



Loan Approval With Machine Learning : A Survey

Soumik Das

PG Student

Department of Master of Computer Applications
School of CS & IT
JAIN (Deemed-to-be-University)

Abstract— In our banking system, the main source of income for the banks comes from the interest that they earn from the loans they give out to people. A bank's profit or loss mostly depends on the loans that are given i.e. whether the person taking the loan will be able to repay it, and repay it on time or not. So the need for a proper measure before handing out a loan of any amount becomes really necessary. Thus the main goal is predicting the chances of getting back the amount given out and hence reducing the chances of having any non-performing assets. Therefore in order to get a proper prediction Machine Learning can be used, which helps in properly identifying the necessary features required for a loan to be granted and hence assure that the mentioned loan amount will be repaid.

Keywords— *Machine Learning, Loan, Testing, Training Prediction, Accuracy*

I. INTRODUCTION

Loan Prediction has been one of the most challenging tasks to tackle in our recent times. As interests and repayment of loans are the main source of income and the only way any bank can certainly make some profit for themselves we can say that this study has become one of the utmost importance.

Thus this study deals with past data for banks which deal with different peoples annual income, marital status, education, location where they live and their credit history and many more to try and accurately make a prediction where they can give some insight whether to grant the applicant the loan amount, and hence loan amount also becomes a huge factor while dealing with such factor.

In these situations where the number of applicants grows due to the over growing population in our country, the need for a proper predictive model which are based on relevant algorithms are needed to cater to the needs of such requests

Pawan Kumar

Assistant Professor,

School of CS & IT
JAIN (Deemed-to-be-University)

comes into play and become really necessary so that it can help us grant loans to those deserving and not to lend out such a lump sum amount to those who don't, and hence few Machine Learning models will allow us to properly guide us through this problem which will give us a proper outcome based on the various factors discussed above.

II. RELATED WORK

During our survey, we found many studies that focus on extracting various components that we find important while selecting the major features while deciding which factors to be taken into consideration while deciding whether to give the loan to a person or not.

We have found various ways which help in predicting the chances of getting the loan back, and among them using Machine Learning Algorithms provided the best results, and hence we have summarized our findings in sub-sections sticking to this topic.

A. Predictive and Probabilistic Approach using Logistic Regression: Application to Prediction of Loan Approval

During our search we found that Logistic Regression [1] also known as Logistic Model is one of the most used algorithms and has a consistent outcome out of all other algorithms that can be used.

The Logistic Regression algorithm [2] being the simplest and yet an effective algorithm to implement has made a wide impact on the whole Machine Learning community to use it in any predictive need where the output is categorical in nature

To find out the number of times a possibility might occur for a categorical output by fitting the selected features in the logistic curve. Shown in Fig. 1

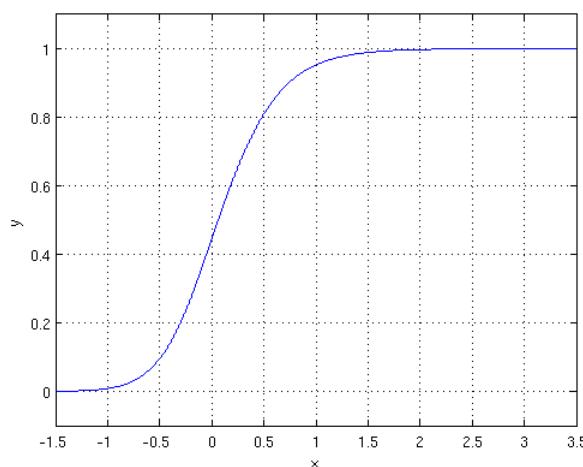


Fig. 1 A General Logit Curve

B. Predictive Analysis of Classification Algorithms on Banking Data

This paper talks about using Logistic Regression [1][2] as one of its methods but also uses Decision trees [2][3], SVM [3] and many other classification based algorithmic models to perform a predictive analysis based on the relation between two factors which are called Recall and Precision to find out the F1 score.

They divided the outcomes based on binary numbers which are 0 if the loan was not approved and 1 if the loan was approved and based on these classes they were able to perform different performance measuring tasks to understand the efficiency of their algorithmic models and the methods they used for it.

The results for these metrics can be calculated after the creation of a confusion matrix [3], which is a table that is used to calculate the accuracy of the classification model on the test data when the actual outcome is known to us.

They have considered the F1 score [3] as their main metrics to be used to conclude their analysis. The above-mentioned metric can be calculated from the confusion matrix (Fig. 2) as such:-

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Fig. 2 Confusion Matrix

F1 Score = $(2 * \text{Precision} * \text{Recall}) / (\text{Precision} + \text{Recall})$
where:-

1. Recall (or how many correct predictions were found) can be calculated as:- $\text{TP}/(\text{TP} + \text{FN})$, and,
2. Precision (or out of the group of correct predictions found how many were actually correct) can be calculated as: $\text{TP}/(\text{TP} + \text{FP})$

After this they were able to see that the Logit Model had the highest accuracy out of all the models based on their methods

C. Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process

In this paper during survey it was found that they were also an avid user of the Logistic Regression classifier [1][2], as it is the simplest and the easiest way of finding out the categorical output for any predictive analysis. But they have also ventured out into other classifier models such as decision trees [3], SVM and Naïve Bayes(NB) Model [5].

SVM [6] is a classifier model where the model tends to draw a separating line called the hyperplane between data points to accurately classify them based on their correlation that is learned by the model based on the data provided. As it keeps on calculating and readjusting the hyperplanes [3] it finally reaches a point where the Support Vectors or the closest points to the hyperplane have been moved away from the line and have a considerable distance between them and thus forms the Decision Boundary. The gap between the support vector or the closest point to the hyper plane determines how good the model is as if the gap is considerable then they can accommodate more new data points accurately into the groups they belong to. This is an unique algorithm as it does both the tasks of a supervised learning algorithm depending on the type of data it is fed and thus is a very flexible classifier (Fig. 3)

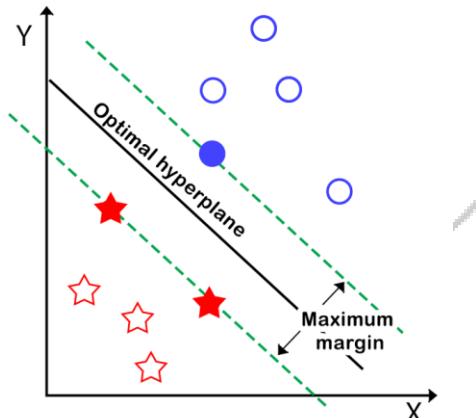
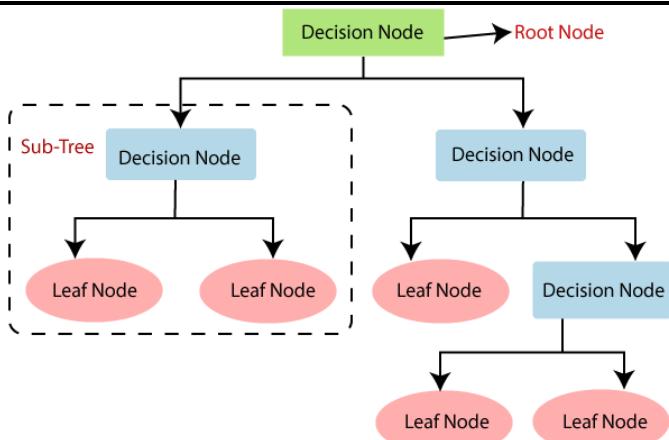


Fig. 3 SVM

D. Modern Approach for Loan Sactioning in Banks Using Machine Learning

This paper in among other Machine Learning models talks about the Decision Tree [11], out of other machine learning which are also used like Linear Logistic and Support Vector Machines (SVM) [5] which is a tree shaped classifier model that is the most human readable out of all.

The decision tree [9] is represented as an upside down tree where the root is the base node (Fig 4) which is used to represent the whole training dataset as a whole which is mentioned later on as we have discussed it further, and is selected as it has the higher information gain and the lowest cost which is calculated using the cost function

*Fig 4. Decision Tree*

The cost function is used to measure how much accuracy it will cost the final output if a certain feature from the dataset is selected, this is used in Recursive Binary Splitting of the dataset.

This model tends to split the data into different parts and then decides whether the tree should be extended based on the outcome from the previous nodes. Decision trees are mainly used to solve classification problems.

A Decision Tree [8] has important parts :-

1. Root Node - is the place where the whole model starts and it represents the dataset as a whole which will get divided as the model keeps on performing
2. Decision Node - This is the node based on which a decision is taken and the algorithm further splits the node into subsets
3. Leaf Node - contain the final output for a decision to be taken once the decision tree gets two leaf nodes as its answer for a decision to be taken the model stops building from there on
4. Splitting - is referred to as splitting the root node into sub-nodes according to the decision taken
5. Pruning - is referred to as the process of removing unwanted branches/sub nodes from the tree that can skew the decision making and making it unfavourable.

The decision tree used the CART algorithm to solve its classification problem which stand for classification and regression algorithm.

But this algorithm has an issue due to the cost function which is that it's an iteratively Greedy algorithm as it tends to reduce the cost of splitting and increase the accuracy exponentially.

Decision trees [10] can further be optimised using the Pruning method as mentioned above. This helps to decide when the decision tree should be stopped from classifying anymore. In this method the decision is stopped from growing to reduce the error of overfitting. It can done by removing the features which are not relevant or are redundant in nature in increasing the accuracy and these decision nodes are then replaced by leaf nodes and then stopped as the decision nodes can't be further split into other nodes.

Due to the simplicity in the understanding of the model it is one of the most preferred algorithm to use when the output to be found in categorical or constant in nature

IV.CONCLUSION

In this paper, we have surveyed and discussed the different ways in which the problem of giving out loans to people have been solved using various machine learning models. This summarized paper gives a proper idea about how important the need of proper predictive analysis has become and classifying different possible candidates has become, in terms of giving a proper insight and knowledge about any banking dataset required for predicting loans and efficiently optimizing the process of sanctioning loans to individuals based on their data. This knowledge will in the end reduce the risk of giving out money to potential NPA for the banks and hence save themselves from incurring a kind of loss and reduce the stress of being losing out their only source of income.

V. REFERENCE

- [1] A. Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval", 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, India, 2017, pp. 1-6, <https://doi:10.1109/ICCCNT.2017.8203946>
- [2] Krishnan k. Pandey, Abhishek Giri, Saket Sharma, Anurag Singh, "Predictive Analysis of Classification Algorithms on Banking Data", International Journal of Innovative Research in Advanced Engineering(IJIRAE) ISSN: 2349-2163, Issue 05, Volume 8 (May 2021)
- [3] E. Chandra Blessie, R. Rekha "Exploring the Machine Learning Algorithm for Prediction the Loan Sactioining Process" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9, November , 2019 DOI:10.35940/ijitee.A4881.119119
- [4] Rath G.B.. Das D.. Acharva B. (2021) Modern Approach for Loan Sanctioning in Banks Using Machine Learning. In: Patnaik S.. Yang XS.. Sethi I. (eds) Advances in Machine Learning and Computational Intelligence. Algorithms for Intelligent Systems. Springer. Singapore. https://doi.org/10.1007/978-981-15-5243-4_15
- [5] Tejaswini ,T. Mohana Kavya , R. Devi Naga Ramya , P. Sai Triveni Venkata Rao Maddumala "ACCURATE LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH", Vol 11, Issue 4, April 2020, ISSN No: 0377-9254
- [6] Yash Divate, Prashant Rana, Pratik Chavan" Loan Approval Prediction Using Machine Learning", e-ISSN: 2395-0056, p-ISSN: 2395-0072 Volume: 08,Issue 05, May 2021
- [7] Kanishk gupta, Binayak Chakrabarti, Aseer Ahmad Ansari, Siddharth S. Rautray, Manjusha Pandey"Loamification – Loan Approval Classification using Machine Learning Algorithms", ICICC- 2021
- [8] Kumar Arun, Garg Ishaan, Kaur Sameet "Loan Approval Prediction based on Machine Learning Approach", IOSR Journal of Computer Engineering(IOSR-JCE), e-ISSN: 2278-0661, p-ISSN:2278-8727, PP 18-21
- [9] Ms. Kathe Rutika Pramod, Ms. Dapse Punam Laxman, Ms. Panhale Sakshi Dattatray Mr. Ghorpade Dinesh B, Ms. Avhad Pooja Prakash "An Approach for Prediction of Loan Approval Using Machine Learning Algorithm", ISSN: 2320-2882
- [10] Ashwini S. Kadam , Shraddha R. Nikam, Ankita A. Aher, Gayatri V. Shelke, Amar S. Chandgude, "Prediction of Loan Approval Using Machine Learning Algorithm ",International Research journal of

Engineering and Technology(IRJET).e-ISSN: 2395-0056,
p-ISSN: 2395-0072, Volume: 08, Issue: 04, Apr 2021

- [11] L. J. Muhammad . . . Ebrahem A. Algehyne . . .
Sani Sharif Usman . . . Abdulkadir Ahmad . . .
Chinmay Chakraborty . . ." A. Mohammed6Supervised
Machine Learning Models for Prediction of COVID-19
Infection using Epidemiology Dataset " SN Computer
Science (2021) 2:11 <https://doi.org/10.1007/s42979-020-00394-7>

