Loan Defaulter System Using Machine Learning

Shraddha Laxman Binawade, Pritee Kailas Chaugule, Pratiksha Suresh Doke, Prof. Suvarna A.Bahir

binawadeshraddha1234@gmail.com
priteechaughule@gmail.com
pratikshadoke5@gmail.com
suvarnabahir.sae@sinhgad.edu

Sinhgad Academy of Engineering Kondhwa

ABSTRACT

In today's world, taking loans from financial institutions has become a very common phenomenon. Everyday a large number of people make application for loans, for a variety of purposes. But all these applicants are not reliable and everyone cannot be approved. Every year, we read about a number of cases where people do not repay bulk of the loan amount to the banks due to which they suffers huge losses. The risk associated with making a decision on loan approval is immense. So the idea of this project is to gather loan data from multiple data sources and use various machine learning algorithms on this data to extract important information. This model can be used by the organizations in making the right decision to approve or reject the loan request of the customers. In this paper, we examine a real bank credit data and conduct several machine learning algorithms on the data for that determine credit worthiness of customers in order to formulate bank risk automated system.

Keywords- Image Process, Mahine Learning, bank credit, classification, confusion matrix, predictive analysis.

I. INTRODUCTION

Bank plays a vital role in market economy. The success or failure of organization largely depends on the industry's ability to evaluate credit risk. Before giving the credit loan to borrowers, bank decides whether the borrower is bad (defaulter) or good (non defaulter).The prediction of borrower status i.e. in future borrower will be defaulter or non-defaulter is a challenging task for any organization or bank. Basically the loan defaulter prediction is a binary classification problem Loan amount; costumers history governs his credit ability for receiving loan. The problem is to classify borrower as defaulter or non-defaulter. However developing such a model is a very challenging task due to increasing in demands for loans. Prototypes of the model which can be used by the organizations for making the correct or right decision for approve or reject the request for loan of the customers. This work includes the construction of an ensemble model by combining different machine learning models. Banks struggle a lot to get an upper hand over each other to enhance overall business due to tight competition. Credit Risk assessment is a crucial issue faced by Banks nowadays which helps them to evaluate if a loan applicant can be a defaulter at a later stage so that they can go ahead and grant the loan or not. This helps the banks to minimize the possible losses and can increase the volume of credits.

Problem Statement:
Banks, Housing Finance Companies and some NBFC deal in various types of loans like housing loan, personal loan, business loan etc in all over the part of countries.
These companies have existence in Rural, Semi Urban and Urban areas. After applying loan by customer these companies validates the eligibility of customers to get the loan or not.

This paper provides a solution to automate this process by employing machine learning algorithm. So the customer will fill an online loan application form. This form consist details like Sex, Marital Status, Qualification, Details of Dependents, Annual Income, Amount of Loan, Credit History of Applicant and others.

II. LITERATURE SURVEY

Vishal Singh, Ayushman Yadav, In this case, the right prediction would be very beneficial using some classes-function algorithm. An example the logistic regression, random forest classifier, support vector machine classifier, etc. A Bank's profit and loss depend on the amount of the loans that is whether the Client or customer is paying back the loan. Recovery of loans is the most important for the banking sector. The improvement process plays an important role in the banking sector. The historical data of candidates was used to build a machine learning model using different classification algorithms. The main objective of this paper is to predict whether a new applicant granted the loan or not using machine learning models trained on the historical data set.

Bhoomi Patel1, Harshal Patil2, Estimation or assessment of default on a debt is a crucial process that should be carried out by banks to help them to assess if a loan applicant can be a defaulter at a later phase so that they process the application and decide whether to approve the loan or not. The conclusion derived from such assessments helps banks and other financial institutions to lessen their losses and eventually increase the number of credits.

P. Maheswari, CH. V. Narayana, This paper uses statistical measures to preprocess the data and build an effective model that will predicts the loan defaulter accurately.

Mohammad Ahmad Sheikh, Amit Kumar Goel, In our banking system, banks have many roducts to sell but main source of income of any banks is on its credit line. So they can earn from interest of those loans which they credits. A bank's profit or a loss depends to a large extent on loans i.e. whether the customers are paying back the loan or defaulting. By predicting the loan defaulters, the bank can reduce its Non- Performing Assets. This makes the study of this phenomenon very important. Previous research in this era has shown that there are so many methods to study the problem of controlling loan default. But as the right predictions are very important for the maximization of profits, it is essential to study the nature of the different methods and their comparison. A very important approach in predictive analytics is used to study the problem of predicting loan defaulters: The Logistic regression model. The data is collected from the Kaggle for studying and prediction.

G. Arutjothi, Dr. C. Senthamarai, Banking Industry always needs a more accurate predictive modeling system for many issues. Predicting credit defaulters is a difficult task for the banking industry. The loan status is one of the quality indicators of the loan. It doesn't show everything immediately, but it is a first step of the loan lending process.

G. Arutjothi, Dr. C. Senthamarai, In this paper, we propose a machine learning classifier based analysis model for credit data. We use the combination of Min-Max normalization and KNearest Neighbor (K-NN) classifier. The objective is implemented using the software package R tool. This proposed model provides the important information with the highest accuracy. It is used to predict the loan status in commercial banks using machine learning classifier.

Ramachandra H V, Balaraju G, The major aim of this project is to predict which of the customers will have their loan paid or not using prominent algorithms like Decision Tree, Logistic Regression and Random Forest. Logistic Regression Confusion matrix analysis is relatively in accordance to Decision Tree and Random Forest algorithm helping us attain an accuracy of 86% with minimum error.

P. Maheswari, CH. V. Narayana, This paper uses statistical measures to preprocess the data and build an effective model that will predicts the loan defaulter accurately.
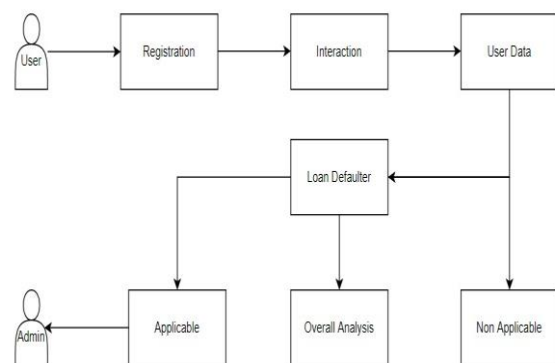
III. ROPOSED SYSTEM

Fig 1. System architecture

1. The whole architecture is made by PyQT library used in python language. PyQT library gives all the necessary stuff related to GUI design. PyQT provides us display screen, buttons and so on. So, In this way PyQT helps us in design GUI.

2. After designing of GUI, another task is to authenticate valid user for operating application. To deal with this task, we are using MySQL database to store data of username and password and through this, user can authenticate easily.

3. Another task is to find out loan defaulters as applicable, non-applicable and overall analysis by using csv files and machine learning algorithms.

4. In this system we detect the loan defaulters with the help of machine learning technique.

Here provide the module for detecting loan defaulters using user data.

IV. IMPLEMENTED PROJECT
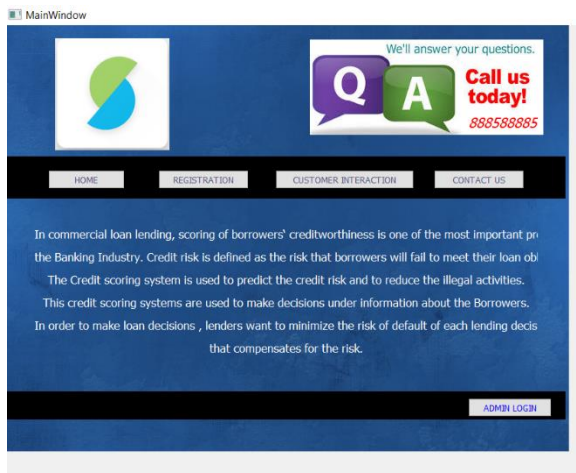
V. CONCLUSIONS

The model concludes that a bank should not only target the rich customers for granting loan but it should assess the other attributes of a customer as well which play a very important part in credit granting decisions and predicting the loan defaulters.

REFERENCES

[1] https://ieeexplore.ieee.org/document/9336801

[2] https://harvestify.herokuapp.com/

[3] https://ieeexplore.ieee.org/document/9155614

[4] Raj, J. S., & Ananthi, J. V., "Recurrent neural networks and nonlinear predict ion in support vector machine" Journal of Soft Computing Paradigm (JSCP), 1(01), 33-40, 2019.

[5] Toby Segaran, "Programming Collective Intelligence: Building Smart Web 2.0 Applications." O'Reilly Media.

[6] Drew Conway and John Myles Whit e," Machine Learning for Hackers: Case Studies and Algorithms to Get you Started," O'Reilly Media.