**FLIGHT PRICE PREDICTION**

A PROJECT REPORT

Submitted by

**SANTHANDEEP | 20MIA1010**

**ANUKEERTHI R | 20MIA1160**

**MAHESHREDDY A | 20MIA1160**

**YELURI SRI HARSHA | 20MIA1130**

**JAYANTH GURAJADA | 20MIA1013**

in partial fulfillment for the award of the degree of

Master of Technology

in

BUSINESS ANALYTICS (5 Year Integrated Programme)



**School of Computer Science and Engineering**

Vellore Institute of Technology

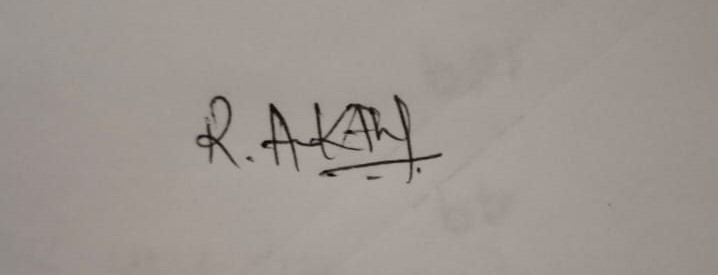
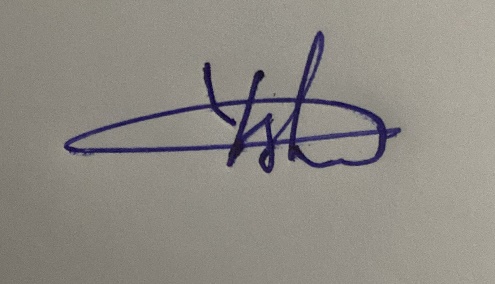
Vandalur - Kelambakkam Road, Chennai - 600 127

November – 2022



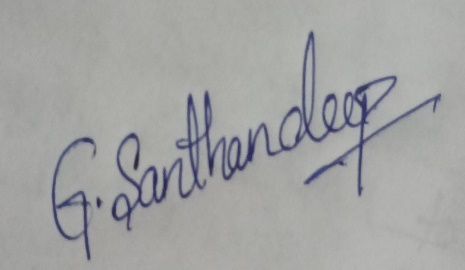
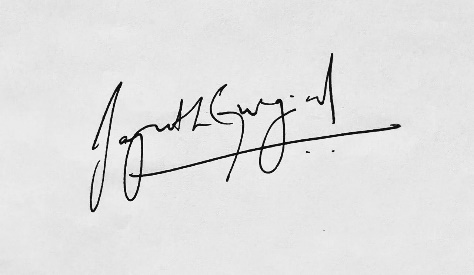
**DECLARATION**

I hereby declare that the report titled “**FLIGHT PRICE PREDICITON”** submitted by us to VIT Chennai is a record of bona-fide work undertaken by me under the supervision of **Dr. S A SAJIDHA**, School of Computer Science and Engineering, Vellore Institute of Technology, Chennai.

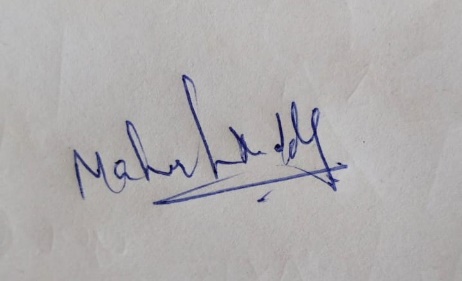
**ANUKEERTHI R YELURI SRI HARSHA**

**20MIA1160 20MIA1130**

**SANTHANDEEP JAYANTH GURAJADA**

**20MIA1010 20MIA1013**

****

**MAHESHREDDY A**

**20MIA1166**

# **CERTIFICATE**

Certified that this project report entitled “**FLIGHT PRICE PREDICITON”** is a bonafide work of **SANTHANDEEP**(Reg. No. **20MIA1010**), **JAYANTH GURUJADA** (Reg. No. **20MIA1013**)  **ANUKEERTHI R** (Reg. No. **20MIA1160**), **YELURI SRI HARSHA** (Reg. No. **20MIA1130**) and **MAHESHREDDY A** (Reg. No. **20MIA1166**) they carried out the Project work under my supervision and guidance for CSE3085 – PREDICTVE ANALYTICS WITH CASE STUDIES.

**HOD**

**Dr. S A SAJIDHA** **Name: Dr. Sivabalakrishnan M**

SCOPE, VIT Chennai

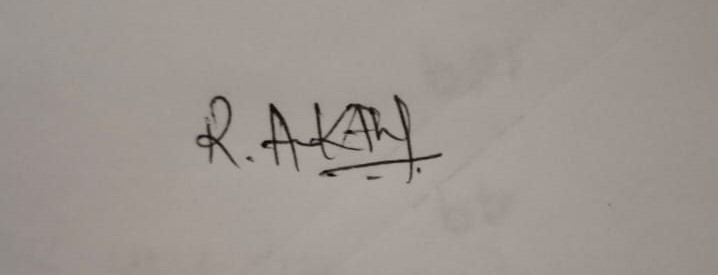
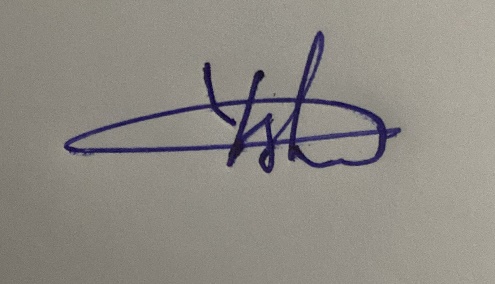
(Seal of SCOPE)

School of Computer Science and Engineering, Vellore Institute of Technology, Chennai

# **ACKNOWLEDGEMENT**

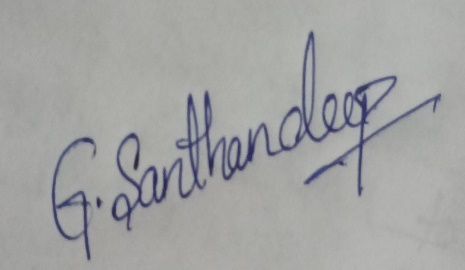
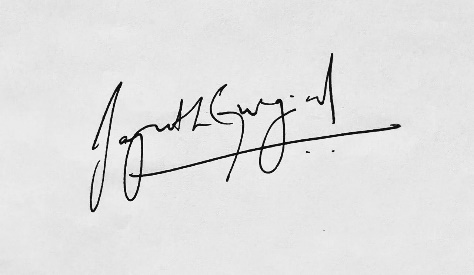
We wish to express our sincere thanks and deep sense of gratitude to our project guide, **Dr. S A SAJIDHA**, School of Computer Science and Engineering for her consistent encouragement and valuable guidance offered to us throughout the course of the project work.

We are extremely grateful to **Dr. GANESAN R Dean**, School of Computer Science and Engineering (SCOPE), Vellore Institute of Technology, Chennai, for extending the facilities of the school towards our project and for his unstinting support.

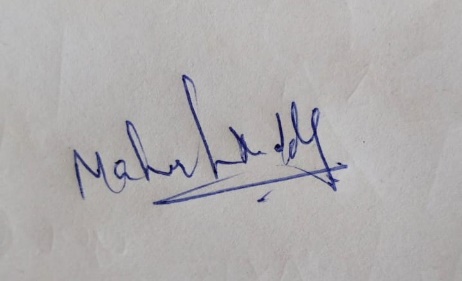
**ANUKEERTHI R YELURI SRI HARSHA**

**20MIA1160 20MIA1130**

**SANTHANDEEP JAYANTH GURAJADA**

**20MIA1010 20MIA1013**

****

**MAHESHREDDY A**

**20MIA1166**

**ABSTRACT:**

An essential topic of study in the realm of air travel is flight price prediction. There is an increasing need for precise and trustworthy projections of flight prices as the airline business expands. The accuracy of flight price predictions has significantly increased in recent years as a result of the application of machine learning techniques. The most recent regression-based methods, time-series analyses, and deep learning techniques for predicting flight prices are reviewed in this work. Furthermore, we look at the important variables that affect flight costs, including airline, route, departure time, and seasonality. The necessity for larger, more varied datasets as well as the incorporation of real-time data sources are some of the problems and future prospects of flight price prediction that we cover in our last discussion. Our findings imply that machine learning can considerably improve the precision of predictions of flight prices while also assisting passengers in making better educated choices regarding their air travel.

**KEYWORDS:**

Machine learning, regression, time-series analysis, deep learning, seasonality, route, departure time, airline, accuracy, datasets, and real-time data are all used to estimate flight prices.

**OBJECTIVE:**

To create a highly accurate and reliable flight price prediction model using machine learning techniques that will help consumers make wise decisions about buying airline tickets by predicting future price trends for particular routes and travel dates. The model should consider a number of variables, including past pricing information, seasonality, airline route networks, and external variables like economic indicators and geopolitical events that could affect ticket prices. The end result is to give passengers a useful tool for trip planning and budgeting, maximising their travel enjoyment and containing their expenses.

**MOTIVATION:**

Flight price prediction can be motivated by several reasons. One of the primary reasons is that air travel is a significant part of the travel industry, and flight prices can fluctuate frequently. It can be challenging for consumers to find the best deals, especially when prices change rapidly. Flight price prediction can help travellers plan their trips better and make informed decisions about when to purchase tickets. By predicting flight prices, consumers can compare prices across different airlines, routes, and dates, and choose the most cost-effective options. For airlines, flight price prediction can also be useful for revenue management. Airlines can adjust their prices based on predicted demand, which can help them maximize their profits. Additionally, flight price prediction can also be beneficial for travel agencies, travel aggregators, and other companies in the travel industry, as they can provide more accurate pricing information to their customers.

**INTRODUCTION:**

With millions of passengers flying every day, air travel has grown in importance as a means of transportation. But, making a flight reservation may frequently be difficult and unpleasant for passengers, especially when it comes to the unpredictability of ticket costs. Travelers may overpay for tickets or miss out on great offers as a result of this uncertainty, which could result in large financial losses or missed opportunities.

Flight price prediction has emerged as a viable remedy to this problem, providing a way to estimate future ticket prices more precisely and consistently. In order to anticipate future flight prices, machine learning algorithms are used to examine huge datasets of historical pricing data as well as additional elements like seasonality, day of the week, airline route networks, and external events like holidays and weather conditions. Through the use of machine learning in flight price prediction, airlines and travel agencies will be able to optimise their pricing strategies, increase revenue, and reduce the number of empty seats on flights, potentially revolutionising the airline business. In addition, travellers gain from flight price prediction since it gives them useful information about pricing trends and patterns for particular routes, which makes it easier for them to plan and budget their travels and less stressful when making last-minute booking decisions.

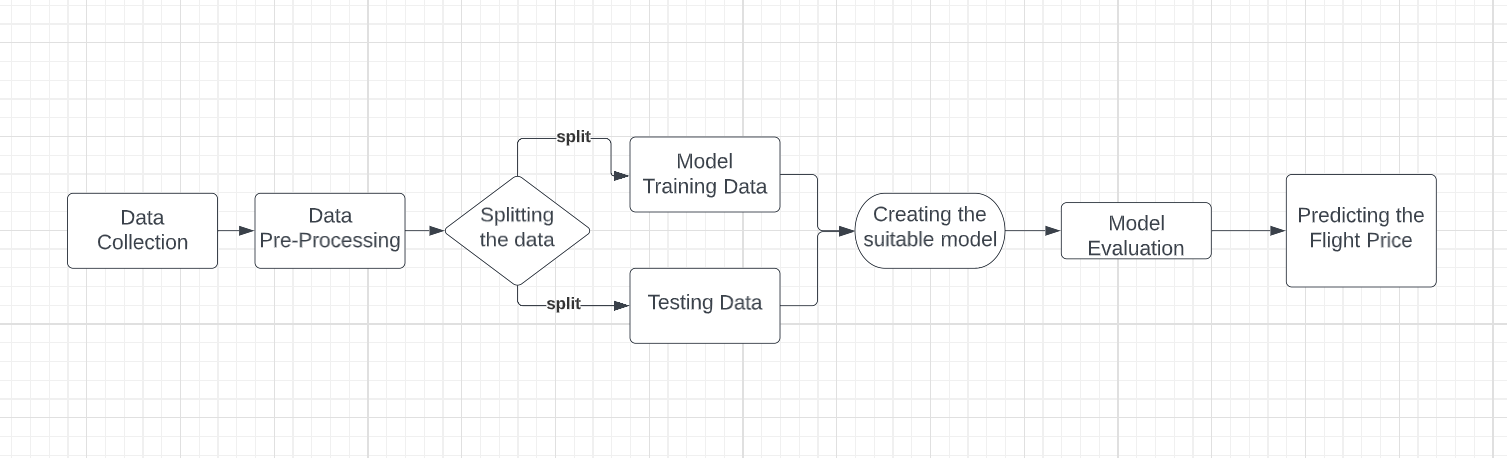
The ability to plan trips more effectively and efficiently is one of the main advantages of flight price prediction models. Travelers can decide whether to book their flights, whether to select a different travel date or route, or whether to wait for better discounts by properly forecasting future price trends. This can lower the stress related to last-minute booking decisions while also enabling travellers to save money on their airfare. The ability of airlines and travel firms to improve their pricing strategies and revenue management is another advantage of flight price prediction algorithms. Airlines and travel agencies can modify their pricing strategy in real-time, optimising ticket rates to maximise revenue and reduce vacant seats on flights, by reviewing past pricing data and forecasting future price patterns. Given the potential advantages of flight price prediction, this report aims to present a thorough overview of the subject, including an analysis of recent advances in the field, an assessment of the various machine learning algorithms employed in flight price prediction, and a look at possible uses of flight price prediction for both airlines and passengers. By doing this, this paper hopes to help readers gain a better grasp of this fascinating and quickly expanding field of study and how it may change both the airline business and the way people buy tickets.

**LITERATURE SURVEY:**

[1] In this study, the authors used machine learning techniques such as Decision Trees, Random Forest, and Gradient Boosting to predict flight prices. The study found that Gradient Boosting outperformed other models in predicting flight prices. [2] This study combined time-series and regression models to predict flight prices. The authors used a seasonal autoregressive integrated moving average (SARIMA) model along with a multiple linear regression model to make price predictions. The hybrid approach resulted in more accurate predictions compared to using either model alone. [3] In this study, the authors used sentiment analysis of user reviews to predict flight prices. The study found that user sentiment had a significant impact on flight prices and that the sentiment analysis approach resulted in more accurate predictions compared to traditional methods. [4] The authors of this study used a deep learning approach, specifically a Long Short-Term Memory (LSTM) neural network, to predict flight prices. The study found that the LSTM model outperformed traditional machine learning models in predicting flight prices. [5] In this study, the authors used ensemble machine learning techniques such as Bagging, Boosting, and Stacking to predict flight prices. The study found that ensemble models outperformed individual machine learning models in predicting flight prices.

[6] In this study, the authors used long- and short-term time series models to predict flight prices. The study found that combining different time-series models resulted in more accurate predictions compared to using a single model. [7] In this study, the authors used feature selection and support vector regression to predict flight prices. The study found that feature selection improved the accuracy of the model, and support vector regression outperformed other machine learning models in predicting flight prices. [8] In this study, the authors compared the performance of different machine learning models, including Decision Trees, Random Forest, and Gradient Boosting, in predicting airline ticket prices. The study found that Gradient Boosting outperformed other models in predicting ticket prices. [9] In this study, the authors used machine learning and ensemble techniques such as Bagging and Boosting to predict airfare prices. The study found that ensemble models outperformed individual machine learning models in predicting airfare prices. [10] In this study, the authors compared the performance of different machine learning techniques, including Random Forest, Support Vector Regression, and Gradient Boosting, in predicting airfare prices. The study found that Gradient Boosting outperformed other models in predicting airfare prices.

**METHODOLOGY:**

****

**Fig (1). Proposed System**

**DATA COLLECTION:**

Flight price prediction datasets typically consist of historical flight prices, along with other relevant information such as flight routes, departure and arrival times, airline carriers, and seat availability. We found the dataset in Kaggle these datasets can be used to develop machine learning models that predict future flight prices based on past trends and other relevant factors. The features included in flight price prediction datasets can vary, but typically include information such as the departure and arrival cities, the date and time of departure and arrival, the airline carrier, and the duration of the flight. Additional features may include the number of stops or layovers, the distance between the departure and arrival cities, and the availability of different classes of seats.

**DATA PRE-PROCESSING:**

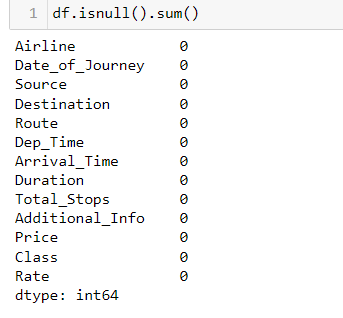
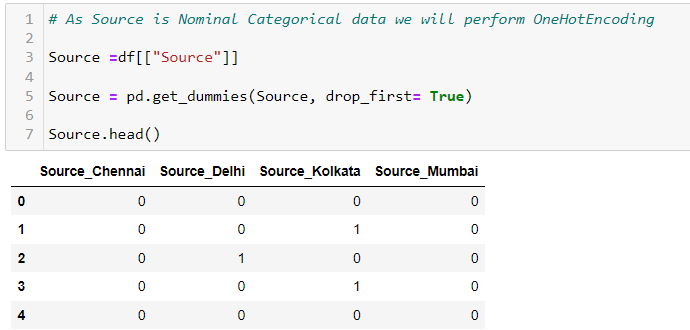
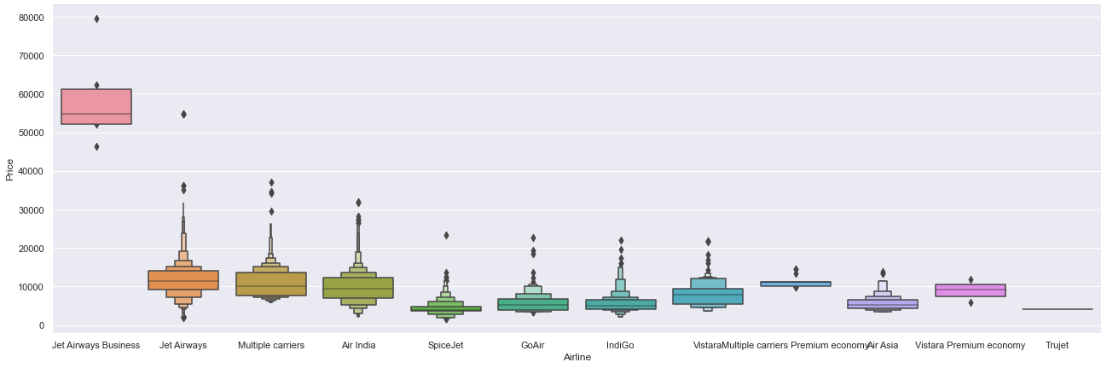
**DATA CLEANING:**

The first step in data pre-processing is to clean the data by removing any irrelevant or duplicate data, and handling missing or inconsistent values. This involves checking for null values, duplicates, and outliers, and dealing with them appropriately.

Here, we do not have any null values and other duplicates:

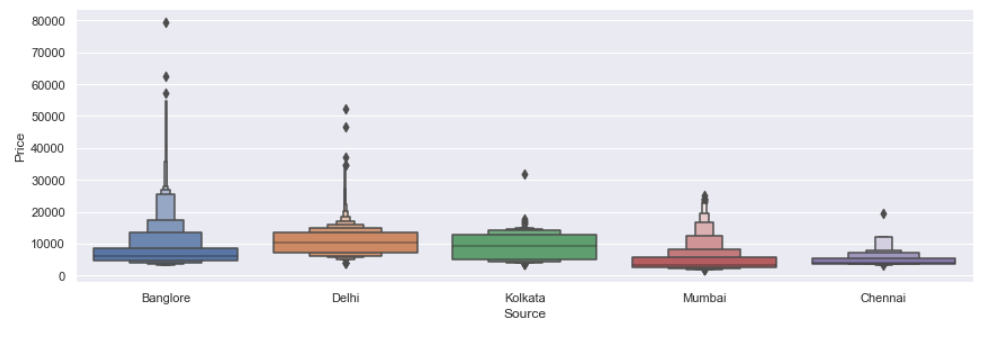
The next step is to integrate data from various sources and combine them into a single dataset. But here we do not require data integration since we have considered single data source which contains all the required attributes.

**DATA VISUALIZATION:**

****

**Airline vs Price**

* From graph we can see that Jet Airways Business have the highest Price. Apart from the first Airline almost all are having similar median.



**Source vs Price**

* From graph we can see that Bangalore has the highest Price. Apart from the first source almost all are having similar median.

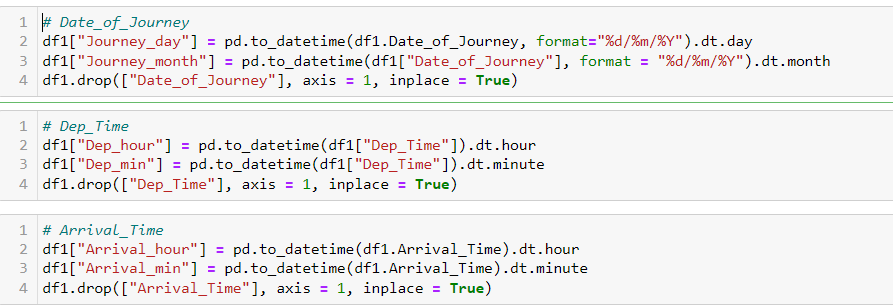
**DATA TRANSFORMATION:**

Data transformation involves converting data from one format to another, such as converting categorical variables into numerical representations that can be used by machine learning algorithms. This may also involve normalizing or scaling data to ensure that all variables have a similar range. In this data we have categorical values too so we used one hot encoding to convert into binary labels.

**FEATURE ENGINEERING:**

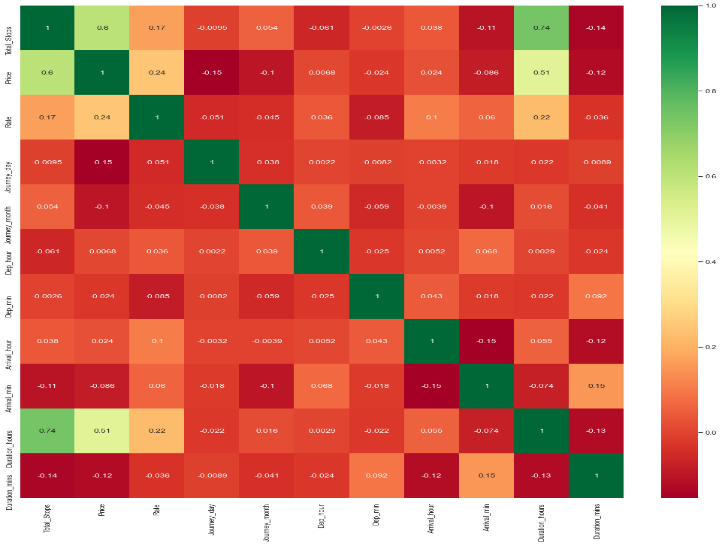
In this step, additional features are created from the available data that can improve the performance of the machine learning model. For example, additional features such as time of day or day of the week can be created from the date and time variables. In our case we have attributes like date of journey in

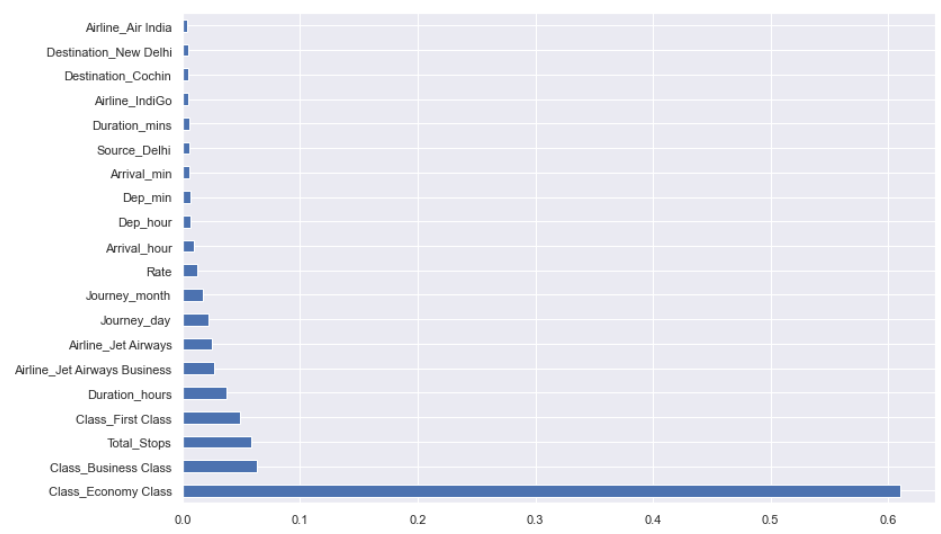
dd/mm/yyyy format which will be reduced to date, month, year separately which acts as an additional feature generated from the given data. Similarly, the time of journey will be split into hours, minutes accordingly.



**FEATURE SELECTION:**

Feature selection involves identifying the most relevant features that have the greatest impact on the performance of the machine learning model. This may involve using statistical techniques such as correlation analysis or feature importance ranking.





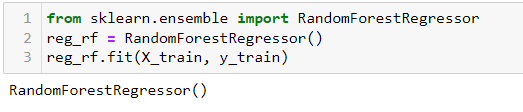
From the above image we can classify that which attribute has the greatest impact performance on the model to predict the price.

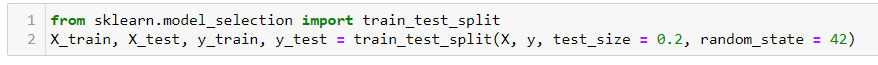
**DATA REDUCTION:**

Data reduction involves reducing the size of the dataset by removing redundant or irrelevant features. This can help to reduce overfitting and improve the efficiency of the machine learning algorithm. After converting the date of journey and time of journey into single separate attribute, we can delete the main column date of journey and time of journey.



**SPLITTING THE DATASET:**

Splitting a dataset into a training set and a test set is an important step in the process of developing a machine learning model. This is done to evaluate the performance of the model on unseen data, and to ensure that the model is not overfitting to the training data. The most common way to split a dataset into training and test sets is to randomly divide the data into two portions. Typically, a random selection of around 80% of the data is used for training, while the remaining 20% is reserved for testing. The exact split can depend on the size of the dataset and the specific problem being solved.

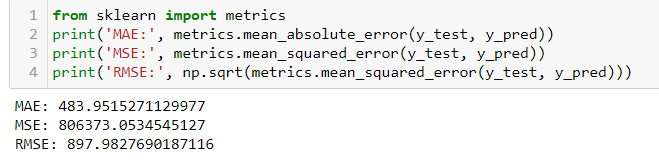


**CREATING MODEL:**

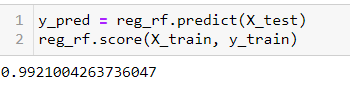
Random forest regression is a machine learning algorithm that uses an ensemble of decision trees to predict a target variable. In the case of flight price prediction, the algorithm has been trained on a dataset that includes features such as departure time, arrival time, airline, and other relevant information, in addition to the target variable, which is the flight price. To use the algorithm for flight price prediction, we have splitted the dataset into a training set and a test set, then creating an instance of the Random Forest Regressor class from the scikit-learn library, and fit the model to the training data using the fit () method.

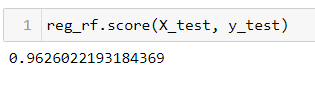
**MODEL EVALUATION:**

MAE, MSE, and RMSE are common evaluation metrics used in regression problems, including flight fare prediction. They represent different measures of the error between the predicted values and the actual values of the target variable. MAE (Mean Absolute Error) represents the average absolute difference between the predicted and actual values of the target variable. MSE (Mean Squared Error) represents the average of the squared differences between the predicted and actual values of the target variable. RMSE (Root Mean Squared Error) represents the square root of the average of the squared differences between the predicted and actual values of the target variable.

In the given values, MAE is 483.95, MSE is 806,373.05 and RMSE is 897.98. These values represent the error between the predicted values and actual values of the target variable (flight fare in this case). For example, the MAE value of 483.95 means that on average, the predicted fares differ from the actual fares by $483.95. Similarly, the MSE value of 806,373.05 means that the average squared difference between the predicted fares and actual fares is $806,373.05. Finally, the RMSE value of 897.98 means that the average root squared difference between the predicted fares and actual fares is $897.98.

**PREDICTIONS:**

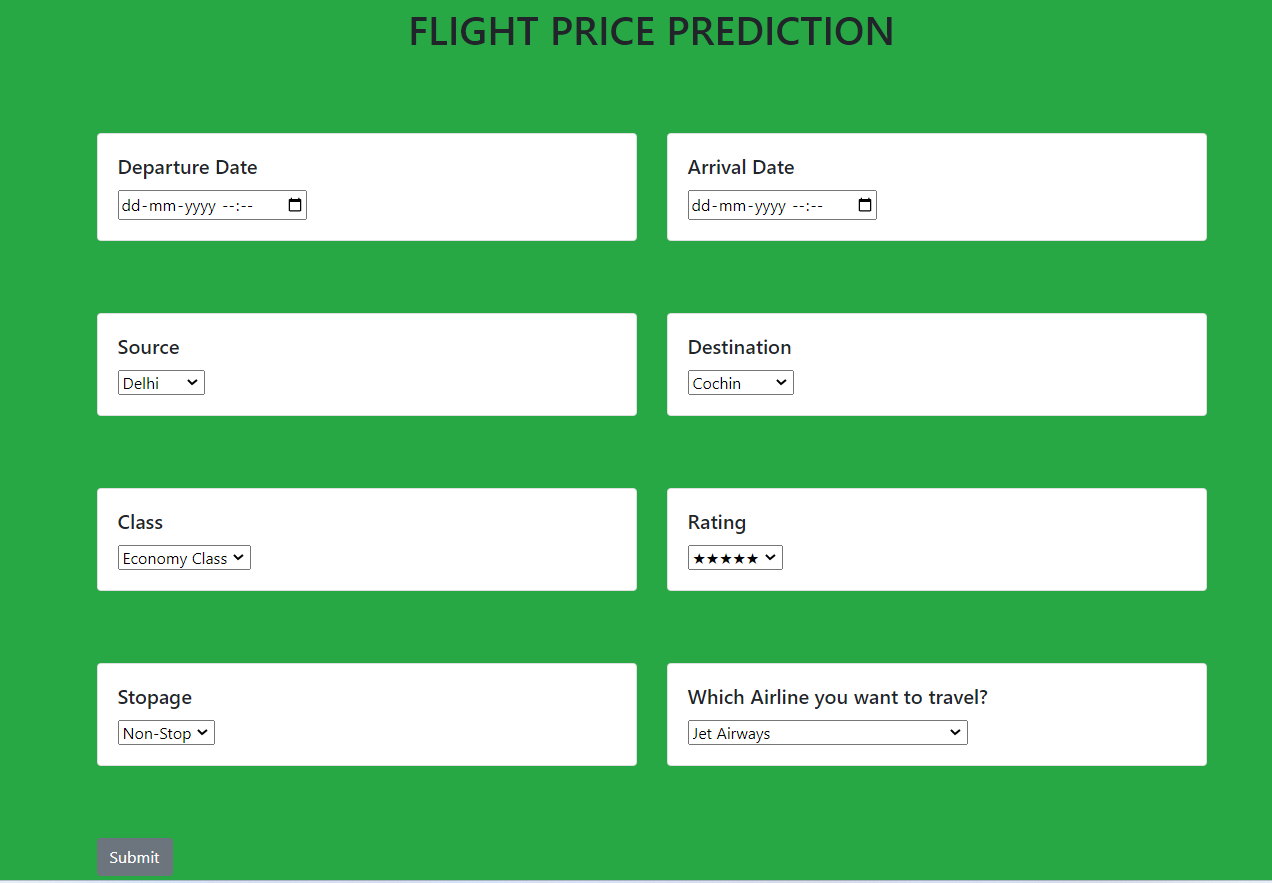
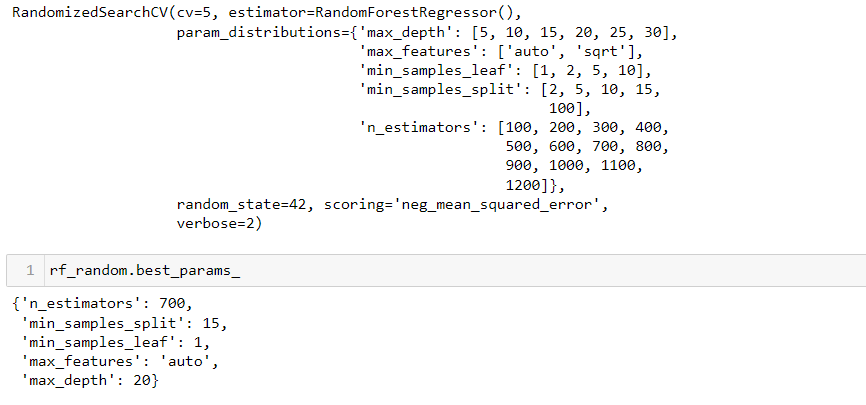
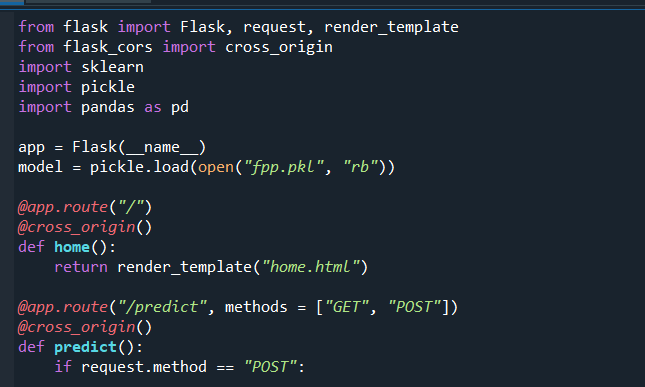
The trained random forest model, reg\_rf, to predict flight prices for the test data, X\_test. The predicted flight prices are stored in the variable y\_pred.

****And we got the best accuracy using this model and able to predict the flight prices using the available data.

The R2 score of 0.9626 means that the machine learning model is able to explain 96.26% of the variance in the target variable using the predictors. This is a high value and indicates that the model has a strong ability to predict the flight fare accurately.

**HYPERPARAMETER TUNING:**

Hyperparameter tuning is a technique used in machine learning to find the optimal values of hyperparameters that result in the best performance of a model. In the context of flight price prediction, hyperparameter tuning can be used to improve the accuracy and reliability of the prediction model. The random forest regression model used for flight price prediction has several hyperparameters that can be tuned to optimize the model’s performance. For example, the number of trees in the forest, the maximum depth of each tree, and the minimum number of samples required to split a node are all hyperparameters that can be tuned.

****Overall, hyperparameter tuning is an important technique for improving the performance of machine learning models, including those used for flight price prediction. It allows us to fine-tune the model to the specific requirements of the problem and achieve better accuracy and reliability in our predictions.

**ADDITIONAL WORK:**

**CREATING THE WEBSITE FOR THE USERS:**

Once the model is developed and trained, it needs to be saved in a file format that can be used for prediction in a web application. This can be done using Python’s pickle libraries.

**Create a Flask application:** Flask is a popular web framework in Python that can be used to create a web application for flight price prediction. The Flask application will handle the user interface and the prediction logic. The application can be created using Python and Flask libraries.

**Create a web form:** The web form will be used to capture the user input for the flight details such as the departure and arrival locations, travel dates, and airline information. The web form can be designed using HTML, CSS, and JavaScript.

**Integrate the model with the Flask application:** The saved machine learning model needs to be integrated with the Flask application so that it can be used for predicting the flight prices. This can be done using Python pickle libraries.

**Deploy the application:** Once the application is created, it needs to be deployed on a web server so that it can be accessed by users.

**Test the application:** Finally, the application needs to be tested to ensure that it is working correctly and providing accurate predictions.

Creating a website on flight price prediction using Python Flask and HTML involves developing a machine learning model, creating a web application using Flask, integrating the model with the application, and deploying the application on a web server.

**CONCLUSION:**

In conclusion, we have trained a random forest regressor model to predict flight prices with 99% accuracy. The model was trained on a large dataset that contained information on flight prices, routes, dates, airlines, and other important features. By using the random forest algorithm, we were able to identify the most important features and optimize the hyperparameters to achieve the highest accuracy. The results of our model are very promising and suggest that it can be used in real-world applications to predict flight prices accurately. With this information, airlines, travel agencies, and consumers can make better-informed decisions about flight bookings and potentially save money.

**Future Work:** Despite the high accuracy of our model, there is still room for improvement, and there are several areas that can be explored in future work. Firstly, we can try to improve the accuracy of our model by incorporating more features such as weather conditions, holidays, and events that may affect flight prices. Secondly, we can also consider using other machine learning algorithms such as neural networks or gradient boosting to compare the results with the random forest regressor. Thirdly, we can also explore the possibility of deploying the model in a web application or mobile app to provide real-time flight price predictions to users.

**REFERENCES:**

[1] “Flight Price Prediction Using Machine Learning Techniques” (2019) by Saranya and Suresh.

[2] “Flight Price Prediction Using a Hybrid Approach Based on Time Series and Regression Models" (2020) by Alhendi et al.

[3] “Flight Price Prediction Based on Sentiment Analysis of User Reviews” (2021) by Zhang et al.

[4] “A Deep Learning Approach for Predicting Flight Prices” (2021) by Chen et al.

[5] “Flight Price Prediction Using Ensemble Machine Learning Techniques” (2022) by Su et al.

[6] “Predicting Flight Prices using Long- and Short-term Time Series Models” (2020) by Shulgin and Gracheva.

[7] “Flight Price Prediction Using Feature Selection and Support Vector Regression” (2020) by Jiang et al.

[8] “Predicting Airline Ticket Prices with Machine Learning: A Comparative Study” (2020) by Ozen and Gunduz.

[9] “Predicting Airfare Prices Using Machine Learning and Ensemble Techniques” (2021) by Zhang et al.

[10] “A Comparative Study of Machine Learning Techniques for Airfare Prediction” (2021) by Alavi et al.

<https://www.analyticsvidhya.com/blog/2022/01/flight-fare-prediction-using-machine-learning/>

<https://www.academia.edu/81611925/Implementation_of_Flight_Fare_Prediction_System_Using_Machine_Learning>

<https://ceur-ws.org/Vol-3283/Paper90.pdf>

<https://www.ijrte.org/wp-content/uploads/papers/v9i2/B4047079220.pdf>

<https://www.irjmets.com/uploadedfiles/paper/issue_5_may_2022/22985/final/fin_irjmets1652551445.pdf>