Flight Delay Prediction Using Machine Learning

Dr. Satish N, Premkumar M, A.N.Karthikeyan, Senthilkumar V M, Sathyaseelan K, Kalaiyarasi V, E.Saranya

Abstract

Numerous businesses rely on various airlines to connect them with other parts of the world, and the aviation industry today plays a crucial role in the global transportation sector. However, flight delays caused by severe weather can have a direct impact on airline services. Correctly anticipating these flight delays enables airlines to anticipate the potential causes of the delays in advance to lessen their impact on passengers and passengers’ ability to prepare for the disruption. This project aims to investigate the methods used to build models that can predict flight delays caused by bad weather. The prediction of airline delays caused by various factors is the primary objective of this project. Commuters, the airline industry, and airport authorities all suffer as a result of flight delays. The Random Forest algorithm is used to predict the flight delay because it is more accurate than other methods.

Key Words: Flight delays, airline delays and Random Forest Algorithm.

I. INTRODUCTION

The air traffic control system, passenger airlines, and cargo airlines are today the most important parts of any transportation system. Countries all over the world have attempted to develop numerous methods for improving the airline transportation system over time. The airline’s operations have been drastically altered as a result. Modern passengers occasionally experience inconvenience as a result of flight delays [1]. About 20% of airline flights are cancelled or delayed annually, costing passengers money and time worth more than $20 billion. One of the most important means of modern versatility is the air transportation system. It is essential to maintain perseverance and resilience in the face of growing passenger and air traffic congestion. The infrastructure of airports is made possible by the availability of resources and land. Maintaining safety, efficiency, capacity, etc., are the norms for technology and procedure improvement. As a result, the National Airspace System (NAS) focuses on minimizing improvisation's impact on the environment. During their journey, passengers can visualize their flight path, altitude, heading, and other relevant parameters using the current technology. However, authorities in charge of air traffic control constantly attempt to reduce flight delays. Even though they worked together, the result is bad because the hours-long delays sometimes cause chaos. Weather, maintenance,
security, and the carrier are important factors that cause delays. The flight transportation system is anticipated to double by 2030 thanks primarily to corporate travel and tourism. Air traffic is also anticipated to rise in the same multiple as this increase. New airports can be built to ease the congestion caused by air traffic. However, the complexity continues to rise exponentially. Therefore, improvising with the existing airports is the only option for reducing the delay. The second option is more logical because of the limited land resources available. The term "delay" basically refers to the time period during which the flight is delayed or cancelled. If their mobility is delayed, it is likely to affect commercial aviation. Trusted customers are dissatisfied as a result of this delay, and sometimes even marketing strategies. The vast amount of data collected throughout a flight journey was stored in order to gain an understanding of the flight system.

1.1 Research motivation
Airport capacity is frequently referred to as average aircraft delay. The issue of flight delays is widespread in this world. It is extremely difficult to justify a delay. Although bad weather appears to be a frequent cause, runway construction and excessive traffic are a few of the rare causes of flight delays. A few flights are postponed due to the traditionalist deferrals, because of the late appearance of the past flight. Because airlines and airports rely on customer loyalty to support their frequent flyer programs, it hurts them and their marketing strategies.

1.2 Problem statement
My case study focused on four major airports in the United States: LaGuardia Airport in New York City, Logan Airport in Boston, San Francisco International Airport in San Francisco, and O'Hare Airport in Chicago. However, we restricted our concept and investigation to LaGuardia International Airport. Although our data was insufficient in comparison to that provided by all US airports, it provided us with valuable insight into the role that weather plays in flight delays. Using exploratory analysis and machine learning models, this project aims to predict airline departure and arrival delays.

II. SYSTEM STUDY
For tabular datasets, the proposed framework serves as a guide for using 2D CNNs. To begin, this framework proposes that feature values from tabular datasets should be mapped into various bins for the purpose of calculating evidence weights and then for the creation of images from the bins. The image dataset is divided into training and test sets following the creation of the images. Second, a 2D CNN must be trained on the training set and evaluated on the test set for its performance. Thirdly, cutting-edge techniques like Grad-CAM, SHAP values, Saliency Map, and LIME must be used to explain the 2D CNN predictions. Fourthly, the sanity check method must be used to verify explanations [32]. Lastly, in order to determine which explanation method is most effective, all methods must be quantitatively compared. The brain that learns everything is known as machine learning. Humans and machines alike learn. Personal experience imparts knowledge. With more data, it's easier to make predictions. For instance, when confronted with a circumstance that is unknown. The pace of progress is lower than in the real circumstance. The same principle applies to machine learning. In order to make a reliable prediction, the machine will take into account a specific instance. If we give it a comparable example, the machine can calculate the result. However, just like it is difficult for humans, machines struggle to make predictions when presented with new examples. Learning and reasoning are the primary objectives of machine learning. The process of machine learning begins with patterns. The data are credited to this discovery. The main thing for an information researcher to do is cautiously select the information they need to take care of the PC with. a list of problem-solving abilities. This is referred to as a vector of functions.

Over the past few decades, there has been an increase in the need for air travel. In 2018, there were an average of 44000 flights per day that carried 2.9 million passengers and 121 billion pounds of freight. The number of commercial flights operating annually decreased by approximately 6.7% between 2011 and 2017, while the number of delayed minutes increased by the same percentage. The ratio of total delay
minutes per year to the number of flights has increased, despite the fact that the number of flights has decreased since 2013. The Government Flight Office (FAA) conjecture predicts that flight activities will increment by almost 35% in the following 20 years, i.e., around 1.5% every year for every business carrier (Flying Organization, n.d.), which is likely to make the total number of delays even worse. The National Centre of Excellence for Aviation Operations Research (NEXTOR) Report, which was released in 2010, estimates that domestic airline delays cost the US economy approximately $32 billion in 2007. In 2017, the airline delay cost the US economy approximately $26.6 billion (Airlines For America, 2018). Between 2007 and 2017, the cost to airlines decreased from $82.2 per minute to $68.48 per minute, while the cost to passengers increased from $37.6 per hour to $49 per hour (Airlines For America, 2018; Ball and co., 2010). According to Bureau of Transportation Statistics (BTS), passenger numbers on commercial airlines are expected to rise from 840.8 million in 2017 to 1.28 billion in the next 20 years.

III. EXISTING SYSTEM
During the most pivotal time in human history, when computers moved from mainframes to personal computers, the cloud, and artificial intelligence. Machine learning, a fundamental subfield of artificial intelligence, has recently gained attention.

3.1 DISADVANTAGES:
- The SVM model takes a lot of time, is inaccurate,
- The worst part is cleaning and formatting the data.

3.2 PROPOSED SYSTEM
Therefore, the factors demonstrate how crucial and pertinent it has become to predict delays across all airline meshes. With the help of Regression Analysis and the regularization technique in Python 3, you can carry out predictive analysis, which entails a variety of statistical methods like supervised machine learning and data mining. Data mining analyses both current and historical data to make predictions or just look at delays in the future. Providing a comprehensive analysis of the performance of individual airlines and airports and then making an informed decision will be made easier with the assistance of this prediction. In addition, delay prediction analysis will assist in crucial decision-making processes that are required by each key player in the air transportation system in addition to the assessment of passengers.

3.4 ADVANTAGES
- The Random Forest approach outperforms the SVM model in terms of performance.
- Some way or another the SVM model is exceptionally tedious and doesn’t be guaranteed to create improved results.
- In the end, 91% of non-delayed flights are correctly predicted by our model.

IV. SYSTEM ARCHITECTURE
Information cleaning is the basic starting move toward assessing the dataset for definite examination. Databases are prone to having data that is noisy, missing, or inconsistent due to the enormous amount of data that is available. which Python uses require the data for. Data cleaning aids in the elimination of noise and inconsistent data. Four variables have been used to present the date format; To make it easier to use, it will be reduced to a single Python format. A missing value can be ignored, manually entered, given a constant or mean value during the data cleaning process, or both. In this instance, it will organize and arrange the entire data frame to keep the relevant attributes and get rid of those with
missing values. This is done to make it easier to use and more readable. From a variety of observations made by the origin airports and all of the airlines, it can be deduced that there is a great deal of variation in average delays, not only between different airlines but also between different airports. This is important because it means that in order to accurately model the delays, a model that is unique to the company and the home airport must be used. After the investigation of dataset, the last intend to accomplish is to devise models for expectation of postponements. The supervised machine learning algorithm that makes the prediction will be accurate.

V. CONCLUSION
The methodology for developing a system to anticipate flight delays is presented in this project. The paper provides specifics about the various methodologies that are utilized or can be utilized to determine the cause of flight delays. Flight delays are currently the topic of conversation due to the significant financial and environmental costs they impose on airlines and passengers. Flight delays drive up costs because they have a significant impact on operations. They may result in higher costs for airlines and their customers. Since the outcome is directly related to the passenger and the airlines, which in turn is related to another group of airlines and passengers, it is very important to have a real time delay for all of the players in the air transportation system. As a result, it is necessary to develop a method for anticipating flight delays in order to reduce financial losses and ensure efficient operation.

VI. FUTURE WORK
In order to obtain the most precise result, additional supporting research is required to correlate the entire problem, scope, and approach. Even though the most common cause of a flight delay is bad weather, major natural or man-made disasters and other unprecedented events can also delay flights significantly.

VIII. References