

## FLIGHT DELAY PREDICTION USING MACHINE LEARNING ALGORITHMS

**P. Rajeshwar Reddy<sup>\*1</sup>, R. Shivani<sup>\*2</sup>, G. Suryateja<sup>\*3</sup>,  
K. Maheswari<sup>\*4</sup>**

<sup>\*1,2,3,4</sup>Affiliated To JNTUH, Department Of Computer Science Engineering, CMR  
Technical Campus, Hyderabad, Telangana, India.

### ABSTRACT

In today's world, larger number of people choose to travel by air because flight travel takes very less time compared to other modes of transport and helps the passengers to save their quality time. There are also few challenges involved in aviation department, one such challenge is flight delay which costs a lot to airlines, operators, and passengers. According to the statistics of the Bureau of Transportation Statistics (BTS), twenty percent of the total flights are delayed. This flight delays can occur due to multiple factors such as weather conditions, airport congestion, technical issues etc. To maintain efficient air travel, a system to be built which predicts flight delays based on machine learning model. In this model we are using different machine learning algorithms used like Random Forest, Naive Bayes, KNeighbors Classifiers and Decision Tree. The goal of the system is to predict the efficient flight delay by using those machine learning. Hence it can help to predict the flight delays and reduce the flight delays.

**Keywords:** Flight Delay Prediction, Random Forest, Decision Tree Algorithm, Kneighbors Classifier, Naïve Bayes, Accuracy, Supervised Learning.

### I. INTRODUCTION

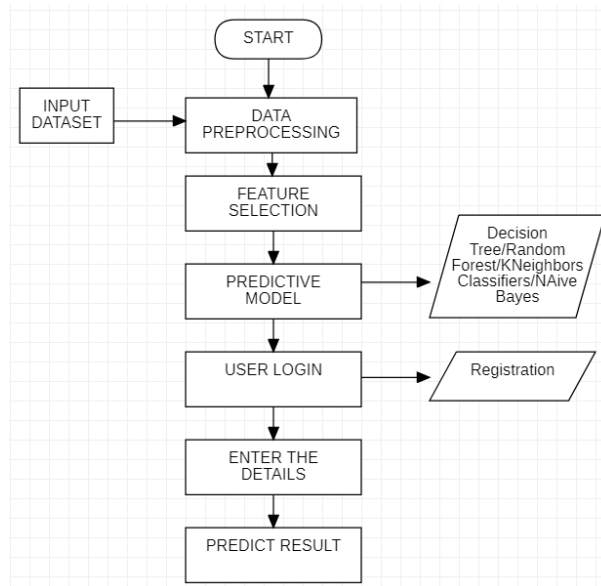
The introduction Air transportation is one of the important modes of transport in the recent times and having a great demand for air transport. But due to increasing air traffic and passenger-traffic, it is very important to maintain resilience. According to few statistics, the delay of a flight was 13.2 minutes in Europe and 13 minutes in the US. And according to the Bureau of Transportation Statistics (BTS), twenty percent of the total flights are delayed and according to Federal Aviation Administration (FAA), more than about \$3 billion wealth lost due to flight delays. The Airline Authorities constantly working to minimize the delay of flights but still the results are not in favour to them. There might be many reasons for the delay of flights. The reasons might be weather conditions, airport congestion, airplane issues etc. These delay of flights results in dissatisfaction to the passengers and lead financial loss to airlines and customers.

Hence to have efficient air travel, a system is required to predict the delay of flights. Here we develop a machine learning model is used to predict flight delays. Here Supervised learning model is used to predict the flight delays. The models include Naïve Bayes, Random Forest, KNeighbors and Decision Tree. Here the flight delay is predicted by considering various factors weather conditions flight scheduling time etc.

### II. METHODOLOGY

In the proposed methodology, we predict the delay of flights which cause a great inconvenience to passengers and cost billion dollars to airlines. Passengers also may face inconvenience and stress and lead to additional expenses for food and lodging. The delays are caused by various delay factors such as air carrier, extreme weather security and technical issues. Here we use the supervised machine learning models to predict the flight delays. We use several machine learning models like Random Forest, KNeighbors Classifiers, Decision Tree, Naïve Bayes. These models helps us to predict efficient delay in flights. Here we can see how each machine learning model performs and can also estimate the accuracy and precision of each model. Here we can compare these models and select the delay value of the most efficient model.

### III. MODELING AND ANALYSIS



**Figure 1: System Architecture**

The above figure shows how the machine learning model was built. Deploying a machine learning model includes the following steps namely data collection, input Dataset, data preprocessing ,feature selection, predictive models such as Random Forest Model, Naïve Bayes, Decision Tree and KNeighbors classifier

**Input Dataset:** Input dataset is given to the system to find the desired predicted output.

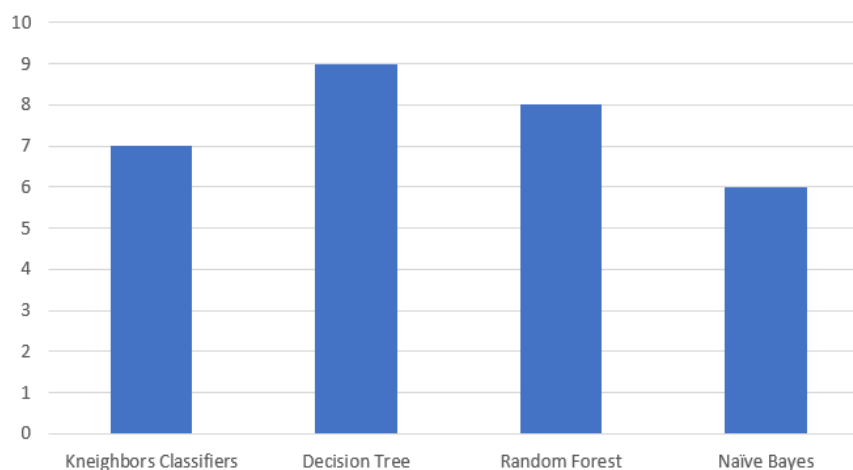
**Data preprocessing:** Data preprocessing is the process in which the collected data is processed and cleaning the data and suitable for machine learning model so as to make execution efficiently.

**Feature Selection:** Feature Selection is the process where we select the suitable features required for the model and avoiding unwanted data.

**Predictive Model** is that to select the efficient model required and it will predict the result by analysing the existing data.

Here after the user registration and then by logging into the system and we have to enter the required essentials. After entering the details in the required fields we can see the result of the flight delay.

### IV. RESULTS AND DISCUSSION



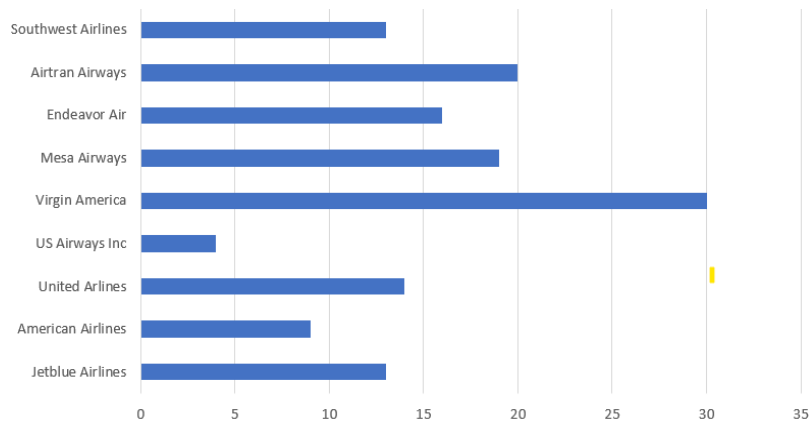
**Figure 2: Comparison of Algorithm Accuracy**

The above graph depicts the accuracy of the machine learning models. Here in the graph we can see the Decision Tree having more accuracy compared to other models while Naïve Bayes is having least among them. Random Forest and KNeighbors Classifiers having accuracy in between Decision Tree and Naïve Bayes.

**Table 1.** Accuracy and Precision of algorithms

Model	Accuracy	precision
Decision Tree	0.9687	0.9543
Random Forest	0.9345	0.9141
KNeighbors Classifiers	0.8912	0.8112
Naïve Bayes	0.8586	0.8213

The above table shows the accuracy and precision of different Machine Learning models. Here Decision Tree model having higher accuracy and Precision compared to the remaining models. Random Forest model having highest accuracy and precision next to Decision Tree. KNeighbors Classifiers having higher accuracy and precision after Decision Tree and Random Forest and having less precision than Naïve Bayes. Here Naïve Bayes is having least precision and accuracy among the models but having precision more than KNeighbors Classifier.


**Figure 3:** Mean delay of airlines

**Table 2.** mean delay of airlines

Flight(from)	Flight(to)	Airways	Delay Prediction (in mins)
Hartsfield Jackson Atlanta International Airport	Orlando International Airport	Southwest Airlines	4
Nashville International Airport	Tampa International Airport	Southwest Airlines	8
Denver International Airport	Hartsfield International Airport	Delta Airlines	7
Chicago O'Hare International Airport	Philadelphia International Airport	Virgin America	6
Orlando International Airport	Phoenix Sky Harbor International Airport	American Airlines	0
Chicago O'Hare International Airport	Cleveland Hopkins International Airport	Virgin America	9
Will Rogers World Airport	William P. Hobby Airport	Southwest Airlines	14
Hartsfield Jackson Atlanta International Airport	Minneapolis-Saint Paul International Airport	Delta Airlines	15

The above graph depicts that the mean delay of the airlines. Here we can find the different airlines and the comparison of the flight delays among them. In the above graph only few airlines are considered. The US Airways has the least mean delay where Virgin America has the highest. The above table shows the result of flight delay. Here are the few airports taken where it shows the journey of the flights from one airport to the

other. Here the route of flight from one airport to other is given. We predict the delay of the flights by using the machine learning models. The models as Naïve bayes, Decision Tree, Random Forest, KNeighbors Classifiers helps us to predict the flight delays. Here the best possible and most efficient delay value is considered. In the above table we can find the flight delay values.

## V. CONCLUSION

In this system it helps us to predict the delay of flights and here we are using the different machine learning algorithms and it gives the range of different methodology that is used to find out the delay in flights. This Flight Delay may increase prices and cost financially a lot to passengers and airways. Hence, it is important to have the machine learning models very good precision and accuracy in predicting the flight delays. The machine learning models used in this system to predict flight delay has good precision and accuracy. Here we used the supervised learning model to predict the flight delay. The models include Random Forest, Naïve Bayes, Decision Tree and KNeighbors. Here the results shows that Decision Tree and Random Forest models has greater precision and accuracy than KNeighbors Classifiers and Naïve Bayes. Hence we predicted the flight delay of efficiently.

## ACKNOWLEDGEMENTS

We thank CMR Technical Campus for supporting this paper titled with "FLIGHT DELAY PREDICTION USING MACHINE LEARNING ALGORITHMS", which provided good facilities and support to accomplish our work. Sincerely thank to our chairman, Director, Deans, Guide and faculty members for giving valuable suggestions and guidance in every aspect of our work.

## VI. REFERENCES

- [1] Yanxiang Zhu, Nilesh Padwal, Mingxuan Li, "Data Analysis of U.S. Airlines Ontime Performance", vol. 5, pp. 55-60
- [2] J. J. Robollo, Hamsa, Balakrishnan, "Characterization and Prediction of Air Traffic Delays".
- [3] Tu Y, Ball MO, Jank WS. Estimating flight departure delay distributions—a statistical approach with longterm trend and short-term pattern. J Am Stat Assoc. 2008;103(481):112–25
- [4] B. Sridhar, N. Chen, "Short term national airspace system delay prediction", Journal of Guidance, Control, and Dynamics, Vol. 32 No. 2, 2009
- [5] Reynolds-Feighan AJ, Button KJ. An assessment of the capacity and congestion levels at European airports. J Air Transp Manag. 1999;5(3):113–34.
- [6] Lawson, D. and Castillo, W. Predicting flight delays. Technical report, Computer Science Department, CS 229, Stanford University, Stanford, CA, 2012.
- [7] Noriko, Etani, "Development of a predictive model for on-time arrival flight of airliner by discovering correlation between flight and weather data," 2019.
- [8] M. M. Najafabadi, F. Villanustre, T. M. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, "Deep learning applications and challenges in big data analytics," J. Big Data, vol. 2, no. 1, pp. 1–21, 2015.
- [9] Evans JE, Allan S, Robinson M. Quantifying delay reduction benefits for aviation convective weather decision support systems. In Proceedings of the 11th Conference on Aviation, Range, and Aerospace Meteorology, Hyannis. 2004.
- [10] E. Cinar, F. Aybek, A. Caycar, C. Cetek, "Capacity and delay analysis for airport manoeuvring areas using simulation," Aircraft Engineering and Aerospace Technology, vol. 86, no. No. 1, pp. 43-55, 2013
- [11] On-Time Flight Departure Prediction System Using Naive Bayes Classification Method (Case Study: XYZ Airline) IJCTT december 2017
- [12] Oza S, et al. Flight delay prediction system using weighted multiple linear regression. Int J Eng Comp Sci. 2015;4(05):11765.
- [13] N, Prabakaran & Kannadasan, Rajendran. (2018). Airline Delay Predictions using Supervised Machine Learning. International Journal of Pure and Applied Mathematics. 119.
- [14] Navoneel, et al., Chakrabarty, "Flight Arrival Delay Prediction Using Gradient Boosting Classifier," in Emerging Technologies in Data Mining and Information Security, Singapore, 2019.