WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Air Traffic System Optimization Through Predictive Analytics A Project Report

Joshini M[1], Deepthika V[2], Afrin Noorjahan S[3], Mrs. Uma Devi G[4], Dr. S. Sharmila[5]

[1,2,3]Student, **[4]Assistant Professor Computer Science and Engineering, [5]Head Of The Department, Computer Science and Engineering**
[1,2,3,4,5]Agni College of Technology, Chennai-600130, Tamil Nadu, India

ABSTRACT:

In response to the increasing air travel delays and their impact on budgetary constraints, operational challenges, and customer dissatisfaction within the aviation industry, we implemented supervised machine learning techniques. Our primary goal was to predict flight delays using a comprehensive dataset that included flight information from JFK airport over a two-year period. The decision tree algorithm was employed to create a binary classification model specifically designed for identifying flight delays. To evaluate the performance of these algorithms, we conducted a thorough analysis using essential metrics such as accuracy, precision, recall, and f1- score. These metrics allowed us to address any disparities within the dataset. Our comparative analysis revealed that the Decision Tree algorithm delivered the highest performance It is worth noting that tree-based ensemble classifiers outperformed other base classifiers in terms of predictive accuracy and overall performance.

INTRODUCTION:

Air travel has become An essential aspect of contemporary society, air travel plays a crucial role in facilitating long-distance journeys for individuals. quickly and efficiently in a given time. However, with the increasing demand for the air travel, flight delays have become a common occurrence that can cause inconvenience, disruptions and frustration, annoyance for passengers, as well as financial losses for the airlines or aviation industries and other stakeholders. There are some important varieties of factors which causes the flight delays. Some types of flight delays are listed below:>>

- Air Traffic Control (ATC) Delays
- Weather Delays
- Mechanical Delays
- Crew Delays
- Security Delays

LITERATURE SURVEY:

According to Alice Sternberg, flight delays have a number of detrimental effects on the aviation industry, airports, and travellers. Their forecast is more important when commercial airline players are making decisions.

Jeorge Soares The creation of precise flight delay prediction models has proven challenging due to the complexity of air travel systems, the abundance of prediction techniques, and the vast amounts of data related to such systems.

According to Diego Carvalho, this study provides a thorough analysis of strategies used to develop flight delay prediction systems from a data science perspective. "We propose a taxonomy and summarise the efforts used to address the flight delay prediction problem, according to scope, data, and computational methods, paying particular attention to an increasing usage of machine learning," says Eduardo Ogasawara..

Existing and proposed system

Existing system:

- In the existing system, the prediction is done which as some inaccurancy and in major cases the flight delays are due to the previous flight delay.
- So the correct reason and confusion still exists in the current system .
- In previous system, they used random forest algorithm which is not as easy to visually interpret as decision trees.

- They have also used Neural regression algorithm (Multi-layer perception) is very slow and the logistic regression algorithm may cause some risk of underfit.

Proposed System:

- In proposed system, the prediction are done based on machine learning and in has a accurate prediction and in addition we use python which is used in the wide range of data analysis and the machine learning.
- In this proposed system, we are using Supervised machine learning model called Decision tree algorithm instead of using neural network algorithm, a random forest algorithm and a logistic regression.
- This decision-tree algorithm used to exact and accurate prediction for flight delays.
- In addition to this, we using label encoding, one hot encoding which is a simple and effective technique for converting categorical variables into numerical forms.

CAUSES FOR DELAYS:

WEATHER DELAYS:

Flight may be delayed or cancelled due to some changes in the climate such as heavy rain, snow fall, violent wind, cloudburst, lightening, thundering, fog or low visibility , storms and so on. These climatic changes does not affect the flight always but the pilot should feel comfortable with these changes otherwise the flight would be delayed. If the sudden climatic changes occurs then the pilot would land the flight at the nearest airport.

CREW DELAYS:

Crew delays are often caused when the particular pilot or co-pilot is currently not available due to any reasons. In this case a colleague will take off the flight and this would take some time for the departure.

MECHANICAL DELAYS:

The mechanical malfunctions and disruptions such asengine/power losses, damaged hydraulics, jamming of the wheels, loss of flight control etc..can cause the mechanical flight delays.

SECURITY DELAYS:

The security delays are caused due to some reasons such as criminal activities, illegal drug carrying, travelling with expired passport that are done by the passengers and can be caught up by the airline securities or guard.

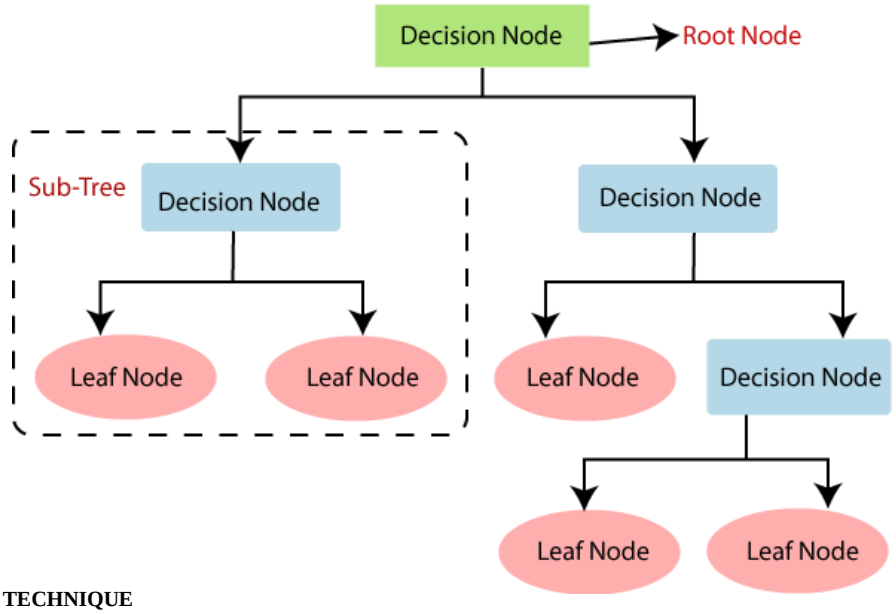
ALGORITHM/TECHNIQUES USED TO PREDICT DELAYS:

DECISION TREE

Decision trees are one form of supervised learning technique that can be used to solve classification and regression problems, however they are most frequently employed to address classification issues. It is a tree-like-structured classifier, with a leaf node that stands in for the result, internal nodes for highlights from a dataset, and branches for the decision-making processes.

- There are 2 following nodes in the decision tree and they are:
  - Leaf node
  - Decision node

It is described as a graphical representation for gathering all feasible answers to a choice problem based on predetermined conditions.



LABEL ENCODING TECHNIQUE

▪We use label encoding in machine learning when we have categorical variables that need to be converted into numerical format.

▪Machine learning algorithms typically working with numerical data, so categorical

variables (such colour, gender, or nationality) need to be converted into numerical form for analysis.

▪Label encoding is a technique that gives a unique integer values to each category in the categorical variable, replacing the original categorical values with numerical labels. For example, consider a categorical variable `color with the categories 'yellow', 'red', and 'black', the label encoding would assign the integer values (for e.g, 0, 1, 2) to each category.

▪This allows machine learning algorithms to process the data more easily and effectively.

ONE HOT ENCODING TECHNIQUE:

■One hot encoding technique is a another technique that used to convert the categorical variables into the numerical format for machine learning.

Consider a categorical variable called "colour," which has the categories "yellow,""red," and "black." Three binary vectors, one for each category, would be produced using a single hot encoding technique:

'yellow':[1, 0, 0]

- 'red': [0,1, 0]
- 'black: [0, 0, 1]

▪Each and every binary vector represents a category and can be used as input to a machine learning. algorithm.

▪One hot encoding ensures that the numerical values for each category are orthogonal to each other, which helps to avoid bias in some machine learning algorithms.

STANDARD SCALER PREPROCESSING TECHNIQUE:

▪ A dataset can be transformed using the preprocessing method known as StandardScaler so that its features have a mean of zero(0) and a standard deviation of one(1)

▪It is a common data transformation technique used in machine learning.

▪The StandardScaler algorithm calculates the mean and standard deviation of each feature of the dataset, then subtracts the mean from each data point and divides by the standard deviation.

▪This centers the data around zero(0) and scales it to the standard deviation of one(1).

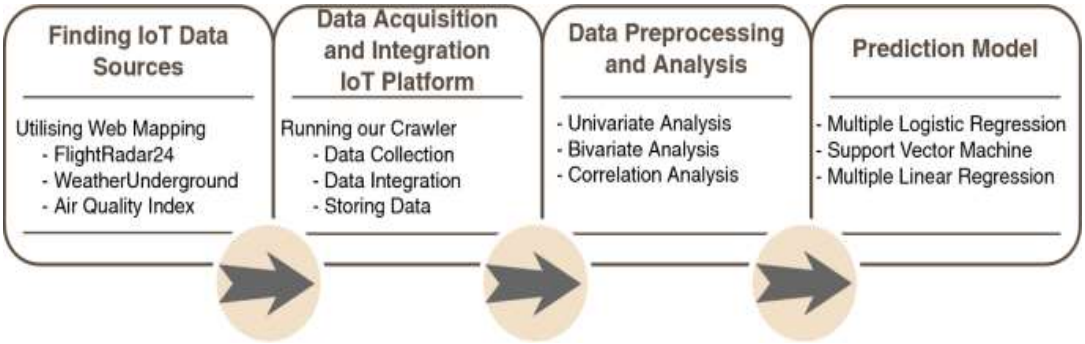
▪Thenormal formula for standardization is:

x_standardized = (x mean) of standard deviation

where x refers to the data point, mean indicates the mean of the features, and the standard deviation refers to the standard deviation of the features.

▪The main advantages of standardizing the data is that it can help improves the performance of some machine learning algorithms, especially those that rely on distance measures or require normally distributed data.

▪Standardizing the data also makes it easier to visualize and interpret the results.



CONCLUSIONS:

By using it to a machine learning problem, the article calculates the likelihood of aircraft delays.

For the binary classification-based flight delay prediction, we employ supervised machine learning.

The estimation methods were weighted to reduce the impact of the dominance of the non-delayed flights over the delayed flights due to the imbalanced nature of the data set.

After using the classifier to estimate latency, we compared the values of the four metrics to determine how well each model performed.

The outcome highlights that the model of the Decision tree method generates the highest values of precision, accuracy, recall, and f1-score.

The Decision Tree works well when predicting flight delays in the highest levels, demonstrating The performance of other tree-based group classifiers is likewise good for flight predictions.

The data set used for this presentation is unevenly distributed, which could lead to unique variations in how each algorithm performs. This issue was resolved in this presentation utilizing weighted assessment metrics.

Future research can benefit from employing methods like SMOTE. We employ a method called SMOTE for the upcoming research to better fix this imbalance and to improve the prediction.

The comparison of this algorithm's results shows that, for this data set, tree-based group algorithms tend to provide the most accurate predictions of flight delays.

To determine the significance of the experiment for predicting flight delays, it would be beneficial to do a similar experiment again using a tree-like ensemble algorithm.

REFERENCES :

▪Khaksar, H., &Sheikholeslami, A. (2017). Airline delay prediction by machine learning algorithms. ScientiaIranica.

▪Esmaeilzadeh, E., &Mokhtarimousavi, S. (2020). Machine learning approach for flight departure delay prediction and analysis. Transportation Research Record: Journal of the Transportation Research Board.

▪M. Al-Tabbakh, S., M. Mohamed, H., & H. El, Z. (2018). Machine learning techniques for analysis of Egyptian flight delay. International Journal of Data Mining & Knowledge Management Process, 8(3), 01–14.

▪Ye, B., Liu, B., Tian, Y., & Wan, L. (2020). A methodology for predicting aggregate flight departure delays in airports based on supervised learning. Sustainability, 12(7), 2749.

▪ATLIOĞLU, M. C., BOLAT, M., ŞAHİN, M., TUNALI, V., & KILINÇ, D. (2020).

Supervised learning approaches to flight delay prediction. Sakarya University Journal of Science.

▪Yu, B., Guo, Z., Asian, S., Wang, H., & Chen, G. (2019). Flight delay prediction for commercial air transport: A deep learning approach. Transportation Research Part E: Logistics and Transportation Review, 125, 203–221.