Project Initialization and Planning Phase

Date	9 July 2024
Team ID	**SWTID1720115788**
Project Title	**Ecommerce Shipping Prediction Using Machine Learning**
Maximum Marks	3 Marks

Project Proposal :

Ecommerce Shipping Prediction with the Help of Machine Learning.Ecommerce shipping prediction is an approach to estimating the successful delivery of the product on time. This shall be based on a variety of factors such as origin and destination of the package, shipping method selected by the customer, carrier used for shipping, or any potential delays or problems risen during the process. These would be the right models that would enable a business organization to make correct predictions of shipment times based on past data and real-time updates from carriers. The model considers many extra factors, such as weather conditions, traffic, or other external variables that might influence delivery time. Overall, any ecommerce enterprise that will deal with the customer by providing an estimated delivery time in real-time and enhancing customer experience must have ecommerce shipping prediction as a critical tool.

Project Overview	
Objective	The core objective is to develop a machine learning model that performs on-time delivery predictions for ecommerce shipments. This shall involve checking on history in shipping and a number of other factors, which may also include the origin and destination, the shipping method chosen by the customer, the career to be used in the shipment, and delays or problems that may be encountered in the process of shipment. The ultimate purpose is enhancing customer experience through delivery estimate accuracy, which would lead to increased levels of customer satisfaction and loyalty. The objective of this project is to harness the power of machine learning in tackling one very important challenge in the ecommerce industry, geared toward the continuous growth and success of the industry.

Scope	A project that takes into consideration historic shipping data, real-time carrier updates, and external factors, but not irrelevant variables like details of the products. The target of the project was areas that have a reliable record of the data, such as North America and Europe, with standard ways of shipment using major carriers, where it excluded niche ways and minor carriers. This project considers weather, traffic, public holidays, and known delays, but excludes infrequent one-time events. Used technologies include ones for machine learning algorithms, data preprocessing tools, and integration technologies. No advanced AI unrelated to predictions is applied.
Problem Statement	
Description	Delivery time estimation is the major challenge that faces e-commerce businesses. if the delivery time is poorly estimated, customers become dissatisfied and leave negative reviews-winning less business. Shipping logistics are indeed very complex and influenced by dint of various factors, including origin location, destination location, shipping method, carrier performance, weather conditions, and traffic, amongst other external factors that make it rather hard to offer a voidable estimate of a precise delivery time. Most traditional methods that are used in delivery time estimation totally ignore the dynamic nature of these factors, hence ending up providing unreliable predictions. There is, hence, the need to have a robust solution that embeds machine learning in analyzing historical data and real-time updates, enabling ecommerce businesses to provide perfect delivery estimates for overall improvement in customer experience.
Impact	The prevention of inaccuracy in delivery prediction, one can definitely pinpoint a basic advancement in customer satisfaction and building trust in e-commerce by such delivery estimation. This could be very effective in positively increasing sales for customer retention in better reviews. For businesses, it will optimize logistics and inventory management to reduce operational costs and improve overall efficiency. Better rapport with the shipping carrier will generate effective collaboration and data sharing among them for developing smooth and efficient supply chain flow.

Proposed Solution	
Approach	The historical shipment data and real-time carrier updates will be harvested systematically, along with other exogenous variables. This comprises weather and traffic among others. Data cleaning, normalization, and feature engineering shall be performed to get meaningful information from the obtained data. Random Forest, Gradient Boosting, Neural Networks are some of the machine learning models that will be selected, trained, and fitted using hyperparameter tuning. Model performance metrics, relevant to the problem at hand, will be included, besides cross-validation. An API will be developed for real-time prediction and further integrated into the ecommerce platform. Now, models are required to be continuously monitored and refreshed so they stay accurate over time, helping customers and businesses have user-friendly dashboards through which they can trace shipments and performance metrics. The uniqueness lies in the fact that, with regard to the real-time data use, API development, and feedback loop underpinning continuous learning itself.
Key Features	The approach shall focus on the following features: real-time integration of weather, traffic, and carrier delay data to improve prediction accuracy, API for seamless delivery time prediction in real time, and user-friendly customer and business dashboards for shipment tracking and performance metrics. In addition, it highlights feature engineering, tuning of hyperparameters for model optimization, and a continuous learning feedback loop for long-term accuracy and reliability.

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	AMD Ryzen 5 5500H with Radeon Graphics
Memory	RAM specifications	16GB
Storage	Disk space for data, models, and logs	e.g., 1 TB SSD
Software		

Frameworks	Python frameworks	Flask
Libraries	Additional libraries	scikit-learn, pandas, numpy, collections, mathplotlib, seaborn, missingno, pickle
Development Environment	IDE, version control	Jupyter Notebook, Git, SPYDER
Data		
Data	Source, size, format	Kaggle dataset, 10999 observations of 12 variables.