Volume-09 Issue 03, March-2025 ISSN: 2456-9348
Impact Factor: 8.232

International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

PREDICTING FLIGHT DELAYS WITH ERROR CALCULATION USING MACHINE LEARNING CLASSIFIERS

K. VARUNIKA
UG Student, Department of Artificial Intelligence and Data Science, J.B. Institute of Engineering and Technology, Hyderabad.

M. RASHMI GUPTA
L. HEMANTH
K. SHIVA KUMAR
UG Students, Department of Artificial Intelligence and Data Science, J.B. Institute of Engineering and Technology, Hyderabad

ABSTRACT
Flight delay is a major problem in the aviation sector. During the last two decades, the growth of the aviation sector has caused air traffic congestion, which has caused flight delays. Flight delays result not only in the loss of fortune also negatively impact the environment. Flight delays also cause significant losses for airlines operating commercial flights. Therefore, they do everything possible in the prevention or avoidance of delays and cancellations of flights by taking some measures. In this paper, using machine learning models such as Logistic Regression, Decision Tree Regression, Bayesian Ridge, Random Forest Regression and Gradient Boosting Regression we predict whether the arrival of a particular flight will be delayed or not.

Keywords:
Flight Prediction, Machine Learning, Error Calculation, Logistic Regression, Decision Tree, Bayesian Ridge, Random Forest, Gradient Boosting, Logistic Regression, U.S. Flight data.

INTRODUCTION

Flight delay is studied vigorously in various research in recent years. The growing demand for air travel has led to an increase in flight delays. According to the Federal Aviation Administration (FAA), the aviation industry loses more than $3 billion in a year due to flight delays [1] and, as per BTS [2], in 2016 there were 860,646 arrival delays. The reasons for the delay of commercial scheduled flights are air traffic congestion, passengers increasing per year, maintenance and safety problems, adverse weather conditions, the late arrival of plane to be used for next flight [3] [4]. In the United States, the FAA believes that a flight is delayed when the scheduled and actual arrival times differs by more than 15 minutes. Since it becomes a serious problem in the United States, analysis and prediction of flight delays are being studied to reduce large costs.

OBJECTIVES

The primary objectives of predicting flight delays using machine learning classifiers with error calculation are:

1. Accurate Delay Prediction: Create machine learning models to forecast flight delays from past data, weather, airport traffic, and other appropriate factors. Enhance airline and passenger planning by minimizing uncertainties in departure and arrival times.

2. Feature Analysis & Selection: Determine the major factors that cause flight delays, including weather, air traffic, airline operations, and airport congestion. Optimize feature selection for enhancing model

Volume-09 Issue 03, March-2025 ISSN: 2456-9348
Impact Factor: 8.232

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

performance and interpretability.
3. Comparison of Machine Learning Classifiers: Compare various machine learning classifiers (e.g., Decision Trees, Random Forest, Support Vector Machines, Neural Networks) for prediction of delays. Compare the performance measures like accuracy, precision, recall, and F1-score.
4. Error Calculation & Model Evaluation: Apply error calculation measures like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and classification-based errors. Compare model strength and dependability under practical applications.
5. Optimization of Model Performance: Hyperparameters should be tuned to improve prediction accuracy and minimize errors. Ensemble methods or deep learning can be used to improve model predictions.
6. Real-World Application & Deployment: Predictive models should be integrated into airline systems or applications for proactive delay management.Insights should be provided for airline operators to optimize scheduling and minimize operational costs.

METHODOLOGY

Predicting Flight Delays with error calculation using Machine Learning classifiers involves:

1. DATASET
2. DATA PREPROCESSING
3. FEATURE EXTRACTION

YEAR	MONTH	DAY	DAY_OF_WEEK	AIRLINE	FLIGHT_NUMBER	TAIL_NUMBER
2015	1	1	4	B6	2023	N324JB
2015	1	1	4	AA	2299	N3LLAA
2015	1	1	4	B6	939	N794JB
2015	1	1	4	AA	1205	N3FKAA
2015	1	1	4	UA	319	N498UA
2015	1	1	4	AA	1103	N3HCAA
2015	1	1	4	AA	1297	N3JYAA
2015	1	1	4	B6	353	N570JB
2015	1	1	4	B6	371	N708JB
2015	1	1	4	B6	583	N531JB
2015	1	1	4	B6	605	N766JB
2015	1	1	4	B6	525	N645JB
2015	1	1	4	DL	421	N967DL

ORIGIN_AIRPORT	DESTINATION_AIRF	SCHEDULED_DEPA	DEPARTURE_TIME	DEPARTURE_DELA	TAXI_OUT	WHEELS_OF
JFK	SJU	535	618	43	13	
JFK	MIA	545	640	55	17	
JFK	BQN	545	545	0	17	
EWR	MIA	559	552	-7	22	
EWR	MCO	600	603	3	14	
LGA	DFW	600				
LGA	MIA	600	708	68	17	
JFK	PBI	600	554	-6	16	
LGA	FLL	600	600	0	22	
JFK	MCO	600	557	-3	16	
EWR	FLL	600	556	-4	12	
JFK	TPA	600	554	-6	21	
JFK	ATL	600	605	5	18	

Snapshot of Dataset

RESULTS AND DISCUSSION

After preprocessing and feature extraction of our dataset, 60% of the dataset was selected for training and 40% of the dataset was selected for testing. For error calculation, we are using scikit-learn metrics [14]. Results are divided between two sections, Departure Delay(A) and Arrival Delay(B).

A. Departure Delay:

Table 1 lists our results for departure delay which compares different Machine Learning models, i.e. Logistic Regression, Decision Tree Regressor, Bayesian Ridge, Random Forest Regressor and Gradient Boosting Regressor, based on various evaluation metrics. Further, we compare each model concerning one evaluation metric at a time and show it as a bar graph.

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

Model	Mean Squared Error	Mean Absolute Error	Explained Variance Score	Median Absolute Error	R2_Score
Logistic Regression	3388.7	26.5	0	7	-0.2
Decision Tree Regressor	3204.7	24.8	-0.1	7	-0.1
Bayesian Ridge	3686.9	37.7	-0.3	24.3	-0.3
Random Forest Regressor	2261.8	24.1	0.2	14.8	0.2
Gradient Boosting Regressor	2317.9	24.7	0.2	13.8	0.2

TABLE I. Departure Delay Evaluation Metrics for various mode

B. Arrival Delay:

Table 2 lists our results for arrival delay which compares different Machine Learning models, i.e. Logistic Regression, Decision Tree Regressor, Bayesian Ridge, Random Forest Regressor and Gradient Boosting Regressor, based on various evaluation metrics. Further, we compare each model concerning one evaluation metric at a time and show it as a bar graph.

Model	Mean Squared Error	Mean Absolute Error	Explained Variance Score	Median Absolute Error	R2_Score
Logistic Regression	4290.2	36.6	-0.1	20	-0.2
Decision Tree Regressor	4501.0	36.4	-0.3	19	-0.3
Bayesian Ridge	4908.8	47.2	-0.4	33	-0.4
Random Forest Regressor	3019.3	30.8	0.2	18.8	0.1
Gradient Boosting Regressor	3132.7	31	0.1	18.2	0.1

TABLE II. Arrival Delay Evaluation Metrics for various

ACKNOWLEDGEMENT

We acknowledge with deep appreciation all the efforts made by everyone in successfully carrying out this project on Predicting Flight Delays with Error Calculation using Machine Learning Classifiers. Above all, we wish to thank our faculty members and institution for the insightful guidance, inspiration, and unfailing support received from them throughout this study. Their input and experience have helped mold the focus of our project. Finally, we recognize the open- source research, datasets, and existing studies that were made available to us as the groundwork for our study. Without them, it would have been impossible to implement our model. Our project has been a worthwhile learning experience, and we appreciate being able to make a contribution towards ML-based Flight Delay Prediction.

CONCLUSION

Machine learning algorithms were applied progressively and successively to predict flight arrival & delay. We built five models out of this. We saw for each evaluation metric considered the values of the models and

iJETRM
International Journal of Engineering Technology Research & Management
Published By:
https://www.ijetrm.com/

compared them. We found out that: - In Departure Delay, Random Forest Regressor was observed as the best model with Mean Squared Error 2261.8 and Mean Absolute Error 24.1, which are the minimum value found in these respective metrics. In Arrival Delay, Random Forest Regressor was the best model observed with Mean Squared Error 3019.3 and Mean Absolute Error 30.8, which are the minimum value found in these respective metrics. In the rest of the metrics, the value of the error of Random Forest Regressor although is not minimum but still gives a low value comparatively. In maximum metrics, we found out that Random Forest Regressor gives us the best value and thus should be the model selected.

REFERENCES

1. N. G. Rupp, "Further Investigation into the Causes of Flight Delays," in Department of Economics, 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 analysis for airport manoeuvring areas using simulation," Aircraft Engineering and Aerospace 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 of weather induced airline delays based 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 delay prediction based on Bayesian Network," Computer Engineering and Design, vol. 5, pp. 1770-1772, 2011.
8. J. J. Robollo, Hamsa, Balakrishnan, "Characterization and Prediction of Air Traffic Delays".
9. S. Sharma, H. Sangoi, R. Raut, V. C. Kotak, S. Oza, "Flight Delay Prediction 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 of Aircraft Flight Delay," Universal Journal of Management, pp. 485 491, 2017.
11. Noriko, Etani, "Development of a predictive model for on-time arrival fight of airliner by discovering correlation between fight and weather data," 2019.
12. [Online]. Available: https://towardsdatascience.com/metrics-to evaluate-your-machine-learning-algorithm-f10ba6e38234.
13. C. J. Willmott, Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square(RMSE) in assessing average model performance," Climate Research, vol. 30, no. 1, pp. 79 - 82, 2005.
14. [Online]. Available: http://scikitlearn.org/stable/modules/classes.html?source=post_page-----f10ba6e38234----------------------#sklearn-metrics-metrics.