Flight Delay Prediction Using Machine Learning

Trisha P, Subiksha N, Prashanth P, Keshav Perumal S, Sudharshanan G, Balasubramani M
School of Computer Science Engineering and Information System
Vellore Institute of Technology

***Abstract*—A Flight delay prediction system that forecasts potential delays prior to departure using statistical or machine learning techniques. For travelers, airlines, and airports alike, flight delays may be annoying and inconvenient. The system gathers information from a number of sources, such as weather information, airline operations, historical flight records, and air traffic control data. The algorithm can identify trends and connections that affect flight delays by examining these data.The use of machine learning models to predict flight delays, focusing on logistic regression, decision tree classifier, random forest classifier, Naive Bayes classifier, and extreme gradient boosting (XGBoost) classifier.**

The Flight Delay Prediction System was built using a technique that comprises feature engineering, data preprocessing, model selection, and evaluation. The study utilizes a comprehensive dataset containing flight details, weather conditions, and other factors influencing delays. A comparative analysis is performed to evaluate model accuracy, precision, recall, and F1 score, determining the most effective approach for flight delay prediction. Preliminary results suggest that ensemble methods, particularly XGBoost, achieve higher accuracy due to their robustness in handling feature interactions. This research provides valuable insights into leveraging machine learning for predictive analytics in aviation, offering a foundation for developing proactive measures to minimize delays and enhance operational efficiency.

***Index Terms*—Machine learning, flight delay, flight cancellation, Weather forecast**

I. Introduction

In the aviation sector, flight delays pose a serious problem since they cause inconvenience to customers, airlines' revenue loss, and additional strain on airport operations. Several variables, including weather, air traffic, security concerns, and operational difficulties unique to individual airlines, can contribute to the unpredictable nature of delays. Because machine learning can examine large, complicated information and find patterns and correlations that traditional statistical methods might overlook, it has emerged as a crucial tool for forecasting flight delays. It is feasible to predict not only whether a flight will be delayed but also how long it might last by utilizing machine learning models. These forecasts offer useful information to travelers, airport officials, and airlines, assisting in scheduling optimization, reducing delays, and improving the overall travel experience. XGBoost, Random Forest classifier, Naive Bayes, and logistic regression are the four machine learning models that shine out in the field of flight delay prediction. Because of their distinct strengths and capacities, each of these models can be used for certain parts of the prediction task.

The Logistic Regression Model is One of the most straightforward and often applied machine learning algorithms for binary classification problems logistic regression. It forecasts the probability of a binary result, in this example, the delay or non-delay of a flight. Using input features such as flight route, departure time, weather, and airline information, logistic regression models the likelihood that a flight falls into a particular class. By applying a logistic function to a linear combination of the input data, it determines the likelihood of a delay and generates a number between 0 and 1, which is then used to determine whether the aircraft is delayed or on time. For many classification tasks, logistic regression offers a solid baseline and is quite interpretable despite its simplicity. Its incapacity to manage intricate, non-linear interactions between input features—which are typical in flight delay scenarios where multiple elements interact in non-linear ways—is its main drawback, though. For preliminary analysis, logistic regression is still a helpful method that enables stakeholders to find fundamental relationships in the data.

The Random Forest Classifier is the order to increase accuracy and lower the possibility of overfitting, the random forest classifier is an ensemble learning model that integrates the predictions of several decision trees. A random subset of the data is used to train each decision tree, and the majority vote across all trees determines the final forecast. Random forests are perfect for identifying patterns in the data because they can manage a lot of input features and intricate relationships between them in flight delay prediction. Because flight delay datasets frequently contain partial records or a variety of data formats, this model is particularly good at handling missing values and can handle both continuous and categorical variables. Feature importance scores, which highlight the elements most closely linked to delays, are another advantage of random forests. For example, if weather is found to have a significant impact, it implies that airlines should focus on addressing weather-related delays. Interpretability is a trait that increases transparency and aids stakeholders in comprehending the reasons behind delays. The XGBoost Classifier is a very effective and potent ensemble learning method. Forecasting aircraft delays is one of the many large-scale prediction jobs that use XGBoost, which is well known for its speed and performance. This model sequentially constructs a number of decision trees, each one aiming to fix the mistakes of the one before it. XGBoost may capture intricate patterns and non-linear correlations between features, such those between weather changes, air traffic, and flight schedules, by fine-tuning hyperparameters to optimize for performance. XGBoost is especially well-suited for situations where predictive accuracy is crucial because of its versatility in handling different loss functions and minimizing prediction mistakes.

II. PROPOSED WORK

A. Data Collection and Preprocessing

The preprocessing and data collection stages are essential to creating a precise flight delay prediction system. Numerous sources of information were used, including weather reports, airline operational data, and previous flight schedules. Preprocessing methods such data cleansing, addressing missing values, and identifying outliers were used to guarantee the quality of the data. For improved model performance, more features were standardized and changed, such as departure time, weather conditions, and airport traffic. By ensuring that the dataset is appropriate for machine learning models, this step improves the prediction accuracy of flight delays.

Historical flight data is collected from various sources, including airlines and public databases such as the U.S. Department of Transportation (DOT), which provides detailed information on flight schedules, delays, and cancellations. In addition to flight data, weather information is sourced from meteorological agencies, as weather conditions play a significant role in flight delays.

B. Feature Engineering

A key factor in raising a flight delay prediction system's accuracy is feature engineering. It entails turning unstructured data into useful features that can improve machine learning models' performance. Features including weather, flight departure time, day of the week, past delay trends, airport congestion, and aircraft turnaround times are all integrated into this system. For improved model processing, categorical variables such as airline and destination are also encoded. More accurate and trustworthy delay forecasts result from feature engineering that makes sure the most pertinent data is recorded.

C. Feature Identification

A feature Finding important characteristics, or features, that could have a big influence on flight delays requires evaluating the data gathered in order to develop a flight delay prediction system. Airports for departures and arrivals are among these features, since some may have more traffic or operational problems than others. Given that delays are more common during peak hours, flight time—including the time of day and duration—is also a crucial consideration. Additional significant factors include weather, past flight delay trends, and the airline, as some may have greater on-time history. Finding and adding these pertinent characteristics significantly enhances the model's capacity to forecast flight delays.

D. Model Selection and Training

The process of choosing and training a model is essential to creating a flight delay prediction system that works. This stage involves assessing the predicted accuracy of several machine learning techniques, including Random Forest, Gradient Boosting, Logistic Regression, Decision Tree classifier and Naive Bayes Classifer. The objective is to choose the model that most accurately depicts the connections between flight delays and input variables. To maximize its capacity to precisely forecast future flight delays, the selected model is then trained using historical flight data, which includes characteristics like weather, airport traffic congestion, and airline performance.

- I. Logistic Regression: It is a popular supervised classification technique that forecasts the likelihood of a binary result, such true/false or yes/no occurrences. Logistic regression uses the logistic function to estimate probabilities that are limited between 0 and 1, in contrast to linear regression, which predicts constant values. For situations where the objective variable is categorical, such forecasting whether the flight will be delayed a "yes" or on time (no), this makes it perfect. Using input features like departure time, weather, and airline performance, the algorithm determines the likelihood of the result. Because of its ease of use, effectiveness, and interpretability, logistic regression is frequently used to solve binary classification issues such as flight delay prediction.

- II. Decision Tree Classifier: is a straightforward yet powerful machine learning technique that bases judgments on input features using a tree-like structure. Every internal node represents a test or decision on a particular feature (weather, departure time, etc.), and each leaf node has a class name, such "delayed" or "on time." It looks like a flowchart. The algorithm determines which feature offers the most valuable information to split the data at each stage as the tree is constructed using a metric such as **information gain** or **Gini index**. Until the data is completely classified, this process keeps on. Choice Although they can be susceptible to overfitting if improperly tuned, trees are helpful for predicting flight delays since they are simple to understand and interpret.

- III. XGBoost Classifier: is a popular supervised classification technique that forecasts the likelihood of a binary result, such true/false or yes/no occurrences. Logistic regression estimates probabilities that fall between 0 and 1, in contrast to linear regression, which predicts constant values. It uses the logistic function to accomplish this. Because of this, it is perfect for situations in where the goal variable is categorical, such as determining whether the flight will be delayed or arrive on time. Using input parameters including weather, departure time, and airline performance, the algorithm determines the likelihood of the result. Because of its ease of use, effectiveness, and interpretability, logistic regression is frequently used to solve binary classification problems, such as flight delay prediction.

- IV. Random Forest: is a well-liked supervised classification method that predicts the probability of a binary outcome, such true/false or yes/no events. Unlike linear regression, which predicts constant values, logistic regression estimates probabilities that are confined between 0 and 1. It does this by using the logistic function. This makes it ideal for scenarios where the objective variable is categorical, like predicting whether the flight will be delayed as a "yes" or on time at all. The algorithm calculates the likelihood of the outcome based on input parameters such as airline performance, weather, and departure time. Logistic regression is often utilized to address binary classification problems, such flight delay prediction, due to its simplicity, efficacy, and interpretability.

- V. Naive Bayes Classifier: A popular stochastic machine learning model for classification problems in a variety of applications, such as spam detection and text classification, is the Naive Bayes classifier. Bayes' Theorem, which connects the conditional and marginal probability of random events, serves as its foundation. The fundamental aspect of Naive Bayes is its assumption of feature independence, which, when assigned a class label, treats every feature (or attribute) in the dataset as independent of the others. Even with enormous datasets, Naive Bayes is especially quick and efficient because of this simplifying assumption, sometimes known as the "naive" assumption. In reality, the Naive Bayes classifier frequently exhibits unexpectedly good performance despite its simplicity, particularly when its presumption of independence is held to be reasonably valid.

E. Model Training:

A key component of machine learning is model training, which enables predictive models to learn from data on their own and carry out tasks efficiently. Data preparation, model selection, training, evaluation, and deployment are some of the crucial elements that make up this methodical process in the context of a flight delay prediction system. Cleaning and preparing historical flight data, which includes variables like weather, airline schedules, and airport traffic, is the first step in data preparation. Depending on their performance criteria, appropriate machine learning algorithms, like Random Forest or Logistic Regression, are chosen once the data is ready. By modifying its parameters to reduce prediction errors, the model learns the complex correlations between input data and the objective variable—flight delay—during the training phase. Before being deployed, the trained model is assessed against an independent test dataset to guarantee its correctness and dependability, enabling it to generate well-informed predictions in real-time situations. This all-encompassing strategy is essential for creating a system that improves aircraft operational efficiency.

F. Model Evaluation and Optimization

In order to construct a flight delay prediction system, this stage is essential. Metrics including accuracy, precision, recall, and F-Measure are used in this phase to evaluate the performance of trained models. Techniques like cross-validation and hyperparameter tuning are used to optimize model performance based on evaluation outcomes, guaranteeing accurate and dependable predictions in real-world situations.

Model Evaluation A critical stage in determining the flight delay prediction system's efficacy is model evaluation. The testing dataset, which the model has not faced during training, is used to thoroughly test the model's performance during this phase. Accuracy, which gauges the total correctness of predictions, and Precision, which assesses the percentage of true positive predictions among all positive classifications, are two key performance metrics used to evaluate its predictive capabilities. To find out how frequently the model misclassifies instances, the error rate is also computed. Stakeholders can assess the model's efficacy and dependability in precisely forecasting flight delays by examining these KPIs, which will help them make the required corrections.

Optimization This crucial stage aims to improve the flight delay prediction model's performance. This method entails adjusting a number of hyperparameters that affect the behavior of the model, such as the learning rate in Gradient Boosting, which regulates the pace at which the model converges, and tree depth in Random Forests, which influences how detailed the trees are. This step might also entail investigating different algorithms to see whether they perform better than the existing model. To make sure the final model has the best accuracy and dependability in predicting flight delays, methods like grid search or randomized search can be used to methodically assess various hyperparameter combinations.

G. Model Deployment and Monitoring

In order to operationalize the flight delay prediction system, this step is essential. In order to provide real-time prediction, this phase entails integrating the trained model into a production environment. To maintain the model's accuracy over time and enable fast updates and modifications in response to shifting flight operational conditions, it is imperative that its performance be continuously monitored. The process of incorporating the completed flight delay prediction model into an operational environment in real time is known as deployment. The model is linked to real-time data sources throughout this phase, including airline schedules, weather reports, and air traffic conditions. The model uses newly incoming data to process it and produce timely forecasts about flight delays. This helps airlines and operational teams to make educated decisions, improving efficiency while improving customer experience by proactively controlling delays and optimizing flight itineraries.

To make sure the flight delay prediction model continues to be accurate and useful over time, monitoring and retraining are crucial processes. Any decline in predicting capacities, which may result from shifting circumstances like seasonal variations, changes in airport operations, or changing flight patterns, can be identified early thanks to ongoing model performance monitoring. The model should be routinely retrained with new data to solve these problems and make sure it stays accurate and relevant. Furthermore, by offering a practical way to train and assess several machine learning models on the dataset, the runner() function expedites this procedure. In order to improve performance and get the best outcomes when predicting flight delays, this function also permits hyperparameter optimization, which allows the model parameters to be fine-tuned.

III. RELATED WORKS

- Over the previous few decades, Germany's air traffic has grown significantly, reaching a record-breaking 3 million flights in 2018. Delays have increased in tandem with this increase in air traffic, resulting in a number of expenses for airlines, travelers, and Air Navigation Service Providers (ANSPs). Abadie et al. have suggested studies to evaluate and predict delay-related expenses in order to address this problem. **[1]** Among these, a recent study suggested using stochastic modeling to predict future air travel, delays, and associated expenses.

- Hatipoğlu and Tosun N presented that Delays in flights have a number of detrimental effects, such as higher expenses, lower customer satisfaction, and contamination of the environment. They want to use machine learning techniques to forecast delayed flights in order to lessen these problems. They use the Bayesian approach to optimize the hyper-parameters of three contemporary algorithms based on gradient boosting: XGBoost, LightGBM, and CatBoost. **[2]** Additionally, they use the Synthetic Minority Over-Sampling Technique (SMOTE) to alleviate the class imbalance of delayed flights. The results show that it is possible to accurately estimate flight delays.

- Due to the quick development of modern aviation, air travel has increased dramatically in recent years. The airline sector has more difficulties with flight delays as demand for flights rises, and these issues need to be resolved. **[3]** Predicting aircraft delays accurately can increase airport operational effectiveness and passenger comfort. They use a dataset from an airline to test three distinct Gradient Boosting techniques.

- In the aviation sector, flight delays pose serious problems that result in operational and financial losses. To address this issue, Anil Ayaydin and Ali Akcayol have proposed three distinct machine learning and deep learning techniques: Random Forest (RF), Long Short-Term Memory (LSTM) networks, and Deep Recurrent Neural Networks (DRNN). They believe that anticipating delays in advance is essential to reducing these effects and putting the appropriate safeguards in place. A real-world dataset comprising aircraft delay data from 368 airports worldwide was used to evaluate and compare these models in great detail **[4]** .

- Large-scale delays are frequently caused by flight delays, which have a ripple effect on the entire air transportation network, affecting connected flights and airports. Predicting individual aircraft delays and managing how these delays spread throughout the network are both necessary to solve this problem. The advanced control of flight delays through delay prediction and schedule optimization is the main emphasis of this study **[5]** . The suggested approach modifies the affected flight plans by combining delay prediction with a delay optimization model.

- **[6]** Both airline operations and airport on-time performance, which are directly related to passenger pleasure, are significantly impacted by flight delays. Therefore, for decision-making to be effective, it is essential to precisely define the factors that cause delays. **[6]** For the investigation of flight delays, traditional statistical models have been employed extensively; however, the problem of unobserved heterogeneity in flight data has not gotten as much attention. Using two modeling techniques, this work empirically analyzes unobserved heterogeneity and evaluates the influence of important variables on flight delays.

- Using the Deep Operator Network (DON) and the Gradient-Mayfly Optimization Algorithm (GMOA), Bisandu and Moulitsas I. forecasted flight delays by combining deep learning with optimization techniques. While GMOA effectively adjusts model parameters to improve prediction accuracy, DONs are excellent at identifying intricate patterns in flight delay data.**[7]** Through efficient management of nonlinear interactions and model parameter optimization made possible by this integration, stakeholders in the aviation sector can gain insightful knowledge.

- Using supervised machine learning algorithms to forecast arrival delays is crucial in order to anticipate aircraft delays. Accuracy is ensured by a thorough review of five different approaches, which use flight data and meteorological information to produce accurate predictions specific to each flying scenario.[8] The study by Krushna Mudigonda et al. **[8]** highlights the value of data analytics and machine learning in maximizing airline performance by improving the accuracy of delay projections, which in turn improves operational efficiency and customer satisfaction in airline transport systems.

- Rajesh K. and Srikanth V.'s said that the machine learning methodology looks into ways to predict flight delays while emphasizing important variables like weather and traffic at airports. Among the evaluated strategies, decision trees and random forests stand out as the best. By empowering airlines and airports to enhance operations, discoveries improve customer happiness and advance the resilience of the aviation industry. By utilizing machine learning, the study highlights how data-driven approaches may be used to control and reduce flight delays, enhancing industry performance as a whole. **[9]**

- Using a variety of machine learning algorithms, such as Random Forest and XGBoost, the problem of forecasting flight delays by utilizing historical data and different contributing elements is investigated; Random Forest obtained the greatest accuracy score of 0.93. In the aviation sector, accurate forecasting facilitates well-informed decision-making, which raises operational effectiveness and passenger pleasure.**[10]** This research helps to improve travel experiences and minimize disruptions by utilizing cutting-edge machine learning techniques.

- Weili Zeng et al presented A deep graph neural network-based delay prediction model called DGLSTM. Airports are represented as nodes in DGLSTM, which improves accuracy and robustness over current approaches by taking delay propagation characteristics into account.[11] The model's superiority for delay prediction inside the airport network is demonstrated by experimental results that prove its efficacy using historical delay data from 325 US airports

- **[12]** An airport-specific causal flight delay prediction model that emphasizes both direct and indirect delay variables. With an average absolute error of roughly 8.15 minutes on the test dataset from Beijing International Airport, LSTMAM, which incorporates an attention mechanism, performs more accurately than baseline algorithms.**[12]** Passengers' anxiety is effectively reduced when LSTM-AM projections are used to notify them ahead of time, and the model's identification of crucial time points helps reduce or eliminate delays by controlling runway and apron flow.

- The effectiveness of the stacking algorithm for predicting airport flight delays, which tackles the machine learning algorithm selection problem. **[13]** It employs Boruta for feature selection and SMOTE for processing imbalanced datasets. The first-level learner uses five supervised algorithms, while the second-level learner uses logistic regression. Multiple evaluation criteria, such as Accuracy, Precision, Recall, F1 Score, ROC curve, and AUC Score, show that stacking increases prediction accuracy while preserving stability in experimental outcomes utilizing 2019 datasets from Boston Logan International Airport.

- Three well-known machine learning methods are used to forecast airline arrival delays on a dataset of domestic flights. Using three situations with different input feature sets, it focuses on how short-term features affect prediction quality.**[14]** The model that incorporates departure delay achieves the best accuracy of 89.9 percent and a recall of 83.4 percent, demonstrating the importance of timing in feature selection for optimal prediction quality. The results indicate that adding short-term information increases prediction accuracy.

- **[15]** An experiment using cluster computing to assess the random forest data mining algorithm's performance. A 35.8 percent speedup in random forest performance is demonstrated by the simulation using one master and three worker nodes with the same hardware specifications.[15] Computing performance in the cluster environment is further accelerated by adding more nodes with better hardware. In the industrial sector, flight delays

  are one of the most common uncertainties that arise during flight planning. Numerous circumstances might lead to these delays, which can cost airlines, operators, and passengers a lot of money.**[16]** Unfavorable weather conditions, excessive demand during holidays or seasonal peaks, airline policies, technical problems including malfunctioning airport infrastructure, baggage handling issues, mechanical faults, or delays from earlier flights can all cause departure delays. Temperature, humidity, rainfall (in millimeters), visibility, and the month of the year are all important elements for this flight delay prediction method, which focuses on weather-related factors that cause delays.

- Commercial airlines suffer large financial losses as a result of flight delays, and passengers become dissatisfied. In order to solve this problem, Bashayer et al. have thoroughly investigated prediction techniques that make use of data mining and machine learning tools. But the accuracy and resilience that are needed are still not there in the models that are now in use. **[17]** The datasets were subjected to three predictive models: logistic regression, decision trees, and our suggested method. The outcomes demonstrate that the suggested approach performs better than the current state-of-the-art models.

- A new prediction framework for flight delay prediction that takes into account both temporal and geographic factors is presented by Qiang Li and Ranzhe Jing. It is called ST-Random Forest. They create a prediction framework utilizing LSTM units to capture the temporal correlations between weather and airport congestion, and they use complex network theory to extract spatial aspects of the aviation network at different levels.**[18]** Airport regulators and aviation authorities have access to a real-time monitoring system through the ST-Random Forest architecture.

- Due to its vital role in guaranteeing effective airline and airport operations, flight delay prediction has drawn more and more attention. The flight delay prediction problem has been approached from a network perspective by Kaiquan et al., taking into account a scenario including multiple airports.**[19]** They create a flight delay prediction method based on graph convolutional neural networks (GCN) to simulate the airport network's time-evolving and periodic graph-structured data.

- Due to growing air traffic congestion worldwide, flight delays have grown to be a major problem for the aviation sector. The goal of the model is to determine the earliest nonstop and connecting flights that are available between a source and a destination.[20] Using open-source/public Application Programming Interfaces (APIs), it obtains flight information from the user. Neo4j then processes the information and converts it into JavaScript Object Notation (JSON) format.

IV. EXPERIMENTAL RESULTS

A. Exploratory Data Analysis (EDA)

Understanding the underlying patterns, trends, and correlations in a dataset before implementing machine learning models is the goal of exploratory data analysis (EDA), a critical phase in the data science process. In order to obtain insights and identify abnormalities, EDA summarizes the data using descriptive statistics, graphics, and correlation analyses. Plotting histograms, scatter plots, box plots, and correlation matrices are common methods for observing feature distribution, locating outliers, and determining correlations between variables. Data scientists can use EDA to find significant patterns that could impact the performance of predictive models, such as missing values, skewed distributions, or feature correlations. EDA also aids in identifying the most pertinent features, directing the preparation of data, and verifying hypotheses. EDA ensures that latter stages of the analysis are based on precise, well-understood data by visually and statistically evaluating the data, which serves as a basis for feature engineering, model selection, and informed decision-making.

B. FACTORS CAUSING DELAY

There are several causes of flight cancellations and delays, and these causes are frequently divided into different categories. These causes are graphically shown by two pie charts that clearly break down the causes of cancellations and delays. Extreme weather phenomena, such as storms or fog, can cause flight schedule disruptions, making weather conditions a common cause of delays. Delays are also caused by airline operations, frequently as a result of inefficient turnaround times, personnel schedule conflicts, or aircraft maintenance. Airport gate availability, runway congestion, and air traffic control limitations are the main causes of National Air System delays. Last but not least, increased security screenings or possible threats cause security-related delays, albeit they are less common. Weather and airline problems are the main causes of cancellations, which frequently follow similar trends. By providing a thorough understanding of the ways in which these different elements impact flight punctuality, these pie charts assist stakeholders in addressing the underlying causes of interruptions.

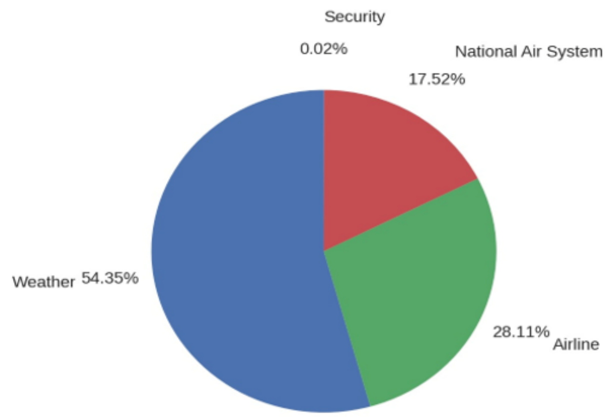

Fig. 1. Factors causing delay

The breakdown of flight delay causes demonstrates the important impact that a number of factors have in interfering with flight schedules.The most significant component, accounting for 54.35 percent of delays, is weather. This is because unfavorable weather patterns like storms, fog, and strong winds have a direct impact on flight safety and scheduling.The significance of effective internal procedures inside airlines is demonstrated by the fact that 28.11 percent of delays are caused by airline-related problems, such as staff scheduling, aircraft maintenance, and operational inefficiencies.The difficulty of controlling air traffic at crowded airports is reflected in the fact that 17.52 percent of delays are caused by National Air System problems, which include runway congestion, restrictions on air traffic control, and other systemic problems. Last but not least, security-related delays account for just 0.02 percent of all delays and are usually caused by infrequent but significant events that demand for stronger security measures. With an emphasis on weather forecasting, operational effectiveness, and air traffic control, stakeholders can better prioritize solutions to reduce delays by having a better understanding of this breakdown. Displays the percentage of flights that are canceled as opposed to those that are not. Of the flights, 97.17 percent were not canceled. The cancellation rate for flights was 2.83 percent.

C. AIRLINE VS DAY OF WEEK

The average arrival delays for each airline on various days of the week will be graphically represented by the heatmap. An effective visual tool for spotting trends and patterns in airline performance concerning arrival delays on various days of the week is the heatmap. It can assist stakeholders (such as customers, airport officials, and airlines) in improving customer happiness, understanding travel reliability, and making operational changes.

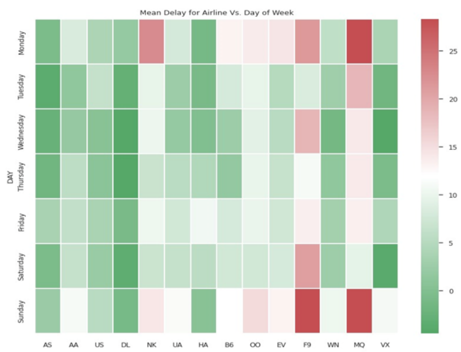

Fig. 2. Heatmap Analysis Airline Vs Day of Week

Fig.2 shows Green: Denotes improved performance with fewer mean delays. Higher mean delays (poor performance) are indicated in red. The color coding makes it easy to determine which airlines do better on particular days. If a certain airline displays a lot of red on Fridays, for example, it can indicate that delays occur more frequently on that day.

D. AIRLINE VS DESTINATION AIRPORTS

Each airline's average arrival delays per destination airport are displayed in the heatmap; low delays are indicated by green, and high delays are indicated by red. Stakeholders can promptly spot trends and possible operational problems thanks to the heatmap, which shows the average arrival delays for airlines at different destination destinations.

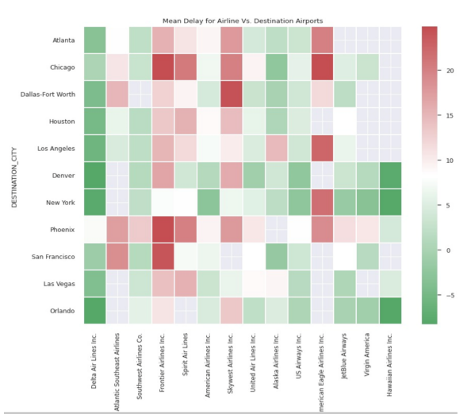

Fig. 3. Heatmap for Airline Vs Destination Airports

Airports may use this information to better understand traffic patterns, airlines can use it to improve service, and travelers can use it to select reputable airlines for their travels. The X-axis in Fig.3 displays the distinct airlines from the dataset, while the Y-axis displays the top 11 destination cities (airports). The average arrival delay for that airline at a particular location is represented by the color of each cell. White cells demonstrate intermediate delays, red cells high mean delays (less favorable performance), and green cells low mean delays (favorable performance). By rapidly identifying airlines and routes that are prone to delays, this graphic helps direct efforts to increase punctuality.

E. DELAY VS DISTANCE

The mean arrival delay may be rather small, say about five minutes, if the line plot yields the following fictitious values for lesser distances (e.g., 0-500 miles). With an increase in distance (for example, 500–1500 miles), the average arrival time may rise to approximately 15 minutes. When traveling greater distances (such as 3000–4000 miles), the average arrival time may increase considerably to about 30 minutes.

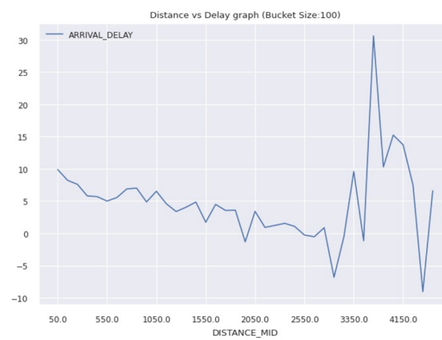

Fig. 4. Delay Vs Distance

The line plot provides information on how delays change with travel length by visualizing the link between mean arrival delay and flight distance. The X-axis gives a clear breakdown of flight distances by showing the midpoint of each distance interval, which ranges from 0 to 5000 miles. The mean arrival delay in minutes for planes within each distance range is displayed on the Y-axis. The average arrival delay for each distance midpoint is shown by a line connecting points as the plot moves forward. This makes it possible to quickly observe trends, such as whether delays are more frequent within a specific distance range or whether longer flights often have more delays.

F. DELAY VS COUNT

Many flights arrive on time or with very little delay, as evidenced by the large number of flights that may fall into the 0–10 minute delay category. Moderate delays may be indicated by a lesser number of flights falling within the 10- to 30-minute range. Longer delays may be less frequent if there are fewer flights with delays longer than thirty minutes.

Based on the given data, the histogram will show the distribution of flight arrival delays, providing a visual depiction of the frequency of various delay times. Both early and late

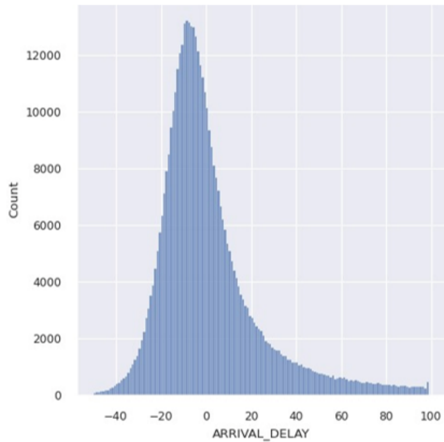

Fig. 5. Delay Vs Count

arrivals are covered by the x-axis, which shows the arrival delay time in minutes, with a range of -50 to 100 minutes. The frequency or count of flights within each designated delay range or bin is displayed on the y-axis. The number of flights that had delays during that particular time period is represented by each bar in the histogram. While shorter bars show fewer instances, taller bars show that more flights experienced delays within range. Common delay intervals and patterns, such as whether most delays are minor or there are frequent, major delays, can be found using this histogram.

G. NO. OF FLIGHTS ON DAYS OF WEEK

Monday through Friday shows an overall rising trend, suggesting that weekday flights are more frequently planned. reduced flight activity on weekends, as evidenced by the decline in flights on Saturday and Sunday. This may visually distort the data representation because the repeated points for Monday and Tuesday will show up at the same y-value as the original days.

Fig. 6. Flights on Days of Week

In Fig.6 The number of flights for each day of the week will be shown on the line graph, giving a clear picture of how flights are distributed throughout the week. With labels like ['Mon', 'Tue', 'Wed', 'Thu', 'Fri', 'Sat', 'Sun', 'Mon.', 'Tue...'], the x-axis will display the days of the week. The y-axis will show the quantity of flights, with values denoting the number of flights on each day. Trends in flight frequency will be revealed by connecting points that correlate to the daily flight counts. It is simple to identify changes in flight volume throughout the week since any discernible patterns, such as increased aircraft activity on weekdays as opposed to weekends, will be shown by changes in the line's slope.

H. NO. OF FLIGHTS ON MONTHS

There was an upward trend from January to July, suggesting that summertime aviation activity increased. There was a minor drop in August and a steady decline toward the end of the year, suggesting that there would be fewer flights in the fall and winter. months. A popular travel season may be indicated by peaks in the summer months (such as July).

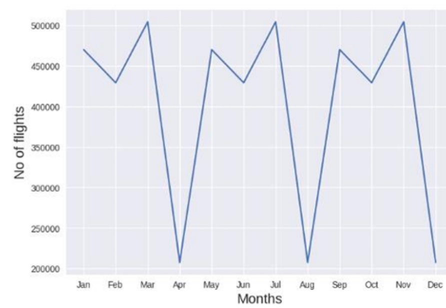

Fig. 7. Flights on Months

A line graph showing the quantity of flights for each month of the year is shown in Fig.7. While the y-axis shows the number of flights, with values corresponding to the number of flights in each month, the x-axis shows the names of the months, illustrating the time frame throughout the year. A line connecting points that show the number of flights for each month makes it possible to see patterns in aviation activity all year round. This graph offers insights on variations in air travel throughout the year by assisting in the identification of any seasonal patterns, such as increased flight volumes during specific months.

I. AIRLINES AND DELAY

Airlines are essential to the aviation sector, and flight delays are greatly impacted by their operational effectiveness. Airline management procedures, fleet maintenance, scheduling, and crew availability are some of the variables that affect delays. Internal airline problems, like incoming flights arriving late, can cause flight delays, which can then affect other departures. Airlines also have to coordinate with airport operations and air traffic control, and any misalignment can cause delays. Additionally, different airlines are affected by weather conditions in different ways; for instance, some may have backup plans that enable them to adjust to bad weather more skillfully than others.

Flight dispersion by airline is depicted in a pie chart in Figure 8. Each of the variously colored slices that make up the graphic represents a distinct airline. Additionally, the corresponding slice shows the percentage of all flights that

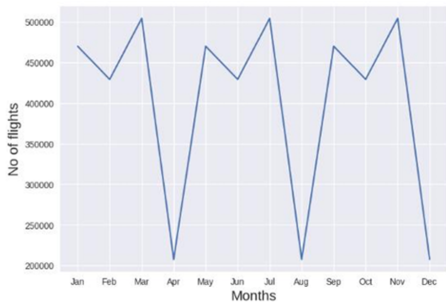

Fig. 8. Airlines and Delay

each airline represents. With 60 percent of the flights, Airline B may have a sizable slice in the resulting pie chart, suggesting that it operates the bulk of the dataset's flights. For Airline A, a moderate slice is 30 percent, whereas for Airline C, a tiny slice is 10 percent. The market share of every airline in the dataset is clearly displayed in this visualization.

V. MODEL TRAINING AND SCORE

there are several causes of flight cancellations and delays, and these causes are frequently divided into different categories. These causes are graphically shown by two pie charts that clearly break down the causes of cancellations and delays. Extreme weather phenomena, such as storms or fog, can cause flight schedule disruptions, making weather conditions a common cause of delays. Delays are also caused by airline operations, frequently as a result of inefficient turnaround times, personnel schedule conflicts, or aircraft maintenance. Airport gate availability, runway congestion, and air traffic control limitations are the main causes of National Air System delays. Last but not least, increased security screenings or possible threats cause security-related delays, albeit they are less common. Weather and airline problems are the main causes of cancellations, which frequently follow similar trends. By providing a thorough understanding of the ways in which these different elements impact flight punctuality, these pie charts assist stakeholders in addressing the underlying causes of interruptions.

A. RANDOM FOREST CLASSIFIER

The model's performance is assessed in the confusion matrix below, which displays the distribution of true positives, true negatives, false positives, and false negatives. 54,756 are the true negatives (top-left). 3,055 false positives (top-right) False negatives = 8,338 (bottom-left). 24,309 true positives (bottom-right) Higher values are represented by darker hues in the color-coded matrix.

A number of performance indicators and their corresponding values with an accuracy of 87.40 percent, the model is generally accurate. The area under the ROC curve, or ROC AUC, is 84.59 percent, which indicates how well the model can differentiate between classes. The precision, or the percentage of genuine positives among all anticipated positives, is 88.84

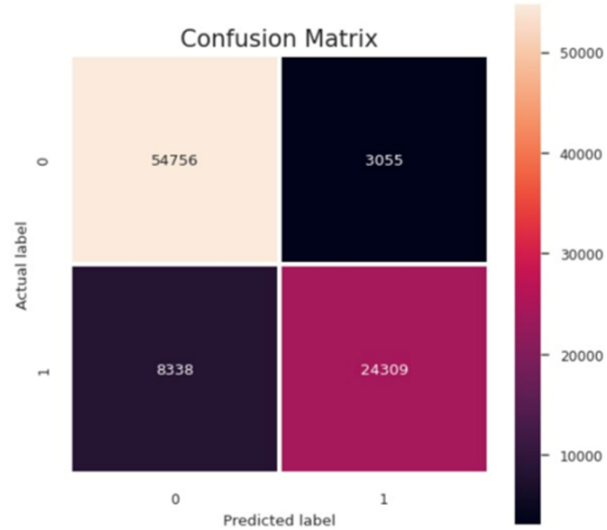

Fig. 9. Confusion Matrix of Random Forest Classifier

percent. The percentage of true positives that were successfully detected is 74.46 percent, or recall. Error: 15 percent, which is the percentage of inaccurate predictions and is determined as the complement of accuracy. The table offers a succinct overview of the main performance indicators, showing that while the Random Forest Classifier has a high precision and a respectable accuracy, its recall may use some work (indicating some missed true positives). Recall and precision appear to be balanced in the overall performance.

PERFORMANCE METRICS	VALUES
ACCURACY	87.40%
ROC_AUC	84.59%
PRECISION	88.84%
RECALL	74.46%
ERROR	15%

Fig. 10. Performance Metrics

A classification model, more especially a Random Forest classifier, is assessed for performance using the Receiver Operating Characteristic (ROC) curve. The True Positive Rate (also known as sensitivity or recall) is plotted on the y-axis of the ROC curve versus the False Positive Rate on the x-axis. A random (no-skill) classifier's performance is shown by the orange dashed line, but the Random Forest model's performance is shown by the blue curve. The blue line shows a strong model with good discriminative capacity as it curves steeply towards the top-left corner. The model performs exceptionally well in differentiating between the positive and negative classes, as indicated by the high area under the ROC curve (AUC) of 0.93. The dashed diagonal line shows the AUC of a model with no skill, which is 0.5, whereas the AUC of a flawless classifier is 1.0. The better the model does, the closer the curve is to the upper-left corner

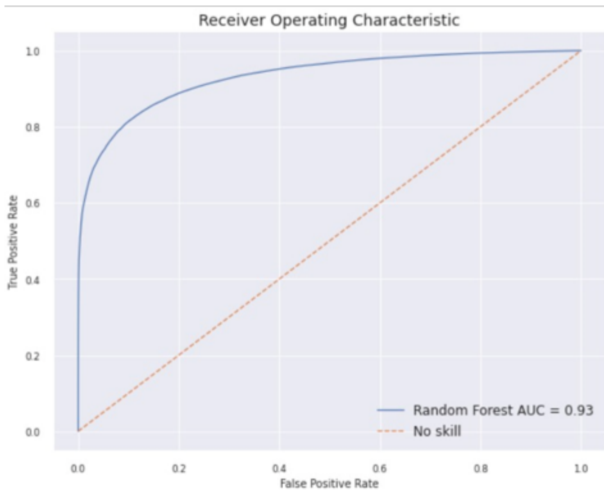

Fig. 11. ROC Curve For Random Forest Classifier

B. LOGISTIC REGRESSION MODEL

Logistic regression is a classification technique that use one or more input features to estimate the likelihood of a binary outcome (such as yes/no or 0/1). Despite linear regression, which predicts a continuous output, logistic regression uses the logistic (sigmoid) function to convert predictions to a 0–1 range. This function applies a nonlinear adjustment to the linear combination of features, resulting in a "S"-shaped curve that limits outputs to probabilities.Logistic regression is a statistical method that predicts binary outcomes using one or more predictor variables.

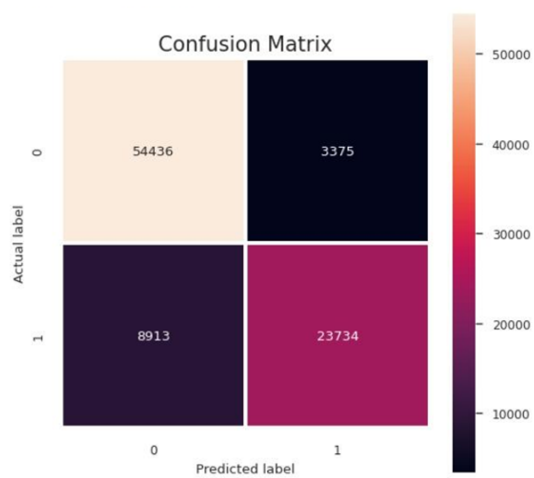

Fig. 12. Confusion Matrix of Logistic Regression

Understanding the classification results is possible thanks to the confusion matrix: The model accurately predicted class 0 in 54,436 cases, which is the True Negatives (Top-left).False Positives (top-right): 3,375 cases in which the model predicted class 1 when it really predicted class 0.8,913 cases of false negatives (bottom-left) occurred when the model predicted class 0 for actual class 1.Bottom-right True Positives: 23,734 cases in which class 1 was accurately predicted by the model.84.41 percent accuracy means that 84.41 percent of the cases were properly classified by the model.

PERFORMANCE METRICS	VALUES
ACCURACY	88.41%
ROC_AUC	86.12%
PRECISION	88.66%
RECALL	77.87%
ERROR	12%

Fig. 13. Performance Metrics

Although the recall indicates that there may be space for improvement in identifying good events, the accuracy and precision are strong. The logistic regression model's Receiver Operating Characteristic (ROC) curve is shown in Fig. 5.4. The logistic regression model, shown by the blue curve, differs dramatically from the diagonal dashed line, which stands for a no-skill classifier. The logistic regression model's AUC (Area Under the Curve) is 0.92, suggesting that it performs well in differentiating between the positive and negative classifications. An ideal model would have an AUC of 1.0, whereas the AUC of a random model would be 0.5.

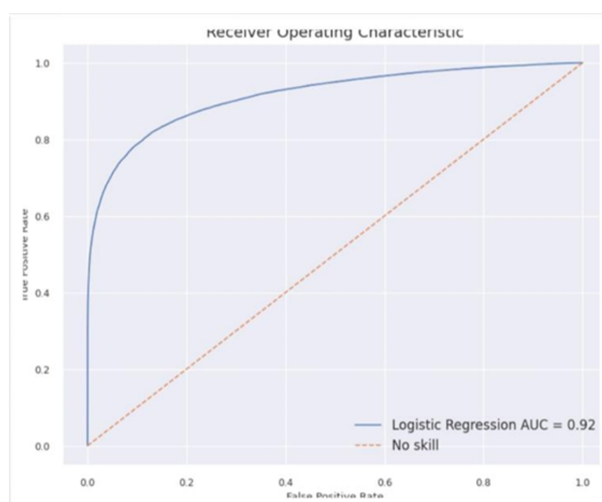

Fig. 14. ROC Curve for Logistic Regression

C. XGBoost Classifier

A confusion matrix generated for an XGBoost model, depicting the model's classification performance on a binary classification task. An XGBoost model's confusion matrix, which shows how well the model classified data on a binary classification problem. The matrix provides a summary of the essential details listed below: 54,561 cases were accurately assigned to the negative class (0), or True Negatives (TN). False Positives (FP): 3,250 cases that were truly negative were mistakenly labeled as positive (1). False Negatives (FN): 7,226 cases that should have been positive were mistakenly labeled as negative (0). True Positives (TP): The positive class (1) was accurately identified in 25,421 cases.

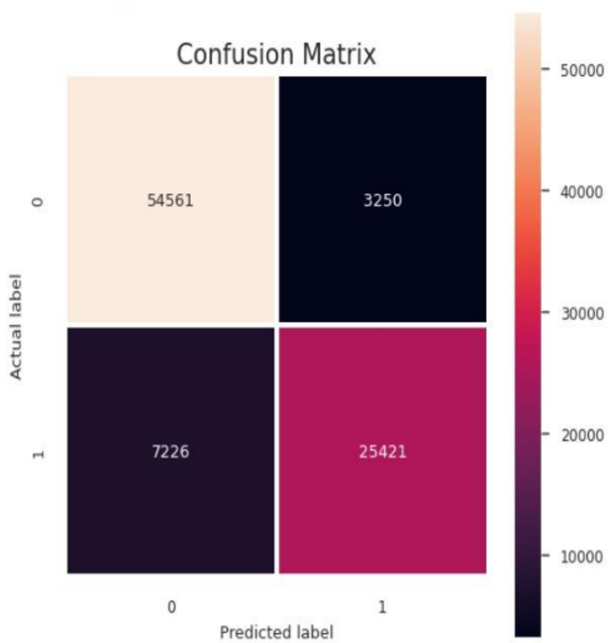

Fig. 15. Confusion Matrix for XGBoost Classifier

The table presents the performance metrics for an XGBoost classification model, highlighting key evaluation indicators: XGBoost classification model, emphasizing important assessment metrics: Accuracy: 88.41 percent: This figure shows that 88.41 percent of all cases were correctly identified by the model. Although accuracy is a broad indicator of a model's overall performance, it does not always accurately represent efficacy in datasets that are not balanced. ROC AUC: 86.12 percent The model's capacity to differentiate between the two classes is assessed by the ROC-AUC score. With a score of 0.8612, one can clearly distinguish between both good and negative examples, where better performance is indicated by higher scores. Precision: 88.66 percent: This means that 88.66 percent of the model's predictions were accurate when the positive class was anticipated. It focuses on whether positive forecasts are accurate. Recall: 77.87 percent: This metric assesses how well the model detects real positive examples. A recall of 77.87 percent indicates that 77.87 percent of all positive cases are captured by the model. Error: 12 percent: The error rate is the percentage of cases that are misclassified, with 12 percent of predictions being off.

PERFORMANCE METRICS	VALUES
ACCURACY	88.41%
ROC_AUC	86.12%
PRECISION	88.66%
RECALL	77.87%
ERROR	12%

Fig. 16. Performance Metrics

According to these criteria, the model has a high overall accuracy of 40 and strikes a fair compromise between precision and recall. An XGBoost (XGB) classifier's Receiver Operating Characteristic (ROC) curve. The XGBoost classifier's ROC curve, which shows how well the model differentiates between classes, is represented by the solid blue line. A superior model has a hugging curve. the upper-left corner. A random or "no skill" classifier is represented by the dashed orange line, which has a baseline Area Under the Curve (AUC) of 0.5. The XGBoost classifier's Area Under the Curve (AUC) is 0.93, indicating exceptionally strong model performance.

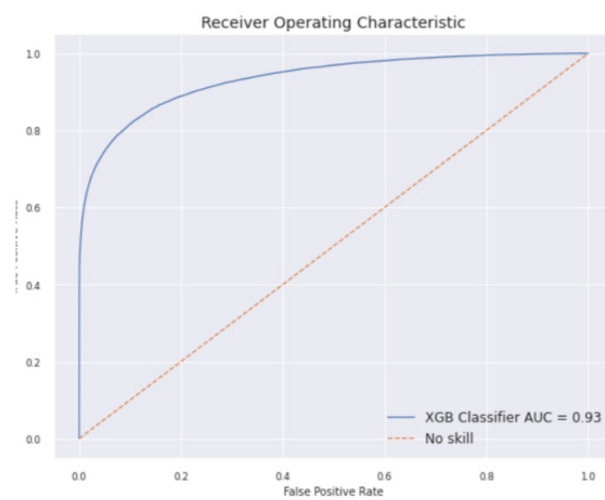

Fig. 17. ROC Curve for XGBoost

D. DECISION TREE CLASSIFIER

This confusion matrix provides a thorough analysis of the decision tree model's performance in splitting a dataset into two groups. True Negatives (48808): These indicate how many cases were accurately assigned to class 0. Untrue Positives (9003): These are instances in which the label was genuinely

0 but the model mispredicted it as 1. False Negatives (8360): Missed positive detections resulted from cases that were actually 1 but were projected to be 0. Instances where the model accurately predicted the actual label as 1 are known as True Positives (24287).

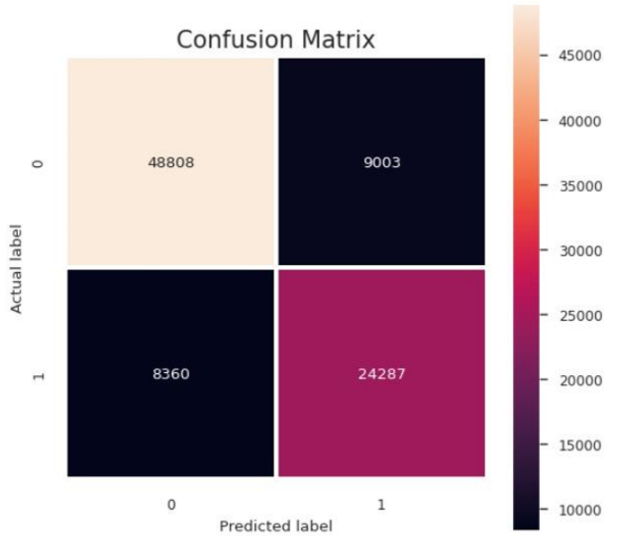

Fig. 18. Confusion Matrix for Decision Tree Classifier

A structured presentation of the Decision Tree Classifier includes a list of important evaluation criteria that are used to gauge how well the model predicts outcomes. The Decision Tree model's performance is clearly shown in the table. demonstrating respectable general classification skills. Nonetheless, the 19 percent mistake rate suggests that there is still opportunity to decrease the quantity of inaccurate forecasts.

PERFORMANCE METRICS	VALUES
ACCURACY	80.80%
ROC_AUC	79.41%
PRECISION	72.96%
RECALL	74.39%
ERROR	19%

Fig. 19. Performance Metrics

A classifier that predicts each instance's class at random is shown by the diagonal line. It has a 0.5 AUC. The ROC curve's form indicates how effectively the model strikes a balance between specificity and sensitivity. A steeper curve A curve that is steeper implies

An improved balance between the twoThe curve Nearer the Upper-Left Corner: This suggests a highly sensitive and particular model. AUC Range: The range of the AUC is 0 to 1.AUC = 1: Complete categorization. AUC = 0.5: Guesswork at random. Interpretation of AUC: AUC greater than 0.5 indicates that the model outperforms random guessing. AUC less than 0.5: The model's performance is inferior to that of random guesswork.

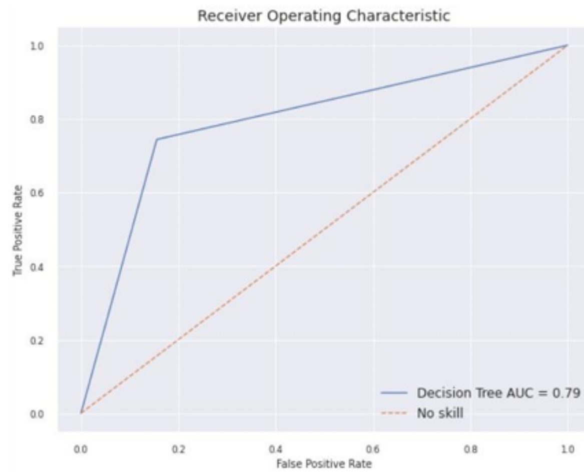

Fig. 20. ROC Curve for Decision Tree Classifier

E. NAIVE BAYES CLASSIFIER

The model's predictions are visualized in the confusion matrix. It demonstrates how many cases were classified accurately (false positives, false negatives) and how many were misclassified (true positives, true negatives). The values found in each The number of instances that fall into a specific combination of actual and expected labels is indicated by the cell.

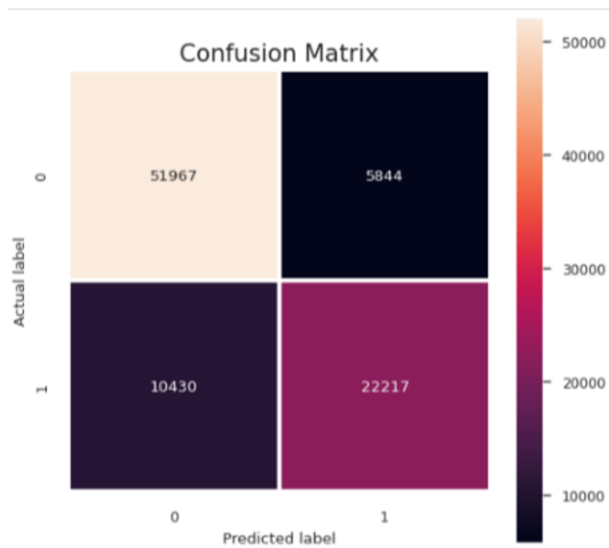

Fig. 21. Confusion Matrix for Naïve Bayes Classifier

A machine learning model's performance metrics are compiled in the table that is presented. Precision: Number: 82.00

percent; ROCAUC: 86.00 percent. Value for Precision: 75.90 percent This quantifies the percentage of cases that are truly positive compared to those that were anticipated to be positive. There are fewer false positives when precision is higher. Value: 73.00 percent, Recall Value: 11 percent error Interpretation: This represents the model's error rate, which is equal to 100 percent minus accuracy.

PERFORMANCE METRICS	VALUES
ACCURACY	82.00%
ROC_AUC	86.00%
PRECISION	75.90%
RECALL	73.00%
ERROR	11%

Fig. 22. Performance Metrics

A classifier that predicts each instance's class at random is shown by the diagonal line in Fig.23. It has a 0.5 AUC. The ROC curve's form indicates how effectively the model strikes a balance between specificity and sensitivity. A steeper curve is one that curve suggests that the two are better traded off. A model with high sensitivity and specificity is shown by a curve that is closer to the top-left corner.

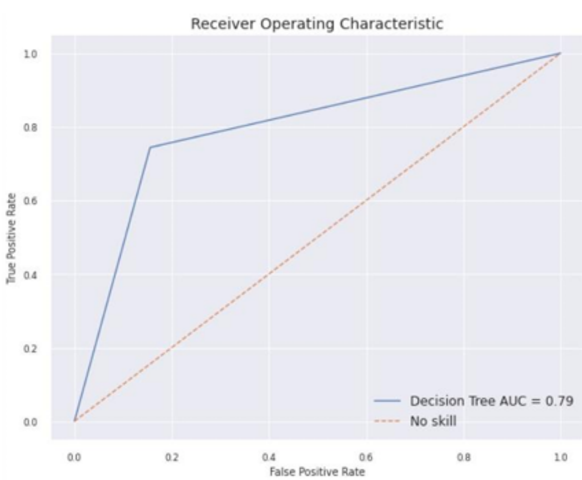

Fig. 23. ROC Curve For Naïve Bayes Classifier

F. RESULT COMPARISON

The result comparison among all the models represented in the table Fig.24 presents the performance metrics — accuracy, precision, and recall of five machine learning models evaluated on a classification task. The Random Forest model is the most effective option for predicting sleep disorders, according to performance measures. It has the highest accuracy (95.9 percent), precision (98.8 percent), and strong recall (93.2 percent). This suggests that Random Forest is very good at reducing false negatives, which is important for thorough disorder prediction, and it is also very accurate. With 98.4 percent precision and 95.1 percent accuracy, Neural Networks likewise performs well, but Random Forest outperforms it due to its slightly lower recall (92.1 percent). While K-Nearest Neighbours and Support Vector Machine produce competitive results, their recall scores are marginally lower (90.2 percent and 94.3 percent, respectively), and Logistic Regression lags behind the other models with an accuracy of 91.1 percent. Random Forest is therefore the best option due to its balanced and reliable performance.

Model	Accuracy	Recall	F1 score
Logistic regression	0.8641	0.73	0.79
Decision tree classifier	0.8080	0.74	0.73
Random forest	0.8740	0.74	0.80
Naïve Bayes Classifier	0.8200	0.73	0.86
XG Boost	0.8841	0.78	0.81

Fig. 24. Result Comparison Table

The highest performance results were achieved by the Random Forest Classifier and the XGBoost Classifier across all evaluation criteria. The accompanying image presents a Receiver Operating Characteristic (ROC) curve Fig. 25, illustrating the effectiveness of various classification models. This curve plots the True Positive Rate (sensitivity) against the False Positive Rate (1-specificity), visually representing each model's ability to distinguish between classes. The Area Under the Curve (AUC) quantifies this performance; a higher AUC value reflects better classification capability. In this case, the decision tree model has the lowest performance, with an AUC of 0.79, while the random forest classifier demonstrates the best results with an AUC of 0.94.

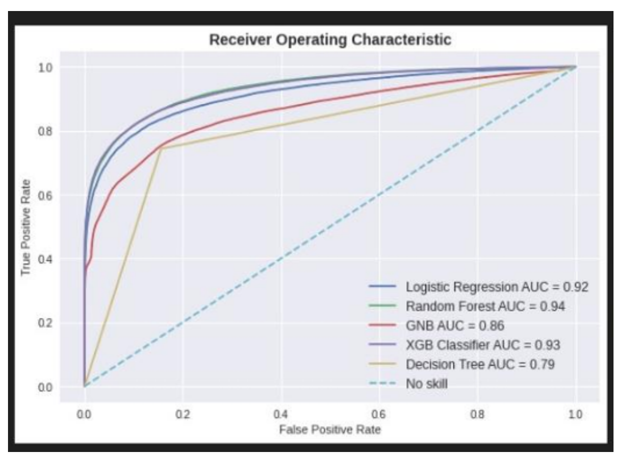

Fig. 25. Combined ROC Curve For All Model

References

[1] Abadie, L. M., Galarraga, I and Ruiz-Gauna, I. (2022). Flight delays in Germany:a model for evaluation of future cost risk. European Journal of Transport and Infrastructure Research, 22(1), 93-117.

[2] Hatıpoğlu, I., Tosun, Ö and Tosun, N. (2022). Flight delay prediction based with machine learning. LogForum, 18(1).

[3] Khan, R and Zahid, T. (2022). Flight Delay Prediction Based on Gradient Boosting Ensemble Models (No. 8511).

[4] Ayaydın, A and Akcayol, M. A. (2022). Deep learning based forecasting of delay on flights. Bilişim Teknolojileri Dergisi, 15(3), 239-249.

[5] Feng'e, L., Xiangyu, K., Fang, Q and Bo, W. (2022). Research on Flight Adjustment Optimization Based on Flight Delay Forecast. Industrial Engineering and Innovation Management, 5(7), 8-15.

[6] Mokhtarimousavi, S and Mehrabi, A. (2023). Flight delay causality: Maching learning technique in conjunction with random parameter statistical analysis. International Journal of Transportation Science and Technology, 12(1),230-244.

[7] Bisandu, D. B and Moulitsas, I. (2024). Prediction of flight delay using deep operator network with gradient-mayfly optimisation algorithm. Expert Systems With Applications, 247, 123306

[8] Sternberg, A., Soares, J., Carvalho, D and Ogasawara, E. (2017). A review on flight delay prediction. arXiv preprint arXiv:1703.06118.

[9] Kokkiligadda Rajesh, Dr. Srikanth V (2022), Predicting Flight Delays Using Machine Learning: An Analysis of Comprehensive Data and Advanced Techniques,International Journal of Advanced Research in Computer and Communication Engineering, Vol. 12, Issue 4.

[10] Vinayagar, M. and Scholar, U. G. A MODEL FOR FORECASTING FLIGHT DELAY USING GRID SEARCH CV TECHNIQUE.

[11] Zeng, W., Li, J., Quan, Z and Lu, X. (2021). A Deep Graph-Embedded LSTM Neural Network Approach for Airport Delay Prediction. Journal of Advanced Transportation, 2021(1), 6638130.

[12] Wang, F., Bi, J., Xie, D and Zhao, X. (2022). Flight delay forecasting and analysis of direct and indirect factors. IET Intelligent Transport Systems, 16(7), 890-907.

[13] Yi, J., Zhang, H., Liu, H., Zhong, G and Li, G. (2021). Flight delay classification prediction based on stacking algorithm. Journal of Advanced Transportation, 2021(1), 4292778

[14] Schösser, D and Schönberger, J. (2022). On the Performance of Machine Learning Based Flight Delay Predictio Investigating the Impact of Short-Term Features. Promet-Traffic and Transportation, 34(6), 825-838.

[15] Paramita, C., Supriyanto, C., Syarifuddin, L. A and Rafrastara, F. A. (2022). The Use of Cluster Computing and Random Forest Algoritm for Flight Delay Prediction. International Journal of Computer Science and Information Security (IJCSIS), 20(2).

[16] Borse, Y., Jain, D., Sharma, S., Vora, V and Zaveri,A. (2020). Flight Delay Prediction System. Int. J. Eng. Res. Technol, 9(3), 88-92.

[17] Alharbi, B and Prince, M. (2020). A hybrid artificial intelligence approach to predict flight delay. International Journal of Engineering Research and Technology, 13(4), 814-822.

[18] Qiang Li, Ranzhe Jing, Flight delay prediction from spatial and temporal perspective,Expert Systems with Applications,Vol 205,2022, 117662,

[19] Cai, K., Li, Y., Fang, Y. P and Zhu, Y. (2021). A deep learning approach for flight delay prediction through time-evolving graphs. IEEE Transactions on Intelligent Transportation Systems, 23(8), 11397-11407.

[20] Kothari, R., Kakkar, R., Agrawal, S., Oza, P., Tanwar, S.,Jayaswal, B and Bokoro, P. N. (2023). Selection of best machine learning model to predict delay in passenger airlines. IEEE Access