Data Collection and Preprocessing Phase

Date	20 July 2024
Team ID	SWTID1720519736
Project Title	Ecommerce Shipping Prediction UsingMachine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysisand decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	The machine learning experiment uses a variety of shipping related characteristics to forecast when goods will arrive on time. To create a model that accurately predicts whether a shipment will arrive on time or not will require the use of a dataset that includes features like warehouse block, mode of shipment, customer care calls, customer rating, cost of the product, prior purchases, product importance, gender, discount offered, and weight in grams. This will enable the logistics and e-commerce sectors to make informed and efficient decisions.
Data Collection Plan	Search for datasets related to e-commerce shipping, logistics, anddelivery times.

	Prioritize datasets with diverse features covering different aspects of the shipping process, such as product details, customer interactions, and shipping methods. Ensure datasets include target variable for predicting on-timedelivery. Focus on obtaining data from reliable sources to ensure data qualityand integrity.
Raw Data Sources Identified	The raw data sources for this project are datasets from Kaggle, a popular platform for data science competitions and archives. The sample data provided represents a subset of the collected data, which includes variables such as inventory space, delivery method, customer service calls, customer rating, product price, past purchases, product importance, gender, discount offered, and weight in grams..

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	ID Warehouse_block, Mode_of_Shipment Customer_care_calls	https://www.kaggle.com/datasets/prachi13/customer-analytics?select=Train.csv	CSV	124 Kb	Public

	Customer_rating, Cost_of_the_Product ,Prior_purchases Product_importance Gender,Discount_offered Weight_in_gm,Reached.on.Time_Y.N				
…	…	…	…	…	…