



Machine Learning Classification and Deep Learning based Credit Prediction and Risk analysis on Loans Data.

Ashutosh Kushwaha¹,

Department of Computer Engineering,
Wagholi, Pune

Mohd. Zaib Nawab¹

Department of Computer Engineering
Wagholi, Pune

Aditya Kardile¹,

Department of Computer Engineering,
Wagholi, Pune

Dr. Vinod Wadne

Department of Computer Engineering
Wagholi, Pune

Abstract: Estimation or assessment of default a debt could be a crucial method that ought to be allotted by banks to assist them to assess if a loan somebody may be a defaulter at a later section in order that they method the applying and judge whether or not to approve the loan or not. Banks have already been creating an attempt to try to this through utilization of FICO score and credit Reports. In this project, using Machine learning techniques, we focus on dealing with imbalance data problem to enhance the performance of loan default and loan approves prediction. Artificial intelligence will facilitate modernize the standards for loan applications and permit banks to try to higher loan management. Equifax, one of the three major credit bureaus, determined through a recent research study that financial institutions sometimes deny good borrowers based on this FICO criteria artificial intelligence can help modernize the criteria for loan applications and allow banks to do better loan management.

Keywords: *Machine Learning, loan, credit, Imbalance data, prediction, etc.*

INTRODUCTION

Take India as an example, the scale of micro loan industry registered 31 % jumps in its loan portfolio to Rs. 2.36 lakh crore for 2019-2020 and expect a moderate growth of about 15 percent in the current financial year. In the recent past, occurrences of financial fraud rapidly reported in India. As compared to traditional ways the frequency, intricacy, diversity, and price of banking frauds have increased exponentially. Consequently, such issues are a grave cause of concern for regulators. The robustness and stability of a country's fiscal structure help ascertain whether the country's economy is worth investing in. It gives an idea about the wellness, security and living standards of its citizens. Distribution of the loans is that the core business a part of virtually each banks. the most portion the bank's assets is directly came from the profit earned from the loans distributed by the banks. The prime objective in banking setting is to take a position their assets in safe hands wherever it's. Nowadays several banks/financial corporations approves loan when a regress method of verification and validation however still there's no surety whether or not the chosen soul is that the worth right soul out of all candidates. Through this method we are able to predict whether that exact soul is safe or not and also the whole method of validation of options is automatic by machine learning technique. Machine Learning is an emerging technology now days. The application of Machine Learning solves the many Real world Problems. Machine Learning is an Artificial

Intelligence technique which is useful in prediction system. Machine learning can create model on training dataset while for making the prediction of model training dataset can be used. This paper applies the machine learning techniques and algorithm to solving loan approval problem of banking sector.

A. Objectives:

- ❑ Optimum results were obtained using different algorithm
- ❑ It runs efficiently on large databases
- ❑ It is very robust and can be simply implemented on classification datasets
- ❑ It works well with data that has highly dependent attributes
- ❑ require little effort for their preparation
- ❑ It runs efficiently on large databases

LITERATURE SURVEY

Bhoomi Patel, Harshal Patil, Jovita Hembram, Shree Jaswal [1] proposed a model which is focused mainly on crime factors that occur each day, rather than focusing on causes of crime occurrence. Using Data mining technique and machine learning Tiannan Deng[2], proposed a study of the prediction of micro-loan default based on Logit model using neural network which is more complex to predict credit risk and use comprehensive discriminant method to analyzing credit and loan default prediction. Dawei Cheng, Zhibin Niu, Yi Tu and Liqing Zhang [3] proposed system on Prediction Defaults for Networked-guarantee Loans using specifically tailored algorithm for prediction, highly imbalanced dataset, removing too many samples leads to loss of information and poor sample representation Jun-Ya Zeng, Jian-Bang Lin, Tian Wang [4] proposed A New Competing Risks Model for Predicting Prepayment and Default Using Data Mining, in this system large number of data cannot easy in the era of Big Data. To predict a model using Data mining techniques

BACKGROUND

The most important background of machine learning algorithms their technique and mathematical formulation are outlined in this section. Analysing the Banking and credit data used these algorithms

1. Machine Learning

Machine learning algorithm can be group into two main categories, they include

1. **Supervised Learning:** supervised learning algorithm main feature is target variable and outcome variable to predict. Supervised

learning technique is achieved using regression and classification problem.

2. **Unsupervised learning:** in unsupervised learning algorithm no target or outcome variable to predict. It is used for clustering entities into an different groups.

2. Classification Algorithms:

Classification algorithms work by predicting the simplest cluster to that a knowledge purpose belongs to by learning from labelled observations; it uses a group of input options for the educational method. Classification algorithms square measure sensible for grouping knowledge that square measure ne'er seen before into their numerous groupings and square measure thus extensively employed in machine learning tasks.

3. Evaluation Matrix:

1. Accuracy:

it is measured how many true positive and true negative cases is correct. Mathematically it is defined as

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN}$$

2. Sensitivity or Recall:

1. Recall: tells us how many of the actual positive cases we were able to predict correctly with our model. Mathematically it is defined as

$$\text{Recall} = \frac{TP}{TP+FN}$$

2. Specificity: tell us how many times classifier gets true negative correct value, mathematically is defined as

$$\text{Specificity} = \frac{TN}{TN+FP}$$

3. Precision:

Precision tells us how many of the correctly predicted cases actually turned out to be positive. Mathematically it is defined as,

$$\text{Precision} = \frac{TP}{TP+FP}$$

PROPOSED SYSTEM

Machine learning application and Web Application can be developed; in the proposed system web application can check the loan default status and approval of loan using customer data. Banking and financial companies can check the loan approval of customer. The proposed model focused on predicting and credibility of customer for loan repayment and analysing their behaviour. The input to the model is that the client behaviour collected. On the output from the classifier, call on whether or not to approve or reject the client

request will be created. Victimization totally different information analytics tools loan prediction and there severity will be forecasted. During this method it's needed to coach the information the info the information victimisation totally different algorithms and so compare user information with trained data to predict the character of loan. To extract patterns from a standard loan approved dataset, and so build a model supported these extracted patterns. The coaching information set is currently equipped to machine learning model; on the premise of this information set the model is trained. Each new mortal detail stuffed at the time of form acts as a take a look at information set. when the operation of testing, model predict whether or not the new mortal could be a match case for approval of the loan or not primarily based upon the logical thinking it conclude on the premise of the coaching information sets. To extract necessary info and predict if a client would be able to repay his loan or not

BLOCK DIAGRAM

The proposed module can be divided in to different sections, machine learning, Flask, HTML, CSS, Anaconda-Jupyter notebook. Architecture used in proposed system are given below

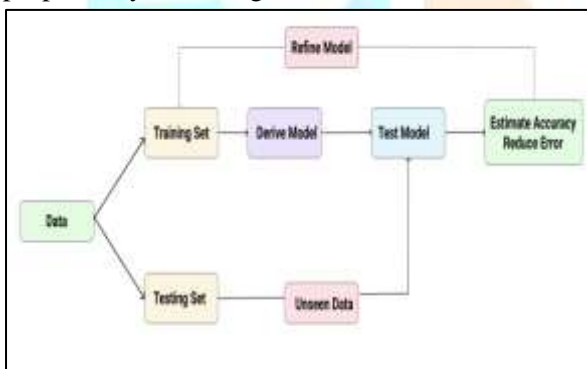


Figure: Architecture of System Design

CONCLUSION

This project aims to develop a robust module for predicting your chances of loan approval and Risk prediction on default loan customer. The features used for prediction are considered to be available at the time of loan origination and thus do not leak any information from the future. The prediction model has been built separately for Individual Applicants and Joint Applicants with a maximum accuracy. Few highly correlated features are used for analyzing and prediction of risk factor and calculating approval of loan. This paper applies Machine Learning algorithm and technique in prediction of loan approval and solving banking default loan problem. Multiple models are used to predict the loan approval status of customer for bank loan using machine learning. Multiple model

can investigate the power of machine learning algorithm for loan approval and loan default prediction.

BIBLIOGRAPHY

- [1] Kacheria, A., Shivakumar, N., Sawkar, S. and Gupta, A. (2016). Loan Sanctioning Prediction System. [online] Ijsce.org.
- [2] <http://www.ijscce.org/wpcontent/uploads/papers/v6i4/D2904096416.pdf>
- [3] A. Gahlaut, Tushar, and P. K. Singh, "Prediction analysis of risky credit using Data mining classification models," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), 2017
- [4] Hamid, Aboobyda & Ahmed, Tarig. (2016). "Developing Prediction Model of Loan Risk in Banks Using Data Mining". Machine Learning and Applications: An International Journal. 3. 1-9. 10.5121/mlaj.2016.3101.
- [5] Mrunal Surve, Pooja Thitme, Priya Shinde, Swati Sonawane, and Sandip Pandit. "Data mining techniques to analyze risk giving loan(bank)" Internation Journal Of Advance Research And Innovative Ideas In Education Volume 2 Issue 1 2016 Page 485-490
- [6] P. Ravikumar and V. Ravi, "Bankruptcy Prediction in Banks by an Ensemble Classifier," 2006 IEEE International Conference on Industrial Technology, Mumbai, 2006, pp. 2032-2036.
- [7] S. Sathyadevan, D. M. S and S. G. S., "Crime analysis and prediction using data mining," 2014 First International Conference on Networks & Soft Computing (ICNSC2014), Guntur, 2014, pp. 406-412
- [8] Lekha, K. and Prakasam, D. (2018). <https://www.researchgate.net/publication/326147494>
- [9] K. C. Lekha and S. Prakasam, "Data mining techniques in detecting and predicting cyber crimes in banking sector," 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, 2017, pp. 1639-1643.
- [10] Yu Jin and Yudan Zhu, "A data-driven approach to predict default risk of loan for online Peer-to-Peer(P2P) lending," School of Information, Zhejiang University of Finance and Economics, 310018 Hangzhou, China. [11] Jannes Klaas, "Loan Default Model Trap,"