iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53038

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Departure Delay Prediction using Machine Learning

Prof. Sakshi Shejole[1], Jaya Rao[2], Pratik Satao[3], Anish Mate[4], Arfat Shaikh[5]
Alard College of Engineering

Abstract: ***The project aims to investigate forecasting strategies for forecasting weather-related planes delays. In order to plan ahead and reduce the effects of disruptions, it is essential for airlines and travelers to be able predict such delays with accuracy. The research focuses on developing an ensemble model-based machine learning flight delay prediction system. The Airline dataset is subjected to the use of three distinct machine learning methods: Random Forest Clasifier, k-nearest-neighbor (KNN), and Support Vector Machine (SVM). The suggested approach is then assessed for efficiency and outcomes using a comparative analysis.***

Keywords: (Random Forest Clasifier , K-Nearest Neighbor Clasifier (KNN), and Support Vector Machine (SVM)

I. INTRODUCTION

Passenger airlines, cargo airlines, and air traffic control systems all play major roles in the modern transportation system. There have been significant changes in how airlines operate as a result of the numerous methodologies that have been developed by countries all over the world to increase the effectiveness and efficiency of airline transportation. However, these developments have also brought about difficulties like flight delays, which can irritate modern travelers. Due to variables like weather, mechanical problems, and passenger concerns, flight operations are becoming more complicated and changing, requiring constant adjustments. These factors may have an effect on flight paths and schedules, leading to more fluctuating flight activity at commercial airports. Airlines and traffic flow managers must successfully manage the complex interactions among passengers, aircraft, airports, and the expectations of aviation stakeholders..

II. LITERATURE SURVEY

LITERATURE REVIEW OF Flight Delay Prediction. According to [1] using the dispersed ADS-B ground stations and the collected ADS-B messages, the initial version of an aviation big data platform is created. The air traffic flow between various cities may be tallied and forecasted by examining the produced dataset and mapping the extracted information to the routes; the prediction job is accomplished using two distinct machine learning techniques, respectively. According to [2] a wider range of variables that might possibly affect the flight delay are considered, and extended flight delay prediction challenges are devised to evaluate various machine learning based algorithms. Automatic dependent surveillance broadcast (ADS-B) signals are collected, preprocessed, and combined with additional data, such as weather conditions, aircraft schedules, and airport details, to provide a dataset for the suggested strategy. Different classification tasks and a regression task are included in the intended prediction tasks. Long short-term memory (LSTM) is capable of managing the acquired aircraft sequence data, according to experimental findings, although over fitting issues arise in our small dataset.

According to [3] a cutting-edge technology for commercial aircraft that predicts flight delays based on stacked Long Short-Term Memory (LSTM) networks. The system gathers essential aspects including weather, air traffic, airspace, and human factors data along later routes by learning from prior trajectories via automated dependent surveillance-broadcast (ADS-B) signals and using the corresponding geolocations.

Our suggested regression model receives these combined characteristics as input. The LSTM architecture learns and abstracts the latent spatiotemporal patterns of the data. According to [4] a Deep Learning-based technique for forecasting aircraft delays (DL). DL is one of the most recent approaches used to address issues with high levels of complexity and vast amounts of data. Additionally, DL has the ability to automatically extract the key characteristics from data. Furthermore, a method based on stack denoising autoencoder is created and incorporated to the suggested model since the majority of flight delay data are noisy. In order to determine the appropriate weight and bias values, the Levenberg-Marquart method is also used. Finally, the output has been adjusted to deliver very accurate findings.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

III. PROPOSED SYSTEM

A. Login Module

User login page.

B. Prediction Page Module

Prediction page module you can enter Flight number, origin airport and destination airport and predict the analysis.

C. Classification Module

Classification module we used the Random Forest Classifier as its accuracy compared to K-Nearest Neighbor and Support Vector Machine is high at the given data set.

1) Register user details and login to the page.
2) Enter Flight details- Flight no, Origin Airport, Destination Airport.
3) Click on the submit button below.
4) Accuracy will be calculated using Random Forest Classifier.
5) Result Generation.
6) Display Prediction Result.

IV. ALGORITHMS

Random Forest : Popular machine learning algorithm Random Forest is a part of the supervised learning approach. It can be applied to ML problems including Classification and Regression.

It is based on the idea of ensemble learning, which is the process of combining different classifiers to solve a complex problem and enhance the performance of the model. As its name implies, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on a single decision tree, the random forest takes the prediction from each tree and predicts the outcome based on the majority votes of prediction.

K Nearest Neighbor (KNN): The K Nearest Neighbors (KNN) algorithm is a popular, straightforward, and versatile machine learning technique used in various fields, including handwriting detection, image recognition, and video recognition. KNN is particularly valuable when obtaining labeled data is challenging or expensive. It can achieve high accuracy in a wide range of prediction problems.

KNN operates by finding the K nearest data points to a new, unlabeled input based on their feature similarities. These nearest neighbors contribute to predicting the label or value of the new data point. The algorithm does not involve learning a specific function or target, but rather uses the local characteristics of the data.

It determines the neighborhood of the unknown input, calculates the distance or similarity measures, and considers other parameters to make predictions. The underlying principle of KNN is the idea that similar data points tend to share similar labels or values. By leveraging the concept of "information gain," the algorithm determines which neighboring data points are most relevant for predicting the unknown value.

Support Vector Machine (SVM): It is a supervised machine learning algorithm utilized for classification and regression tasks. It is widely recognized and employed for classification purposes. The main objective of SVM is to discover a hyperplane in an N-dimensional space that distinctly separates different classes of data.

The SVM algorithm identifies the closest data points, known as support vectors, from each class. It determines the hyperplane by considering these support vectors.

The margin, which is the distance between the support vectors and the hyperplane, is maximized by the SVM algorithm. By maximizing the margin, SVM aims to achieve better generalization and robustness. One of the notable characteristics of SVM is its ability to handle outliers effectively. The algorithm is capable of ignoring outliers while finding the optimal hyperplane that maximizes the margin. This helps in reducing the impact of noisy or irrelevant data points. SVM is particularly useful in cases where the number of features or dimensions is high. It can effectively handle high-dimensional data and make accurate predictions. Additionally, SVM is memory efficient as it only uses a subset of training points, the support vectors, in the decision function. This property allows SVM to scale well to large datasets.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

A. System Architecture

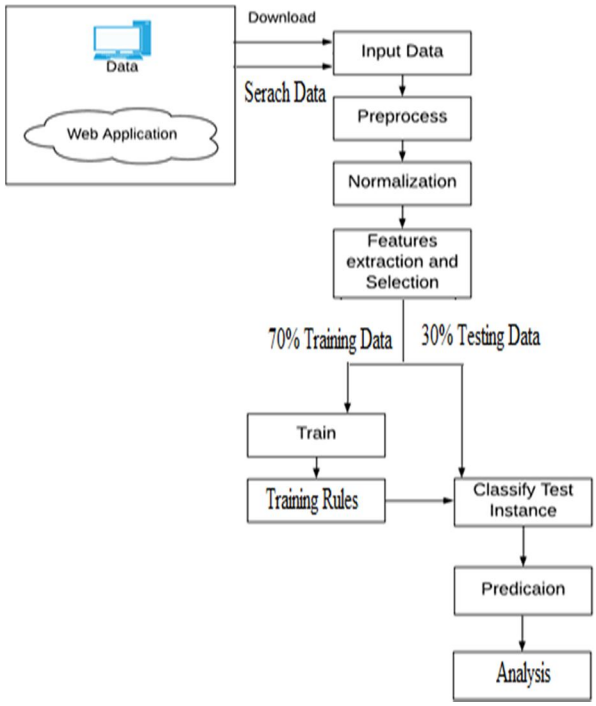
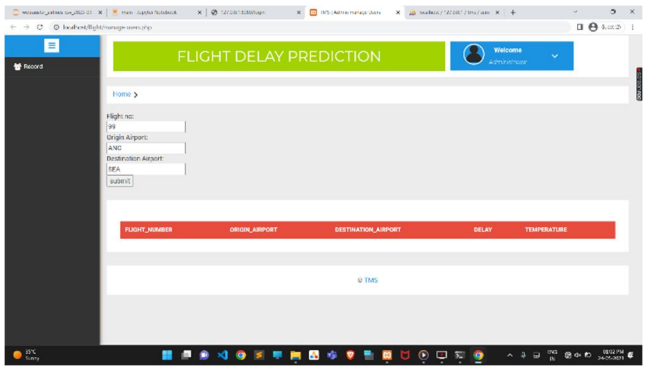
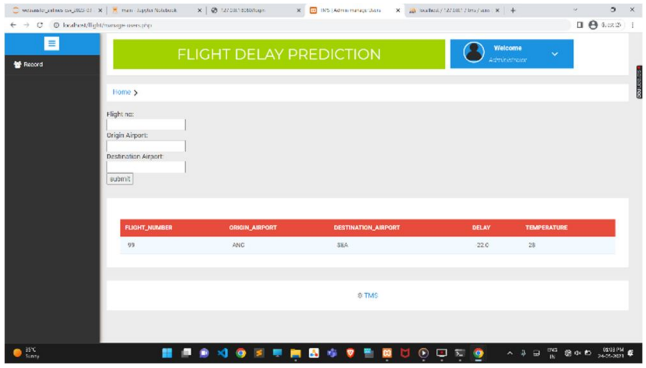


Fig. 1. System Architecture

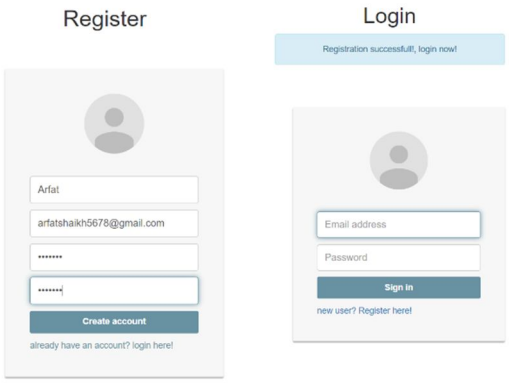
V. RESULTS



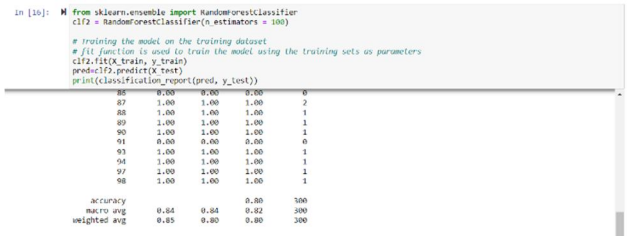
Enter data



Prediction



User Register/Login



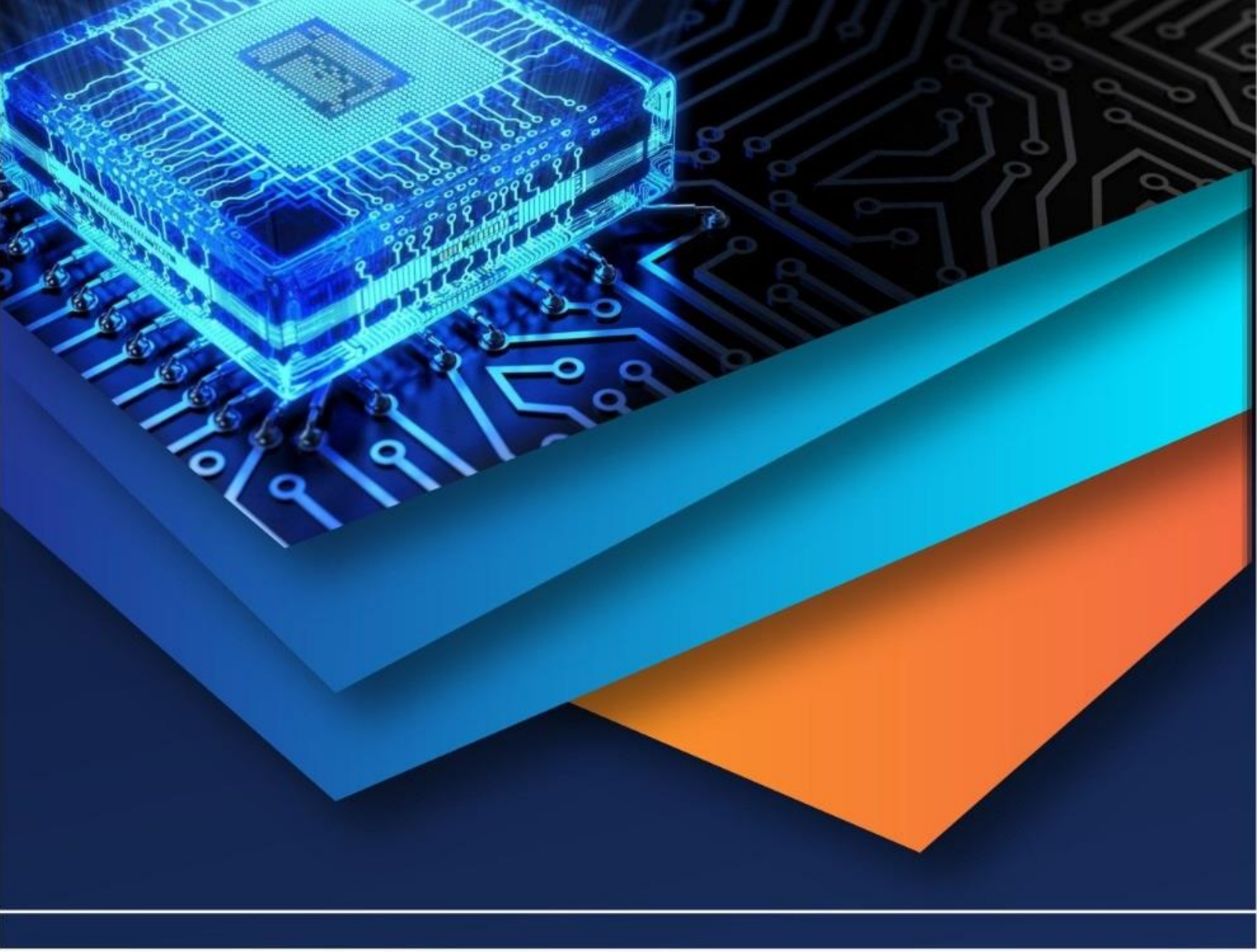
Random Forest Classification Accuracy.

VI. CONCLUSIONS.

Flight delays are a prevalent problem in the aviation industry, consuming valuable time and resources. In order to identify the primary causes of aircraft delays, this research analyzed airline data related to delays. Additionally, machine learning ensemble models were explored to predict future delays. The findings indicate that the originating airport holds the highest significance, closely followed by the choice of airline, in determining the likelihood of a delay occurrence.

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 PerformanceComputing 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 Datadriven Machine Learning Approach.” 2020 IEEE International Conference on Industrial Engineering and Engineer- ing Management (IEEM). IEEE, 2020.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)