S

**Flight delay prediction for aviation industry using machine learning**

**FLIGHT DELAY PREDICTION FOR AVIATION INDUSTRY USING MACHINE LEARNING**

TEAM ID :NM2023TMID32693

TEAM SIZE :4

TEAM LEADER :ANITHA A

TEAM MEMBERS :1) NALINI R

2)PRIYADHARSHINI K

3)PUSHPA K

CATEGORY :MACHINE LEARNING WITH PYTHON

**INTRODUCTION:**

The primary goal of this paper is to predict the airline delays caused by various factors such as technical issue with aircraft, bad weather, air traffic control problems, bird strikes. Major problem arrises in the current aviation system due to flight delay. Some methods are required to introduce how delay propagates in the airport networks. Flight delays becomes an important subject and problem for air transportation system all over the world. The aviation industry is continuing to suffer from economic losses associated with the flight delays all the time. Many popular data driven methods have been used to predict flight delay, including the random forest algorithm, artificial neural network ,logic probability and deep learning. The critical aim of these methods is to extract appropriate influential factors from real system to build reliable and high- efficient prediction model.

Air ways is one of the crucial modes of transportation in our modern words, and with the increasing number of air vehicles it’s leading to simultaneous increase in the air traffic. So it’s important to maintain a flexible system. the Corporate travels and tourism are the two major contributors to flight transportation which is expected to be doubled by 2030, As a result the air traffic is also expected to increase in the same multiple .If we consider the US, where the airlines are handled by federal aviation administration ,they handle about 16,405,000 flights every year and handling the air traffic became a crucial part for safe movement. the air traffic authorities continuously try to disparage the delay in departure and arrival of the flights. Despite their best efforts the outcome is undesirable as sometimes the delays are hours causing chaos for the days schedule. Some of the important parameters that cause delay include weather, carrier, maintenance, security. These delays causes congestion in the air traffic. One of the solution is to minimize the air traffic congestion is to construct new airports, but the complexity increases .we could improvise the existing airports but considering the limited availability of land resources, the ultimate logical solution would be predicting the delay of the flights. Delay basically represents the period by which the aircraft is late or has been cancelled. The delay results in complexity in air traffic and dissatisfaction of customers and increase in costs for the company .If a flight is delayed by 10 minutes the flight is considered delay. In the USA from march 2021 to march 2022 itself the delay in flights is about 20.29 % of which air carrier delay is 7.04%, delay due weather is 0.72% , delay due to navigation system is 4.26 and delay due to aircraft arriving late and security delay is 6.12% .we cannot exactly predict the reason for the delay but after the arrival we can predict the delay time for reaching the destination

**PURPOSE :**

This project purpose is therefore,predicting flight delays can improve airline operations and passenger satisfaction,which will result in a positive impact on the economy.Impact on the economy. In this study, the main goal is to compare the performance of machine learning classification algorithms when predicting flight delays.

**METHODOLOGY :**

In this project the machine learning flow can be explained in three stages

1.Gathering of data

2.Data Pre-processing

3.Training & Testing of Model

**1.GATHERING OF DATA :**

The three databases are selected for the best combination based features as per the previous experience of the data analysed. The BTS provides the database on the real time performance data. To explain the arrival delay of the flights, causes of the delay are reported in five categories: air carrier, extreme weather, national aviation system, late arriving aircraft and security. In our model, only the hourly departures and hourly arrivals are applied in corresponding airports.

**2.DATA PRE-PROCESSING :**

Instead of the precise delay time, we use the delay group in our research for random forest classification. The application of the delay group is suitable for the uncertainty of the actual air traffic control. All the non-positive delay groups are classified as on-time flights. The maximum of the delay group is 11 (delay time greater than 170 minutes) per day . The Following details shown in Table 1. Reference table of delay -related data are extracted from the BTS groups Delay groups Delayed time (min)-2 (-inf, -15) -1 [-15,0) 0 [0,15) 1 [15,30) 2 [30,45) 3 [45,60) 4 [60,75) 5 [75,90) 6 [90,105) 7 [105,120) 8 [120,135) 9 [135,150) 10 [150,165) 11 [165,180) 12 [180,+inf) The following delay database. • Day of month • Day of week • Scheduled departure time • Scheduled arrival time • Scheduled elapsed time .

**3.TRAINING & TESTING MODEL :**

• DATE\_MONTH\_YEAR: date, month,

• FLYING\_AIRLINE : flight selection by user

• FLYING\_FROM & FLYING\_TO: departure destination & arrival destination.

**DATA SOURCE :**

We have used the following data sets:

1. Statistical Computing & Statistical Graphics Administration (RITA)

database and structured for our use.

2. For weather data we have: From NOAA the weather observation data has been fetched at hourly land based weather format. Daily & hourly data of various airports is included in the source.

**SOFTWARE PACKAGES USED :**

**Python Packages**

1. Pandas : It is an open source python package that is most widely used for data analysis.

2. Numpy & Scipy : These are multidimensional arrays and general mathematical algorithms, etc.

3. Sci-kit Learn : It is a free machine learning library for python. It has features like K scientific libraries like Numpy & Scipy (which we have used here).

4. Matplotlib : It is a cross - platform , data visualization and graphical plotting library for python an numerical extension NumPy

5. CSV : “Comma-Separated Values” is a simple file format used to store tabular data like spreadsheets or databases. It is generally imported and exported from programs.

**FUTURE SCOPE :**

Currently the dataset is limited to only flight and weather data of USA. However, In Future, we can include datasets from Other International Countries Expanding the scope of this project we can also add the flight data for the domestic flights. In addition, we can implement weather data to get real time and results that are more accurate.

**CONCLUSION** :

Machine learning and deep learning algorithms can be used to predict flight delays. The purpose of doing the above classification and analysis is to gauge the delay to not only suffice the various purposes of humankind, but also analyze factors affecting delay. The Proposed System is build using Support Vector Machine, Random Forest and KNN. The overall precision, recall and accuracy is calculated by using above techniques As seen from the articles and papers These methodologies offer virtually identical accuracy however we want an algorithmic rule that is good with real world prediction and analysis and thus: naïve- Bayes. Except being clever with real time prediction algorithmic rule that considers or assumes independence among predictions that makes the system scalable. As other independent attribute may be superimposed up to the algorithmic rule for computation of the delay, the anticipated postpone can thus facilitate the floor personnel for implementing accurate and easy operation plans and consequently the facts, if sent to the passengers will benefits the airways and also the passengers.