TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 2798 - 2805

Evaluation of Machine Learning Models for Credit Scoring

[1]Ramesh Cheripelli, [2]Kovvuri Ramya Sri
[1] G.Narayanamma Institute of Technology & Science,[2] G.Narayanamma Institute of Technology & Science

Article Info
Volume 82
Page Number: 2798 - 2805
Publication Issue:
January-February 2020

Article History
***Article** Received: 14 March 2019*
***Revised**: 27 May 2019*
***Accepted**: 16 October 2019*
***Publication**: 18 January 2020*

Abstract:
One of the most important application of machine learning is credit scoring.It is uber important for banks and financial institutions to develop credit card or loan services to compete with foreign capital and obtain profits, On the other side it is urgent to improve the ability to control credit risks. Customers are the valuable assets of any bank. The payment of timely bills is important for the running of banks. But if the customers do not pay on time, it may incur huge loss to any financial organization. In this paper we try to build several models which will predict the credit score of customers. Credit score is calculated on banking and finance datasets. To show the relation between attributes, the correlation matrix is generated. In the experimental part, the graphs are generated, which shows the contrast for better analysis. This paper predicts and proposes the factors or attributes which optimize the profits of any banking organization. The best model will be selected based on the accuracy, sensitivity and specificity values obtained.

Keywords:*Credit Scoring, Machine Learning, Weight Of evidence analysis, Crisp DM Framework.*

I. INTRODUCTION

The objective and purpose of this paper is to propose a methodology,architecture and algorithms of machine learning, to findlearning patterns, relationships, and interaction and effective analysis for a complex system assessment involving many datasets with lakhs of records and variables that can be obtained from an online database . Here we have obtained from bank and credit bureau.

Analysis may take between days or months with huge team members to find appropriate solutions for wide real problems or events that occur in financial institutions and to discuss concepts andimprovement with collaborations. Modeling an interactive system that involves multi variable data sets can lead to many obstacles and challenges.

This paper concentrates on proposing an efficient and interactive Machine Learning based system architecture to understand the behavior of different customers in banking and finance sector by classifying them into buckets. Our models integrate techniques from statistics and machine learning models that include logistic regression and random forests. Our algorithms also identify the behavior of the components at an absolute nano level and can help us in modelling the behaviors of customers involved in learning. Usage of machine learning has made the problem solving easier.

In the era of emerging technologies, advancements in machine learning techniques have created a great impact in our lives. These advancements made us to develop new way of approaches to solve problems in different areas such as cancer diagnosis, predictive forecasting, speech recognition etc. Various machine learning techniques like Reinforcement learning and Supervised learning, Unsupervised learning have been extensively used to transform a computer into an intelligent machine to solve complex and challenging problems in real world.

The advancements in our technologies in terms of computational power and acquiring large data made machine learning models to grow complex day-by-day. So building a good model is not sufficient since evaluating a model is equally important. So a good metric is needed to estimate the prediction error. Although plethora of metrics are available in the community of machine learning, confusion arises very often in choosing the right metric. However there is no common theory in choosing a metric to evaluate a model.

This paper describes popular supervised learning algorithms along with the mathematics behind them. We analyze and evaluate the important evaluation metrics in classification and regression algorithms. After assessing the discussed evaluation metrics, a conclusion is made on choosing the right evaluation metric for the given problem. The integration of these mathematical models with machine learning makes the job easier.

We propose the usage of three algorithms to explore and study customer behaviors which will help us in finding the credit score in this paper. The results obtained by the models of each algorithm will then be compared based on their sensitivity, specificity and accuracy values. Then the final selected model which has more stable values will be selected.

Usage of machine learning for effective assessment and R programming language for visualization and analysis of large amount of data increases in Machine Learning(ML) is a class of algorithms provides various techniques to analyse the huge data in a better possible way. ML provides classification techniques, clustering mechanisms and Recommender systems to name a few.

The work provides the key benefits of such integration and future scope of the integration along with possible research constraints in the reality. We believe the work gives a platform to researchers so as to extract the future scope of the integration and difficulties faced in the process.

This paper is an extension of our work. We mentioned details regarding the variables, records and Machine Learning architecture for intensive computing and very efficient analysis and assessment.

Our paper is arranged as follows: in Section II we briefly discuss our related work on Machine Learning for credit scoring and data analytics. In Section III we see the variables in the datasets.In section IV Exploratory data analysis for calculating predictor variables are discussed. We discussed the predictive modelling. In Section VI we see the application score card in Section V. In Section VII we discuss the losses avoided and profits gained. Section VIII concludes with directions for future work.

II. RELATED WORKS

Due to the increasing popularity and usage of credit cards in banking and finance sector. As customers provide their information, large amount of personal information can be inferred from the customers' whereabouts.

In [1] We have used logistic regression, decision trees and random forests and decided that random forest has given better results in accuracy, specificity and sensitivity. further changes have been made in this paper to improve the results

In [2] The authors have used Artificial Immune System for Credit Scoring and classification. They have also used AIRS, CLONALG and Immunos. This algorithm is applicable only for small datasets which cannot be used in practical purposes. The accuracy in huge datasets is very low. In [3] the authors have proposed usage of data mining instead of machine learning model to predict the customer defaultness. Their model is built using data mining, hence they can detect the customers with high propensity to be a defaulter, but not necessarily providing the reason of defaultness. In [4] in their examination on data set has considered active

January - February 2020
ISSN: 0193 - 4120 Page No. 2798 - 2805

defaulters in the banking and finance industry by applying various methods of data mining. As they have not used machine learning there is more involvement of human and can be difficult to scale up the project or extend the project. In [5] the authors have used data mining again for prediction, because of which we can predict the patterns of existing customers and not new customers. We can make rules but cannot make the machine understand the rules and models for automatic prediction.
In [6] we have used the same algorithms but changes performed on the models has yielded much better and stable results. In [7] authors have used neural network and multi-layer perceptron. In [8] authors have used neural networks and evolutionary techniques, but the paper failed to mention about the credit scoring strategy. In [9] the authors have used as neural networks, multivariate discriminant analysis, and logistic regression analysis, CART, C5.0. Which failed to work for banks with huge datasets. In [10] the authors have used Confusion Matrix, Credit Risk Evaluation, Support vector machine and Extreme learning machine but the paper has been published for calculating the banks with the largest number of customers with good credit scores.

III. DATASETS AND DATA UNDERSTANDING

2 data sets are required for the analysis.Datasets are provided by the bank.

Table 1. Demographic Data.

Variables	Description
Application.ID	Unique ID of the customer
Age	Age of customers
Gender	Gender of customers
Marital.Status	Marital status of customers (at the time of application)
No.of.dependents	No. of children's of customer
Income	Income of customer
Education	Education of customer
Profession	Profession of customers
Type of residence	Type of residence of customers
No of months in current residence	No of months in current residence of customers
No of months current any Company	No of months in current company of customers
Performance Tag	Status of customer performance ("1"

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 2798 - 2805

Table 2. Credit Bureau Data.

Variables	
Application ID	Customer application ID
No.of.times.90.DPD.or.worse.in.last.6.months	Number of times customers has not paid dues since 90days in last 6 months
No.of.times.60.DPD.or.worse.in.last.6.months	Number of times customers has not paid dues since 60 days last 6 months
No.of.times.30.DPD.or.worse.in.last.6.months	Number of times customers has not paid dues since 30 days last 6 months
No.of.times.90.DPD.or.worse.in.last.12.months	Number of times customers has not paid dues since 90 days last 12 months
No.of.times.60.DPD.or.worse.in.last.12.months	Number of times customers has not paid dues since 60 days last 12 months
No.of.times.30.DPD.or.worse.in.last.12.months	Number of times customers has not paid dues since 30 days last 12 months
Average.CC.Utilization.in.last.12.months	. Average utilization of credit card by customers
No.of.trades.opened.in.last.6.months	No of times the customers has done the trades in last 6 months
No.of.trades.opened.in.last.12.months	No of times the customers has done the trades in last 12 months
No.of.PL.trades.opened.in.last.6.months	No of PL trades in last 6 month of customers
No.of.PL.trades.opened.in.last.12.months	No of PL trades in last 12 month of customers
No.of.Inquiries.in.last.6.months.	No of times the customer has inquired in last 6 months
No.of.Inquiries.in.last.12.months	No of times the customer has inquired in last 12 months
Presence.of.open.home.loan.	Is the customers has home loan (1 represents "Yes")
Outstanding.Balance	Outstanding balance of customers
Total.No.of..trades	Number of times the customer has done total trades
Presence.of.open.auto.loan.	Is the customer has auto loan (1 represents "Yes")
Performance.Tag	Status of customer performance (" 1 represents "Default")

I. DATA CLEANING & PREPARATION

First step is to load the data from the both CSV files: Credit Bureau Data and Demographic Data into two different data frames.

A. *Preliminary Checks*

- We will then perform some necessary preliminary data quality checks on each of these data frames.
- Check the structure of data. This will show us the number and type of these variables (continuous or discrete).
- Check the summary of data. This will tell us if any variables have NA values.
- We will also check the number of rows and columns in each data frame.
- We will check if output variable Performance tag is same for both Demographic and Credit Bureau Data.
- We will then check if the number of rows are same between both data frames.
- We will check the proportion of the Default vs Non-Default customers. In the current dataset we have an issue of class imbalance. The number of default customers is only 4% of the total customers.

Note: This may pose some challenges in building models like decision trees. As decision trees have not performed well on the same data set based on the previous research. We have not performed it again.

B. *Duplicate Detection*

We will then verify if there are any duplicate ids (Application.Id) in both demographic and Credit Bureau Data. If there any duplicates we will first remove them from each of the data frames.

Merging

Then we will merge both Demographic and Credit Bureau Data using Application.Id as the key and create a “Master Data Frame”. We will now

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 2798 - 2805

carry out further data quality checks on this master data frame.

C. Missing Value Treatment

On the Master data frame we will then check if dependent variable (Performance.Tag) has any missing values. As per the project guideline if there are missing values, we will treat such records as data for applicants that have been rejected. So, we will remove these records from our dataset and use it later for our model building.

Then we will check if there are any columns which are entirely blank. If there are no such columns, we will then check how many variables have missing data. As per the guideline we need to do WOE transformation. Since WOE transformation takes care of missing value treatment, we will not remove such rows from our dataset.

D. Outlier Detection

We will then analyze each variable independently. For continuous variableswe will check if there are any outliers (like negative values or large values). If there are such values we will try to do outlier treatment for each of the variables. We will identify the IQR and try to cap the values.

For discrete variables we will check the number of distinct values in each variable and see if there are any spelling mistakes or inconsistencies.

E. Derived Metrics

We will analyze if there is a possibility of calculating some derived metrics which might help in our data analysis.

IV. EXPLORATORY DATA ANALYSIS

In this step we will do exploratory data analysis of the demographic and credit bureau data which will help us in identifying the important variables for prediction of defaulters.

$$WeightofEvidence = ln\left(\frac{DistributionGood_i}{DistributionBad_i}\right)$$

Based on the IV (information value) of each variable we predict the importance of that variable)

$$IV = \sum\left(DistributionGood_i - DistributionBad_i\right) \times WOE_i$$

Table 3. Information value and their predictive power.

Information Value	Predictive Power
<0.02	Not useful for prediction value
0.02 to 0.1	Weak Predictor value
0.1 to 0.3	Medium Predictor value
0.3 to 0.5	Strong Predictor value
>0.5	Suspicious or too good to be true value

From the table 4 we will observe the influence of each of the variables on the dependent variable i.e., default rate is very high for some categories in the variable. With this manual analysis of finding the most important variables we listed out some variables which are important. We will also draw correlation plots of the continuous data in the demographic and Credit bureau data

Table 4. Information value of predictive variables

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 2798 - 2805

Predictor Variable	IV
Avgas.CC.Utilization.in.last.12.months	0.3118158
Number.of.trades.opened.in.last.12.months	0.2992422
Number.of.PL.trades.opened.in.last.12.months	0.2976330
Number.of.Inquiries.in.last.12.months..excluding.home...auto.loan	0.2965392
Outstanding.Balance	0.2469674
Number.of.times.30.DPD.or.worse.in.last.6.months	0.2420549
Total.Number.of.Trades	0.2378859
Number.of.PL.trades.opened.in.last.6.months	0.2203559
Number.of.times.90.DPD.or.worse.in.last.12.months	0.2142245
Number.of.times.60.DPD.or.worse.in.last.6.months	0.2062044
Number.of.Inquiries.in.last.6.months..excluding.home...auto.loans	0.2052807
Number.of.times.30.DPD.or.worse.in.last.12.months	0.1987550
Number.of.trades.opened.in.last.6.months	0.1864486
Number.of.times.60.DPD.or.worse.in.last.12.months	0.1858931
Number.of.times.90.DPD.or.worse.in.last.6.months	0.1603274

V. WOE TRANSFORMATION

In credit scoring, weight of Evidence & equivalently Information Value are often used to compare predictive power among variables. weight of Evidence is also helpful in treatment of missing values. So, we will use WOE to impute missing values from the data.

Using "Information" Package we will calculate the IV (information value), WOE on the master

data frame. This will give WOE and IV for each variable.

1. Then using a utility function, we will transform each column (by identifying the bins) and substituting the actual value with the WOE value (for the bin).
2. We will then store this master data in a separate file. Let us call it **"Master WOE Frame".**

3.We will also generate some graphs based on the WOE data.

Note: We need to remove the Application.Id column for calculating the WOE matrix as this column is used only to identify each row.

VI. MODEL BUILDING

The models were developed after the data has been classified into different buckets.

Based on the performance of logistic regression and random forest, Random forest model has got better values and has detected 75% of defaulters correctly and 80% of defaulters in top 4 deciles

Table 5. Values obtained for different buckets

bucket	total	Total Bad	Cum- Bad	Gain	Lift
1	6951	1739	1739	59.2	5.9
2	6950	272	2011	68.4	3.4
3	6950	195	2206	75.1	2.5
4	6950	169	2375	80.8	2.0
5	6950	155	2530	86.1	1.7
6	6950	122	2652	90.3	1.5
7	6950	95	2747	93.5	1.3
8	6950	85	2832	96.4	1.2
9	6950	58	2890	98.4	1.1
10	6950	48	2938	100.0	1.0

VII. MODEL EVALUATION

A. Application Scorecard

Application Scorecard (master population): Score varies between 200 to 530; Cut-off score – 338.

Which means Cut-off: 338 is the baseline for providing credit card to the customers.

January - February 2020
ISSN: 0193 - 4120 Page No. 2798 - 2805

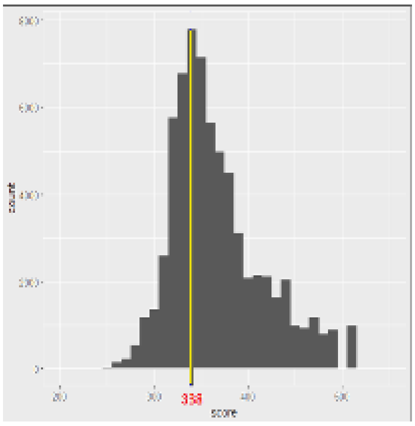


Fig. 1. Graph representing baseline cutoff.

Application Scorecard (Rejected population): 70% of defaulters correctly identified.
Average score of rejected population is less than the average score of approved* population
Total rejected applications by bank: 1423
Identified correctly at cut-off score by model: 1006

*Approved population (master data) is a population for which the application is accepted by bank

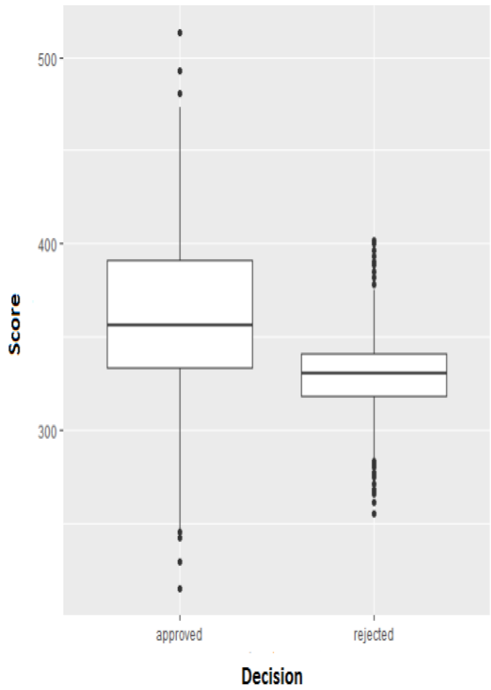


Fig. 2. Box plot showing approved and rejected candidates.

B. Financial Benefits

Credit loss* : Reduced credit loss from 4% customers to 1% customers.

Because

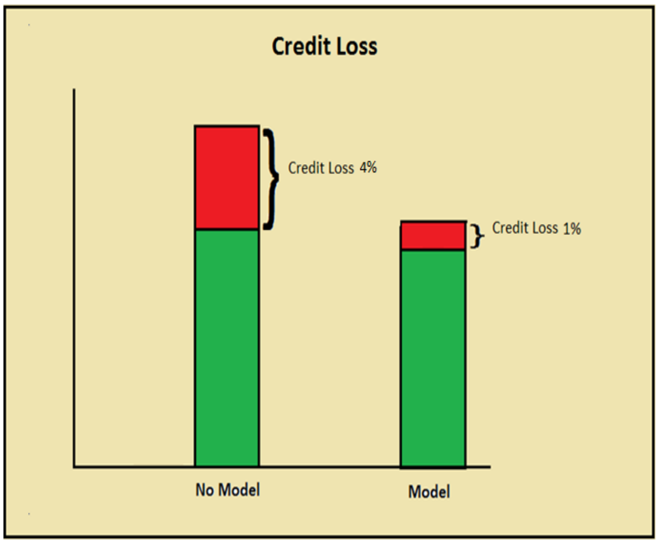
- Credit loss no model = **4%**
- Credit loss with model = **1%**

Credit Loss Saved: 3%
* The loss occurred from the bad customers

Table 6. Predicted and Actual defaulters using confusion matrix

Confusion Matrix		Actual Defaults	
		Good Customers(0)	Bad Customers(1)
Predicted Defaults	Good Customers(0)	47938	732
	Bad Customers (1)	18625	2206

Fig. 3. Credit loss without and with model.



Revenue Loss*: Reducing 30% revenue (Auto-approval)
Because,

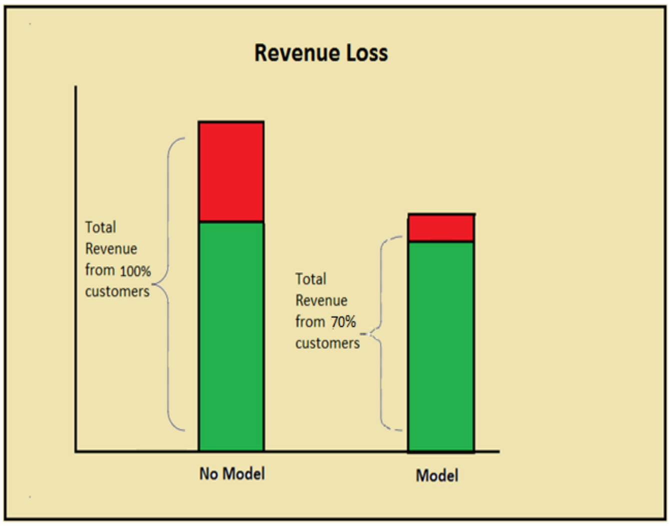
- Revenue no model = **100%**
- Revenue with model = **70%**

Revenue Loss : 30%

TEST Engineering & Management
January - February 2020
ISSN: 0193 - 4120 Page No. 2798 - 2805

*The revenue loss is occurred by wrongly identified "bad" to the good customers

Fig. 4. Revenue loss with and without model



VIII. CONCLUSION

The work provides the key benefits of such integration and future scope of the integration along with possible research constraints in the reality. We believe the work gives a platform to researchers so as to extract the future scope of the integration and difficulties faced in the process, and the factors that involve in prediction. In our future work, we will attempt to build models using support vector machines and check if the results obtained are better or worse.

IX. REFERENCES

[1] K Ramya Sri and Ch Ramesh "Credit Risk Valuation Using Machine Learning Algorithm" Springer Learning and Analytics in Intelligent Systems series, vol. 4, chapter 74, June 2019 https://link.springer.com/chapter/10.1007%2F978-3-030-24318-0_74

[2] Antariksha Bhaduri, "Credit Scoring using Artificial Immune System Algorithms: A Comparative Study.

[3] Wei Li, Jibiao Liao, "An Empirical Study on Credit Scoring Model for Credit Card by using Data Mining Technology".

[4] Aihua Li, Wenbin Li, Yong Shi, "Credit Card Management

[5] Yi Peng, Gang Kou "A Comparative Study of Classification Methods in Financial Risk Detection"

[6] Trilok Chand Sharma and Manoj Jain, "WEKA approach for comparative study of classification algorithm", International journal of advanced research in computer and communication engineering, vol.2, issue 4, April 2013.

[7] Antariksha Bhaduri, "Credit Scoring using Artificial Immune System Algorithms: A Comparative Study.

[8] Su-Lin Pang, Yan-Ming Wang, Yuan-Huai Ba, "Credit Scoring Model Based On Neural Network"

[9] M. B. Yobas, J. N. Crook, P. Ross,"Credit scoring using neural and evolutionary techniques"

[10]Wei Xu, Shenghu Zhou, Dongmei Duan, Yanhui Chen," A Support Vector Machine Based Method For Credit Risk Assessment"

[11]Jongsik Yoon, Young S. Kwon, Tae Hyup Roh," Performance Improvement of Bankruptcy Prediction using Credit Card Sales Information of Small & Micro Business"

[12]Hongming Zhou, Yuan Lan, Yeng Chai Soh, Guang-Bin Huang and Rui Zhang," Credit Risk Evaluation with Extreme Learning Machine"

[13]Anne Kraus: Recent Methods from Statistics and Machine Learning for Credit Scoring

[14]Mehnaz Khan and S.M. K. Quadri, "Effects of using filter based feature selection on the performance of machine learners using different datasets", Bharati vidyapeeth's institute of computer applications and management's International Journal of Information Technology, vol.5 July-December 2013.