Loan Approval Prediction using Machine Learning Algorithms Approach

Nitesh Pandey[1], Ramanand Gupta[2], Sagar Uniyal[3], Vishal Kumar[4]
[1,2,3,4] *Department of Computer Science and Engineering, Raj Kumar Goel Institute of Technology, AKTU*

Abstract **- Banking system have large number of products to earn profit, but their vital source of income is from its credit system. Because Credit system can earn from interests of that loans which they credit. Banking system always need accurate modelling system for large number of issues. The prediction of credit defaulters is one of the difficult task for any bank. But by forecasting the loan defaulters, the banks definitely may reduce its loss by reducing its 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. This makes the study of this loan approval prediction important. Machine Learning techniques are very crucial and useful in prediction of these type of data. In this research paper four algorithms of classification-based machine learning that is Logistic Regression, Decision tree, Support vector Machine and Random forest is applied and among them Support Vector Machine algorithm is most accurate to predict the loan approval with large accuracy.**

Index Terms **- Loan, Machine Learning, Prediction, Testing, Training.**

I.INTRODUCTION

One of the most important factors which affects our country's economy and the financial condition is the credit system governed by the banks. The process of bank credit risk evaluation is recognising at banks across the globe. "As we know credit risk evaluation is very crucial, there are variety of techniques are used for risk level calculation. In addition, credit risk is one of the main functions of the banking community" [4]. "In this paper we have taken the information of past clients of different banks to whom on a bunch of boundaries advance were endorsed. So, the AI model is prepared on that record to get precise outcomes as it is known to all that circulation of credits is the fundamental business of each bank. Our fundamental goal of this examination is to anticipate the wellbeing of credit" [4]. The primary segment of bank's beneficial resource straightforwardly comes from the benefit acquired from the advances being circulated by the bank [4].

Loan Approval Prediction is extremely handy for employee of banks as well as for the applicant also. "The purpose of this Paper is to offer a quick, immediate, and simple method for selecting deserving applicants. "It is the responsibility of bank to ensure their assets is in the right hand. By implementation of this system we will be able to predict and ensure that applicant for the loan is safe or not by this automatic loan approval prediction system" [3]. "One of disadvantage of this model is that it emphasize different weights to each factor but in real life sometime loan can be approved on the basis of single strong factor only, which is not possible through this system" [4].

There may be various benefits that the bank can obtain, such as setting a time limit for applicants to check and ensure whether their loan will be sanctioned or not. This prediction system may be helpful in the sense that, it gives the right to bankers to more focus on valuable assets for the bank not focus on the poor applicants. It will reduce the time for the loan application process of the applicant. "Result against particular Loan Id can be send to various department of banks so that they can take appropriate action on application. This helps all others department to carried out other formalities" [4].

In this research paper, there are eight sections. In the next section we have defined the problem statement. After that, there is a brief description of our dataset. In the next section there is a literature survey. In the next, the algorithms used to make the model. In next, a brief discussion about feature engineering and then result and analysis and then conclusion.

II.PROBLEM STATEMENT

Account firms and banks need to automatize the credit qualification activity (continuously) essentially dependent on data given by customers when rounding out an online structure. Sex, Marital Status, Education, Number of Dependents, Salary, Loan Amount, Credit History, and different subtleties are incorporated. To digitize this interaction, they made an issue to group the client sections that can apply for a credit sum, permitting them to focus on these clients explicitly. They have introduced a fractional informational collection for this situation.

"Approval of Loan is a very common real-life problem that every company faces in their lending operations. If the loan approval process is automated, it can save a lot of man hours and improve the speed of service to the customers. The increase in customer satisfaction and savings in operational costs are significant" [9]. "However, the rewards can only be realised if the bank has a sturdy model in place to accurately forecast which client's loans it should accept and which it should reject, in order to reduce potential risk" [2].

III.DATA SET

Variable	Description	Type
Loan_ID	Unique Loan ID (Primary Key)	Integer
Gender	Female or Male	Character
Married	Yes or No	Character
Dependents	Number of dependencies	Integer
Education	Graduate /Un-graduate	String
Self Employed	Yes or No	Character
Applicant Income	Applicant's Income	Integer
Co Applicant Income	Co-applicant's Income	Integer
Loan_Amount	Loan Amount	Integer
Loan_Amount _Term	Term of Loan (In months)	Integer
Credit_History	Applicant's Credit history	Integer
Property_Area	Urban/Semi Urban/Rural	String
Loan_Status	Yes or No	String

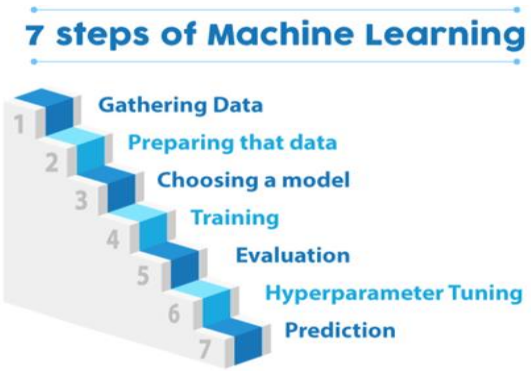

Fig 1: Methodology for Typical Machine Learning approach

IV.LITERATURE SURVEY

"Vaidya had suggested a method for approving loan forecasts using logistic regression" [6]. "Logistic Regression is one of the most popular and very useful classification based algorithm" [6]. "The purpose or the importance of using Logistic Regression was that it uses the concept of predictive analysis which was suitable enough for describing the data" [6].

"M. Bayraktar et al. proposed a method for credit risk analysis using machine learning. Boltzman machine was used to make the analysis for risk calculation of loan" [7].

"Y. Shi and P. Song proposed a method for evaluating project loans using risk analysis. The method evaluate the risk involved in loans of commercial banks" [8].

"V. C. T. Chan et al. proposed a credit approval system using web-services. For clients loan as approved by the system. The consumer provides extra relevant information with the credit application. This information's are processed by Credit Approval System which finally give credit score to the applicant. The paper developed a web services based solution of this problem" [9].

After going through this, it is found that loan approval prediction task is very crucial for banking system. Machine learning algorithm are very helpful in predicting outcomes even when data is huge in size.

V.MACHINE LEARNING ALGORITHMS USED

A.RANDOM FOREST

Random Forest is a prominent learning method in Supervised Machine Learning, and it is effective for Regression & Classification tasks in ML.

1. It generates random forests and then uses these random forests to seek the solutions.
2. It is an ensemble learning in which significant number of classifiers are used to solve a complex problem.
3. Random forest analyzes each tree for prediction rather than just one to avoid overfitting issues.
4. The greater the number of trees, the greater the accuracy in problem solving.

B.SUPPORT VECTOR MACHINE (SVM)

SVM is a well known Supervised Machine Learning algorithm. Currently, the SVM classifier is the most common classifier

1.SVM has demonstrated a wide range of exceptional abilities, particularly in classification problems.

C.DECISION TREE

Decision Tree is a non-parametric supervised machine learning algorithm.

1. It can also be used for both classification and regression, although it is more widely used for classification technique.
2. It works with categorical and continuous variables.

D.LOGISTIC REGRESSION (LR)

One of the most common algorithm for machine learning is logistic regression, which also falls within the supervised learning method.

1. For LR, the sample size should be big.
2. It is used for predicting categorical target variable.
3. It returns the probablistic value instead of producing 0 or 1.

VI.FEATURE ENGINNERING

In Feature Engineering cleaning of dataset takes place. A clean and well-prepared dataset results in a more accurate and faster model. Various techniques such as Imputation, Binning, One Hot Encoding, Handling outliers comes under this.

VII.RESULT AND ANALYSIS

With the model trained, it needs to be tested. The data which we split during test trained module is used for evaluation the model. Mainly confusion metrics, precision, recall, accuracy and F1 score methods are used for evaluating the classification problem.

a. CONFUSION MATRIX

TRUE POSITIVE (TP)

The forecasted value matches the actual value i.e the positive value was predicted as positive.

Algorithm	True Positive
Random Forest	21
Support Vector Machine	21
Decision Tree	23
Logistic Regression	27

TRUE NEGATIVE (TN)

- The predicted value by the model exact matches the actual value.
- The actual value was negative and the model also predicted a negative value.

Algorithm	True Negative
Random Forest	73
Support Vector Machine	77
Decision Tree	63
Logistic Regression	66

FALSE POSITIVE (FP)

- The predicted value was incorrectly or falsely predicted.
- The model predicted a positive value despite the fact that the actual value was negative.

Algorithm	False Positive
Random Forest	24
Support Vector Machine	24
Decision Tree	22
Logistic Regression	18

FALSE NEGATIVE (FN)

- The predicted value was incorrectly or falsely predicted.
- The actual value was already positive but the model predicted a negative value

Algorithm	False Negative
Random Forest	5
Support Vector Machine	1
Decision Tree	15
Logistic Regression	12

b. ACCURACY

It is one of the method of evaluating the classification problem. It is the ratio of Number of correct prediction to the total number of prediction.

Algorithm	Accuracy (in %)
Random Forest	76.42
Support Vector Machine	79.67
Decision Tree	70
Logistic Regression	75.60

c. PRECISION

Precision actually tells us how many of the correctly predicted cases were found to be positive.

$$precision = \frac{tp}{tp + fp}$$

Algorithm	Precision
Random Forest	0.46
Support Vector Machine	0.46
Decision Tree	0.51
Logistic Regression	0.6

d. RECALL

Recall tells us how many actual positive cases our model was able to predicted correctly.

$$recall = \frac{tp}{tp + fn}$$

Algorithm	Recall
Random Forest	0.80
Support Vector Machine	0.95
Decision Tree	0.60
Logistic Regression	0.69

e. F1 SCORE

It is computed by the calculation of harmonic mean of precision and recall.

$$F_1\text{-}score$$

$$\frac{1}{\frac{1}{2}\left(\frac{1}{\text{recall}} + \frac{1}{\text{precision}}\right)}$$

Algorithm	F1 Score
Random Forest	0.58
Support Vector Machine	0.61
Decision Tree	0.55
Logistic Regression	0.64

VIII.CONCLUSION

In this paper, machine learning was used to predict loan acceptance. The prediction method begins with data pre-processing, filling the missing values, experimental data analysis. After evaluating model on test dataset, each of these algorithms obtained a precision rate between 70% and 80%. Although here it can be concluded with certainty that the Support Vector Machine model is very efficient and produces superior results than other models.

REFERENCES

[1] Ashwini S. Kadam, Shraddha R Nikam, Ankita A. Aher, Gayatri V. Shelke, Amar S. Chandgude. “Prediction for Loan Approval using Machine Learning Algorithm”, Apr 2021 International Research Journal of Engineering and Technology (IRJET)

[2] Mohammad Ahmad Sheikh, Amit Kumar Goel,Tapas Kumar. "An Approach for Prediction of Loan Approval using Machine Learning Algorithm", 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), 2020

[3] X.FrencisJensy, V.P.Sumathi,Janani Shiva Shri, “An exploratory Data Analysis for Loan Predict ion based on nature of clients”, International Journal of Recent Technology and Engineering (IJRTE),Volume-7 Issue-4S, November 2018

[4] J. Tejaswini1, T. Mohana Kavya, R. Devi Naga Ramya, P. Sai Triveni Venkata Rao Maddumala. “ACCURATE LOAN APPROVAL PREDICTION BASED ON MACHINE LEARNING APPROACH” Vol 11, www. jespublication.com, page 523 Issue 4, April/ 2020 ISSN NO: 0377-9254

[5] “Prediction for Loan Approval using Machine Learning Algorithm” International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 08 Issue: 04 | Apr 2021 www.irjet.net p-ISSN: 2395-0072

[6] Vaidya, "Predictive and probabilistic approach using logistic regression: Application to prediction of loan approval," 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Delhi, 2017, pp. 1-6.doi: 10.1109/ICCCNT.2017.8203946

[7] M. Bayraktar, M. S. Aktaş, O. Kalıpsız, O. Susuz and S. Bayracı, "Credit risk analysis with classification Restricted Boltzmann Machine," 2018 26th Signal Processing and Communications Applications Conference (SIU), Izmir, 2018, pp. 1-4.doi: 10.1109/SIU.2018.8404397

[8] Y. Shi and P. Song, "Improvement Research on the Project Loan Evaluation of Commercial Bank Based on the Risk Analysis," 2017 10th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, 2017, pp. 3-6.doi: 10.1109/ISCID.2017.60

[9] V. C. T. Chan et al., "Designing a Credit Approval System Using Web Services, BPEL, and AJAX," 2009 IEEE International Conference on e-Business Engineering, Macau, 2009, pp. 287-294.doi: 10.1109/ICEBE.2009.46