International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 09 Issue: 11 | Nov 2022 www.irjet.net p-ISSN: 2395-0072

Departure Delay Prediction using Machine Learning.

Prof.Shardha Thete[1], Jaya Rao[2], Pratik Satao[3], Anish Mate[4], Arfat Shaikh[5]

ALARD COLLEGE OF ENGINEERING & MANAGEMENT
(ALARD Knowledge Park, Survey No. 50, Marunji, Near Rajiv Gandhi IT Park, Hinjewadi, Pune-411057)
Approved by AICTE. Recognized by DTE. NAAC Accredited. Affiliated to SPPU (Pune University).

Abstract - *Many companies rely on different airlines to link them with other regions of the world nowadays since the aviation industry is so important to the global transportation sector. Extreme weather, however, can have a direct impact on the airline. In order to address this problem, reliable forecasting of these flight delays helps travelers to prepare for the disruption to their travel plans and enables airlines to address the causes of the flight delays in advance to lessen the impact. This project's goal is to examine the methods for creating forecasting models for aircraft delays brought on by adverse weather. In the present research, a machine learning flight delay prediction model is established with the help of machine learning ensemble models. Three different machine learning techniques such as Naive Bayes, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) applied to the Airline dataset. To validate the per- formance and efficiency of the proposed method, a comparative analysis is performed*.

Key Words: (*Naive Bayes, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM*)**:** Machine learning, Prediction, reprocess) …

1. Introduction:

Passenger airlines, cargo airlines, and air traffic control systems are the main elements of any transportation system in the modern world. Nations all around the world have attempted to develop various methodologies over time to enhance the airline transportation system. This has significantly altered how airlines operate. Modern travelers may experience annoyance from flight delays. Around 20Flight operations on commercial aircraft have become increasingly complex and dynamic. The daily operations of airlines require adjustments to be made in response to factors such as weather conditions, mechanical issues, or passenger complaints. These variables can also impact routes and schedules, increasing variability in flight activity at commercial airports. Managing the complex interaction between passengers, planes, airports and the demands of aviation stakeholders is challenging for airlines and traffic flow managers at commercial airports who must respond quickly to unexpected changes in demand.

1.1 Pre-Process:

There may be a lot of useless information and gaps in the data. Data preparation is carried out to handle this portion. Numerous data pre-processing techniques, such as data cleansing, data transformation, and data reduction.

Preprocessing is the process of converting unprocessed characteristics into information that a machine learning algorithm can comprehend and learn from. As illustrated, selecting the appropriate methods and preprocessing approaches for machine learning needs careful examination of the raw data and is somewhat of an art form.

1.2 Normalization

Normalization is a scaling method, a mapping method, or a first stage of processing. From an existing range, we can build a new range. It can be quite beneficial for objectives of forecasting or prediction. As we all know, there are several techniques to predict or forecast, but they can all differ significantly from one another. Therefore, the Normalization approach is needed to bring prediction and forecasting closer while maintaining their wide range. Data should be transformed such that they are either dimensionless or have equal distributions in order to be normalized. Other names for this normalizing procedure include standardization, feature scaling, etc. In every machine learning application, model fitting, and data pre-processing, normalizing is a vital step. Each variable is given an equal amount of weight during normalizing to ensure that no one variable can affect model performance just because they have larger amounts

1.3 Feature Extraction:

The process of translating raw data into numerical features that may be processed while keeping the information in the original data set is referred to as feature extraction. It produces better outcomes than applying machine learning straight to raw data.

It is possible to extract features manually or automatically:

Identification and description of the characteristics that are important to a particular situation are necessary for manual feature extraction, as well as the implementation of a method to extract those features. A good understanding of

the context or domain may often aid in making judgments about which characteristics could be helpful.

2. Algorithms:

2.1 Naïve Bayes

The supervised learning technique is known as the Naive Bayes algorithm, that is built on the Bayes theorem, is often used to resolve classification issues. It is mostly applied in text summarization with a large training set. The Naive Bayes Classifier is one of the most suitable and efficient classification algorithms available today. It assists in the development of rapid machine learning models capable of making accurate predictions. Being a classification algorithm, it makes predictions based on the chances that an object will appear. It is a family of algorithms rather than a single method, and they are all based on the idea that every pair of features being classified is independent of the other.

2.2 SVM

Support Vector Machine is a common Supervised Learning technique for Classification issues. However, it is largely employed in Machine Learning. The SVM algorithm's objective is to establish the optimal lines or distance measure that can divide n-dimensional area into classes, allowing us to quickly classify fresh data points in the ahead. A hyperplane is the name given to this optimal decision border. SVM selects the maximum vectors and points that helps in the creation of the hyperplane. Support vectors, which are used to represent such high instances, form the basis for the SVM method.

2.3 KNN

One of the simplest machine learning algorithms, K-Nearest Neighbor uses the supervised learning method. A new data item is classified using the K-NN algorithm based on similarity after all the existing data has been stored. This means that utilizing the K-NN method, fresh data may be quickly and accurately sorted into a suitable category. Although the K-NN approach is most frequently employed for classification issues, it may also be utilized for regression.

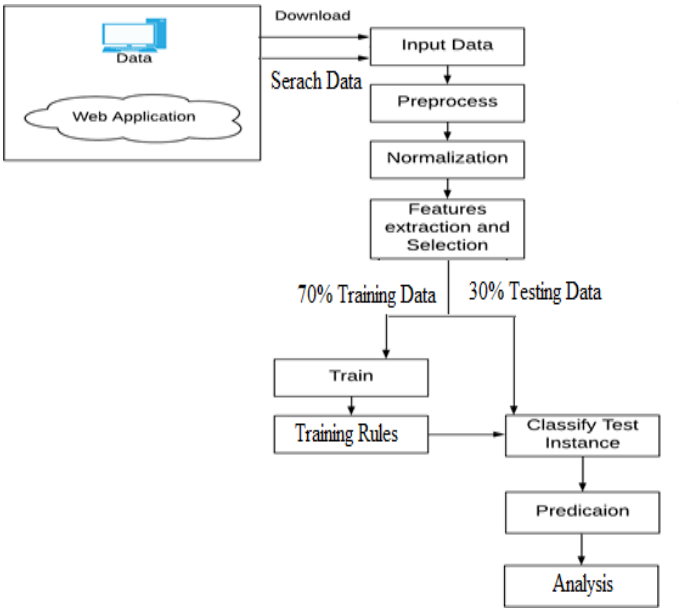
3.1. Objectives of the study

The Main Goal is to study and analyze Flight Delay Prediction system using machine learning techniques and to design and develop feature extraction and selection techniques to achieve better classification accuracy and to design and develop machine learning techniques for effective flight delay prediction, also we need to explore and validate the proposed system results with various existing systems and show the system's effectiveness.

3.2. Methodology

We imported the dataset out from directories using one of the machine learning libraries called Pandas because it offers an express, robust, and an array that simply tries to manipulate the data throughout many other environments. The dataset is in the csv format used in data science. Comma-separated values, or CSV in this case, may be imported into pandas from a local directory.

4. System Architecture:



5. Scope of the study

5.1 Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. Activity Diagrams explain how activities are organized to create a service at various levels of abstraction. Typically, an event must be accomplished by some operations, especially when the operation is designed to do a number of different tasks that require coordination, or how the events in a single use case connect to one another, especially in use cases where activities may overlap and require coordination. It is also suitable for modelling how a set of use cases interact to represent business workflows.

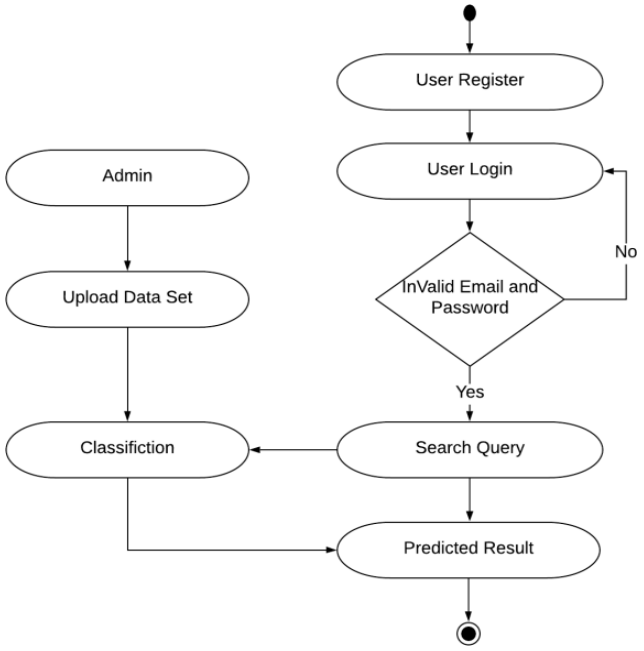


Diagram 1: Activity Diagram

5.2 The scope and high-level functions of a system are described in use-case diagrams. The interactions between the system and its actors are also depicted in these diagrams. Use case diagrams show how users and other external entities interact with internal software systems. Use case diagrams, a behavioural UML diagram type, are frequently used to assess various systems. They enable you to observe the many roles in a system and how they interact with it. A graphical representation of a user's potential interactions with a system is called a use case diagram. A use case diagram, which is frequently complemented by other types of diagrams, displays the numerous use cases and user types the system has. Either circles or ellipses are used to depict the use cases.

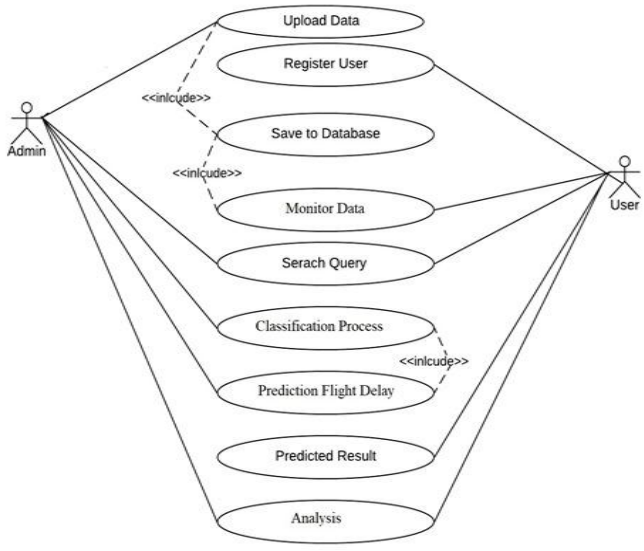


Diagram -2: Use Case Diagram

6. CONCLUSION:

A very common issue in the aviation sector that costs important time and energy are flight delays. To identify the most important causes of aircraft delays, this study examined airline data on delays. For the purpose of determining if a delay would occur in the future, we also investigated machine learning ensemble models. The data shows that the originating airport, followed by the airline a customer chooses to travel with, are the two factors that are most important in determining whether a delay will occur or not.

ACKNOWLEDGEMENT

Efforts have been taken in this project. It would not, however, have been possible without the kind support and assistance of many individuals and organizations. We would like to extend our sincere thanks to all of them. We are highly indebted to Prof. Shardha Thete for his guidance and constant supervision as well as for providing relevant project details, as well as their assistance in finishing the project. We would like to express our gratitude towards my parents & member of the Computer Department of Alard College of Engineering, Pune for their kind cooperation and encouragement which help us in the completion of this project. We would like to express our special gratitude and thanks to All other Professors in the Department for giving us such attention and time. Our thanks and appreciations also go to our colleagues in developing the project and people who have willingly helped us out with their abilities. Last but not the least, we are grateful to all our friends and our parents for their direct or indirect constant moral support throughout the course of this project.

REFERENCES

1. Gui, Guan, et al. "Machine learning aided air traffic flow analysis based on aviation big data." IEEE Transactions on Vehicular Technology 69.5 (2020): 4817-4826.

2. Gui, Guan, et al. "Flight delay prediction based on aviation big data and machine learn- ing." IEEE Transactions on Vehicular Technology 69.1 (2019): 140-150.

3. Zhang, Kai, et al. "Spatio-temporal data mining for aviation delay prediction." 2020 IEEE 39th International Performance Computing and Communications Conference (IPCCC). IEEE, 2020.

4.Yazdi, Maryam Farshchian, et al. "Flight delay prediction based on deep learning and Levenberg-Marquart algorithm." Journal of Big Data 7.1 (2020): 1-28.

5.Huo, Jiage, et al. "The Prediction of Flight Delay: Big Data-driven Machine Learning Approach." 2020 IEEE International Conference on Industrial Engineering and Engineer- ing Management (IEEM). IEEE, 2020.

6.Gui, Guan, et al. "Flight delay prediction based on aviation big data and machine learn- ing." IEEE Transactions on Vehicular Technology 69.1 (2019): 140-150.

7.Alla, Hajar, Lahcen Moumoun, and Youssef Balouki. "A multilayer perceptron neural network with selective-data training for flight arrival delay prediction." Scientific Programming 2021 (2021).

8.Esmaeilzadeh,Ehsan, and Seyedmirsajad Mokhtarimousavi. "Machine learning ap□proach for flight departure delay prediction and analysis." Transportation Research Record 2674.8 (2020): 145-159.

9. Yi, Jia, et al. "Flight delay classification prediction based on stacking algorithm." Journal of Advanced Transportation 2021 (2021).

10. Liu, Fan, et al. "Generalized flight delay prediction method using gradient boost□ing decision tree." 2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring). IEEE, 2020