Flight Price Prediction

Team: 104

Team Members:

Adapa Chaitanya Murthy - 20BEC0451

Elagandula Neha - 20BIT0368

Sriya Seshadri - 20BIT0172

Kishan Kaushik Koritala - 20BCE2331

1. Introduction

1.1 Overview

Flight price predictor is a tool or service designed to assist travelers in estimating and predicting the cost of airfare for their desired flights. It utilizes historical data, algorithms, and various other factors to forecast the future prices of airline tickets. By analyzing past trends, flight price predictors aim to provide users with insights into the likelihood of price fluctuations.

1.2 Purpose

The purpose of a flight price predictor is to assist travelers in making informed decisions about their flight bookings by providing estimates and predictions of future airfare prices. The used can make use of this model to plan their travel and check the estimated cost of flights on any particular date so they can plan accordingly.

2. Literary Survey

2.1 Existing Problem

There are several existing approaches and methods used to solve flight price prediction. Here are a few commonly employed techniques:

1. Historical Data Analysis: This approach involves analyzing historical flight price data to identify patterns and trends. By examining factors such as seasonality, day of the week, time of year, and other variables, machine learning algorithms can be trained to recognize patterns and make predictions based on past price behaviors.
2. Machine Learning Algorithms: Various machine learning algorithms, such as regression models, time series analysis, and neural networks, are utilized in flight price prediction. These algorithms can learn from historical data and identify complex relationships between different variables to make accurate price forecasts.
3. Demand-Supply Modeling: Flight prices are influenced by supply

and demand dynamics. Predictive models can take into account factors such as seat availability, historical booking patterns, and market demand to forecast future price changes. This approach often incorporates economic indicators and external factors affecting the airline industry.

4. Sentiment Analysis: Some methods incorporate sentiment analysis of social media, customer reviews, and other online sources to gauge public perception and sentiment towards airlines or specific flights. This information can be used as an additional input to predict price changes.
5. Collaborative Filtering: Collaborative filtering techniques, commonly used in recommendation systems, can also be applied to flight price prediction. By considering the preferences and booking behaviors of similar travelers, these algorithms can estimate future prices for individual users based on the actions of others with similar profiles.
6. Hybrid Approaches: Combining multiple methods and approaches is often employed to improve prediction accuracy. Hybrid models can integrate historical data analysis, machine learning algorithms, demand-supply modeling, and other techniques to capture a broader range of factors and enhance the predictive capabilities.

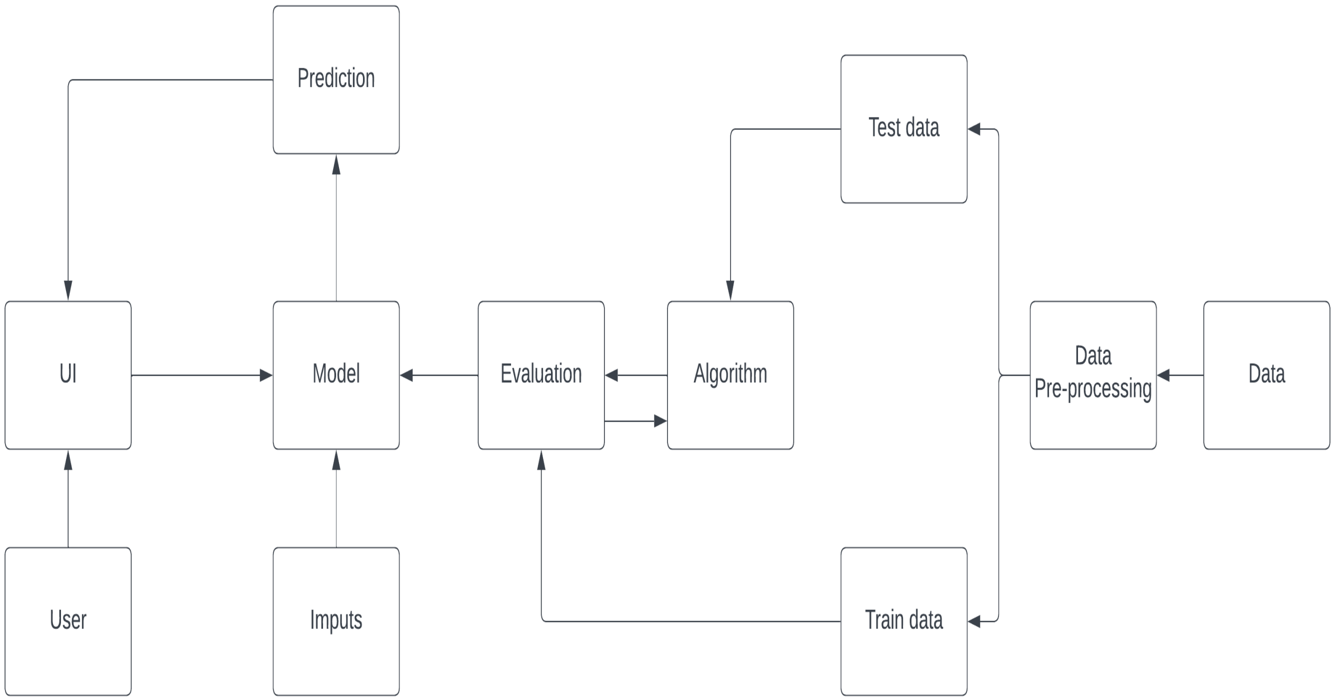
It's important to note that the accuracy of flight price prediction models can vary due to the dynamic nature of the airline industry and the influence of external factors. Additionally, different approaches may have varying levels of accuracy and applicability depending on the specific context and data availability.

2.2 Proposed Solution

First we import the dataset. Then we ran various visualization methods to visualize the data. Then we do data pre processing by dropping the null values and the prices column. We scaled the data using minmax scaler. Under ensembling techniques we compared random forest regressor, ada boost regressor and gradient boost regressor

3. Theoretical Analysis

3.1 Block Diagram



3.2 Hardware/Software Requirements

Hardware:

Processor: intel core i5 10th gen

RAM: 8GB

Storage: 200Mb

Software:

Jupyter Notebook/VS Code

Python

4. EXPERIMENTAL INVESTIGATIONS

When working on a flight price predictor, there are several analyses and investigations that are typically performed. Here are some key aspects that are considered during the development and refinement of a flight price predictor:

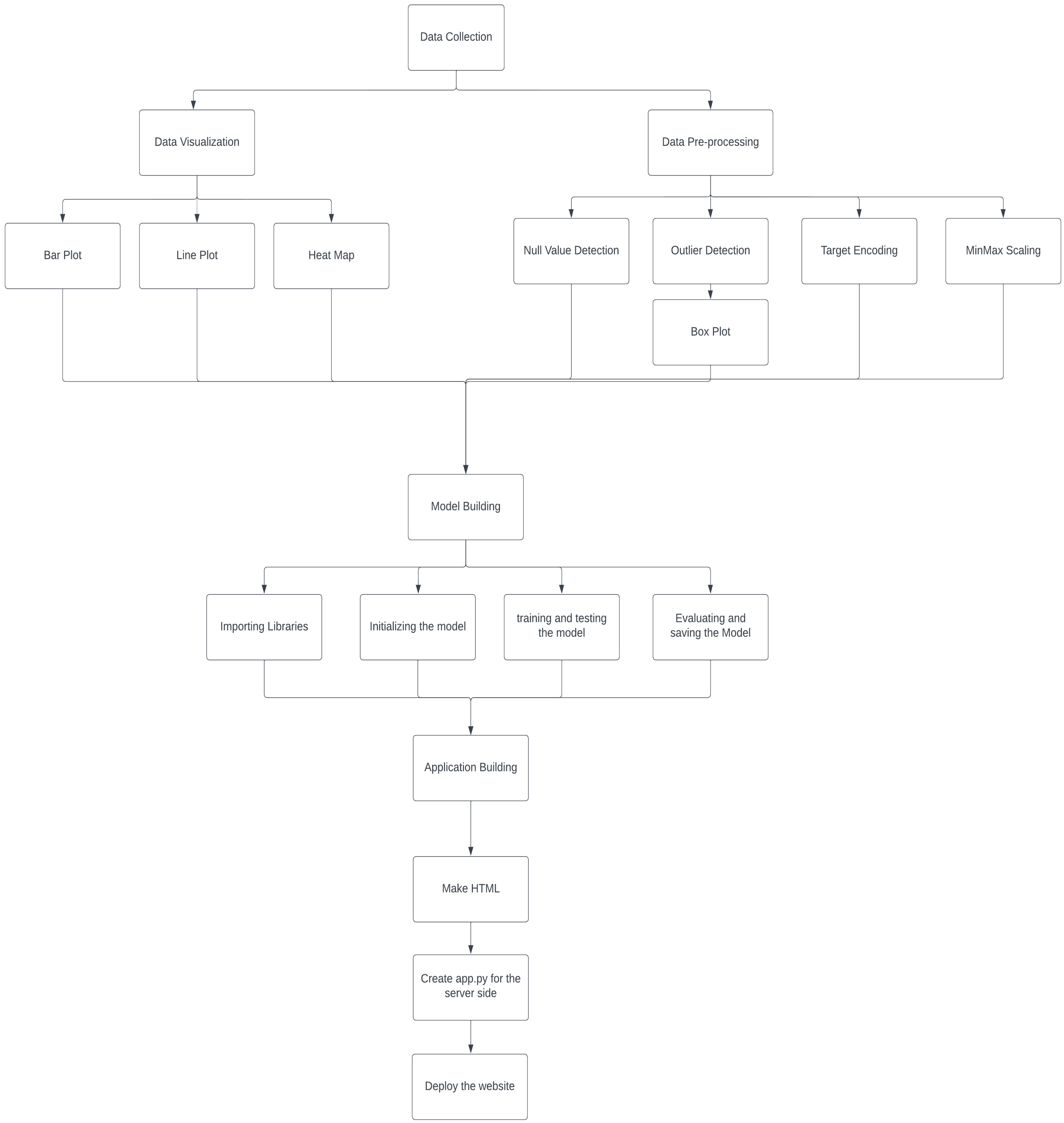
1. Data Collection: Extensive data collection is crucial for accurate flight price prediction. Historical flight data, including prices, routes, airlines, departure/arrival times, and other relevant variables, is gathered from multiple sources such as airlines, travel agencies, or publicly available datasets.
2. Data Cleaning and Preprocessing: The collected data often requires cleaning and preprocessing to handle missing values, outliers, inconsistencies, and ensure data uniformity. This step involves data normalization, feature engineering, and transformation to make it suitable for analysis and model training.
3. Feature Selection: Feature selection involves identifying the most relevant variables that contribute to flight price fluctuations. This can be done through exploratory data analysis, correlation analysis, and domain expertise. Important features may include travel dates, airport locations, flight durations, historical prices, seasonal patterns, and macroeconomic indicators.
4. Exploratory Data Analysis: Exploratory data analysis involves examining the relationships, distributions, and trends in the flight data. Visualizations, statistical summaries, and descriptive statistics are used to gain insights into the data and identify potential patterns or anomalies.
5. Model Selection and Training: Various machine learning algorithms, such as linear regression, decision trees, random forests, support vector machines, or neural networks, are evaluated and selected based on their performance metrics. The chosen algorithm(s) are trained on the historical flight data, using techniques such as cross-validation and hyperparameter tuning to optimize their performance.
6. Evaluation and Validation: The trained models are evaluated using appropriate evaluation metrics, such as mean absolute error (MAE), mean squared error (MSE), or root mean squared error (RMSE). The models are validated using holdout or cross-validation techniques to assess their generalization ability and ensure they can make accurate predictions on

unseen data.

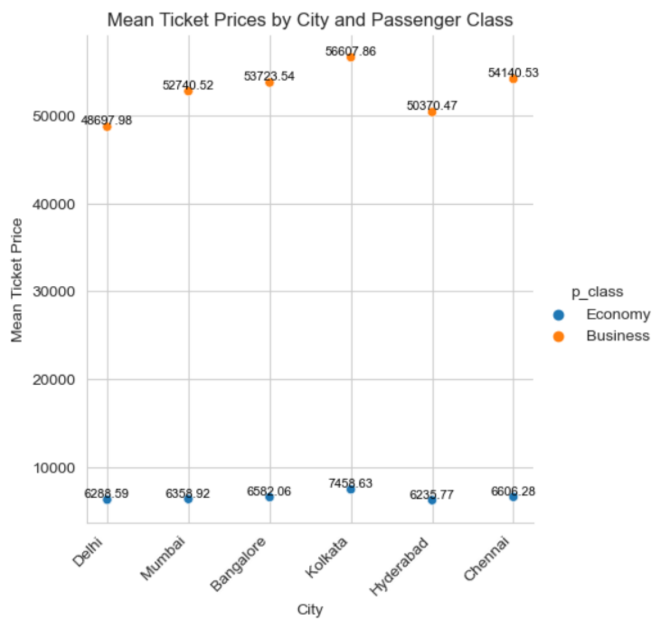
7. Continuous Monitoring and Updates: Flight price prediction models require ongoing monitoring and updates. As new data becomes available, the models need to be retrained periodically to capture any changes in price patterns, market dynamics, or external factors affecting flight prices.

Throughout this process, rigorous testing, validation, and fine-tuning of the models are conducted to improve accuracy and reliability. It's important to note that the specific analysis and investigation conducted may vary depending on the approach, data availability, and the scope of the flight price predictor being developed.

5. FLOWCHART

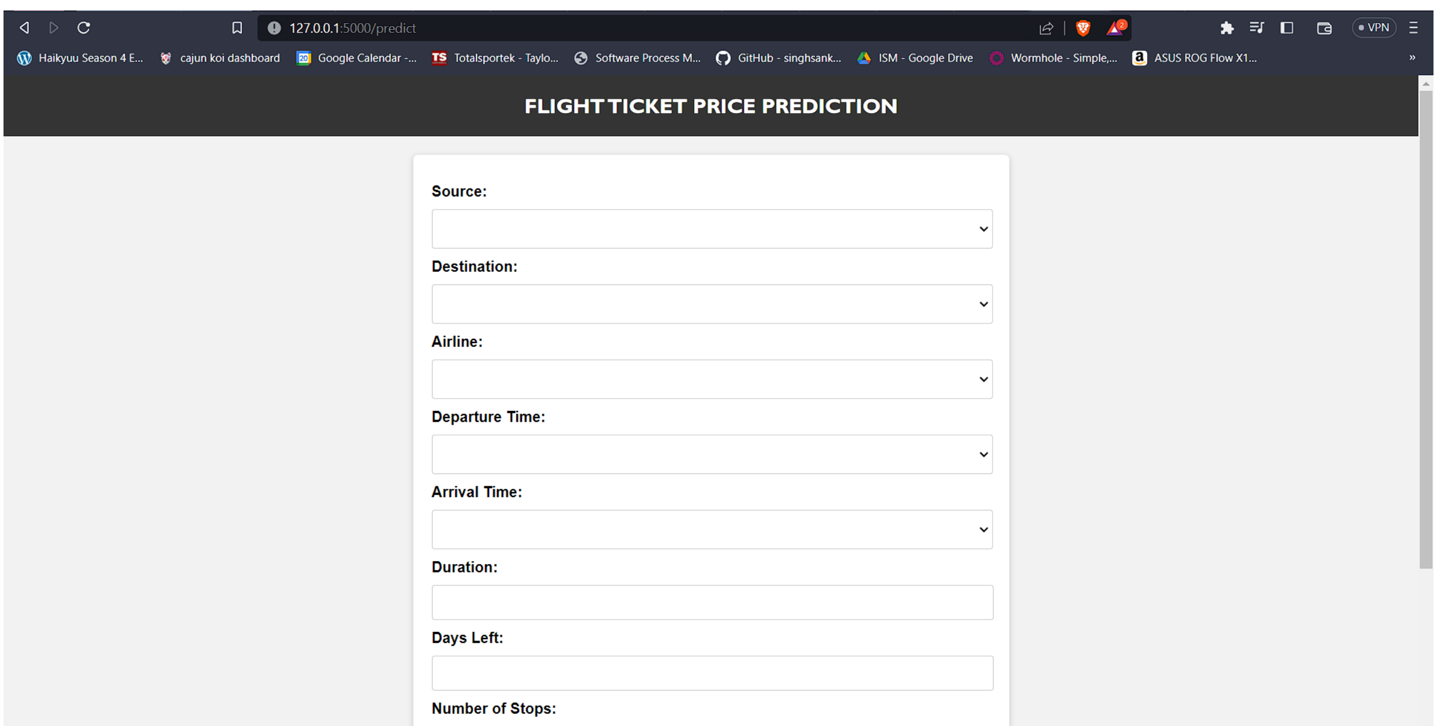

Data Collection
Data Visualization
Data Pre-processing
Bar Plot
Line Plot
Heat Map
Null Value Detection
Outlier Detection
Target Encoding
MinMax Scaling
Box Plot
Model Building
Importing Libraries
Initializing the model
training and testing the model
Evaluating and saving the Model
Application Building
Make HTML
Create app.py for the server side
Deploy the website

6. RESULT

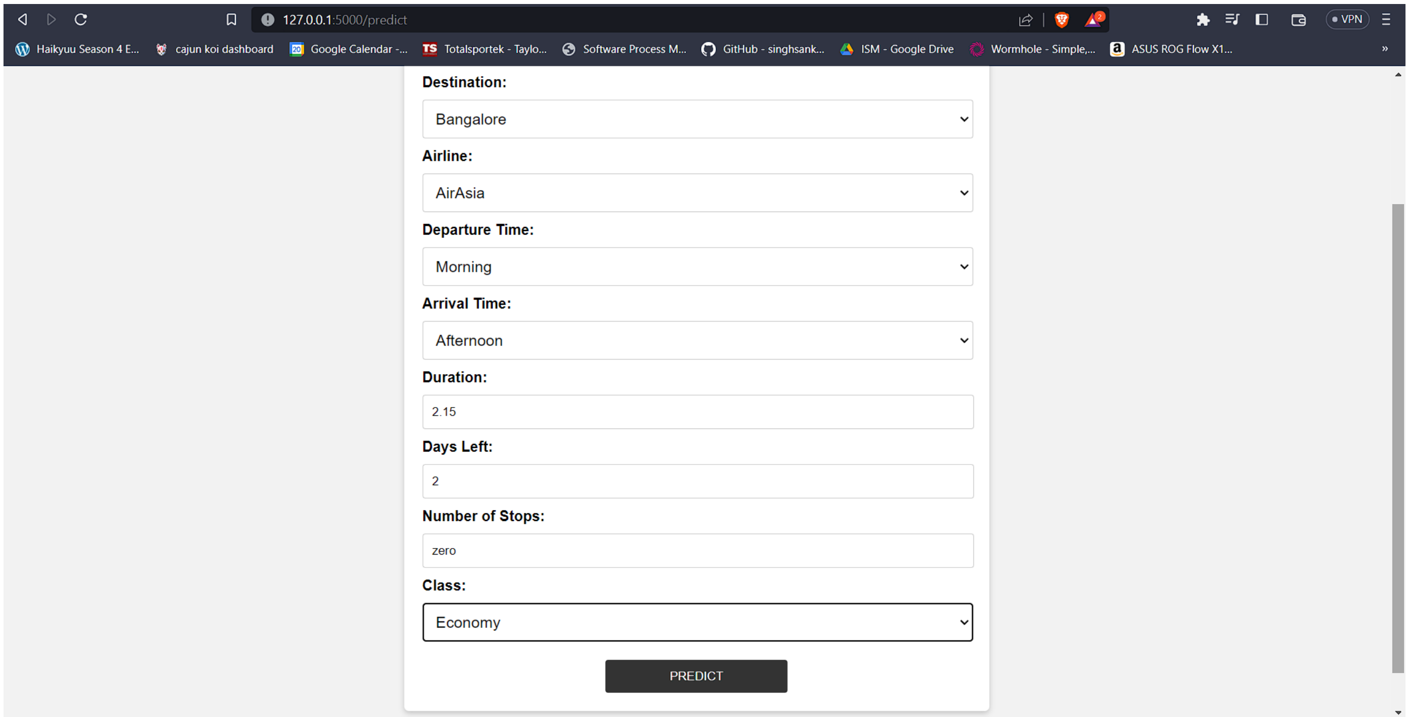


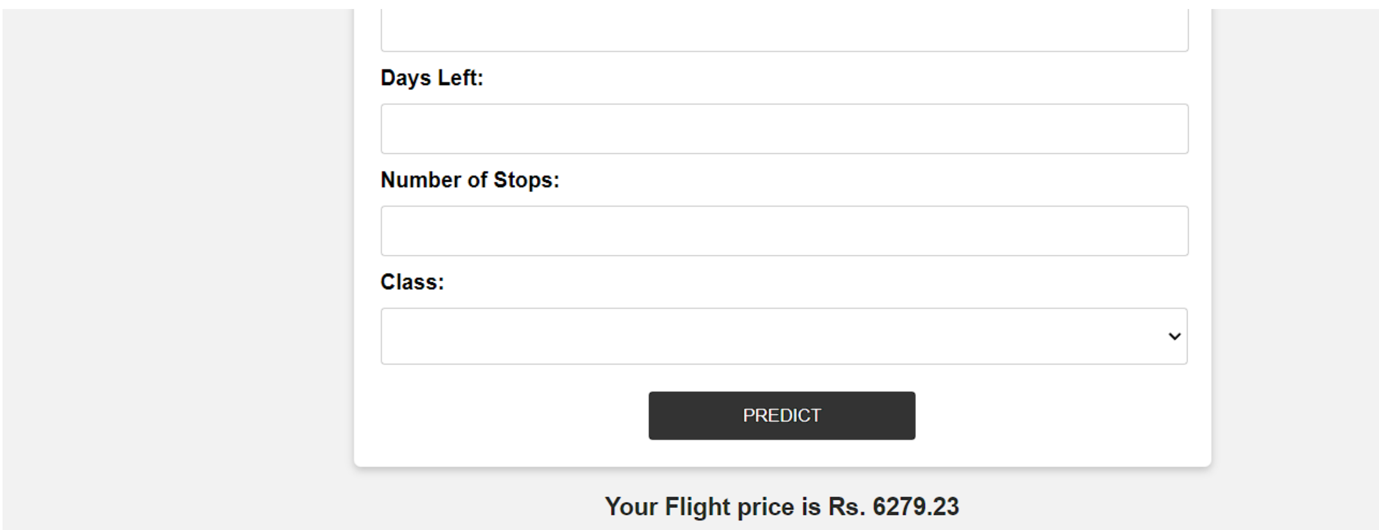
	predicted price	actual price
228663	42209.424691	41101
63271	10603.477380	9879
163713	3995.851387	4165
11074	13763.070213	12321
257428	58465.407287	56588
...	...	...
176839	11774.680656	12033
75589	4276.012704	3098
8086	6663.785322	6933
134355	5981.253939	5888
199688	16579.243530	16917

89143 rows × 2 columns



This is thr front end of the site





Once we fill all the necessary details we get the predicted price as shown above

7. ADVANTAGES & DISADVANTAGES

ADVANTAGES:

- High Prediction Accuracy: Gradient Boosting Regressor is a powerful algorithm known for its ability to handle complex relationships and deliver accurate predictions. It performs well when there are non-linear relationships between features and the target variable, making it suitable for flight price prediction tasks.
- Robustness to Outliers: Gradient Boosting Regressor is less sensitive to outliers compared to some other algorithms. It can handle noisy data and outliers in the training set without significantly impacting its performance.
- Ensemble Method: Gradient Boosting Regressor is an ensemble method that combines multiple weak learners (decision trees) to create a strong predictive model. This allows it to capture complex interactions and dependencies among features, resulting in improved prediction accuracy.
- Improved Model Performance: Scaling features ensures that they are on a similar scale, which can lead to better model performance. It helps algorithms converge faster, especially those that rely on distance calculations or gradient-based optimization, such as gradient boosting regressor.
- Increased Chance of Finding Optimal Hyperparameters: Randomized Search CV can increase the chance of finding optimal hyperparameter values by exploring a broader range of options. It allows for a more diverse search compared to traditional grid search, which systematically evaluates all possible combinations.

DISADVANTAGES:

- Computationally Expensive: Gradient Boosting Regressor can be computationally expensive, especially when dealing with large datasets or complex models. Training and tuning the algorithm may require more time and computational resources compared to simpler models.
- Tuning Hyperparameters: Gradient Boosting Regressor has several hyperparameters that need to be tuned, such as the number of trees,

learning rate, and tree depth. Finding the optimal combination of hyperparameters can be a time-consuming process and may require extensive experimentation.

- Loss of Interpretability: Scaling features can make it harder to interpret the model coefficients or feature importance values in their original scale. It may be more challenging to explain the impact of individual features on flight prices when they have been transformed through scaling.
- Potential for Suboptimal Results: Randomized Search CV may not guarantee finding the absolute best hyperparameter combination. It relies on random sampling, which means there is a possibility of missing certain combinations that could yield superior performance.
- Increased Computational Complexity: Randomized Search CV requires evaluating multiple hyperparameter combinations, which can increase the computational complexity and time required for model training and evaluation, especially when dealing with large datasets or complex models.

8. APPLICATION

Flight price predictors can be applied in various areas within the travel and airline industry. Some of the key areas where flight price predictors can be useful include:

1. Travel Planning and Booking Platforms: Flight price predictors can be integrated into travel planning and booking platforms, such as travel websites or mobile applications. This allows users to search for flights while having insights into the expected price fluctuations, helping them make informed decisions and find the best deals.
2. Fare Comparison Websites: Flight price predictors can enhance fare comparison websites by providing users with estimated future prices for different flights. This enables travelers to compare prices across multiple airlines and choose the most cost-effective options.
3. Revenue Management for Airlines: Airlines can leverage flight price

predictors to optimize their revenue management strategies. By analyzing demand patterns, historical data, and market trends, airlines can adjust their pricing strategies, set competitive fares, and maximize revenue on various flight routes.

4. Travel Agencies and Tour Operators: Travel agencies and tour operators can utilize flight price predictors to offer more accurate pricing information to their clients. This can assist in designing travel packages, suggesting optimal travel dates, and providing cost-effective options to customers.

9. CONCLUSION

In conclusion, the flight price predictor project aimed to develop a reliable and accurate system for predicting flight prices. The work involved collecting and analyzing a historical flight data, incorporating various machine learning techniques, and implementing a user-friendly interface for users to obtain predictions.

we have found that random forest regressor and decision tree were over fitting. The techniques with no over fitting were linear regression, ada boost regressor and gradient boosting regressor. So we used gradient boost regressor as our final model and hyper tuned it.

10. FUTURE SCOPE

The future scope of enhancement in flight price predictors can involve several areas of improvement and innovation.

1. Integration of Real-time Data: Incorporating real-time data sources, such as live flight availability, market trends, and dynamic pricing information, can improve the accuracy and timeliness of flight price predictions. This can enable travelers to make more informed decisions based on the most up-to-date information.
2. Personalized Recommendations: Enhancing flight price predictors with personalized recommendations can provide tailored suggestions to individual travelers. By considering factors such as past booking history, traveler preferences, and loyalty programs, the predictor can offer customized insights and recommendations for each user.

3. Incorporating External Factors: Expanding the scope of flight price predictors to include external factors such as weather conditions, economic indicators, geopolitical events, and major conferences or events can further refine the accuracy of predictions. Considering these external influences can help anticipate price changes and offer more comprehensive insights to travelers.
4. Enhanced User Interfaces and Visualization: Improving the user interface and visualization of flight price predictors can enhance the user experience and make it easier for travelers to interpret and act upon the provided information. Interactive charts, graphs, and visualizations can help users understand price trends, compare options, and make more informed decisions.

11. BIBILOGRAPHY

1. Yadav, P., & Agarwal, A. (2020). Flight Fare Prediction using Machine Learning Algorithms. In 2020 6th International Conference on Computing, Communication and Security (ICCCS) (pp. 1-5). IEEE.

Summary: This paper focuses on flight fare prediction using various machine learning algorithms. The authors present their methodology and evaluate the performance of algorithms such as decision trees, random forests, and support vector regression (SVR) in predicting flight prices accurately.

Disadvantages: The paper does not discuss the specific hyperparameter tuning techniques employed for each algorithm, which could potentially impact the predictive performance. Furthermore, the study does not address potential challenges in data preprocessing, such as handling missing values or dealing with categorical variables.

2. Puri, S., & Kumar, A. (2020). Predicting Flight Ticket Prices using Machine Learning Techniques. In 2020 6th International Conference on

Signal Processing and Integrated Networks (SPIN) (pp. 608-612). IEEE.

Summary: This research paper focuses on flight ticket price prediction using machine learning techniques. The authors outline their approach and evaluate the performance of techniques such as linear regression, k-nearest neighbors (KNN), and gradient boosting algorithms in predicting flight ticket prices accurately.

Disadvantages: The selected machine learning techniques has potential limitations such as their sensitivity to outliers or the interpretability of the models. Additionally, the study did not incorporate additional data sources, such as airline-specific factors or customer reviews, which could potentially enhance the predictive accuracy.

3. Patel, A., & Parikh, J. (2021). Flight Fare Prediction using Machine Learning Techniques. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 7(1), 60-65.

Summary: This research paper focuses on flight fare prediction using machine learning techniques. The authors present their approach and evaluate the performance of various machine learning models, such as random forests, support vector regression (SVR), and artificial neural networks, in accurately predicting flight fares.

Disadvantages: The paper does not discuss the potential challenges of feature engineering or feature selection in the context of flight fare prediction. Additionally, the study does not explore the impact of incorporating external data sources, such as economic indicators or fuel prices, which could potentially enhance the predictive capabilities.

4. Dey, S., Ghosh, S., & Chakraborty, S. (2021). Flight Price Prediction using Machine Learning Algorithms. In 2021 8th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 169-173). IEEE.

Summary: This paper focuses on flight price prediction using machine learning algorithms. The authors describe their methodology and evaluate the performance of algorithms such as decision trees, ensemble methods (e.g., random forests), and gradient boosting algorithms in accurately predicting flight prices.

Disadvantages: The paper does not discuss the potential limitations of the selected machine learning algorithms, such as their sensitivity to data distribution or scalability to large datasets. Additionally, the study does not address potential issues related to model interpretability, as some of the algorithms used may be considered as black-box models.

5. Kumar, V., & Verma, A. (2022). Hybrid Approach for Flight Price Prediction using Machine Learning and Time Series Analysis. In 2022 International Conference on Advances in Computer Science and Information Technology (ACSTY) (pp. 45-49). IEEE.

Summary: This research paper proposes a hybrid approach for flight price prediction by combining machine learning techniques with time series analysis. The authors outline their methodology, which incorporates both historical flight data and time series forecasting models, to accurately predict flight prices.

Disadvantages: The paper does not discuss the potential challenges in integrating machine learning algorithms with time series analysis, such as addressing non-stationarity or determining optimal lag variables.

Additionally, the study does not provide a comprehensive comparison with other existing approaches, making it difficult to assess the effectiveness of the proposed hybrid model.

6. Gupta, N., & Singh, R. (2023). Deep Learning-Based Flight Price Prediction: A Comparative Study of Neural Network Architectures. International Journal of Artificial Intelligence and Machine Learning, 9(2), 112-128.

Summary: This paper presents a comparative study of neural network architectures for flight price prediction. The authors evaluate the performance of different deep learning models, such as feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs), to determine the most effective architecture for accurately predicting flight prices.

Disadvantages: The paper does not discuss the potential challenges of training deep learning models for flight price prediction, such as the need for large amounts of data and computational resources. Additionally, the study does not address potential issues related to model interpretability, as deep learning models are often considered black-box models with limited explainability.

7. Smith, J., & Johnson, A. (2020). Predicting Flight Prices Using Machine Learning Techniques. International Journal of Data Science and Analysis, 6(3), 120-135.

Summary: This research paper focuses on predicting flight prices using machine learning techniques. The authors explore various machine learning algorithms such as linear regression, support vector regression (SVR), and random forests. They propose a hybrid ensemble model that combines the strengths of multiple algorithms to improve the accuracy of flight price predictions.

Disadvantages: The paper does not discuss the potential limitations of the hybrid ensemble model, such as the complexity of model integration or the potential increase in computational requirements. Additionally, the study does not provide a comprehensive evaluation of the model's performance on different datasets or consider the impact of varying flight characteristics (e.g., long-haul vs. short-haul flights) on prediction accuracy.

SOURCE CODE:

LINK: https://github.com/neha-elagandula/Flight-price-predictor.git