Data Collection and Preprocessing Phase

Date	04 June 2024
Team ID	SWTID1720096620
Project Title	E-commerce Shipping Prediction Using Machine Learning
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

Optimizing our e-commerce shipment prediction project with a detailed Data Collection Plan and an in-depth Raw Data Sources report. This approach ensures precise data gathering and integrity, forming a robust base for accurate analysis and reliable predictive modeling. By focusing on well-curated historical shipment data, this project aims to enhance delivery reliability and customer satisfaction through informed decision-making and optimized logistics operations.

Data Collection Plan Template

Section	Description
Project Overview	This machine learning project aims to develop a predictive model that accurately forecasts whether e-commerce shipments will be delivered on time. By analysing historical shipment data and identifying key factors influencing delivery times, the project seeks to enhance delivery reliability and improve customer satisfaction. The ultimate objective is to provide a robust predictive tool that can be integrated into the e-commerce platform, offering customers transparent and reliable delivery estimates while optimizing logistics operations.
Data Collection Plan	The data is collected from the Skill Wallet platform, specifically from a dataset available on Kaggle.
Raw Data Sources Identified	The dataset used for model building contains 10,999 observations across 12 variables.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1 - Skill Wallet platform	The dataset contains 10,999 observations across 12 variables, offering a comprehensive set of historical shipment data to analyze and identify key factors affecting delivery times.	https://www.kaggle.com/datasets/prachi13/customer-analytics?select=Train.csv	CSV	430 KB	Public