Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720097765
Project Title	Ecommerce Shipping Prediction Using Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report

Gathering, validating, and storing all necessary data systematically to ensure accuracy and completeness for predicting e-commerce shipping outcomes.

Data Collection Plan

Section	Description
Project Overview	• This machine learning project aims to predict whether e- commerce shipments reach their destinations on time based on some features like Warehouse block, Product importance, Mode of Shipment, Weight in grams. • The objective is to improve decision-making processes and customer satisfaction through accurate shipment predictions.
Data Collection Plan	• Search for datasets which have shipment details, product details • Prioritize datasets with diverse demographic information.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as Gender, shipment details, customer details, product details for machine learning analysis.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The data set contains customer details (ID, Customer care calls, Customer rating, Prior purchases, Gender), product details (Cost of the Product, Product importance, Discount offered, Weight in gms, Shipment details (Warehouse block, Mode of Shipment, Reached on Time)	https://www.kaggle.com/datasets/prachi13/customer-analytics?select=Train.csv	CSV	440KB	Public