High Level Design

Flight Fare Prediction

Submitted By:

Vijay Kumar Shah

Date:

02/09/2024

High Level Design

Abstract:

The rapid growth in air travel has made it essential for travellers to anticipate flight fare fluctuations and make informed decisions. This project focuses on predicting flight fares based on various factors such as flight timing, destination, duration, and seasonal variations. Accurate fare predictions can assist travellers in optimizing their travel plans, saving both time and money.

The project employs classical machine learning techniques, beginning with data exploration and cleaning to ensure quality inputs. Feature engineering is then used to extract meaningful variables that influence fare prices. Multiple machine learning algorithms are applied to develop a robust prediction model, which is thoroughly tested for performance. The outcome is a predictive model that provides reliable flight fare estimations, offering a valuable tool for travel planning.

High Level Design

1. Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like: o Security
    - Reliability
    - Maintainability
    - Portability
    - Reusability
    - Application compatibility
    - Resource utilization
    - Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

High Level Design

1.3 Definitions

Terms	*Description*
Database	*Collection of all the information monitored by this system*
IDE AWS	*Integrated Developer Amazon Web Services*

2. General Description

2.1 Product Perspective

The Flight Fare Prediction system is a standalone application designed to predict flight fares using machine learning. It can be integrated into travel platforms, offering fare predictions based on factors like flight details, timing, and historical data. The system operates independently, providing users with an intuitive interface to input travel details and receive fare predictions. It is adaptable to changing data, ensuring ongoing accuracy and relevance.

2.2 Problem Statement

- Frequent fare fluctuations: Flight ticket prices vary based on multiple factors.
- Impact of timing and destination: Prices are influenced by flight timing, destination, and duration.
- Seasonal variations: Occasions like vacations and festivals affect fare prices.
- Need for fare prediction: Predicting flight fares can help travellers save money and plan efficiently.

2.3 Proposed Solution

Utilize machine learning techniques to predict flight fares by analysing factors such as flight timing, destination, and duration. The solution involves data cleaning, feature engineering, and applying various algorithms to develop a predictive model that accurately forecasts flight prices.

High Level Design

2.4 Further Improvements

- Incorporate real-time data for dynamic fare prediction.
- Enhance the model with advanced algorithms like ensemble methods.
- Integrate additional features like airline reputation, seat class, and booking window.
- Implement a user-friendly interface for broader accessibility.

2.5 Technical Requirements

1. **Programming Language:** Python 3.x

2. **Libraries:**

   - Data Processing: pandas, NumPy
   - Visualization: matplotlib, seaborn
   - Machine Learning: scikit-learn

3. **Development Environment:** Jupyter Notebook, PyCharm, or VS Code

4. Dataset: CSV file containing flight details (e.g., date, time, destination, fare)

5. Hardware:

   - Minimum: 4 GB RAM, Dual-core processor
   - Recommended: 8 GB RAM, Quad-core processor for faster processing

6. Software:

   - Python installed with required libraries
   - IDE or Notebook environment for development
   - Excel/CSV reader for dataset exploration

2.6. Data Requirements

The data requirements for this project include a dataset containing relevant features that influence flight fares. Essential features may include flight date, departure and arrival times, airline, source and destination cities, flight duration, number of stops, and any special occasions (e.g., holidays or festivals). The dataset should be clean, accurate, and sufficiently large to capture various patterns and trends. Additionally, the data should cover a broad time range to account for seasonal variations and provide a solid foundation for training and testing machine learning models

High Level Design

2.6. Tools Used

- **Python:** Utilized for data manipulation, model building, and evaluation with libraries such as pandas, numpy, and scikit-learn.
- **Jupyter Notebook:** Employed for interactive coding and visualization throughout the development process.
- **Matplotlib/Seaborn:** Used for data visualization to understand trends and patterns in the dataset.
- **Scikit-learn:** Applied for implementing machine learning algorithms, including regression and classification models.
- **Pandas:** Utilized for data cleaning, transformation, and handling large datasets efficiently.

High Level Design

2.7 Constraints

1. Data Quality: The accuracy of predictions depends on the quality and completeness of the dataset. Missing or erroneous data can impact model performance.

2. Feature Limitations: The available features may not capture all factors affecting flight fares, potentially limiting prediction accuracy.

3. Algorithm Performance: Different machine learning algorithms may yield varying results; optimal performance depends on the choice of algorithm and parameter tuning.

4. Seasonal Variations: The model may struggle to predict fares during unusual events or changes in travel patterns not represented in historical data.

5. Computational Resources: Training complex models and processing large datasets require significant computational power and memory.

2.8 Assumptions

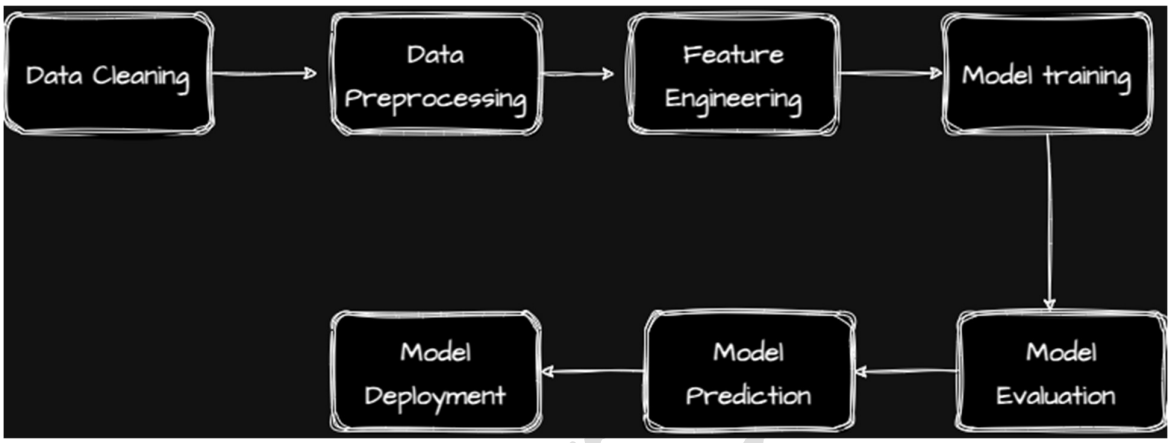
1. **Data Accuracy:** The dataset used for training and testing the model is assumed to be accurate and representative of real-world flight fare data.
2. **Feature Relevance:** The selected features, such as flight timing, destination, and duration, are assumed to significantly influence flight fare predictions.
3. **Model Generalization:** The machine learning models are assumed to generalize well to unseen data and provide reliable fare predictions for various scenarios.
4. **Data Consistency:** The dataset is assumed to be consistent in terms of formatting and units across different entries.
5. **Seasonal Effects:** The model assumes that seasonal variations and special occasions have a predictable impact on flight fares.

High Level Design

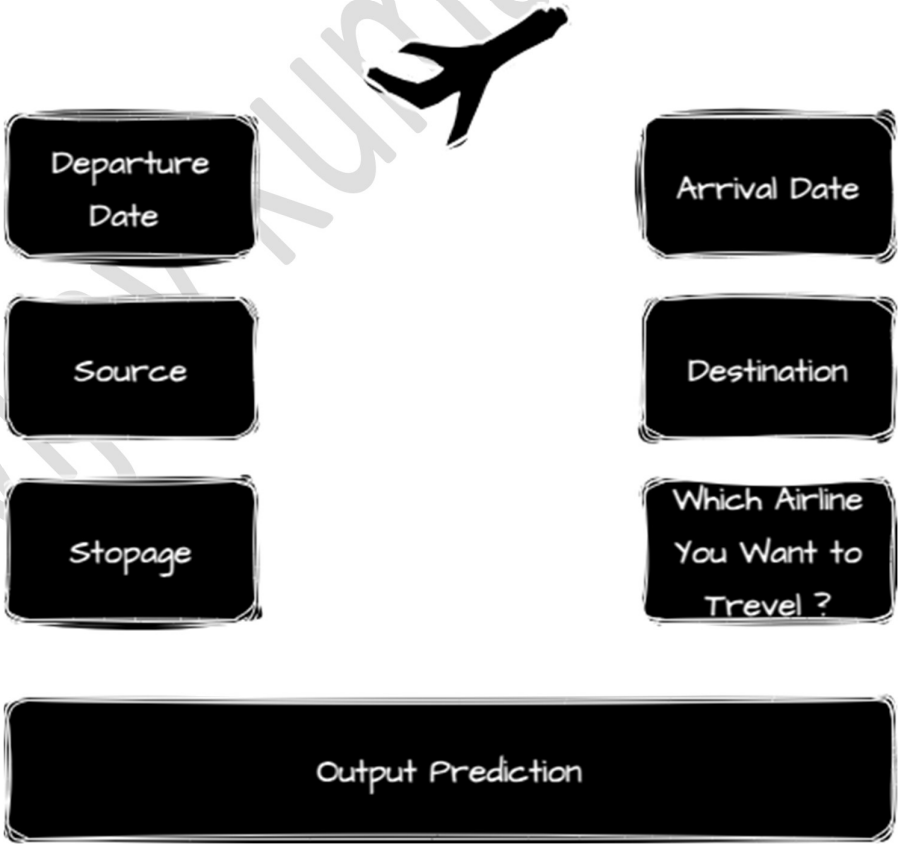
3. Design Details

3.1 Process Flow

Here is various process involve in data process like dataset download, data cleaning, anomaly detection, outliers detection and many.

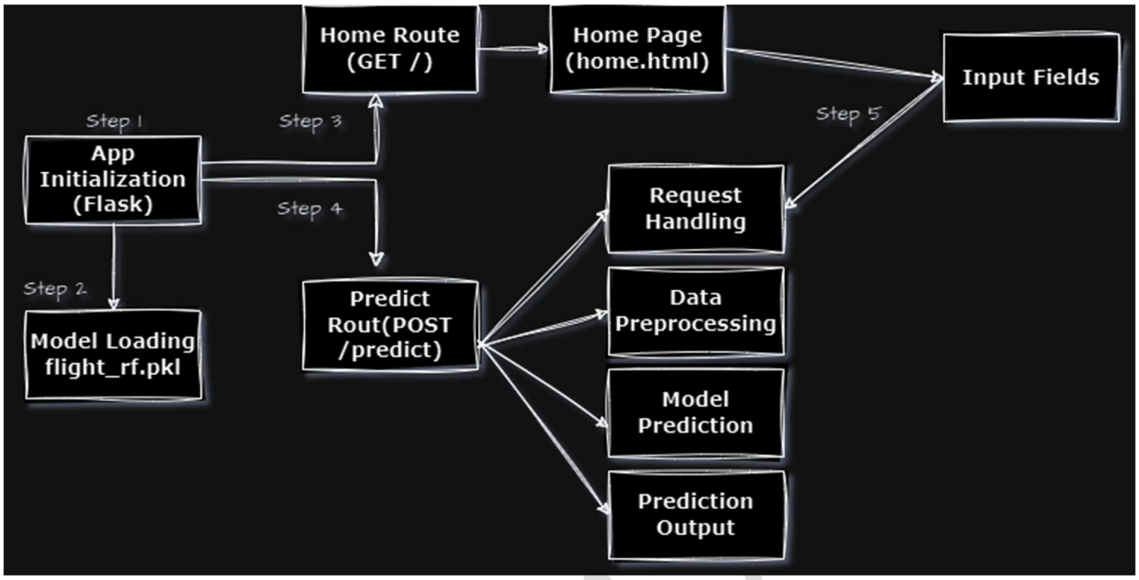


3.1 Fronted Design



High Level Design

3.1 Backend Design



4. Conclusion

The Flight Price Prediction project successfully implements a machine learning model to predict flight prices based on various features such as date of journey, departure and arrival times, total stops, airline, and source/destination locations. By leveraging Flask for the backend and HTML for the frontend, the system allows users to input flight details through a simple web interface and receive accurate fare predictions in real time. The solution addresses the problem of fluctuating airfares by providing users with data-driven insights, enabling them to make informed travel decisions.

This project demonstrates the effective application of machine learning and web development technologies to solve a real-world problem. The model can be further enhanced with additional features, more extensive datasets, or advanced algorithms to improve prediction accuracy and provide even more value to users.