

Data Collection and Preprocessing Phase

Date	8 August 2025
Skillwallet ID	SWUID20250188620
Project Title	Anemia Sense: Leveraging Machine Learning for Precise Anemia Recognition
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

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 prediction endeavor.

Data Collection Plan:

Section	Description
Project Overview	The machine learning project aims to predict anemia risk based on patient blood parameters. Using a dataset with features such as gender, hemoglobin, MCH, MCHC, and MCV, the objective is to build a model that accurately classifies anemia status (present or absent), enabling early detection and proactive health management.
Data Collection Plan	<ul style="list-style-type: none"> ● Search for datasets related to anemia diagnosis and hematological health parameters. ● Prioritize datasets with diverse demographic and medical information.
Raw Data Sources Identified	The raw data sources for this project include dataset obtained from smartinternz. The provided dataset contains key variables such as gender, hemoglobin level, MCH, MCHC, and MCV values for machine learning analysis.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
smartinternz	The dataset comprises patient gender, hemoglobin levels, MCH, MCHC, MCV, and anemia diagnosis results.	https://drive.google.com/file/d/1KMJFNFGwoaQoAouIPabMEHcT1bvqEXau/view	CSV	34 kB	Public