Data Collection and Preprocessing Phase

Date	15 June 2024
Team ID	739732
Project Title	Loan Sanction Amount Prediction Data With Ml
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 Sanction Amount Prediction model project aims to develop a machine learning tool to predict the sanction amount for loan applications based on customer data such as age, gender, income, credit score, and property details. The project involves collecting and preprocessing data, developing and evaluating a predictive model using algorithms like linear regression or decision trees, and deploying the model in a user-friendly application for use by financial institutions. The goal is to provide a reliable prediction mechanism that improves loan approval processes and supports financial decision-making.

Data Collection Plan	• **Methods:** Extract data from existing loan management systems, conduct customer surveys, or gather data from financial applications. • **Techniques:** API integration for loan management systems, online survey tools, data extraction scripts.
Raw Data Sources Identified	Gathered a dataset from Kaggle containing loan application information such as age, gender, income, credit score, and property details for loan sanction amount prediction. The dataset includes features relevant for building and training the prediction model, enabling accurate loan amount assessments and analysis.

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises loan application details and target column	https://www.kaggle.com/datasets/altruistdelhite04/loan-prediction-problem-dataset	CSV	1348 kb	Public