**­­­­­­**

­­

AMITY UNIVERSITY UTTAR PRADESH

on

**Analysis of Machine learning techniques**

**for Airfare prediction**

Submitted to

## Amity Institute of Information Technology

**In partial fulfilment of the requirements for the award of the degree of**

Bachelor of Computer Applications

# Submitted To: Submitted By:

Dr. Deepa Gupta Jaskirat Singh

(Assistant Professor) A1049520020

Batch: 2020-2025

**Analysis of Machine learning techniques**

**for Airfare prediction**

Jaskirat Singh (Amity Institute of Information Technology, Amity University, Noida, India, [jaskirat1311@gmail.com](mailto:jaskirat1311@gmail.com) )

Deepa Gupta(Amity Institute of Information Technology, Amity University, Noida, India, [dgupta@amity.edu](mailto:dgupta@amity.edu) )

Lavanya Sharma(Amity Institute of Information Technology, Amity University, Noida, India, [shm.lavanya@gmail.com](mailto:shm.lavanya@gmail.com) )

**Abstract**

This paper contains that how data is processed in order to predict prices of things but in this report, we will learn about prediction of prices of airlines tickets is done from scratch. Along with the accuracy in predicting prices of each model, this paper contains how every accuracy of the derived features of the dataset dependent for prediction of prices of different airline tickets and deals with the computation of airline tickets using machine learning algorithms. For this process a dataset on previous scheduled airline tickets in a time from different Airline Companies is used. The given data is pre-processed by using data cleaning, data wrangling and different data science techniques to make the data ready to gain insights. To understand the algorithm or method used for predicting fares of previous airline tickets. To calculate the possible minimum price of the tickets, data for a certain air route with different flights and their details has been collected including the features like time of departure time of arrival and of which airways flight is etc. over a certain period of time to implement algorithms on it and extract useful features. To obtain different useful features from the data by applying different Machine Learning algorithms. It includes the machine learning algorithms which are used to predict the prices at the given time with the limitations they have [6].

**KEYWORDS:** Data Pre-processing, Prediction models, Airfare prices, Label encoding, Random Forests, Decision Tree, Linear Regression, KNN algorithm, Cross validation.

**1. INTRODUCTION**

Social media today is an important part for everyone nowadays. As a result, buying and selling stuffs online is too normal these days, people use applications and websites to purchase stuffs satisfying their needs by sitting at home using online payment methods and other methods with proper secure ways. Analysis of specific product ideas can inform companies the level of satisfaction a customer got with their services. The system of buying tickets is used to purchase flight tickets days before the flight is scheduled prebooking for the flights that are planned to make a trip between two destinations in which a buyer can select the date, time and a flight with stoppage or no stoppage. Some aviation lines don’t agree with the procedure because plane organization might change the prices depending on availability of tickets and demand for tickets. It helps in booking of tickets from anywhere through APIs and online procedures and payment to avoid last minute rush for tickets [1-5].

Most of the airlines use difficult ways and rules for managing revenue for execution of typical systems for estimation. Mostly customers try to buy the tickets in advance of the departure date so that they will not have to pay high prices of tickets as date comes closer and demands rises for the left tickets. But sometimes this results in spending more money than expected for the same seat. The ideal aim of the airways is to gain profit on the tickets whereas the customer searches for the minimum rate as per their budget. But this system confirms that person a ticket so that they could travel according to the date they have selected. At the point of purchasing tickets the middle upper-class people look for a cheaper and under budget tickets. The rate of airline tickets at less cost is consistently rising due to competition among different airlines [6,7,8]. Many websites have been introduced on internet for booking of tickets at cheaper cost. Using machine learning algorithms, they usually show different tickets by different airlines for same route based on the previous airfare prices.

The data collected is proper raw data with values more than one datatype but when the training of data will take place the machine learning algorithm needs the data in numerical type (i.e. integers) and the data is needed to be changed into the numerical format for reducing time complexity and for better accuracy from each algorithm we are going to apply that is why different IT professionals coordinate with each other like business analyst and data scientist or data analyst they put efforts and make the data useful for gaining insights so that the result we get can be used in many different prediction or analysis therefore step by step each professional performs according to their respective fields. So, this is not as simple as we think it is, it includes a lot of working on raw data, to understand data by methods such as data visualization, data modelling etc. Different algorithms will be applied, and we will get the predictions with their accuracy percentages and also the parameters required [9-16].

**2. Background**

It is not an easy task for a customer to buy a plane ticket at very low cost. These few tricks are being investigated to determine the time and date to get airline tickets with the lowest billing rate. Most of these frameworks use a state-of-the-art framework which are known as Machine Learning. To make the right decision before buying an airplane tickets Gini and Groves violated the Partial Least Square Regression (PLSR) modeling. Details are collected at key locations for booking travel experiences for 2019 in the months of March, April, May and June. Additional data were compiled and used to monitor the demonstration relationships of the final model. Janssen has developed an aspiring model using the Linear Quantile Blended Regression process for the San Francisco - New York study where daily flights are provided by www.infare.com. Two key points, for example, the number of departure days and whether the flight is over a week or a working day is considered a model promotion. The model accurately balances flights ahead of time from the date of flight. In any case, the model does not interfere with the wide range of time frame, closing the date of flight [17 - 20].

Wohlfarth introduced a model used to improve time while purchasing a ticket online ticketing based on major preparations known as macked point processors, information mining frameworks (strategy and collection) and countless test frameworks. The plot is proposed to transform a variety of recreational activities into integrated themes of a game program that can support the test of individual integration. This relevant topic is included in the discussion based on a close assessment of the behavior. The progress model measures the appropriate change schemes. A tree-based investigation was used to select the best planned collection and a short time later it was considered a disagreement on the movement model. A study conducted by Dominguez-Menchero suggests that the optimal purchase time depends on a nonparametric isotonic strategy for rebellion in a particular course, senders, and duration. The model offers the most satisfying number of days before buying a plane ticket. The model looks at two types of variables such as travel and acquisition date [22 - 27].

**3. LOOKING INTO THE DATA BY THE TEAM**

It is a piece of work of more than just one IT professional. A proper cycle works for the thing we just do by entering our details in minutes. Many data science field IT professionals collaborate with each other to gain insights from the data. First a dataset of previous bookings of tickets of different airlines is collected. Which consist of features like timing, destination, any stoppage or not, airline, dates etc. Then a business analyst will look into the data and some insights. Then a data scientist / analyst will collaborate with the business analyst to decide from where to extract the data [like from 3rd party APIs, databases etc.]. Then the Data Scientist have the duty to check if the data is in right format or not because before applying machine learning algorithms the data should be in right format so as to make the process less complex and more precise. To make it ready for machine learning algorithms data pre-processing, data wrangling, data cleaning, feature encoding and other methods are applied on the raw dataset. For these procedures mostly two programming languages are used Python and R. But in this report, I used python and its libraries to perform every step of converting raw data into a right format for analysis and after all the mentioned steps we will apply different algorithms to the model [21-27].

**4. DATA COLLECTION & PRE-PROCESSING**

**4.1. DATASET**

The dataset consists of 200 rows and 11 columns with different features as mentioned: **Airline, date of journey, Source, Destination, Route, departure and arrival time, Duration, Total\_stops, Price and additional information for meals** as shown in fig.1**.** The dataset is prepared according to fulfill the needs it is totally raw data as entered in the database and it shows planned flight tickets with their price and every detail and now, we will process this raw data through various steps and procedures in short, we have to analyses this raw data and chive to change it to a suitable form because machine learning algorithms only understand the numerical values.

Table, Excel

Description automatically generated

A picture containing graphical user interface

Description automatically generated

**Fig .1** Dataset based on previous flight bookings with features Airline, Date of journey, Source, Destination, Route, Departure time, Arrival time, Duration, Total stops, Additional Info and Price

**4.2 UNDERSTANDING DATA AND PRE-PROCESSING**

As I have named the dataset **“Data\_Flight.xlsx”** to **“train\_data”** as to easily access and to implement codes on it so it should be assigned after importing python libraries like pandas, NumPy, seaborn, matplotlib etc.

In this step, we have to deal with mostly two things:

**A. Missing values**

First, we have to check the missing values in the raw dataset. Using python libraries, the command used for searching missing values, Program line to find missing values:

**train\_data.isna().sum()**

Now we have to deal with the missing values as per above result only Route and Total\_Stops have one missing value so we have to apply a code line to correct it [13]. Program line to correct missing values:

**train\_data.dropna(inplace=True)**

**train\_data.isna().sum()**

**B. Data cleaning to make our data ready for analytics as well as modelling**

First, we have to check the data type:

**train\_data.dtypes**

Second, we have to change the datatypes of times. In order to change first we have to declare a function in which the algo of changing datatype is mentioned. So, the function is:

**def change\_into\_datetime(col):**

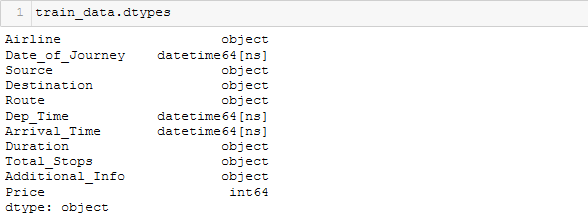
**train\_data[col] = pd.to\_datetime(train\_data[col])**

Now, use loop for changing the datatype of the whole column.

**For i in [‘Date\_of\_Journey’,’Dep\_Time’, ‘Arrival\_Time’]:**

**change\_into\_datetime(i)**

After using this function run the code line **train\_data.dtypes** again and you will get the results in which the datatypes of times will be changed as shown in fig. 2.

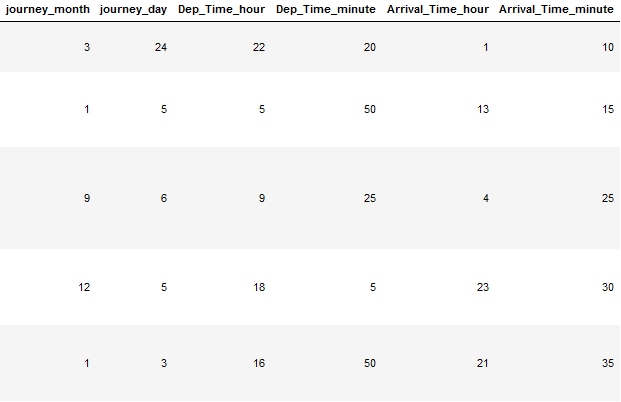


**Fig. 2** Datatypes of each feature is shown in this image. Airline, Source, Destination, Route, Duration, Total stops, Additional info are of object datatype whereas All time related features are of datetime datatype.

**4.3. EXTRACTING DERIVED FEATURES FROM THE DATA**

As it is difficult for machine learning algorithm to understand the dates and time entered in the raw data so we have to split into day, month as well as year from the journey date, arrival time, duration etc.

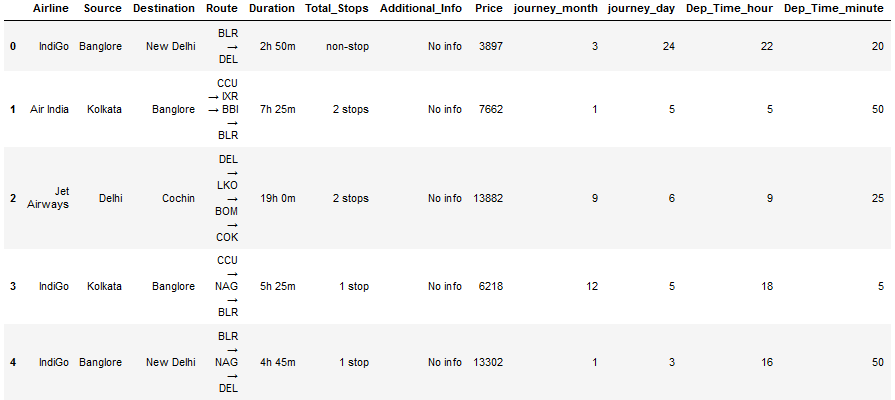
After applying the python code for splitting up the column into the result as shown in fig. 3.1:



**fig. 3.1** This image shows the data of arrival & departure times is separated w.r.t hours and minutes for better understanding

This will help in reducing in time and space complexity and more precisely the algorithm will work on this dataset. Similarly, we have modified the duration column for better understanding [5].

As we know duration of flights is in hours and minutes. Therefore, suppose duration time for the flight from Delhi to Cochin is 19 hours then it will show 19 hours 0 minutes to because there may be many duration times which contain some minutes too. For example, the duration time of flight from Bangalore to Delhi is 2 hours 50 minutes. So, now both the duration time is in same format whether the duration time has any of the parameter as zero. It will be shown in same format. Hence, easy to understand. An output after implementation is shown in fig. 3.2.



**Fig. 3.2** Duration column has been corrected here as seen in image in value ‘19h 0m’ minute is null but it will appear as 0 minute so that algorithm can understand.

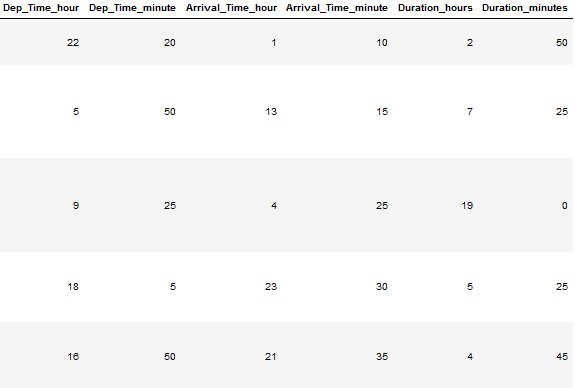
And to perform this method on data we have used split function to split the duration considering the duration time as list which have two parameters hours and minutes where the values can vary. The algorithm behind splitting the duration is

**IF: Length of duration == 2 (after splitting), then the same value will be printed.**

**ELSE: IF there is a ‘h’ (for hours) in length of duration.**

**THEN, duration = duration + ‘0m’**

And now, the Duration column in dataset is further divided into two columns ‘Duration\_hours’ and ‘Duration\_minutes’ for better understanding while machine learning algorithm works on the dataset.



**Fig. 4** For better understanding of data by machine learning the duration column is separated into hours and minutes columns.

As we can see in the above image (fig. 4) that every column containing more than two parameters are now divided into more columns according to number of parameters the columns had. This is all about data preprocessing and extracting features from the data so that to reduce the complexity to understand data by the machine learning algorithms. Now we will proceed to further steps and we have divided the duration column so now we have to remove the combined one.

For removal, code line:

**drop\_column(train\_data, ‘Duration’)** #to remove ‘Duration’ column

**train\_data.head()** #to print the updated data

and to change the datatypes of ‘Duration\_hours’ and ‘Duration\_minutes’ to integer we can use astype() function.

For conversion, code line:

**train\_data[‘Duration\_hours’] = train\_data[‘Duration\_hours’].astype(int)**

**train\_data[‘Duration\_minutes’]=train\_data[‘Duration\_minutes’].astype(int)**

**4.4. HANDLING CATEGORIAL DATA AND FEATURE ENCODING**

Now, we will work on the categorical data. Firstly, we will deal with the columns containing objects as their values to check that we have to pass a code line that is:

**cat\_col = [col for col in train\_data.columns if train\_data[col].dtype == ‘O’]**

**cat\_col**

This will print the name of columns that contain data as object in the form of concatenated string.

Similarly, columns not having values an object will be categorized separately bye passing code line that is:

**cont\_col = [col for col in train\_data.columns if train\_data[col].dtype != ‘O’]**

**cont\_col**

We have just changed the condition here; the datatype should not be an object. Hence, a concatenated string is printed.

**Feature encoding** is done with the categorized data. In simple words it is process to transform categorical variable into a continuous variable or numbers and then using them in the model to achieve results and this results in better implementation of methodology for prediction and getting fewer complex results [3][7].

For visual representation of data usually through graphs we can run a code and get the graph in respect to two parameters. For example, if I want a graph on airlines w.r.t their prices. Code line will be:

**plt.figure (figsize = (15,5))** #declared the size of figure we want

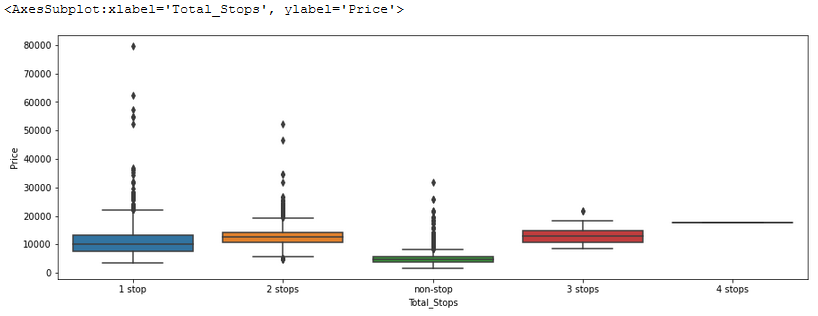
**sns.boxplot(x=‘Airline’,y=‘Price’,data = train\_data.sort\_values(‘Price’,ascending=False)),** result is shown in fig. 5.1.

Chart, box and whisker chart

Description automatically generated

**Fig. 5.1** Graph that shows the relation between Airline and their prices, graph shows how Jet airways is the most expensive airline whereas SpiceJet is cheaper among the airlines.

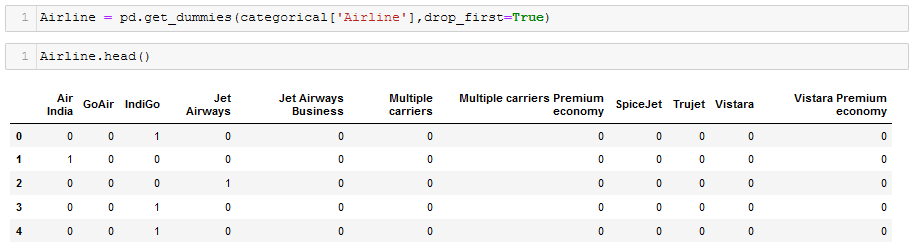
Another example as shown in fig. 5.2, just changed one of the parameter x = ‘Total\_Stops’ as total stops is a column in dataset.



**Fig. 5.2** Graph that shows the relation between Airline and the number of stops it has. It shows tickets with less cost has less stoppage.

Now we will use one hot encoding to convert categorial variables into integers. One hot encoding is a process of representing the categorial variables mapped as integer value because ML algorithms work on integers value.

In pandas we have to include get\_dummies which is used for data manipulation which is a part of data preprocessing.



**Fig. 6** In this image each airline shows numerical value of occurrence in a case like how many stoppages it has.

As we can see the categorical variables of column ’Airline’ is converted into integer format (fig. 6).

**4.5. ANALYSIS OF DATASET**

Preparing the data for further steps is done firstly by analyzing the dataset, uncovering the useful hidden features inside other features. Also, some features have been derived from other features and some columns have been divided into more than one column as it contained more than one parameter and those parameters are important factors, analysis becomes necessary for recognize good and bad features from the data [5].

Table

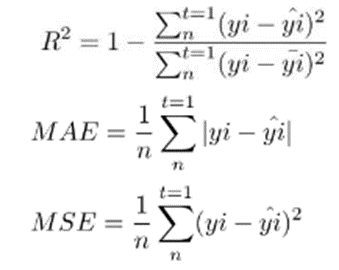
Description automatically generated with medium confidence

**Fig. 7** This image shows how a prepared data looks like before implementing the algorithms, each data must have numeric values.

After all data pre-processing and analysis of the data above image shows how a prepared data look like (as shown in fig.7). All data in numerical form and ready for implementing machine learning algorithms on it.

**5. Analysis of Machine Learning Techniques**

To develop a model that will predict the prices of airlines we have different machine learning algorithms to implement on dataset after data pre-processing. They are: Linear Regression, Random Forest, KNN algorithm, Decision tree. To do the evaluation of performance of this model, some definite parameters are examined like: **1. R-squared value 2. Mean Absolute Error (MAE) 3. Mean Squared Error (MSE)** [4] formulas for the three parameters are:



(**Eq. 1**)

Text, letter

Description automatically generated

(**Eq. 2**)

Text, letter

Description automatically generated

(**Eq. 3)**

**5.1 RANDOM FOREST**

It is another machine learning algorithm that contains multiple decision trees. This algorithm usually combines the version which is hard to predict to construct better predictive models for usage. It accumulates the lower version to make a massive version. Its capabilities are illustrated and passed to the trees without any substitution to attain the distinctly selection trees which are cannot be correlated. To pick the exceptional cut up its miles needed to have more slight correlation among the trees. The fundamental theory that distinguishes random forest algorithm from the selection tree is accumulated uncorrelated trees [4][14]. Random Forest has almost all the identical parameters similar to a decision tree. Low integration between models is key. Just as low-correlated investments (such as stocks and bonds) come together to form a portfolio larger than the sum of its shares, inconsistent models can produce more accurate forecasts than any other prediction. The reason for this positive effect is that the trees protect each other from their own mistakes (as long as they do not always fail in one place). While some trees may not be good, many other trees will be good, so as a group the trees are able to move in the right direction. The requirements for a random forest to do well are therefore:

1. There needs to be a certain signal in our features for the models built using those features to perform better than simply guessing.

2. The predictions (and therefore errors) made by each tree need to have a low correlation with each other.

The dimensionality and nature of theta depends upon its usage withinside the improvement of a tree. After the creation of countless trees, they choose the most popular category. This technique is known as random forests [28] and the implemented result is shown in fig. 8.1, 8.2 and 8.3 in section 6.

**5.2 LINEAR REGRESSION**

To find out the relation between two parameters which are continuous linear regressions analysis is preferred to used. It is method of modelling a value set as target and basically it is basically based on the number of independent variables we have in a data. As well as how correlated dependent variables and independent variables are in the selected data. In this method of analysis, the number of independent variables is only one and the connection among the independent and the dependent variables can differ in linear manner. Before diving deep into machine learning the major concept, we should know about are gradient decent and cost function.

**y(pred) = b0 + b1 \* x** (**Eq. 4**)

Values in the above equation b0 and b1 are used so to reduce the error as much as possible. The squared value which is predicted and the actual value difference results in the error. The mean square error (MSE) is used for dealing with the data which is negative, as squaring of negative or positive turns the value to positive. In this equation, b1 is called as bias which shows the positive or negative relation among the variables x and y. MSE, MAE and R-squared value are used to measure the accuracy of the linear regression problem [19][4]. The implemented results are shown in fig. 9.1, 9.2, 9.3 and 9.4 in section 6.

**5.3 DECISION TREE**

It is a tool that supports decisions it works using a tree-like model of the decisions. It is an algorithm that only have conditional controlling statements. In this, the tree count segregates the data or information which has been collected into small modules, and at the similar time makes the data determined. The results at the end shows that the tree have decision nodes as well as the leaf nodes. Each node shows the test done on an extracted feature of data and each leaf node constitutes label of the class (or class label) and the branches of the model show the conjunction between feature leading to those class labels [14]. The decision might contain at least two branches w.r.t rate. At first, the collected data is considered as root. There are two major parameters while calculating decision tree first is Information Gain and the second is Gini index [3]. Proportional change in entropy is known as Information Gain, whereas Gini Index is a component which is responsible for measuring how regularly an arbitrarily picked components may be differentiated wrongly by mistake results in errors. It implicates a characteristic with less Gini Index should be liked. For regression tree, capacity of the cost could be basically a squared equation:

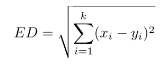
**E = ∑ (y - yᶺ)2** (**Eq. 5**)

In the above equation, y denoted original value from the data and the other parameter y-cap is the value that is predicted. Presence of a class that contains highest number of values of an assumed value acquired by the function used for splitting is known as Information Gain. If in case the value is kept splitting again and again with no cause at the leaf node [15][16]. The algorithm will be slow in completing, huge as well as over fitted. To overcome this, a slightest count on the example that is trained of the leaf node is allocated. The implementation results are shown in fig. 10.1, 10.2, 10.3, and 10.4.

**5.4 K-NEAREST NEIGHBOUR (KNN) ALGORITHM**

In K-nearest neighbor relapse algorithm, the result is the mean of its k-nearest neighbors. Similar to Support Vector Machine (SVM), KNN is likewise a technique without parameters. Thinking about couple of qualities, outcomes are registered to accomplish the perfect value. KNN algorithm is a managed order calculation that can likewise be utilized as a regressor. It relegates another information highlight the class. It is non-parametric in light of the fact that it doesn't take any presumption. It figures the distance between each preparation model and another information point [3][20]. To process this distance following distance estimation techniques are utilized:

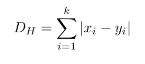
**1. Euclidean Distance**

** (Eq. 6)**

**2. Manhatten distance**

**** (**Eq. 7**)

**3. Hamming distance**

**** (**Eq. 8**)

K is the count of entries according to dataset that are selected by the training model that are nearly around fresh data point. The implementation of this algorithm [27] along with the results are shown in fig. 11.1, 11.2, 11.3 and 11.4.

**6. ALGORITHMS IMPLEMENTATION & EVALUATION**

After data preprocessing and getting the data prepared to apply machine learning algorithms. We can easily apply the algorithms to the model by declaring a function and then importing few algorithm functions from scikit learn.

**1. RANDOM FOREST ALGORITHM**

**CODE**

Graphical user interface, text, application, email

Description automatically generated

**Fig. 8.1** Function in python language is created to implement the Random Forest Algorithm to predict the airfare prices and to give the performance scores of the algorithm.

**OUTPUT**

Text

Description automatically generated

**Fig. 8.2** This image shows Predictions made by the Random Forest Algorithm along with training score, R-squared, MAE, MSE and RMSE scores.

Chart, histogram

Description automatically generated

**Fig. 8.3** X-Y axis Graph with X-axis showing Variation of Price of tickets for predictions made by Random Forest Algorithm and Y-axis showing density of the price.

**2. LINEAR REGRESSION ALGORITHM**

**CODE**

Graphical user interface, text, application, email

Description automatically generated

**Fig. 9.1** Function in python language is created to implement the Linear Regression Algorithm to predict the airfare prices and to give the performance scores of the algorithm.

Text

Description automatically generated

**Fig. 9.2** Importing Linear Regression in Python from sklearn package and implementing using the predict keyword.

**OUTPUT**

Text

Description automatically generated

**Fig. 9.3** This image shows Predictions made by the Linear Regression Algorithm along with training score, R-squared, MAE, MSE and RMSE scores.

Chart, histogram

Description automatically generated

**Fig. 9.4** X-Y axis Graph with X-axis showing Variation of Price of tickets for predictions made by Linear Regression Algorithm and Y-axis showing density of the price.

**3. DECISION TREE ALGORITHM**

**CODE**

Graphical user interface, text, application, email

Description automatically generated

**Fig. 10.1** Function in python language is created to implement the Decision Tree algorithm to predict the airfare prices and to give the performance scores of the algorithm.

A picture containing chart

Description automatically generated

**Fig. 10.2** Importing DecisionTreeRegressor in Python from sklearn package and implementing using the predict keyword.

**OUTPUT**A picture containing chart

Description automatically generated

**Fig. 10.3** This image shows Predictions made by the Decision Tree Algorithm along with training score, R-squared, MAE, MSE and RMSE scores.

A picture containing chart

Description automatically generated

**Fig. 10.4** X-Y axis Graph with X-axis showing Variation of Price of tickets for predictions made by Decision Tree Algorithm and Y-axis showing density of the price.

**4. K-NEAREST NEIGHBOR ALGORITHM**

**CODE**

Graphical user interface, text, application, email

Description automatically generated

**Fig. 11.1** Function in python language is created to implement the K-nearest neighbor algorithm to predict the airfare prices and to give the performance scores of the algorithm.

Text

Description automatically generated

**Fig. 11.2** Importing KNN algorithm in Python from sklearn package and implementing using the predict keyword.

**OUTPUT**

Text

Description automatically generated

**Fig. 11.3** This image shows Predictions made by the K-nearest neighbor Algorithm along with training score, R-squared, MAE, MSE and RMSE scores.

Text

Description automatically generated

**Fig. 11.4** X-Y axis Graph with X-axis showing Variation of Price of tickets for predictions made by K-nearest neighbor Algorithm and Y-axis showing density of the price.

**ALGORITHM’S EVALUATION TABLE:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Random forest** | **Linear Regression** | **Decision Tree** | **KNN Algorithm** |
| **Training Score** | 0.9345 | 0.6828 | 0.1000 | 0.7738 |
| **R-Squared** | 0.82581 | 0.5549 | 0.4838 | 0.2786 |
| **MAE** | 1914.4497 | 1883.7805 | 2725.575 | 2181.297 |
| **MSE** | 4976585.149 | 7119150.076 | 11270523.625 | 778473.767 |
| **RMSE** | 43.75 | 43.51907 | 52.20 | 46.70 |

**Table 1** Performance accuracy table which shows training score, r-squared, MSE, MAE and RMSE of all the algorithms. Which shows Random Forest have best training score and Decision Tree has poor training score.

**7. Limitations**

1. RANDOM FOREST ALGORITHM

This algorithm requires a lot of computational power also the resources as the process includes of building a many numerous trees to merge the outputs. Also, more time is required for the algorithm to train the model.

2. DECISION TREE ALGORITHM

Unstable character of the algorithm that means if there is a small change in the data it may lead to drastic changes in the structure of the optimal tree with will lead to false or deviated results.

3. LINEAR REGRESSION ALGORITHM

The data we are using must be independent without independency algorithm will not work properly and complexity will rise inversely. It only looks for the mean of the dependent variable.

4. K-NEAREST NEIGHBOUR ALGORITHM

Accuracy of this algorithm mainly depends on in what condition data is also it is used for small datasets because for large datasets it is very slow in processing as well as it requires more memory.

**8. CONCLUSION**

In this report the general data for the dynamic charge modifications within side the tickets of airlines are presented. This put forward the statistics approximately the peaks and valleys within side the fares of airlines tickets according to the days, whether it is end of week, daytime, during summer or winter vacations, festivals, morning or night, additionally the gadget gaining knowledge of fashions within side the artificial intelligence or field of computation that are already being obtained on unique datasets are measured. Their accuracy and performances are estimation also comparison is done which will show higher end result. For predicting the prices, price tag outlays perfectly different prediction fashions are examined so as to get the best predicting accuracy. As the manner of getting prices of different company are developed aiming to make the management of sales as higher as possible. For getting the end result more accurate, regression evaluation is used. Studies shows the characteristic that affects the expenses of the price tag are to be pondered.

9. **FUTURE WORK**

In future, the approximate count of available tickets can enhance the overall performance of the model. The best weakness of this work is the deficiency of information. Anybody wishing to develop it ought to look for elective wellsprings of chronicled information, or be more systematic in gathering information physically over a timeframe [13]. Moreover, a more shifted set of flights ought to be investigated, since it is totally conceivable that aircrafts fluctuate their evaluating methodology as indicated by the attributes of the flight (for model, admissions for local trips out of little air terminals might act uniquely in contrast to the major, well-flown courses we considered here) [2]. At long last, it is intriguing to think about our framework's exactness against that of the business frameworks accessible today (ideally throughout some stretch of time). The way covid-19 had packed us in our homes is the time period we get to know how important and safe online purchasing is and by advancing technology there will be more features and easy ways to purchase tickets and any other stuff by just sitting at home and payment options are also safe ways there are more than two-factor authentication and to not get robbed confirmation of purchase of tickets are being sent through many ways like sending permanent soft copy of whole ticket through emails, messages etc. with every detail about the booking so that the buyer can print the hard copy and get assured about reservation[21].

**REFERENCES**

1. Tianyi Wang∗, Samira Pouyanfar∗, Haiman Tian∗, Yudong Tao†, A Framework for Airfare Price Prediction: A Machine Learning Approach, July 2019
2. Manolis Papadakis, “Predicting Airfare Prices”, 2014.
3. Supriy Rajankar & Neha Sakharkar, Pricing Prediction using Machine Learning‖, International Journal of Engineering Research and Technology, vol 8, issue 12, December 2019.
4. Supriya Rajankar and Neha Sakharkar, A Survey on Flight Pricing Prediction using Machine Learning, Vol. 8 Issue 06, June-2019.
5. Akanksha V. Jain, Aanal S. Raval, Ruchi K. Oza, “Airfare Price Prediction Based on Reviews

Using Machine Learning Techniques”, Vol. 9, Issue 3, March 2020.

1. K. Tziridis, Th. Kalampokas, G.A. Papakostas, K.I. Diamantaras, Airfare Prices Prediction Using Machine Learning Techniques, EURASIP 2017.
2. J. Stavins, “Price discrimination in the airline market: The effect of market concentration,” Review of Economics and Statistics, vol. 83, no. 1, pp. 200–202, 2001.
3. B. Mantin and B. Koo, “Dynamic price dispersion in airline markets,” Transportation Research Part E: Logistics and Transportation Review, vol. 45, no. 6, pp. 1020–1029, 2009.
4. P. Malighetti, S. Paleari, and R. Redondi, “Has Ryanair’s pricing strategy changed over time? an empirical analysis of its 2006–2007 flights,” Tourism Management, vol. 31, no. 1, pp. 36–44, 2010.
5. T. H. Oum, A. Zhang, and Y. Zhang, “Inter-firm rivalry and firm-specific price elasticities in deregulated airline markets,” Journal of Transport Economics and Policy, vol. 7, no. 2, pp. 171–192, 1993.
6. K. S. Gerardi and A. H. Shapiro, “Does competition reduce price dispersion? new evidence from the airline industry,” Journal of Political Economy, vol. 117, no. 1, pp. 1–37, 2009.
7. S. A. Rhoades, “The herfindahl-hirschman index,” Federal Reserve Bulletin, vol. 79, p. 188, 1993.
8. G. Francis, A. Fidato, and I. Humphreys, “Airport–airline interaction: the impact of low-cost carriers on two european airports,” Journal of Air Transport Management, vol. 9, no. 4, pp. 267–273, 2003.
9. International Civil Aviation Organization, “List of lowcost-carriers (LCCs),” cited July 2018. [Online]. Available: https://www.icao.int/sustainability/Documents/LCC-List.pdf
10. C. Koopmans and R. Lieshout, “Airline cost changes: To what extent are they passed through to the passenger?” Journal of Air Transport Management, vol. 53, pp. 1–11, 2016.
11. S. Lee, K. Seo, and A. Sharma, “Corporate social responsibility and firm performance in the airline industry: The moderating role of oil prices,” Tourism Management, vol. 38, pp. 20–30, 2013.
12. V. Nair and G. E. Hinton, “Rectified linear units improve restricted Boltzmann machines,” in the 27th international conference on machine learning, 2010, pp. 807–814.
13. D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in the 3rd international conference on learning representations, 2015.
14. Fare prediction. Available: <http://arxiv.org/abs/1412.6980> [accessed on 12 June 2021]
15. T. Liu, J. Cao, Y. Tan, and Q. Xiao, “ACER: An adaptive context-aware ensemble regression model for airfare price prediction,” in the international conference on progress in informatics and computing, 2017, pp. 312–317.
16. V. Pai, “On the factors that affect airline flight frequency and aircraft size,” Journal of Air Transport Management, vol. 16, no. 4, pp. 169–177, 2010.
17. M. S. Ryerson and H. Kim, “Integrating airline operational practices into passenger airline hub definition,” Journal of Transport Geography, vol. 31, pp. 84–93, 2013.
18. H. Baik, A. A. Trani, N. Hinze, H. Swingle, S. Ashiabor, and A. Seshadri, “Forecasting model for air taxi, commercial airline, and automobile demand in the United States,” Transportation Research Record, vol. 2052, no. 1, pp. 9–20, 2008.
19. T. Janssen, T. Dijkstra, S. Abbas, and A. C. van Riel, “A linear quantile mixed regression model for prediction of airline ticket prices,” Radboud University, 2014.
20. R. Ren, Y. Yang, and S. Yuan, “Prediction of airline ticket price,” University of Stanford, 2014.
21. T. Wohlfarth, S. Clemenc¸on, F. Roueff, and X. Casellato, ´ “A data-mining approach to travel price forecasting,” in the 10th international conference on machine learning and applications and workshops, vol. 1, 2011, pp. 84–89.
22. H.-C. Huang, “A hybrid neural network prediction model of air ticket sales,” Telkomnika Indonesian Journal of Electrical Engineering, vol. 11, no. 11, pp. 6413–6419, 2013.
23. L. Breiman, “Random forests,” Machine Learning, vol. 45, pp. 5-32, 2001.