**Data Collection and Preprocessing Phase**

|  |  |
| --- | --- |
| Date | 19 June 2025 |
| Team ID | SWTID1749821186 |
| Project Title | Enhancing Product Reliability: Leveraging Transfer Learning for Fault Detection |
| Maximum Marks | 2 Marks |

**Data Quality Report**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Source** | **Data Quality Issue** | **Severity** | **Resolution Plan** |
| Kaggle - Real-life Industrial Dataset of Casting Product | Uneven class distribution (more good than defective images) | Moderate | Apply data augmentation to the minority class (defective) to balance the dataset. |
|  | Grayscale images need to be converted to RGB for VGG16 | Low | Use cv2.cvtColor(image, cv2.COLOR\_GRAY2RGB) during preprocessing. |
|  | Varying image sizes | Moderate | Resize all images to 224×224 using cv2.resize() or Image.resize(). |
|  | Presence of noise or low-contrast images | Moderate | Apply Gaussian blur or histogram equalization to enhance image clarity. |
|  | Noisy backgrounds in some images | Low | Use cropping or masking to focus on the casting region. |
|  | File naming inconsistencies or duplicates | Low | Standardize filenames and remove duplicates using hashing or checksum methods. |