

Data Collection and Preprocessing Phase

Date	15 March 2024
Team ID	SWTID1720076203
Project Title	Anemia Sense: Leveraging Machine Learning for Precise Anemia Recognitions
Maximum Marks	2 Marks

Data Quality Report

The Data Quality Report summarizes data quality issues from the selected source, including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Female count is observed to be more than male count	Moderate	<p>The dataset has more females than males, it means the data is imbalanced. To address this issue, Undersampling can be used.</p> <p>Undersampling involves reducing the number of samples in the majority class (females, in this case) to match the number of samples in the minority class (males). This creates a balanced dataset where both classes have an equal number of samples.</p>