IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10850

FLIGHT DELAY ARRIVAL PREDICTION

Mukta S Bharadwaj[1], Dr. Sanjay Kumar C.K[2]

PG Scholar, Dept. of MCA, The National Institute of Engineering (NIE), Mysuru, Karnataka India[1]

Assistant Professor& HoD, Dept. of MCA, The National Institute of Engineering (NIE), Mysuru, Karnataka India[2]

Abstract: Flight delays have significant implications for both travelers and the aviation industry. In this project, we address the challenge of predicting flight delays by harnessing the power of machine learning regression algorithms. By analyzing historical flight data and relevant influencing factors, we aim to provide accurate estimates of delay times for specific flights. The project begins with a comprehensive collection and pre-processing of relevant data, including departure and arrival times, weather conditions, airport congestion, and historical delay records. Feature engineering techniques are employed to extract valuable insights from the raw data, enhancing the predictive capabilities of the models. Several machine learning regression algorithms are explored, including Linear Regression, Support Vector Regression (SVR), and Decision Tree Regression. The selection of these algorithms is based on their suitability for capturing complex relationships and patterns within the data. Hyper parameter tuning and model evaluation are conducted rigorously to ensure optimal model performance. Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE) are employed to quantify the accuracy of the predictions The contribution to our calculation is columns of highlight vector like flight date, take-off delay, separation between the two air terminals, planned appearance time and so forth We at that point use choice tree classifier to foresee if the flight appearance will be delayed or not. Besides, we compare decision tree classifier and calculated.

I. INTRODUCTION

Flight delays have become a persistent concern for travellers and the aviation industry as a whole. These delays can lead to inconvenience, missed connections, financial losses, and operational challenges for airlines and airports. Addressing this issue is crucial to improving passenger satisfaction and operational efficiency. Machine learning methods have become more prevalent recently has emerged as powerful tools for predicting various phenomena, including flight delays. This project focuses on utilizing machine learning regression algorithms to predict flight delays with enhanced accuracy and reliability. The goal of this project is to create forecasting models that calculate the time of flight delays built upon a variety of contributing factors. These factors include historical delay data, departure and arrival times, weather conditions, air traffic congestion, and other relevant variables. By leveraging historical data and real-time inputs, these models can offer timely and accurate predictions, allowing airlines, passengers, and airport authorities to make informed decisions. The project employs a range of regression algorithms, such as 'Linear Regression', 'Support Vector Regression (SVR)', and 'Decision Tree Regression', to capture the intricate relationships between the predictor variables and the target variable, which is the duration of the flight delay. Each algorithm brings its unique strengths to the table, enabling the project to explore different approaches to predicting flight delays. The predictive models are developed in a systematic manner, beginning with data collection and pre-processing. This involves cleaning and transforming the data into a suitable format for analysis. Feature engineering techniques are then applied to extract meaningful information from the raw data, potentially revealing hidden patterns that contribute to flight delays. Once the data is prepared, the project delves into the selection and configuration of regression algorithms. Hyper parameters are fine-tuned through techniques like grid search or random search to optimize model performance. Various metrics are used to assess the models' predictive abilities, allowing for a quantitative assessment of their accuracy and effectiveness. Furthermore, the project recognizes the importance of model interpretability. Transparent models like Linear Regression and Decision Trees offer insights into the elements that contribute to flight delays, which can be valuable for stakeholders seeking to understand the causes of delays and devise strategies for their mitigation. In conclusion, this project aims to harness the capabilities of machine learning regression algorithms to provide accurate and timely predictions of flight delays. By doing so, it contributes to the improvement of passenger experiences and operational efficiency within the aviation industry. The subsequent sections of this report delve into the methodology, experimentation, results, and implications due to the delayed flight prediction models developed in this project. In response to growing concerns of fuel emissions and their negative impact on health, there is an active research from the aviation for finding techniques to predict flight delays accurately to ensure smooth flight operations and reduce delays. The input to our algorithm is rows of feature vector like departure date, departure delay, distance with respect to the two airports, scheduled arrival time etc, Then we apply decision tree classifier to predict if the flight arrival will be delayed or not. In the US, FAA considered a flight to be delayed when difference between Arrivals times both predicted and actual is greater than 15 minutes.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10850

II. PROBLEM STATEMENT

Flight delays continue to be a significant issue in the aviation industry, causing inconvenience to passengers and operational challenges for airlines and airports. Predicting flight delays accurately is a complex task because of the volume of factors involved, including weather conditions, air traffic, and operational constraints. Developing a reliable and accurate 'flight delay prediction system' is essential to address this problem and enhance passenger satisfaction and operational efficiency.

III. RELATED WORK

[1]Because of the inherent complexity of flight data, departure and arrival delays are unpredictable. Numerous flights are postponed or cancelled each year for a variety of reasons. These factors include the environment, safety, carrier delays, and more. Long lines at screening areas, faulty screening equipment, and reboarding of aircraft as a result of security breaches can all contribute to flight delays. Blizzards, hurricanes, tornadoes, and other severe weather events will inevitably cause flight delays or even cancellations. The performance of several different In many different machine learning applications, including pattern recognition, data mining, and machine translation, supervised machine learning algorithms have been used extensively. Similar to this, numerous attempts have been made to apply different supervised or Using unsupervised machine learning techniques, analyze air traffic data. To analyze those air traffic data, Gradient Boosted Decision Tree, one of the well known machine learning tools, has not yet been tried. Modelling sequential data has proven to be extremely accurate using gradient-boosted decision trees. This model makes it possible to accurately predict the daily patterns of arrival and departure flight delays at a specific airport. In comparison to other methods, it demonstrates greater accuracy.

[2]This paper offers a wealth of knowledge on various machine learning algorithms for anticipating delayed flights. Flight delays are a serious issue for both airlines and passengers as air travel grows rapidly. In addition to losing time, passengers also lose faith in airlines. The airline companies will suffer a significant financial loss as a result, and they will also lose their good name. Therefore, accurate tracking and foreseeing of flight delays are crucial. To model flight arrival delays, our study primarily focused on departure delay time, distance, and weather variables. This foreseen outcome lessens both the inconvenience for passengers and the loss experienced by airlines. This work's predictive model is intended to forecast airline arrival. by using supervised machine learning algorithms to reduce delays. The weather data from July 2019 to December 2019 and domestic US flight data were collected and used to train the predictive model. The predictive model that aims to forecast flight delays was developed using the XG Boost and linear regression algorithms. Each algorithm's performance was examined. The model received flight data as well as weather data. Using this data, the XG Boost trained model performed binary classification to determine whether or not there would be an arrival delay, and a linear regression model then predicted the length of the delay for the flight, This project requires the usage of Tensor flow algorithms

[3]This paper thoroughly examines the challenge of optimizing taxiing routes between taxiways and runways within a hub airport. It establishes a comprehensive optimization model centred on safety, efficiency, and fairness in aircraft surface taxiing. Notably, the model takes into account the viewpoints of air traffic controllers, airlines, and airport authorities, while also incorporating the influential aspect of runway occupancy on surface taxiing dynamics.

[4]The paper titled "[8] A Comparative Analysis of Models for Predicting Delays in Air Traffic Networks" by K. Gopalakrishnan and H. Balakrishnan, presented at the USA/Europe Air Traffic Management Seminar in 2017, conducts a comprehensive comparison of different models used for predicting delays within air traffic networks. The study likely evaluates various predictive models to determine their effectiveness in estimating and anticipating delays within the context of air traffic operations.

IV. PROPOSED SYSTEM

The proposed solution involves the development of advanced machine learning regression models to significantly enhance the accuracy of flight delay predictions. Through the integration of historical data, weather insights, airport congestion data, and other relevant factors, the system aims to provide timely and accurate predictions that consider the dynamic nature of flight operations. This represents a marked improvement over current systems by utilizing machine learning techniques capable of identifying intricate patterns and relationships.

Machine Learning Models:
The intended system will leverage a variety of machine learning algorithms, such as regression, decision trees, random forests, or neural networks, to effectively capture complex flight delay patterns.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10850

Feature Engineering: The system will involve the identification and creation of pertinent features, including flight routes, times of day, airlines, airport congestion, and historical delay trends.

Integration of Real-Time Data: The system will seamlessly incorporate real-time data streams, encompassing live flight status updates, current weather conditions, and air traffic information. This integration is poised to significantly enhance the precision of predictions.

V. METHODOLOGY

1. Data Preparation: Input: Flight dataset Process: The pre-processing phase involves identifying and handling missing values, as well as eliminating unnecessary features. Output: Processed dataset without missing values and irrelevant features Error-Handling: In case the input dataset is not valid, appropriate error handling mechanisms are implemented. This stage focuses on enhancing the dataset's quality by addressing gaps and eliminating unimportant attributes, while also accounting for potential input errors.

2. Feature Selection: Input: Processed data Process: Select only pertinent data essential for analysis. Output: Display the chosen data. Feature selection involves identifying and retaining the most influential attributes that significantly impact the analysis outcomes. This aids in reducing noise and enhancing dataset efficiency.

3. Data Separation: Training and Test Sets Input: Feature-selected data Process: Divide the dataset into distinct training and test sets. Output: Display the designated training and test sets for subsequent analysis. Segmenting the data into separate training and test subsets is essential to gauge the performance of algorithms on previously unseen data.

4. Prediction: Utilizing Training and Test Data Input: Trained dataset Process: Execute prediction algorithms on the trained data. Output: Present the obtained prediction outcomes. This phase involves applying the chosen prediction algorithms to the trained dataset and showcasing the resultant predictions.

Regression Algorithm A regression algorithm is designed to find the historical relationship between an independent and a dependent variable to predict the future values of the dependent variable. A regression models the past relationship between variables to predict their future behaviour. The Algorithm uses the linear regression techniques based on the data set collected for the project. The linear regression technique helps in predicting the future behavior of road with help of the statistical methods. The algorithm find the mean and variance value of the dependent variables, and apply the formula Y=b0+b1*x

to predict the future behaviour STEPS for Polynomial Regression:
Step 1: Import libraries and dataset
Step 2: Dividing the dataset into 2 components
Step 3: Fitting Linear Regression to the dataset
Step 4: Fitting Polynomial Regression to the dataset
Step 5: Fitting the Polynomial Regression model on two components X and y.
Step 6: we are visualizing the Linear Regression results using High Chart.
Step 7: Visualising the Polynomial Regression results
Step 8: Predicting new result for Polynomial Regression

VI. EXPERIMENTAL RESULTS

result analysis of the "Flight Arrival Delay Prediction using Regression Algorithms" project demonstrates the efficacy of the implemented system in accurately predicting flight arrival delays. The system's performance was evaluated through rigorous testing and comparison of different regression algorithms, leading to valuable insights into its predictive capabilities. Algorithm Performance Comparison: The implemented regression algorithms, including Decision Tree Regression, Linear Regression, and Polynomial Regression, were subjected to comprehensive testing using historical flight data. The evaluation metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2), were utilized to quantify each algorithm's performance. Accuracy and Generalization: The system's predictive accuracy was impressive, with all algorithms achieving relatively low error values. Linear Regression demonstrated strong performance for capturing linear relationships, while Polynomial Regression excelled in modeling non-linear patterns. Decision Tree Regression exhibited adaptability to intricate data patterns, although it occasionally showed signs of over fitting.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 8, August 2023

DOI: 10.17148/IARJSET.2023.10850

Feature Importance Insights: Feature importance analysis revealed that factors such as departure time, weather conditions, and airline information significantly influenced the prediction of flight arrival delays. This insight aligns with domain knowledge and showcases the interpretability of the chosen algorithms. Real-world Scenario Testing: The system's real-world testing involved predicting flight delays for unseen data points. The predictions closely aligned with actual flight delay observations, reinforcing the models' reliability for practical usage. Visualization and User Interaction: Visualizations, including graphs illustrating feature dependencies and predicted versus actual delay values, enhanced the user experience and facilitated a clearer understanding of the prediction outcomes.

VII. CONCLUSION

This project and the analysis retrieved are beneficial not just for passengers point of view, but for every decision maker in the aviation industry. Apart from the financial losses incurred by the industry, flight delay also portray a negative reputation of the airlines, and decreases their reliability. It causes various sustainability issues, for example, increase in fuel consumption and gas emissions. The analysis carried here not only predicts delays according to the previous available data, but also give statistical description of airlines, their rankings considering their on-time performance, and delays with respect to time, showing the peak hours of delay. This project able to serve as prototype by any aviation authority for their benefit, in the Indian Scenario too, it can work as an efficient model or a proper prototype to study delay analysis, based on the real dataset provided. This project has encompassed and showed the importance of Regression Analysis in Machine Learning, Data Mining concepts for efficient data cleaning, Cross Validation technique and Regularization in ML for making proper models and its predictive analysis

FUTURE ENHANCEMENT

In the future enhancement In advance, we can more number of flight and collect for large number of data for better result use deep learning techniques for the prediction. We are able to collect datasets for all international and national flights data and can make prediction.

REFERENCES

[1]. Koren, Y., Bell, R., &Volinsky, C. (2009). Matrix [1] A Statistical Approach to Predict Flight Delay Using Gradient Boosting Decision Tree by Suvojit Manna, Sanket Biswas
[2]. Machine Learning Model based Prediction of Flight Delay by N Lakshmi Kalyani, Dr.B.V.Kiranmayee
[3]. Y. Jiang, X. Xu, H. Zhang, and Y. Luo, “Taxiing route scheduling between taxiway and runway in hub airport,” Mathematical Problems in Engineering, vol. 2015, 2015.
[4]. K.Gopalakrishnan and H. Balakrishnan, “A comparative analysis of models for predicting delays in air traffic networks,” in USA/Europe Air Traffic Management Seminar, 2017.
[5]. C. Cetek, E. Cinar, F. Aybek, and A. Cavcar, “Capacity and delay analysis for airport manoeuvring areas using simulation,” Aircraft Engineering and Aerospace Technology, vol. 86, no. 1, pp. 43–55, 2013.
[6]. K. B. Nogueira, P. H. Aguiar, and L. Weigang, “Using ant algorithm to arrange taxiway sequencing in airport,” International Journal of Computer Theory and Engineering, vol. 6, no. 4, p. 357, 2014.
[7]. R. R. Clewlow, I. Simaiakis, and H. Balakrishnan, “Impact of arrivals on departure taxi operations at airports,” 2010.
[8]. H. Balakrishnan, “Control and optimization algorithms for air transportation systems,” Annual Reviews in Control, vol. 41, pp. 39–46, 2016.
[9]. N. Pyrgiotis, K. M. Malone, and A. Odoni, “Modelling delay propagation within an airport network,” Transportation Research Part C: Emerging Technologies, vol. 27, pp. 60– 75, 2013.
[10]. I. Simaiakis and H. Balakrishnan, “Queuing models of airport departure processes for emissions reduction,” in AIAA Guidance, Navigation and Control Conference and Exhibit, vol. 104, 2009.

