International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

The Novel Approach of Flight Delay Prediction Using Machine Learning Techniques

Mr. Nithish Raagav[1], J. Sundar[2], Dr. R. Bharathi[3]

[1]Student, Computer Science, SRM University

Abstract:
The issue of delays in flights remains one of the major challenges in modern systems of flight management all across the globe due to low productivity, high costs, and unsatisfactory rates of customer satisfaction. This research presents a new approach to the prediction of delays in flights based on advanced machine learning technologies which use several different databases, namely those regarding previous flights, weather conditions, air traffic data, and airport congestion indicators.

I. INTRODUCTION

This The importance of air transport in modern times cannot be taken for granted because it connects different nations with one another and facilitates international business transactions in an efficient manner. The increasing popularity of air transport has resulted in an increased number of flights taking place in a day; therefore, effective management of flights becomes critical. Negatively, one of the most frequently occurring problems with flight operations is flight delays and timings. Flights can be delayed due to a variety of factors like bad weather conditions, excessive air traffic, aircraft-related problems, and other operational reason,etc. Flights delays not only affect airlines plans and schedules but also shows imapct on negative experience for passengers. Recent developments in aviation technology have made huge amounts of data available while providing new opportunities to harness this data using machine learning techniques.Predictive models based on machine learning algorithms have made it possible to forecast future events based on the analysis of past trends.

II. LITERATURE REVIEW

For quite some time now, flight delay prediction has received a many attention from various researchers, as this topic has direct implications on airline business operations and passenger satisfaction. In earlier times, almost all studies focused on using traditional statistical methods to come up with solutions for flight delay predictions including regression models, probabilistic models, etc.

However, with advancements in computing technologies, there have been attempts to develop more precise and reliable flight delay prediction models using various machine learning tools and algorithm for better performance. Some common algorithms used include Decision Tree, Random Forest and SVM. In recent times, there have been various experiments done using deep learning models. Deep learning models like Neural Networks that specifically include LSTM algorithms have been found quite successful in flight delay prediction tasks because they take into consideration time-based correlations within data samples.

Furthermore, including new sources of real-time data, such as weather and air traffic information, has also helped in improving the effectiveness of flight delay predictions. However, despite numerous advances in flight delay predictions, there are still some gaps and simple mistake in existing literature.

III. METHODOLOGY

- The proposed system provides prediction of flight delays using Machine Learning.The proposed system uses flight data as input and predicts whether a flight will get delayed or not
- Data Collection Module: Data is gathered from multiple sources like flights schedule, weather forecast and airport traffic.
- Data Preprocessing Module: This module converts the gathered data into an organized form by cleaning,normalizing and transforming the data(such as handling missing or inconsistent data,converting categorical variables to numeric etc.).
- Feature Selection Module: Relevant features needed for prediction are selected.Based on intuition features like departure time,weather,carrier and delays are selected.
- Training Module: The module where trained models use datasets to learn about various patterns and flight delays.
- Prediction Module: The test data is given as input to trained models to predict or analyse if the next flight will get delayed.
- Evaluation Module: Evaluating the system by using various parameters such as accuracy,precision,recall,F1-score etc.
- This system follows a pipeline architecture in which each module builds upon the previous module.

Techniques used

Metric this project takes advantage of Machine Learning technologies in order to increase prediction precision and reliability. The main focus lies on using supervised algorithms because the historical data is already labelled, and in each case, the status regarding the delay is known. Random Forest and Decision Trees are used, because they have proven effective with complex data and are relatively easy to understand.

Apart from those, we also make use of Ensemble algorithms to combine the outputs of several models, in order to reduce errors and to improve the quality of predictions as well. Feature Engineering plays an important part in the development of a precise prediction model by generating new features from old data. Preprocessing measures such as normalization, dealing with missing values, and balancing the data set can also be used in the course of development. In some cases, we also make use of time-related analysis in order to detect possible patterns in relation to the time of day, day of the week, or seasonal changes.

Design approach

The development of the system follows a systematic and progressed approach that to strike a balance between accuracy and usability. Initially, the problem is stated that explicitly and the necessary data is collected from reliable sources. Second, the collected dataset is analyzed to get an understanding of its nature and structure, and also potential problems with its use.

Afterwards, the data is preprocessed and key features are selected, after which the data is ready to be used in the training process. In this stage, multiple machine learning models are implemented and analyzed based on their performances, after which the best performing one(s) are chosen.

Also, the end-user is taken into consideration when designing the system where the output predictions are displayed in a user-friendly manner.The design method of this project is created with practicality,

precision, and ease of execution in mind. The framework is designed with an organized approach, guaranteeing that each step makes a significant contribution to the predictive process.

Data preparation is another essential element in the design process. It is important since the model's performance may be significantly improved via appropriate data processing methods. The initial information is usually inaccurate or incomplete, making data preparation necessary before any model can be built.

Flow chart

Methodology for the development of the system can be presented by means of a flowchart which describes all the stages of the flight delay prediction process. Initially, there should be collected data on the issue. After data collecting, there should be performed data pre-processing stage. Feature selection and extraction are to be completed at the third stage. After this stage, there follows the training stage, when a machine learning algorithm learns from the data.

Evaluation stage comes after training. In case the performance of the model is good enough, then prediction and deployment can be considered. In case there is some need for further improvement, the whole process can be repeated once again.

IV. IMPLEMENTATION

In carrying out this project, it was done in an orderly and practical way whereby the emphasis was laid on the creation of a dependable flight delay prediction model through machine learning tools and techniques for better enhanced perfomance. The whole process was conducted in a programming environment where data analyses could be done easily, thus ensuring that large amounts of data were processed without difficulties. This project follows a particular pipeline whereby the first step involves the gathering of data, followed by prediction.

A. Tools used

The various tools and applications used in this study proved to be instrumental in the success of the research process. Python programming language was one of the main languages that were employed in this study because it provides a plethora of libraries used in data manipulation and machine learning.

Libraries like NumPy and Pandas were used to process the dataset while Matplotlib and Seaborn provided visual analysis of the data. Scikit-Learn library was applied in the development of prediction models due to its ease of use when dealing with decision trees, random forest, and support vector machines among others.

TensorFlow/Keras library was at times employed to provide solutions on complex neural network technique.The dataset utilized in this experiment contains historical data regarding flights, which contains many factors affecting the delay. The common factors may contain flight numbers, departure time, arrival time, source airport, destination airport, weather conditions, and whether there is delay or not.

The dataset was collected from open sources or generated through simulation experiments. Prior to running the machine learning model on the data, proper preprocessing was done. The preprocessing phase involved treating the missing value problem, deleting duplicates, encoding categorical variables, and scaling the data when necessary.

B. Result of Dataset used

After cleaning the data and making it ready for processing, it was split into training and testing subsets. The training subset was utilized for creating and training the models, while the testing one was applied for evaluating the performance of each of them. Several models were developed and compared to select

the most optimal approach. The efficiency of each model is evaluated using different metrics, such as accuracy, precision, detailing and recall.

In general, the experimental design aimed at ensuring that the model was accurate and accessable to implement. The use of efficient machine learning tools, appropriate data management, and correct model choice led to the development of an effective flight delay prediction system with no lagging.

In this regard, for the research titled "A Novel Approach of Flight Delay Prediction Using Machine Learning Techniques," the dataset used in this study has been obtained from readily available sources in the fields of aviation and meteorology, with a view to incorporating diversity and relevance. The dataset used herein involves historical data of flights, including features such as flight number, company name, airport of departure and arrival, scheduled time of departure and arrival, delay time, and cancellations. Additional features have also been added to improve the prediction accuracy of the model, namely weather data, air traffic intensity, and time-related variables.

ACKNOWLEDGMENT

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try “R. B. G. thanks”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

The following is an example of an acknowledgment.

The authors gratefully acknowledge the contributions of T. Edison, G. Westinghouse, N. Tesla, A. Volta and A. Ampere to the electric power industry.

Project Outcome

Thus, a model was produced during the research that has been proven to be able to accurately forecast the delay of aircraft more effectively than any other statistical method known before.

Having analyzed different factors, such as previous flights' schedules, information about the weather, and air traffic, it became evident that through the implementation of machine learning methods such as ensemble and neural networks, the predictions would become possible owing to their capacity to assess complex non-linear dependencies.

REFERENCES

1. M.L.Hansen , J.Zhang ,”Predicting flight delays using machine learning techniques”, 2019 ,IEEE xplore
2. AS.Choi , Y.J.Kim, “ Prediction of flight arrival delays using machine learning models”, 2016 , Springer.com
3. JK.R.Patil , S.V.Bhosale , “Flight delay prediction using data mining techniques”, 2018 , IEEE xplore.org
4. EE.Frank , M.A.Hall ,”Data mining: tools and techniques for flight delay analyse”, 2016, Science direct.com
5. R.K.Sharma , A.Kumar , “Analyze of flight delay prediction using machine learning algorithms”, 2021 , arxiv.org
6. DP.Balakrishnan , R.Ganesan .”Accuracy of reinforcement learning algorithms for predicting flight delays”, September 2022, Reseaech gate.net
7. H.Lee , D.Lee , S.Kim ,” Deep learning model for flight delay prediciton”, 2024 , IEEE Xplore.org

8. B. Yu , H.Wang , G.Chen ,” Flight delay prediction for commercial air transport: deep learning approach” , 2019 , Science direct.com
9. X. Liu , M.Zhang ,” A Hybrid machine learning model for flight delay prediction “ , January 2020 , IEEE Xplore.org
10. D. Rupasinghe , T.Fernando , “ Machine learning approach for airline delay prediction”, 2021 , IEEE Xplore.org