Airfare Price Detection Using Machine Learning

Overview

This project focuses on building a machine learning model that can **predict airfare prices** based on various flight-related features. Accurate price prediction is beneficial for travelers, travel agencies, and airline companies to plan effectively and stay competitive.

Objectives

- Analyze the factors affecting flight prices.
- Apply machine learning regression techniques to predict airfare.
- Evaluate models and select the most accurate one.
- Build a reliable and efficient predictive system.

Dataset

- **Source**: Kaggle
- **Features**:
    - Airline
    - Date of Journey
    - Source & Destination
    - Route
    - Duration
    - Total Stops
    - Additional Info
    - Price (Target)

Technologies Used

- **Language**: Python
- **Libraries**:
    - Pandas & NumPy (Data manipulation)
    - Matplotlib & Seaborn (Data visualization)
    - Scikit-learn (ML models)
    - XGBoost (Gradient boosting)

Exploratory Data Analysis (EDA)

Key steps:

- Converted Date of Journey, Dep_Time, Arrival_Time to datetime features.
- Extracted day, month, hour for deeper insights.
- Encoded categorical variables using Label Encoding and OneHot Encoding.
- Handled null values and inconsistencies in data.

Models Implemented

Model	MAE	MSE	RMSE	R^2 Score
Linear Regression	~1800	~6000000	~2450	~0.68
Random Forest Regressor	~1100	~1500000	~1225	~0.85
ExtraTrees Regressor	~1050	~1450000	~1204	~0.86

✅ **Best Model**: ExtraTrees Regressor

📈 Performance Metrics

- **MAE (Mean Absolute Error)**: Average absolute error between predicted and actual prices.
- **RMSE (Root Mean Squared Error)**: Penalizes large errors more than MAE.
- **R^2 Score**: Indicates how well the model explains the variance in target values.

📊 Visualization

- Heatmaps for correlation analysis.
- Count plots for airline frequency, source/destination distribution.
- Distribution plots of predicted vs actual prices.

🚀 Future Work

- Deploy the model using Flask/Streamlit as a web app.
- Use LSTM or Deep Learning for time-based predictions.
- Integrate real-time pricing data using APIs.
- Apply to hotel or travel package prediction systems.

🗂️ Project Structure

```
Airfare-Price-Detection/
│
├── Airfare_Price's_Detection.ipynb     # Main code implementation
├── Research Paper.pdf                  # Detailed research documentation
├── Airfare Price Detection(ppt).pdf    # Project presentation
├── README.md                           # GitHub documentation
└── Dataset/                            # Raw and cleaned datasets
```

👨‍💻 Author

Harsh Kumar
✉️ hk6528259@gmail.com
📍 Greater Noida, Uttar Pradesh, India

📄 License