Credit Scoring Techniques: A Survey

Thabiso Peter Mpofu[1], Macdonald Mukosera[2]

[1]M.Tech. Student, Department of Computer Science, School of Information Technology, Jawaharlal Nehru Technological University Hyderabad, India

[2]Jawaharlal Nehru Technological University Hyderabad, School of Information Technology, Kukatpally, Hyderabad 500085, India

Abstract: *Credit scoring is a numerical expression of the credit worthiness of an individual. A Value with a specific creditworthiness associated is assigned to an individual. Overall objective is to determine the creditworthiness of an individual. Ability of an individual to repay is determined in the credit scoring process. The credit scoring process looks at specific criteria such as income, credit history and many others. All this is done with the intent to reduce the overall default rate thereby decreasing the overall risk of financial institutions such as banks and micro lending institutions. Several credit scoring methodologies have been proposed and implemented and are varied from statistical based methods to Artificial Intelligence based techniques.*

Keywords: credit scoring, credit worthiness, default rate, risk, credit scoring techniques.

1. Introduction

Financial Institutions such as banks and micro lending institutions issue out loans to individuals and corporate institutions. The general public still forms the core of the financial institutions clients and thus the need for credit scoring. For the financial institutions to remain competitive and as go on as a going concern, the loans issued out have to be paid back. It is not everyone who borrows that has the capacity or ability to repay back the loans they have taken. There is therefore a certain risk associated in the issuance of loans by the financial institutions. Due diligence is therefore supposed to be observed by the financial institutions issuing out the loans. Credit scoring is one of the most successful applications and operations research techniques used in banking and finance, and is also one of the earliest financial risk management tools developed [1]. Credit scoring was developed by Fair and Isaac in the early 1960s and in simple terms corresponds to producing a score that can be used to classify customers into two separate groups: the credit-worthy or"good" group (likely to repay the loan), and the non credit-worthy or"bad" group (rejected due to its high probability of defaulting) [4]. A risk assessment and analysis is supposed to be carried out in the issuance of loans. The creditworthiness of the individual to whom the loan is being issued out to is supposed to be assessed and analysed by the issuance of a credit score. The credit score issued to an individual will determine on whether to issue out the loan to the person or not. Several credit scoring techniques have been proposed and implemented over the years. The credit scoring techniques vary from statistical based techniques to Artificial Intelligence based techniques. Statistical based techniques range from Linear Discriminant Analysis to Logistic regression. Artificial Intelligence (AI) based techniques have a learning and memory ability. AI techniques are able to learn from specific data and pick up trends. They are also able to remember what has been previously learnt. AI techniques have proven to be more flexible compared to statistical based techniques which operate on a rigid set of conditions. AI based techniques such as Artificial Neural Networks, Genetic Algorithm and Artificial Immune Systems will be looked at in this paper. With the advent of the global financial crisis which was largely as a result of defaulting loans or bad loans where people were unable to repay loans that they took, the area of credit scoring is therefore of the essence.

2. Factors considered in Credit Scoring

In the credit scoring process, there are certain and specific criteria that are looked at. The main determinants of whether a default will take place or not can be classified into the following four areas [3], [13].

1. Financial Indicators
2. Demographic Indicators
3. Employment Indicators
4. Behavioral Indicators

2.1 Financial Indicators
Financial indicators essentially indicate financial status or position of the loan applicant in repaying the loan. Cash inflows and outflows give a realistic position or potential maximum possible monthly payment the loan applicant can pay.

2.1.1 Total assets of borrower
The total assets of borrower reflect the borrower's repayment ability. If the borrower is hit with financial troubles, assets can be used to cover the loan taken.

2.1.2 Gross income of borrower
Cash inflows should be known in order to calculate the effective capacity of repayment

2.1.3 Gross income of household
If borrower is unable to repay, family member can assist in the repayment. The higher the household income is the better your chances of loan acceptance

2.1.4 Monthly costs of household
Cash outflows should be known in order to calculate the effective capacity of repayment

2.2 Demographic Indicators

These variables typically do not have the highest consideration in the loan consideration process.

2.2.1 Age of borrower
Age to determine the maturity levels

2.2.2 Sex of borrower
The gender of the applicant is hardly ever looked at anymore.

2.2.3 Marital status of borrower
Depending on the purpose of the loan marital status can also be considered. If married, it can be coupled with analysis of house hold income.

2.2.4 Number of dependants
The more kids, the more the load and the more likely to default.

2.2.5 Home status
Is the house rented or owned, if the house is owned it increases the likelihood of getting a loan relative to when house is rented

2.2.6 District of address
The area you stay also signifies your social status, the higher your social status the more likely you are to repay

2.3 Employment Indicators

Employment details are analysed in this segment

2.3.1 Type of employment
Are you self employed or employed somewhere. Such factors are such as the nature of work that you do are looked at**.**

2.3.2 Length of current employment
The more the years the more it signifies on your stability. More stable people are more likely to repay.

2.2.3 Number of employments over the last *x* years
Depending with the organizations worked for, if the change in employment moves up the ladder, the more likely you are to repay.

2.4 Behavioral Indicators

The financial behavioral patterns are looked at in this segment

2.4.1 Checking account (CA)
Does the applicant have an account with the financial institution issuing the loan or not? If a client has some history with a financial institution, then that bank can easily verify the behavioral activities of the loan applicant.

2.4.2 Average balance on CA
What is your average account balance over a period of time? If the average balance per month is higher than the monthly installment, the more likely for the loan to be approved.

2.4.3 Loans outstanding
Your current load of loans is looked at. If u have other loans that you are servicing it then decreases your chances or ability to pay back other loans

2.4.4 Loans defaulted or delinquent
Loans defaulted are loans taken but never fully paid back or never paid back at all. Your loan history will give an indication of the likelihood of repayment in the loan currently being applied for.

2.4.5 Number of payments per year
At what intervals do you make payments?

2.4.6 Collateral/ guarantee
If loan applicant has collateral, the more likely for the loan to be approved. The loan applicant will work diligently to repay for fear of losing the asset that has been put up as collateral.

3. Credit scoring Techniques

Credit Scoring techniques are divide into statistical and Artificial intelligence based methods.

3.1 Statistical based methods

3.1.1 Linear regression
Linear regression is a credit scoring technique concerned with describing the relationship between a response variable and one or more independent variables.[3]

3.1.2 Discriminant analysis
Discriminant analysis is a credit scoring technique developed to discriminate between two groups. It is widely agreed that the discriminant approach is still one of the most widely established techniques to classify customers as good credit or bad credit.[3]

3.1.3 Probit analysis
Probit analysis is a credit scoring technique that finds coefficient values, this in turn is the probability of a unit value of a dichotomous coefficient. A linear combination of the independent variables is transformed into its cumulative probability value from a normal distribution [3].

3.1.4 Decision trees
A classification tree is a non-parametric method to analyse dependent and/or categorical variables as a function of continuous explanatory variables (Breiman et al. 1984; Arminger et al, 1997). In a classification tree, a dichotomous tree is built by splitting the records at each node based on a function of a single input. The system considers all possible splits to find the best one, and the winning sub-tree is selected based on its overall error rate or lowest cost of misclassification (Zekic-Susac et al, 2004) [3]

3.1.5 Logistic Regression

Logistic regression (LR) is a probabilistic statistical classification model.Logistic regression is one of the most widely used techniques in credit scoring like discriminant analysis. The main differences between a logistic regression model from a linear regression model are that the outcome variable in logistic regression is dichotomous (a 0/1outcome) [3]. LR model can fit various kinds of distribution functions such as Gamble, Poisson, and normal distributions, unlike other statistical tools (e.g. discriminant analysis or ordinary linear regression). It is more suitable for the fraud detection problems. In addition, in order to increase its accuracy and flexibility, several methods have been proposed to extend the traditional binary logistic regression model including multinomial logistic regression model and logistic regression model for ordered categories [5].

3.2 Artificial Intelligence based techniques

3.2.1 Artificial Neural Networks (ANNs)

ANNs are computational models inspired by an animal's central nervous systems which are capable of machine learning as well as pattern recognition [12]. ANNs are inspired by the functionality of the nerve cells in the brain. Just like humans, ANNs can learn to recognise patterns by repeated exposure to many different examples. They are non-linear models that can classify based on pattern recognition capabilities. This gives them an advantage over conventional statistical techniques used in industry which are primarily linear. In the field of credit scoring, studies have shown that neural networks perform significantly better than statistical techniques. [1], [5]. ANN have been used in credit rating and credit scoring quite extensively as illustrated in the following papers : "Artificial Neural Networks for Corporation Credit Rating Analysis"[7], "Personal Credit Rating Assessment for the National Student Loans based on Artificial Neural Network"[8], "Personal Credit Rating Using Artificial Intelligence Technology for the National Student Loans" where a Back Propagation neural network was used [9], "Research of electronic commercial credit rating based on Neural Network with Principal Component Analysis" [10]

3.2.2 Genetic Algorithms (GAs)

GAs try and replicate the natural selection process where genes are passed from one generation to the next generation.. GAs are inspired by biological evolution and offer efficient problem-solving mechanisms. A problem's solution is evolved over many processing cycles, each time producing better solutions. Application of GAs is rapidly expanding with successful applications in finance trading, fraud detection and other areas of credit risk. Desai et al. investigated the use of GAs as a credit scoring model in a credit-union environment while Yobas et al. compared the predictive performances of four techniques, one of which is GAs, GA faired quite well coming in second place[1][5]. GAs can perform at least as good or better than traditional techniques [2].

3.2.3 Artificial Immune Systems (AIS)

AIS are an Artificial intelligence technique inspires natural immune system of the body. AIS have a learning and memory component, and perform pattern recognition. AIS were implemented in a paper titled "An Artificial Immune System for Extracting Fuzzy Rules in Credit Scoring" [11]. Weka data mining software was used to classify and in turn compared with other well-known classifiers. They used the clonal selection algorithm to implement the AIS. Competitive results with high accuracy were obtained.

4. Future Work

The area of credit scoring is a very critical area for the financial sector. For the financial sector to survive formidable credit scoring techniques are required.. The area of credit scoring is a very interesting area with a lot of research still going on. Algorithms and techniques which attain higher detection rates continue to be found. Particularly in the area of Artificial Intelligence (AI). AI techniques are an area to look out for particularly in the area of credit card scoring.

5. Conclusion

Various credit scoring techniques were looked at. The first review looked at statistical based methods and the rest were nature inspired. The nature inspired methods were ANN, based on the model of brain, GA, based on natural selection modelled on the evolution of species, and the last one is Artificial Immune System. Some studies found statistical techniques to perform better than AI techniques, while others concluded just the opposite.

References

[1] Kevin Leung, France Cheong, Christopher Cheong "Consumer Credit Scoring using an Artificial Immune System Algorithm," IEEE, pp. 3377-3384, 2007

[2] David J. Fogarty, "USING GENETIC ALGORITHMS FOR CREDIT SCORING SYSTEM MAINTENANCE FUNCTIONS," International Journal of Artificial Intelligence & Applications (IJAIA), Vol.3, No.6, pp 1-8, November 2012

[3] Abdou, H. & Pointon, J. (2011) 'Credit scoring, statistical techniques and evaluation criteria: a review of the literature ', Intelligent Systems in Accounting, Finance & Management, 18 (2-3), pp. 59-88.

[4] Antonio I. S. Nascimento and Germano C. Vasconcelos ,"An Experimental Investigation of Artificial Immune System Algorithms for Credit Risk Assessment Applications" In Proceedings of the WCCI IEEE World Congress on Computational Intelligence, pp 1-8, 2012

[5] Antariksha Bhaduri "Credit Scoring using Artificial Immune System Algorithms: A Comparative Study," World Congress on Nature & Biologically Inspired Computing, pp. 1540-1540, 2009

[6] Yang Zong-chang, Kuang Hong "Credit Evaluation for Mobile Customers Using Artificial Immune

[7] Liu Yijun, Cai Qiuru, Luo Ye, Qian Jin, Ye Feiyue "Artificial Neural Networks for Corporation Credit Rating Analysis," International Conference on Networking and Digital Society,pp. 81-84, 2009\

[8] Xiao_ jie Zhang, Jian Hu "Personal Credit Rating Assessment for the National Student Loans based on Artificial Neural Network," International Conference on Business Intelligence and Financial Engineering, pp. 53-56, 2009

[9] Jian HU "Personal Credit Rating Using Artificial Intelligence Technology for the National Student Loans," In Proceedings of the 4th International Conference on Computer Science & Education, pp. 103-106, 2009

[10] XUE Xiang-hong, XUE Xiao-feng "Research of electronic commercial credit rating based on Neural Network with Principal Component Analysis," IEEE

[11] , pp. 1-4, 2010

[12] Ehsan Kamalloo ,Mohammad Saniee Abadeh "An Artificial Immune System for Extracting Fuzzy Rules in Credit Scoring;" IEEE, pp 1-8, 2010

[13] Wikipidea, wikipedia.org, [Online]. Available: http://en.wikipedia.org/wiki/Artificial_neural_network,[Accessed: July.15, 2014]

[14] Martin VOJTEK, Evžen KO ČENDA www.risknet.de [Online]. Available: https://www.risknet.de/uploads/tx_bxelibrary/CreditSco ring.pdf, [Accessed: July.15, 2014]

Author Profile

Thabiso Peter Mpofu received B. Tech degree in Computer Science at Harare Institute of Technology (HIT), Zimbabwe in 2010. He is currently pursuing M. Tech Computer Science at JNTUH, India. He is a HIT staff development research fellow. His research interests are in the area of Data Mining, Network Security and Mobile Computing.

Macdonald Mukosera received the B.Tech Hons degree in Computer Science from Harare Institute of Technology in 2010. During 2011-2012, he worked as a Teaching assistant at Harare Institute of Technology in the Software Engineering Department. He is now studying M Tech Computer Science at Jawaharlal Technological University Hyderabad in School of Information Technology.