Project Initialization and Planning Phase

Date	04 June 2024
Team ID	SWTID1720183095
Project Title	Ecommerce Shipping Prediction Using Machine Learning
Maximum Marks	3 Marks

Project Proposal (Proposed Solution)

The difficulty of estimating the shipping time for eCommerce orders is addressed in this project proposal. The defined scope includes the analysis of important data variables like Warehouse_block, Mode_of_Shipment, Customer_care_calls, Customer_rating, Cost_of_the_Product, Prior_purchases, Product_importance, Gender, Discount_offered, Weight_in_gms, and Reached_on_Time_Y.N. with the clear goal of increasing delivery efficiency and customer satisfaction. The clear problem statement centers on precisely estimating, given these conditions, whether an order will arrive on schedule. The suggested remedy describes a machine learning strategy that uses past data to train models of prediction. The incorporation of real-time data processing and intuitive interfaces for tracking forecasts are important elements. to design, build, and maintain the system with engineers.

Project Overview	
Objective	The goal of this project is to create a machine learning model that can forecast shipping times for online purchases, enhancing both operational and customer efficiency.
Scope	Developing a machine learning model to predict shipping times based on historical order data.
Problem Statement	
Description	It aims for improving customer experience and operational efficiency using machine learning.
Impact	This would enable companies to enhance order fulfilment procedures overall and give clients accurate delivery estimates.
Proposed Solution	
Approach	1. Data Collection and Preprocessing: Collect and clean data, handle missing values, encode categorical variables. 2. Exploratory Data Analysis (EDA): Conduct statistical analysis and visualizations to identify patterns. 3. Model Selection: Evaluate algorithms like Random Forest, KNN, and SVM with cross-validation. 4. Model Training and Tuning: Train models and optimize hyperparameters using grid search or random search. 5. Model Evaluation: Assess models with metrics such as MAE, RMSE, Confusion Matrix and F1-score to select the best model. 6. Integration and Deployment: Develop and deploy RESTful APIs or microservices for real-time predictions. 7. Monitoring and Maintenance: Implement continuous monitoring, periodic retraining, and address model drift and data quality

	issues.
Key Features	1. Comprehensive Data Integration: Incorporates a wide range of variables, including Warehouse block, Mode of Shipment, Product importance, and more, to capture all relevant factors affecting delivery times. 2. Real-time Processing: Supports real-time data integration and prediction, enabling dynamic updates and timely decision-making. 3. User-friendly Interface: Provides an intuitive interface or API for seamless integration with existing eCommerce systems, ensuring ease of use for non-technical users. 4. Continuous Learning: Employs continuous monitoring and periodic retraining of models to maintain accuracy and adapt to changing conditions. 5. Scalability: Designed to handle large datasets and high transaction volumes, ensuring the solution can scale with business growth.

Resource Requirements

Resource Type	**Description**	**Specification/Allocation**
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	T4 GPU
Memory	RAM specifications	8 GB
Storage	Disk space for data, models, and logs	1 TB SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, matplotlib, pycharm
Development Environment	IDE, version control	Jupyter Notebook, Spyder
Data		
Data	Source, size, format	Kaggle dataset, 614, csv UCI dataset,690, Performance Monitoring