

Data Collection and Preprocessing Phase

Date	21 July 2024
Team ID	SWTID1720530286
Project Title	Ecommerce Shipping Prediction Using Machine 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 analysis and decision-making endeavor.

Data Collection Plan Template

Section	Description
Project Overview	The machine learning project aims to predict the timely delivery of shipments based on various features related to the shipping process. Using a dataset with features such as 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, the objective is to build a model that accurately classifies whether a shipment will reach on time or not, facilitating efficient and informed decision-making in the logistics and e-commerce sectors.
Data Collection Plan	Search for datasets related to e-commerce shipping, logistics, and delivery times.

	<p>Prioritize datasets with diverse features covering different aspects of the shipping process, such as product details, customer interactions, and shipping methods.</p> <p>Ensure datasets include target variable for predicting on-time delivery.</p> <p>Focus on obtaining data from reliable sources to ensure data quality and integrity.</p>
Raw Data Sources Identified	<p>The raw data sources for this project include datasets obtained from Kaggle, a popular platform for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as 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.</p>

Raw Data Sources Template

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

	Customer_rating,Co st_of_the_Product,P rior_purchases Product_importance Gender,Discount_of ferred Weight_in_gm,Reac hed.on.Time_Y.N				
...