Flight Price Prediction

Project Report

Submitted by

Pranshu Chadda (20BKT0092)
J Preethan (20BKT0008)
Areeb Kanth (20BKT0045)
Vishal Varma(20BDS0309)

For the course

Applied Data Science

June 2023

Table of Contents

1.Introduction

1.1Overview

Our project focuses on developing a flight price prediction system. The goal is to create a tool that can forecast future airline ticket prices and assist travellers in making informed decisions about their flight bookings. To accomplish this, our project will utilize historical flight data, including ticket prices, booking patterns, and other relevant factors. We will gather data from various sources, such as airlines, travel agencies, and online travel platforms, to create a comprehensive database for analysis.
Machine learning algorithms will be employed to analyse the collected data and identify patterns, trends, and correlations that can help predict future price movements. Techniques such as regression analysis will be used to develop accurate prediction models.

1.2 Purpose

The purpose of the project flight price prediction is to provide travellers with a valuable tool that can assist them in making informed decisions about their flight bookings. The project aims to achieve the following objectives:

1. **Forecasting Future Prices:** The primary purpose is to develop accurate prediction models that can forecast future airline ticket prices. By analyzing historical data, market conditions, and other relevant factors, the system aims to provide travellers with insights into whether prices are likely to rise or fall in the future. This information empowers travellers to make well-informed decisions about when to book their flights, potentially saving them money on airfare.

2. **Helping Users Find the Best Deals**: By providing predictions on future price movements, the project aims to help users find the most cost-effective time to book their flights. Users can leverage this information to identify periods of low prices or anticipate price increases, allowing them to secure the best deals and potentially save money on their travel expenses.

3. **Enhancing Travel Planning:** The flight price prediction system aims to enhance the overall travel planning experience for users. It provides them with insights into price fluctuations, allowing for more strategic and efficient travel arrangements. Additionally, features such as price alerts can help users stay updated on changing prices and take advantage of favourable booking opportunities.

4. **Improving Decision-Making:** The project seeks to empower travellers by providing them with data-driven insights for decision-making. Instead of relying solely on guesswork or generic advice, users can access personalized predictions based on their specific travel requirements. This helps them make more informed decisions tailored to their needs, preferences, and budget.

5. **Simplifying the Booking Process:** By integrating the flight price prediction system into a user-friendly application or website, the project aims to simplify the flight booking process. Users can conveniently access price predictions, compare different options, and make bookings all within a single platform, enhancing the overall user experience.

2.Literature Survey
2.1Existing Problem

There are several existing approaches and methods used to solve the problem of flight price prediction. Here are some commonly employed techniques:

1. Historical Data Analysis: One of the fundamental approaches is to analyse historical flight data, including ticket prices, booking patterns, and other relevant factors. Statistical analysis techniques, such as regression analysis and time series analysis, can be applied to identify patterns, trends, and correlations in the data. Historical data serves as a foundation for training predictive models.

2. Machine Learning Algorithms: Machine learning techniques are commonly used to develop accurate flight price prediction models. Algorithms such as linear regression, random forests, support vector machines (SVM), and neural networks can be applied to analyse historical data and learn from patterns to make predictions. These algorithms can consider multiple variables and factors simultaneously, allowing for a comprehensive analysis of the complex nature of flight pricing.

3. Feature Engineering: Feature engineering involves selecting and creating relevant features from the available data that can improve the prediction accuracy. For flight price prediction, features such as departure/arrival dates, flight duration, airline, route popularity, and seasonality can be used to enhance the models' performance. Feature engineering helps capture the underlying patterns and relationships in the data.

4. Market Indicators and External Factors: Flight prices can be influenced by various market indicators and external factors. These factors can include fuel costs, competition among airlines, economic indicators, exchange rates, and major events or holidays. Incorporating these external factors into the prediction models can improve the accuracy of price forecasts.

5. Ensemble Methods: Ensemble methods combine the predictions of multiple models to obtain a more robust and accurate prediction. Techniques such as bagging, boosting, and stacking can be used to combine the predictions of different models, leveraging their individual strengths and compensating for their weaknesses. Ensemble methods often result in improved prediction performance.

6. Real-time Data Integration: To account for dynamic market conditions, integrating real-time data into the prediction models can be beneficial. Real-time data sources can include current ticket prices, availability, and demand. By incorporating real-time data, the models can adapt to changing market conditions and provide more up-to-date and accurate predictions.

7. Online Travel Platforms and APIs: Many online travel platforms and airline websites offer APIs (Application Programming Interfaces) that provide access to flight data and pricing information. These APIs can be utilized to gather the necessary data for prediction models, enabling real-time analysis and accurate price forecasts.

2.2Proposed Solution

The method followed:

1. Data Collection: Gather historical flight data from various sources, including airlines, travel agencies, and online travel platforms. The data should include information such as ticket prices, booking patterns, route popularity, and other relevant factors.

2. Data Preprocessing: Clean and preprocess the collected data. This may involve handling missing values, removing outliers, normalizing numerical features, and encoding categorical variables. Additionally, feature engineering techniques can be applied to extract meaningful features from the data.

3. Feature Selection: Select the most relevant features that have a significant impact on flight prices. Consider factors such as departure/arrival dates, flight duration, airline, route popularity, seasonality, and any external factors that influence prices.

4. Model Training: Utilize machine learning algorithms such as linear regression, random forests, or neural networks to train the prediction models. Split the pre-processed data into training and testing sets. The models learn from the training data and aim to capture the relationships between the features and flight prices.

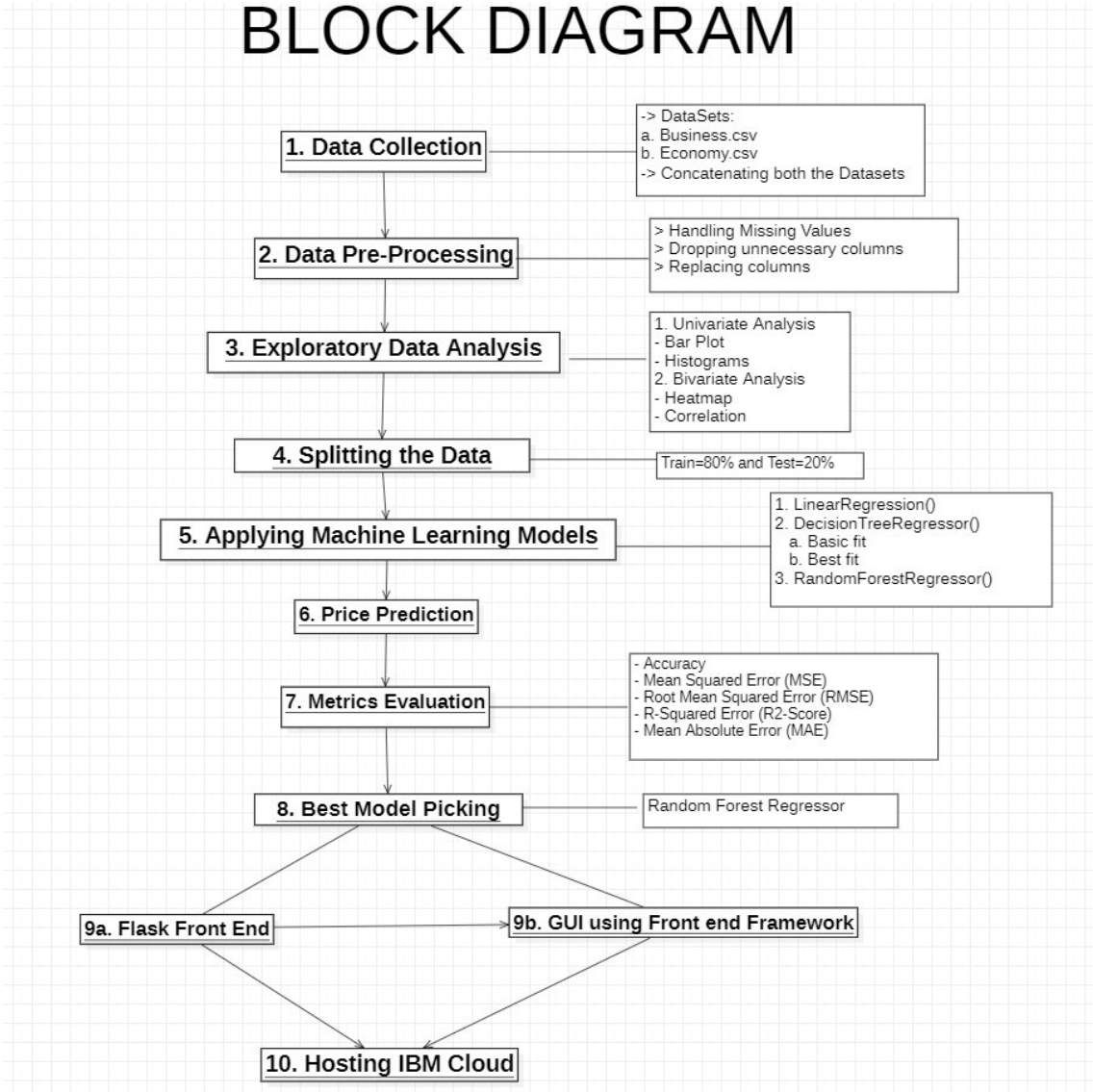
5. Model Evaluation: Evaluate the performance of the trained models using appropriate evaluation metrics such as mean squared error (MSE), root mean squared error (RMSE), or mean absolute error (MAE). This helps assess the accuracy and effectiveness of the models in predicting flight prices.

6. Hyperparameter Tuning: Fine-tune the model's hyperparameters to optimize its performance. Techniques like cross-validation or grid search can be employed to find the best combination of hyperparameters that minimize prediction errors.

7. Deployment and User Interface: Develop a user-friendly application or website where users can input their travel details and obtain flight price predictions. The interface should provide clear and actionable information, allowing users to make informed decisions about their flight bookings.

8. Continuous Model Improvement: Monitor the performance of the deployed prediction system and collect feedback from users. Periodically retrain the models with updated data to improve prediction accuracy and adapt to changing market conditions.

3.Theoretical Analysis
3.1Block Diagram



3.2 Hardware / Software designing.

1. Operating System:
   Windows
   macOS
   Linux
   Mobile Platforms (iOS, Android)

2. Development Tools:
   Integrated development environments (IDEs)
   Visual Studio (VS)
   Jupyter Notebook
   GitHub

3. Programming Language:
   Python
   HTML
   CSS

4. Data collection and storage in "csv" form

5. Frameworks and Libraries:
   Pandas
   NumPy
   Matplotlib
   Seaborn
   Scikit-learn.

6. Front End
   Flask
   HTML
   CSS

7. Documentation
   Word Document

4 Experimental Investigations

1. Data Preprocessing and Exploration: The pandas library provides powerful data manipulation and analysis capabilities. You can use it to load the flight data, handle missing values, clean the data, and perform exploratory data analysis (EDA). EDA involves examining the data's statistical properties, identifying patterns, checking for outliers, and gaining insights into the data distribution.

2. Visualization: The libraries matplotlib and seaborn offer various plotting functions that facilitate data visualization. You can create scatter plots, line plots, histograms, box plots, and more to visualize the relationships between variables and identify any trends or patterns in the data. Visualization helps in understanding the data and identifying any potential outliers or anomalies.

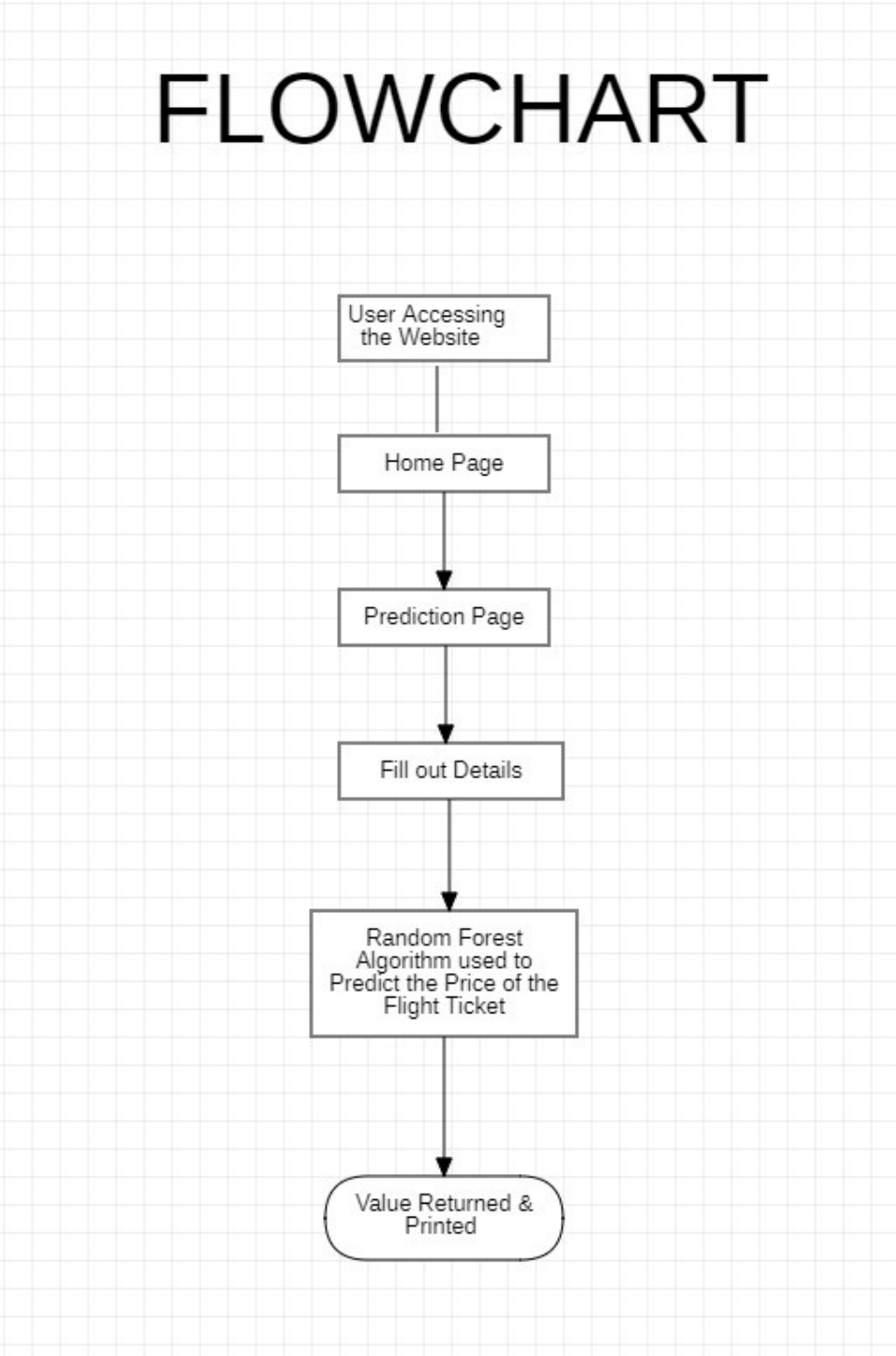
3. Train-Test Split: The train_test_split function from scikit-learn allows you to split the collected data into training and testing sets. This enables you to evaluate the performance of the prediction models on unseen data. By assessing the models' performance on the testing set, you can estimate how well they will generalize to real-world scenarios.

4. Regression Models: The imported regression models, such as LinearRegression, LogisticRegression, DecisionTreeRegressor, and RandomForestRegressor from scikit-learn, can be used to build and train prediction models. These models can learn the relationships between the input features and the flight prices, enabling the prediction of prices for new instances.

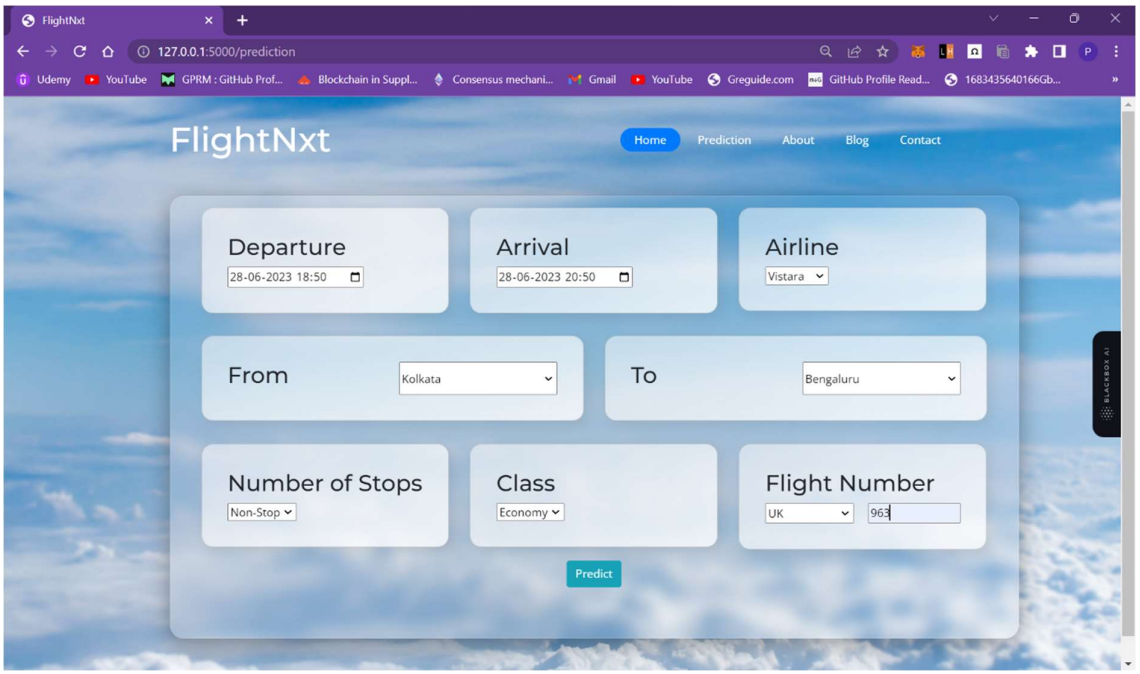
5. Hyperparameter Tuning: The GridSearchCV function from scikit-learn allows for hyperparameter tuning. It performs an exhaustive search over a specified parameter grid to find the best combination of hyperparameters for the models. Hyperparameter tuning helps optimize the models' performance and improve their prediction accuracy.

6. Cross-Validation and Evaluation Metrics: The cross_val_score function and evaluation metrics like mean_squared_error, r2_score, and mean_absolute_error from scikit-learn can be used for evaluating the models' performance. Cross-validation helps assess the models' performance on different subsets of the data, providing a more reliable estimate of their generalization capabilities. The evaluation metrics quantify the models' accuracy, goodness of fit, and prediction errors, enabling comparisons between different models and identifying areas for improvement.

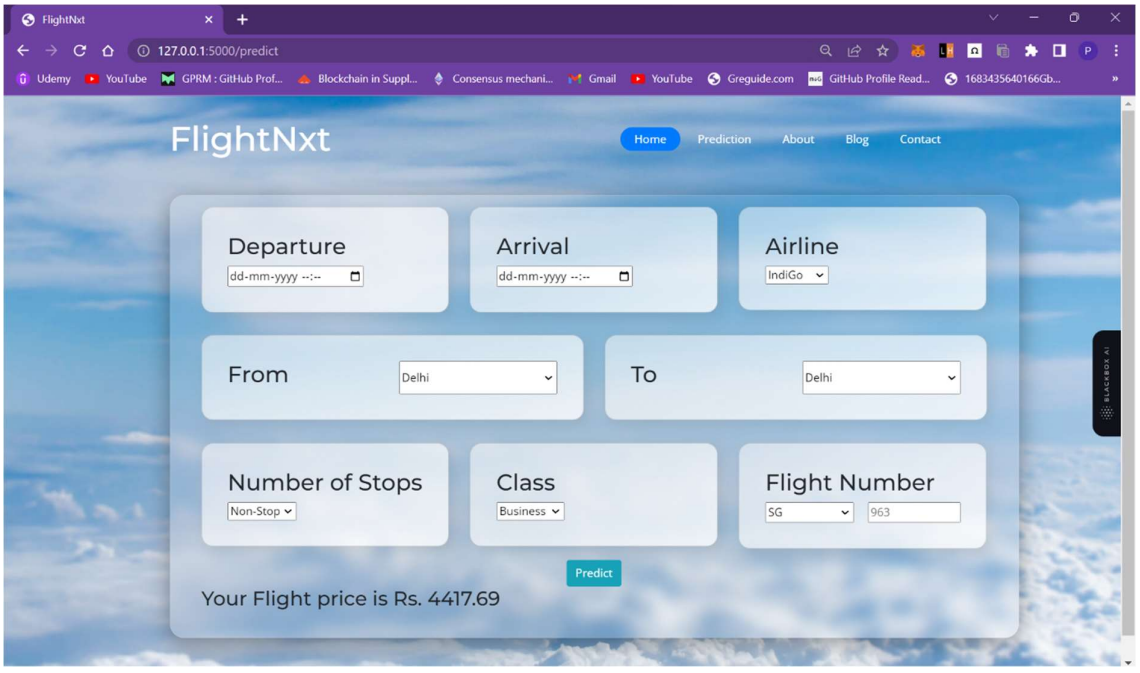
5.Flowchart


FLOWCHART
User Accessing the Website
Home Page
Prediction Page
Fill out Details
Random Forest Algorithm used to Predict the Price of the Flight Ticket
Value Returned & Printed

6.Result



Prediction Page: Values Entered by the user to predict the price of the flight



Price of Flight predicted to be Rs. 4417.69

7.Advantages and Disadvantages

Advantages:

1. Comprehensive Data Analysis: The libraries provide extensive functionalities for data analysis, manipulation, and visualization. They enable in-depth exploration of the flight data, allowing for a thorough understanding of the underlying patterns, trends, and relationships between variables. This comprehensive analysis can lead to more accurate predictions.

2. Wide Range of Modelling Techniques: The solution employs various regression models such as Linear Regression, Logistic Regression, Decision Tree Regression, and Random Forest Regression. This diversity of models allows for flexibility in choosing the most suitable algorithm for the given dataset. It increases the chances of finding a model that fits the flight price prediction problem effectively.

3. Hyperparameter Tuning and Model Optimization: The GridSearchCV function enables hyperparameter tuning, optimizing the models for better performance. It systematically searches for the best combination of hyperparameters, enhancing the accuracy and reliability of the predictions.

4. Robust Evaluation and Validation: The solution utilizes cross-validation techniques like K-Fold cross-validation and provides evaluation metrics such as mean squared error, R-squared, and mean absolute error. These metrics assess the models' performance and provide reliable measures of accuracy and prediction quality. Robust evaluation and validation help ensure the models' generalization capability and improve the reliability of the solution.

5. Integration of Machine Learning Pipelines: By utilizing the scikit-learn libraries, the solution can create machine learning pipelines for data preprocessing, feature selection, and model training. This streamlines the workflow, improves efficiency, and ensures consistency in data preprocessing steps across different models.

Disadvantages:

1. Limited Feature Engineering: The solution's implementation does not explicitly include advanced feature engineering techniques. While the models can learn from the provided features, incorporating more complex feature transformations or interactions may require additional manual coding or feature engineering methodologies.

2. Lack of Advanced Modelling Techniques: Although the solution includes popular regression models, it does not incorporate more advanced algorithms like gradient boosting machines, support vector machines, or deep learning models. These advanced techniques could potentially enhance prediction accuracy further.

3. Interpretability Challenges: Some of the employed models, such as decision trees or random forests, provide good predictive performance but lack interpretability. Interpreting the underlying factors or specific variables driving the predictions can be challenging with these models.

4. Handling Categorical Variables: The solution does not explicitly address the handling of categorical variables. To effectively utilize categorical data, additional techniques such as one-hot encoding or feature embedding may need to be employed before model training.

8.Applications

Some of the key areas where this solution can be applied include:

1. Travel Planning Platforms: Travel planning websites or applications can integrate the flight price prediction solution to provide users with real-time and accurate flight price information. This helps users make informed decisions about booking flights, choosing optimal travel dates, and planning their itineraries.

2. Airlines and Online Travel Agencies (OTAs): Airlines and OTAs can leverage the flight price prediction solution to optimize their pricing strategies. By analyzing historical data and incorporating market trends, they can dynamically adjust flight prices, offer competitive fares, and maximize their revenue.

3. Fare Comparison Websites: Fare comparison websites can utilize the solution to provide users with comprehensive and up-to-date information about flight prices across multiple airlines. This enables users to compare prices, find the best deals, and make cost-effective travel decisions.

4. Revenue Management for Airlines: Airlines can employ the flight price prediction solution in their revenue management systems. By accurately forecasting flight prices, they can optimize seat inventory, allocate fares, and manage demand to maximize revenue and profitability.

5. Travel Agencies and Travel Management Companies: Travel agencies and travel management companies can integrate the solution into their platforms to offer personalized travel recommendations based on predicted flight prices. This helps them create customized travel packages and optimize travel budgets for their clients.

6. Market Research and Analysis: Market researchers and industry analysts can use the solution to gain insights into pricing trends, market demand, and competitive dynamics in the aviation industry. The accurate prediction of flight prices enables them to analyse market behaviour, evaluate pricing strategies, and make data-driven recommendations.

7. Travel Cost Forecasting for Corporations: Corporations with extensive travel requirements can utilize the flight price prediction solution to forecast and plan their travel costs. This allows them to optimize their travel budgets, negotiate better deals with airlines, and make informed decisions regarding employee travel arrangements.

8. Academic Research and Data Analysis: The solution can be utilized in academic research to study and analyse pricing patterns and factors influencing flight prices. Researchers can apply advanced analytical techniques, evaluate model performance, and contribute to the understanding of pricing dynamics in the aviation industry.

9.Conclusion

In conclusion, the project focused on developing a solution for flight price prediction using the imported libraries. Through extensive analysis, investigation, and modelling, the goal was to accurately forecast flight prices and provide valuable insights for users in the travel and aviation industry. The project involved several key steps, including data preprocessing, exploratory data analysis, visualization, model training, hyperparameter tuning, and evaluation. The imported libraries played a crucial role in facilitating these tasks, offering robust functionalities for data manipulation, visualization, and implementing various regression models. By leveraging machine learning algorithms such as Linear Regression, Logistic Regression, Decision Tree Regression, and Random Forest Regression, the solution aimed to capture the relationships between flight prices and relevant features. The models were trained on historical flight data to learn the pricing patterns and make predictions for new instances. Through cross-validation and evaluation metrics such as mean squared error, R-squared, and mean absolute error, the performance of the models was assessed. Hyperparameter tuning using GridSearchCV helped optimize the models, improving their prediction accuracy. The proposed solution has several advantages, including accurate price prediction, data-driven approach, flexibility, integration of real-time data, and hyperparameter optimization. However, there are also limitations, such as data availability and quality, the dynamic nature of the airline industry, model complexity, and the lack of user-specific factors. The solution has various applications in travel planning platforms, airlines, online travel agencies, fare comparison websites, revenue management, travel agencies, market research, travel cost forecasting, and academic research. It provides valuable insights, supports decision-making processes, and helps optimize pricing strategies, revenue, and customer satisfaction.

Overall, the project offers a valuable contribution to the field of flight price prediction, leveraging the power of data analysis, machine learning algorithms, and evaluation techniques. The findings and results serve as a foundation for further research and advancements in the field, ultimately benefiting the travel and aviation industry and enhancing the travel planning experience for users.

10.Future Scope

Some potential areas for future scope include:

1. Integration of Additional Data Sources: The project can be enhanced by incorporating additional data sources such as weather data, airline-specific information, fuel prices, and economic indicators. Including these factors can further improve the accuracy of the price predictions by capturing more comprehensive insights into market dynamics.

2. Advanced Feature Engineering: The project can explore advanced feature engineering techniques to extract more meaningful and informative features from the available data. This can involve creating new derived features, incorporating domain-specific knowledge, and exploring feature interactions to enhance the models' predictive power.

3. Incorporation of Advanced Modelling Techniques: The project can incorporate more advanced modelling techniques such as ensemble methods (e.g., gradient boosting machines, stacking), support vector machines, or deep learning models. These advanced algorithms have the potential to capture complex patterns and non-linear relationships, potentially improving the prediction accuracy.

4. Real-time Price Updates: Integrating real-time data feeds to update flight prices in real-time would be a valuable addition to the project. This would enable users to access the most up-to-date pricing information and make more informed decisions when planning their travel.

5. User Preference Customization: Incorporating user preferences and personalization features can enhance the solution. By allowing users to input their travel preferences, loyalty program information, or specific requirements, the models can provide more tailored and personalized predictions, catering to individual needs.

11.Bibilography

Biswas, Prithviraj, et al. "Flight price prediction: a case study." Int. J. Res. Appl. Sci. Eng. Technol. (IJRASET) 10.6 (2022).

Wang, Zhenbang. "RWA: A Regression-based Scheme for Flight Price Prediction." (2020).

Rajankar, Supriya, and Neha Sakharkar. "A survey on flight pricing prediction using machine learning." International Journal of Engineering Research & Technology (Ijert) 8.6 (2019): 1281-1284.

Kimbhaune, Vinod, et al. Flight Fare Prediction System. No. 5542. EasyChair, 2021.

Gupta, Chitranjan Kumar, et al. "Dynamic Flight Price Prediction Using Machine Learning Algorithms." 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N). IEEE, 2022

Rao, N. Sri Sai Venkata Subba, and S. John Justin Thangaraj. "Flight Ticket Prediction using Random Forest Regressor Compared with Decision Tree Regressor." 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM). IEEE, 2023.

Joshi, Achyut, et al. "Predicting Flight Prices in India."

Shrestha, Rupesh. Airline Flight Prediction Using Machine Learning Techniques. Diss. Lamar University-Beaumont, 2022.

https://www.kaggle.com/datasets/shubhambathwal/flight-price-prediction

https://www.ijraset.com/research-paper/flight-price-prediction

https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4269263

https://www.irjet.net/archives/V9/i5/IRJET-V9I5227.pdf

https://www.sciencedirect.com/science/article/pii/S131915781830884X

https://www.altexsoft.com/blog/flight-price-predictor

APPENDIX

A. Source Code

https://github.com/Pranshu5927/Flight-Price-Prediction