USING MACHINE LEARNED CLASSIFIERS TO PREDICT FLIGHT DELAYS WITH ERROR CALCULATION

G. Mahesh[1], P.Jagadeesh[2], G.Navyu[3], Sk. Imtiyaj Babu[4], M. David Raju[5]
[1] *Asst. Professor, Krishna Chaitanya Institute of Technology & Sciences , Markapur, A.P, India*
[2,3,4,5] *Scholar, Krishna Chaitanya Institute of Technology & Sciences , Markapur, India*
**E-mail: maheshlucky13@gmail.com*

ABSTRACT:

A significant issue in the aviation industry is flight delays. The expansion of the aviation industry during the past two decades has increased air traffic, which has delayed flights. Not only do flight delays cost money, but they also have a bad effect on the environment. Airlines that operate commercial flights suffer huge losses as a result of flight delays. In order to minimise or avoid flight delays and cancellations, they thus take all reasonable precautions. In this research, we forecast whether a certain flight's arrival will be delayed or not using machine learning models including Logistic Regression, Decision Tree Regression, Bayesian Ridge, Random Forest Regression, and Gradient Boosting Regression.

Keywords: Logistic Regression, Decision Tree Regression, Bayesian Ridge, Random Forest Regression and Gradient Boosting Regression.

[1] INTRODUCTION

A mathematical technique for generating approximations from raw data is statistical modelling. Then forecasts are made using these approximations. Based on historical statistical data, statistical models can assist forecast the probability behaviour of a system in the future. Numerous sectors have employed predictive modelling, such as criminal investigations to identify the likelihood of spam emails and aircraft delays. Regression models have been found effective in predicting flight delays because they highlighted the various causes of flight delays, according to an evaluation of how well different models perform in this area. They were unable to classify complicated data, though. Econometric models have been used to simulate the cancellation of planned flights and to demonstrate how delays at one airport spread to other locationsSince they didn't take into account elements that were hard to quantify, these models didn't offer a full defence. The models produced discriminatory and arbitrary conclusions when they were applied to social and economic circumstances. Random forest has been discovered to perform better than the other models. The accuracy of the prediction may vary depending on variables like the forecast time and airline dynamics. According to a created multiple regression model, the length of the flight, the time of day, and the scheduled departure are the main predictors of delay.

The model did highlight the important aspects, but its prediction accuracy was subpar. Additionally, the model is restricted to a single flight path.Fourier fit model was shown to be very accurate in predicting flight delays when compared to other models, such as the K-means clustering algorithm. The two models were discovered to be appropriate for a single airport, but not for the prediction of several airports. The normal distribution and the Poisson distribution are two examples of probability models that have been used to simulate aeroplane arrival and departure delays. However, based on factors like time length and the number of airports taken into account, the prediction's accuracy changed. It was found that the Poisson distribution was better at modelling flight arrival delays than the normal distribution was at modelling flight departure delays. These models, nevertheless, are parametric and presuppose that the reaction will have a specific functional shape.The final model will not fit the data well and its estimations will be subpar if the training data set does not conform to this shape. Flight on-time performance has been modelled using a logistic regression model. With both the training and testing data sets, the model performed well.

The model's variance was likewise minimal. However, if the training data set does not conform to the predicted functional form, its parametric character may be a drawback. In the emergency room, neural networks outperformed the logistic regression model in predicting mortality in patients with suspected sepsis.This was ascribed to the neural networks' capacity to fit non-linear relationships between dependent and independent variables and its few characteristics that needed to be confirmed before model development. It was determined that the Support Vector Machine (SVM) model was fitted and appropriately allocated the training data set. SVM outperformed multiple linear regression and back propagation neural network in predicting the auto-ignition temperatures of organic substances. Models of delay innovation have been created using random forests. The study's findings suggested that, up to a certain crucial value, more decision trees were better. Random forest outperformed decision tree in the computational toxicology prediction of new vehicle approach, according to the data. SVM and random forests are under the category of machine learning.

The training data for machine learning is split up into many samples. A model is fitted and evaluated against the testing data set for each sample. A plot of the train errors and the test errors against the sample size reveals the sample that produces the best model. The non-parametric feature of the SVM and the random forest, in which they do not presume a certain functional form of the answer under consideration, is their main benefit. As a result, they can suit a greater variety of response forms, making them exceedingly adaptable. There are no modelling studies on flight delays available for the Kenyan aviation sector. This research compares the accuracy of models that have been used to forecast aircraft delays at Nairobi's Jomo Kenyatta International Airport. Detailed information about flights at Jomo Kenyatta International Airport that was collected from Kenya Airports Authority.

The information related to the 2017–2018 fiscal year, which began in March 2017 and ended in March 2018. In addition to the day of the flight (Monday through Sunday), the month (January to December), the airline, the flight class (domestic or international), the season (summer (March through October) or winter (October through March), the aircraft's capacity, the flight ID (tail number), and whether the flight had flown at night or during the day were among the variables used.R-Score statistical analysis software was used to examine the data. The time discrepancy between the planned time and the exact time for planes was computed. A delay was defined as a time difference of more than 15 minutes, and was assigned the value 1, whereas a non-delay was defined as a time difference of less than 15 minutes, and was assigned the value 0. Machine learning was used to fit the three models: the Random Forest, the SVM, and the logistic regression. A training data set of 15,000 flights and a testing data set of 5,000 flights were created from the whole data set.

[2] LITERATURE SURVEY

An airport maneuvering area is an example of a very complex system that may be investigated by using a two-stage method based on real-time and quick simulation techniques. The baseline model used to identify the congestion locations is analyzed using quick and real-time simulations in the first step. Improvements to the design of the maneuvering area are suggested based on the analysis. The second stage involves the generation and evaluation of alternative scenarios that incorporate these upgrades in a fast-time simulation environment. The primary congested locations in the baseline airport model are identified using the results of simulations of various runway layouts. Both the departure queue points and the taxiway system have congestion nodes.Three alternative models, including reconfigured taxiways and fast-exit taxiways, are examined using the fast-time simulation approach to reduce congestion at these spots. For additional testing in real-time simulations, the alternate solution that performed the best in these tests is chosen. It is demonstrated that the approach would lead to a rise in hourly operations and a sharp drop in overall ground delays.

Simulation approaches save cost and time when undertaking the investigations required to identify congestion and design changes. It is demonstrated that it is required to also test the outcomes of the fast-time simulations in real-time simulations even if fast-time simulations are often sufficient for identifying solutions when critical settings for the airport are taken into account.The models neglect the impacts of weather phenomena like rain, fog, snow, etc. The runways being utilised have a big impact on ground movements in manoeuvring zones. Therefore, three different runway usage scenarios are explored in the study to provide a thorough evaluation. At order to pinpoint the regions of congestion in the large-scale airport manoeuvring zones and come up with methods to reduce the congestion, this study combines fast- and real-time simulation approaches. This strategy aims to lessen the drawbacks of both strategies while combining their benefits. There is no study in the literature that combines these two methods to analyse the capacity of airport manoeuvring areas.

The fundamental goal of the suggested work is to analyse flight arrival delay using data mining and four supervised machine learning algorithms: random forest, Support Vector Machine (SVM), Gradient Boosting Classifier (GBC), and k-nearest neighbour algorithm, and compare their performances to find the best performing classifier. Data from BTS and the US Department of Transportation was gathered to train each predictive model. The information included all of American Airlines' flights between the top five busiest airports in the United States—Atlanta, Los Angeles, Chicago, Dallas/Fort Worth, and New York—during the years 2015 and 2016. To forecast the arrival delay of specific scheduled flights, the aforementioned supervised machine learning algorithms were assessedTo properly determine if a certain aircraft will be delayed by more than 15 minutes or not, all the algorithms were utilised to develop the prediction models and compared to one another. As a consequence, when compared to KNN, SVM, and random forest, the gradient boosting classifier performs the best in terms of predicting arrival delay for 79.7% of all booked American Airlines flights. A GBC-based prediction model like this one has the potential to prevent significant costs for commercial airlines, who suffer from planned aircraft arrival delays.

The main objective of the model put forward in this research is to forecast aircraft delays brought on by bad weather using supervised machine learning and data mining techniques. In order to train the model, meteorological data from 2005 to 2015 was combined with domestic US flight data. Utilizing sampling strategies, the consequences of unbalanced training data are mitigated. The AdaBoost, k-Nearest-Neighbors, decision trees, random forests, and the AdaBoost were used to develop models that can forecast flight delays individually. Then, the receiver operating characteristic (ROC) curve and the prediction accuracy of each method were contrasted. Flight information and the weather forecast were acquired and included into the model during the prediction stage.Usingthosedata,thetrainedmodelperformedabinary classification topredictedwhether a scheduledflight will be delayed oron-time.

Flight delays are bad for travellers, airports, and airlines. All participants in commercial aviation must consider their forecast while making decisions. Additionally, the complexity of air travel made it difficult to create precise prediction models for flight delays.

Increasing client happiness is a key component of the airline company. Flights are delayed and cause consumer displeasure due to inclement weather, a mechanical issue, and the delayed arrival of the aircraft at the place of departure. With the use of weather and flight data, a prediction model for flights arriving on time is put forth. The investigation of the relationship between flight data and weather data is the main study topic in this work. It is discovered that the sea-level pressures of three weather observation spots, namely Wakkanai as the most northern spot, Minami-Torishima as the most eastern spot, and Yonagunijima as the most western spot, can classify the relationship between pressure pattern and flight data of Peach Aviation, an LCC (low-cost carrier) in

Japanpressurepatterns.As a consequence, utilising the Random Forest Classifier of machine learning, on-time arrival flight is predicted with a 77 percent accuracy. Additionally, a tool for predicting on-time arrival of flights is being developed to assess the viability of the predictive model.

[3] SYSTEM ARCHITECTURE

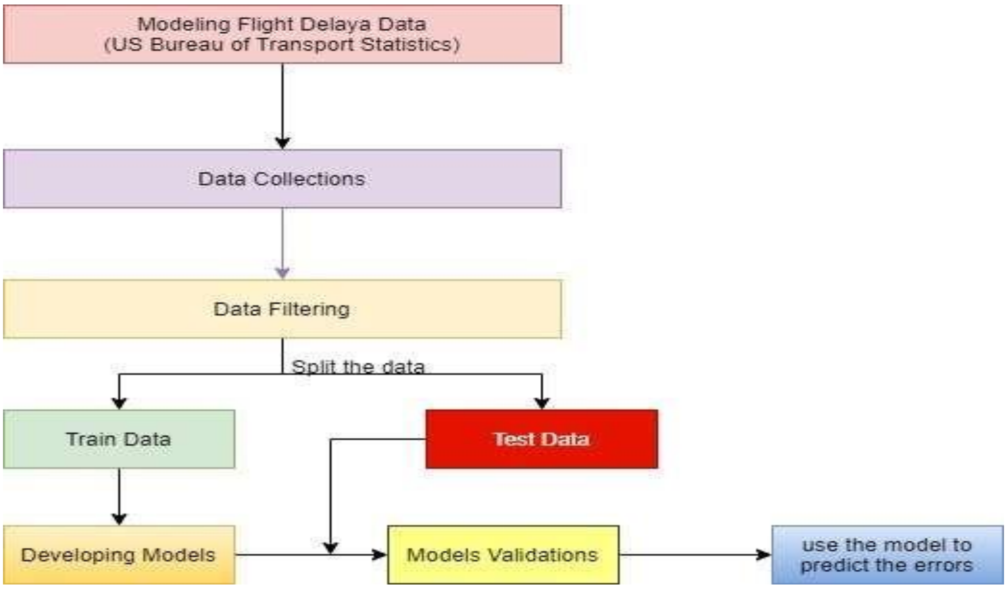


Fig.1 System Architecture of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation

[4] IMPLEMENTATION

4.1 Modules Description

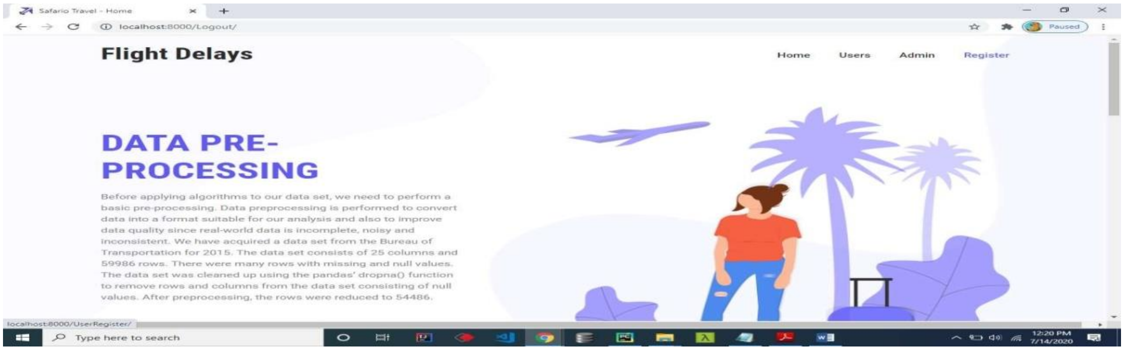
i) User: The user can sign up initially. He needed a working user email and cellphone upon registration for more conversations. After registering, the user can be activated by the admin. Once the user has been activated by the admin, they may log in to our system. The US Bureau of Transportation dataset is not directly processed. The data has to be cleaned before the procedure. Once the data has been cleaned, the user may evaluate the departure delay performance using the models they have chosen. The user's browser displays the findings. Both a visual depiction and all mistake scores can be shown.

ii) Admin: With his login information, Admin may log in. He can activate the users after logging in. Only our applications allow the enabled user to log in. We have researched a variety of sources to determine which variables will be most useful in predicting departure and arrival delays. We get to the conclusion that the dataset's parameters are Day, Departure Delay, Airline, Flight Number, Destination Airport, Origin Airport, Day of Week, and Taxi out after conducting a number of searches. We thus take this data into account for the next step.

iii)Data Preprocess: The SQL Lite database has been used to hold the data that the admin gave. Our technique requires data cleansing, which must be done. We may fill in the missing numbers with the mean type using a pandas data frame. The data will be displayed on the browser once it has been cleared.

iv)**Model Execution:** With the use of machine learning models like gradient boosting regression, logistic regression, decision tree regression, bayesian ridge, random forest regression, and Bayesian ridge, we can predict the outcome. The MSE is suitable for our regression issues since it is differentiable, which adds to the algorithmic stability. Additionally, it severely penalises larger errors relative to minor errors. An indicator of risk, MAE provides the predicted value of the absolute error loss. This method measures the Explained Variance Score, or the degree to which our machine learning model explains the scattering of the dataset. R2 Rating This statistic assesses the likelihood that the model will be able to predict unknown samples by the proportion of explained variance. It indicates the quality of fit.Thebestscorecanbe 1.0and thescorecan also be negative.

4.2 Screenshots

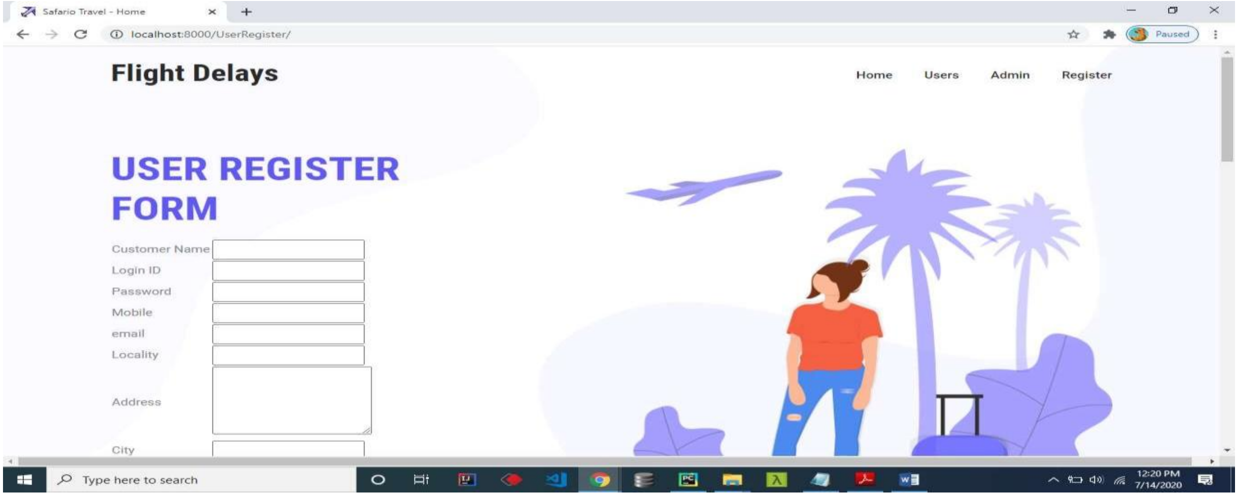


Flight Delays

DATA PRE-PROCESSING

Before applying algorithms to our data set, we need to perform a basic pre-processing. Data preprocessing is performed to convert data into a format suitable for our analysis and also to improve data quality since real-world data is incomplete, noisy and inconsistent. We have acquired a data set from the Bureau of Transportation for 2015. The data set consists of 25 columns and 59986 rows. There were many rows with missing and null values. The data set was cleaned up using the pandas' dropna() function to remove rows and columns from the data set consisting of null values. After preprocessing, the rows were reduced to 54486.

Fig. 2 Home Page of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation

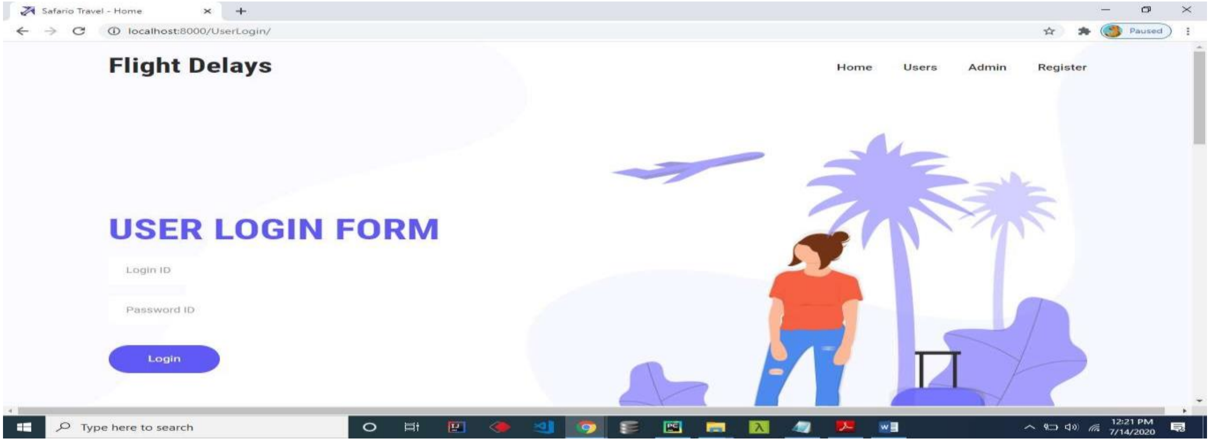


Flight Delays

USER REGISTER FORM

Customer Name
Login ID
Password
Mobile
email
Locality
Address
City

Fig. 3 Register Form of Using Machine Learned Classifiers to Predict Flight Delays with ErrorCalculation



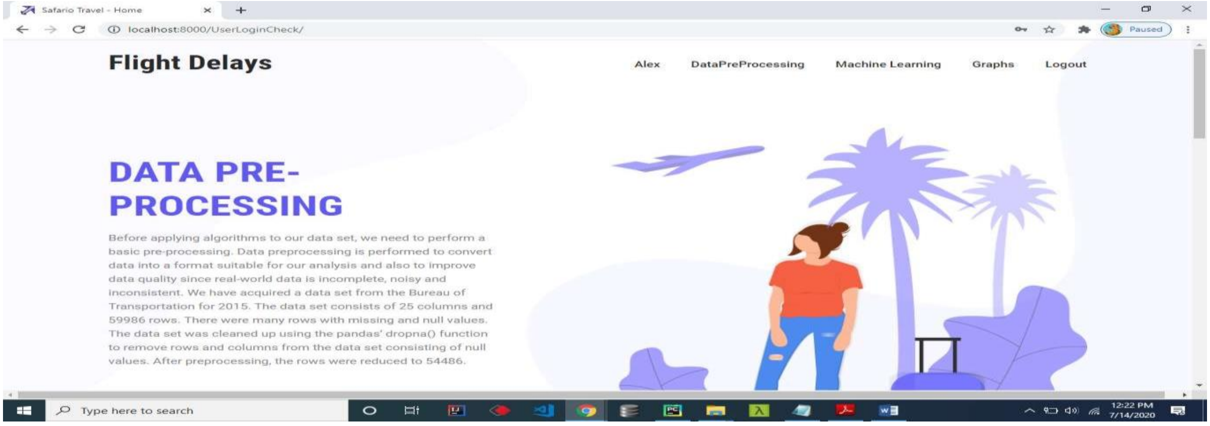
Flight Delays

USER LOGIN FORM

Login ID

Password ID

Fig. 4 User Login Form of Using Machine Learned Classifiers to Predict Flight Delays withErrorCalculation

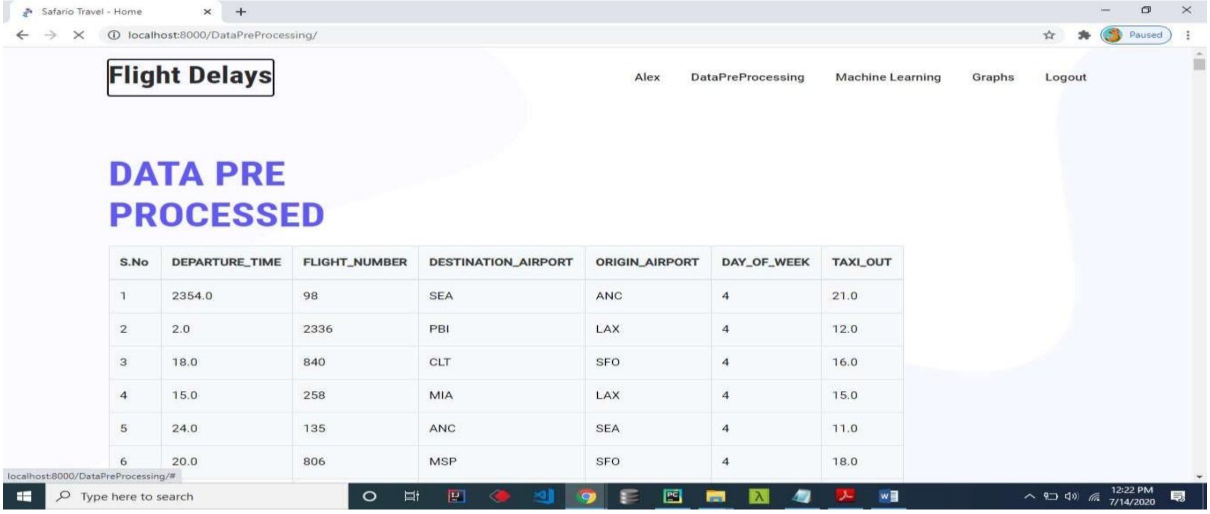


Flight Delays

DATA PRE-PROCESSING

Before applying algorithms to our data set, we need to perform a basic pre-processing. Data preprocessing is performed to convert data into a format suitable for our analysis and also to improve data quality since real-world data is incomplete, noisy and inconsistent. We have acquired a data set from the Bureau of Transportation for 2015. The data set consists of 25 columns and 59986 rows. There were many rows with missing and null values. The data set was cleaned up using the pandas' dropna() function to remove rows and columns from the data set consisting of null values. After preprocessing, the rows were reduced to 54486.

Fig.5 User Home Page of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation

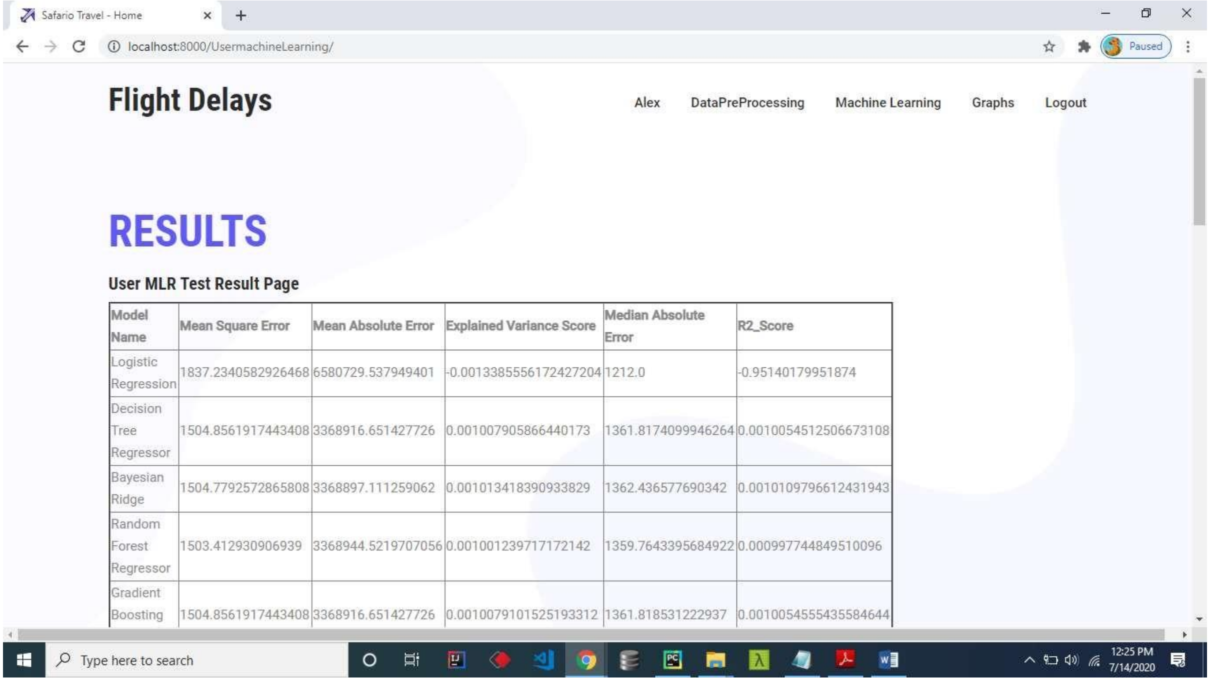


Flight Delays

DATA PRE PROCESSED

S.No	DEPARTURE_TIME	FLIGHT_NUMBER	DESTINATION_AIRPORT	ORIGIN_AIRPORT	DAY_OF_WEEK	TAXI_OUT
1	2354.0	98	SEA	ANC	4	21.0
2	2.0	2336	PBI	LAX	4	12.0
3	18.0	840	CLT	SFO	4	16.0
4	15.0	258	MIA	LAX	4	15.0
5	24.0	135	ANC	SEA	4	11.0
6	20.0	806	MSP	SFO	4	18.0

Fig. 6 : Preprocessed Data of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation



Flight Delays

RESULTS

User MLR Test Result Page

Model Name	Mean Square Error	Mean Absolute Error	Explained Variance Score	Median Absolute Error	R2_Score
Logistic Regression	1837.2340582926468	6580729.537949401	-0.0013385556172427204	1212.0	-0.95140179951874
Decision Tree Regressor	1504.8561917443408	3368916.651427726	0.001007905866440173	1361.8174099946264	0.0010054512506673108
Bayesian Ridge	1504.7792572865808	3368897.111259062	0.001013418390933829	1362.436577690342	0.0010109796612431943
Random Forest Regressor	1503.412930906939	3368944.5219707056	0.001001239717172142	1359.7643395684922	0.000997744849510096
Gradient Boosting	1504.8561917443408	3368916.651427726	0.0010079101525193312	1361.818531222937	0.0010054555435584644

Fig. 7: Algorithm Codes of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation

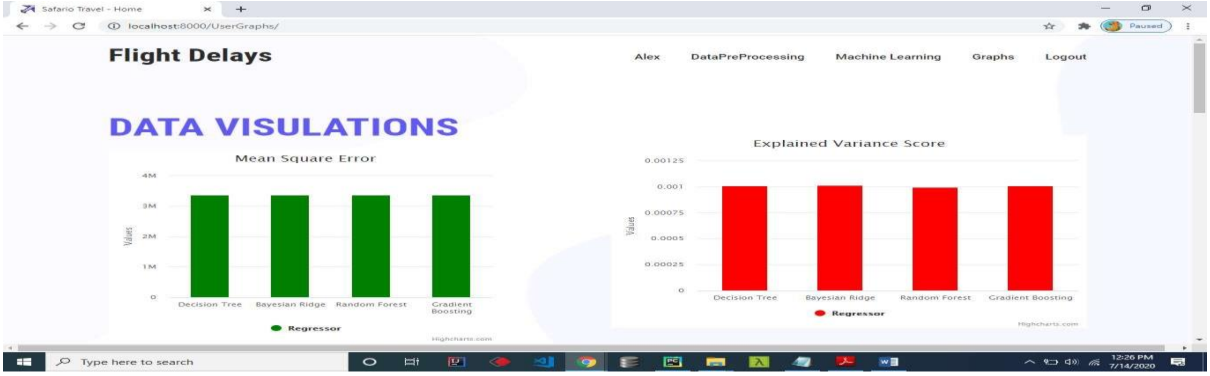


Fig. 8 : User side Graphs of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation Graph

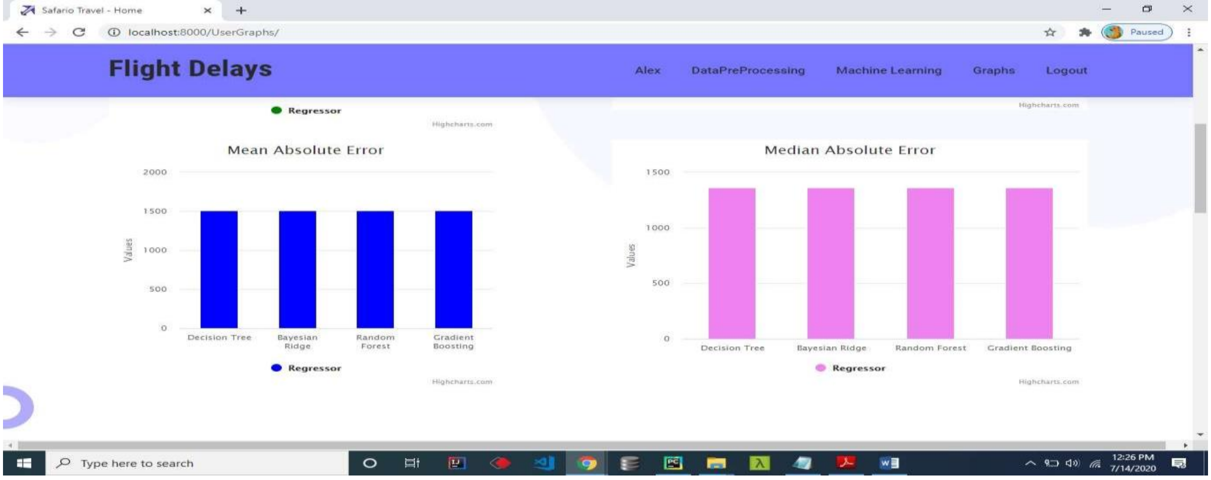


Fig. 9 : Graph of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation

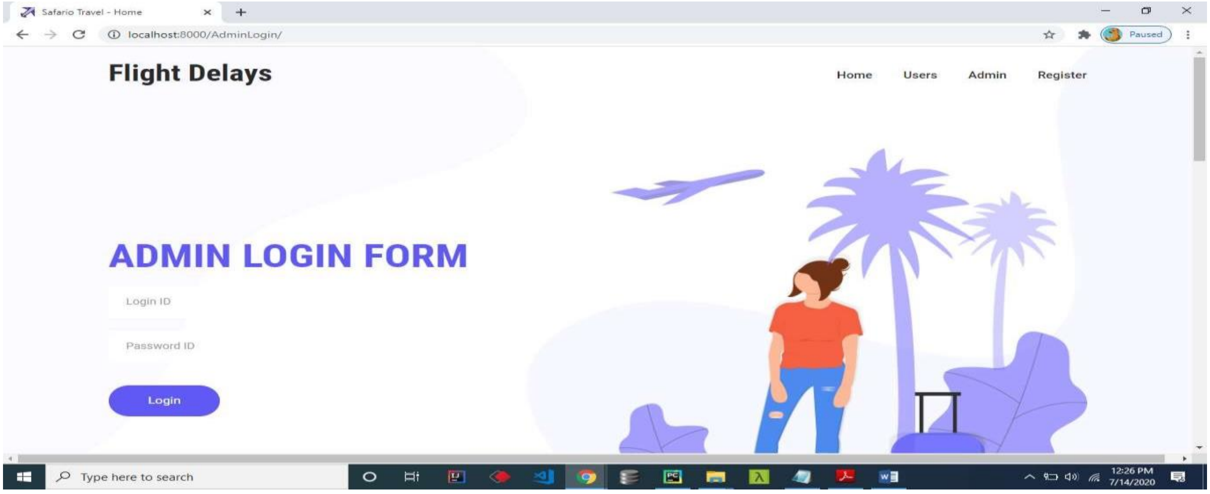
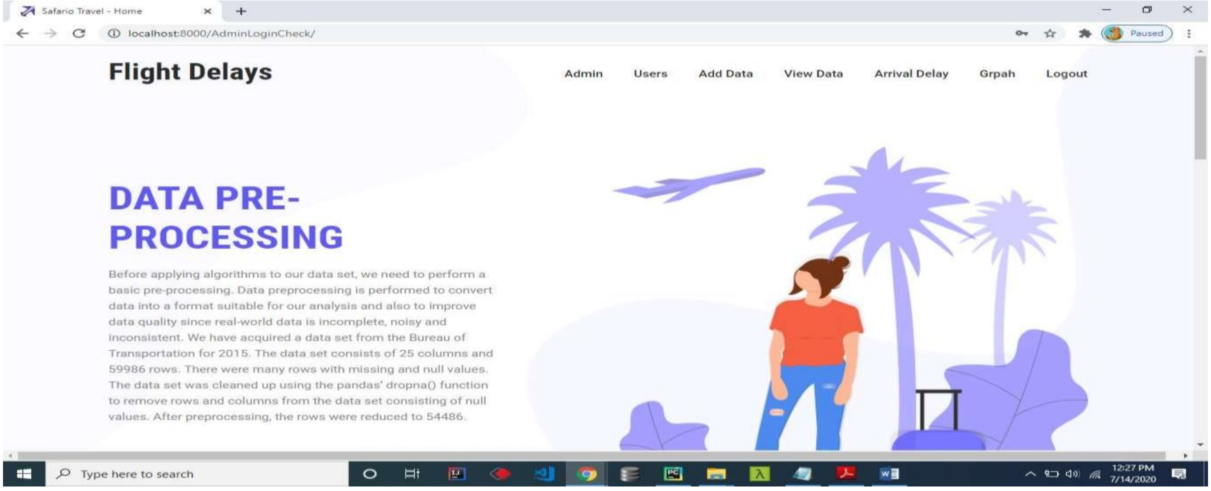


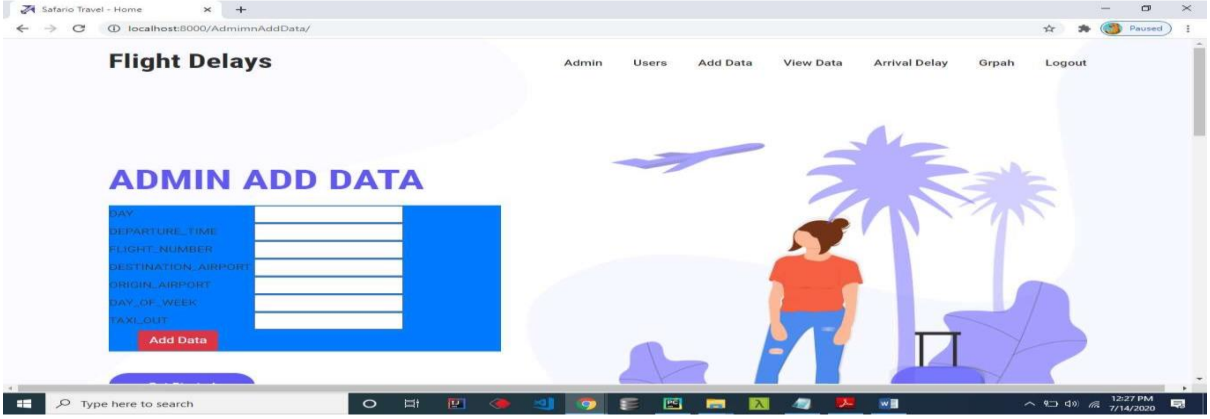
Fig. 10 : Admin Login Page of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation



DATA PRE-PROCESSING

Before applying algorithms to our data set, we need to perform a basic pre-processing. Data preprocessing is performed to convert data into a format suitable for our analysis and also to improve data quality since real-world data is incomplete, noisy and inconsistent. We have acquired a data set from the Bureau of Transportation for 2015. The data set consists of 25 columns and 59986 rows. There were many rows with missing and null values. The data set was cleaned up using the pandas' dropna() function to remove rows and columns from the data set consisting of null values. After preprocessing, the rows were reduced to 54486.

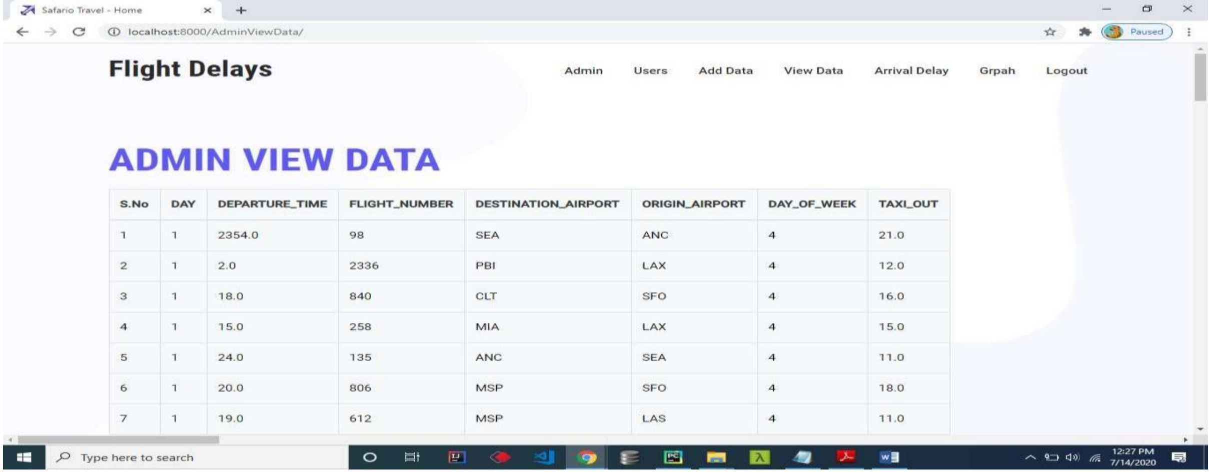
Fig. 11 Admin Home Page of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation



ADMIN ADD DATA

DAY
DEPARTURE_TIME
FLIGHT_NUMBER
DESTINATION_AIRPORT
ORIGIN_AIRPORT
DAY_OF_WEEK
TAXI_OUT

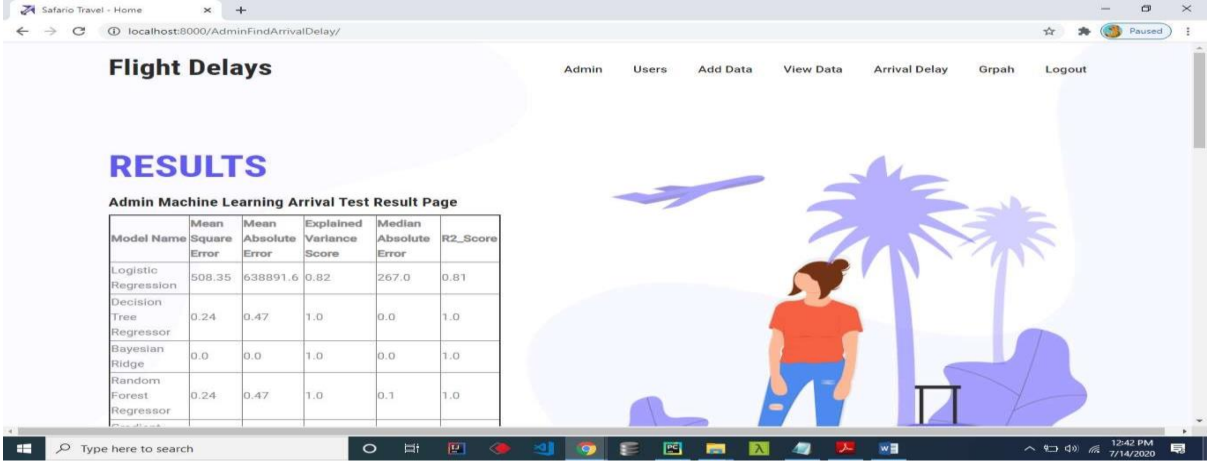
Fig. 12 :Admin Adding Data of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation



ADMIN VIEW DATA

S.No	DAY	DEPARTURE_TIME	FLIGHT_NUMBER	DESTINATION_AIRPORT	ORIGIN_AIRPORT	DAY_OF_WEEK	TAXI_OUT
1	1	2354.0	98	SEA	ANC	4	21.0
2	1	2.0	2336	PBI	LAX	4	12.0
3	1	18.0	840	CLT	SFO	4	16.0
4	1	15.0	258	MIA	LAX	4	15.0
5	1	24.0	135	ANC	SEA	4	11.0
6	1	20.0	806	MSP	SFO	4	18.0
7	1	19.0	612	MSP	LAS	4	11.0

Fig. 13 View Data of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation



Admin Machine Learning Arrival Test Result Page

Model Name	Mean Square Error	Mean Absolute Error	Explained Variance Score	Median Absolute Error	R2_Score
Logistic Regression	508.35	638891.6	0.82	267.0	0.81
Decision Tree Regressor	0.24	0.47	1.0	0.0	1.0
Bayesian Ridge	0.0	0.0	1.0	0.0	1.0
Random Forest Regressor	0.24	0.47	1.0	0.1	1.0

Fig. 14 Admin View Results of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation

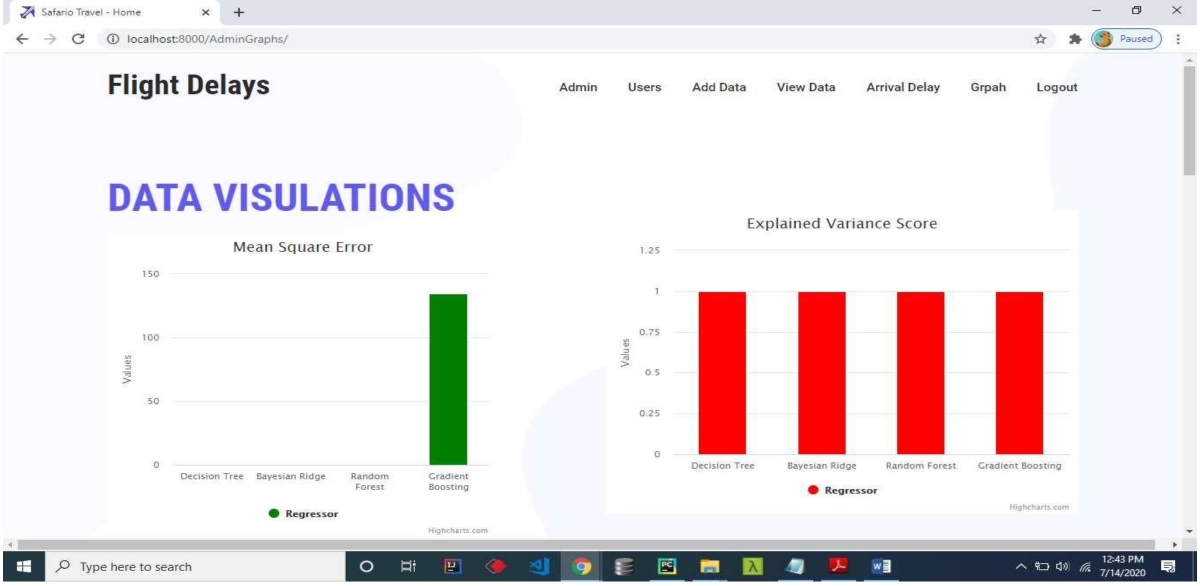


Fig. 15 Arrival Graph of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation

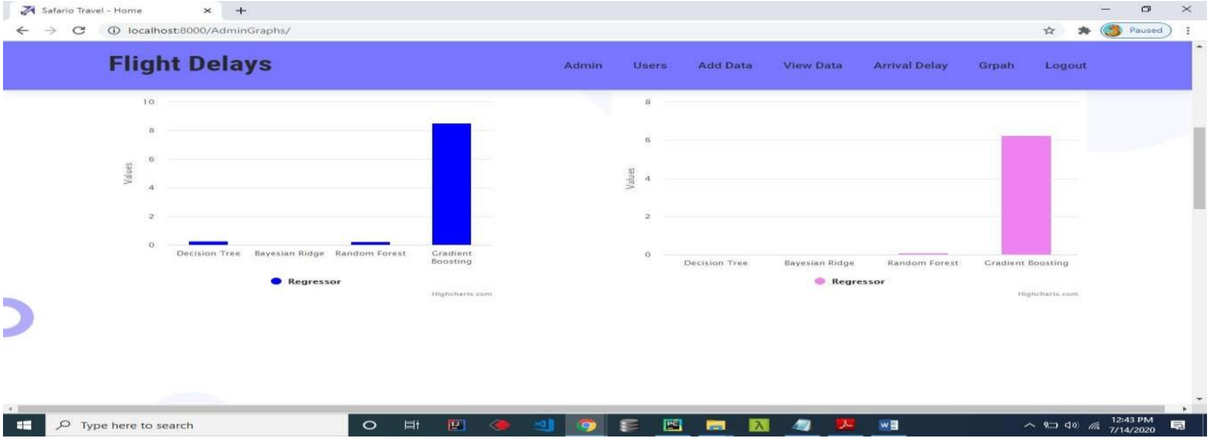
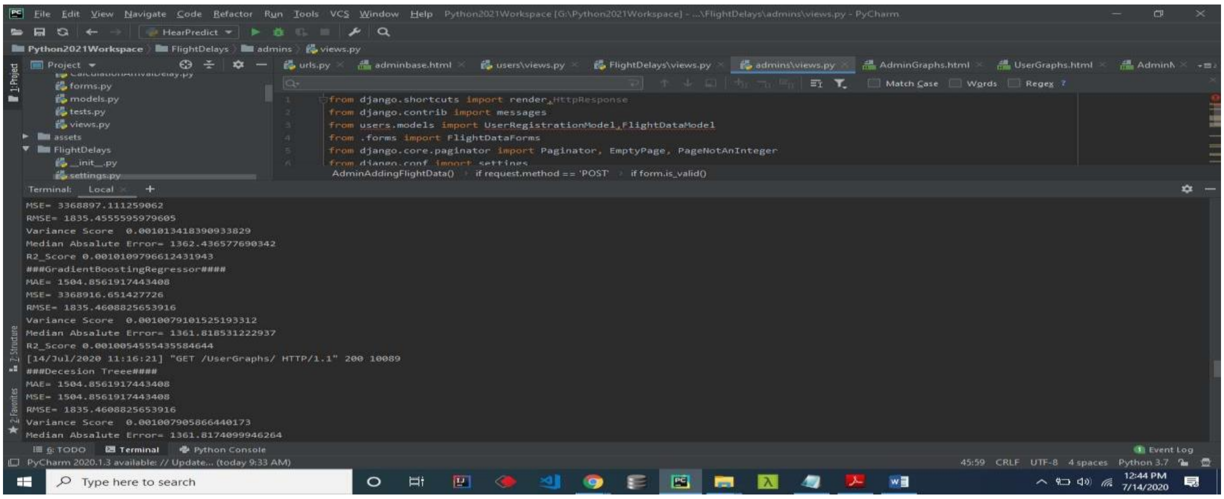


Fig. 16 Arrival Graph of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation



```
from django.shortcuts import render,HttpResponse
from django.contrib import messages
from users.models import UserRegistrationModel,FlightDataModel
from .forms import FlightDataForms
from django.core.paginator import Paginator, EmptyPage, PageNotAnInteger
from django.conf import settings
```

```
MSE= 3368897.111259062
RMSE= 1835.4555595979605
Variance Score  0.001013418390933829
Median Absalute Error= 1362.436577690342
R2_Score 0.0010109796612431943
###GradientBoostingRegressor####
MAE= 1504.8561917443408
MSE= 3368916.651427726
RMSE= 1835.4608825653916
Variance Score  0.0010079101525193312
Median Absalute Error= 1361.818531222937
R2_Score 0.0010054555435584644
[14/Jul/2020 11:16:21] "GET /UserGraphs/ HTTP/1.1" 200 10089
###Decesion Treee####
MAE= 1504.8561917443408
MSE= 1504.8561917443408
RMSE= 1835.4608825653916
Variance Score  0.001007905866440173
Median Absalute Error= 1361.8174099946264
```

Fig. 17 Server Side Results of Using Machine Learned Classifiers to Predict Flight Delays with Error Calculation

[5] CONCLUSION

To anticipate aero plane arrival and delay, machine learning techniques were employed progressively and consecutively. We used this to create five models. We saw that the values of the models were taken into account and compared for each assessment metric. We discovered that: - In terms of departure delay, the Random Forest Regressor was determined to be the best model, with Mean Squared Error of 2261.8 and Mean Absolute Error of 24.1—both of which are the lowest values recorded in their respective metrics. The best model for Arrival Delay was the Random Forest Regressor, which had Mean Squared Error 3019.3 and Mean Absolute Error 30.8—the lowest values for each of these measures.Although the Random Forest Regressor's error number is not the smallest in the other measures, it still provides a low value in comparison. We discovered that the Random Forest Regressor provides the best value in terms of maximum metrics and ought to be the model chosen.

The implementation of more sophisticated, contemporary, and cutting-edge preprocessing approaches, automated hybrid learning and sampling algorithms, and deep learning models modified to achieve greater performance can all fall under the future purview of this study. Additional factors can be included to help a prediction model develop. For instance, one model uses meteorological information to create error-free models for aircraft delays. Since we only utilized US data in this work, the model may now be trained using data from other nations as well. More accurate predictive models may be built with the use of models that are complex and hybrids of many different models when given the necessary processing capacity, as well as with the use of larger, more detailed datasets.Additionally, the model can be set up to forecast flight delays at other airports, therefore data from such airports would need to be incorporated into this study.

REFERENCES

[1] N. G. Rupp, "Further Investigation into the Causes of Flight Delays," in Department of Economics, East Carolina University, 2007.

[2] "Bureau of Transportation Statistics (BTS) Databases and Statistics," [Online]. Available:http://www.transtats.bts.gov.

[3] Airports Council International,World Airport Traffic Report,"2015,2016.

[4] E. Cinar, F. Aybek, A. Caycar, C. Cetek, "Capacity and delay analysis for airport mano euvringareas using simulation," Aircraft Engineering and Aerospace Technology, vol. 86, no. No. 1,pp. 43-55, 2013.

[5] Navoneel, et al., Chakrabarty, "Flight Arrival Delay Prediction Using Gradient BoostingClassifier," inEmergingTechnologiesinDataMiningand InformationSecurity,Singapore,2019.

[6] Y. J. Kim, S. Briceno, D. Mavris, Sun Choi, "Prediction of weather induced airline delays based on machine learning algorithms," in35th DigitalAvionics Systems Conference(DASC), 2016.

[7] W.-d. Cao. a. X.-y. Lin, "Flight turnaround time analysis and delay prediction based on Bayesian Network, "Computer Engineering and Design, vol. 5, pp. 1770-1772, 2011.

[8] J.J.Robollo, Hamsa,Balakrishnan, "Characterization and Prediction of Air Traffic Delays".

[9] S.Sharma, H. Sangoi, R. Raut, V. C. Kotak, S. Oza,"FlightDelay Prediction System Using Weighted Multiple Linear Regression," International Journal of Engineering and Computer Science,vol.4, no. 4, pp. 11668 - 11677, April2015.

[10] A. M. Kalliguddi, Area K., Leboulluec, "Predictive Modelling of Aircraft Flight Delay, "Universal Journal of Management, pp. 485 - 491,2017.

[11] Noriko,Etani, "Development of apredictive model foront time arrival fligh to fair liner by Discovering correlartion between flight andweather data",2019.

[12] [Online]. Available:https://towardsdatascience.com/metrics-toevaluate-your-machine-learning-algorithm-f10ba6e38234.

[13] GNR Prasad, SK Althaf Hussain Basha, Mallikharjuna Rao K M GnanaVardhan "A Review of Predictive And Descriptive Data Mining Techniques In Higher Education Domain, International Journal of Computer Engineering and Applications(IJCEA),Volume 13, Issue 6, January. 21, ISSN2321-3469.

[14] B Sasidhar, Sk Althaf Hussain Basha , " A Comparative Study of Educational Data Mining Methods Used to Forecast Student Success and Failures", International Journal Computer Science Information and Engineering Technologies (IJCSIET), International Conference 2014,ISSN:2277-4408,2014.

[15] Ch. Prakash, Sk Althaf Hussain Basha, D. Mounika, G. Maheetha, "An Approach for Multi Instance Clustering of Student Academic Performance in Education Domain", IIJDWM Journal, Volume 3,Issue 1,pp.1-9,Feb.2013,ISSN: 2249-7161.

[16] Sd.Muneer , Sk Althaf Hussain Basha, A.Govardhan, V.Uday Kumar " Generate Eligible Students using Decision Trees- A Frame work for Employee Ability" International Journal of Advanced Computing(IJAC), Volume 4,Issue 2,2012,pp.68-76, ISSN:0975-7686.

[17] Mohd. Zaheer Ahmed , Sk Althaf Hussain Basha, A.Govardhan ,Y.R.Ramesh Kumar , "Predicting Student Academic Performance Using Temporal Association Mining" International Journal of Information Systems and Education (IJISE), Vol.2, No.1(2012),pp.21-41,ISSN: 2231- 1262.

[18] A.Govardhan , SK Althaf Hussain Basha, Y.R.Ramesh Kumar , Mohd. Zaheer Ahmed, "Study of Education Patterns Using Association Mining" International Journal Data Warehousing (IJDW), Vol.3 ,No.2,2011, pp. 53-64, ISSN: 0975-6124.

[19] SK Althaf Hussain Basha, A.Govardhan, "MICR: Multiple Instance Cluster Regression for Student Academic Performance in Higher Education", International Journal of Computer Applications(IJCA), Volume 14– No.4,2011,pp.23-29, ISSN: 0975-8887 (Impact Factor : 0.8

[20] SK Althaf Hussain Basha, A. Govardhan "A Comparative Analysis of Prediction Techniques for Predicting Graduate Rate of University", European Journal of Scientific Research (EJSR) ,Vol.46 No.2,2010, pp.186-193, ISSN No:1450-216X . (Impact Factor0.783, Citations: 12)

[21] Sk. Althaf Hussain Basha, A.Govardhan "Rank Analysis Through Polyanalyst using Linear Regression" , International Journal of Computer Science and Network Security(IJCSNS), VOL.9 No.9,2009, pp. 290-293, ISSN: 1738-7906. (Impact Factor:2.512, Citations:7)

[22] T Naveen Kumar, SK Althaf Hussain Basha, V. Anand , DonapatiSrikanth, "Categorization of Academic Student Performance using Hybrid Techniques" International Conference on Advanced Computing Methodologies (ICACM-2013), Hyderabad, pp.325- 330,2013.

[23] Y. Vijayalata, Sk Althaf Hussain Basha, V. Anand ,Donapati Srikanth, " Study of Education patterns using Rare Association Mining-A case Study " , IEEE International Conference on Engineering for Humanity (ICEH-2013), Hyderabad, pp. 53-61,2013,ISSN: 978-93-82880- 53-0.

[24] Y Ramesh Kumar, Sk Althaf Hussain Basha, Y Vijayalata, " Predicting Student Academic Performance using Temporal Association Mining-A case Study on Educational Data " , IEEE International Conference on Engineering for Humanity (ICEH-2013), Hyderabad, pp. 21- 27,2013, ISSN:978-93-82880-53-0.

[25] B Sashidhar, SK Althaf Hussain Basha, Y R Ramesh Kumar , A Govardhan, "A Case Study: Data Mining and Data Modelling Techniques Applied to Student Enrollment", National Conference on Data Modeling, Image Analysis Pattern Recognition (DMIAPR) 2011 at GITAM Institute of Technology, GITAM University, Vizag.

[26] N Kartiek, Sk Althaf Hussain Basha, "Forecasting the Academic Results of Students using Artificial Neural Networks", National Conference in Modern. Trends in Computer Science and Technology (NCMTCSCT2013), ECET, Hyderabad,2013,ISSN:978-162776537-4.

[27] Y R Ramesh Kumar, SK Althaf Hussain Basha, A Govardhan, B.Sasidhar, " Mining Educational Data to Analyze Academic Student's Performance", National Conference in Modern.TrendsinComputerScienceandTechnology(NCMTCSCT2013),ECET,Hyderabad,201 3,ISSN:978-162776537-4.

[28] D.Mounika, Sk Althaf Hussain Basha, Y.R. Ramesh Kumar, Y Vijayalatha, A. Govardhan, " Study of Education Patterns using Rare Association Mining", National Conference on Emerging Trends of Computing Technologies, (NCECT2013), GRIET,Hyderabad,pp.93- 103,2013.

[29] C. J. Willmott, Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square (RMSE) in assessing average model performance," Climate Research, vol. 30, no. 1, pp. 79 -82, 2005.