

A large, light gray watermark of the 'PW' logo is centered on the page, behind the title text.

Flight Fare Prediction

Contents

Document Version Control.....	3
Abstract.....	4
1 Introduction	5
1.1 Why this High-Level Design Document?	5
1.2 Scope.	5
2 General Description.....	6
2.1 Product Perspective	6
2.2 Problem statement.....	6
2.3 PROPOSED SOLUTION	6
2.4 FURTHER IMPROVEMENTS	6
2.5 Technical Requirements.....	7
2.6 Data Requirements	7
2.7 Tools used.	7
2.7.1 Hardware Requirements.....	8
3 Design Details.....	9
3.1 Process Flow.	9
3.1.1 Model Training and Evaluation.....	9
3.1.2 Deployment Process.....	9
3.2 Event log.....	9
3.3 Error Handling.....	10
4 Performance.....	11
4.1 Reusability.....	11
4.2 Application Compatibility	11
4.3 Resource Utilization	11
4.4 Deployment.	11
5 Conclusion	12

Document Version Control

Date Issued	Version	Description	Author
16/10/202	1	Initial HLD — V1.0	Ankit Zanzmera

Abstract

The Flight Fare Prediction Project embodies a commitment to best practices in data science and machine learning. At its core, the project is characterized by a modular code structure designed to enhance maintainability and facilitate collaboration among team members. This modular approach simplifies code readability, reusability, and the seamless integration of new features or improvements.

Two significant pillars of the project's success are MLflow and DVC. MLflow, an open-source platform, is employed for model monitoring, enabling real-time tracking of model performance, parameters, and metrics. This empowers stakeholders with valuable insights into model behavior and provides a mechanism for maintaining model health over time. DVC, Data Version Control, is used for efficient data versioning and storage. It ensures the reproducibility of data pipelines and systematic tracking of data changes, ultimately enhancing data quality and mitigating the risk of inconsistencies in training and evaluation data.

The project's documentation spans a comprehensive journey from data collection and preprocessing to model development, evaluation, and deployment. It features a meticulous exploratory data analysis that uncovers critical insights, robust model selection, hyperparameter tuning to ensure model accuracy and reliability, and clear visualizations and performance metrics to aid stakeholder understanding and evaluation of project outcomes.

The Flight Fare Prediction Project sets a high standard in the data science community. It is not merely a prediction system, but a demonstration of excellence in data science, promoting transparency, collaboration, and reproducibility. The documentation provides a complete guide to the project, underlining the importance of following best practices, and discusses potential future enhancements to guide further improvements and the addition of new features.

1. Introduction:

1.1. Why this High-Level Design Document?

A High-Level Design (HLD) document is an essential component of any project, serving as a foundational guide for the entire project lifecycle. This document is indispensable for several reasons. Firstly, it aids in project planning and organization by outlining the project's architecture, components, and their interactions. This structural clarity allows for efficient project management and execution.

1.2. Scope :

In the project scope, we delineate the project's boundaries, objectives, and parameters to provide a clear understanding of its purpose and limitations. We begin by offering an overarching description of the project, elaborating on its goals and what it aims to accomplish. The specific objectives are then outlined, accompanied by a list of tangible deliverables, such as reports, software, and documentation. Within this scope, we also specify the activities to be undertaken as part of the project, clearly articulating the work processes involved.

2. General Description:

2.1. Product Perspective:

From a product perspective, the Flight Fare Prediction Project offers a sophisticated and modular solution for accurately predicting flight fares. It incorporates cutting-edge technologies, such as MLflow for model monitoring and DVC for data versioning, to ensure the highest level of performance, reliability, and data integrity. The modular code structure enhances maintainability and facilitates collaboration, while meticulous data preprocessing, model development, and performance evaluation make it a robust and trustworthy tool for users. By leveraging this solution, stakeholders can make informed decisions and anticipate airfare costs effectively, thereby optimizing travel planning and budget management. Future enhancements are also on the horizon to further enhance the capabilities of this valuable predictive tool.

2.2. Problem Statement:

The aviation industry faces a recurring challenge of fluctuating and often unpredictable flight fares, leading to traveler frustration and budgetary concerns. Passengers and travel companies require a solution to accurately forecast flight fares and streamline travel planning.

2.3. Proposed Solution:

My Flight Fare Prediction Project leverages machine learning and data science techniques to tackle this problem. By collecting and preprocessing historical flight data, we train a predictive model that estimates future fares accurately. Utilizing MLflow for model monitoring and DVC for data versioning, we ensure the reliability of our predictions. This solution empowers travelers to make well-informed decisions and businesses to optimize their offerings. Future enhancements will further refine and expand the model's capabilities, providing an even more robust tool for addressing this long-standing issue in the travel industry.

2.4. Further Improvements:

Continuing to enhance our Flight Fare Prediction Project, we can implement several improvements. These include expanding the dataset to cover more airlines and routes, integrating real-time data updates for up-to-the-minute fare predictions, and exploring the incorporation of more advanced machine learning algorithms for increased accuracy. Additionally, the development of a user-friendly mobile application or web platform would offer a more accessible and practical interface for travelers. These improvements will further elevate the project's utility and user experience, making it an even more valuable resource for both individual travelers and travel-related businesses.

2.5. Technical Requirements:

The successful operation of our Flight Fare Prediction Project necessitates several technical requirements. These include robust server infrastructure to handle data processing and model training, sufficient storage capacity for historical data and model artifacts, and a reliable web hosting environment for deployment. Additionally, the use of MLflow for model monitoring requires adequate computational resources. The project also relies on a range of software libraries and tools, including Python, scikit-learn, and MLflow for machine learning, and DVC for data version control. Furthermore, integration with APIs to access real-time data updates is crucial for maintaining data accuracy. Meeting these technical requirements ensures the project's functionality and reliability, delivering accurate flight fare predictions to users and businesses alike.

2.6. Data Requirements:

To ensure the accuracy and reliability of the Flight Fare Prediction Project, specific data requirements are essential. These encompass access to comprehensive historical flight data, including fare details, routes, airlines, and seasonal variations. Real-time flight data feeds are crucial for continuous model updates. Data preprocessing capabilities, including data cleaning, normalization, and feature engineering, are required for optimal model performance. Moreover, integration with DVC for data versioning and management is necessary to maintain data integrity. Adherence to data privacy regulations and security measures is essential to safeguard user information. Meeting these data requirements ensures the project can generate precise and dependable flight fare predictions for travelers and industry stakeholders.

2.7. Tools used:



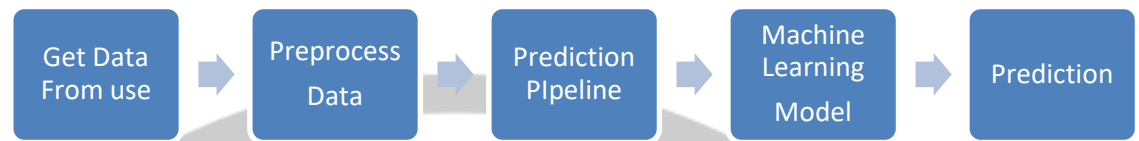
2.7.1. Hardware Requirements:

- Windows 7 or above
- Internet Connection

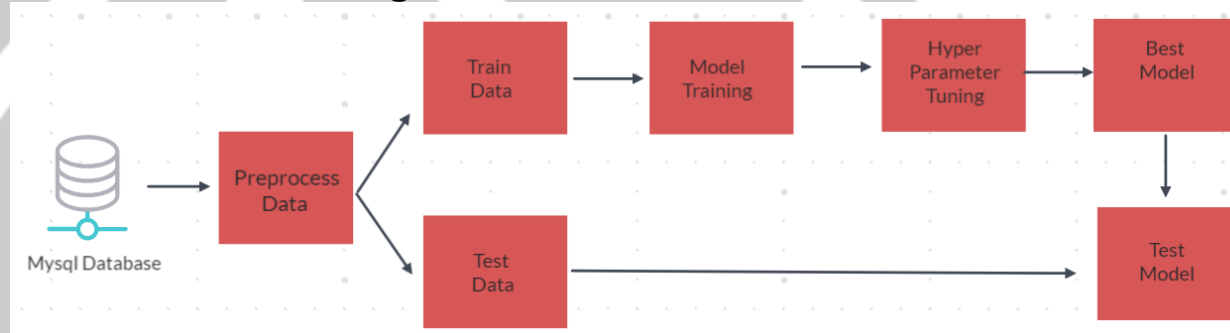


3. Design Details:

3.1. Process Flow:

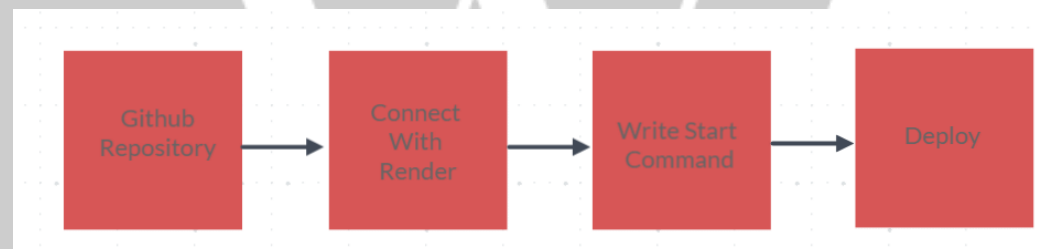


3.1.1. Model Training and Evaluation:



3.1.2. Deployment Process:

I depoloyed my Project on [Render](#) Which is Fastest Way to host Project..



3.2. Event Log :

An event log is a record of events and actions within a computer system or software application. These logs are crucial for monitoring, diagnosing issues, and enhancing security. They contain details like event type, source, and time, making them valuable for system maintenance and troubleshooting. Event logs help in understanding what's happening in a digital environment.

3.3. Error Handling:

Error handling is the practice of dealing with unexpected issues in software. This includes using techniques like exception handling, error codes, and logging to manage errors and ensure a program runs smoothly even when things go wrong. Proper error handling helps with debugging and improves user experience.



4. Performance:

4.1. Reusability:

we emphasize the importance of writing code in a modular, structured, and reusable manner. This approach significantly enhances the efficiency and maintainability of the project.

The codebase is designed to maximize reusability in several key ways:

- Modular Structure
- Functions and Libraries
- Standardized Interface
- Object-Oriented Design
- Version Control

4.2. Application Compatibility:

Compatibility is a critical aspect of our project, and we address it through the following strategies:

- Cross-Browser Compatibility
- Mobile Responsiveness
- Cross-Platform Compatibility

4.3. Resource Utilization:

When any task is performed, it will likely use all the processing power available until that function is finished.

4.4. Deployment:



5. Conclusion:

In conclusion, the Flight Fare Prediction Project represents a significant step forward in the realm of data-driven travel planning and fare forecasting. By leveraging advanced technologies like MLflow and DVC, the project demonstrates its commitment to best practices in data science and machine learning. Its modular code structure, meticulously organized and documented, not only enhances code maintainability but also encourages collaboration among team members.

The project's primary goal of accurately predicting flight fares has been achieved through thorough data preprocessing, robust model development, and in-depth performance evaluation. The resulting predictive model offers travelers and industry stakeholders a valuable tool for making informed decisions, optimizing travel plans, and effectively managing budgets.

In summary, the Flight Fare Prediction Project not only delivers a powerful predictive system but also serves as a beacon of best practices in data science, fostering transparency, collaboration, and reproducibility. As it continues to evolve, it will contribute to a more informed and efficient travel landscape, making it an invaluable asset for both individual travelers and the travel industry as a whole.