

## Data Collection and Preprocessing Phase

Date	7th July 2025
Team ID	SWTID1750822736
Project Title	Fault Detection using transfer learning
Maximum Marks	2 Marks

### Data Quality Report:

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

### Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	Missing or corrupt image files, inconsistent image sizes	Moderate	Checked for corrupt/missing images, ensured consistent resizing to (224, 224, 3)
Kaggle Dataset	Class imbalance – more examples in one category (Faulty vs Good)	Moderate	Applied SMOTE (Synthetic Minority Oversampling Technique) to balance classes
Kaggle Dataset	Categorical label ("Faulty", "Non-Faulty") not interpretable directly by ML models	Moderate	Applied label encoding: "Faulty" → 1, "Non-Faulty" → 0
CNN Features	Features extracted from CNN layers vary in scale and distribution	Moderate	Applied Standard Scaling to normalize the feature vectors