Data Collection and Preprocessing Phase

Date	3 october 2024
Team ID	LTVIP2024TMID24947
Project Title	SmartLender - Applicant Credibility Prediction for Loan Approval
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 **SmartLender** aims to solve this problem by providing an automated system that evaluates applicants based on multiple factors like their number of dependents, education level, employment status, loan amount, loan term, CIBIL score, and assets. This model will ensure fast, fair, and accurate loan approval decisions, helping both the lender and applicant experience a smoother loan approval process.
Data Collection Plan	• Search for datasets related to loan approvals, financial information, and applicant details. • Prioritize datasets with diverse demographic information.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle the popular platforms for data science competitionsand repositories. The provided sample data represents a subset of the collected information, encompassing variables such as their number of dependents, education level, employment status, loan amount, loan term, CIBIL score, and assets.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details (gender, marital status), financial metrics (income, loan amount), and loan approval outcomes.	https://www.kaggle.com/datasets/architsharma01/loan-approval-prediction-dataset/data	CSV	384 KB	Public