Project Definition:

Loan Default Prediction Analysis

January 2023

1. Introduction:

The aim of this project is to develop a loan default prediction model by analyzing historical data and identifying key factors that contribute to loan defaults. The model will enable lenders to assess the risk associated with potential borrowers and make more informed lending decisions.

2. Objective:

The primary objective of this project is to accurately predict loan defaults by analyzing various factors that influence borrower behavior. By understanding the types of defaults and reasons behind loan defaults, this project will provide valuable insights into risk management and help lenders minimize potential losses.

3. Scope:

The project will utilize historical loan data, including borrower information, loan characteristics, and repayment behavior, to build a predictive model. The analysis will focus on identifying patterns, trends, and risk factors associated with loan defaults. The project will also explore the impact of different loan types and borrower profiles on default rates.

4. Methodology:

The project will follow the following steps:

a. **Data Collection:** You can download the dataset from this.

b. **Data Preprocessing:** Clean and preprocess the data by handling missing values, removing duplicates, and transforming variables if necessary.

c. **Feature Engineering:** Identify relevant features and create new variables that capture borrower characteristics, loan attributes, and other factors that may influence loan defaults.

d. **Exploratory Data Analysis:** Conduct a comprehensive analysis of the data to identify patterns, trends, and correlations between variables. Visualizations and statistical techniques will be used to gain insights into the data.

e. **Model Development:** Utilize machine learning algorithms such as logistic regression, decision trees, or random forests to build a loan default prediction model. The model will be trained and evaluated using appropriate techniques such as cross-validation and performance metrics.

f. **Model Evaluation and Optimization:** Assess the performance of the prediction model by different metrics (specially Lift curve and H-Measure) and fine-tune it by adjusting parameters or trying alternative algorithms if necessary.

g. **Results and Recommendations:** Present the findings of the analysis, including key risk factors, model performance, and recommendations for lenders to mitigate loan default risks.

5. **Deliverables:**

The project will deliver the following:
- A comprehensive analysis of loan default trends, types, and reasons.
- A loan default prediction model with documented methodology and performance evaluation.
- Visualizations, reports, codes, and presentations summarizing the findings and recommendations for lenders.

6. **Timeline:**

The project is estimated to be completed within 3 days.

7. **Resources:**

The project will require access to historical loan data, data analysis and visualization tools, programming languages (Python), machine learning libraries, and expertise in data analysis and predictive modeling.

By conducting this loan default prediction analysis, lenders will be better equipped to assess the creditworthiness of potential borrowers and make informed decisions, leading to improved risk management and reduced loan default rates.