International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT1952198

Loan Default Identification and its Effect

Gopal Choudhary[1], Yash Garud[1], Akshil Shetty[1], Rumit Kadakia[1], Sonali Borase[2]

[1]Department of Computer Science and Engineering, MPSTME, NMIMS, Shirpur, District: Dhule, Maharashtra, India

[2]Assistant Professor, Department of Computer Science and Engineering, MPSTME, NMIMS, Shirpur, District: Dhule, Maharashtra, India

ABSTRACT

Now a days banking sector is on boom everyone is applying for loan but banks have limitation that they have limited assets so they can provide loan to limited loan applications but when they provide loan, they must assure that loan is being granted to only genuine customers. So, this paper focuses on we will try to lessen the uncertainty factor and assure the loan approval to genuine customers only and save the bank assets. That is performed by way of mining the massive data of the earlier data of the human beings to whom the loan become acknowledged earlier than and on the idea of those records/reviews the machine was skilled the use of the system mastering version which provide the maximum correct result. The main focus of the paper will be on the loan to be approved of those customers only who will be able to pay it back.

Keywords : Loan, Machine Learning, Training, Testing, Prediction.

I. INTRODUCTION

When a client takes loan from a bank and financers, he needs to pay it on time but he not paid the loan amount to that bank or financer and default that loan amount. Now that same customer willing is to take loan from more financers and banks and default that loan amount. So, these needs to be stopped somewhere.

Loan portfolio are the largest source of revenue of most of the banks and financers but it's also leads to NPA (Non-Performing Assets) or end up with loan default. And this affect the financial performance of the financers and banks. The loan defaulters are trying to fool multiple financers and banks at a time. So, they needed to be stopped at right time.

The very obvious effect of loan default is that it causes the monitory growths of financers and banks. This result in the increase in the rate of interest on the loan, and this leads to overall failure to the economic growth.

II. METHODS AND MATERIAL

The implementation of the project can be divided into two parts.

1) Data analysis and data cleaning

Selecting relevant features

In machine learning and information, feature selection, additionally known as variable choice, attribute selection or variable subset choice, is the procedure of choosing a subset of applicable capabilities (variables, predictors) to be used in model creation. function choice techniques are used for four motives:

- simplification of models to cause them to simpler to interpret with the aid of researchers/customers
- reduces training time
- To avoid the curse of dimensionality

- enhanced generalization by reducing overfitting (formally, reduction of variance)

Null value imputation

Missing data can occur while no records is supplied for one or greater items or for a complete unit. missing records is a totally big problem in actual existence scenario. missing information can also talk to as NA (now not available) values. In Data Frame now and again many datasets sincerely arrive with lacking information, either as it exists and turned into no longer amassed or it in no way existed. for instance, suppose unique user being surveyed may also pick out now not to share their income, a few users may select no longer to proportion the cope with in this manner many datasets went lacking.

Handling Outliers

An outlier can be referred as a data point this is remote from different comparable factors. they'll be due to variability inside the measurement or might also suggest experimental errors. If feasible, outliers must be excluded from the data set. but, detecting that anomalous times might be very difficult, and isn't usually viable.

Different methods of dealing with outliers:

- Univariate method: This technique looks for data points with extreme values on one variable.
- Multivariate method: right here we search for unusual combinations on all the variables.
- Minkowski error: This approach reduces the contribution of ability outliers within the training method.

Training a machine learning model:

Logistic Regression

Rather than predicting precisely 0 or 1, logistic regression generates a possibility—a value between zero and 1, exclusive. for instance, don't forget a logistic regression version for unsolicited mail detection. If the version infers a cost of 0.932 on a particular electronic mail message, it implies a 93.2% chance that the email message is junk mail. extra precisely, it way that inside the limit of infinite schooling examples, the set of examples for which the model predicts 0.932 will certainly be spam ninety 3.2% of the time and the last 6. 8% will nolonger.

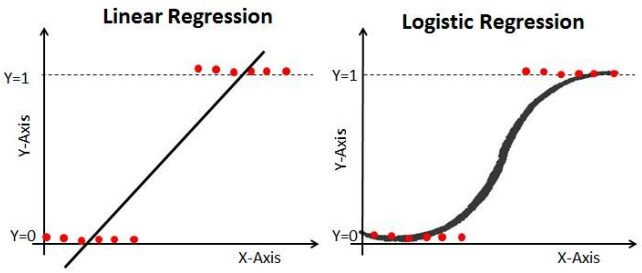

Fig 1. Logistic Regression

Random Forest

Random forest is very easy and flexible, easy to apply ML algorithm that produces, even without hyper-parameter tuning, a top-notch result maximum of the time. it is also one of the maximum used algorithms, as its simplicity and the truth that it can be used for both category and regression duties. in this put up, you'll analyze, how the random forest algorithm works and numerous different vital things approximately it.

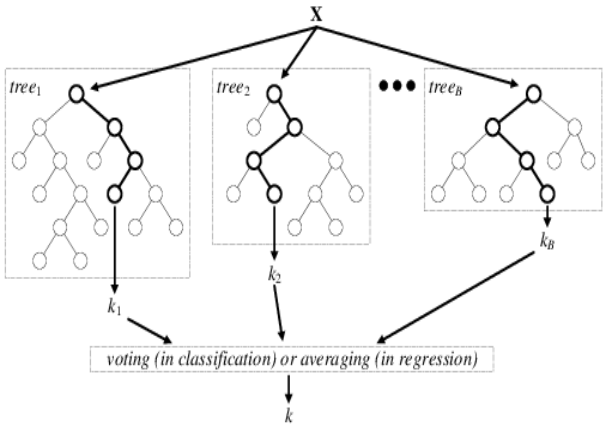

Fig 2. Random Forest

Random Forest is a supervised studying set of rules. Like you can already see from its call, it creates a forest and creates it somehow random. The forest it builds, is an ensemble of decision tree, maximum of the time skilled with the "bagging" technique. the overall concept of the bagging technique is that a combination of learning models increases the overall result.

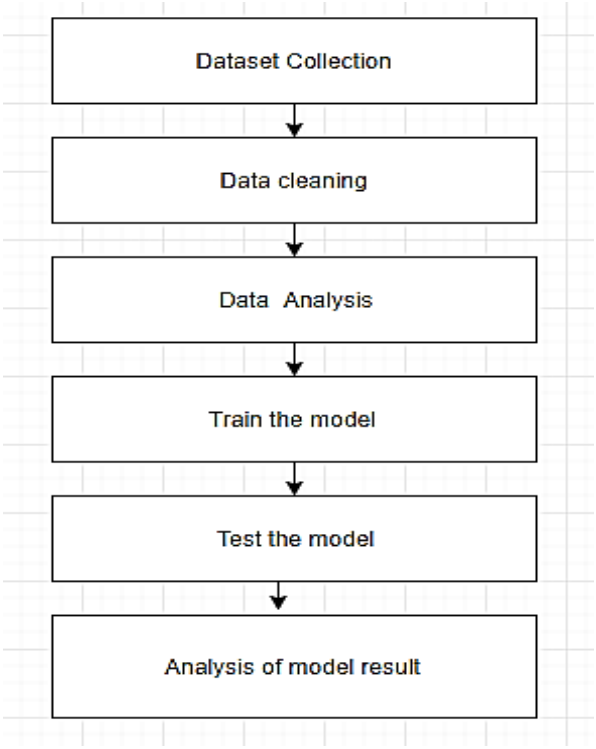

Fig 3. Flow Diagram of the System

III. EXPERIMENTAL RESULT

Data Set

The data which has been trained is being applied to machine learning model, whenever the new customer enters the detail in application form that act as test data set. So, after testing operation is performed, then afterwards the model will predict that will the new customer will be able to pay the loan on time or not on the process of loan approval on the basis of the training dataset

Table 1

Attribute Name	Description	Type
Cust_id	It's a Unique id	Integer
loan_amnt	The Loan amount that has been sanctioned	Integer
term	Duration time of Loan	Character
int_rate	The annual interest rate on the loan amount	Integer
installment	The emi that are monthly paid by the borrower	Integer
grade	LC assigned loan grade	Character
annual_inc	annual_inc: The self-claimed annual income given by the borrower during registration.	Integer
Credit_History	credit history meets guidelines	Integer
Property_Area	Type of area whether it is urban or rural	String
loan_status_coded	Loan Approved(Y/N)	Integer

Accuracy Measure

The model is able to predict 61% of defaulters this can be observed by ROI (Return on Investment). To calculate ROI here we have used fully paid, charged off and loan in grace period So how to calculate ROI = 'total_payment' / 'funded_amount' ROI calculated without using this model=-4.57 and after using this model ROI calculated is 2.22

Grade wise analysis of ROI

As you can see in below table the ROI of C, D, E, G become positive using this model.

	ROI	% Picked	% Default	ROI_w/o_model	% Picked_w/o_model	% Default_w/o_model
A	0.0250712	99.9774	8.64002	0.025071	100	8.63806
B	0.0200738	96.1866	16.1057	0.0152713	100	16.7564
C	0.017765	53.9985	18.7983	-0.0447373	100	25.9779
D	0.0349775	24.597	20.8825	-0.096151	100	34.7948
E	0.0703003	7.08117	22.2561	-0.143831	100	42.5518
F	-0.0234227	3.26508	22.0339	-0.16447	100	47.316
G	0.0961287	2.6694	38.4615	-0.163535	100	50.1027

Table 2. ROI Index Table

IV.CONCLUSION

From a proper analysis of tremendous factors and constraints at the aspect, it is able to be appropriately concluded that the product is a distinctly green aspect. This utility is running nicely and meeting to all Banker necessities. This factor can be without difficulty plugged in lots of different systems. There have been numbers instances of computer system faults, mistakes in content and most essential weight of features is fixed in automatic prediction machine, so in the close to future the so –known as software program may be made extra at ease, dependable and dynamic weight adjustment. In close to future this module of prediction can be combine with the module of automatic processing device. the device is trained on antique schooling dataset in destiny software program may be made such that new testing date have to additionally take part in education records after a few fix times.

V. REFERENCES

[1]. A. Malali, "Predicting Loan Outcomes using Machine Learning," Aug 25, 2016.

[2]. A. J. Hamid, "DEVELOPING PREDICTION MODEL OF LOAN," An International Journal (MLAIJ) Vol.3, No.1, March 2016.

[3]. A. Goyal, "Loan Prediction Using Ensemble Technique," International Journal of Advanced Research in Computer and Communication Engineering, March 2016.

[4]. A. Goyal, "A survey on Ensemble Model for Loan Prediction," International Journal of Engineering Trends and Applications (IJETA) – Volume 3, Feb 2016.

[5]. R. S. T. Sivasree M S, "Loan Credibility Prediction System," International Journal of Engineering Research & Technology (IJERT), Sept,2015.

[6]. S. G, "Credit Risk Analysis and Prediction," International Journal of Engineering and Technology (IJET) , 2014.

[7]. E. I.Altman, "Predicting performance in the savings and loan association industry," Journal of Monetary Economics, oct,1977.

[8]. M. R. Islam, "A DATA MINING APPROACH TO PREDICT," International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5, March,2015.

[9]. D. K. Kavitha, "Classifying Data and Predicting Risk towards Multi -," International Journal of Advance Research in, Feb,2016.

Cite this article as :

Gopal Choudhary, Yash Garud, Akshil Shetty, Rumit Kadakia, Sonali Borase, "Loan Default Identification and its Effect", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 865-868, March-April 2019. Available at doi :
https://doi.org/10.32628/CSEIT1952198
Journal URL : http://ijsrcseit.com/CSEIT1952198