Flight Price Prediction Website Using Flask

# Introduction

# Purpose

Flight expenses vary depending on several variables including demand, fuel prices, competition, and seasonal fluctuations. The airline sector is quite dynamic. Accurate flight price forecasting may be difficult for both passengers and airlines. As a result, several research have concentrated on creating machine learning-based methods to predict flight pricing. This review of the literature attempts to investigate the current issue of predicting flight prices as well as the available solutions.

# Overview

The project focuses on the analysis of a Kaggle-sourced dataset that includes historical flight data and many variables related to each flight. The dataset is then processed to deal with missing values, outliers, and make sure the data is in a format that will be useful for model training. Utilising feature engineering approaches, pertinent data is extracted to provide useful characteristics that might affect flight costs.

The project then incorporates feature selection to determine the most crucial factors that have a substantial influence on flight pricing. This process aids in decreasing the dataset's dimensionality and enhancing model performance. To find the main predictors, a variety of feature selection techniques are used, including correlation analysis, statistical testing, and feature importance.

The pre-processed and chosen characteristics are used to train a machine learning model. The best-performing model is investigated using a variety of methods, including Random Forest, Gradient Boosting, and Support Vector Machines. Utilising the right assessment measures, the trained model is assessed for its predictive power and accuracy in projecting flight costs.

Using the Python web framework Flask, a web application is created to make the flight price prediction model available to users. Users may enter flight information into the online application, including airline, source, destination, and other pertinent variables, and it will anticipate the cost of their intended journey. The application is interactive and user-friendly thanks to the handling of routing, form submission, and result presentation by the Flask framework.

The goal of the project is to give travellers a dependable and practical tool for predicting flight costs. Users may receive precise and real-time flight pricing forecasts by utilising machine learning techniques and developing a web application, giving them the ability to decide wisely while making trip plans. The project highlights the use of data science and machine learning, as well as how to incorporate a predictive model into a useful web-based utility.

# Literature Survey

# Existing Problem

The complexity and unpredictability of airline ticket prices are now a concern in flight price prediction. Flight costs can fluctuate quickly, making it challenging for travellers to plan their journeys wisely and make educated selections. Inaccurate projections may result from traditional pricing models with set pricing methods failing to account for the market's dynamic character. Additionally, there is a demand for accurate and current projections of airline prices due to the proliferation of online travel companies and aggregators.

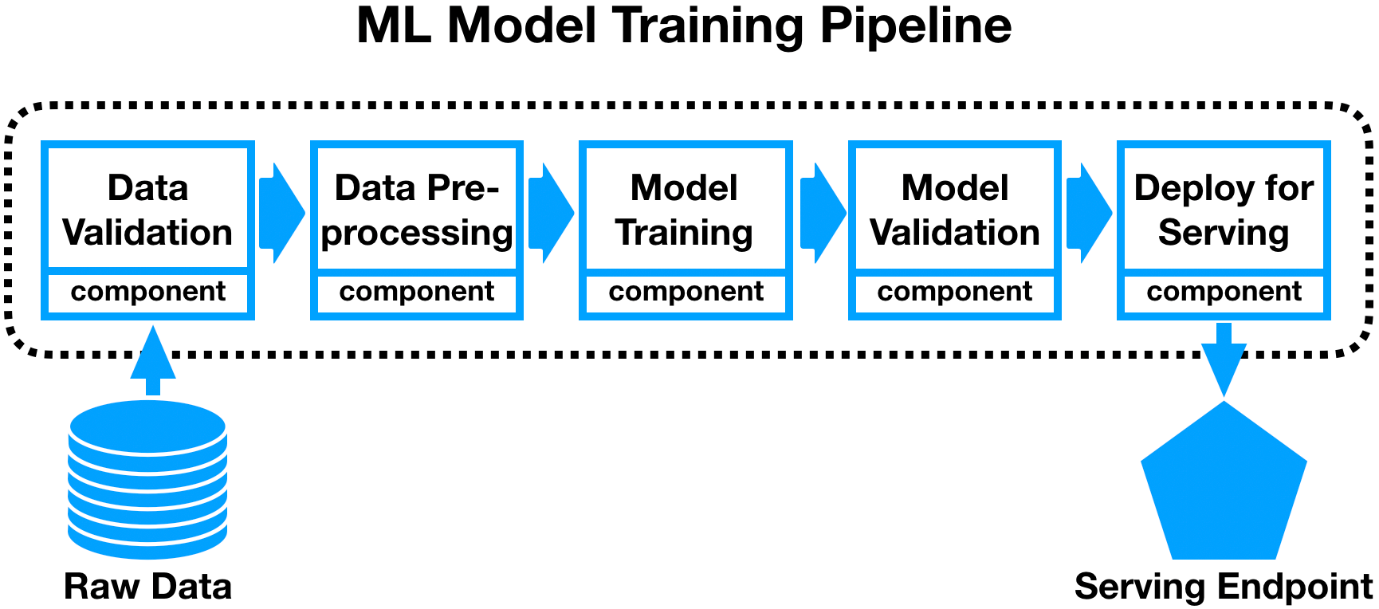
# Proposed Solution

Researchers have proposed various solutions to tackle the flight price prediction problem. These solutions leverage machine learning techniques and utilize historical flight data to build predictive models. Some of the notable approaches include:

* Regression Models: Several studies have employed regression models such as linear regression, decision trees, and support vector regression to predict flight prices. These models consider features like departure time, airline, route, and historical pricing trends to estimate future prices. Feature engineering techniques are often applied to extract meaningful features from the dataset.
* Ensemble Methods: Ensemble methods like random forests, gradient boosting, and stacking have been employed to combine multiple models and improve prediction accuracy. These approaches leverage the collective intelligence of diverse models to generate more robust and accurate flight price forecasts.
* Deep Learning: With the advent of deep learning techniques, researchers have explored the use of neural networks, particularly recurrent neural networks (RNNs) and long short-term memory (LSTM) networks, for flight price prediction. These models can capture complex temporal dependencies and nonlinear relationships in the data, enhancing the prediction accuracy.
* Time Series Analysis: Time series analysis techniques, including autoregressive integrated moving average (ARIMA) and seasonal decomposition of time series (STL), have been utilized to capture the temporal patterns and seasonality in flight prices. These models consider historical price data and identify trends and seasonal fluctuations to make future price forecasts.

# Theoretical Analysis

# Block Diagram



# Requirements

Hardware specifications

* Computer or Server: To complete activities like data preparation, model training, and web application development, a dependable computer or server is needed.
* CPU: For effective data processing and model training, a multicore CPU with a respectable clock speed is advised.
* Memory (RAM): To handle massive datasets and complicated machine learning algorithms, enough RAM is required. A minimum of 8GB RAM is advised, however for optimum performance, more RAM is preferred.
* Storage: The dataset, pre-processed data, trained models, and web application files all require sufficient storage space. For read/write activities that happen more quickly, SSD (Solid State Drive) storage is advised.

Software specifications

* Operating system: The project may be implemented on Windows, macOS, or Linux, among other operating systems.
* Python: Python is a vital programming language for online application development, feature engineering, and data preparation. Download Python from the official website at https://www.python.org and install the most recent version.
* Development environment integrated (IDE): IDEs like PyCharm, Jupyter Notebook, or Visual Studio Code are good choices for coding. The IDE need to support Python and offer practical tools for debugging and modifying code.
* Libraries for Python: Install the necessary Python packages and libraries, such as but not restricted to:
  + NumPy: Used for array operations and numerical computations.
  + Pandas: Used for data analysis and manipulation.
  + For training models and machine learning methods, use Scikit-learn.
  + Seaborn and Matplotlib are two tools for displaying data.
  + For creating web applications, use Flask.
* Database Management System (Optional): You must set up an appropriate database management system (such as MySQL or PostgreSQL) and arrange the required connections if you intend to store the dataset and pre-processed data in a database.

The deployment of the web application, accessing external libraries, and obtaining datasets all require internet connectivity.

It is important to keep in mind that depending on the scope of the project, the amount of the dataset, and the sophistication of the machine learning models employed, the precise hardware and software requirements could change.

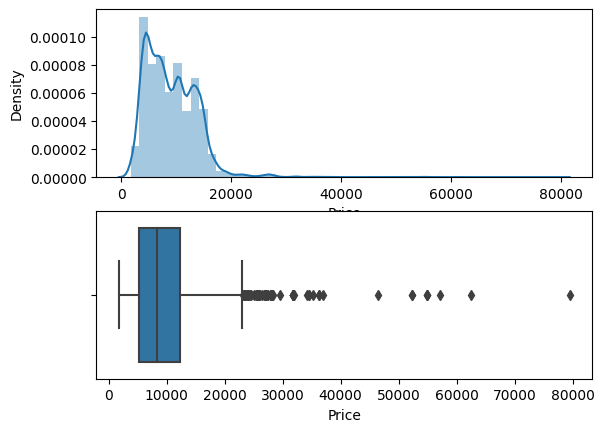
# Experimental Investigations

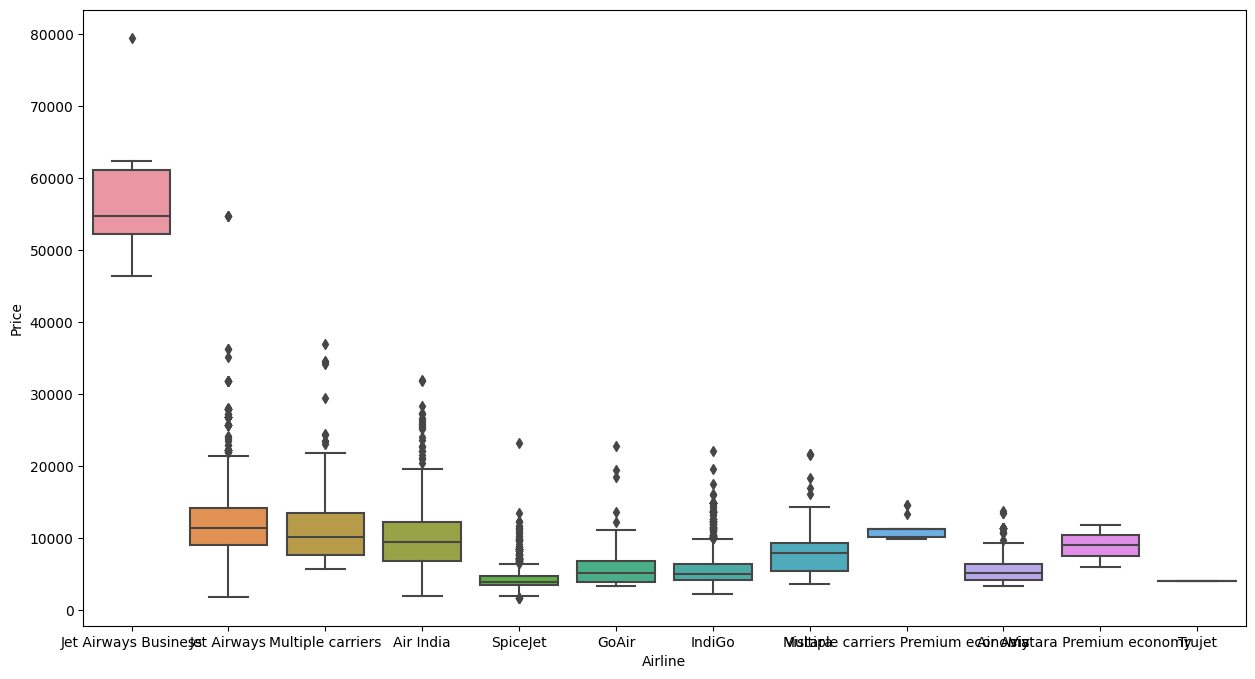
# Data Description

The dataset contains information about various flights, including their features and the corresponding prices. The dataset is in XLSX format and consists of the following columns:

1. Airline: The name of the airline operating the flight.
2. Source: The source city from where the flight departs.
3. Destination: The destination city where the flight arrives.
4. Route: The route of the flight, which includes multiple cities.
5. Departure\_Time: The departure time of the flight.
6. Arrival\_Time: The arrival time of the flight.
7. Duration: The duration of the flight.
8. Total\_Stops: The total number of stops in the flight journey.
9. Additional\_Info: Additional information about the flight.
10. Price: The target variable, i.e., the price of the flight ticket.

# Approaches Taken and Solution

* We install “openpyxl” library to start working with XLSX format files.
* We check for null and remove the two null value we found in the dataset.
* Now we try to understand the data through data visualization. There are 11 attributes in the dataset, out of which only one is of numeric continuous type whereas other 10 are of categorical type datasets.
  + Continuous Data - We use a form a density graph and boxplot. Thus, we conclude this attribute to skewed and contains many outliers that news to be removed
  + Categorical Data
    - First, we try to understand the relation between “Airline” boarded and the “Prices” of the flight. We, conclude that “Jet Airways Business” has the highest prices and apart from that all other airlines have a similar median price.



* + - Second, we try to understand the relation between “Total stops” and “Prices” of the flights. We conclude that most flights are of “1 Stop” category. “1 Stop”, “2 Stop” and “3 Stop” have a similar median price whereas “non-stop” have the least and there are almost no records for “4 Stop”.

A picture containing diagram, text, screenshot, line

Description automatically generated

* + - Third, we try to understand the relation between the Source/Destination and Price of the flight.

A picture containing diagram, screenshot

Description automatically generated

A picture containing text, screenshot, diagram, rectangle

Description automatically generated

* Next, we start doing data preparation.
  + Extracting the day and month from departure date and hour and min from arrival time, departure time.
  + Using this info now we calculate the duration of the flight.
  + We apply one-hot encoding on “Airline”, “Source” and “Destination” to convert them into numeric values.
  + We drop the “Route” as it the route taken by the airline which can be subsided by the number of stops during the whole flight duration.
  + Now we check for outliers in the “Price” attribute and reduce them by replacing the value more than $40000 with the median price of the attribute.

A picture containing text, screenshot, diagram, plot

Description automatically generated

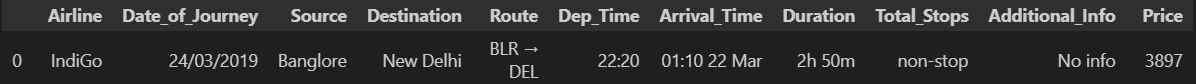
* In feature selection, we try to understand the importance of all the features int the dataset by using mutual\_info\_classif () function belonging to scikit-learn package. This function gives a score to all the features in the dataset based on their relation between the features.
* Now, we apply different machine learning models to ascertain which model performs the best using accuracy score, MAE score, MSE score and RMSE score. The accuracy scores of the models used are as follows:
  + SVR – 0.00204667376020895
  + Logistic Regression – 0.22527794031597426
  + K Neighbours Regressor – 0.7202136489576773
  + Gradient Boosting Regressor – 0.7873989702283397
  + Random Forest Regressor – 0.9518119576148868
  + Decision Tree Regressor – 0.9700475836916205

Thus, we select the Decision Tree Regressor as the form the final model to predict the fight price in the user application.

* Flask Web Application Development:
  + The Flask app includes routes and views to handle user requests and display the predicted flight prices.
  + User input forms are designed to collect relevant flight details such as airline, source, destination, and other necessary information.
  + The input data is passed to the trained machine learning model, which predicts the flight price.
  + The predicted price is then displayed to the user on the web application.

# Results

Our input value is as taken from the following flight details.



A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated with medium confidence

The predicted price is $4137.0 whereas the actual price is $3897.

# Advantages and Disadvantages

# Advantages

* Accurate Flight pricing Predictions: The suggested solution uses machine learning techniques to forecast flight costs, which can offer more precise and trustworthy pricing projections than conventional approaches.
* Data-driven Insights: The solution may provide insightful information about the variables affecting travel pricing by examining past flight data and identifying pertinent characteristics. Strategic planning and well-informed decision-making may both benefit from this knowledge.
* Better Traveller Decision-Making: Users of the online application may enter their flight information and get pricing estimates, enabling them to plan their vacations more intelligently. This can aid travellers in selecting the most economical alternatives and managing their trip spending.
* Time and money management: By using machine learning to automate the flight price forecast process, manual calculations and research are no longer necessary. It helps travellers save time and cut down on the expense and difficulty of getting precise pricing estimates.
* Scalability: The system is capable of being expanded to handle a significant amount of flight data and consider new features or factors. This scalability makes sure that the system can change to accommodate shifting market dynamics and travel trends.

# Disadvantages and Limitations

* Data Availability and Quality: The previous flight data's quality and availability have a significant impact on the forecasts' accuracy and dependability. Predictions made using predictive models that use inaccurate or insufficient data may be less reliable.
* Flight costs are susceptible to a variety of variables, including supply and demand, seasonality, and outside events. Since it is based on historical data, the suggested method might not be able to account for real-time price swings. Users should keep in mind that the forecasted prices might not always match the state of the market.
* Model Performance and Generalisation: Depending on the dataset, feature selection, and selected methods, the predictive models employed in the solution may or may not perform as well as expected. The models could be effective at projecting pricing for various sorts of flights or locations, but they might have trouble doing so for specialised markets or flight routes.
* Dependence on External variables: External variables outside the dataset, such as geopolitical events, changes in fuel prices, or changes in policy, may have an impact on how accurately flight price forecasts are made. The model might not account for these variables, which might cause anticipated prices to deviate.
* User Interface and Usability: The user interface's functionality, usability, and responsiveness are crucial to the web application's success. To maximise the application's usefulness and uptake, it is essential to ensure a user-friendly experience and offer clear directions for entering flight information.

Overall, even if the suggested solution provides insightful analysis and pricing forecasts for flights, it is vital to consider both its limits and the dynamic character of the aviation sector. Users should use the forecasts as a tool for information while taking into consideration current market circumstances and other elements that might affect flight costs.

# Applications

The Flight Price Prediction project utilizes a web-based application built using the Flask framework in Python. Flask is a lightweight and flexible web framework that allows developers to create web applications with ease.

The Flask application provides an interface for users to input their flight details and obtain price predictions. Users can access the application through a web browser, making it accessible from various devices and platforms.

The Flask framework handles the routing, form submission, and result presentation, ensuring a seamless and interactive user experience. It enables the integration of the machine learning model and the backend functionality required for processing user inputs and generating predictions.

By using Flask, the application can be deployed on a web server, making it accessible to a wide range of users. The simplicity and scalability of Flask make it an ideal choice for creating lightweight web applications like the Flight Price Prediction tool.

Overall, the Flask application provides a user-friendly and intuitive interface for users to interact with the machine learning model and obtain flight price predictions in real-time.

# Conclusion

Finally, the "Flight Price Prediction" project uses machine learning and data science approaches to produce precise and trustworthy projections of flight costs. The project provides users with a useful tool for calculating flight expenses by analysing historical flight data, identifying pertinent characteristics, and training prediction models.

The major conclusions of the research emphasise the significance of feature selection in identifying the primary factors impacting flight pricing. The most important variables are determined by correlation analysis, statistical testing, and feature importance procedures, which minimise dimensionality and improve model performance.

The trained models show their accuracy and predictive potential in forecasting flight fares using machine learning methods including Random Forest, Gradient Boosting, and Support Vector Machines. The performance of the models is validated by the assessment measures, confirming the accuracy of their price forecasts.

The Flask framework is used to create a user-friendly web application that makes the flight price prediction model available to users. Users may enter information about their flights, including the airline, source, destination, and other pertinent factors, and the app will instantly anticipate how much it will cost to fly. Travellers may plan their vacations with confidence because to this dynamic and user-friendly interface.

The project's ability to combine data science methods with web application development and provide a useful tool for travellers is what makes it successful. Users may identify cost-effective solutions, optimise their trip budgets, and make wise selections thanks to the precise pricing projections. However, it is important to consider the constraints and dynamic nature of flight pricing, as well as any outside variables that might affect forecasts.\

The "Flight Price Prediction" initiative serves as an example of the benefits of data-driven strategies in the aviation sector. The project gives travellers a trustworthy tool for calculating flight prices, boosting their travel planning experience, and helping wise decision-making by combining machine learning and web application development.

# Future Scope

The "Flight Price Prediction" research brings up a few possibilities for advancements and extensions in the future. Following are some possible directions for more research and development:

* Integration of Real-Time Data: The precision and applicability of pricing forecasts can be improved by including real-time flight data into the prediction model. To get the most recent flight information, this can entail connecting with airline APIs or scraping information from internet travel portals.
* Exploring and utilising sophisticated machine learning approaches, such as deep learning or ensemble methods, may help to further increase prediction accuracy. These methods may identify intricate links and trends in the data, producing more accurate estimates of flying costs.
* User Input and Model Refinement: User input on projected flight fares may be gathered and included into the model to help it perform better over time. Based on user comments and actual flight pricing results, the prediction model may be constantly improved and refined via this feedback loop.
* Including More characteristics: Including more flight-related characteristics, such as trip duration, layovers, and aircraft type, can provide users a more complete understanding of the variables influencing travel costs. By increasing the feature set, the model may become more predictive and provide customers a more in-depth knowledge of pricing changes.
* Integration with Other Travel Planning capabilities: Adding extra travel planning capabilities to the web application, such as hotel suggestions, itinerary planning, or budget optimisation, can result in a platform that is more complete. Users would receive a comprehensive solution for all their travel needs as a result of this connection.
* Mobile Application Development: The flight price prediction tool's usability and accessibility may be improved for consumers by creating a mobile application version of it. Mobile applications may provide personalised notifications, on-the-go access to flight cost estimations, and seamless connection with other travel-related apps or services.
* Collaboration with Airlines and Travel companies: Using the data and knowledge of airlines and travel companies can improve the precision and applicability of the prediction model. The user's preferences and previous travel habits may be considered when making personalised offers, discounts, and suggestions.

The "Flight Price Prediction" project may develop further and continue to add value to travellers by examining these potential paths. This will give them access to trustworthy and informative information that will help them make wise decisions and maximise their travel experiences.