Project Initialization and Planning Phase

Date	04 June 2024
Team ID	SWTID1720096620
Project Title	E-commerce Shipping Prediction Using Machine
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

This project proposal presents a comprehensive solution to tackle a specific problem. It includes a clear objective, a well-defined scope, and a succinct problem statement. The proposed solution elaborates on the methodology, key features, and necessary resources, encompassing hardware, software, and personnel requirements.

Project Overview	
Objective	The primary goal of this project is to develop a predictive model capable of accurately forecasting whether an e-commerce shipment will be delivered on time.
Scope	This project focuses on developing and validating a machine learning model using historical shipment data to predict on-time deliveries. It encompasses data preprocessing, feature engineering, model selection, and evaluation. The project is confined to the provided dataset and does not include real-time data integration or external factors beyond the dataset's scope.
Problem Statement	
Description	The issue at hand is the unpredictability of e-commerce shipment delivery times, leading to customer dissatisfaction and a loss of trust in the platform. Despite promises of timely delivery, various factors such as logistical inefficiencies, product characteristics, and shipment methods result in delays. This project aims to identify and analyze these factors to create a predictive model that accurately forecasts whether a shipment will be delivered on time, thereby improving the reliability of delivery estimates and enhancing overall customer satisfaction.
Impact	Addressing the problem of unpredictable e-commerce shipment delivery times will have significant positive implications. Accurate

	and reliable delivery predictions will enhance customer satisfaction by setting clear and realistic expectations, thereby building trust in the e-commerce platform. This can lead to increased customer loyalty, positive reviews, and repeat business. Additionally, better predictability can optimize logistics operations, reduce costs associated with delays and returns, and improve overall efficiency.
Proposed Solution	
Approach	The methodology for this project involves several key steps. Initially, data exploration and preprocessing will be conducted to clean the dataset and handle any missing values or inconsistencies. Feature engineering will be performed to create meaningful variables that could enhance the predictive power of the model. Various machine learning techniques, such as logistic regression, decision trees, random forests, and gradient boosting, will be employed to develop the predictive model. The dataset will be split into training and testing sets to evaluate model performance using accuracy, precision, recall, and F1-score metrics. Cross-validation techniques will be used to ensure the model's robustness and prevent overfitting. Finally, the best-performing model will be selected for integration into the e-commerce platform, providing real-time delivery predictions to improve customer satisfaction.
Key Features	The proposed solution is unique in its comprehensive approach to addressing the unpredictability of e-commerce shipment delivery times. By leveraging a combination of advanced machine learning techniques and thorough data analysis, the model will not only predict delivery times but also identify the key factors contributing to delays. This dual focus allows for more precise and actionable insights, enabling the e-commerce platform to address root causes and optimize logistics operations.

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	Intel Core i7-12700H, NVIDIA GeForce RTX 3060 (6GB GDDR6)
Memory	RAM specifications	16 GB DDR5 RAM

Storage	Disk space for data, models, and logs	1 TB PCIe NVMe SSD
Software		
Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, seaborn, pickle, sklearn, matplotlib, xgboost
Development Environment	IDE, version control	Jupyter Notebook, Git
Data		
Data	Source, size, format	Kaggle dataset (10999, 12) 430KB