RESEARCH ARTICLE OPEN ACCESS

Credit risk analysis using regression and Classification techiniques

G. CHANDUNI [1], Dr. E. SREEDEVI [2]
MCA VIth semester, Asst, Professor Dept. Master of Computer Applications
Sree Vidyanikethan Institute of Management
A.Rangampet, Tirupati - India

ABSTRACT
Banking industry has the key activity of loaning money to people who are in need of cash. Once a monetary institute lends cash to a client, they're taking some reasonably risk. So, before loaning, monetary institutes check whether or not they would be paid back the loan by customer in the future. Considering the factors like current financial gain and expenditure of the client, By these we can able to analysis whether or not the client is eligible to take up the loan or not. This type of research is manual and long. So, it required some automation. Here, we wish to assist monetary corporations, like banks, NBFS, lenders, and so on. We'll build different types of classifiers and regression algorithms to predict to whom monetary institutes ought to provide loans or credit.
Here we are using different kinds of regression and classification algorithms to predict customer eligibility.
Then the next question that strikes the mind is "what's the output of our algorithms?" Our algorithm can generate probability, these can indicate the possibilities of borrowers defaulting. Defaulting means that borrowers cannot repay their loan during a certain quantity of your time. Here, chance indicates the possibilities of a client may or may not paying their loan EMI on time, leading to default. So, the next chance to indicate the probability of the client that the they will repay the loan or not.
Keywords: Preprocessing Data, Selection Process, classification and regression Algorithms, Cross Validation, HyperParameter Tuning, Probability.

I. INTRODUCTION

Banking industry has the foremost activity of lending money to those that area unit in would like of money. so as to payback the principle borrowed from the investor bank collects the interest created by the principle borrowers. Credit risk analysis is changing into a vital field in money risk management.
As credit risk prediction plays a very important role inside the banking sector and it's a really important and largest challenge faced by all the banks, accuracy plays a really necessary role in classification of credit info to avoid the loss of Banks. It increase the defaulter's rate within the credit risk information set that isn't reliable provides motivation towards this sector.
Yet, so far, several lenders are slow to fully utilize the prediction of digitizing risk. this is often despite a recent report from McKinsey showing that machine learning could scale back credit losses by up to ten per cent, with over half risk managers expecting credit calling times to fall by twenty five to fifty percent.
Several credit risk analysis techniques area unit used for the analysis of credit risk of the client dataset. The analysis of the credit risk datasets results in the choice to issue the loan of the client or reject the appliance of the client, that the troublesome task that involves the deep analysis of the client credit dataset or the information provided by the client.
The industry evaluates the accuracy of the datasets so as to classify the loan candidates into sensible and unhealthy categories. The candidates the unit of measurement within the sensible categories have the high likelihood of returning the money to the bank. The candidates that area unit within the unhealthy categories have the low likelihood of returning of the money to the bank therefore, they are the defaulters of the loans.
To reduced the loan defaulter's rate within the credit
Risk dataset differing kinds of regression and classifier techniques area unit used. some day large losses are usually reduced even with a little improvement within the accuracy of credit analysis. the advantages of the reliable credit risk dataset is it reduces the value of credit marking, sensible higher cognitive process in terribly less time and avoid less risk associates with art collection.

II. LITERATURE SURVEY

Numerous literatures relating offer the likelihood of a client that has been revealed already and are out there for public usage. A comprehensive understanding of credit risk analysis are often useful to the bank and cut back the loss.

- Shishi Dahita, N.P, Singh during this paper hybrid approach is employed to reinforce the classification accuracy for higher credit analysis of client loans. 2 benchmarked of MLP Neural network Technique with FS and cloth for higher classification accuracy so up loan granting. during this paper ANN is ar introduced with a special stress on Multilayer preception design this is often followed by an outline of technique|the tactic|the strategy} used for ensemble of classifiers 'with a stress on cloth method and on FS.
- Bhuvaneswari, This study deals with the analysis of an information set comprising of sumptuous vehicle credit portfolios characterised by relevant variables. It aims at assessing the danger related to these portfolios and at last presents a prophetic model that highlights the vital variables and depicts the mix of these variables that classify a consumer below defaulter or no defaulter. The study starts with the employment of standard applied math techniques and after presents machine learning approach exploitation 3 totally different call tree classifiers.
- Peter Martey Add, Gallus gallus Guegan, and Bertrand Hassani during this paper, {we will|we'll|we ar going to} specialise in the algorithms that are accustomed create these choices. Algorithms ar employed in {different|totally totally different|completely different} domains with different objectives. as an example, they're employed in enterprises to recruit persons appropriate for the profile projected. Algorithms will alter the method, create it faster and a lot of fluid, etc. yet,

III. SYSTEM ANALYSIS

EXISTING SYSTEM

- In the existing system they're victimization the conventional Classifying Algorithms to predict Credit Repaying nature of a client.
- The strength of current ancient approaches is predicated on formalized hypothesis that aren't capturing the fraud behavior of the client.
- The distortions within the existing model isn't simply specifiable.
- This model remains faraway from attaining mature levels each at the method and at the credit granting , watching and management method.

DRAWBACKS

- In the existing system they are using the normal Classifying Algorithms to predict Credit Repaying nature of a client.
- Not capturing the fraud behavior of the Customer.

PROPOSED SYSTEM

The aim of the proposed system is detecting using regression and classifiers of supervised learning. In this proposed model we are building the models which will accurately predit the risk factor using the boosting algorithms using mechine learning. In these paper can give probability of each customer by applying machine learning algorithms at least more than
one algorithm.Based on accuracy, this solution will select the best classifier (such as Gradient and Ada boosting algorithms). The algorithms which we are using give more acuurate result when compare to other algorithms.

ADVANTAGES

The system is effective in design and to implement. There are some expected advantages of proposed system. It has following features :

- Lend to right type of customers.
- Monitor collections.
- Predict and reduce delinquencies.
- Reduce NPA and increase profitability.

IV. IMPLEMENTATION

MODULES

User: I these module contains all the details of the client like id, name, income, loans, debit ration etc…

System: System can take the dataset, system can perform preprocessing the data, find client probability.

Proposed Algorithm Steps:

1. *Calculate the average of the target label
2. Calculate the residuals
3. Construct a decision tree
4. Predict the target label using all types of trees with in the ensemble
5. Calculate new residuals

6. Repeat step three and five until the number of iterations matches the number specified by the hyper parameter
7. Once trained, use all the trees in the ensemble to make a final prediction as to the value of the target variable.

V. RESULT AND DISCUSSION

To find the probability of the each customer, we implement it in python with anaconda- windows-x64. The experiments are executed on HP Think Centre M8400 (Intel (R) Core(TM) i5-2770 CPU @ 3.40GHz (8 CPUs)

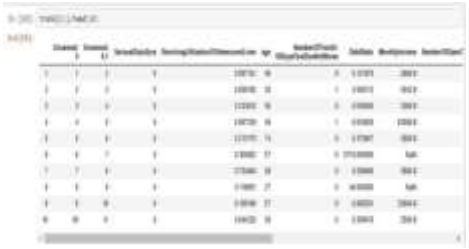

FIG 1: Representation of rows and colus

In these figure represents all customer details
Are represented in the form of rows and columns.

FIG 2: Statistical Representation Of data

In above figure represents all the customer details in the form of statistical representation.

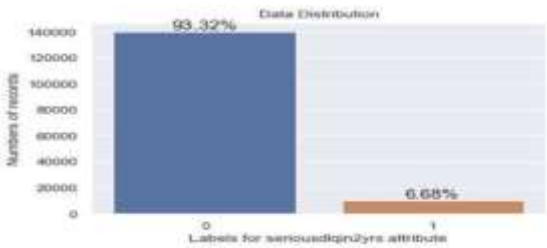

FIG 3: Statistical representation Of Targeted Data

In below figure it represents the targeted data in statistical format
The number of records are 140000

FIG 4: Classification And regression report and Acu-roc result

The classification and regression report and Acu-roc result is 0.861

FIG 5: Cross Validation Score

In above figure represents to cross validation scores all algorithms

Random Forest Classifier=0.78

Logistic Regression=0.69

KNeighbours Classifier=0.57

Ada boost Classifier=0.85

Gradient boosting Classifier=0.86

FIG 6: Hyperparameter Tuning

The hyperparameter tuning of the ada boosting algorithm is 0.85

ID of the Customer	probability to re-pay credit
0	21%
1	20%
2	17%
3	22%
4	23%
5	18%
6	18%
7	18%
8	18%
9	41%
10	17%
11	17%
12	17%
13	20%
14	20%
15	17%

FIG 7: Probability Of the customer

Finally we find the probability of the customer.

VI. CONCLUSION

Credit card risk analysis is one of the most important in now a days , the bank has to know who are the customers can really repay amount back before giving money to the customers . By this the bank can reduce the financial loss and increase its revenue. Here our project can give how much probability can the customer can have to repay the money

In this paper we are using different types of classifiers and regression algorithms and ensembling techniques which might increase the accuracy of client loan. The percect assessment of client credit risk is of outermost importance for lending organizations. Increasing the demand for client credit has led to the competition in credit industry. For future development we can able to use more effective algorithms and techniques to get more accuracy which will be useful in financial field.

Gradient Boosting Algorithm will gives more accurate 0.86

REFERENCES

[1] Credit Risk Analysis by using Machine Learning Classifier by TN Pandey - International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS- 2017).

[2] Dhaiya, S., Singh, N.P., 'Impact of Bagging on MLP classifier', International Conferences On Computing For Sustainable Global Development, pp. 3794-3799. 2016.

[3] Danenas,P., Grasva,G., 'Selection Of Support Vector Machine Based Classifier For Credit Risk', Expert System With Application , Vol. 42, pp. 3194-3204, 2015.

[4] Pandey, T.N., Jagadev, A.K., Choudhury, D. and Dehuri, S.,'Machine learning-based classifiers ensemble for credit risk assessment',Int. J. Electronic Finance, Vol. 7(3/4), pp.227–249, 2013.

[5] Chorowski, J., Wang, J., Zurada, M.J., 'Review and comparison of SVM and ELM based classifiers', Neurocomputing, Vol. 128, pp. 506-516,2014.

[6] Bask, A., Merisalo-Ratanen, H., Tinnila, M. and Lauraeus, T., 'Towards e-banking: the evolution of business models in financial services',International Journal of Electronic Finance, Vol. 5(4), pp. 333–356,2011.

[7] Bekhet, H.A., Al-alak, B.A., 'Measuring e-statement quality impact on customer satisfaction and loyalty', International Journal of Electronic Finance, Vol. 5(4), pp.299– 315, 2011.