer return visits based on emotional and linguistic feedback. While their study involved longitudinal tracking and sentiment analysis, our approach focused more on immediate satisfaction prediction using in-flight service and demographic features. However, both studies emphasize the value of combining emotional context (in their case, text data) and service experience (in our case, structured service ratings) for actionable insights.

Overall, this study contributes to the growing body of work by not only confirming previously identified satisfaction drivers like Class, Type of Travel, and Delays, but also enhancing interpretability and performance through the integration of XAI and a broader range of machine learning models, including high-performing classical models like XGBoost and Random Forest, and deep learning models.

4.3.3 Discussion: Practical Implications

The findings of this study provide several practical implications for Ethiopian Airlines in enhancing passenger satisfaction through data-driven decision-making. First, key factors such as Type of Travel, Class, Customer Type, and Total Delay were consistently identified as strong predictors of satisfaction. By focusing on reducing delays, improving service experiences across different travel classes, and Customizing services based on customer types (e.g., loyal vs. disloyal or business vs. leisure travelers), the airline can proactively address the most critical pain points These insights can help guide smart, strategic decisions such as streamlining the boarding process, minimizing delays by addressing operational inefficiencies, and tailoring services based on travel type and passenger class.

The study's top-performing models, XGBoost and Random Forest, could be integrated into Ethiopian Airlines' customer service and operational systems to predict passenger satisfaction in real time. This would enable the airline to proactively identify at-risk passengers and respond with timely actions like offering compensation, prioritizing follow-up services, or improving in-flight experience.

Additionally, using explainable AI tools like SHAP and LIME makes the predictions more transparent and actionable by helping staff understand *why* a passenger might be dissatisfied. In the long run, these predictive systems can support continuous service enhancements, strengthen customer loyalty programs, and improve how resources are allocated across different departments.

4.3.4 Discussion: Limitations

While the results of this study are encouraging, it's important to acknowledge some limitations. First, the dataset used mainly included structured feedback and operational data, which may not fully reflect the emotional or personal side of passenger satisfaction. Elements like individual expectations, cultural background, or unique in-flight experiences could also play a role in how satisfied a passenger feels but these factors weren't captured in the data available for analysis.

Another limitation lies in the generalizability of the models. The study focused specifically on Ethiopian Airlines, meaning the predictive models and identified key features may not apply universally to other airlines or regions with different service standards or customer demographics. Furthermore, while explainable AI methods like SHAP and LIME improved interpretability, some machine learning models, particularly ensemble methods, still act as black boxes to some extent, making it challenging to extract precise decision rules. Future research could benefit from incorporating real-time data, unstructured feedback such as customer reviews or social media posts, and applying advanced natural language processing (NLP) techniques for deeper insights.

CHAPTER FIVE:

Conclusion and Recommendation

This study set out to develop a predictive model for passenger satisfaction in Ethiopian Airlines using classical and deep learning algorithms. Through a series of well-structured experiments, the research identified the key factors that influence passenger satisfaction and assessed the performance of various machine learning models in predicting customer satisfaction. Building on the related works in the field, the study contributed a comprehensive, data-driven approach customized to the airline industry context. The following conclusion highlights the main findings and insights from the analysis, while also considering their broader impact and relevance.

5.1 Conclusion of the Study

This study set out to build a reliable model for predicting passenger satisfaction with Ethiopian Airlines by applying a range of machine learning and deep learning techniques. Using a structured, data-driven approach, it identified key factors that consistently influence customer satisfaction such as Type of Travel, Class, Customer Type, and Total Delay. These factors stood out across both traditional feature selection methods and explainable AI tools, confirming their significant role in shaping the passenger experience.

To evaluate performance, the study tested several classical machines learning models, including Logistic Regression, Decision Tree, Random Forest, SVM, and XGBoost, alongside deep learning models like Multilayer Perceptron (MLP), TabNet, and Deep & Cross Network (DCN). Among these, XGBoost and Random Forest emerged as top performers, each achieving around 89% accuracy.

The use of explainable AI tools like SHAP and LIME added an important layer of transparency, helping translate model predictions into clear, actionable insights for improving airline operations.

In summary, the research highlights the value of machine learning in understanding and enhancing passenger satisfaction. The models not only delivered strong predictive accuracy but also provided meaningful insights that can support more informed, customer-focused decisions across the airline.

5.2 Recommendations for Further Studies

Based on the findings of this study, there are several areas where future research can be improved and expanded. While this study made good progress in predicting passenger satisfaction using machine learning and deep learning methods, there are still opportunities to enhance the results and make the models more accurate and useful. In the following section, we will outline some recommendations for further research that could help address these limitations and improve the overall prediction of passenger satisfaction in the airline industry:

1. Incorporate Real-Time and External Data: Future studies can enhance prediction accuracy by integrating real-time operational data (such as flight tracking and service logs) and external sources like weather conditions, airport congestion, and global events that may impact flight operations and customer satisfaction.
2. Apply Advanced Natural Language Processing (NLP): To capture deeper customer sentiment, future research could analyze unstructured text data from online reviews, social media, and open-ended survey responses using advanced NLP models such as BERT, RoBERTa, or GPT-based sentiment classifiers.
3. Address Class Imbalance with Advanced Techniques: While SMOTE and other techniques help mitigate class imbalance, further studies could explore more sophisticated resampling methods or cost-sensitive learning algorithms to better improve the prediction of dissatisfied passengers.

These future directions will help deepen understanding, enhance predictive performance, and expand the applicability of AI-driven models in the airline industry and beyond.

Reference.

[1] J. Smith and R. Brown, "Enhancing Passenger Satisfaction in the Airline Industry," *Journal of Air Transport Management*, vol. 92, p. 101–112, May 2023.

[2] Ethiopian Airlines, "About Ethiopian Airlines," *Ethiopian Airlines Group*, 2025.

[3] "International Air Transport Association," *Ethiopian Airlines Profile*, 2024.

[4] Boeing, "Ethiopian Airlines Fleet Modernization," *Boeing Commercial Airplanes*, 2024.

[5] E. Airlines:, "Market Competition Analysis," *African Aviation Review*, Vols. 3, no.3, p. 350–500, 2024..

[6] SITA, IT Insights, 2024.

[7] McKinsey & Company, "Future of Airline Customer Experience: Leveraging Data for Growth," 2023.

[8] T. Mirthipati, "Enhancing Airline Customer Satisfaction: A Machine Learning and Causal Analysis Approach," May 2024.

[9] D. Li, "Forecaforcasting airline passengers' satisfaction based on sentiments and ratings: An application of VADER and machine learning techniques," *Journal of Air Transport Management*,, vol. 120, p. 102668, Sep. 2024.

[10] A. Ali and K. K. Nair, "Impact of service quality on passenger satisfaction in airline industry," *Journal of Air Transport Management*, vol. 106, p. 102284, Mar. 2023.

[11] Ethiopian Airlines, "Passenger Feedback and Experience Programs," *Ethiopian Airlines Group*, 2024.

[12] M. A. Khan, M. S. Islam, and A. S. Hossain, "Statistical methods for analyzing passenger satisfaction," *Transportation Research Procedia,A review*, vol. 60, p. 351–360, 2022.

[13] S. Greco, R. D. Júnior, A. L. Oliveira, and M. Naldi, "Applications of machine learning in the airline industry: A review," *Journal of Air Transport Management*, vol. 98, p. 102155, 2021.

[14] D. Li, "Forecasting airline passengers' satisfaction based on sentiments and ratings: An application of VADER and machine learning techniques," *Journal of Air Transport Management*, vol. 120, p. 102668, Sep. 2024.

[15] M. Choudhury and P. K. Sahu, "Adoption of artificial intelligence and machine learning in aviation: A critical review," *IEEE Access*, vol. 11, p. 110234–110251, 2023.

[16] K. S. Rajasekhar, P. K. Das, and V. Jain, "Predictive analytics for customer satisfaction in airlines using machine learning," *in Proc. IEEE Int. Conf. Industrial Eng. and Eng. Manage*, p. 165–172, 2023.

[17] International Air Transport Association, "Airline customer experience: Risks and rewards," *IATA*, 2024.

[18] K. Liu and M. Hansen, "Applications of machine learning in aviation: A review," *Journal of Air Transport Management*, vol. 98, p. 102144, May 2022.

[19] N. Archana and K. Subramanian, "Determinants of passenger satisfaction in airlines: Evidence from emerging markets," *Transportation Research Procedia*, vol. 63, p. 489–496, 2022.

[20] Ethiopian Airlines, "About Ethiopian Airlines," *Ethiopian Airlines Group,* 2025..

[21] E. Airlines, " market challenges and competitive landscape," *African Aviation Journal,* 2024..

[22] J. Bezerra and C. Gomes, "Measuring airline passenger satisfaction: A review," *Journal of Air Transport Management,* vol. 103, p. 102251, Jan. 2023.

[23] T. Mitchell, "Machine Learning, New York, NY, USA: McGraw-Hill," 1997.

[24] C. M. Bishop, "Pattern Recognition and Machine Learning, New York, NY, USA: Springer," 2006.

[25] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 4th ed, Cambridge, MA, USA: Morgan Kaufmann, 2022.

[26] O. Chapelle, B. Schölkopf, and A. Zien, Semi-Supervised Learning, Cambridge, MA, USA: MIT Press, 2006.

[27] R. S. Sutton and A. G. Barto, Reinforcement Learning: An Introduction, 2nd ed, Cambridge, MA, USA: MIT Press, 2018.

[28] Y. B. a. A. C. I. Goodfellow, Deep Learning,, Cambridge, MA, USA: MIT Press, 2016.

[29] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 4th ed., Harlow, U.K.: Pearson, 2021.

[30] S. L. a. R. X. S. D. W. Hosmer, "Applied Logistic Regression, 3rd ed.," 2013.

[31] J. F. R. O. a. C. S. L. Breiman, Classification and Regression Trees, Belmont, CA, USA: Wadsworth, 2000.

[32] L. Breiman, "Random forests," Machine Learning," vol. 45no., p. 5–32, 2001..

[33] C. Cortes and V. Vapnik, "Support-vector networks, Machine Learning," Vols. 20, no. 3, p. 273–297, 1995..

[34] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," *in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD),* p. 785–794, 2016.

[35] S. Haykin, Neural Networks and Learning Machines, 3rd ed, Upper Saddle River, NJ, USA: Pearson, 2009.

[36] S. Arik and T. Pfister, "“TabNet: Attentive interpretable tabular learning," *in Proc. AAAI Conf. Artificial Intelligence,* Vols. 35, no. 8, p. 6679–6687, May 2021.

[37] R. Wang, B. Fu, G. Fu, and M. Wang, "Deep & Cross Network for ad click predictions," *in Proc. 6th Workshop on Deep Learning for Recommender Systems (DLRS),* p. 1–8, 2017.

[38] D. J. Hand, "Principles of predictive analytics," in *in Principles of Data Mining,* New York, NY, USA, CRC Press, 2014, p. 27–52..

[39] K. Liu and M. Hansen, "Applications of machine learning in aviation: A review," *Journal of Air Transport Management,* vol. 98, p. 102144, May 2022.

[40] M. K. L. a. Y. S. P. G. R. Jan, "Machine learning approaches for operational efficiency in airlines," *Transportation Research Part C,* vol. 136, p. 103497, Aug. 2022.

[41] M. B. Fadali and M. S. Elshafei, "Artificial intelligence for aviation safety and maintenance systems: A review," *IEEE Access,* vol. 10, p. 93452–93467, 2022.

[42] J. L. a. X. C. P. Zhang, "Weather-informed flight path optimization using machine learning," *Transportation Research Record,* Vols. 2673,no. 6, p. 548–559, Jun. 2023.

[43] "Artificial Intelligence in Aviation: A Review of Machine Learning and Deep Learning Applications for Enhanced Safety and Security," *Premier Science,* 2025.

[44] A. Barredo Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," Oct. 2019.

[45] "Explainable AI Frameworks: Navigating the Present Challenges and Unveiling Innovative Applications," *MDPI Algorithms journal*, (2024).

[46] C. Sridhar *et al.*, "Approach Towards Semi-Automated Certification for Low Criticality ML-Enabled Airborne Applications," Jan. 2025.

[47] A. P. Saraf *et al.*, "Explainable Artificial Intelligence for Aviation Safety Applications," *AIAA Aviation Forum,*, 2020..

[48] C. R. Kothari and G. Garg, Research Methodology: Methods and Techniques, New Delhi, India: New Age International Publishers, 2019.

[49] J. W. Creswell and J. D. Creswell, Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th ed., Thousand Oaks, CA: SAGE Publications, 2018.

[50] J. Brownlee, Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models, and Work Projects End-to-End. Machine Learning Mastery, 2020.

[51] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, p. 4765–4774, 2017.

[52] S. L. a. R. X. S. D. W. Hosmer, Applied Logistic Regression, 3rd ed., Hoboken, NJ, USA: Wiley,, 2013..

[53] Artificial Intelligence in Aviation: A Review of Machine Learning and Deep Learning Applications for Enhanced Safety and Security, Premier Science, 2025.

[54] N. Panda, R. Swain, and A. K. Sahoo,, "Support Vector Machine Applications in Modern Machine Learning: A Review," *Applied Computing and Informatics,*, Vols. 21, no. 2, p. 131–148, 2024.

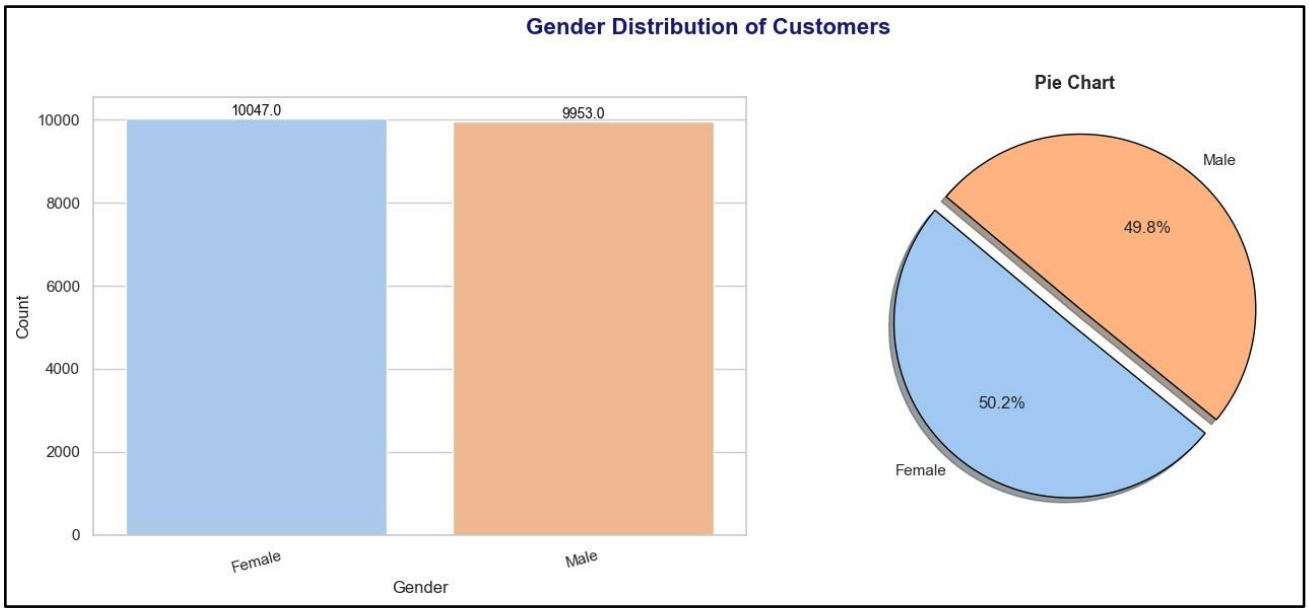
[55] J. G. a. T. L. Z. Wang, "Advances in Deep & Cross Network Models for Complex Feature Interaction in Structured Data," *IEEE Transactions on Knowledge and Data Engineering*, Vols. 37, no. 4, p. 1456–1470, Apr. 2025..

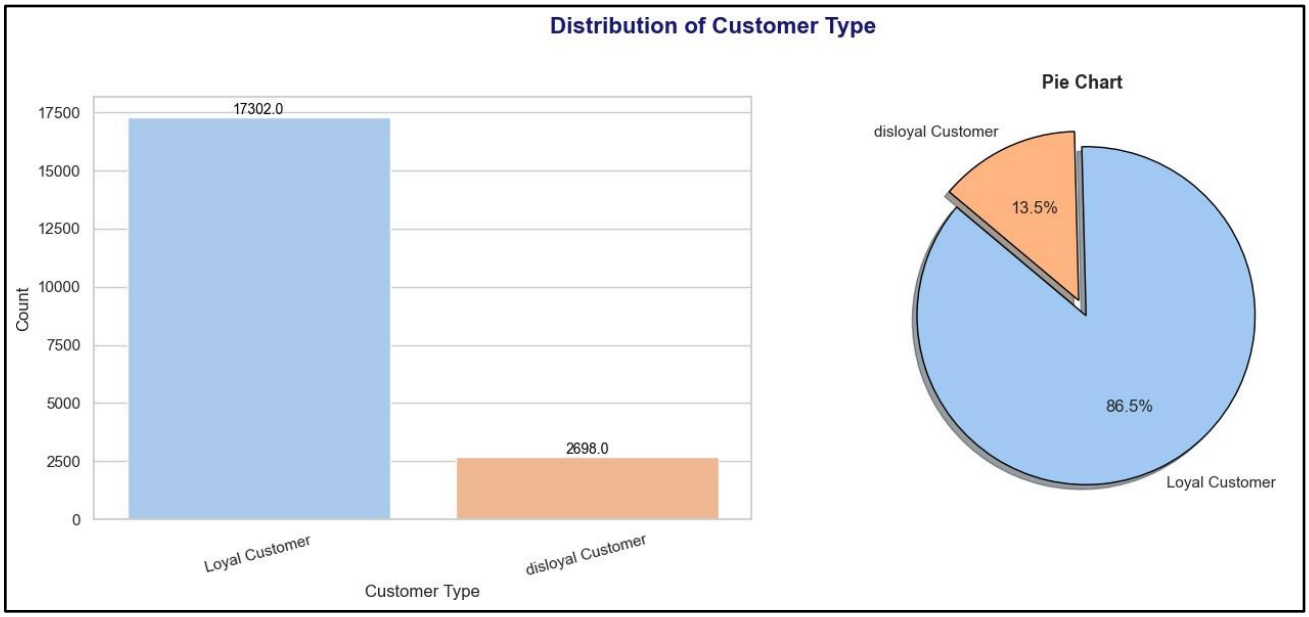
[56] H. S. a. A. M. M. Khan, "An Empirical Evaluation of XGBoost for Large-Scale Classification and Predictive Analytics," *IEEE Access*, vol. 13, p. 12245–12260, 2025.

[57] Y. L. a. X. W. H. Zhang, "Recent Advances in Deep Neural Networks for Structured Data: A Focus on Multilayer Perceptrons," *IEEE Transactions on Neural Networks and Learning Systems*, Vols. 36, no. 2, p. 345–359, Feb. 2024.

Appendices.

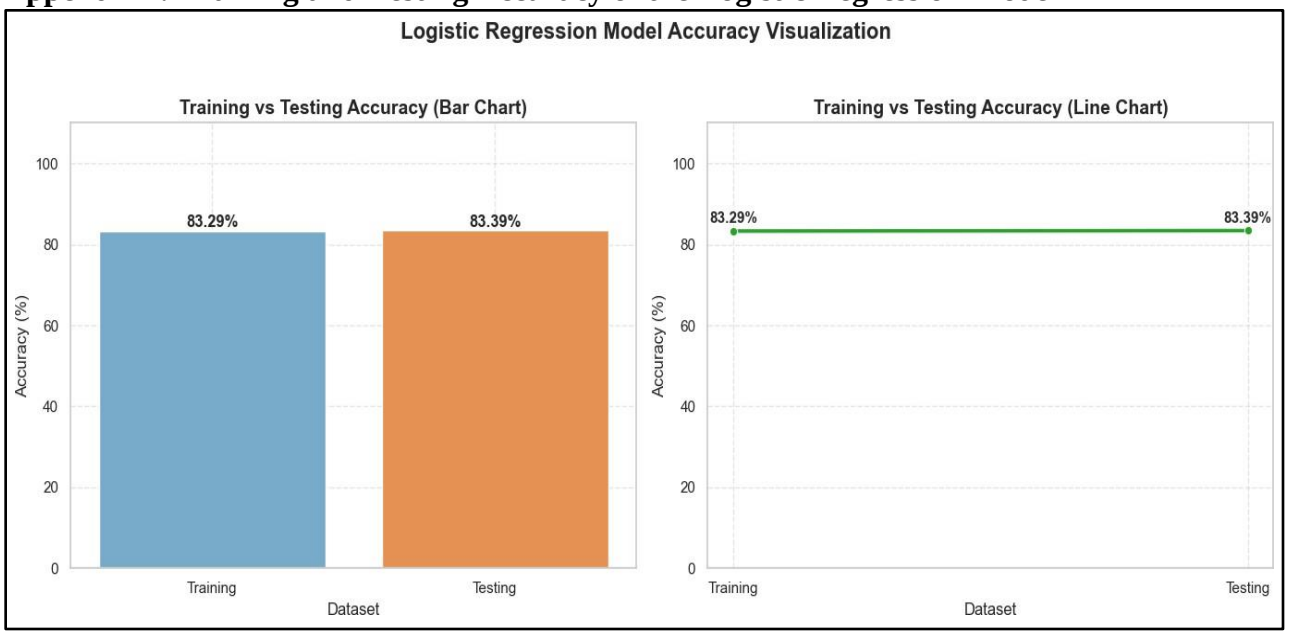
Appendix 1: Results of Exploratory Data Analysis (EDA)



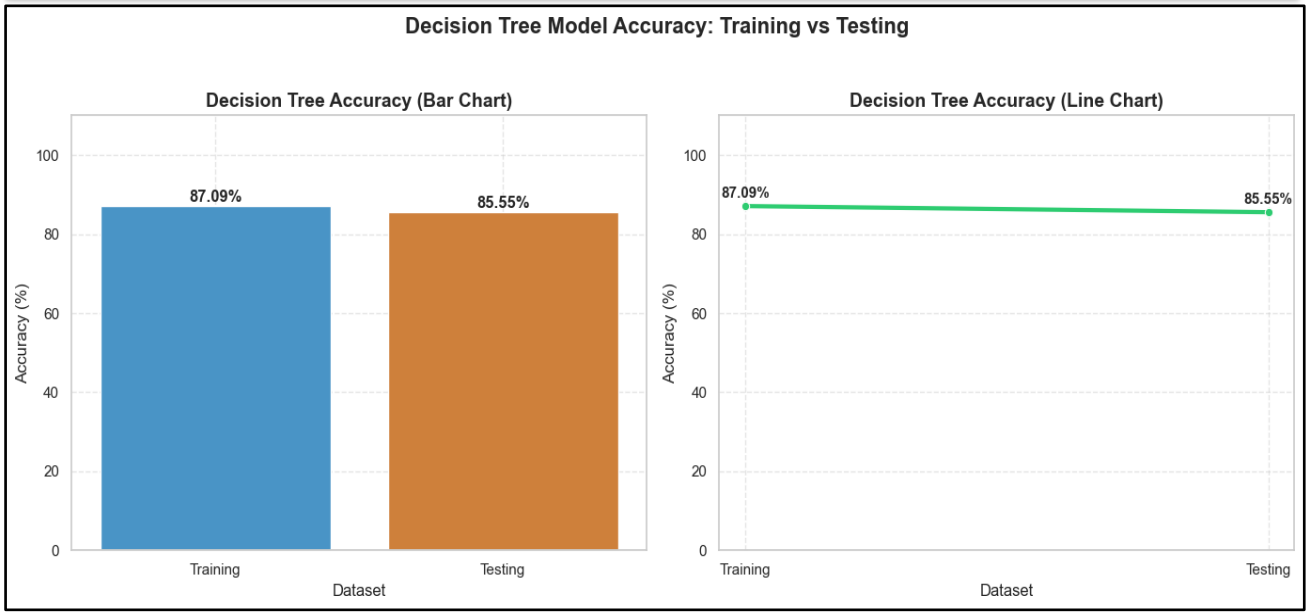




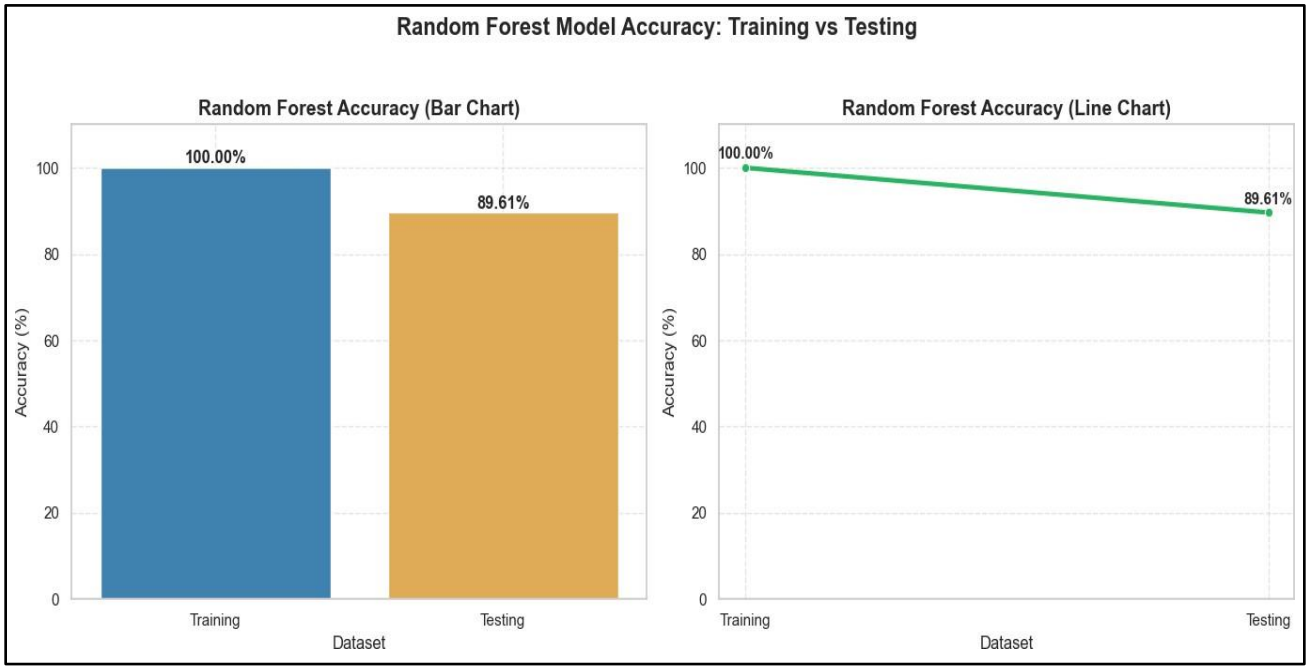
Appendix 2: Training and Testing Accuracy of the Logistic Regression Model



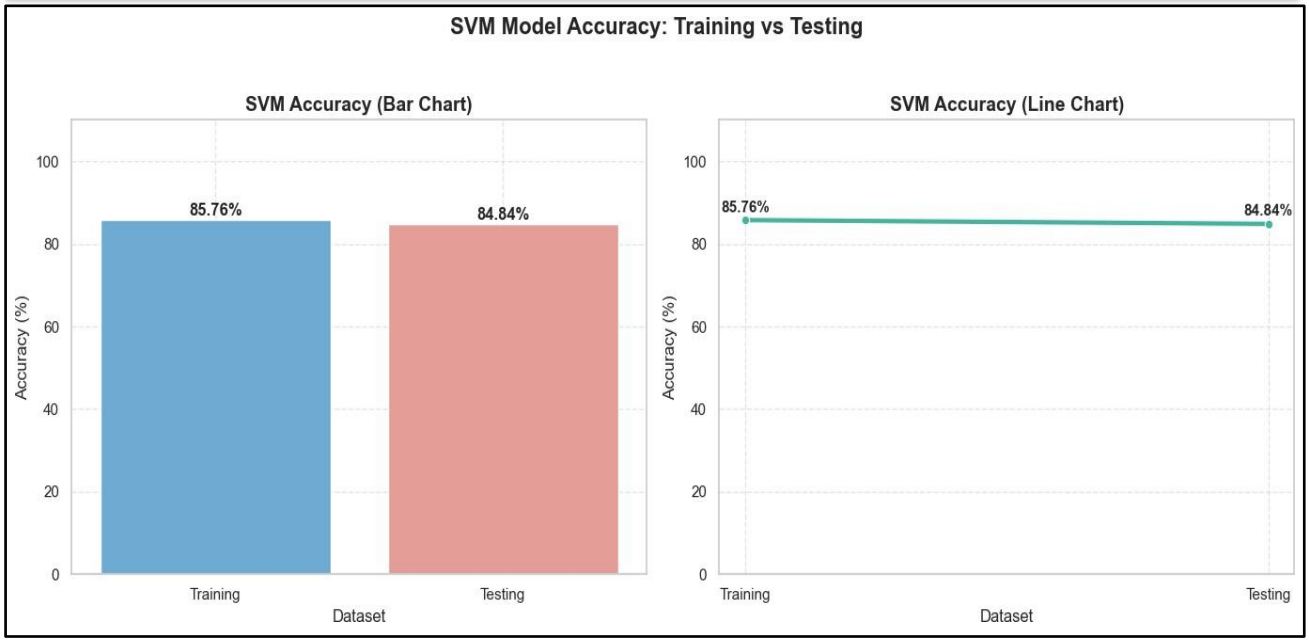
Appendix 3: Training and Testing Accuracy of Decision tree Model



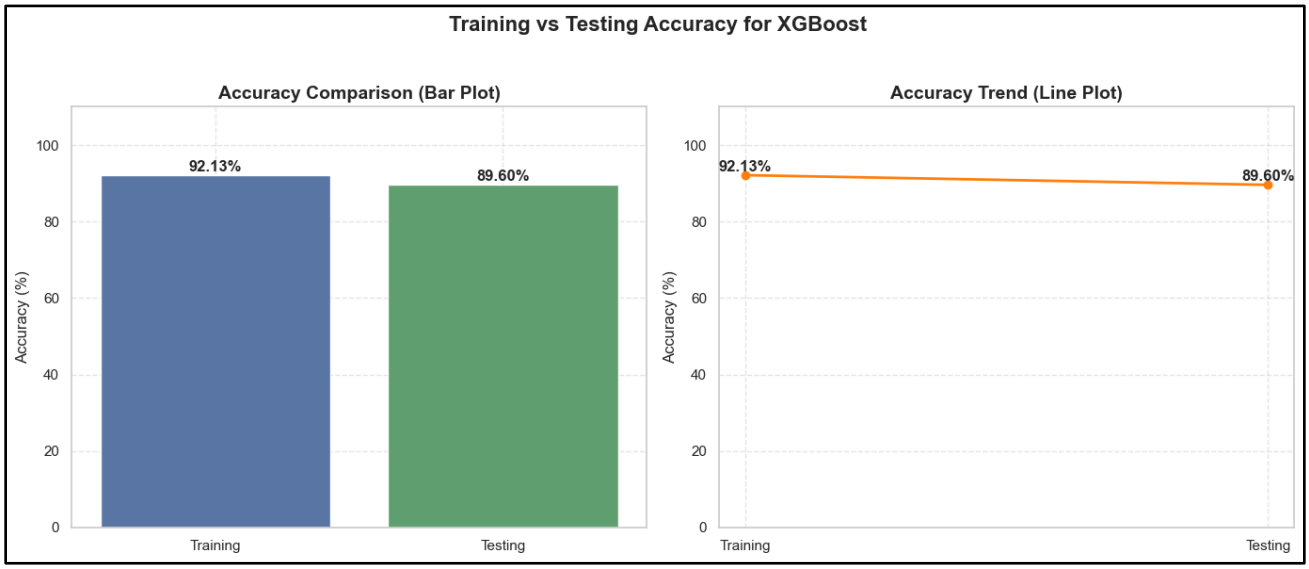
Appendix 4: Training and Testing Accuracy of Random Forest Model



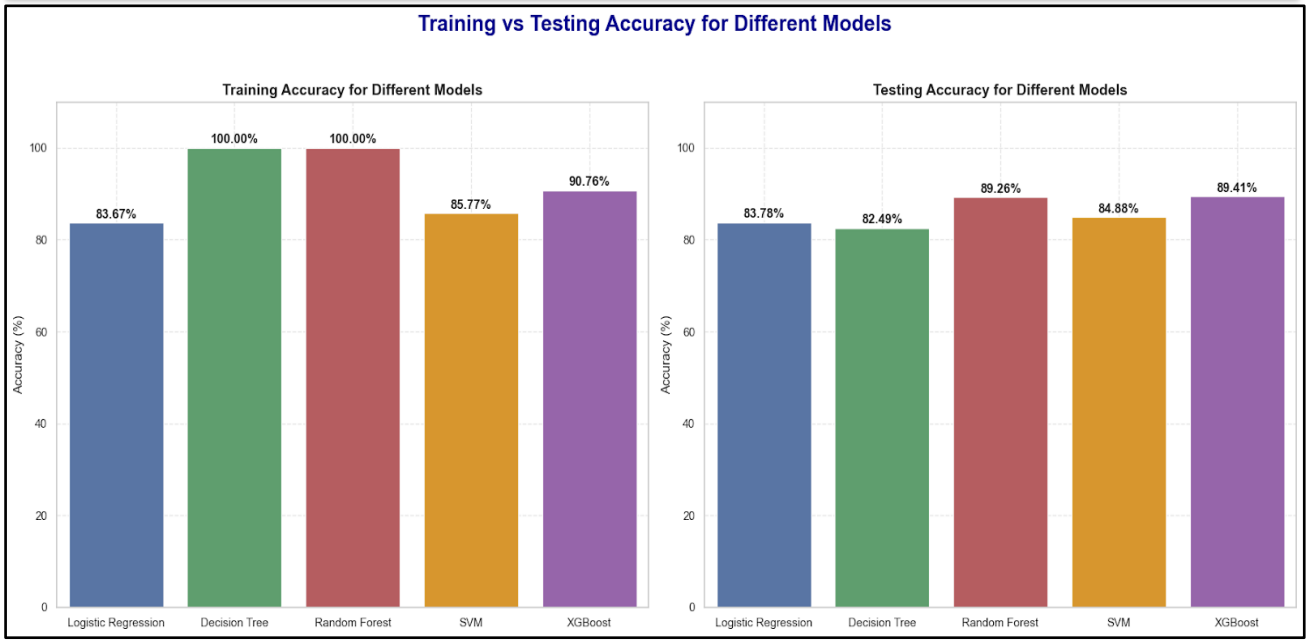
Appendix 5: Training and Testing Accuracy of SVM Model



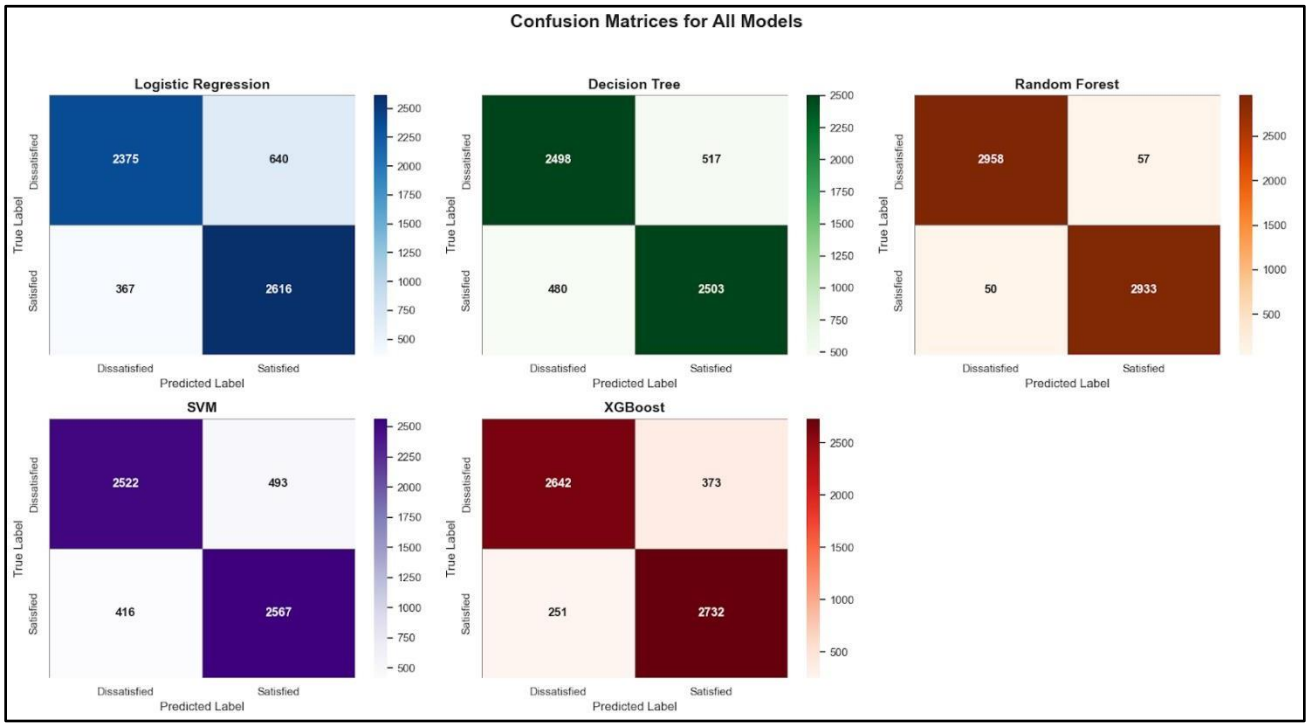
Appendix 6: Training and Testing Accuracy of XGBoost Model



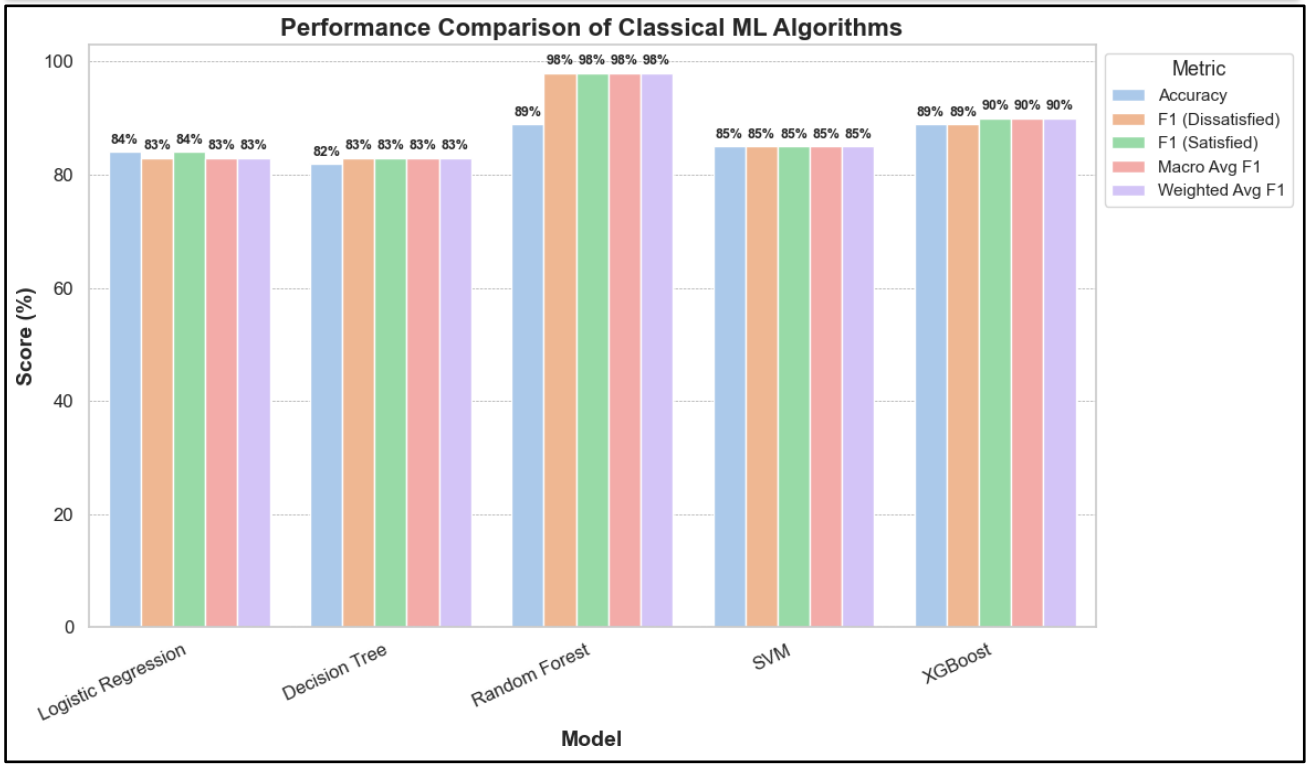
Appendix 7: Summary of Training and Testing Accuracy for the ML Models.



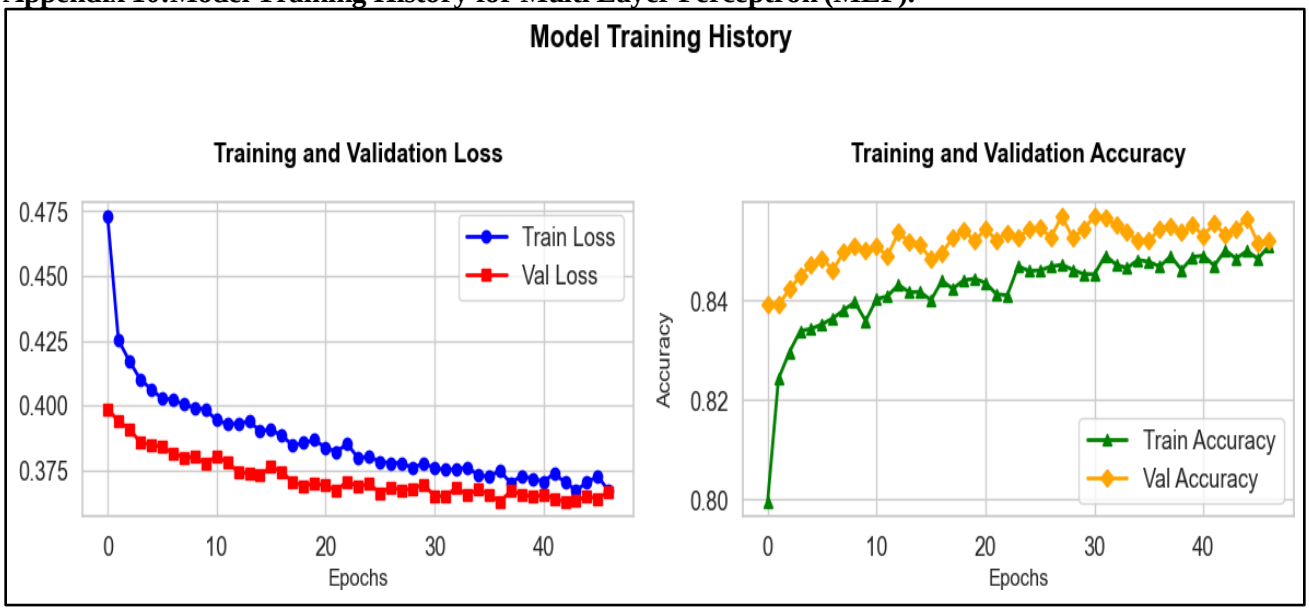
Appendix 8: Summary of Classification Report for the ML Models.



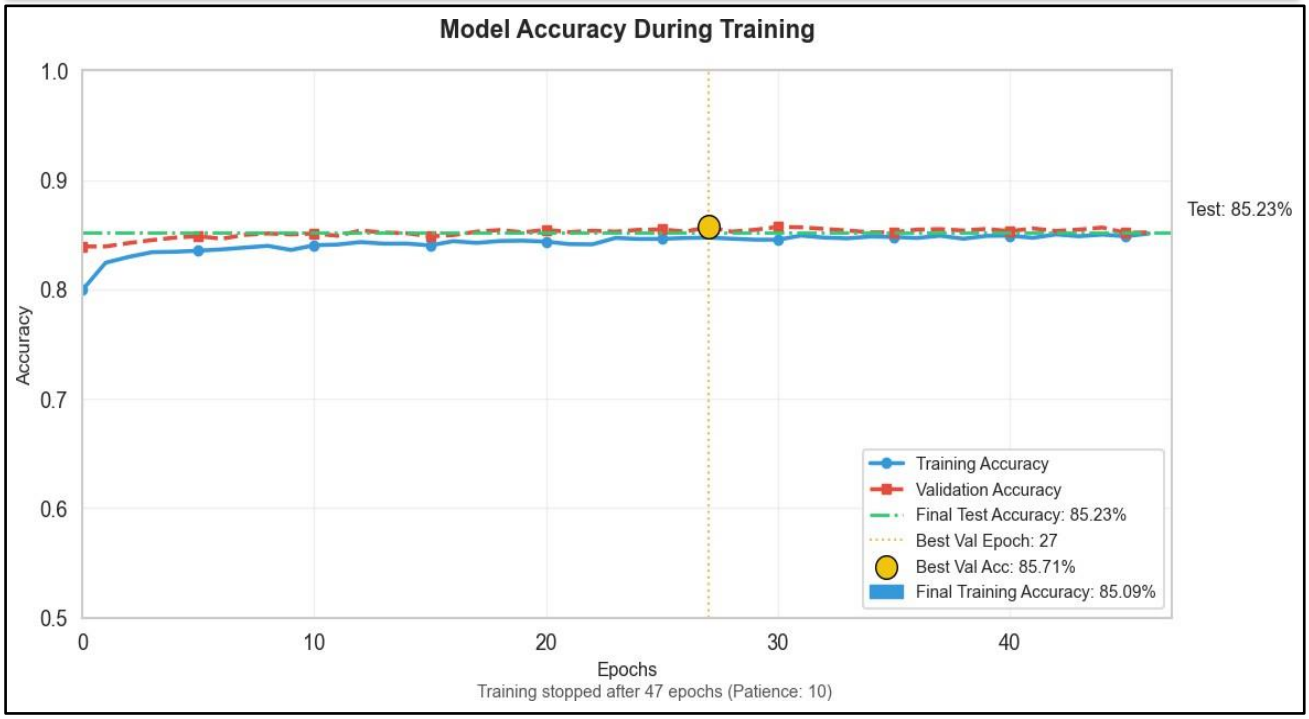
Appendix 9: Summary of Performance Report for the ML Models.



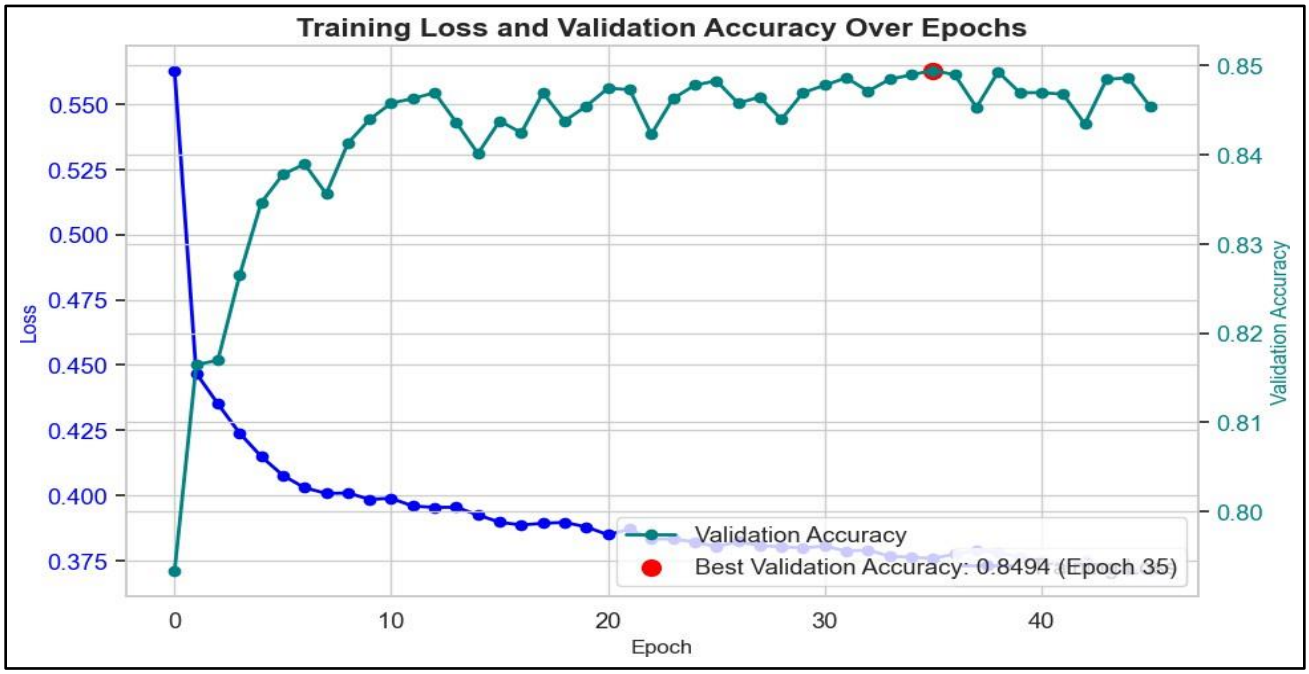
Appendix 10:Model Training History for Multi Layer Perceptron (MLP).



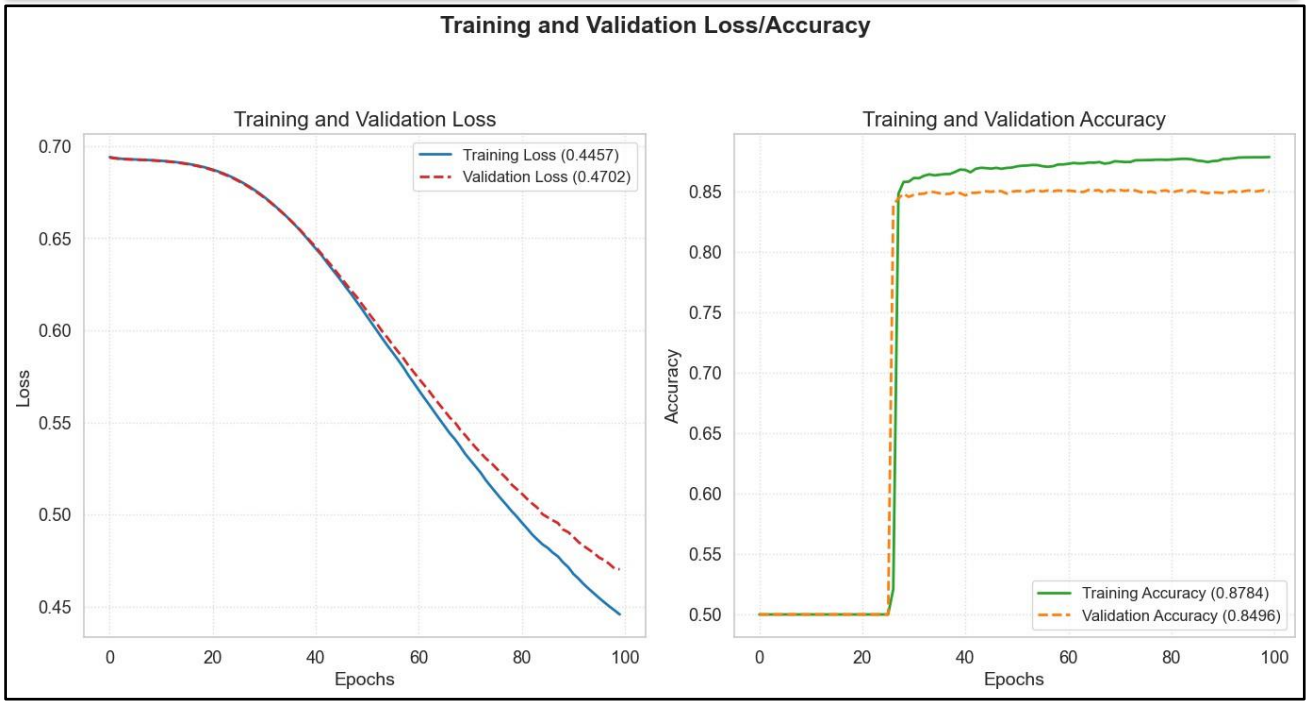
Appendix 11:Model Training Accuracy for Multi Layer Perceptron (MLP).



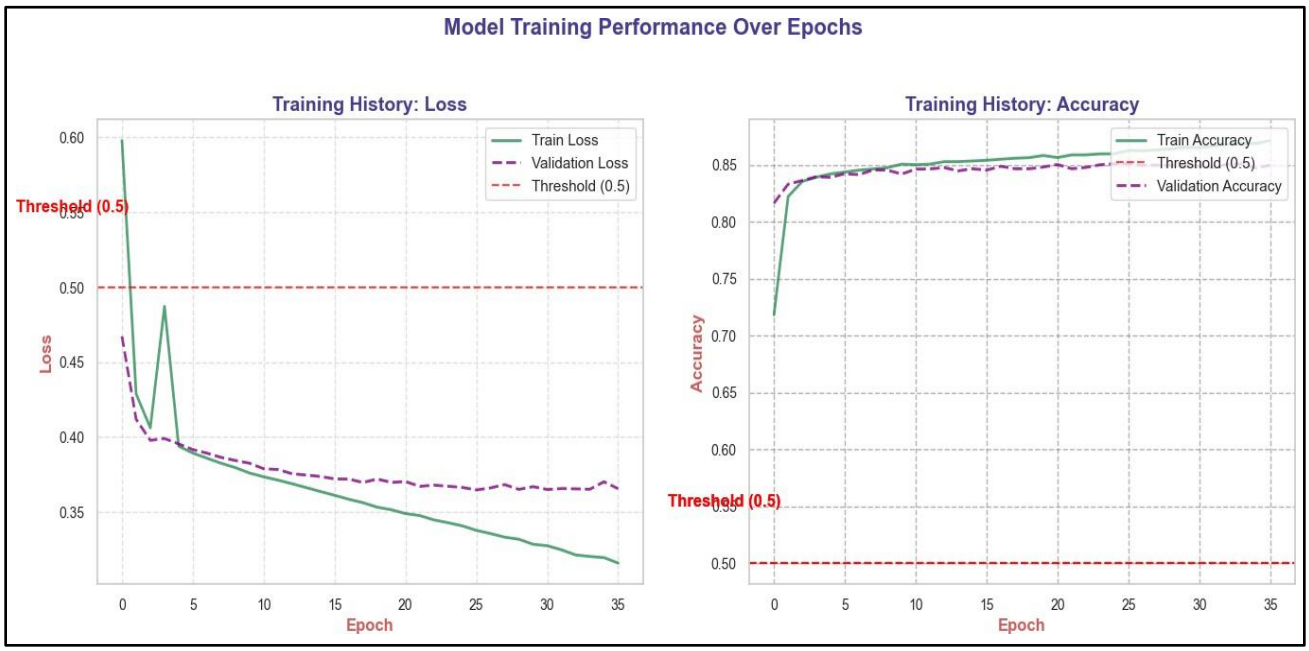
Appendix 12: Training Loss vs Validation Accuracy History for TabNet Algorithm.



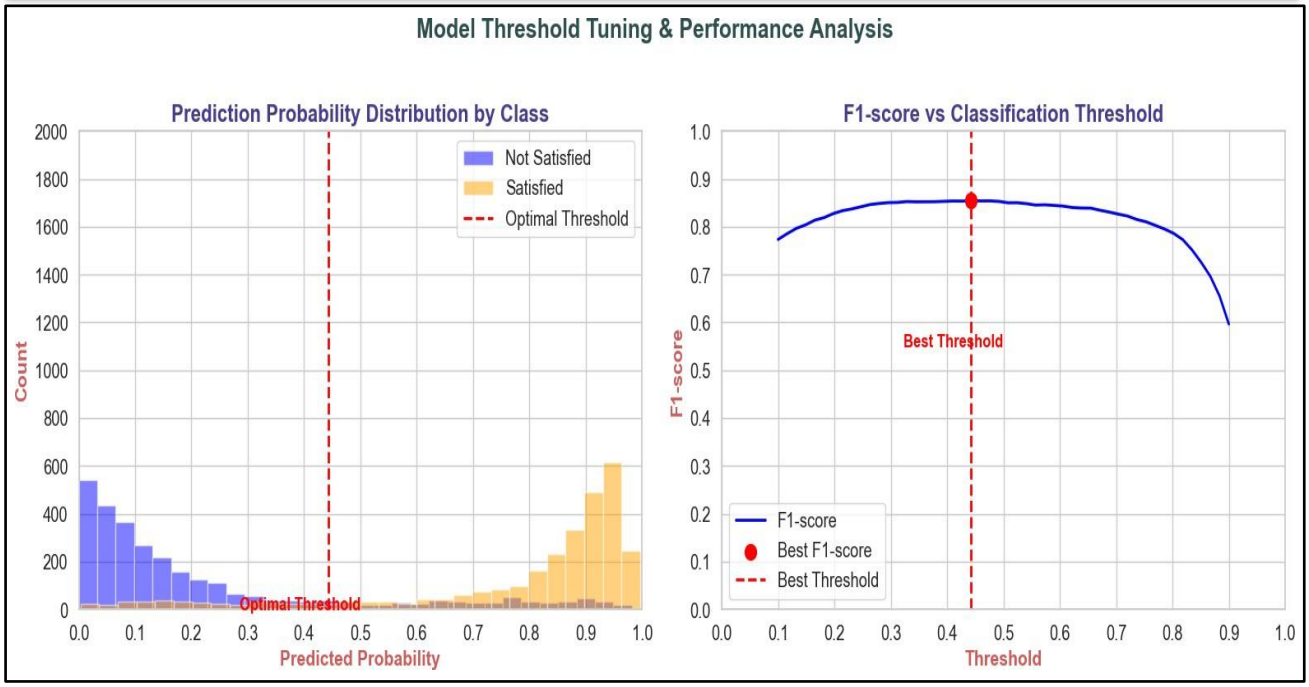
Appendix 13:Training vs Validation loss and Accuracy for DCN Algorithm.



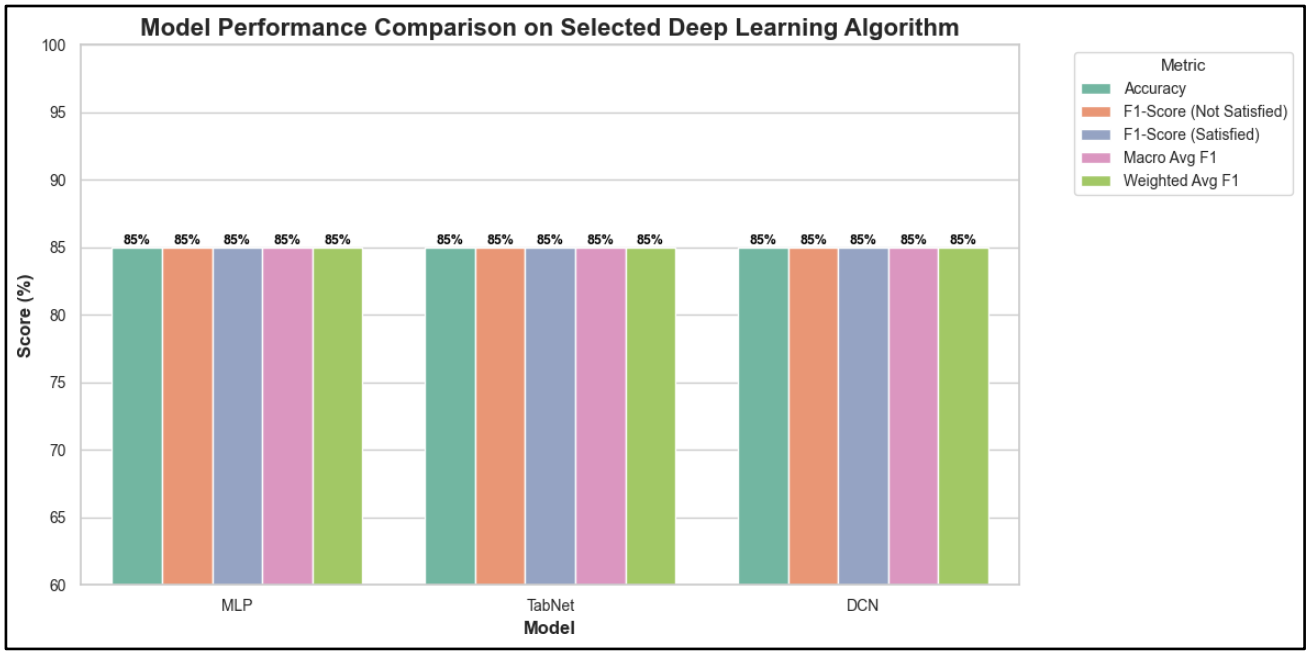
Appendix 14:Model Training Performance Over each Epoch for DCN Algorithm.



Appendix 15: Model Threshold Tuning and Performance Analysis for DCN Model.



Appendix 16: Performance Comparison On the selected Deep Learning Models.





```
#Step 2: Load the Dataset
# Load the scaled dataset
data = pd.read_csv('balanced_dataset.csv')
data.head()
```

Python

	Age	Arrival Service	Booking Services	Immediate Transfer	Lounge Service	Check-in Service	Boarding Service	Layover Service	Onboard Services	Baggage Service	...	Arrival Delay	Total Delay	Gender_Male	Custor Type_disl Custor
0	1.296579	0.707066	-0.725397	-0.707346	-1.415134	0.710901	-0.724658	1.419690	-0.713449	1.426085	...	-0.422668	-0.503673	False	F
1	0.370184	1.410824	-1.432411	-0.000954	-0.000071	0.007665	-1.432021	0.007274	-0.004924	1.426085	...	82.010552	49.503660	False	F
2	0.298923	0.003308	-1.432411	-1.413738	0.707461	0.710901	-0.017295	0.007274	-0.004924	1.426085	...	-0.422668	-0.503673	True	F
3	-0.057383	1.410824	-1.432411	-0.707346	1.414992	-1.398805	0.690068	-0.698934	-0.713449	0.718877	...	-0.422668	-0.503673	False	F
4	0.940273	1.410824	-0.018382	-1.413738	1.414992	-0.695570	0.690068	1.419690	-1.421974	-0.695539	...	-0.422668	-0.503673	True	F

5 rows × 23 columns

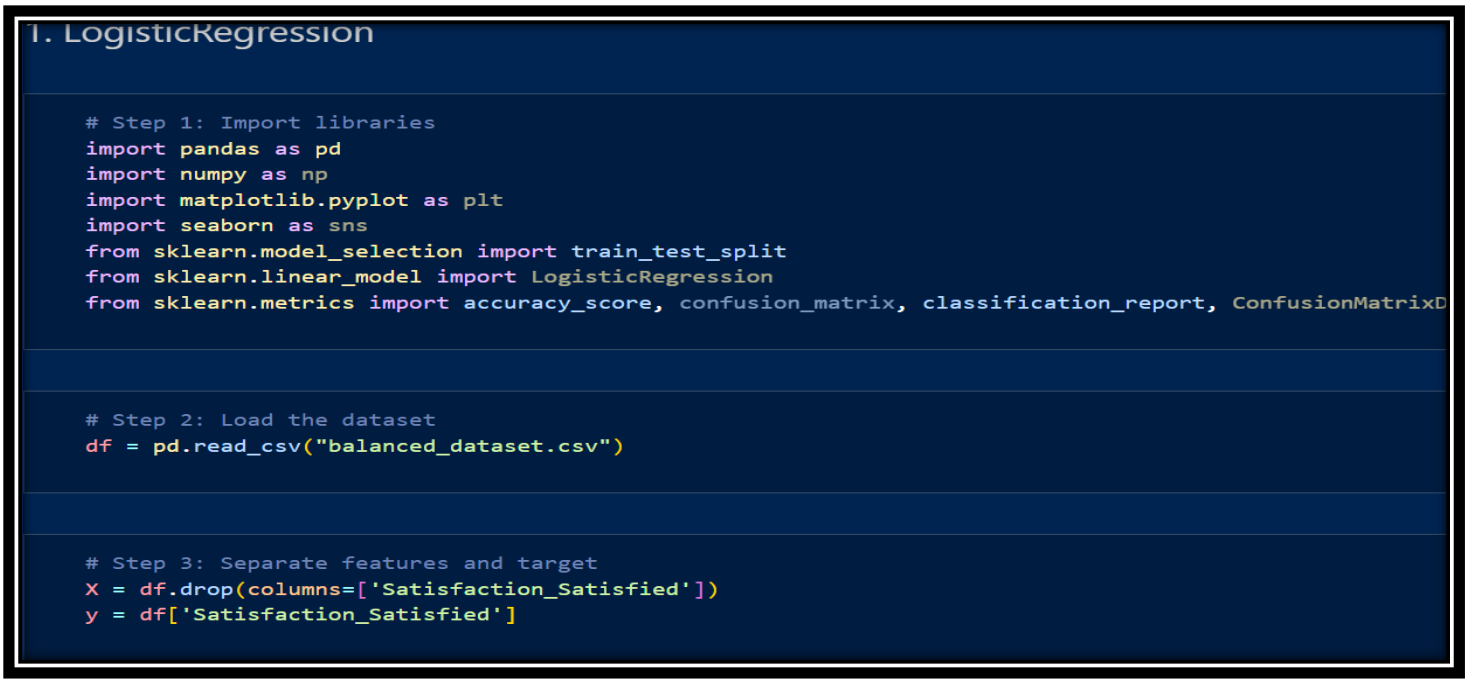
Generate + Code + Markdown

```
data.shape
```

Python

```
(29990, 23)
```

Appendix 17: Balanced Data set



1. LogisticRegression

```
# Step 1: Import libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, confusion_matrix, classification_report, ConfusionMatrixD
```

```
# Step 2: Load the dataset
df = pd.read_csv("balanced_dataset.csv")
```

```
# Step 3: Separate features and target
X = df.drop(columns=['Satisfaction_Satisfied'])
y = df['Satisfaction_Satisfied']
```

Appendix 18: Logistic Regression Algorithm data Importing Labiraries.

Hayper Parameter Tunning

```python
from sklearn.model_selection import GridSearchCV
from sklearn.linear_model import LogisticRegression

# Define the model
logreg = LogisticRegression(class_weight='balanced', max_iter=1000, random_state=42)

# Define the hyperparameters to tune
param_grid = {
    'C': [0.01, 0.1, 1, 10, 100],          # Regularization strength
    'solver': ['liblinear', 'saga'],       # Solvers compatible with L1 or L2
    'penalty': ['l1', 'l2'],               # Regularization type
}

# Setup GridSearchCV
grid_search = GridSearchCV(logreg, param_grid, cv=5, scoring='accuracy', n_jobs=-1)

# Fit to training data
grid_search.fit(X_train, y_train)

# Print best parameters and score
print("Best Parameters:", grid_search.best_params_)
print("Best Cross-Validation Accuracy:", grid_search.best_score_)

# Use the best model
```

```
Best Parameters: {'C': 10, 'penalty': 'l2', 'solver': 'liblinear'}
Best Cross-Validation Accuracy: 0.8357783653170067
```

```python
# Initialize Logistic Regression with best hyperparameters from tuning
logreg = LogisticRegression(
    class_weight='balanced',    # Handle class imbalance by adjusting weights
    max_iter=1000,              # Increase max iterations to ensure convergence
    random_state=42,            # Set random seed for reproducibility
    C=0.01,                     # Regularization strength (smaller = stronger regularization)
    penalty='l2',               # Use L1 regularization (feature selection)
    solver='liblinear'          # 'saga' solver supports L1 with large datasets
)

# Fit the model to the training data
logreg.fit(X_train, y_train)   # Train the logistic regression model
```

Appendix 19: Logistic Regression Algorithm hayper parameter tuning.

Visualizing Accuracy for Logistic Regression

```
import matplotlib.pyplot as plt
import seaborn as sns

# Step 1: Filter only the class rows (exclude accuracy, macro avg, etc.)
class_metrics = report_df.loc[['Dissatisfied', 'Satisfied'], ['precision', 'recall', 'f1-score']]

# Step 2: Reset index and melt DataFrame for plotting
class_metrics = class_metrics.reset_index().melt(id_vars='index', var_name='Metric', value_name='Score')
class_metrics.rename(columns={'index': 'Class'}, inplace=True)

# Step 3: Convert scores to percentage
class_metrics['Score'] = class_metrics['Score'] * 100

# Step 4: Extract and format accuracy
accuracy_percent = float(report_df.loc['accuracy', 'precision']) * 100
accuracy_label = f"Overall Accuracy: {accuracy_percent:.2f}%"

# Step 5: Plot
plt.figure(figsize=(12, 6))
ax = sns.barplot(data=class_metrics, x='Metric', y='Score', hue='Class', palette='coolwarm')
```

Appendix 20: Visualizing Accuracy for Logistic Regression Algorithm.

2. Decision Tree

Generate | + Code | + Markdown

```
# Step 1: Import Libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.tree import plot_tree
```

```
# Step 2: Load and Split the Dataset
# Load dataset
dataset = pd.read_csv('balanced_dataset.csv')

# Split into features and target
X = dataset.drop(columns=['Satisfaction_Satisfied'])
y = dataset['Satisfaction_Satisfied']
```

Appendix 20: Decision Tree Algorithm data Importing Labiraries.

Hyperparameter Tuning for Decision Tree

Generate + Code + Markdown

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import GridSearchCV

# Define the model
dt_model = DecisionTreeClassifier(random_state=42)

# Define the hyperparameters to tune
param_grid = {
    'criterion': ['gini', 'entropy'],               # Split criteria (Gini or Information Gain)
    'max_depth': [None, 5, 10, 15, 20],              # Max depth of the tree
    'min_samples_split': [2, 5, 10, 20],             # Minimum samples required to split an internal node
    'min_samples_leaf': [1, 2, 5, 10],               # Minimum samples required to be at a leaf node
    'max_features': [None, 'sqrt', 'log2'],          # Number of features to consider at each split
}

# Setup GridSearchCV
grid_search = GridSearchCV(estimator=dt_model, param_grid=param_grid, cv=5, n_jobs=-1, scoring='accuracy')

# Fit the grid search to the training data
grid_search.fit(X_train, y_train)

# Print best parameters and score
print("Best Parameters:", grid_search.best_params_)
```

Step 3: Train the Decision Tree Model

Generate + Code + Markdown

```
# Initialize and train Decision Tree with best tuned parameters
dt_model = DecisionTreeClassifier(
    criterion='entropy',            # Use Gini impurity (best from tuning)
    max_depth=10,                  # Limit depth to prevent overfitting (best from tuning)
    min_samples_split=20,          # Require at least 20 samples to split (best from tuning)
    min_samples_leaf=5,            # Require at least 5 samples per leaf (best from tuning)
    max_features=None,             # Consider all features at each split (best from tuning)
    random_state=42                # Set random seed for reproducibility
)

# Train the model on the training data
dt_model.fit(X_train, y_train)
```

```
DecisionTreeClassifier
DecisionTreeClassifier(criterion='entropy', max_depth=10, min_samples_leaf=5,
                       min_samples_split=20, random_state=42)
```

```
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.tree import plot_tree
from sklearn.tree import DecisionTreeClassifier

# Train a Decision Tree (replace with your actual model)
dt_model = DecisionTreeClassifier(max_depth=3, random_state=42)  # Limiting depth for clarity
dt_model.fit(X_train, y_train)

# Get top N important features
importances = pd.Series(dt_model.feature_importances_, index=X_train.columns)
top_features = importances.sort_values(ascending=False).head(10).index.tolist()

# Train new model using only top features
X_train_top = X_train[top_features]
dt_model_top = DecisionTreeClassifier(max_depth=3, random_state=42)
dt_model_top.fit(X_train_top, y_train)

# Plot the simplified tree
plt.figure(figsize=(20, 10))
plot_tree(
    dt_model_top,
    feature_names=top_features,
    class_names=['Not Satisfied', 'Satisfied'],  # Replace with your label list
    filled=True,
    rounded=True,
    fontsize=12,
    node_ids=True,
```

Appendix 22: Decision Tree Algorithm hayper parameter tuning and Train The Decision Tree Model, visualize decision tree

Random Forest

```
# Step 1: Import Libraries
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, accuracy_score
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.metrics import confusion_matrix
```

```
# Step 2: Load and Split the Dataset
# Load dataset
dataset = pd.read_csv('balanced_dataset.csv')

# Split into features and target
X = dataset.drop(columns=['Satisfaction_Satisfied'])
y = dataset['Satisfaction_Satisfied']

# Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, stratify=y, random_state=42)
```

Hyperparameter Tuning

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import GridSearchCV

# Define the Random Forest model
rf_model = RandomForestClassifier(random_state=42)

# Define the hyperparameters grid
param_grid = {
    'n_estimators': [50, 100, 200],                # Number of trees in the forest
    'criterion': ['gini', 'entropy'],              # Function to measure quality of a split
    'max_depth': [None, 10, 20, 30],               # Max depth of the tree
    'min_samples_split': [2, 5, 10],               # Minimum samples required to split an internal node
    'min_samples_leaf': [1, 2, 4],                 # Minimum samples required to be at a leaf node
    'max_features': ['sqrt', 'log2', None],        # Number of features to consider at each split
    'bootstrap': [True, False],                    # Whether to use bootstrap sampling
    'class_weight': ['balanced', None]             # Whether to balance class weights
}

# Setup GridSearchCV to find the best combination of parameters
grid_search = GridSearchCV(estimator=rf_model, param_grid=param_grid, cv=5, n_jobs=-1, scoring='accuracy')

# Fit the grid search to the training data
grid_search.fit(X_train, y_train)
```

```python
from sklearn.ensemble import RandomForestClassifier

rf_model = RandomForestClassifier(
    n_estimators=100,                # Number of trees in the forest
    criterion='gini',                # Function to measure the quality of a split: 'gini' or 'entropy'
    max_depth=None,                 # Max depth of the tree (None means nodes are expanded until all leaves are pure)
    min_samples_split=2,            # Minimum number of samples required to split an internal node
    min_samples_leaf=1,             # Minimum number of samples required to be at a leaf node
    min_weight_fraction_leaf=0.0,   # Minimum weighted fraction of the sum total of weights required to be at a leaf node
    max_features='sqrt',            # Number of features to consider when looking for the best split
    max_leaf_nodes=None,            # Grow trees with max_leaf_nodes in best-first fashion
    min_impurity_decrease=0.0,      # Node will be split if this split induces a decrease of impurity greater than or equal to this value
    bootstrap=True,                 # Whether bootstrap samples are used when building trees
    oob_score=False,                # Whether to use out-of-bag samples to estimate the generalization score
    n_jobs=-1,                      # Number of jobs to run in parallel (-1 means using all processors)
    random_state=42,                # Seed for reproducibility
    verbose=0,                      # Controls the verbosity when fitting and predicting
    warm_start=False,               # Reuse the solution of the previous call to fit and add more estimators to the ensemble
    class_weight='balanced',        # Automatically adjust weights inversely proportional to class frequencies
    ccp_alpha=0.0,                  # Complexity parameter used for Minimal Cost-Complexity Pruning
    max_samples=None                # Number of samples to draw from X to train each base estimator (only if bootstrap=True)
)

rf_model.fit(X_train, y_train)
```

Classification Report

```python
from sklearn.metrics import confusion_matrix

# Step 5: Make predictions on the test set
y_pred = rf_model.predict(X_test)

# Calculate the confusion matrix
cm = confusion_matrix(y_test, y_pred)

# Calculate correctly and incorrectly classified predictions
correctly_classified = (y_pred == y_test).sum()
incorrectly_classified = (y_pred != y_test).sum()

# Print the results
print(f"Correctly classified predictions: {correctly_classified}")
print(f"Incorrectly classified predictions: {incorrectly_classified}")

# Labels for the confusion matrix plot
labels = ['Dissatisfied', 'Satisfied']

# Plot the confusion matrix as a heatmap
plt.figure(figsize=(11, 7))
sns.heatmap(cm, annot=True, fmt='d', cmap='Greens', xticklabels=labels, yticklabels=labels)
```

Appendix 23: Rendom Forest Algorithm data importing and hayper parameter tuning Classification report.

SVM Hyperparameter Tuning

```python
from sklearn.svm import SVC
from sklearn.model_selection import GridSearchCV

# Create a base SVM model
svm_model = SVC(class_weight='balanced')  # Handle imbalance if needed

# Define the parameter grid
param_grid = {
    'C': [0.01, 0.1, 1, 10],                  # Regularization strength
    'kernel': ['linear', 'rbf', 'poly'],       # Different kernel types
    'gamma': ['scale', 'auto', 0.001, 0.01, 0.1],  # Kernel coefficient
    'degree': [2, 3, 4]                        # Only used for 'poly' kernel
}

# Setup GridSearchCV
grid_search = GridSearchCV(estimator=svm_model, param_grid=param_grid,
                           cv=5, n_jobs=-1, scoring='accuracy', verbose=1)

# Fit to training data
grid_search.fit(X_train, y_train)

# Print best parameters and best cross-validation score
```

```python
# Step 4: Train the Support Vector Machine (SVM) Model
svm_model = SVC(
    C=1.0,                      # Regularization parameter (higher = less regularization)
    kernel='rbf',               # Kernel type: 'rbf' (radial basis function)
    degree=3,                   # Degree for 'poly' kernel (ignored by other kernels)
    gamma='scale',              # Kernel coefficient for 'rbf', 'poly', and 'sigmoid'
    coef0=0.0,                  # Independent term in kernel function
    shrinking=True,             # Use shrinking heuristic (faster training)
    probability=True,           # Enable probability estimates
    tol=1e-3,                   # Tolerance for stopping criterion
    cache_size=200,             # Size of kernel cache (in MB)
    class_weight='balanced',    # Handle unbalanced class weights
    verbose=False,              # Disable verbose output
    random_state=42             # Reproducibility
)

# Fit the model to the training data
svm_model.fit(X_train, y_train)
```

Appendix 24: SVM Algorithm hayper parameter tuning.

XGBoost

```python
# Step 1: Import necessary libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from xgboost import XGBClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, accuracy_score, confusion_matrix
```

Generate + Code + Markdown

```python
# Load the dataset from the previous session (already loaded as 'df') and prepare X and y
dataset = pd.read_csv('balanced_dataset.csv')
X = dataset.drop(columns='Satisfaction_Satisfied')  # Features
y = df['Satisfaction_Satisfied']  # Target variable
```

```python
# Step 2: Train-test split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Hyperparameter Tuning

```python
# Import necessary modules
from xgboost import XGBClassifier  # XGBoost classifier for gradient boosting
from sklearn.model_selection import GridSearchCV  # Grid search for hyperparameter tuning

# Initialize the base XGBoost model
# - use_label_encoder=False: avoids deprecation warning
# - eval_metric='logloss': evaluation metric for binary classification
# - random_state=42: ensures reproducibility
xgb_model = XGBClassifier(use_label_encoder=False, eval_metric='logloss', random_state=42)

# Define the hyperparameter grid for tuning
# Each key corresponds to a hyperparameter of XGBClassifier
param_grid = {
    'n_estimators': [100, 200],          # Number of trees (boosting rounds)
    'max_depth': [3, 5, 7],              # Maximum tree depth for base learners
    'learning_rate': [0.01, 0.1, 0.2],   # Step size shrinkage used in update
    'subsample': [0.8, 1.0],             # Subsample ratio of training data
    'colsample_bytree': [0.8, 1.0],      # Subsample ratio of columns when constructing each tree
    'gamma': [0, 1, 5],                  # Minimum loss reduction to make a split
    'reg_alpha': [0, 0.1, 1],            # L1 regularization term on weights
    'reg_lambda': [1, 5, 10]             # L2 regularization term on weights
}
```

```
import warnings
warnings.filterwarnings("ignore", category=UserWarning, module='xgboost')

# Import necessary libraries
import pandas as pd
import numpy as np
from xgboost import XGBClassifier
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler

# Step 1: Load dataset
dataset = pd.read_csv('balanced_dataset.csv')

# Step 2: Encode target labels (Dissatisfied → 0, Satisfied → 1)
dataset['Satisfaction_Satisfied'] = dataset['Satisfaction_Satisfied'].map({'Dissatisfied': 0, 'Satisfied': 1})

# Step 3: Define X and y
y = dataset['Satisfaction_Satisfied']
X = dataset.drop('Satisfaction_Satisfied', axis=1)

# Step 4: One-hot encode categorical variables (if any)
X = pd.get_dummies(X, drop_first=True)

# Step 5: Split the dataset into train and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Model Interpretability

- XGBoost provides built-in feature importances and works efficiently with SHAP.

```
import xgboost as xgb
import matplotlib.pyplot as plt
import shap
import pandas as pd

# Ensure feature names are set
xgb_model.get_booster().feature_names = X_train.columns.tolist()

# --- XGBoost Feature Importance Plot ---
# Get raw feature importances
importance_df = xgb_model.get_booster().get_score(importance_type='gain')
importance_df = pd.DataFrame.from_dict(importance_df, orient='index', columns=['Gain'])
importance_df.index.name = 'Feature'
importance_df.reset_index(inplace=True)

# Get top 10 features by gain
top10 = importance_df.sort_values(by='Gain', ascending=False).head(10)

# Plot manually for more control
plt.figure(figsize=(10, 7))
```

Appendix 25: XGBOOST Algorithm Impoting data, hayper parameter tuning model Interpretability.

```python
# Step 1: Import Necessary Libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report, confusion_matrix
import matplotlib.pyplot as plt
import seaborn as sns
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout
import warnings
warnings.filterwarnings('ignore')rr



#Step 2: Load the Dataset
```

```
Epoch 1/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 5s 4ms/step - accuracy: 0.7576 - loss: 0.5273 - val_accuracy: 0.8391 - val_loss: 0.3987
Epoch 2/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8217 - loss: 0.4312 - val_accuracy: 0.8391 - val_loss: 0.3939
Epoch 3/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 3s 4ms/step - accuracy: 0.8366 - loss: 0.4080 - val_accuracy: 0.8423 - val_loss: 0.3909
Epoch 4/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8299 - loss: 0.4141 - val_accuracy: 0.8450 - val_loss: 0.3858
Epoch 5/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8324 - loss: 0.4054 - val_accuracy: 0.8473 - val_loss: 0.3848
Epoch 6/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8375 - loss: 0.3989 - val_accuracy: 0.8485 - val_loss: 0.3840
Epoch 7/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8396 - loss: 0.3984 - val_accuracy: 0.8462 - val_loss: 0.3815
Epoch 8/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8397 - loss: 0.4027 - val_accuracy: 0.8498 - val_loss: 0.3800
Epoch 9/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8372 - loss: 0.4036 - val_accuracy: 0.8510 - val_loss: 0.3800
Epoch 10/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8367 - loss: 0.3963 - val_accuracy: 0.8502 - val_loss: 0.3777
Epoch 11/100
600/600 ━━━━━━━━━━━━━━━━━━━━ 2s 3ms/step - accuracy: 0.8418 - loss: 0.3902 - val_accuracy: 0.8510 - val_loss: 0.3803
```

```
# ------ TRAINING ------

# Set up EarlyStopping callback to monitor validation loss
# Will stop training if val_loss doesn't improve for 10 consecutive epochs
# restore_best_weights=True ensures we keep the best model weights
early_stop = EarlyStopping(
    monitor='val_loss',  # Metric to monitor (validation loss)
    patience=10,         # Number of epochs to wait before stopping
    restore_best_weights=True  # Restore weights from best epoch
)

# Train the model using fit() method
history = model.fit(
    # Training data features (input)
    X_train,
    # Training data labels (target)
    y_train,
    # Maximum number of training epochs
    epochs=100,
    # Number of samples per gradient update
    batch_size=32,
    # Fraction of training data to use for validation
    validation_split=0.2,
    # Verbosity mode (1 = progress bar)
    verbose=1,
    # List of callbacks to apply during training
    callbacks=[early_stop]  # Using our EarlyStopping callback
```

```
e a Sequential model (linear stack of layers)
 Sequential([

irst hidden layer with 64 neurons and ReLU activation
nput_shape specifies the number of features in the input data
se(64, activation='relu', input_shape=(X_train.shape[1],)),

atch normalization layer to normalize activations from previous layer
elps stabilize and accelerate training
chNormalization(),

ropout layer randomly sets 30% of input units to 0 during training
elps prevent overfitting by introducing noise
pout(0.3),

econd hidden layer with 32 neurons and ReLU activation
se(32, activation='relu'),

nother batch normalization layer
chNormalization(),

utput layer with 1 neuron and sigmoid activation
igmoid outputs probability between 0 and 1 for binary classification
se(1, activation='sigmoid')
```

Appendix 26: Importing necessary libraries, data preprocessing, hyperparameter tuning, and training the MLP model.