Data Collection and Preprocessing Phase

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

Data Collection Plan & Raw Data Sources Identification Report:

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavor.

Data Collection Plan:

Section	Description
Project Overview	The machine learning project aims to automate fault detection in manufacturing syst image-based data. Utilizing transfer learning with pre-trained CNN models like VG ResNet50, and InceptionV3, the objective is to classify products as faulty or non-fau on visual features, thus improving inspection accuracy, reducing human error, and e real-time quality control in high-speed production environments.
Data Collection Plan	• Search for image datasets that include labeled examples of faulty and non-faulty manufactured products. • Prioritize datasets with sufficient variability in lighting, angle, and defect types to enhance model generalization. • Collect additional custom image data if needed to represent domain-specific faults. • Ensure datasets are balanced and include metadata where applicable (e.g., fault type, location).
Raw Data Sources Identified	The raw data sources for this project include publicly available datasets from platforms like Kaggle and Mendeley Data, which host manufacturing defect images across domains such as casting, PCB, and textile production. Sample datasets include the Casting Product Image Dataset and Severstal Steel Defect Dataset. These sources

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	This dataset contains grayscale images of casted components, labeled as "defective" or "okay." Useful for binary classification tasks in manufacturing fault detection.	https://www.kaggle.com/ravirajsinh45/real-life-industrial-dataset-of-casting-product	JPG	~360 MB	Public
Severstal: Steel Defect Detection (Kaggle)	A high-resolution steel surface image dataset with defect annotations in CSV format. Suitable for segmentation and classification of defects.	https://www.kaggle.com/c/severstal-steel-defect-detection/data	JPG, CSV	~2.5 GB	Public