Data Collection and Preprocessing Phase

Date	15 july 2024
Team ID	740040
Project Title	Predicting CO2 emissions by countries 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 CO2 emissions of countries based on various socio-economic and environmental factors. Using a dataset with features such as GDP, population, energy consumption, and industrial output, the objective is to build a model that accurately predicts CO2 emissions, facilitating efficient and informed decision-making for environmental policy and strategy.
Data Collection	Search for datasets related to CO2 emissions. - Prioritize datasets with comprehensive global coverage and diverse socio-economic factors
Raw data resources identified	The raw data sources for this project include datasets obtained from platforms like kaggle, uci, world bank, and international environmental agencies. The provided sample data represents a subset of the extensive datasets available in these repositories.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset	It is the actual data set used to train the model for performing various actions . There are many features which are responsible for CO-2 Emission of Countries, e.g. Country Name, Country Code, Indicator Name etc.	https://www.kaggle.com/code/ashukr/exploring-co2-emission/notebook?select=Indicators.csv	CSV	574.31 MB	Public