

# Loan Repayment Prediction

## 1. Introduction

This project applies machine learning techniques to predict whether a borrower will successfully repay a loan, addressing the critical challenge of credit risk assessment in the banking and financial services industry. Accurate loan default prediction helps lenders minimize defaults and optimize lending decisions.

## 2. Methodology

- **Data Source:** The dataset, sourced from LendingClub via Kaggle, includes detailed loan and borrower information, such as loan amount, interest rate, credit score, income, and debt-to-income ratio.
- **Task:** The problem is framed as a binary classification task: predicting whether a loan will be fully paid or defaulted.
- **Machine Learning Approaches: Random Forest/XGBoost/ Artificial Neural Networks.** These models are trained and compared to assess their effectiveness in loan default prediction.
- **Model Evaluation:** The models are evaluated using metrics such as AUC-ROC, accuracy, precision, and recall.

## 3. Key Components/Output

- Conducted exploratory data analysis (EDA) to understand data distributions and identify key patterns.
- Developed and trained predictive models to assess loan repayment risk.
- Compared the performance of multiple machine learning algorithms to identify the most effective model for loan default prediction.

## 4. Limitations and Future Improvements

- **Data Constraints:** The dataset is clean and well-structured. However, real-world scenarios may introduce more complexity, with models needing further refinement to adapt to such conditions.
- **Data Expansion:** Incorporating additional data, such as macroeconomic factors, could enhance the model's robustness and predictive power.
- **Model Enhancements:** Future work could focus on further tuning the models or exploring ensemble methods to achieve even better performance.

## Data Description

	Variables	Description
0	loan_amnt	The amount of loan owed by the borrower.
1	term	The number of payments on the loan. Values are in months (e.g. 36 or 60).
2	int_rate	Interest rate on the loan.
3	installment	The monthly payment owed by the borrower.
4	grade	Loan grade assigned by Lending Club

5	sub-grade	Loan subgrade assigned by Lending Club
6	emp_title	Borrower's job title. Provided by the borrower in the loan application.
7	emp_length	Borrower's employment lengths in years. Possible values are between 0 and 10 where 0 means less than one year and 10 means ten or more years.
8	home_ownership	Borrower's home ownership status. Provided by the borrower in the loan application or obtained from the credit report. Possible values are RENT, OWN, MORTGAGE, OTHER.
9	annual_inc	Self-reported annual income. Provided by the borrower in the loan application.
10	verification_status	Indicates if borrower's income information is verified by Lending Club.
11	issue_d	The month in which the loan was funded
12	loan_status	Current status of the loan.
13	purpose	What the loan is used for.
14	title	Loan title provided by the borrower.
15	zip_code	The first 3 numbers of the zip code provided by the borrower in the loan application.
16	addr_state	The state of address provided by the borrower in the loan application.
17	dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested Lending Club loan, divided by the borrower's self-reported monthly income.
18	earliest_cr_line	The month the borrower's earliest reported credit line was opened.
19	open_acc	The number of open credit lines in the borrowers' credit file.
20	pub_rec	Number of derogatory public records.
21	revol_bal	Total credit revolving balance
22	revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
23	total_acc	The total number of credit lines currently in the borrowers' credit file.
24	initial_list_status	The initial listing status of the loan.
25	application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers.
26	mort_acc	Number of mortgage accounts.
27	pub_rec_bankruptcies	Number of public record bankruptcies.