

Flight Price Prediction for Users by Machine Learning Techniques

Pavithra Maria K¹, Anitha K L²

Student, Department of Computer Science, Mount Carmel College, Bangalore, India¹

Lecturer, Department of Computer Science, Mount Carmel College, Bangalore, India²

Abstract: People who frequently travel through flight will have better knowledge on best discount and right time to buy the ticket. For the business purpose many airline companies change prices according to the seasons or time duration. They will increase the price when people travel more. Estimating the highest prices of the airlines data for the route is collected with features such as Duration, Source, Destination, Arrival, Departure. Features are taken from chosen dataset and in this paper, we have used machine learning techniques and regression strategies for prediction of the price wherein the airline price ticket costs vary overtime. We have implemented flight price prediction for users by using decision tree and random forest algorithms. Decision tree shows the best accuracy of 80% for predicting the flight price. Also, we have done correlation tests and ANOVA test for the statistical analysis.

Keywords: Feature selection, Airfare price, Machine learning, Pricing Models, Prediction Model, Random Forest.

I. INTRODUCTION

Perfect time for purchasing plane ticket by the passenger's view is difficult since passengers get very less information of future business price rates. Different models figure out future business price on plane and categorise the best time to obtain flight ticket. Airlines use different strategies of pricing for their tickets, later taking the decision on price because order shows higher value for the approximation models. The causes behind the difficult system is each Planes has limited number of seats to be filed, so airlines must regulate demand. Suppose when demand is expected to increase capacity, the airline may increase prices, to decrease the rate at which seats fill.

Also, seating arrangements in flight which is not occupied shows the loss of the amount invested for the business airline companies and making them purchase the ticket to fill the seats for any price this would be the best idea to get profit in loss too. Passengers should be compatible with the airline companies to get adjusted for the increase and decrease of the price. Passengers or customers should make their own planning to get the best offers available on different airlines and travel through less price. Planes ticket prices changes as time passes, pulling out the elements which creates the difference. Reporting the correlated and models which is used to price the flight tickets. Then, using that information, building the model which helps passengers to make pull out the ticket to buy and predicting air ticket prices which progresses in the future. Duration, Arrival time, Price, Source, Destination and much more these are the attribute used for flight price prediction.

II. OBJECTIVE

In conceptual level, there are two subtasks for selecting features and taking decision about feature combination. Execution of Building the model through Random Forest, decision tree and test the results of continuous as well as categorical data. Removing the elements which are of not required, duplicate and redundant information from the data collected.

Limiting the caused error and through classification results we are increasing the accuracy. Choosing the subdivision of suitable attribute from the whole dataset. Performing statistical analysis through various test as ANOVA, Correlation and Chi-Square test.

III. LITERATURE SURVEY

In the preceding work on improving prediction models for airline prices by using Machine Learning (ML) techniques, the different exploration team has concentrated on various attributes and have trained the models on various kinds of Airlines. Specific trend is that they are trying to predict the price. Specifically, categorizing flight price with two divisions of elements helps the studied impact on mean price of the plane. Authors have examined the airline profit by applying pricing modes and have found that after a time duration of 70 days, categorical cases for a flight are observed

as flight departure and the discount opportunities also tend to increase over time. Through the analysis we identify equal pricing techniques applied by the airline companies to positively manage the airline offers and demand to increase their business profit. Results shows airlines worry about the price changes according to the season in websites.

The point should be noted down that the importance between the online pricing, and the realised price dispersion on a flight. At the end using prices from actual transactions, the authors have found that online price division is more highly present in lower business airline competition [1].

Predicting the plane ticket prices and limiting the price for passengers. Price models with reliability can assist passengers to determine the scope of future prices for the airline companies. Present business airline companies will not give passengers to estimate the future reliable costs of any departure requirements. For the usually travelling passengers in this model was developed with price attribute from their own history.

First, the high price duration will not be the minimum price accessible for a plane, the passenger data is specifically scheduled that may need to be changed for the price. The attribute from the dataset with subject information also as mentioning high understanding is not required. The Final model can be inspected for subject understanding. In the final work there are extra price limitations that pull out to have outcomes nearer to the accurate solution [2].

Price attribute allows grouping into clusters, also on the similarity of their behaviour. By this representation, the clustering stage follows statistical analysis in group thus formed. The probability distribution for observed trajectories has been estimated for each cluster. From previous data, we learn a decision tree method, by visualising through different attributes to assign new flight. A variant, considering the first points of the trajectory in the list of predictor variables, has also been considered, to obtain more accurate predictions. In future, the data should be improved more by including with stops, prices found on online flight booking websites, and the increase in prices are collected from 28 days to 90 days [3].

Improving the ML structure to predict the mean plane price for the business purpose. For predicting the mean plane price with modification of R squared score, feature selection techniques were proposed in our model. Comparing the production of various ML classifiers which tells the greater plane price prediction task [4]. Facts gathered from website that sells the planes ticket through internet apps. Authors have reported that there is limited public information access which will miss the main target attribute. Final accountable prediction model is improved by two unrelated prediction models such as Random Forest and Multilayer Perceptron. Weights for drifting with R-square value and the main estimation of the metric was used [5].

Authors have reported on the work that collecting plane data from a Greek flight company from the website, also have visualised that it is workable to forecast prices for plane on previous collected data. Outcome tells that ML models are acceptable for forecasting plane prices and various other elements is collecting the data and selecting the features from which authors have drawn positive results.

From the experiment's authors have told which elements influence the plane forecasting at most and the other element is that it could improve the forecasting accuracy. In coming days paperwork will be expanded to forecast the plane prices for the whole plane companies. Extra observation on huge data sets is required and passengers must buy airline ticket in the best business time duration [6]. Forecasting air prices regarding the timeline issue and ML ideas and strategies have been used to solve plane price issue. Here it is the combination of clustering, modelling and approach to propose ACER. ACER finds attribute which are acceptable for content data so that models can be trained. Tested conclusion tells that ACER execute better on all chosen routes [7].

Authors have initiated to address for improving the plane estimate favour, also considering the plane estimation as a timeline issue. Concentrated on extracting possible figures of the fare modification by ML techniques. Data model is first presented by the authors to arrange the price values and pull out attributes.

Fitting the flowing way and understanding the idea of floating in the sequence, the authors have presented Learn++. This paper study has still in its early stage. instead of the limited price on a whole travelling places, a perfect view of independent plane will be worth as passengers might get partiality while buying the tickets [8].

Predicting the fight fares in the limited time by improving the study in a present scenario. Improvement is done in a linear model and algorithm used in this has drawn information and the practical data as time, week, day, date etc, are given as input to the forecast value.

In detail monitoring, the passenger gets an approximation of plane price with date to choose the best blend of date and price. The price for weekend on Sunday is not possible to calculate in this presented model, as weekend on Sundays the most accidental price difference compared to other days in the week and needs more elements, nonlinear model for successful forecast which will be the upcoming range of study to be done for this presented technique [9]. To forecast the mean plane ticket amount on the business area, machine learning support was evolved. Selecting feature techniques authors have presented model to forecast the mean flight amount with R squared score of 80% accuracy.

According to the profit of ticket with time, departure, arrival the datasets has detailed information. Also, price prediction can be calculated on a quarterly base. Framework can be extended to include air ticket transaction information which gives more detail information about airlines, such as time and departure and arrival, etc., it builds more airfare price prediction model for effectively [10].

Regression techniques are used such as logistic regression which alter its outcomes using the sigmoid logistic function to give back a probability value which can then be assigned to two or more discrete classes. Based on this author have used it for airline industry because most of the time the ticket price keeps changing for any day. For example - if you want to buy a ticket for a flight in ten days the ticket price may increase or decrease according to the day and the difference between travel date and booking date. As they have used a training dataset for the prediction of ticket price, so it gives us a good result. The accuracy of logistic regression model is up to 70-75%. The conclusion of the given model is that most of the plane ticket price vary from day to day. Authors have reported that the ticket price is high for a certain period and then it gradually decreases to a certain level. When the flight is at a difference of 2-3 days' time the ticket price starts increasing again [11].

IV. METHOD AND MATERIALS

A. Proposed Model

Through Regression Analysis the visualization and forecasting are performed for the presented model. Blending of technologies, processing is called Conceptually the Intelligence, that is cloud computing, machine learning and virtualization etc. ML is in trend to build our skills and it is one of the highest growth field in computer science and health care informatics.

As the time passes by the algorithm should be learnt is the main goal in Machine. Also, used for predicting algorithm that makes the communication with agent and makes easier for learning. In this paper, random forest and decision tree algorithms is to find solutions for flight price problems in machine learning tasks and a hybrid method is formulated from Chi square, anova and Correlation tests is performed.

The data collecting is performed followed by data pre-processing. Before data modelling is done, data must be split into train and test dataset to ignore the data leakage. Based on the various attributes in the dataset for example departure and arrival features play the important role for predicting the price. Running the random forest and decision tree models, grouping the maximum price of airlines. Next performing the feature engineering and calculating the accuracy.

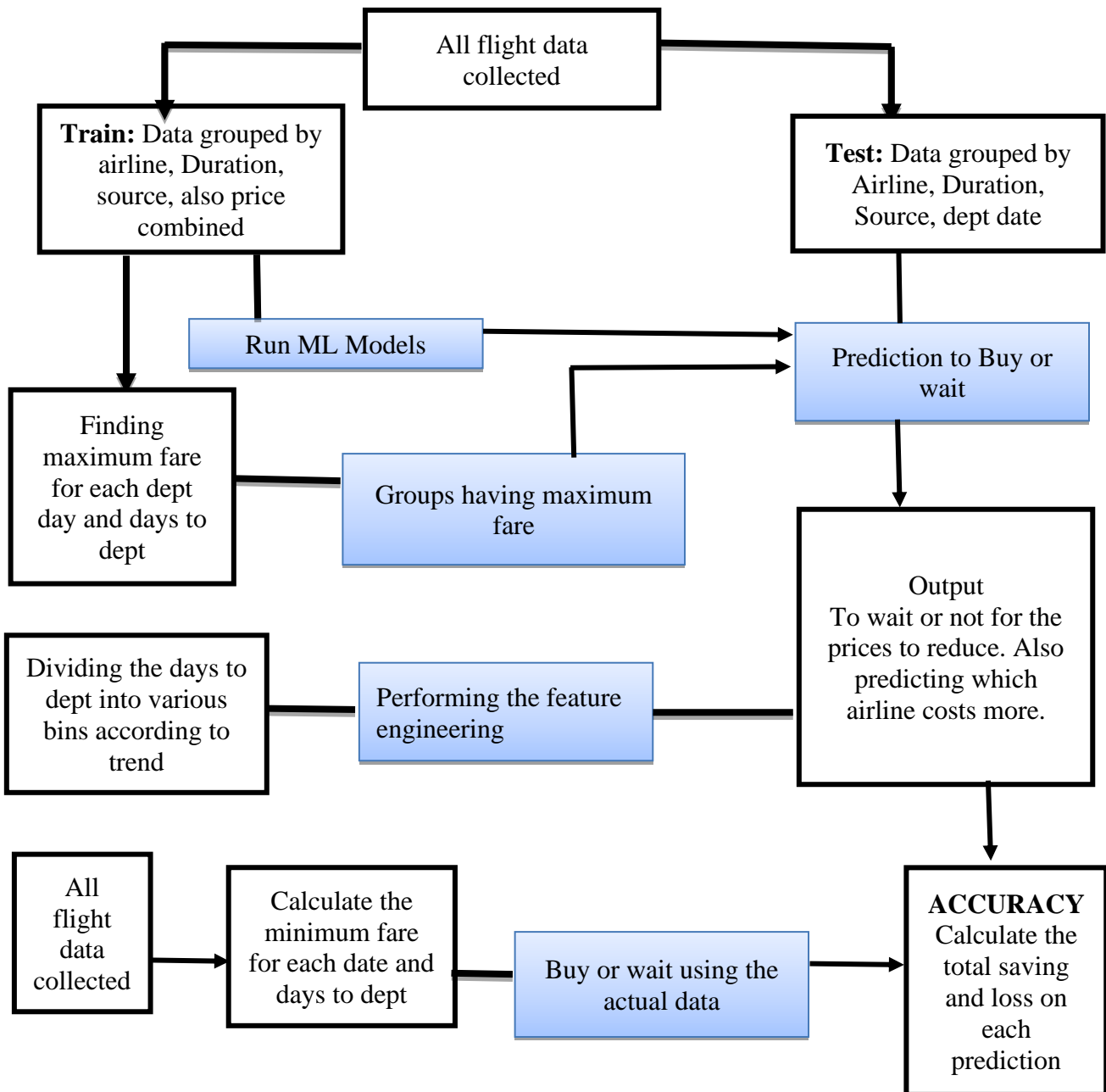


Fig. 1 Workflow for Flight Price Prediction

B. DATA COLLECTION

The important part of the project is data collection. Data on different websites is gathered with unique attribute to provide the best accuracy. The data is collected from website kaggle.com and the models are implemented using python. The python-script helps to easily pre-process the data and forecast the output. The duplicate values are avoided in the pre-processing step. This dataset is more concentrated on calculating the plane price value. The dataset contains the data with attributes such as

- Journey_Date
- Departure
- Designation
- Arrival

- Airline
- Duration
- Source
- Price

1. Cleaning and preparing data

The gathered data must be cleaned and pre-processed and after improving the data, it is read to run on the algorithms. The duplicate values are removed, data is arranged with numerical values by pre-processing and by this model building and selecting the features becomes easier. Pre-processing plays the vital role for the whole dataset.

TABLE I: CLEAN AND PREPARED DATASET

Total_Stops	Journey_day	Journey_month	Dep_hour	Dep_min	Arrival_hour	Arrival_min	Duration_hours	Duration_mins	Air India	GoAir	IndiGo	Jet Airways	Jet Airways Business	Multiple carriers	Premium economy	
0	1	6	6	17	30	4	25	10	55	0	0	0	1	0	0	0
1	1	12	5	6	20	10	20	4	0	0	0	1	0	0	0	0
2	1	21	5	19	15	19	0	23	45	0	0	0	1	0	0	0
3	1	21	5	8	0	21	0	13	0	0	0	0	0	0	1	0
4	0	24	6	23	55	2	45	2	50	0	0	0	0	0	0	0

Dataset in table I shows the information which is needed for the analyzing the data. Extra features is added to create best results. Feature like Dep_hour and Arrival_hour and Duration_hour is created to analyze the data for time duration of the day and other factors.

TABLE II: CLEAN AND PREPARED DATASET

Multiple carriers	Multiple carriers Premium economy	SpiceJet	Vistara	Vistara Premium economy	Chennai	Delhi	Kolkata	Mumbai	Cochin	Delhi	Hyderabad	Kolkata	New Delhi
0	0	0	0	0	0	1	0	0	1	0	0	0	0
0	0	0	0	0	0	0	1	0	0	0	0	0	0
0	0	0	0	0	0	1	0	0	1	0	0	0	0
1	0	0	0	0	0	1	0	0	1	0	0	0	0
0	0	0	0	0	0	0	0	0	0	1	0	0	0

Dataset in Table II shows the complete information of pre-processed data as mention in Table I.

2. Analyzing Data

Constructing of the data is the huge task, by knowing the various patterns of data visualization and later using the required machine learning models. Also, from the current attribute the new small elements can be acquired. If it is on holiday, festival day or a weekday or weekend, plane date plays main role. Travelling during weekends is costlier

than the planes on weekdays and time is considered in classes as: Morning, afternoon, evening and night, so time plays important role. Travelling days is computed with plane date and the date on which data is collected.

V. MACHINE LEARNING MODELS

Flight price forecasting using various algorithms in machine learning. The algorithms for forecasting purpose are: Support Vector Machine, Linear regression, K-Nearest neighbors, Multilayer Perceptron, Gradient Boosting and Random Forest Algorithm, Decision tree. Traversing the python library and parameters like R-square, MAE and MSE area unit to verify the production of those models.

1. Linear regression

Variable quantity of that price is to be found for this we are employing statistical regression analysis such as correlation between 2 continuous variables, from the 2 variables. The equation for statistical regression is:

$$y(\text{pred}) = b_0 + b_1 * x \quad (1)$$

The two major factors to grasp statistical regression is gradient descent and price operate area unit. It gives the simplest match line to the given data that the forecast error is minimum and provides the applied mathematics relationship not the settled relationship between 2 variables. The sq. of expected and actual price distinction gives the error. To alter the negative values, the mean sq. error is taken (MSE). Value of the coefficients b_1 and b_0 area unit chosen in order that the error value is as little as doable.

Choosing a random data from a dataset with replacement is called Bootstrap aggregating. By gradient boosting and random forest strategies achieves the greater accuracy.

2. Decision Tree

It is used to make any decision and have multiple branches which are the Decision Node and Leaf Node. Decision Tree used for both classification and Regression problems, but mostly it is preferred for solving Classification problems which is a Supervised learning technique. Decision tree has the two nodes, represents the features of a dataset, each leaf node shows the outcome is Internal nodes and branches shows the decision rules. Based on features of the given dataset the test must be performed. For getting all the possible solutions to a problem on the given condition's visualization is done. It starts with the root node that expands on branches and builds structure as tree is called as decision tree. CART algorithm helps in Classification and Regression Tree algorithm, that is used to build the tree. The two essential properties for tree computation is Gini index and data Gain. Lot of successful of the substance tells that it has Higher entropy. The Decision Tree gives the best accuracy with 80% contrast to random forest algorithm.

3. Random Forest

One of the popular machine learning algorithms which belongs to the supervised learning technique is Random Forest. Process of combining multiple classifiers to solve complicated problem and increasing the performance of the model it is based on the concept of learning. To avoid the overfitting problem and the greater number of trees in the forest tends to greater accuracy. Random forest is used for both classification and regression problems is the huge advantage. The resultant accuracy the random forest gives are 70% as shown in the result and analysis graph and table.

VI. STATISTICAL ANALYSIS

1. Chi-Square Test

Relationship between two categorical variables is used to do statistical test that is Chi-Square. One variable having the frequency compared against the second variable's categories is done by executing Chi-Square. That defines the data is shown as a frequency table, rows show the independent variables and columns shows the dependent variables.

2. Correlation Test

The correlation or bivariate relationship between two independent variables, so correlation plot is used. Identifying the correlation of one independent variable with a group of other variables, VIF is used. So, VIF is used for best understanding. When VIF is equal to 1, it is No Correlation. VIF is equal to 1 to 5, it is Moderate Correlation. VIF is

greater than 10, it is called Highly Correlated.

3. Anova Test

To compare two or more contrasting together to determine the analysis of variance Analysis of Variance statistical test is performed. Analyzing the differences between groups and signifying the difference statistically is the one-way ANOVA tests. The other way to compare two or more independent group which is using when at least three independent groups are available.

VII. EXPERIMENTAL RESULTS AND PERFORMANCE ANALYSIS

Python is a high-level object-oriented scripting language, designed to be readable it uses English keywords more and uses indentation, whereas other languages use punctuation. Functions gives best modularity for our application and a high amount of reusability of the code. In python classes and objects are easily used.

Many built-in functions like print(), etc. We can also create your own functions in python, so these functions is called user-defined functions. Python libraries for data analysis by making use of Numpy, Pandas and Scipy for the selected dataset. Open-source library pandas is used to manipulate, analyze, load, and visualize the selected datasets. The other open-source library scikit-learn builds smart models and make cool predictions and is used in machine learning algorithms.

Output of the model is visualized graphically across the test dataset for chosen test dataset. The visualization represents study of real value, also the prediction of results. Decision Tree, Bagging Tree, Random Forest and Linear regression tells the results gained by the analysis. Also, gives the attribute price to purchase the flight ticket at the right time for predicting the price values. Decision Tree algorithm has more accuracy compared to other algorithms for the given dataset. It gives the highest R-Square value with maximum accuracy in the regression analysis. First column gives the values for R Square, were table III shows R-square, MSE and MAE values.

TABLE III: ALGORITHM EVALUATION

Machine Learning(ML)	R-squared	MAE	MSE
Random Forest	0.79	1166.1987291481917	4054043.514563705
Decision Tree	0.80	1166.1987291481917	4054043.514563705

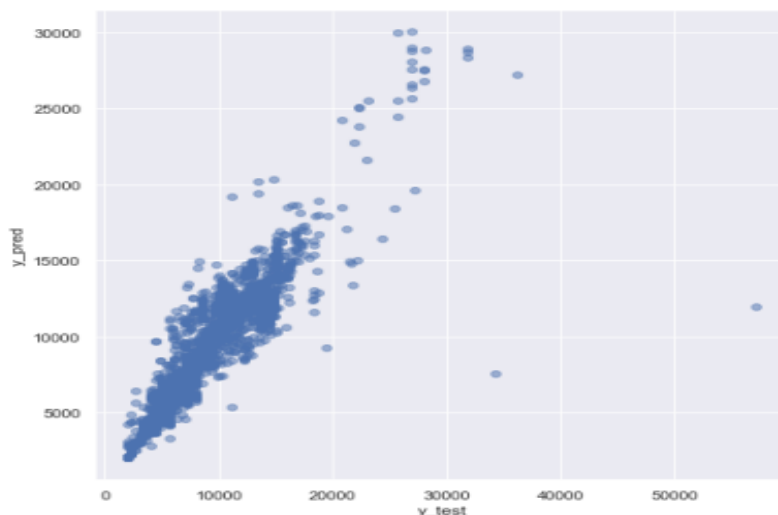


Fig. 2 Graphical results for Flight price prediction using Random Forest.

The Fig 2 for Random Forest, graph is plotted with Y-test data, it shows the price value is getting higher. Here the observation is according to the prediction the passengers who have purchased the ticket with highest amount of price is around thirty thousand and most of the people have purchased the tickets between five thousand to fifteen thousand. Through this analysis the advantage is that people will have more idea about the frequent price offers in festival and holiday season and will choose the best price for travelling.

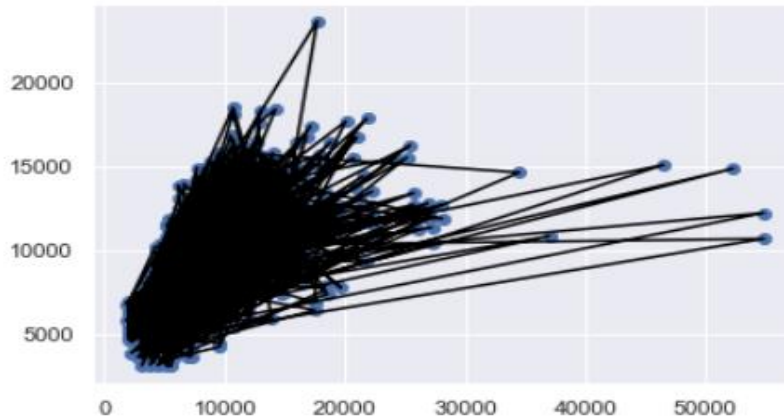


Fig. 3 Graphical results for Flight Price Prediction using Decision Tree.

The Fig 3 for Decision Tree tells with Y_Test set again the price increases as the time varies. In this graph also we observe passengers gave purchased the ticket from five thousand to fifteen thousand and highest ticket price purchased is twenty-five thousand. So, we can see that the accuracy level is low here itself. Through this analysis the advantage is that passengers can ask for the review to their friends or relatives who travel mostly, this helps to take better decision for travelers.

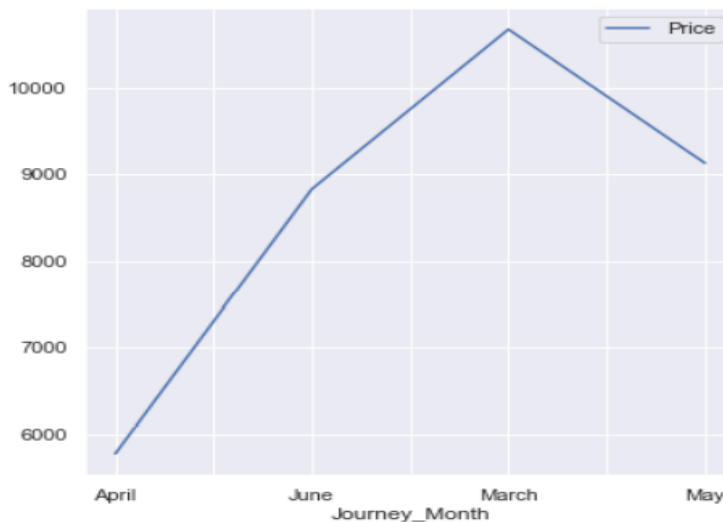


Fig. 4 Graphical results of Analysis between Journey month and Price.

The Fig 4 considering the features Price and Journey Month we see that the prices are higher at the month of the march as people travel more the company increases the expenses. People should make their own strategies to when to travel and make use of the best offer from the airlines. Since airline company increase the price for the business purpose, people should be also smart to travel with best price. Through this analysis the advantage is that passengers can see which month is best to travel with affordable prices because season to season price changes. People are smart to choose the cheapest price and comfort for travelling.

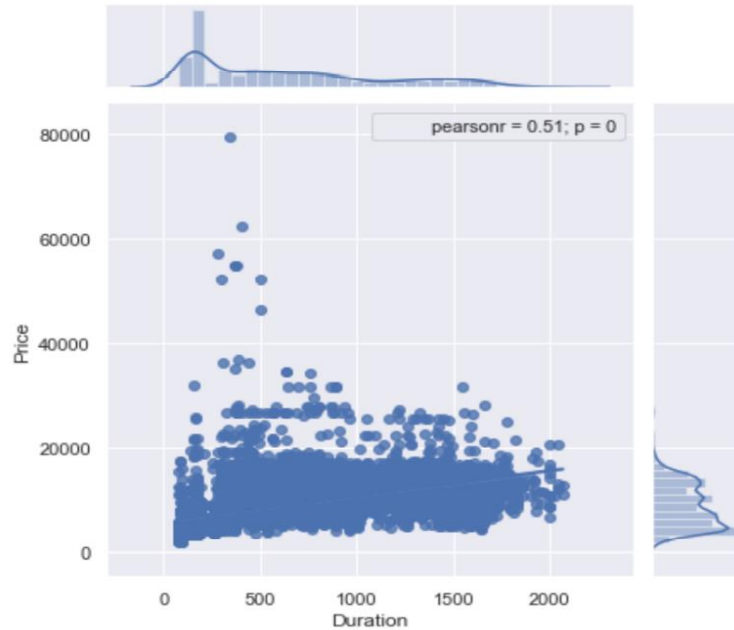


Fig. 5 The Statistical Analysis between Duration and Price as correlation test.

The Fig 5 correlation test gets p-value < 0.05 , hence we accept H_1 and say the target variable and continuous independent variable are correlated. $r = 0.51$ says they are moderately related. Through this analysis the advantage is that the time duration plays the important role for making the decision to board the flight with best price. With the limited amount of time the best price can be chosen by the passengers. Everybody can afford the flight ticket with best price and best offers.

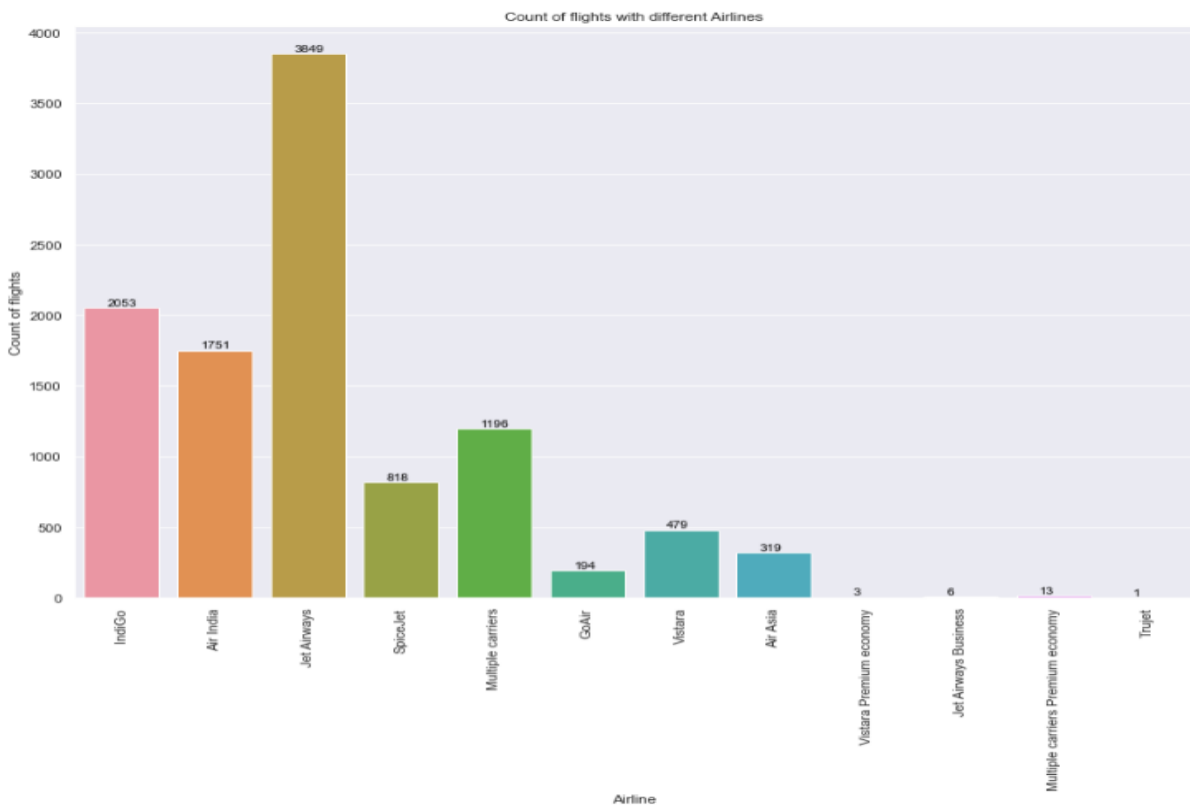


Fig. 6 Graphical results for Airlines and the number of flights with price.

The Fig 6 the graphical representation shows the price value is getting higher. The observation is Jet Airways is having the highest price of airline compared to other airlines with the price 3849 rupees. The second highest is IndiGo airline with 2053 rupees and all other airline prices are almost similar. Through this analysis the advantage is that the customers has the option to choose the various airlines with different price and comfort with their budget to travel the new places and explore the world.

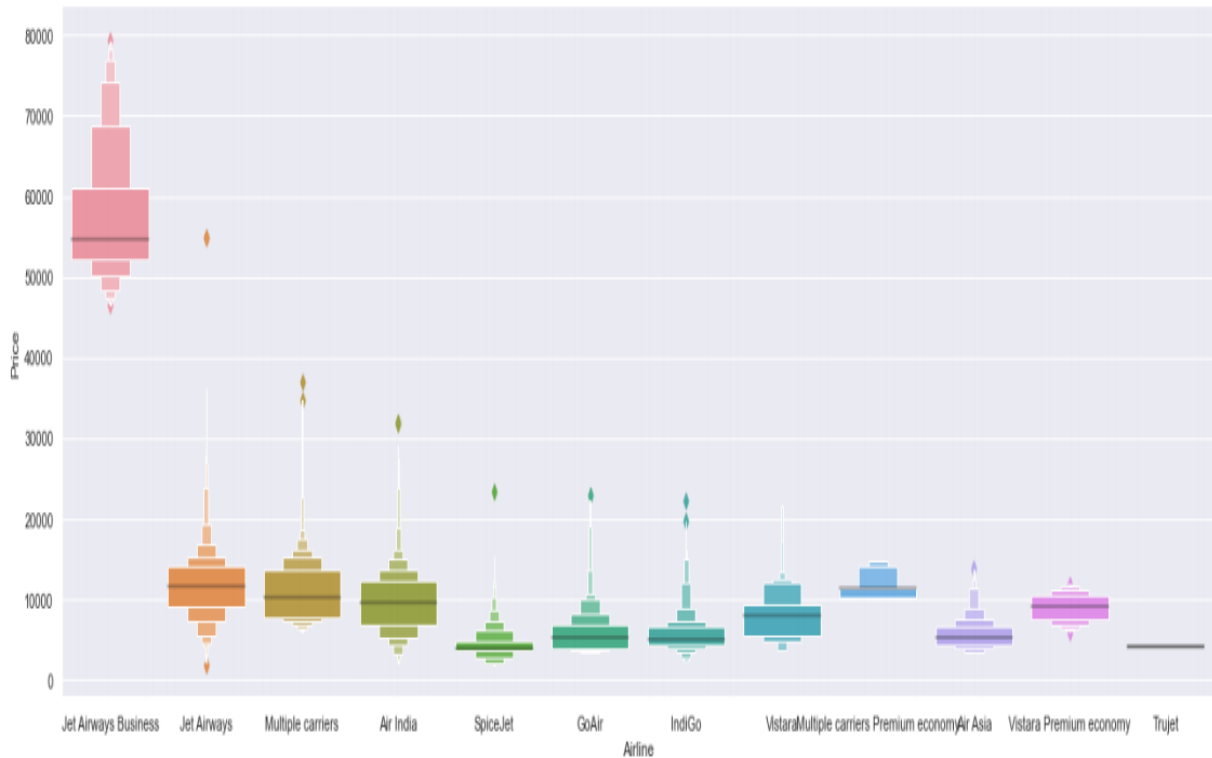


Fig. 7 The Analysis between Airline and Price

The Fig 7 Jet airways and Air India are full-service airlines are and always highly priced due to various amenities they provide. Low-cost carriers like indigo and SpiceJet have a lower and similar fare. Through this analysis the advantage is that travellers can see the highest to lowest prices to know the price differences in each airline. Passengers can share the review of the flight on the each airline websites, so that the other passengers gets the benefit out of it.

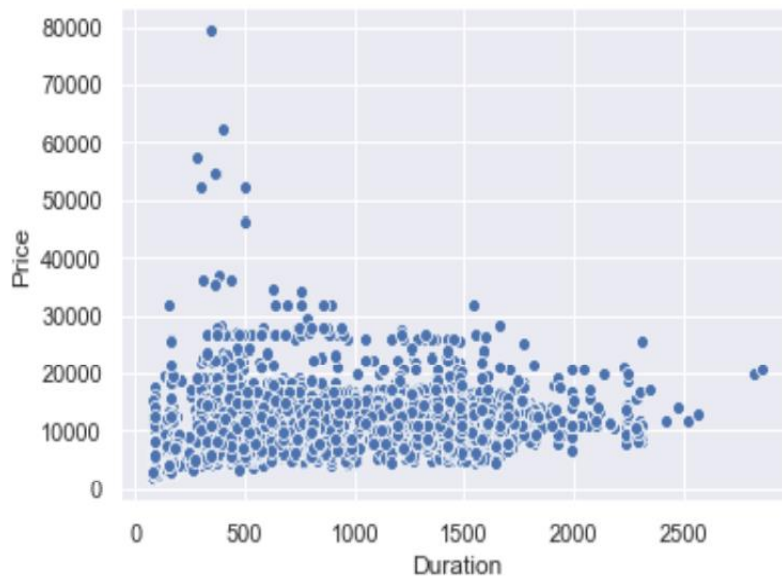


Fig. 8 The Analysis between Duration and Price

We know that duration (or distance) plays a major role in affecting air ticket prices, but we see no such pattern here, as there must be pattern here and other significant factors affecting air fare like type of airline, destination of flight, date of journey of flight (higher if collides with a public holiday). Through this analysis the advantage is that passengers might want to reach the destination sooner, so the duration plays the important role to reach sooner with good amount of price.

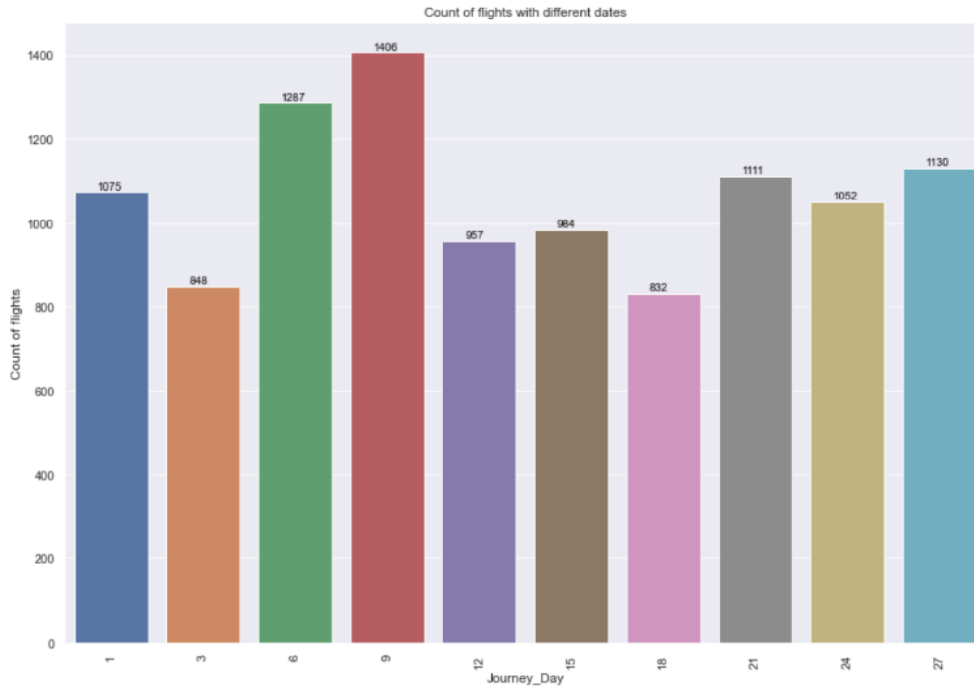


Fig. 9 The analysis between Journey_Day and Count of flight

It shows the price count of flights on the journey day, here the maximum count is on 9th and second highest is on 6th. The Visualisation should be attractive and easy to understand so multiple colours are used to make the difference between each date. Through this analysis the advantage is that the Journey Day plays the important role because the flight charges for weekdays in might be lesser and weekends the price might be higher based on the offers and airlines chosen to board the flight.

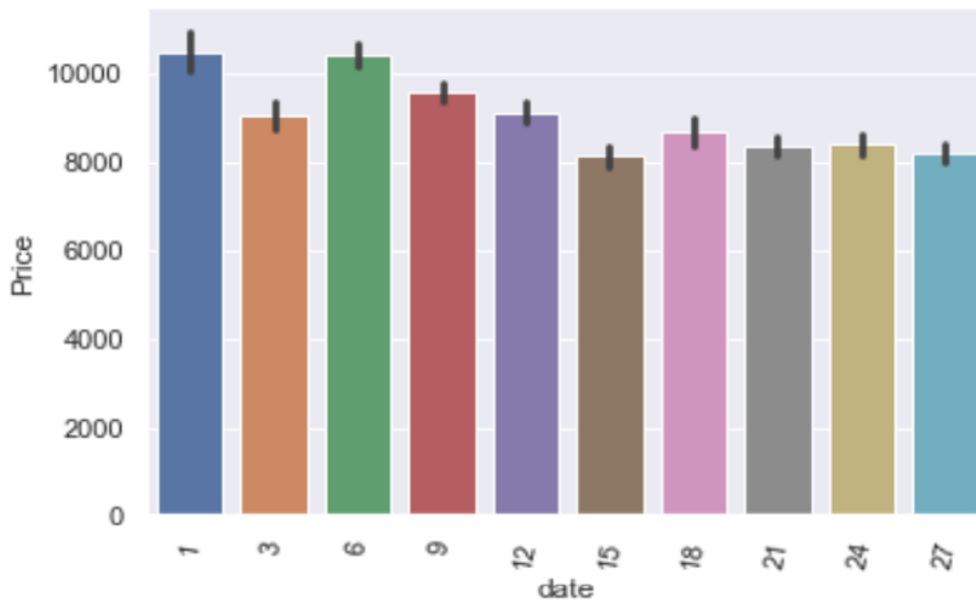


Fig. 10 The Analysis between Date and Price

It looks like that there's a trend in the plane price when contrasted to the day of respective months, prices are higher in the start of month, but this is not a trend if you see from the broader perspective as this might be due to various reasons. For e.g. the date of Journey is 10th March and people are booking towards 5th March or so, this will lead to higher flight prices. Prices increase as near you date of booking is to the date of journey. So, flight prices don't follow any pattern towards any time of the month. Through this analysis the advantage is that the passengers can know the which date and day they are travelling by this people can plan accordingly and book the tickets, which makes the passengers very convenient and organised without confusion.

VIII. CONCLUSION AND FUTURE SCOPE

Evaluating the algorithmic rule, a dataset is collected, pre-processed, performed data modelling and studied a value difference for the number of restricted days by the passengers for travelling. Machine Learning algorithms with square measure for forecasting the accurate fare of airlines and it gives accurate value of plane price ticket at limited and highest value. Information is collected from Kaggle websites that sell the flight tickets therefore restricting data which are often accessed. The results obtained by the random forest and decision tree algorithm has better accuracy, but best accuracy is predicted by decision tree algorithm as shown is the above analysis. Accuracy of the model is also forecasted by the R-squared value.

In Upcoming days when huge amount of information is accessed as in detailed information in the dataset, the expected results in future are highly correct. For further research anyone desire to expand upon it ought to request different sources of historical data or be a lot of organized in collection knowledge manually over amount of your time to boot, a lot of different combination of plane are going to be traversed. There is whole possibility that planes differ their execution ideas consisting characteristics of the plane. At last, it is curious to match our model accuracy with that of the business models accuracy offered nowadays.

REFERENCES

- [1] Bachis, E., & Piga, C. A. (2011). "Low-cost airlines and online price dispersion. *International Journal of Industrial Organization*", 29(6), 655–667. doi:10.1016/j.ijindorg.2011.02.006
- [2] Groves, W. and Gini, M., 2021. "A Regression Model For Predicting Optimal Purchase Timing For Airline Tickets.". Available at: <<https://conservancy.umn.edu/handle/11299/215872>>
- [3] T. Wohlfarth, S. Clemencon, F. Roueff and X. Casellato, "A Data-Mining Approach to Travel Price Forecasting," doi: 10.1109/ICMLA.2011.11.
- [4] T. Wang et al., "A Framework for Airfare Price Prediction: A Machine Learning Approach," doi: 10.1109/IRI.2019.00041.
- [5] V. H. Vu, Q. T. Minh and P. H. Phung, "An airfare prediction model for developing markets," doi: 10.1109/ICOIN.2018.8343221.
- [6] K. Tziridis, T. Kalampokas, G. A. Papakostas and K. I. Diamantaras, "Airfare prices prediction using machine learning techniques," doi: 10.23919/EUSIPCO.2017.8081365.
- [7] Y. Chen, J. Cao, S. Feng and Y. Tan, "An ensemble learning based approach for building airfare forecast service," doi: 10.1109/BigData.2015.7363846
- [8] T. Liu, J. Cao, Y. Tan and Q. Xiao, "ACER: An adaptive context-aware ensemble regression model for airfare price prediction," doi: 10.1109/PIC.2017.8359563.
- [9] Boruah A., Baruah K., Das B., Das M.J., Gohain N.B. (2019) "A Bayesian Approach for Flight Fare Prediction Based on Kalman Filter," https://doi.org/10.1007/978-981-13-0224-4_18
- [10] T. Wang et al., "A Framework for Airfare Price Prediction: A Machine Learning Approach," doi: 10.1109/IRI.2019.00041.
- [11] G.A. Papakostas, K.I. Diamantaras and T. Papadimitriou, "Parallel pattern classification utilizing GPU-Based kernelized slackmin algorithm," doi:10.1016/j.jpdc.2016.09.001
- [12] G. Francis, A. Fidato, and I. Humphreys, "Airport–airline interaction: the impact of low-cost carriers on two european airports," doi:10.1016/s0969-6997(03)00004-8
- [13] C. Koopmans and R. Lieshout, "Airline cost changes: To what extent are they passed through to the passenger?" doi:10.1016/j.jairtraman.2015.12.013
- [14] S. Lee, K. Seo, and A. Sharma, "Corporate social responsibility and firm performance in the airline industry: The moderating role of oil prices," doi:10.1016/j.tourman.2013.02.002
- [15] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines." Available: <http://arxiv.org/abs/1412.6980>
- [16] L. Breiman, "Random forests," *Machine Learning*, vol. 45, pp. 5-32, 2001. doi:10.1023/a:1010933404324
- [17] H. Drucker, C.J.C. Burges, L. Kaufman, A. Smola and V. Vapnik, "Support vector regression machines," doi:10.1016/s0165-0114(02)00514-6
- [18] K. S. Gerardi and A. H. Shapiro, "Does competition reduce price dispersion? new evidence from the airline industry," doi:10.1086/597328