

Data Collection and Preprocessing Phase

Data Quality Report:

The data quality report outlines the accuracy, completeness, consistency, and reliability of the dataset used in the project. It helps in identifying missing values, outliers, and potential data entry errors. Regular analysis ensures the dataset is clean, standardized, and suitable for processing, which improves the performance and validity of the model or application built upon it.

Data Quality Report:

Column Name	Data Type	Missing Values	Outliers Detected	Value Range	Remarks
Gender	Categorical (0: Female, 1: Male)	No	No	0 – 1	Binary encoded, evenly distributed
Hemoglobin	Float	No	Possibly (low values < 8)	6.6 – 16.9 g/dL	Normal range: 12–16 (F), 13–17 (M)
MCH	Float	No	Possibly	16 – 30 pg	Normal range ~27–31 pg; some low values
MCHC	Float	No	Possibly	27.8 – 32.5 g/dL	Normal range ~32–36 g/dL; may be slightly low
MCV	Float	No	Possibly	69.4 – 101.6 fL	Normal range ~80–100 fL; some borderline cases
Result	Categorical (0: Non-Anemic, 1: Anemic)	No	No	0 – 1	Balanced target variable