Data Collection and Preprocessing Phase

Date	6 August 2025
Skill wallet ID	SWUID20250185217
Project Title	Anemia Sense: Leveraging Machine Learning for Precise Anemia Recognition
Maximum Marks	2 Marks

Data Quality Report:

The **Data Quality Report** provides a comprehensive summary of all issues identified in the selected anemia dataset, including their type, frequency, and severity level, along with well-defined resolution plans. It plays a crucial role in assessing the overall health of the dataset by highlighting potential problems such as missing values, inconsistent data formats, duplicate records, or class imbalances. Addressing these issues ensures that the dataset is clean, consistent, and reliable, thereby improving the accuracy and robustness of the machine learning models. This systematic process not only enhances data integrity but also lays a strong foundation for effective analysis, preprocessing, and model training.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
anemia.csv	Gender column stored as categorical values (Male/Female)	Moderate	Applied label encoding (Male = 1, Female = 0)
anemia.csv	Outliers in Hemoglobin and MCV values	Low	Retained medically valid extreme values after verification
anemia.csv	No missing values in dataset	None	No action needed