Data Collection and Preprocessing Phase

Date	11 July 2024
Team ID	740023
Project Title	Smart Lender - Applicant Credibility Prediction For Loan Approval
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	The loan approval prediction project aims to build a machine learning model to assess the likelihood of loan applications being approved based on historical data. This system enhances decision-making efficiency for financial institutions by providing accurate and data-driven predictions. The project seeks to reduce the risk of defaults and streamline the loan approval process.
Data Collection Plan	It involves gathering historical loan application data, including applicant demographics, financial status, loan details, and approval status. This data can be sourced from internal banking systems and publicly available financial datasets. Ensuring data quality and completeness is crucial for building a robust predictive model.

Raw Data Sources Identified	Raw data will be sourced from internal banking systems, including historical loan application records with detailed applicant and loan information. Public datasets, such as those from Kaggle and the UCI Machine Learning Repository, will supplement this data. Additionally, credit histories and scores from credit bureaus like Experian, Equifax, and TransUnion will be integrated to enhance the dataset.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Excel Dataset	The dataset comprises applicant details(gender, marital status), financial metrics(income,loa namount),and loan approval outcomes.	"C:\Users\shiva\ OneDrive\Docum ents\miniproject\f lask\loan_predicti on.csv"	CSV	40 GB	Public