

Data Collection and Preprocessing Phase

Date	July 2024
Team ID	739887
Project Title	Drug classification 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	Machine learning has become a powerful tool in the field of drug discovery and development. It is used extensively for drug classification, which involves categorizing drugs based on various criteria such as their chemical structure, therapeutic use, mechanism of action, or adverse effects. Here's an overview of how machine learning is applied to drug classification:
Data Collection Plan	<ol style="list-style-type: none"> 1. Search for datasets related to Abalone Age. 2. Prioritize datasets with diverse information.

Raw Data Sources Identified	The raw data sources for this project include 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, encompassing variables such as Age, Sex, BP, Cholesterol, Na_to_K and drug
-----------------------------	---

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size
Dataset 1	Dataset contains features like Age, Sex, BP, Cholesterol, Na_to_k and Drug.	https://www.kaggle.com/datasets/prathamtripathi/drug-classification	CSV	72 KB