

Data Collection and Preprocessing Phase

Date	15 June 2025
Team ID	SWTID1749621188
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	The project aims to build a machine learning model to detect anemia in patients using hematological data such as hemoglobin, MCV, MCH, and others.
Data Collection Plan	The dataset was made available to us on the Smart Internz platform
Raw Data Sources Identified	A single CSV dataset containing features like Haemoglobin, MCV, MCH, and Result (anaemic or not)

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions

Dataset 1	<p>A structured dataset containing anonymized blood test records used for anemia detection.</p> <p>Columns include Hemoglobin, MCH, MCV, MCHC, and binary result label.</p>	https://drive.google.com/file/d/1KMJFNFGwoaQoAouIPabMEHcT1bvqEXau/view?usp=sharing	CSV	34 KB	Public
-----------	---	---	-----	----------	--------