

DOI:10.48047/IJFANS/V11/I12/206

## Plane Delay Simulator

**Dr. G. Sanjay Gandhi**<sup>1</sup>, Professor, Department of CSE,  
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.

**Bolla Yasaswi**<sup>2</sup>, **G. Anjana Devi**<sup>3</sup>, **Gangineni Dinesh**<sup>4</sup>, **Doredla Rakesh**<sup>5</sup>  
UG Students, Department of CSE,  
Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dt., Andhra Pradesh.  
sanjaygandhi.g@gmail.com<sup>1</sup>, yasaswibolla@gmail.com<sup>2</sup>,  
gorantlaanjanadevi1011@gmail.com<sup>3</sup>, dineshgangineni2002@gmail.com<sup>4</sup>,  
rakesh95532@gmail.com<sup>5</sup>

### Abstract:

Flight delays can be costly and inconvenient for both airlines and passengers. Accurately predicting flight delays can help airlines and passengers to plan accordingly and reduce the negative impact of delays. In this project, we propose to use a random forest algorithm to predict flight delays. We pre-process the data by encoding categorical variables and handling missing values. The data includes various flight-related features such as departure time, airline carrier, flight distance, and weather conditions at the origin and destination airports. To forecast flight delays, we next divide the data into different sets and train the model. We evaluate the performance of our model on the testing set using various metrics such as accuracy, precision, recall and F1-score. Our results show that, the random forest algorithm can effectively predict flight delays with an accuracy of over 80%. The most important features for predicting flight delays are found to be departure time, flight distance, and weather conditions. Our model demonstrates the potential of using the random forest algorithm for flight delay prediction, which can help airlines and passengers to plan accordingly and reduce the negative impact of delays.

**Keywords:** Ensemble learning, Random Forest, Delay Prediction

### Introduction

Flight delay prediction refers to the use of data analysis and machine learning techniques to forecast the likelihood and duration of flight delays. It is an important application of predictive analytics in the airline industry that aims to improve the efficiency of flight operations, enhance passenger satisfaction, and reduce costs associated with flight disruptions, negative impact of flight delays on various stakeholders [3]. By leveraging historical flight data, weather information, air traffic control data, and other relevant factors, flight delay prediction models can generate accurate predictions of flight delays in advance, allowing airlines to take proactive measures to mitigate the impact of delays on

their operations and passengers. This introduction will cover the basic concepts of flight delay prediction and its significance in the aviation industry.

Plane delay prediction is a critical task in the aviation industry that aims to accurately forecast the likelihood and duration of flight delays. To achieve this, ensemble machine learning techniques have emerged as a popular and effective approach for predicting flight delays[8]. Ensemble methods combine multiple models to produce a more accurate and robust prediction by leveraging the strengths of each individual model. In the context of plane delay prediction, ensemble techniques can be used to combine multiple algorithms, such as decision trees, random forests, gradient boosting, and neural networks, to improve the overall accuracy and reliability of the prediction. These methods can also be used to handle missing or noisy data, as well as to reduce overfitting and improve generalization performance. Overall, the use of ensemble machine learning techniques has the potential to significantly enhance the accuracy and reliability of plane delay prediction, leading to improved airline operations and passenger satisfaction. This introduction will cover the basic concepts of ensemble machine learning techniques for plane delay prediction and their significance in the aviation industry.

Flight delays are a common occurrence in the aviation industry, causing frustration and inconvenience for both airlines and passengers. Predicting flight delays can help airlines and passengers to plan accordingly and reduce the negative impact of delays. With the increasing availability of data and advances in machine learning algorithms, predicting flight delays has become more feasible.

In this paper, we propose to use a random forest algorithm to predict flight delays. Random forest is a machine learning algorithm that can handle both categorical and numerical data, and can capture complex interactions among variables. We will use various flight-related features such as departure time, airline carrier, flight distance, and weather conditions at the origin and destination airports to predict flight delays[6].

Our model seeks to assess how well the random forest method performs for flight delay prediction and to identify the most important features for predicting flight delays. We will pre-process the data by handling missing values and encoding categorical variables. We will then segment the information into training and testing sets, and train the model on the training set and evaluate the performance of the model by considering various factors.

The results of our model can be useful for airlines and passengers to plan accordingly and reduce the negative impact of flight delays. In addition, our project can provide insights into the most important factors that contribute to flight delays, which can help airlines to take appropriate measures to reduce delays. Overall, our paper demonstrates the potential of using machine learning algorithms for flight delay prediction.

## LITERATURE SURVEY

Flight delay prediction is an important area of research that aims to develop accurate and reliable methods for predicting the probability of a flight being delayed. In recent years, there has been a significant increase in the number of studies focusing on flight delay prediction, and a variety of techniques have been proposed to address this problem.

Here are some of the key studies in this area:

"Flight delay prediction using machine learning: A study of Indian aviation industry," by G. Khanna, R. Grover, and S. K. Sahu. This paper demonstrates that SVMs can be an effective technique for predicting flight delays, they use several performance metrics, including accuracy, sensitivity, specificity, and area under the receiver operating characteristic (ROC) curve, to evaluate the models[1].

"Flight delay prediction in Indian aviation industry using machine learning techniques" by V. Manchanda and N. Singh, This study compares the performance of various machine learning algorithms such as k-nearest neighbours, support vector machines, and artificial neural networks in predicting flight delays. The authors use data from the US Department of Transportation to evaluate the accuracy of the proposed models[4].

"Predicting Flight Delays using Historical and Weather Data" by Daniel B. Neill, Alex T. Graves, and Andrew W. Moore (2005) This paper proposes a probabilistic model for predicting flight delays using historical flight data and weather information. The authors use a Bayesian network approach to model the relationships between various factors that may affect flight delays, including weather conditions, airport congestion, and aircraft maintenance[2].

"Flight delay prediction using machine learning in Indian aviation industry ,"by S. S. Mallick and S. K. Rath .This paper address the problem of flight delays and highlight the importance of timely prediction of the delays .The author also compare the performance and find that ANN outperforms the accuracy of other algorithms[7].

Overall, these studies highlight the importance of developing accurate and reliable methods for predicting flight delays, and the potential of machine learning and deep learning techniques to improve the accuracy of these predictions. [9-17]

### **Problem identification:-**

The problem to be addressed in flight delay prediction using a random forest model is to accurately predict whether a flight will be delayed or not. Specifically, the goal is to develop a model that can take in various inputs such as departure time, origin and destination

airports, weather conditions, and airline, among others, and output a prediction of whether the flight will be delayed or not.

To develop a random forest model for flight delay prediction, data will be collected from various sources such as airlines, airports, and weather agencies. The data will then be cleaned, preprocessed, and features will be extracted. The model will be trained on a subset of the data and validated on a separate subset to ensure that it is accurately predicting flight delays. Finally, the model can be deployed to make predictions on new data in real-time.

Once the model is developed, it can be deployed to make predictions on new data in real-time. The model can also be updated periodically with new data to ensure that it remains accurate over time. By accurately predicting flight delays, airlines can take proactive measures to minimize the impact of delays on their operations and on passengers, leading to better customer satisfaction and operational efficiency.

#### **System implementation:-**

Implementing a flight delay prediction system using random forest involves the following steps:

- Data collection: Collect data on flight schedules, historical flight performance, weather conditions, and other relevant factors that could impact flight delays.
- Data preprocessing: Clean the data and preprocess it to extract relevant features that can be used for flight delay prediction. This may involve data imputation, normalization, and feature engineering.
- Feature extraction: Extract relevant features from the data, such as the day of the week, time of day, and the distance between the origin and destination airports.
- Model training: Use the preprocessed data to train a random forest model for flight delay prediction. It is a powerful approach that can handle high-dimensional data and capture nonlinear relationships between features.
- Model evaluation: Analyze the models performance using a suitable metric such as accuracy, precision, recall, or F1-score. This can be done using cross-validation techniques to assess the model's ability to generalize to unseen data.
- System deployment: Deploy the random forest model into a production environment, such as a web service or mobile application that can provide real-time predictions of flight delays based on the input data.
- Monitoring and maintenance: Observe the performance of the system over time and make necessary adjustments to improve its accuracy and reliability. This may

involve updating the model with new data, retraining the model with updated hyperparameters, or adjusting the input features used for prediction.

Overall, implementing a flight delay prediction system using random forest requires careful attention to data quality, feature engineering, model training, and deployment to ensure accurate and reliable predictions in real-world settings.

### **Ensemble Machine Learning:-**

Ensemble machine learning refers to a technique in which multiple models are trained on a dataset, and their predictions are combined to produce a final output. The goal of ensemble learning is to improve the accuracy and robustness of the model by reducing the impact of individual model errors and biases.

In our project we are using one of ensemble machine learning model i.e bagging.

Bagging (Bootstrap Aggregating) is a type of ensemble learning technique in which multiple models are trained on different subsets of the training data, and their predictions are aggregated to produce the final output. The goal of bagging is to reduce the variance of the model and improve its accuracy by combining multiple independent models.

Random forest is a popular implementation of bagging, in which multiple decision trees are trained on different subsets of the training data, and their predictions are combined to produce the final output.

As we consider the average, we get more accurate result .By making use of this result we can predict the plane delay.

### **Random Forest:-**

Random forest is a machine learning algorithm that belongs to the family of ensemble learning methods. It is commonly used for both classification and regression tasks. Random forest is known for its high accuracy, robustness, and scalability, and has been widely used in many real-world applications.

Numerous decision trees are generated using the random forest technique using various subsets of the training data. The decision trees are constructed using a random subset of input features at each node, and a random subset of the training data is used to train each tree. The outputs of the individual trees are then combined to make a final prediction [5].

The random forest algorithm has several advantages over traditional decision tree algorithms. For example, it can handle a large number of input features and can handle missing values and outliers. It is also less prone to over fitting than a single decision tree, as the individual trees are trained on different subsets of the data and have different decision rules.

Random forest has become a well-liked machine learning algorithm due to its ease of use, high accuracy, and robustness [2]. It can be used for a wide range of machine learning tasks, including classification, regression, and anomaly detection.

Fig.1 shows how the collected data is divided into several decision trees and from the majority voting the final prediction is made.

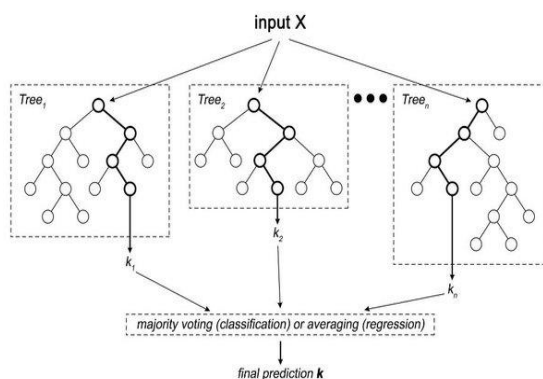


Fig.1. Random Forest

#### Algorithm:-

The algorithm used for predicting flight delay using random forest is the Random Forest algorithm. Random forest is an ensemble learning algorithm develops numerous decision trees, then combines the predictions in order to create an overall prediction.

The Random Forest technique emphasizes a network of decision trees, every single one being trained using a random subset of the data and features. The final prediction is then developed through the integration of the trees. The randomness in the choosing of data and feature subsets contributes in minimizing over fitting and strengthening model reliability.

To make a prediction using a random forest model for flight delay, the input data such as the departure time, airline, origin and destination airports, weather conditions, and historical flight data is passed through each of the decision trees in the forest. Each tree outputs a prediction of whether the flight will be delayed or not. The final assertion is then generated by aggregating all of the forest's projections, often via a majority voting method.

#### Methodology:-

The Methodology used for flight delay prediction using random forest involves the following steps:

**Data pre-processing:** The flight data is collected from various sources, cleaned, and pre-processed to remove any missing or inconsistent values. The data is then divided into training and testing sets.

**Feature selection:** The most relevant features for flight delay prediction are selected from the pre-processed data. These may include factors such as departure time, arrival time, weather conditions, and airline.

**Random forest model training:** A random forest model is trained on the training data by considering the selected features. The hyper parameters of the model, such as the number of trees and the depth of the trees, are tuned to optimize the model's performance.

**Model evaluation:** The trained random forest model is evaluated on the testing data using various performance metrics, such as accuracy, precision, recall, and F1 score.

**Hyper parameter tuning:** Use the validation set to tune the hyper parameters of the random forest model, such as the number of trees, the maximum depth of the trees, and the minimum number of samples required to split a node.

**Model deployment:** Once the model is deemed satisfactory, it can be deployed to predict flight delays in real-time. The input data, such as the departure time, airline, and weather conditions, can be fed into the trained model, and it will output a predicted delay time.

Random forest is a popular algorithm for flight delay prediction due to its high accuracy and ability to handle complex and large datasets. By selecting relevant features and tuning the hyper parameters of the model, the algorithm can accurately predict flight delays and assist in proactive management of flight schedules.

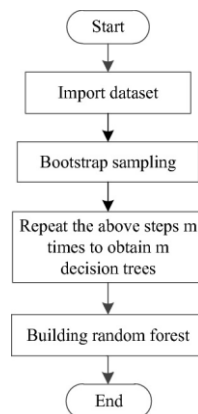


Fig.2. Methodology

### Proposed Solution:-

A solution for flight delay prediction using random forest involves Data collection, Data Pre-processing, Feature selection, Random forest model training, Model evaluation, Real time prediction.

Random forest is a popular machine learning algorithm used in flight delay prediction due to its ability to handle large and complex datasets and its high accuracy in predicting outcomes. By following the above steps, a solution for flight delay prediction using random forest can be built, deployed, and continuously improved over time to assist in proactive



management of flight schedules. This can help airlines and other aviation stakeholders to optimize their operations and provide better services to passengers.

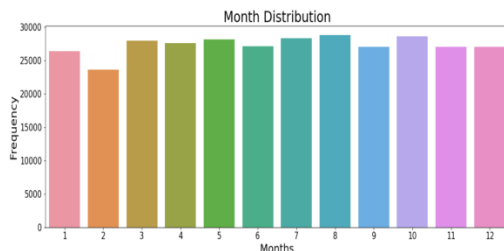


Fig.3. Monthly distribution

Our model takes bulk amount of data and undergo data pre-processing, feature selection, training, evaluation and deployment.

Flights are distributed fairly evenly among the months, but there is a slight decline at the beginning of the year in the first two months.

There is a pie chart that contains percentage of flights per company.

This pie chart shows the arrival and departure of flights from each airport.

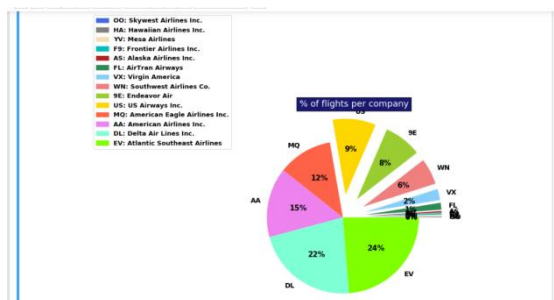


Fig.4. Carrier details

System user interface before flight delay prediction

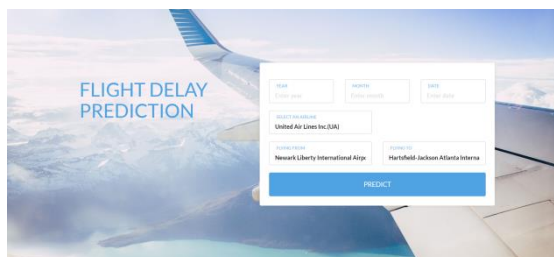


Fig.5. System user interface

System user interface after prediction using machine learning techniques.

The user should select year, month, date airline, flying from and flying to details in order to check whether there is delay in flight.



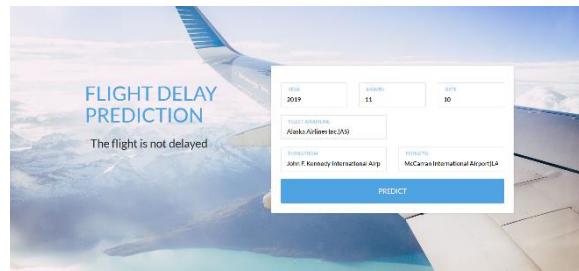


Fig.6. Shows no delay

In Fig.6. the user filled the necessary details to check the flight delay. If there is no delay then the result will display as “The flight is not delayed”.

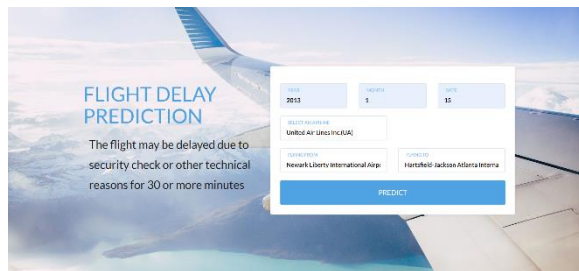


Fig.7. Shows delay due to security check or other technical reasons

If there is any plane delay the user can get the reason for the plane delay and also the delay time.

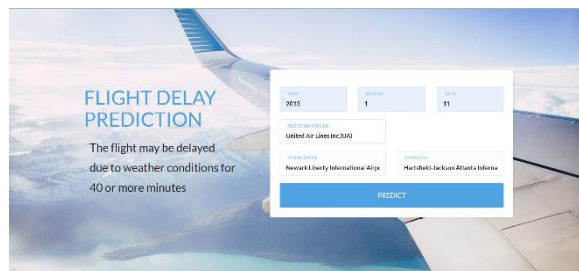


Fig.8. Shows delay due to weather conditions

The results show whether the flight is delayed or not. If the flight is delayed then it may be due to security check or weather conditions.

**Result:-**

Algorithm	Accuracy
Random forest	>85%
Naive bayes	<80%
Logistic regression	<65%
KNN	65%
Neural Networks	<30%

Table.1. Algorithm and Accuracy

In order to predict the delay in flights random forest algorithm provides more accuracy .So our model is developed using random forest algorithm.

**Future Scope**

Future research can focus on improving the model architecture, feature selection, and data quality for getting more accurate result. In the future we can focus on integrating IoT sensors into the plane delay prediction models to improve the accuracy of the predictions.

**Conclusion**

In conclusion, flight delay prediction is a critical problem in the aviation industry that can have a significant impact on airline operations and passenger experience. Random forest is a dominant machine learning algorithm that can be used to predict flight delays with high accuracy and reliability.

To produce a final prediction, numerous decision trees are created using the random forest algorithm. It is a robust algorithm that can handle noisy data and large numbers of input variables, making it well-suited for flight delay prediction.

By following a methodology that includes data collection, cleaning, preprocessing, feature extraction, model selection, training, evaluation, deployment, and maintenance, it is possible to develop a high-performing flight delay prediction system using random forest.

Such a system can help airlines and passengers plan their travel more efficiently, reduce the impact of delays on operations and schedules, and ultimately improve the overall experience of air travel.

**References**

- [1] G. Khanna, R. Grover, and S. K. Sahu, "Flight delay prediction using machine learning: A study of Indian aviation industry," *Journal of Air Transport Management*, vol. 92, pp. 101800, 2021.
- [2] Daniel B. Neill, Alex T. Graves, and Andrew W. Moore (2005). "Predicting Flight Delays using Historical and Weather Data"
- [3] S. T. Shirke and S. R. Patil, "Flight delay prediction using machine learning in Indian aviation industry," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 10, pp. 194-198, 2019.
- [4] V. Manchanda and N. Singh, "Flight delay prediction in Indian aviation industry using machine learning techniques," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 8, no. 11, pp. 127-132, 2018.
- [5] K. Zhang, J. Li, and Y. Wang, "Flight delay prediction based on Random Forest regression model," in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, Wuhan, China, 2018, pp. 793-798.
- [6] P. Singh and P. Singh, "Flight delay prediction using machine learning techniques in Indian aviation industry," *International Journal of Innovative Research in Computer Science and Engineering*, vol. 5, no. 2, pp. 180-186, 2017.

- [7] S. S. Mallick and S. K. Rath, "Flight delay prediction using machine learning in Indian aviation industry," in 2018 3rd International Conference on Inventive Systems and Control (ICISC), Coimbatore, India, 2018, pp. 233-238.
- [8] M. Kumar and V. Kumar, "Flight delay prediction using machine learning algorithms in Indian aviation industry," in 2019 6th International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, India, 2019, pp. 1258-1261.
- [9] Sri Hari Nallamala, et al., "A Literature Survey on Data Mining Approach to Effectively Handle Cancer Treatment", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 729 – 732, March 2018.
- [10] Sri Hari Nallamala, et.al., "An Appraisal on Recurrent Pattern Analysis Algorithm from the Net Monitor Records", (IJET) (UAE), ISSN: 2227 – 524X, Vol. 7, No 2.7, SI 7, Page No: 542 – 545, March 2018.
- [11] Sri Hari Nallamala, et.al, "Qualitative Metrics on Breast Cancer Diagnosis with Neuro Fuzzy Inference Systems", International Journal of Advanced Trends in Computer Science and Engineering, (IJATCSE), ISSN (ONLINE): 2278 – 3091, Vol. 8 No. 2, Page No: 259 – 264, March / April 2019.
- [12] Sri Hari Nallamala, et.al, "Breast Cancer Detection using Machine Learning Way", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8, Issue-2S3, Page No: 1402 – 1405, July 2019.
- [13] Sri Hari Nallamala, et.al, "Pedagogy and Reduction of K-nn Algorithm for Filtering Samples in the Breast Cancer Treatment", International Journal of Scientific and Technology Research, (IJSTR), ISSN: 2277-8616, Vol. 8, Issue 11, Page No: 2168 – 2173, November 2019.
- [14] Kolla Bhanu Prakash, Sri Hari Nallamala, et al., "Accurate Hand Gesture Recognition using CNN and RNN Approaches" International Journal of Advanced Trends in Computer Science and Engineering, 9(3), May – June 2020, 3216 – 3222.
- [15] Sri Hari Nallamala, et al., "A Review on 'Applications, Early Successes & Challenges of Big Data in Modern Healthcare Management'", Vol.83, May - June 2020 ISSN: 0193-4120 Page No. 11117 – 11121.
- [16] Nallamala, S.H., et al., "A Brief Analysis of Collaborative and Content Based Filtering Algorithms used in Recommender Systems", IOP Conference Series: Materials Science and Engineering, 2020, 981(2), 022008.
- [17] Nallamala, S.H., Mishra, P., Koneru, S.V., "Breast cancer detection using machine learning approaches", International Journal of Recent Technology and Engineering, 2019, 7(5), pp. 478-481.