Data Collection and Preprocessing Phase

Date	June 2024
Team ID	Team-739859
Project Title	Auto Insurance Fraud Detection Using Machine 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: Insurance Claim Database

Data Source	Data Quality Issue	Severity	Resolution Plan
Insurance Claim Table	Missing values in the 'Claim Date' field	High	Implement a mandatory check for 'Claim Date' field before data entry
	Inconsistent formatting in the 'Claim ID' field (e.g., mix of numbers and letters)	Medium	Apply a standard format and use a script to reformat existing entries
	Outdated 'Policyholder Address' information	Low	Regularly update addresses through automated address validation APIs

Data Source: Policy Details Dataset

Data Source	Data Quality Issue	Severity	Resolution Plan
Policy Details Table	Incorrect 'Premium Amount' values due to manual entry errors	High	Implement validation rules and automate data entry where possible
	Missing 'Agent ID' identifiers in some records	Medium	Perform a data audit to identify missing values and rectify them
	Redundant columns with no relevant data	Low	Conduct a data review to identify and remove redundant columns
	'Policy Type' field has ambiguous names leading to misclassification	Medium	Standardize policy type names and ensure consistent use across all datasets