

Data Collection and Preprocessing Phase

Date	4 July 2024
Team ID	SWTID1720093035
Project Title	TechPart Vision: Personal Computer Parts Image Classification Using EfficientNet Transfer Learning
Maximum Marks	2 Marks

Data Quality Report Template

The Data Quality Report Template 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 Source	Data Quality Issue	Severity	Resolution Plan
https://www.kaggle.com/datasets/asaniczka/pc-parts-images-dataset-classification	Inconsistent Image Sizes: Images may have different dimensions.	Moderate	Solution: Resize all images to a uniform size before feeding them into the model. This can be done using image preprocessing libraries like OpenCV .

https://www.kaggle.com/datasets/asaniczka/pc-parts-images-dataset-classification	<p>The dataset has so many missing values which leads to incorrect prediction and bias in the model.</p>	<p>High</p>	<p>Solution: The dataset has so many missing values which leads to incorrect prediction and bias in the model. So we are using imputation techniques to fill in the gaps. such as mean/median/mode replacement. K-Nearest Neighbors (KNN) imputation</p>
---	--	-------------	--