

# Bank Credit Risk Analyses Using Machine Learning Algorithms

ISSN (e) 2520-7393 ISSN (p) 2521-5027 Received on 14<sup>th</sup> Sept, 2019 Revised on 28<sup>th</sup> Sept, 2019 www.estirj.com

Summair.Alam<sup>1</sup>,Dr.Noor-ul-.Zaman<sup>2</sup>, Dr. Shahnawaz Talpur<sup>3</sup>,Rajesh.Kumar<sup>4</sup>, Komal<sup>5</sup>

<sup>1,2,3</sup>Institute of Information & Communication Technologies (IICT)Mehran University of Engineering and Technology <sup>4</sup>Department of Computing Hamdard University UET Karachi

\_\_\_\_\_

Abstract: Earlier computers was just categorized as a want of an individual but now it becomes a need of an individual. Machine learning serves as a major part in field of computer, machine is not able to think over different situations but it can draw different sort of relationships between different features and characteristics. The major part of our life is to avoid fraudulent activities but till now we are unable to control over it. Loan business is one of the major businesses of commercial banks. Fraudulent activities can be handle through embedding machine learning algorithms in our daily life. In this research we have used supervised machine learning and for that we have to provide labeled data to the machine learning algorithm, and we have used SVM classifier algorithm. In this research we have analyze the credit data in different perspectives.

Keywords: Credit risk, MDA, support Vector Machine (SVM), Loan Evaluation, analyses

# **1. Introduction**

t this moment world has becomes computerized A almost. All the work done through computers, from the small game programming to the heavy transactions, all things are possible just because of computers. Computers plays vital role in every aspect of life and earlier it was just categorized as a want of an individual but now it becomes a need of an individual. Machine learning serves as a major part in field of computer. Now computer becomes essential part of our life so we should also embed machine learning in our daily life work usage. Men is powerful as a thinker but machine is more powerful than men in calculation. Machine is not able to think over different situations but it can draw different sort of relationships between different features and characteristics. The major part of our life is to avoid fraudulent but till now men is unable to control over it. But it can be handle through embedding machine learning algorithms in our daily life so that it can be easily predict earlier about any kind of fraudulent occurrence in future before trusting to anyone. Machine learning is of two types, first one is supervised machine learning and another one is unsupervised machine learning. In supervised machine learning there should be labeled data available as a training data set for machine learning algorithm but in unsupervised machine learning there should not be any sort of labeled data available, machine have to learn without any training data set. But in this project we will use supervised machine learning and for that we have to provide labeled data to the machine learning algorithm. Borrowers are one of the major commercial banks. Today, as bad debt becomes a financial enemy of the financial industry in every country, credit risk is closely related to the financial community. [2].

# 2. Problem Statement

In our world, almost all the valuable transactions are done through banks and our banking system are now introducing different kinds of credit or loan offers for their customers. As these kinds of offers now came in our system therefore the rate of fraudulent also increases in terms of these offers. With the development of the economy, the real estate market is growing rapidly. As a result, domestic loans are increasing, which leads to some risks together. Since breach of contract may lead to financial risks for the banking system and, ultimately, to the national economy, an assessment model is needed to eliminate potential risks for any sort of credit [3]. So how to handle such kind of situations? Many banking systems and other systems that gives such kind of offers (credit or loan), are in huge trouble to handle fraudulent. For an individual, it is not an easy task to predict that the customer will return the loan + interest on time or after due date or do any fraudulent after taking loan. In recent years, a large number of commercial banks, non-performing loans (NPL) have become one of the biggest obstacles to the prevention of renewal and development in Chinese commercial banks. [1]. Through the face of the customer or any other sort of data, men cannot be able to draw any relationship or do any prediction so therefore it is essential to use loan/credit data analysis to handle the data and use machine learning algorithms to predict about the customers.

# 3. Aims and Objective

The main objectives of this research is to handle huge amount of credit data, analyze the credit data in different perspectives, represent data in graphical manner for more understanding, to compare credit data, apply machine learning algorithm on credit data and test and verify machine learning algorithms on credit data

#### 4. Related Work

Evaluation of residential loan by combining RVM and Logistic Regression [3]. It's about only residential loan RVM and Logistic Regression combine in it. Classification methods of credit rating- A comparative Analysis on SVM and MDA[1] SVM 70% accuracy result. MDA 50% accuracy result. Draws a CRM(customer relationship management).Fuzzy Pattern Recognition Model of Bank Loan Risk and its Application[6]. Data is divided into five categories in it. No machine learning algorithm used. In [8]adopted the approach of Logistic Regression and Classification and Regression Trees (CART) with techniques such as under sampling, Prior Probabilities, Loss Matrix and Matrix Weighing to deal with imbalanced data. In [9] proposed a data mining approach for credit cardholders' behavior analysis. And we compared the performance of general MCLP classification method and MCLP based on PCA dimension reduction method in the end. The performance shows the potential application ability of this approach for credit cardholders' behavior analysis. In [10] this paper they proposed propose an optimal credit scoring model to reassess the default risk of credit card holders for credit card issuing banks in Taiwan. This paper adopted four credit scoring models which are the linear discriminate analysis, decision tree, backpropagation neural network. In [11] suggests the decisionmaking personnel to establish a decision-making support system to assist their judgment by using the classification model.

In [12] It discussed how to employ the incremental learning method, on the basis of keeping the former study results, only to study the newly-rising data, which formed a continuous learning process and offer technology supports to settle the problem of the bank customer credit evaluation on real time

## 5. Methodology

The data of credit or loan will be collected from different sources where it is possible to be collect and then to check data in terms of compatibility. If the data is not compatible then work has to be done upon that data to make that data compatible and do some analysis on that data by drawing different sort graphs based on data so that an individual can understand the whole data easily. After doing analysis and checking the compatibility now the comparison can be done on different perspectives of data. When the data is said to be fit for doing further analysis then machine learning algorithm will be apply on it. As we are working on supervised machine learning so therefore data plays a vital role in it. After that the machine learning algorithm has to be tested and verified. Thefigure.1 illustrate the whole scenario.



Figure 1: the figure illustrate the whole scenario



Figure.2. flow chart of whole process results

### 6. Proposed Solution

Figure.2 Shows the missing values in dataset in terms of percentage.

	df_bool	df_amt	missing_ratio_percent
<pre>paid_off_time</pre>	True	100	20.0
past_due_days	True	300	60.0
['PAIDOFF' 'CO	LLECTION'	'COLLEC	TION_PAIDOFF']

# 7 Results and Discussion

7.1Count of loan status:

Counting number of loan members according to their status that is

#### PAIDOFF

#### COLLECTION,

COLLECTION PAIDOFF.



Figure 3 Count of Loan Status

7.2 Amount Of loan Figure 4 Count of Loan Status 7.3 Count Terms Of Loan Status



For how much terms/period loan taken showing through graph according to category

PAIDOFF COLLECTION

# COLLECTION PAIDOFF.



Figure 5 Count of Loan Status

# 7.3 By Means Of Effective Date



Figure 6 effective date means by which date loan taken

# 7.4 Applying RFC

Applying random forest classifier for the classification of data set as well as prediction technique.

Accuracy rate is about 0.68 which means 68%..



7.5 Applying Svm Classifier With Linear Kernel

Applying SVM classifier with linear kernel technique. Result accuracy is about 0.66 that means 66%.



Figure 7 effective date means by which date loan taken

# 7.6 Applying SVM Classifier With Radial Basis Function

Applying svm classifier with RBF kernel. Result accuracy rate is 0.71 and it means 71%. Its result fluctuates and depends upon gamma value. Train set is 80% and test set is 20%.



Figure 9 effective date means by which date loan taken

#### **8** Conclusions

This paper proposed, to avoid fraudulent activates we have used supervised machine learning and for that we have to provide labeled data to the machine learning algorithm to avoid fraudulent activities. Also tested, analyze and verify machine learning algorithms on credit data. we have collected the loan data, checked data in terms of compatibility Analyze data by drawing different sort of graphs so that an individual can understand the whole data easily. Also we have used SVM support vector machine algorithm and Random forest classification.

# References

- Chun F. Hsu, "Classification Methods of Credit Rating A Comparative Analysis on SVM, MDA and RST" in Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference, Wuhan, China.
- [2] M.J. Pazzani, "Knowledge discovery from data?" in IEEE Intelligent Systems and their Applications (Volume: 15, Issue: 2, Mar/Apr 2000).
- [3] Qinrong Meng, "Evaluation of Residential Loan by Combining RVM and Logistic Regression" in Management and Service Science (MASS), 2011 International Conference, Wuhan, China.
- [4] Aslı Çaliş, Ahmet Boyaci, Kasım Baynal, "Data mining application in banking sector with clustering and classification methods" in Industrial Engineering and Operations Management (IEOM), 2015 International Conference, Dubai, United Arab Emirates.
- [5] Yu Zhang, Guang Yu, Yong-sheng Guan, Dong-hui Yang, "Feature selection of nonperforming loans in Chinese commercial banks" in Management Science & Engineering (ICMSE), 2014 International Conference, Helsinki, Finland.
- [6 Changling Yin," Fuzzy pattern recognition model of bank loan risk and its application" in Fuzzy Systems and Knowledge Discovery (FSKD), 2010 Seventh International Conference, Yantai, Shandong
- [7] Sonali Bakshi, Credit Card Fraud Detection, IEEE (2018)
- [8] Shiivong Birla, Kashish Kohli, Machine Learning on Imbalanced Data in Credit Risk IEEE 2016
- [9] Aihua Li ; Yong Shi ; Meihong Zhu , A Data Mining Approach to Classify Credit Cardholders' Behavior IEEE 2007
- [10] Hui-Chung Yeh ,Min-Li Yang . An Empirical Study of Credit Scoring Model for Credit Card.
- [11] Chun F. Hsu, "Classification Methods of Credit Rating A Comparative Analysis on SVM, MDA and RST" in Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference, Wuhan, China

# **About Authors**

**Summair Alam** received his B.E in software engineering degree from Mehran UET Jamshoro. His interest lies in Artificial intelligence and image processing.

**Dr. Noor ul Zaman** received his PhD in Data mining and his research interest lies in big data and Computer Vision.

**Dr. Shanwaz Talpur received** his PhD in Computer Architecture from Beijing Institute of Technology, China, He is an active researcher and teacher since more than past 15 years. He is among first few Higher Education Commission (HEC) approved Supervisors. He teaches undergraduate, masters and Ph.D. degree levels. Currently working as Professor and Co-Director IICT MUET Jamshoro.

**Engr.Rajesh Kumar** received his M.E in Computer Information Engineering from MUET Jamshoro. His research interest lies in Machine learning, deep learning and Artificial Intelligence .Currently Working as a Lecturer in Computing Dept. Hamdard University.

**Engr. Komal** received his M.E in Information Technology MUET Jamshoro. Her research interest lies in Image processing, deep learning and Artificial Intelligence .Currently working as a Lecturer in Computing Dept. Hamdard University.