Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720452383
Project Title	Ecommerce Shipping Prediction Using Machine
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

To enhance the accuracy of shipping predictions using machine learning, a meticulous data collection plan is crucial. This involves gathering historical shipping data, order details, and external factors such as weather conditions, traffic patterns, and carrier performance. The template will guide the identification and validation of raw data sources, ensuring data integrity and quality. By systematically preprocessing and curating this data, we create a robust foundation for developing and deploying predictive models, leading to informed decision-making and improved ecommerce logistics.

Data Collection Plan Template

Section	Description
Project Overview	The goal of this machine learning project is to enhance ecommerce shipping predictions by leveraging historical shipping data and external factors such as weather, traffic, and carrier performance. By developing an accurate predictive model, we aim to provide real-time shipping time estimates, improving customer satisfaction and operational efficiency
Data Collection Plan	The data sources have been collected from sources like git which envolves the details of the vehicle and there effective shipping time
Raw Data Sources Identified	The raw dataset being used in this model training derived from Kaggle dataset and ucl which are highly popular platform for collecting dataset for datascience competition and repositories

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	This dataset ivolves the details regarding the goods like there warehouse id mode of transportation customer rating discount rates etc.	https://www.kaggle.com/datasets/prachi13/customer-analytics?select=Train.csv	CSV	15 kb	Public