

# Smart Lender - Applicant Credibility Prediction for Loan Approval

## Literature Survey

<b>Industry Mentor(s):</b>	Sowjanya Sandeep Doodigani	<b>Faculty Mentor:</b>	Lakshmi Priya S
<b>Team Members:</b>	Anirudh T E Pranav Raj SB Tushar Shah Yashwanth M	<b>Roll Numbers:</b>	SSNCE195001016 SSNCE195001078 SSNCE195001117 SSNCE195001130

### Problem Statement:

One of the most important factors which affect our country's economy and financial condition is the credit system governed by the banks. The process of bank credit risk evaluation is recognized at banks across the globe. As we know credit risk evaluation is very crucial, there are a variety of techniques used for risk level calculation. In addition, credit risk is one of the main functions of the banking community.

The prediction of credit defaulters is one of the difficult tasks for any bank. But by forecasting the loan defaulters, the banks definitely may reduce their loss by reducing their non-profit assets, so that recovery of approved loans can take place without any loss and it can play as the contributing parameter of the bank statement.

### Introduction:

A loan default occurs when a borrower takes money from a bank and does not repay the loan. People often default on loans due to various reasons. Borrowers who default on loans not only damage their credit but also risk being sued and having their wages garnished.

Many individuals utilize debt to pay for things they wouldn't be able to buy otherwise, such as a home or a vehicle. While loans may be beneficial financial instruments when utilized correctly, they can also be formidable foes.

Lending is a vital tool that propels all enterprises and individuals worldwide to greater financial success. The need for capital has risen dramatically as the world's economies become increasingly integrated and interdependent.

In the last decade, the number of retail borrowers, SMEs, and commercial borrowers has increased dramatically. Though most financial institutions have seen an increase in revenue and profit due to this rising trend, not everything is green. In recent years, there has been an increase in loan defaults, which has already begun to affect the bottom lines of several financial institutions.

We aim to make use of machine learning to make better financial predictions and understand the banking sector's lending applications and the creditworthiness of individuals and organizations. Machine Learning techniques are very crucial and useful in the prediction of these types of data.

We will be using classification algorithms such as Decision tree, Random forest, KNN, and xgboost. We will train and test the data with these algorithms. From this the best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.

**Key Points:**

1. This project aims to work towards predicting the loan defaulters in the banking ecosystem which play a vital role in the financial functioning of our country, primarily- Credit Risk Analysis.
2. We would be using machine learning with Python to make predictions and draw meaningful conclusions from the dataset presented to us.
3. Algorithms like Decision tree, Random forest, KNN, and xgboost will be used. We will train and test the data with these algorithms. From this the best model is selected and saved in pkl format. We will be doing flask integration and IBM deployment.
4. We would be performing data preprocessing, model building, working on training the model, and deploying it during the course of the project.
5. Tasks would be divided between the team members efficiently to ensure quality and efficiency.

## **Literature review:**

### **1. Classification of Applicants**

The loan credibility prediction system analyzes the applicant's information and classifies him into either applicant with safe credit or applicant with risky credit. Applicants with safe credit have a higher possibility to repay the loan promptly while applicants with risky credit have a higher possibility of failing to repay the loan amount. The financial institution's profitability definitely depends on the accuracy of the model.

Applying Random Forest classification algorithm, it is very effective to build a successful predictive model that helps the bankers to take the proper decision regarding safe and risky loan applicants. Random Forest classifier produces high accuracy in both weka and R under the credit data set.

### **2. Improving the model Accuracy**

To improve the accuracy of Loan Credibility Prediction System (LCPS) several Classification techniques are commonly used. The major techniques are ensemble, optimization and feature selection. Ensemble method improves the accuracy to a certain extent in multiple data sets. The main idea behind ensemble methodology is to combine a set of multiple learning algorithms in order to obtain better predictive performance. Optimization has also produced better classification accuracy by changing the values of various selected parameters..

### **3. Determining the criteria for applicants**

Feature selection is one of the most important data preprocessing techniques in data mining and it is closely related to dimensionality reduction. The main idea behind feature selection is ranking the individual features based on some criteria and then searching for an optimal feature subset based on evaluation criteria to test the optimality. The accuracy level considerably increased after feature selection methods were applied to the classifier.

Subset search evaluates the quality of subsets of features. The best performance of the selected features can be achieved when both the feature selection and classification stages are optimized together using the same criterion function

### **4. Predicting high-risk applications**

Using a wrapperfisher feature selection algorithm to select the most significant features which will improve the accuracy of Random Forest (RF) classification. After studying various past data from the bank it is possible to identify several attributes that can influence the customer behavior. The most influencing attribute can be considered while a new customer approaches the bank for

a loan and thus we can identify the potential of the customer. Here by enabling the bank officers to identify fraud applicants.

Wrapper methods are based on greedy search algorithms as they evaluate all possible combinations of the features and select the combination that produces the best result for a specific machine learning algorithm.

**References:**

1. Soni P M, Varghese Paul, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1S4, June 2019.
2. VivekBhambri “Application of Data Mining in Banking Sector”, IJCSt Vol. 2, ISSue 2, June 2011, ISSN : 2229- 4333(Print) | ISSN : 0976-8491(Online)
3. Challita N., Khalil M., Beausero P. 2016 IEEE IntMultidiscipConfEng Technol. 2016. New feature selection method based on neural network and machine learning.