Data Collection and Preprocessing Phase

Date	11-03-2025
Team ID	739955
Project Title	AI-POWERED VEHICLE DAMAGE ASSESSMENT FOR COST ESTIMATION AND INSURANCE CLAIMS.
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template

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 Template

Section	Description
Project Overview	This project focuses on developing an AI-powered solution for automated vehicle damage assessment and cost estimation, primarily aimed at improving the insurance claim process. Currently, damage evaluation is manual, time-consuming, and often inconsistent. By leveraging computer vision and deep learning, the proposed system can analyze images of damaged vehicles, classify the type and severity of damage, and estimate repair costs in real time.
Data Collection Plan	Collect high-quality vehicle damage images from various angles, with labels for damage type, severity, and location.

	Use annotation tools like LabelImg for labeling and apply data augmentation to improve model performance. Gather metadata such as vehicle details and repair costs, and split the dataset into training, validation, and testing sets.. Temperature, Humidity and Raw H2 related details for machine learning analysis.
Raw Data Sources Identified	Use publicly available datasets like Stanford Cars, CompCars, and Kaggle's car damage datasets. Supplement with real-world images from insurance companies or repair shops, if accessible.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises of various damaged car images.	https://drive.google.com/drive/folders/1lpOrcULlx5mHuSbbLRCrvmCwXKkbDVKE?usp=sharing	CSV	4 MB	Public