International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Airline Flight Delay Prediction Based on Machine Learning and Aviation Big Data

Dr. M.A. Chaudhari[1], Rothe Sharad[2], Kasar Sanika[3], Pawar Shubham[4], Ballal Purva[5]

[1,2,3,4,5] **Department of Information Technology, Amrutvahini College of Engineering, Sangamner, India**

ABSTRACT-.

Flight delay is a challenging problem in the aviation industry, as it can cause significant costs for airlines, operators, and travelers. Machine learning methods have been applied in various studies to predict flight delay, but relying only on a single route or airport may not be sufficient to forecast the future of flights. Therefore, a broader scope of factors needs to be analyzed to identify potential influences on flight delays. One study used a flight delay dataset and applied supervised learning algorithms to predict flight departure delays. The model was evaluated to identify the best model and determine which features were more important in predicting delays. Factors that can potentially influence flight delays include bad weather conditions, seasonal and holiday demands, airline policies, technical issues such as problems in airport facilities, luggage handling, and mechanical apparatus, and the accumulation of delays from preceding flights. In one flight delay prediction system, weather parameters such as temperature, humidity, rain in mm, visibility, and month number were considered essential parameters for predicting delays.

Keywords- prediction, machine learning, dataset, decision tree, random forest

I. INTRODUCTION

The field of aviation has seen a significant increase in the volume of data generated in recent years due to advancements in sensor technology, data storage, and data processing capabilities. This has led to the emergence of the concept of aviation big data, which refers to the large volumes of data generated by the aviation industry that can be leveraged to gain insights and improve operational efficiency. One area where aviation big data can be particularly useful is in predicting flight delays. Flight delays significantly impact the airline industry, causing inconvenience to passengers, increasing operational costs, and reducing revenue. By accurately predicting flight delays, airlines can take proactive measures to mitigate their impact, such as rescheduling flights, reallocating resources, and communicating with passengers in advance. Machine learning is a powerful tool that can be used to analyze and make predictions based on large volumes of data. In-flight delay prediction, machine learning algorithms can be trained on historical data to identify patterns and factors contributing to flight delays. These algorithms can then be used to make predictions based on real-time data, such as weather conditions, airport congestion, and flight schedules. Some machine learning algorithms used for flight delay prediction include decision trees, random forests, support vector machines, and neural networks. These algorithms can be trained on various types of data, including flight schedules, weather data, and airport data, to predict flight delays with varying levels of accuracy. Overall, the use of aviation big data and machine learning for flight delay prediction holds great promise for improving operational efficiency in the aviation industry. As data collection and processing capabilities continue to improve, we can expect to see even more accurate and effective prediction models in the future.

II. PROBLEM STATEMENT

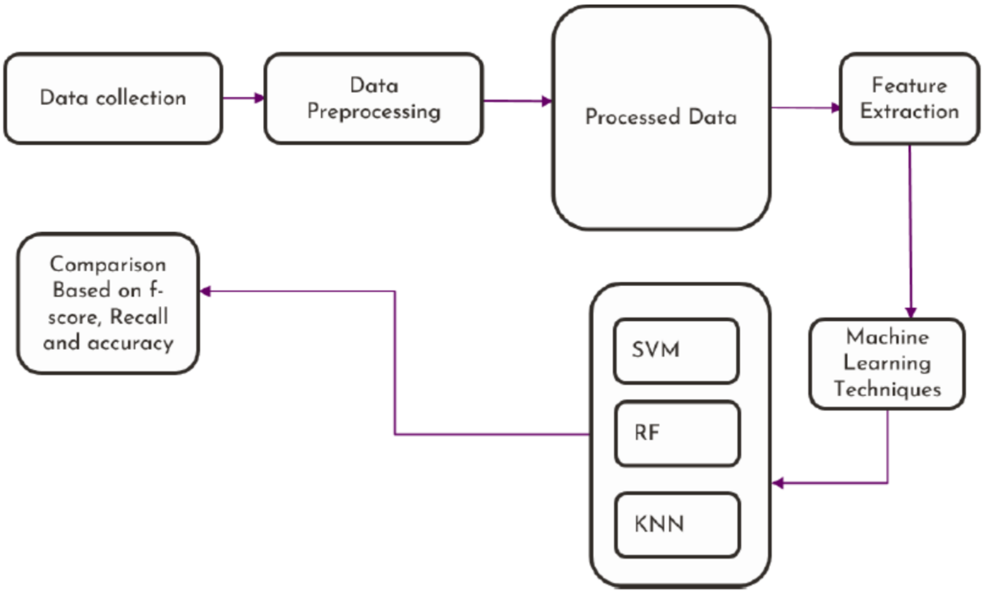
To design about airlines' increase in prices of customers and operational prices due to flight delay. Flight delay cost a lot to the airlines as well as passengers in financial and environmental terms. Flight delay causes to prices by costing a lot on operational purpose. System is based on the weather parameters which can result in delays.

III. MODULE DESCRIPTION

1. Data collection: Collecting flight data from various sources such as flight schedules, flight tracking systems, and weather forecasts. This data may also include information about the airline, the aircraft, the origin and destination airports, and any potential delays.
2. Data pre-processing: Cleaning and filtering the collected data to remove any missing or irrelevant data. This step may also involve transforming the data to make it suitable for use in machine learning algorithms.
3. Feature engineering: Identifying the most relevant features that can potentially influence flight delays, such as weather conditions, airline policies, and airport infrastructure.

4. Model selection: Choosing an appropriate machine learning algorithm to train on the collected and pre-processed data. Common algorithms used for flight delay prediction include decision trees, random forests, neural networks, and support vector machines.
5. Model training: Splitting the dataset into training and testing sets, and using the training set to train the chosen machine learning algorithm.
6. Model evaluation: Evaluate the trained model on the testing set to measure its accuracy in predicting flight delays. This step may also involve tuning the hyperparameters of the model to improve its performance.
7. Deployment: Integrating the trained model into a flight delay prediction system that can be used by airlines, airports, and other stakeholders in the aviation industry.

IV. SYSTEM ARCHITECTURE



V. CONCLUSION

The paper highlights the importance of developing a system to predict flight delay and presents various methodologies for the same. It emphasizes the need for accurate and real-time prediction of flight delays to reduce monetary losses for airlines and passengers, improve operational efficiency, and enhance the overall reputation of the aviation industry. The paper discusses the use of classification and regression techniques like Naïve Bayes, Feed-forward Networks, Random Forests, Decision Trees, Regression Tree, etc., and concludes that Naïve Bayes offers a scalable solution with high accuracy for real-time prediction. The paper also presents the implementation of Random Forest-based and LSTM-based architectures to predict individual flight delays and evaluates their performance. The Random Forest-based architecture is found to be more adaptable, but with lower training accuracy, while the LSTM-based architecture is effective in handling time sequences but suffers from overfitting problems. The paper suggests collecting or generating more training data, integrating more information, and designing more delicate networks to improve testing accuracy and overcome the overfitting problem. Overall, the paper offers valuable insights for decision-makers in the aviation industry to improve their operational efficiency, reduce financial losses, and address sustainability issues caused by flight delays. The analysis presented in the paper can be used as a prototype by aviation authorities to develop their own flight delay prediction systems.

REFERENCES

[1] Guan Gui and Fan Liu, "Flight Delay Prediction Based on Aviation Big Data and Machine Learning ", IEEE 2020.
[2] SP Lakshmi Narayanan1, S Sharbini2, R Priyanka3, R Kamalendran4, Mohit kumar5, C Murale, "Flight Delay Prediction Using Supervised Learning", IRJET 2021.
[3] Rahul Garg*1, Soham Gosavi*2, Tejas Choulwar*3, Onkar Vallal*4, Prof. Sarika Aundhakar*5, "Flight Delay Prediction", International Research Journal of Modernization in Engineering Technology and Science 2022.
[4] N Lakshmi Kalyani and Bindu Sri Sai U, "Machine Learning Model–based Prediction of Flight Delay", IEEE 2020.
[5] Jiage Huo and K.L. Keung , "The Prediction of Flight Delay: Big Data-driven Machine Learning Approach", IEEE 2020.
[6] Predicting Flight Delays Dieterich Lawson jdlawson@stanford.edu William Castillo will. castillo@stanford.edu.
[7] K. Tziridis, Th. Kalampokas, G.A. Papakostas, K.I. Diamantaras "Airfare Prices Prediction Using Machine Learning Techniques", IEEE 2020.
[8] Mrs. Yogita Borse*1, Dhruvin Jain*2, Shreyash Sharma*3, Viral Vora*4, Aakash Zaveri*5, "Flight Delay Prediction System", International Journal of Engineering Research Technology (IJERT) 2020.