**Project on optimizing flight price prediction using machine learning**

# **CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S. No** | **Description** | **Page no.** |
| 1 | Introduction |  |
| 2 | Overview of the project |  |
| 3 | Purpose of the project |  |
| 4 | Literature Survey & existing problems |  |
| 5 | Proposed solution |  |
| 6 | Theoretical analysis |  |
| 7 | Result |  |
| 8 | Advantages &Disadvantages |  |
| 9 | Applications |  |
| 10 | Conclusions |  |
| 11 | Future scope |  |
|  |  |  |

**1.Introduction**

Nowadays, the airline corporations are using complex strategies for the flight ticket fare calculations. This highly complicated methods makes the flight ticket fare difficult to guess for the customers, since the fare changes dynamically.

* Our project “Optimizing Flight Booking Decisions Through Machine Learning Price Prediction” which resolve this problem and provide a facility where people will be able to predict the flight-ticket price before purchasing the ticket.

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

* 1. **overview**

Optimizing flight price prediction using machine learning involves utilizing various techniques and algorithms to accurately forecast flight ticket prices. The goal is to provide travelers with reliable and timely information about when to book flights to get the best deals, while also assisting airlines and travel agencies in pricing strategies. Here's an overview of the process:

* **Data Collection:** Gather historical flight data, including ticket prices, travel dates, departure and arrival locations, airline information, seat availability, booking times, and other relevant factors. Additional data sources like weather conditions, holidays, and economic indicators can also be considered.
* **Data Preprocessing:** Clean the collected data, handle missing values, and perform feature engineering. Feature engineering involves selecting and creating relevant features that might influence flight prices, such as day of the week, time of day, distance between airports, and more.
* **Data Splitting:** Divide the dataset into training, validation, and testing sets. The training set is used to train the machine learning models, the validation set helps tune hyperparameters, and the testing set evaluates the final model's performance.
* **Feature Scaling and Normalization:** Scale and normalize numerical features to ensure that they are on a similar scale. This helps prevent certain features from dominating the learning process due to their larger magnitudes.
* **Model Selection:** Choose appropriate machine learning algorithms for the prediction task. Common choices include regression algorithms (linear regression, decision trees, random forests, gradient boosting), time series models (ARIMA, LSTM), and more advanced methods like neural networks.
* Model Training: Train the selected models on the training dataset. The models learn the relationships between features and flight prices during this stage.
* **Hyperparameter Tuning:** Optimize the model's hyperparameters to improve its performance on the validation set. This can involve techniques like grid search, random search, or Bayesian optimization.
* **Model Evaluation**: Evaluate the models using appropriate evaluation metrics such as mean squared error (MSE), mean absolute error (MAE), or root mean squared error (RMSE) on the testing dataset. This step helps assess how well the models generalize to new, unseen data.
* **Ensemble Techniques:** Combine multiple models to create an ensemble that often performs better than individual models. Techniques like stacking, bagging, and boosting can be used.
* **Deployment and Monitoring:** Deploy the optimized model into a real-world application, such as a website or mobile app, to provide flight price predictions to users. Continuously monitor the model's performance and retrain it periodically with new data to ensure accuracy.
* **User Feedback and Iteration:** Collect feedback from users and monitor their interactions with the application. Use this feedback to improve the model and the user experience iteratively.
* **Business Strategy:** Collaborate with airlines, travel agencies, and other stakeholders to develop pricing strategies based on the predicted flight prices. Help travelers make informed decisions about when to book flights to get the best value.

By following these steps, optimizing flight price prediction using machine learning can lead to more accurate and timely predictions, benefiting both travelers and the travel industry as a whole.

**1.2 purpose**

The purpose of optimizing flight price prediction using machine learning is to develop accurate and reliable models that can forecast the prices of airline tickets. This can offer several benefits to both consumers and the airline industry:

* **Cost Savings for Consumers:** Accurate flight price predictions allow travelers to make more informed decisions about when to book their flights. This can help them find the best deals and avoid overpaying for tickets, leading to cost savings for individuals and families.
* **Revenue Generation for Airlines:** Airlines can use accurate price prediction models to adjust their pricing strategies dynamically. By understanding demand patterns and market trends, airlines can optimize their revenue by offering competitive prices while maximizing seat occupancy.
* **Demand Forecasting:** Machine learning models can analyze historical data and current market conditions to predict demand for specific flights. This enables airlines to allocate resources more efficiently and plan their operations better.
* **Inventory Management:** Airlines can optimize their inventory and seat allocation based on predicted demand. This can help prevent underbooking or overbooking, which can lead to revenue loss or customer dissatisfaction.
* **Personalized Pricing and Offers:** Machine learning can help airlines tailor pricing and offers to individual customers based on their preferences, purchase history, and browsing behavior. This can enhance customer satisfaction and loyalty.
* **Competitive Advantage:** Airlines that successfully implement accurate price prediction models can gain a competitive advantage by attracting cost-conscious travelers and maintaining a strong market position.
* **Data-Driven Decision Making:** Machine learning models provide data-driven insights that can guide strategic decisions related to route planning, marketing campaigns, and revenue management.
* **Enhanced Customer Experience:** By helping customers find the best deals and options, accurate flight price prediction contributes to an improved overall travel experience.

**2.LITERATURE SURVEY**

**2.1 Existing problem**

Flight prices exhibit temporal patterns, and time series analysis techniques such as Autoregressive Integrated Moving Average (ARIMA), Seasonal Decomposition of Time Series (STL), and Exponential Smoothing methods can be used to capture these patterns and make predictions.

* **Regression Models:** Classical regression models like Linear Regression, Lasso Regression, and Ridge Regression can be applied to predict flight prices by considering various features such as departure/arrival locations, dates, day of the week, time of day, and airline.
* **Ensemble Methods:** Ensemble techniques like Random Forest and Gradient Boosting can combine multiple weak models to improve prediction accuracy. They can capture complex relationships between features and flight prices.
* **Neural Networks:** Deep learning models, particularly neural networks, can learn intricate patterns from large datasets. Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks are effective for sequence data like time series.
* **Support Vector Regression (SVR):** SVR is a regression technique that finds a hyperplane that best fits the data points. It can capture nonlinear relationships between features and flight prices.
* **XGBoost and LightGBM:** These gradient boosting frameworks are widely used for structured data prediction tasks. They handle missing values well and can capture feature interactions effectively.
* **Feature Engineering:** Creating informative features can significantly improve prediction accuracy. Features like historical price trends, seasonality indicators, and special events (holidays, festivals, conferences) can provide valuable information.
* **Data Fusion:** Combining flight price data with external data sources like weather data, economic indicators, and social events can enhance prediction accuracy by capturing additional influencing factors.
* **Hyperparameter Tuning**: Fine-tuning model hyperparameters using techniques like grid search or random search can optimize model performance.
* **Cross-Validation:** Using techniques like k-fold cross-validation helps in estimating model performance and generalization to unseen data.
* **Anomaly Detection:** Identifying anomalies or unusual patterns in flight price data can help in understanding and predicting sudden price changes.
* **Reinforcement Learning:** In some cases, reinforcement learning can be applied to optimize the timing of flight bookings, treating it as a sequential decision-making problem.
* **Online Learning:** Flight prices can change frequently, and online learning techniques allow the model to adapt and learn from new data points in real-time.
* **Model Stacking**: Combining predictions from multiple models using techniques like model stacking or blending can lead to better overall performance.
* **Interpretable Models:** Sometimes, using interpretable models like Decision Trees or Linear Regression can provide insights into the factors influencing flight prices.

Remember that the effectiveness of these approaches depends on the quality and quantity of the data available, the complexity of the problem, and the specific characteristics of the flight price prediction task.

**2.2 proposed approach**

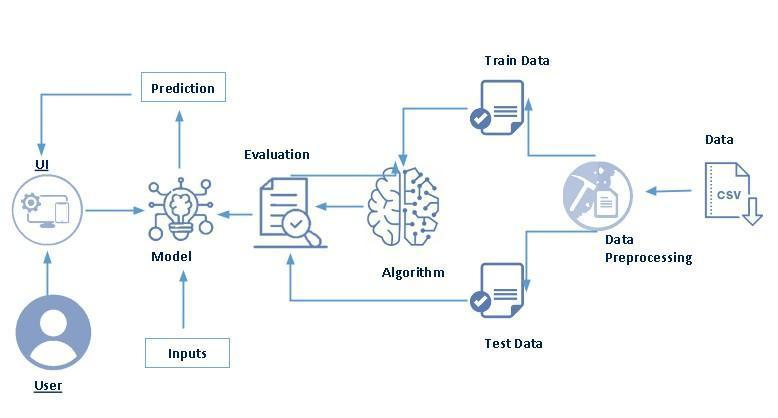
* The proposed approach is using machine learning algorithm and we are using supervised learning.
* We are gathering our data from a site. The data is containing some of the details of Indian flight of a short duration.
* This project involves the feature engineering for processing the dataset(data) to convert it into data frame. When we have the processed data frame we move to normalizing the data frame.
* The regression model which we have selected for our prediction is

“Extreme Gradient Boosting”.

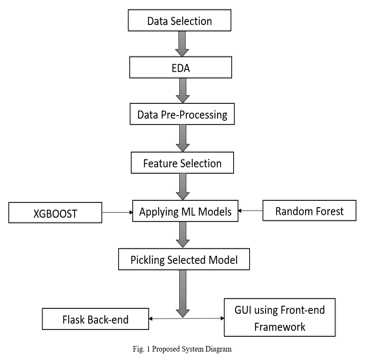
* We are training the model with the normalized data frame. After experimenting and tuning the hyper parameters we are obtaining the predicted results and the accuracy.

**3.THEORITICAL ANALYSIS**

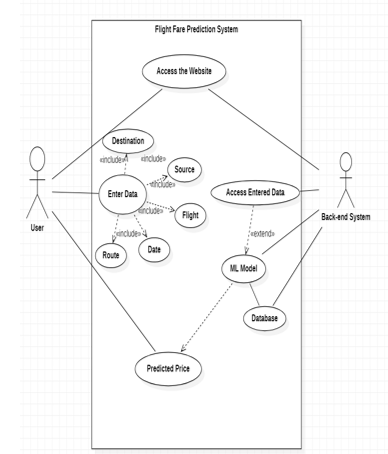
**3.1 technical Architecture**

****

**Basic proposed system:**



**Use case diagram:**



**3.2Hardware / Software designing**

**Technology:**

* Machine Learning (Supervisedlearning).
* Python.
* HTML.
* CSS.
* Bootstrap.
* Django (Framework)

**Software:**

* Jupyter Notebook.
* Pycharm

**Hard ware:**

* Processing power
* Memory (RAM)
* Storage

**4.Result**

**5.Advantages:**

1. **Improved Accuracy:** An optimized prediction system can provide more accurate forecasts of flight prices, helping travelers make better informed decisions.
2. **Cost Savings:** Travelers can potentially save money by booking flights at optimal times when prices are lower, leading to cost savings.
3. **Better Planning**: Accurate price predictions enable travelers to plan their trips well in advance, taking advantage of lower prices and availability.
4. **Competitive Edge:** Airlines and travel agencies can gain a competitive edge by offering customers accurate and timely price predictions.
5. **Enhanced Customer Experience:** Travelers appreciate transparency and accurate information, leading to a better overall experience.
6. **Data-Driven Insights:** The prediction system can provide valuable insights into pricing trends, seasonality, and demand patterns, helping airlines adjust their strategies.
7. **Automation**: Once set up, the system can automate the process of monitoring and predicting flight prices, saving time and resources.
8. **Integration Possibilities:** The system can be integrated into travel apps, websites, or platforms, enriching their offerings.

**Disadvantages:**

1. **Data Quality Dependence**: Accurate predictions rely heavily on high-quality and up-to-date data. Inaccurate or incomplete data can lead to flawed predictions.
2. **Complexity:** Designing and implementing an optimized prediction system can be complex, requiring advanced data science and technical expertise.
3. **Algorithm Performance:** The chosen prediction algorithm(s) might not always perform well, leading to less accurate forecasts.
4. **Variable Factors:** Flight prices can be influenced by a multitude of ever-changing factors, making accurate predictions challenging.
5. **External Shocks:** Unforeseen events like geopolitical issues, natural disasters, or economic crises can disrupt price patterns, affecting prediction accuracy.
6. **Ethical Concerns:** Privacy and ethical concerns may arise if the system relies on user data to make predictions.
7. **User Behavior Changes:** If users change their behavior based on predictions (e.g., booking more during predicted low-price periods), it could alter the accuracy of the model.
8. **Maintenance and Updates:** The prediction model requires ongoing maintenance, updates, and monitoring to adapt to changing market conditions.
9. **Overfitting:** Depending on the complexity of the model, there's a risk of overfitting to historical data and losing generalization capability.

**6.Applications**

An optimized flight price prediction system can have numerous applications across the travel and aviation industry, as well as for travelers and businesses. Here are some key areas where such a system can be applied:

1. **Travel Booking Platforms:** Online travel agencies (OTAs) and booking websites can integrate the prediction system to provide users with real-time insights into flight price trends, helping them make informed booking decisions.
2. **Airlines:** Airlines can use optimized price prediction to dynamically adjust their pricing strategies based on demand and market conditions, optimizing revenue management.
3. **Travel Apps:** Mobile apps focused on travel planning can incorporate the prediction system to offer users personalized travel itineraries and options for cost-effective flight bookings.
4. **Corporate Travel Management**: Businesses that manage corporate travel arrangements can benefit from accurate price predictions to optimize travel budgets and reduce expenses.
5. **Tour Operators:** Companies organizing tours and vacation packages can use price predictions to offer customers more attractive and cost-effective travel options.
6. **Price Tracking Tools:** Price tracking tools can alert travelers when flight prices drop or reach a certain threshold, enabling them to book flights at optimal times.
7. **Frequent Flyers:** Frequent flyers can use the prediction system to plan their travel schedule and redeem miles or loyalty points when flight prices are projected to be lower.
8. **Travel Deal Websites:** Websites that aggregate travel deals and discounts can incorporate price prediction to showcase the best times to book flights for maximum savings.
9. **Vacation Planners:** Travel planners can optimize itineraries by suggesting alternative travel dates based on predicted price fluctuations.
10. **Market Research:** Researchers and analysts in the travel industry can use the prediction system to study market trends, pricing dynamics, and consumer behavior.
11. **Airfare Alerts:** Services that send airfare alerts and notifications to subscribers can utilize price prediction to enhance the accuracy of their notifications.
12. **Travel Insurance Providers**: Travel insurance companies can factor in predicted flight prices when calculating insurance coverage.

**7.Conclusion**

This System customer can predict the flight price of a particular seat on a day before booking the flight tickets. It will provide an ease to the customer for the flight ticket booking. Travelers can save money if they choose to buy a ticket when its price is the lowest. The problem is how to determine when is the best time to buy flight ticket for the desired destination and period.

**8.Scope**

* For purchasing an airplane ticket, the traditional purchase strategy is to buy a ticket far in advance of the flight’s departure date to avoid the risk that the price may increase rapidly before the departure date. However, this is usually not always true, airplane companies can decrease the prices if they want to increase the sales.
* Airline companies use many different variables to determine the flight ticket prices that indicates whether the travel is during the holidays, the number of free seats in the plane etc., or even in which month it is, some of the variables are observed, but some of them are hidden.
* In this context, buyers are trying to find the right day to buy the ticket, and on the contrary, the airplane companies are trying to keep the overall revenue as high as possible. Airline companies have the freedom to change the flight ticket prices at any moment. Travelers can save money if they choose to buy a ticket when its price is the lowest.