***Flight Price and Customer Satisfaction Prediction***

by

Chandhini N

**Flight Price Prediction**

**Problem Statement**

To build a robust machine learning model capable of predicting flight ticket prices based on historical data, leveraging various features like departure time, source, destination, airline type, and total stops.

**Introduction**

Flight ticket prices fluctuate based on multiple factors such as demand, airline, route, and time of booking. Predicting flight prices helps travellers plan cost-effective trips, assists travel agencies in offering competitive pricing, and enables airlines to optimize revenue management**.**

**Domain:**

* Travel and business

**Business Use Cases:**

* Helping travellers plan trips by predicting flight prices based on their preferences.
* Assisting travel agencies in price optimization and marketing strategies.
* Enabling businesses to budget for employee travel by forecasting ticket prices.
* Supporting airline companies in identifying trends and optimizing pricing strategies.

**Data Preprocessing / Data Cleaning:**

* **Handling missing values:** Missing values are dropped from the dataset.
* **Feature Encoding:** Data columns such as Route are encoded by Label Encoding and One-Hot Encoding is used for data columns such as Airline, Destination, Additional\_Info, Source.

**Exploratory Data Analysis:**

* Histograms are used to check the flight price distribution.
* Count plots are used to perform the flight count based on the number of stops.
* Scatter plots are used to check the relation between the
* Flight duration and price
* Total stops and Price
* Variation of price between airlines
* Variation of price between airlines w.r.t route
* Variation of price based on duration w.r.t total stops
* **Findings:**
* Airlines, total stops, and departure time were significant factors affecting ticket prices.
* Flights with more layovers tended to be cheaper compared to direct flights.

**Feature Engineering**

* Extracted separate hour and minute values from timestamps.
* Extracted year, month, day from the date of journey.
* Created "price per minute" as a new feature to analyze cost efficiency.
* Performed feature selection to prioritize the most important features which are used in Streamlit application.

**Model Building**

* Machine learning models such as ‘Linear Regression’, ‘Random Forest’, ‘XGBoost’, ‘Decision tree’, ‘Gradient boosting’ are used.
* Grid Search CV techniques optimized hyperparameters.

**Evaluation metrics**

* Evaluation metrics such as Mean square error, mean absolute error, R2 score are used to evaluate the efficiency of the models.

**MLflow**

* MLflow was integrated to:
* Track experiments and log parameters, metrics, and model artifacts.
* Compare model performance across multiple iterations.
* Store trained models for future deployment.

**Final Model Selection**

* **XGBoost** was chosen as the best-performing model due to its lowest RMSE and highest R-squared value.

**Streamlit**

* A Streamlit application was built to:
  + Provide an interactive UI for users to input flight details and receive price predictions.
  + Display visualizations of flight price trends and influencing factors.

**Key Business Insights**

* Airlines can optimize pricing strategies by analyzing demand fluctuations and seasonal trends.
* Travel agencies can improve pricing recommendations by leveraging predictive models.
* Passengers can save on flight costs by booking flights with more stops or with low price per minute.

**Customer Satisfaction Prediction**

**Problem Statement**

To build a classification model that predicts whether a customer is satisfied or dissatisfied based on service-related attributes such as inflight service, seat comfort, and boarding experience.

**Introduction**

Customer satisfaction is a critical metric for airlines as it directly impacts customer retention, brand reputation, and revenue. Predicting customer satisfaction based on various service-related factors enables airlines to improve customer experience proactively.

**Domain:**

* Customer Experience

**Business use cases:**

* Enhancing customer experience by predicting and addressing dissatisfaction.
* Providing actionable insights for businesses to improve services.
* Supporting marketing teams in identifying target customer groups.
* Assisting management in decision-making for customer retention strategies.

**Data preprocessing / Data cleaning:**

* **Handling Missing Values:** Missing values are dropped from the dataset.
* **Feature Encoding:** Label encoding was used for categorical featuressuch as gender, type of travel, customer type, class, satisfaction.

**Exploratory Data Analysis (EDA):**

* Histogram is used to check the customer satisfaction rates.
* Count plots are used to find the satisfaction rate of the customers based on
* Gender
* Customer type
* Type of travel
* Class
* Bar plot is used to find the satisfaction rate based on flight distance.
  + **Findings:**
    - Online boarding, inflight entertainment, and seat comfort had the highest impact on customer satisfaction.
    - Business class passengers reported significantly higher satisfaction compared to economy class passengers.

**Feature Engineering**

* **Feature Selection:**
  + Removed less significant variables to streamline model efficiency.

**Model Building**

* Machine learning models such as ‘Logistic Regression’, ‘Random Forest’, ‘XGBoost’, ‘Decision tree’, ‘Gradient boosting’ are used.

**Class Imbalance Handling:**

* SMOTE (Synthetic Minority Over-sampling Technique) was used to balance the dataset.

**Evaluation metrics**

* Evaluation metrics such as accuracy score, precision score, recall, F1 score and confusion matrix are used to evaluate the efficiency of the models.

**MLflow**

* Log model training metrics and parameters for tracking performance.
* Store trained classification models for future evaluation and deployment.
* Automate model versioning and comparison.

**Final Model Selection**

* **XGBoost** was chosen as the best model due to its superior accuracy and robustness in handling imbalanced data.

**Streamlit**

* A Streamlit application was developed to:
  + Allow users to input customer details and predict satisfaction levels.
  + Provide graphical insights on customer satisfaction trends and influential factors.

**Key Business Insights**

* Airlines can enhance customer experience by focusing on inflight wifi service, inflight entertainment, and online boarding.
* Marketing strategies can be improved by targeting dissatisfied customers with personalized offers.
* Operational improvements can be made by addressing service aspects that negatively impact satisfaction.

**Conclusion**

The successful implementation of machine learning models for flight price prediction and customer satisfaction analysis provides actionable insights for various stakeholders in the airline industry.

* **Flight price prediction models** enable dynamic pricing strategies, helping airlines maximize revenue while allowing travellers to book cost-effective flights.
* **Customer satisfaction models** highlight critical service factors, empowering airlines to enhance passenger experiences and increase brand loyalty.
* **MLflow integration** ensures seamless model tracking, experiment logging, and version control for reproducibility.
* **Streamlit deployment** bridges the gap between data science and user interaction, offering a real-time predictive tool for end-users.

By leveraging these models, airlines can optimize operations, improve customer satisfaction, and make data-driven decisions for a competitive advantage in the aviation industry.