RESEARCH ARTICLE OPEN ACCESS

Applying Machine Learning Algorithm for Predicting Flight Delays

Dr Subba Reddy Borra[1], Swetcha Pandla[2], Mahalaxmi Raccha[3], Alekhya Sooram [4]

1(Professor and Head, Information Technology, Malla Reddy Engineering College for Women, Hyderabad-500100
Email: bvsr79@gmail.com)
2, 3, 4 (Student, Information Technology, Malla Reddy Engineering College for Women, Hyderabad-500100.
Email: mahalaxmiracha19@gmail.com)

Abstract:

Flight delays are a serious problem in the aviation sector. The expansion of the aviation industry during the past two decades has increased air traffic, which has delayed flights. Not only do flight delays cost money, but they also have a bad effect on the environment. Flight delays cause significant losses for airlines that run commercial flights. In order to prevent or avoid flight delays and cancellations, they thus take all reasonable precautions. In this study, we use machine learning models such logistic regression, decision tree regression, Bayesian ridge, random forest regression, and gradient boosting regression to predict whether a given flight's arrival would be delayed or not.

Keywords —Flight Delay, Machine Learning, Logistic Regression, Bayesian Ridge, Gradient Boosting Regression

I. INTRODUCTION

Flight delays have been extensively studied in recent years. The rising demand for air travel has led to a rise in flight delays. According to BTS [2], there were 860,646 arrival delays in 2016. The Federal Aviation Administration (FAA) estimates that aircraft delays cost the aviation industry more than $3 billion yearly [1].

Commercial scheduled flights frequently have delays due to a variety of factors, such as air traffic congestion, an annual increase in passengers, maintenance and safety concerns, inclement weather, and the delayed arrival of the aircraft that will be utilised for the upcoming voyage [3] [4]. The FAA in the US determines that an aircraft is delayed if there is a difference of more than 15 minutes

between the scheduled and actual arrival times. because it grows.

II. LITERATURE REVIEW

Flight delays have been the subject of extensive research. Air traffic control, airline decision-making, and ground delay response programmers have all experienced significant difficulties with the forecasting, analysis, and causation of aircraft delays. On the sequence's delay propagation, research is being done. Research into the forecast model for arrival delay and departure delay utilising meteorological factors is also encouraged. In the past, researchers have experimented with using machine learning to forecast aircraft delays. In order to anticipate delays in the arrival of operating flights, including the five busiest US airports, Chakrabarty et al. [5] employed supervised automatic learning methods (random forest, Gradient Boosting Classifier, Support Vector Machine, and the k-nearest neighbor algorithm). The maximum precision with gradient booster as a classifier with a restricted training set was 79.7%.

III. EXISTING SYSTEM

Automated learning algorithms under supervision. Support Vector Machine and the k-nearest neighbour algorithm were used to estimate delays in the arrival of operating planes, including the five busiest US airports. When using gradient booster as a classifier with a small amount of data, the precision was very low. k-Nearest Neighbours machine learning methods were used to anticipate aircraft delays. The model has been updated with information from weather forecasts and flight schedules. Sampling techniques were employed to balance the data, and it was found that the trained classifier with sampling techniques had greater accuracy than the classifier trained without sampling.

DISADVANTAGES

- The non-parametric nature of the response under examination data does not presume a specific functional form.
- Other variables, such as the number of origin-destination pairs and the forecast horizon, may also affect predictability.
- The forecasts were based on a few crucial characteristics.
- K-nearest Neighbor, Support Vector Machine, and Multiple Linear Regression are the algorithms used.

IV. PROPOSED SYSTEM

We used data gathered by the Bureau of Transportation; U.S. Statistics of all domestic flights taken in 2015; to anticipate flight delays and train models. The US Bureau of Transport Statistics gives statistics on arrival and departure, including wheels-off time, departure delay, and taxi-out time per airport. It also includes actual departure time, scheduled departure time, and scheduled elapsed time. The airport and the airline both offer

cancellation and rerouting information, along with the date, time, flight labelling, and airline airborne time. The data collection has 31 columns and 20277 records, and thanks to our technology, it is expandable. We may fill in the missing numbers by utilising the pandas package, which is crucial for processing data for models.

ADVANTAGES

1. A guided learning method to discover the benefits of having a schedule and an actual arrival time
2. Algorithms have a low computational cost.
3. Based on a set of factors, we design a system that forecasts airline departure delays.Algorithms Used: Logistic Regression, Decision Tree, Random Forest, Gradient Boosting.

MODULES

- User
- Admin
- Data preprocess
- Model Execution

DESCRIPTION OF MODULES:

User:

The user can sign up initially. He needed a working user email and mobile upon registration for future communications. Once the user register then admin can activate the customer. Once admin activated the User then user can login into our system. The dataset collected from US Bureau of Transport is not directly processed. Before process we need to clean the data. Once clean the data then user can test the departure delay performance based on selected models. The user can see the results in the browser. The all error scores displayed and graphical representation can be displayed.

Admin:

With his login information, Admin can log in. He can activate the users after logging in. Only our applications allow the activated user to log in. We have researched a variety of sources to determine which variables will be most useful in predicting departure and arrival delays. We get to the conclusion that the dataset's parameters are Day, Departure Delay, Airline, Flight Number, Destination Airport, Origin Airport, Day of Week, and Taxi out after conducting multiple searches. We therefore take this data into account for the next step.

Data Pre-process:

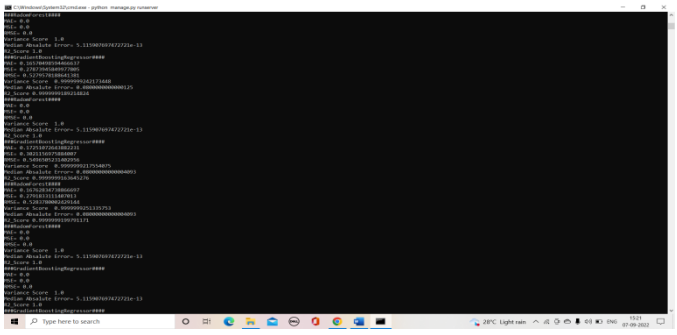
The admin provided data has been stored in the sqlite database. To process our methodology we need to perform data cleaning process. We can fill in the missing numbers with the mean type using a pandas

data frame. The data will be displayed on the browser once it has been cleared.

Model Execution:

With the help of machine learning models like gradient boosting regression, logistic regression, decision tree regression, bayesian ridge, random forest regression, and Bayesian ridge, we can predict the outcome. The MSE is suitable for our regression issues since it is differentiable, which adds to the algorithmic stability. Additionally, it severely penalises larger errors relative to smaller errors. An indicator of risk, MAE provides the predicted value of the absolute error loss. This method measures the Explained Variance Score, or the degree to which our machine learning model explains the scattering of the dataset. R2 Rating This statistic assesses the likelihood that the model will be able to predict unknown samples through the proportion of explained variance. It indicates the quality of fit. The best possible result is 1.0 and score can also be negative.

V. RESULTS



VI. CONCLUSION

We developed a machine learning model through this project that can forecast flight departure delays. Logistic Regression was the most effective model. A total of 62% of the departure delays were predicted by the model. Additionally, it was shown that the departure airports had a significant impact on flight delays. This obliquely implies that the likelihood of flight delays will be higher at crowded, big airports than at lesser airports. By conducting this analysis, it will be possible to ensure that the schedules are properly handled and that the airport's operations are enhanced to prevent such delays. I think that travelling by plane is the quickest option, and I think that speed is really important.

VII.FUTURE WORK

Future applications of this work may include the use of more advanced, modern, and cutting-edge preprocessing techniques, automated hybrid learning and sampling methods, and deep learning models adjusted to obtain better performance. To assist in the development of a prediction model, additional factors might be added. For instance, one model develops error-free models for flight delays using meteorological statistics. The model can now be trained using data from other countries because we only used US data in this study. When given the requisite processing power, complicated models and hybrids of many distinct models, as well as larger, more detailed information, can be used to create more accurate prediction models. The model can further be modified.

REFERENCES:

1. N. G. Rupp, "Further Investigation into the Causes of Flight Delays," in Department ofEconomics, East Carolina University, 2007.

2."Bureau of Transportation Statistics (BTS) Databases and Statistics,"[Online]. Available:http://www.transtats.bts.gov.

3. "Airports Council International, World Airport Traffic Report," 2015,2016.

4. E. Cinar, F. Aybek, A. Caycar, C. Cetek, "Capacity and delay analysisfor airport manoeuvring areas using simulation," Aircraft Engineering andAerospace Technology, vol.86, no. No. 1,pp. 43-55, 2013.

5.Navoneel, et al., Chakrabarty, "Flight Arrival Delay Prediction Using Gradient Boosting Classifier," in Emerging Technologies in Data Mining

and Information Security, Singapore,2019.

6. Y. J. Kim, S. Briceno, D. Mavris, Sun Choi, "Prediction ofweatherinduced airline delaysbased on machine learning algorithms," in 35th Digital Avionics Systems Conference (DASC), 2016.

7. W.-d. Cao. a. X.-y. Lin, "Flight turnaround time analysis and delayprediction based onBayesian Network," Computer Engineering and Design,vol. 5, pp. 1770-1772, 2011.

8. J.J. Robollo, Hamsa, Balakrishnan, "Characterization and Prediction of AirTraffic Delays".

9. S. Sharma, H. Sangoi, R. Raut, V. C. Kotak, S. Oza, "Flight DelayPrediction System Using Weighted Multiple Linear Regression," International Journal of Engineering and Computer Science, vol. 4, no. 4,pp. 11668 - 11677, April 2015.

10.A. M. Kalliguddi, Area K., Leboulluec, "Predictive Modelling ofAircraft Flight Delay,"Universal Journal of Management, pp. 485 - 491,2014