

Data Collection and Preprocessing Phase

Date	09 July 2024
Team ID	SWTID1720080033
Project Title	Anemia Sense: Leveraging Machine Learning For Precise Anemia Recognitions
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	<p>The Anemia Sense project is a model designed to leverage machine learning algorithms to accurately predict cases of anemia, a condition characterized by the deficiency of RBCs or hemoglobin in the blood, when given a set of relevant data.</p> <p>By predicting such a condition, diagnosis, treatment plans and overall monitoring can become more effective and efficient.</p>
Data Collection Plan	The Data will be collected from kaggle.com.

Raw Data Sources Identified	Anemia.csv – kaggle dataset that holds fields such as Gender, Hemoglobin, MCH, MCHC, MCV and dependent field Result
-----------------------------	---

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Dataset 1 (Anemia.csv)	The kaggle dataset on anemia patients containing relevant data points and correct classifications.	https://drive.google.com/file/d/1KMJFNFGwoaQoAouIPabMEHcT1bvqEXau/view?usp=sharing	CSV	34 KB	Public