

Prediction of Modernized Loan Approval System Based on Machine Learning Approach

Vishal Singh
B.Tech

School of Computing Science
Greater Noida, India
vishalsingh220799@gmail.com

Ayushman Yadav
B.Tech

School of Computing Science
Greater Noida, India
ayushman.up10@gmail.com

Rajat Awasthi
B.Tech

School of Computing Science
Greater Noida, India
awasthirudra3333@gmail.com

Guide: N.Partheeban
Professor

School of Computing Science &
Engineering (SCSE),
Galgotias University
n.partheeban@galgotiasuniversity.edu. in

Abstract: Technology has boosted the existence of humankind the quality of life they live. Every day we are planning to create something new and different. We have a solution for every other problem we have machines to support our lives and make us somewhat complete in the banking sector candidate gets proofs/ backup before approval of the loan amount. The application approved or not approved depends upon the historical data of the candidate by the system. Every day lots of people applying for the loan in the banking sector but Bank would have limited funds. 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.

Keywords:- Machine learning, Data, Loan, Training, Testing, Prediction

I. INTRODUCTION

Prediction of modernized loan approval system based on machine learning approach is a loan approval system from where we can know whether the loan will pass or not. In this system, we take some data from the user like his monthly income, marriage status, loan amount, loan duration, etc. Then the bank will decide according to its parameters whether the client will get the loan or not.

So there is a classification system, in this system, a training set is employed to make the model and the classifier may classify the data items into their appropriate class. A test dataset is created that trains the data and gives the appropriate result that, is the client potential and can repay the loan.

Prediction of a modernized loan approval system is incredibly helpful for banks and also the clients. This system checks the candidate on his priority basis. Customer can

submit his application directly to the bank so the bank will do the whole process, no third party or stockholder will interfere in it. And finally, the bank will decide that the candidate is deserving or not on its priority basis. The only object of this research paper is that the deserving candidate gets straight forward and quick results.

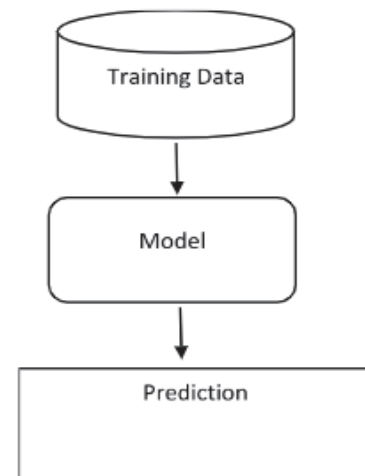


Fig. 1. Basic Machine Learning Model

II. MACHINE LEARNING ALGORITHMS

In this research paper we are using three Machine Learning algorithms which are used to find out the correct prediction of Data set.

- (a) **XGBoost** – XGBoost is a Decision tree based open source software library. It implements machine learning algorithms that uses a gradient boosting framework. It works on Linux, Windows, and macOS.
- (b) **Random Forest**– Random forests is a classification algorithm which builds big number of Decision tree, whose prediction is more accurate than any of individual decision tree.

(c) **Decision Tree** – A Decision tree split the dataset in to smaller parts. And then predict the every chances.

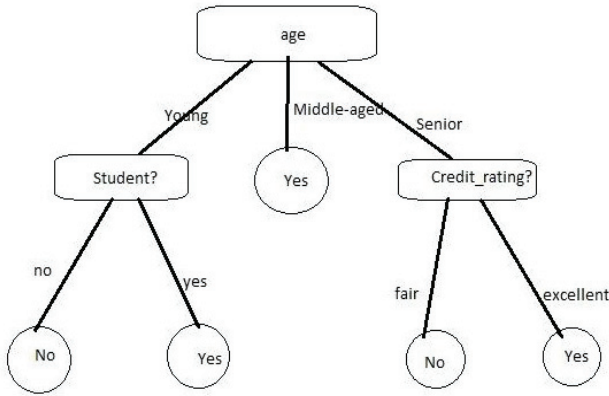


Fig. 2. Decision Tree

III. PROBLEM FORMULATION

There is a major problem that many people not able to back the loans to banks. And banks are going in losses.

Banks received many applications for loan approval day by day and not everyone gets approved. Most of the banks have their own credit score and risk assessment techniques so as to check that the loan is approved or not. Why this loan problem arises this question will get resolved in just a few minutes. The main reason to get a loan is to fulfill the needs of something. For a businessman he/she wants to increase the business or if that company is at loss to get over from that he/she needs a loan. In middle-class people wants to fulfill their needs so they want a loan. So, the main thing of this to fulfill the needs of someone or for something.

Again the question arises that what are the problems that are forming in providing the loans. The answer to this question that not everybody can loan because if he/she is not able to return then who is providing the loan he/she or the company or the bank that is providing the loan will get in the loss. So, first who is providing the loan they have to verify or set some criteria that who is taking the loan is able to return or not. Like in banks like we have a credit card facility but not everybody gets a credit card. For that, a credit score is there to check whether eligible or not. For credit score one should have a good credit score then he/she be able to get a loan. Some criteria like a source of income should be there for getting a credit card. Banks provide loans on behalf of one who is taking the loan he/she should provide some documents and verify. Like some company not able to provide the loans then banks get in loss and they called it NBFC's.

During this project data processing algorithms are going to study loan-approved data might help in predicting the like defaulters thereby helping the banks for creating better decisions within the future.

IV. REQUIRED TOOLS

- MS Office
- Jupyter notebook
- Python3

- Data set
- Numpy
- Pandas
- XGBoost
- Machine learning algorithms
- Matplotlib

V. LOAN PREDICTION DATA ANALYSIS

The question arises that on what basis we analyze that we should provide the loan or not. We have two target variables on that basis we provide the loan to our customer. We have to check all the formalities like income proof, address proof, id proof, etc. Then we provide the loan that the customer is eligible to return or not. In the middle class, there is a major need for loans as parents need for their child's education, for business also.

In some cases, people suddenly undergo financial crises, while some try and scam money from banks.

So, the reason we have to check all the things because banks are not undergoing an NPA loan.

Better the customer, chances of loan to be back are high.

Background verification should be high so that we can expect a return of the loan at the perfect time.

So, we analysis on several bases and these are called our target variable

- Data set

TABLE I. MAIN DATA SET

Variable Name	Description	Type
Loan_ID	UniqueLoan_ID	Integer
Gender	Male/ Female	Character
Married	Applicant	married (Y/N) Character

TABLE II. DATA SET

Variable Name	Description	Type
Dependents	Number of dependents	Integer
Education	Graduate/ Under Graduate	String
Self_Employed	Self Employed	(Y/N) Character
Applicant_Income	Applicant income	Integer
Co_Applicant_Income	Coapplicant income	Integer
Loan_Amount	Loan amount in thousands	Integer
Loan_Amount_Term	Term of loan in months	Integer
Credit_History	credit history guidelines	Integer
Property_Area	Urban/ Semi Urban/ Rural	String
Loan_Status	Loan Approved(Y/N)	Character

VI. LOAN PREDICTION METHODOLOGY

This proposed model will characterize the behavior of customers on the Basis of their record.

These records is taken from the customers, and create a data set. With the help of These data sets and training machine learning model, we predict that the customer's loan will passed or not.

This Machine algorithms predict the possibility of a customer would be able to repay the loan or not

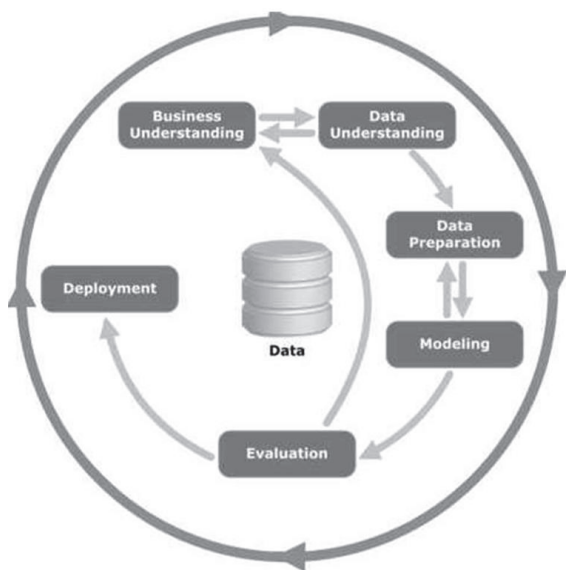


Fig. 3. Process diagram

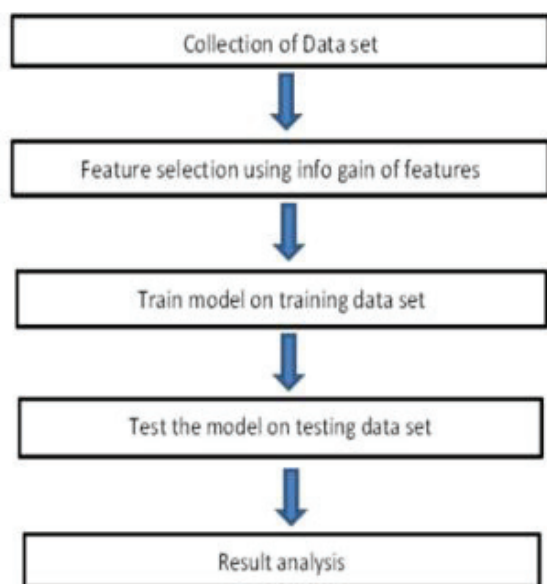


Fig. 4. Loan Prediction Methodology

VII. MERITS OF PROPOSED SYSTEM

In this, we are going to discuss the advantage of loan prediction. In this system, we are going to predict that the person who is applying for a loan can repay or not. If the client can repay then we predict that yes, eligible for a loan. And if the candidate fails then we predict that client is not eligible.

The advantage of this system is that we provided some conditions by setting the algorithms and just by evaluating the details, we get to know eligibility criteria that client is eligible or not.

This system may be built which is able to take various inputs from the users like salary, address, loan amount, loan duration, etc and provide a prediction of whether their application will be approved by the bank or not.

This Research paper helps the banks to minimize the possible losses and can increase the volume of credits.

VIII. ARCHITECTURE DIAGRAM FOR PROPOSED METHOD

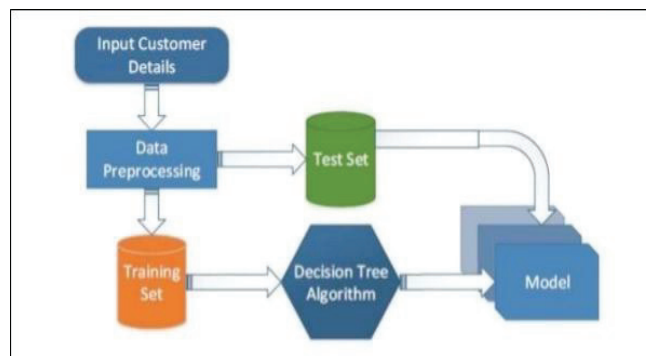


Fig. 5. Architecture Diagram

IX. CONCLUSION:

According to this research paper prediction accuracy is sweet for both datasets. In some situations like client going through some disaster so here the algorithm cannot predict the appropriate result. This research paper can find out the client is potential and repay the loan and the accuracy is good. loan duration, loan amount, age, income are the most important factors for finding out there (whether the client would have been). 'zip code' and 'credit history' are the foremost important factors for predicting the category of the loan Applicant.

Submission message	Score	Code File	Solution File	Final Solution
XGBoost	0.777777777777778	--	Download	<input type="radio"/>
Random Forest	0.763888888888889	--	Download	<input type="radio"/>
Decision Tree	0.645833333333333	--	Download	<input type="radio"/>
Third submission	0.777777777777778	--	Download	<input type="radio"/>
Second submission	0.777777777777778	--	Download	<input type="radio"/>
First submission	0.784722222222222	--	Download	<input type="radio"/>

Fig. 6. Result

ACKNOWLEDGMENT

I would like to thank Galgotias University for proposing the Capstone Project in our curriculum so that we can learn new concepts easily by doing hands-on. I would like to thank our Vice-Chancellor Mrs. Preeti Bajaj for encouraging us to learn more by doing practical. I would like to extend my regards to my mentor and my project guide Dr. N Partheeban who have helped me in getting this project ready. They have always stood by me, whenever I needed them, and they have

done a lot for me in order to get this project completed. Without their valuable help, this project would not have been completed. I would also like to thank the Internet for providing a large source of information that I needed for completing this project. Once, again I would like to thank all the people involved in this project from the bottom of my heart, and please forgive me if I have forgotten to mention any name. This project is completed with your help only. THANK YOU.

REFERENCES

- [1] Amruta S. Aphale and R. Prof. Dr. Sandeep. R Shinde, "Predict Loan Approval in Banking System Machine Learning Approach for Cooperative Banks Loan Approval", International Journal of Engineering Trends and Applications (IJETA), vol. 9, issue 8, 2020)
- [2] Loan Prediction Using Ensemble Technique, International Journal of Advanced Research in Computer and Communication Engineering Vol. 5, Issue 3, March 2016
- [3] Exploratory data analysis
https://en.wikipedia.org/wiki/Exploratory_data_analysis
- [4] Pandas Library <https://pandas.pydata.org/pandas-docs/stable/>
- [5] MeanDecreaseAccuracy
<https://dinsdalelab.sdsu.edu/metag.stats/code/randomforest.html>