Data Collection and Preprocessing Phase

Date	June 2024
Team ID	Team-739856
Project Title	Auto Insurance Fraud Detection Using Machine Learning
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	Aims to provide accurate fraud detection by considering factors like claim history, policy details, and customer information, offering real-time alerts, integrating with insurance platforms.
Data Collection Plan	Data will be collected from insurance claim databases, policy details, customer information, historical fraud records, and external verification sources.
Raw Data Sources Identified	Gathering data from claim histories, policy documents, customer profiles, fraud reports, and integration with insurance platforms.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Access Permissions
Dataset 1	In this project we have used .csv data. This data is downloaded from kaggle.com.	https://www.kaggle.com/datasets/buntyshah/auto-insurance-claims-data	CSV	Public
Dataset 2	Access Google Drive with a Google account (for personal use) or Google Workspace account	https://colab.research.google.com/drive/18JQ885csjB2LetKjdkKFQBXRjOJiiWMx?usp=sharing	CSV	Private (with access)