Data Collection and Preprocessing Phase

Date	**10 July 2024**
Team ID	**SWTID1720116037**
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 main aim of this "Ecommerce Shipping Prediction Using Machine Learning" project is to predict the shipping time whether a product in e-commerce platform can reach us at specific timing. Therefore, proper estimations of the time a good will reach are expected to result in governed cost overruns and an increased level of customer satisfaction+loyalty which can be only achieved by analyzing tons of data about popularity ranks(P), logistic distributions(L) where currently ML techniques used widely(Characterization analysis for example using product type, Customer location or methods shipments as factors influencing delivery timing. From satisfying every customer to also ensuring

	that the buyer has a richer experience with its ecommerce platform by delivering products on time. The best part is that the same becomes a sure shot way of making such an ecommerce store successful and maintainable business in future.
Data Collection Plan	Sales data: Extract all data that will include product ID, order date, shipping date, delivery date, customer location, warehouse location, and shipping method from the sales records of the company. Product Info: This includes the weight of the sold products, dimension, category
Raw Data Sources Identified	The raw data sources for this project are: datasets obtained from Kaggle , one of the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, ware house block, customer care calls, cost of product , customer rating, prior purchase, product importance, reached on time, discount offer , Gender , mode of shipment machine learning analysis.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1	ware house block, customer care calls, cost of product , customer rating, prior purchase, product	https://www.kaggle.com/datasets/prachi13/customer-	CSV	440.46 KB	Public

	importance, reached on time, discount offer , Gender , mode of shipment	analytics?select= Train.csv			