**FLIGHT PRICE**

**AND**

**CUSTOMER SATISFACTION**

**PROJECT**

**DOCUMENT**

# **INDEX**

* **Objective**
* **Dataset Description**
* **Data Preprocessing**
* **Modeling**
* **ML Flow**
* **Stream lit Application**

# FLIGHT PRICE PREDICTION

**Objective:**

The primary goal of this project is to predict flight prices based on various features such as departure and arrival time, airline, stops, and others. This project uses machine learning models to provide accurate flight price predictions.

**Dataset Description:**

**Dataset Name:** Flight\_Price.csv and Passenger\_Satisfaction.csv.

**Features:**

* **Airline:** Name of the airline.
* **Date\_of\_Journey:** Date of takeoff.
* **Source:** Starting airport location.
* **Destination:** Final landing airport location.
* **Route:** The route from where the plane will go and stops.
* **Dep\_Time:** Departure time.
* **Arrival\_Time:** Arrival time of the plane landing.
* **Duration:** How long the flight lasted.
* **Total\_Stops:** Number of stops between flights for fuel, etc.
* **Additional\_Info:** Additional notes from the airline (e.g., meal not included).

**Data Preprocessing:**

**Handling Missing Values:** The dataset was cleaned by dropping any rows with missing values using df.dropna().

**Feature Engineering:**

* The Date\_of\_Journey feature was split into three separate features: Day, Month, and Year.
* The Duration feature was split into two features: Duration\_Hours and Duration\_Minutes.
* The Dep\_Time feature was split into two features: Dep\_Hour and Dep\_Minute.
* The Arrival\_Time was cleaned using regular expressions and split into Arr\_Hour and Arr\_Minute.

**Label Encoding:** Categorical features like Airline, Source, Destination, and Additional\_Info were converted into numerical values using Label Encoding to make them suitable for machine learning models.

**Feature Dropping:** Unnecessary columns like Route, Date\_of\_Journey, Duration, Dep\_Time, Arrival\_Time, and clean\_Arrival\_Time were dropped.

**Modeling:**

**Independent Variables (X):** All columns except for the target variable Price.

**Target Variable (Y):** The price of the flight.

**Train-Test Split:** The data was split into training and testing datasets with a test size of 20% using train\_test\_split().

**Models Used:**

* Linear Regression
* Lasso Regression
* Ridge Regression
* Random Forest Regressor
* XGBoost Regressor

**Model Training:** Each model was trained on the training data (x\_train, y\_train).

**Evaluation Metrics:** For model evaluation, the following metrics were used:

* Root Mean Squared Error (RMSE)
* Mean Squared Error (MSE)
* R-squared Score (R²).

**ML Flow:**

* Logged experiments and metrics for each regression model using ML flow.
* Tracked parameters, metrics (e.g., RMSE, R-squared), and artifacts (e.g., model files, visualizations).

**Stream lit Application:**

* Developed a stream lit app to filter and visulaize the data from the dataset.
* Based on user interest filters are added for route, airline and time.
* Predicting the flight ticket prices based on some user inputs.

**Customer Satisfaction Prediction**

**Objective:**

This project aims to predict passenger satisfaction based on various features, such as gender, customer type, travel type, class, etc. Using different classification models, the objective is to determine the factors influencing passenger satisfaction and predict satisfaction outcomes.

**Dataset Description:**

**Dataset Name:** Passenger\_Satisfaction.csv.

**Features:**

* **Gender:** Gender of the passengers (Female, Male)
* **Customer Type:** The customer type (Loyal customer, disloyal customer)
* ***Age:*** The actual age of the passengers
* **Type of Travel:** Purpose of the flight of the passengers (Personal Travel, Business Travel)
* **Class:** Travel class in the plane of the passengers (Business, Eco, Eco Plus)
* **Flight distance:** The flight distance of this journey
* **Inflight wifi service:** Satisfaction level of the inflight wifi service (0:Not Applicable;1-5)
* **Departure/Arrival time convenient:** Satisfaction level of Departure/Arrival time convenient
* **Ease of Online booking:** Satisfaction level of online booking
* ***Gate location:*** Satisfaction level of Gate location
* **Food and drink:** Satisfaction level of Food and drink
* **Online boarding:** Satisfaction level of online boarding
* **Seat comfort:** Satisfaction level of Seat comfort
* **Inflight entertainment:** Satisfaction level of inflight entertainment
* **On-board service:** Satisfaction level of On-board service
* **Leg room service:** Satisfaction level of Leg room service
* **Baggage handling:** Satisfaction level of baggage handling
* **Check-in service:**Satisfaction level of Check-in service
* **Inflight service:**Satisfaction level of inflight service
* **Cleanliness:** Satisfaction level of Cleanliness
* **Departure Delay in Minutes:** Minutes delayed when departure
* **Arrival Delay in Minutes:** Minutes delayed when Arrival
* **Satisfaction:** Airline satisfaction level(Satisfaction, neutral or dissatisfaction).

**Data Preprocessing:**

**Handling Missing Values:** The dataset had missing values which were removed using df.dropna() to ensure that the model is trained on a clean dataset.

**Feature Engineering:**

**Label Encoding:** Categorical features such as Gender, Customer Type, Type of Travel, Class, and Satisfaction were label-encoded. Label encoding transforms each unique value in a categorical feature into an integer.

**Dropping Unnecessary Columns:** The columns Unnamed: 0 and id were dropped as they did not provide any meaningful information to the model.

**Independent Variables (X):** All the features, except for the target variable satisfaction.

**Target Variable (Y):** The target variable satisfaction, which indicates whether a passenger is satisfied or dissatisfied.

**Modeling:**

**Train-Test Split:** The dataset was split into a training set and a testing set using train\_test\_split(). 80% of the data was used for training and 20% for testing.

**Classification Models Used:** Several machine learning classification models were employed to predict passenger satisfaction:

**Logistic Regression:** A simple model to classify binary outcomes.

**AdaBoost Classifier:** An ensemble method that combines weak classifiers to form a strong classifier.

**Random Forest Classifier:** A robust ensemble method using multiple decision trees for classification.

**K-Nearest Neighbors (KNN):** A simple algorithm that classifies based on the closest data points.

**XGBoost Classifier:** A gradient boosting algorithm that is known for its performance in classification tasks.

**Model Training:** Each model was trained using the training data (x\_train, y\_train).

**Evaluation Metrics:** For model evaluation, the following metrics were used:

* Accuracy Score
* Precision Score
* Recall Score
* F1 Score.

**ML Flow:**

* Logged experiments and metrics for each classification model using ML flow.
* Tracked  parameters, metrics (e.g., accuracy, F1-score), and confusion matrices.

**Stream lit Application:**

* Developed a stream lit app to filter and visulaize the data from the dataset.
* Based on user interest filters are added for inflight services feedback.
* Predicting the customer satisfaction based on user feedback.

**THANK YOU**