Project Initialization and Planning Phase

Date	08 July 2024
Team ID	SWTID1720084775
Project Title	Ecommerce Shipping Prediction using Machine Learning
Maximum Marks	3 Marks

Project Proposal (Proposed Solution)

This project proposal outlines a solution to address the challenge of predicting the shipping time for eCommerce orders. With a clear objective of improving delivery efficiency and customer satisfaction, the defined scope includes the analysis of key data variables such as 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. The concise problem statement focuses on accurately forecasting whether an order will reach on time based on these factors. The proposed solution details a machine learning approach, leveraging historical data to train predictive models. Key features include the integration of real-time data processing and user-friendly interfaces for monitoring predictions. Resource requirements encompass robust hardware for data storage and processing, sophisticated software for machine learning algorithms, and a skilled team of data scientists and software engineers to develop, implement, and maintain the system.

Project Overview	
Objective	Develop a machine learning model to accurately predict eCommerce order delivery times, enhancing operational efficiency and customer satisfaction.
Scope	The project scope includes creating, deploying, and assessing a machine learning model to forecast eCommerce delivery times with and end-user applications excluding broader logistics optimization.
Problem Statement	
Description	The problem to be addressed is the uncertainty in predicting whether eCommerce orders will be delivered on time, which leads to decreased customer satisfaction and operational efficiencies.

Impact	Solving the problem of accurately predicting eCommerce order delivery times has significant implications. It will enhance customer satisfaction by providing reliable delivery estimates, reduce operational costs by optimizing logistics, and improve inventory management by aligning stock levels with delivery schedules. Additionally, it can lead to better resource allocation, higher customer retention, and a competitive advantage in the market due to improved service reliability.
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	Tesla T4 GPU
Memory	RAM specifications	15GB
Storage	Disk space for data, models, and logs	cloud storage used via Colab, 78.2GB
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, seaborn
Development Environment	IDE, version control	Google Colab, Git
Data		
Data	Source, size, format	"kaggle.com/datasets/prachi13/customer_analytics", 440.46kB, csv file