Loan Amount Prediction Based on Credit Historyand Other Variables

Anup Lakra, Rupesh Kumar Yadav Amity Institute of Information Technology
Amity University Raipur, Chhattisgarh

ABSTRACT

Loan approval is a very important process for banking organizations. The system approved or reject the loan applications.
Recovery of loans is a major contributing parameter in the financial statements of a bank. It is very difficult to predict the possibility of payment of loan by the customer. In recent years many researchers worked on loan approval prediction systems. Machine Learning (ML)techniques are very useful in predicting outcomes for large amount of data. In this paper three machine learning algorithms, Logistic Regression (LR), Decision Tree (DT) and Random Forest (RF) are applied to predict the loan approval of customers. The experimental results conclude that the accuracy of Logistic Regression machine learning algorithm is better as compared to Decision Tree and Random Forest machine learning approaches.

I. INTRODUCTION

A loan is the core businesspart of banks. The main portion the bank's profit is directly come from the profit earned from the loans. Though bank approves loan after a regress process of verification and testimonial but still there's no surety whether the chosen hopeful is the right hopeful or not.

This process takes fresh time while doing it manually. We can prophesy whether that particular hopeful is safe or not and the wholeprocess of testimonial is automated by machine literacy

style. Loan Prognostic is really helpful for retainer of banks as well as for the hopeful also.

II. EXISTING SYSTEM

Bank employees check the details of applicant manually and give the loan to eligible applicant. Checking the details of all applicants takes lot of time. The artificial neural network model for predict the credit risk of a bank. The Feed- forward back propagation neural network is used to forecast the creditdefault. The method in which two or more classifiers are combined together to produce a ensemble model for the better prediction.

They used the bagging and boosting techniques and then used random forest technique. The process of classifiers is to improvethe performance of the data and itgives better efficiency. In this

work, the authors describe various ensemble techniques for

binary classification and also for multi class classification. The new technique that is described by the authors for ensemble is COB which gives effective performance of classification, but it also compromised with noise and outlier data of classification. Finally, they concluded that the ensemble-based algorithm improves the results for training data set.

III. PROPOSEDSYSTEM

To deal with the problem, we developed automatic loan prediction using machine learning techniques. We will train the machine with previous dataset. so, machine can analyse and understand the process. Then machine will check for eligibleapplicant and give us result.
Advantages

- Time period for loan sanctioning will be reduced.

- Whole process will be automated, so human error will be avoided Eligible applicant willbe sanctioned loan without any delay.

IV. ARCHITECTURE TECHNIQUES

Decision Tree

Decision tree algorithm in machine erudition how's which efficiently performs both family and retrogression tasks. It creates decision trees. Decision trees are universally used in the banking assiduousness due to their high exactitude and capableness to formulate a statistical model in plain language. In Decision tree each knot represents a criterion (diagnostic), each link (branch) represents a decision (rule) and each chip represents an outce (categoricalor continues value).

V. IMPLEMENTATION METHODLOGY

Data Collection:
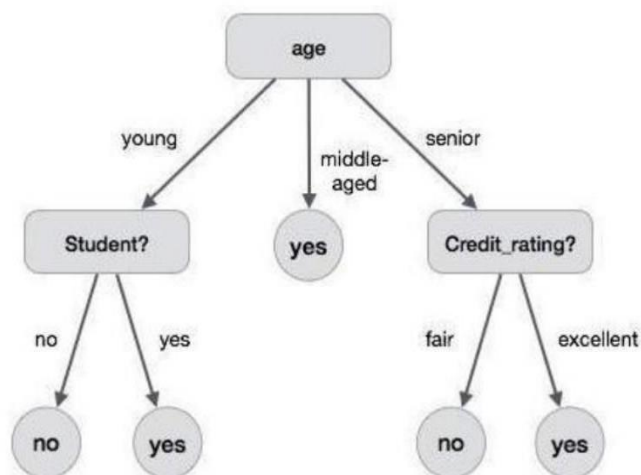
The dataset collected for foretelling loan failure clients is foretold into Training set and testing set. Generally, 8020 proportion is applied to dissociate the training set and testing set. The data model which was created using Decision tree is applied on the training set and hung on the test take fineness, Test set forecasting is done. Following are the attributes:

Variable	Description
Loan_id	Unique loan id
Gender	Male/ female
Married	Application married(Y/N)
Dependents	Number of dependents
Education	Applicant education(graduate/under graduate)
Self_employed	Self employed(Y/N)
Applicant income	Applicant income
Co Application income	Co application income
Loan_Amount	Loan amount in thousands
Loan_Amount_term	Term of loan in months
Credit_history	Credit history meets guidelines
Property_area	Urban/semi Urban/rural
Loan_status	Loan approved(Y/N)

Preprocessing:

The collected data may contain missing values that may
lead to inconsistency. To gain betterresults data need to be preprocessed and so it'll better the effectiveness of the algorithm. We should remove
the outliers and we need to convert the variables. In order to flooring these issues, we use chart function.

Train model on training dataset

Now we should train the model on the training dataset and make soothsaying's for the test dataset. We can divide our train dataset into two tract train and testimony. We can train the model on this training part and using that make soothsaying's for the testimony part. In this way, we can validate our soothsaying's as we've the true soothsaying's for the testimony part (which we don't have for the test dataset)

Correlating attributes:

Grounded on the correlation among attributes it was observed more likely to pay back their loans. The attributes that are individual andsignificant can include Property area, education, loan measure, and originally credit History, which issince by insight it's considered as important. The correlation among attributes can be associated usingcorplot and boxplot in Python platform

VI. RESULTS

Decision Tree:

Decision tree is a type of supervised education algorithm (having a pre- defined target variable) that is generally used in category problems. In this approach, we disassociate the population or sample into two or added homogeneous sets (or sub-populations) based on the most significant splitter/ differentiator in input variables.

Decision trees use multiple algorithms to decide to disunite a bump into two or added sub- knots. The creation of sub- knots increases the unsophistication of attendant sub- knots. In other words, we can say that chasteness of the bump increases with respect to the target variable.

Predicting the outcomes:

Using decision tree algorithm, the outcomes of all applicants can be stored in any file.

Algorithm:

1. Import all the required python modules

2. Import the database for both TESTING and TRAINING.

3. Check any NULLVALUES are exists

4. If NULLVALUES exits , fill the table with corresponding coding

5. Exploratory Data Analysis for all ATTRIBUTES from the table

6. Plot all graphs using MATPLOTLIB module

7. Build the DECISIONTREE MODEL for the coding

8. Send that output to CSVFILE

	Loan_ID	Loan_Status
0	LP001015	Y
1	LP001022	Y
2	LP001031	Y
3	LP001035	N
4	LP001051	Y
5	LP001054	N
6	LP001055	Y
7	LP001056	N
8	LP001059	Y
9	LP001067	Y
10	LP001078	N
11	LP001082	Y
12	LP001083	Y
13	LP001094	N
14	LP001096	Y
15	LP001099	Y
16	LP001105	N
17	LP001107	Y
18	LP001108	Y
19	LP001115	N
20	LP001121	Y
21	LP001124	Y
22	LP001128	Y
23	LP001135	Y
24	LP001149	N

VII.   CONCLUSION

From a proper analysis of positive points and constraints on the member, it can be safely concluded that the product is a considerably productive member. This use is working duly and meeting to all Banker requisites. This member can be freely plugged in numerous other systems. There have been mathematics cases of computer glitches, violations in content and most important weight of features are fixed in automated prophecy system, so in the near future the so – called

software could be made more secure, trustworthy, and dynamic weight conformation. In near future this module of prophecy can be integrated with the module of automated processing system. The system is trained on old training dataset in future software can be made resembling that new testing date should also take part in training data after some fix time.

VIII.   REFERENCES

[1] Kumar Arun, Garg Ishan, Kaur Sanmeet, May-Jun. 2016. Loan Approval Prediction based on Machine Learning Approach, IOSR Journal of Computer Engineering (IOSR-JCE)

[2] Wei Li, Shuai Ding, Yi Chen, and Shanlin Yang, Heterogeneous Ensemble for Default Prediction of Peer-to-Peer Lending in China, Key Laboratory of Process Optimization and Intelligent Decision-Making, Ministry of Education, Hefei University of Technology, Hefei 2009, China

[3] Short-term prediction of Mortgage default using ensembled machine learning models, Jesse C.Sealand on july 20, 2018.

Clustering Loan Applicants based on Risk Percentage using K- Means Clustering Techniques, Dr.
K. Kavitha, International Journal of Advanced Research in