

Data Mining and Machine Learning for Financial Analysis

Thitimanan Damrongsakmethee^{1,2*} and Victor-Emil Neagoe²

¹Department of Business Computer, Suratthani Rajabhat University, Thailand; manancc@gmail.com

²Department of Applied Electronics and Information Engineering, Polytechnic University of Bucharest, Romania; victoremil@gmail.com

Abstract

Data mining is the process of discovering patterns, corresponding to valuable information from the large data sets, involving methods at the intersection of machine learning, statistics and database systems. Evolving from the fields of pattern recognition and artificial intelligence, machine learning explores the study and construction of algorithms that can learn from sample inputs. Financial data analysis is used in many financial institutes for accurate analysis of consumer data to find defaulters, to reduce the manual errors involved, for fast and saving time processing, to reduce the mis judgements, to classify the customers directly and to reduce the loss of the financial institutions. We have analysed a lot of machine learning techniques for financial analysis, namely models of supervised classification (Artificial Neural Networks, Support Vector Machine, Decision Trees), those of prediction (Cox survival model, CART Decision Trees), and also modelsof clustering (K-means clustering).

Keywords: Artificial Neural Networks, Bankruptcy Prediction, Classification, Clustering, Credit Risk, Credit Scoring, Data Mining, Financial Analysis, Machine Learning, Risk Management

1. Introduction

Since 1980, a lot of researchers had as aim of their work to find solutions to search a pattern in a large amount of data. There are used statistical, machine learning and computational intelligence techniques grouped under the umbrella *called data mining*. Data mining is a technique for discovering interesting patterns from a large amount of data stored in the databases, data mart, data warehouse or other information repositories¹. Data mart and data warehouse are tools that help in the management of business information. In the era of using the data warehouse through the development of the data mart, even though it will be limited to use of any of the departments, but

the information stored in the data warehouse is more important for a specific organization². Data warehouse is a set of data mart that provides information from the different operations in the company. It can contain information about the daily operations of the various factions of the company⁴. Data mart as part of a data warehouse can provide transaction reports and analysis on some departments or operations in the company. Each company can keep the information of each department in its database, such as database of financial department, database of sales department, database of production and that of marketing department^{5,6}. Data warehouses have a basic architecture that can create applications from data mining⁷. Data mining can be considered as

*Author for correspondence

a result of the natural evolution of information technology from multiple disciplines as database and data warehouse technology, statistics, high performance computing, machine learning, computational intelligence (implying neural networks, fuzzy systems, evolutionary computing, swarm intelligence and so on), pattern recognition, data visualization, information retrieval, image processing, and spatial or temporal data analysis⁸. Data mining and Knowledge Discovery from the Database (KDD) are recent developments in the field of data management technologies⁹. KDD is a kind of data mining designed to extract knowledge from a large amount of data¹⁰. The standard procedure in performing data mining based on Cross-Industry Standard Process of Data Mining (CRISP-DM) involves six phases, see Figure 1¹¹. These are the following:

1. *Business understanding* phase, consisting in: choosing the objectives, understanding the business goal, learning situation assessment and developing a project plan.
2. *Data understanding* phase, which consists of considering the data requirements and initial data collection, exploration and quality assessment.
3. *Data preparation* phase, consisting in: selection of required data, data integration and formatting, data transformation and data cleaning.
4. *Modeling* phase, consisting in: selection of appropriate modeling techniques, development and examination of alternative modeling algorithms and parameter settings and finding the tuning of model setting according to an initial assessment of the model's performance.
5. *Evaluation* phase, corresponding to evaluation of the model experiment results.
6. *Deployment* phase, representing an implementation step, where a model report is performed.

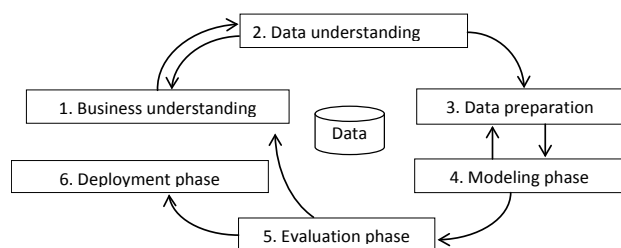


Figure 1. Cross-Industry Standard Process of Data Mining (CRISP-DM).

Data mining and machine learning are very necessary as a way of managing big data, enterprise efficiency and busi-

ness intelligence^{1,12}. Data mining provides enormous value in finance and banking¹³. Banks need to find the hidden patterns in the large sets of data and thus they can monitor the data in their database¹⁴. Such data can be personal data that describe the financial status and the financial behavior before and by the time the client receives a credit¹⁵. Most banks and financial institutions have different services for customers such as that of monitoring data for opening a savings account for each client's business. The schedule gives credit to customers in transactions such as mortgage business, car loan services, the investment services, insurance services and stock investment services⁸. Other financial applications of data mining and machine learning are prediction of financial events that will happen in the future, such as: stock markets, foreign exchange rates, bankruptcy, credit rating of the bank's customer information, predictive financial and investment analysis, trading futures, understanding and managing financial risk in banks¹⁶.

As technology develops, it starts with bringing Artificial Intelligence (AI) technology to be used in fund managements, asset managements and other more financial institutions. Machine learning algorithms are used to isolate and analyze data from the large database^{17,18}. Using this tool, one can find some patterns and can predict the outcome¹⁹. However, there were many kinds of research using machine learning in banking to forecast future events that can help in decision-making processes. Nowadays, financial institutions and most banks are investing in information technology to bring data mining and machine intelligence techniques to manage the group of datasets in order to successfully operate in the presence of a competitive business^{12,20}.

2. Risk in banks

The banks are aware of the various risks. That may occur and adversely affect the business of the bank²¹. Banks analyze the risk factors that are important^{22,23}. The quality of risk analysis may affect the financial performance of the business. There are risks of all institutions and organizations that may bring numerous direct and indirect losses^{21,24}. There are three major risks in banks corresponding to credit risk, operation risk and market risk^{19,22,25}. Financial institutions should monitor credit risk management as appropriate. Banks are required to manage the credit risk compared to the risk of credit management individually²⁶. Credit risk management

efficiency is very important and essential to the long-term success of banks^{28,29}. Popular tool used to evaluate the credit risk of individuals is Credit scoring²⁷. Credit scoring uses a report to evaluate some external components. The external reports process information on the status of credit risk data from credit bureaus and reliability party credit attributed together with the financial history and the current financial state of borrowers individually. Financial institutions have to remove unwanted features to distinguish between “good” and “bad” policies to manage the credit risk of each entity³⁰.

3. Related Work

The Bank is an agency that has a major role in the development of the economy of the country. The unknown future behaviors of the customers are quite important to Customer Relationship Management (CRM). It becomes increasingly important for the bank to predict their customer's future decisions in order to take suitable actions in time²⁹. There are various areas where data mining and machine learning can be used in financial sectors like customer segmentation and profitability, credit analysis, predicting payment default, marketing, fraudulent transactions, ranking investments, optimizing stock portfolios, cash management and forecasting operations, high risk loan applicants, most profitable credit card customer and cross selling³¹⁻³³. In 2010, M. C. Lee and C. To³⁴ described the application of novel data mining techniques for evaluation of the enterprise financial distress and credit prediction; there are improved the performance of algorithms by using Support Vector Machine (SVM) with 3-folds cross validation and Back Propagation Neural Network (BPN) by the four measured attributes. The data for this study have been collected from the database of a security firm in Taiwan. In this research, there are used 20 experimental samples for training data and 25 samples for testing data. By comparing the results, there has been shown that SVM gives higher precision of about 100% prediction accuracy and classification accuracy, implying low error rates, while BPN has led to 96% of prediction accuracy and 95% of classification accuracy.

Many kinds of research about customer credit policy analysis were performed in 2012. K. Chopde et al.²³ have studied the data mining techniques for credit risk analysis - in particular, the decision tree techniques. This research used data mining for credit risk analysis enabling the bank

to reduce the manual errors. This decision-making process is fast, it saves time processing and it helps the bank to reduce the misjudgments. The research result found by the Meta Decision Tree (MDTs) used a base level classifier and the Random Forest (RF) classifier, leading to a more accurate classification score than the CART decision tree. Overall, the decision tree has proved to be a technique that can classify the customers directly with a good score and thus it can reduce loss for the financial institutions in the best way.

I. G. Ngurah et al.²⁴ used to suggest a decision tree model for credit assessment. This paper aims to identify factors which are necessary for a rural bank in Bali to assess credit applications. Current decision criteria in credit risk assessment are evaluated. The credit risk assessment model has been applied to PT BPR X and it has used C5.0 methodology; this model has used 84% of 1028 data as evaluation data to suggest the new criteria in analyzing the loan application. The result showed that PT BPR X can reduce nonperforming loans to less than 5% and the bank can be classified or not as a well-performing one by applying data mining technology. In the same year, W. Chen et al.²⁹ proposed a hybrid data mining technique to build an accurate credit scoring model to evaluate credit risk based on the credit data set provided by a local bank in China. This research has proposed two processing stages: the first (clustering stage), meaning that the samples of accepted and new applicants are grouped into a homogeneous cluster by using K-means clustering. The second processing stage is the classification with Support Vector Machines (SVM). By comparison with other credit scoring models, here the samples the previous model uses three or four classes rather than two (good and bad credit). In addition, data mining ideas and algorithms can be applied to panel data to find a knowledge which is different from the regression knowledge discovered by the traditional linear regression. For this reason, G. Nie et al.¹² proposed a new distance measurement with real panel data about credit card application in China; it can be used in panel data clustering with K-means clustering method. This research analyzed the groups of different customers by the behavior of credit card holders. The result has shown that more accurate knowledge can be found with the panel data structure; distance measurement can reflect the information of different periods and panel data can be used in clusters to provide new knowledge.

In 2015, A. Byanjankaret al.³⁰ described the application of Artificial Neural Networks (ANNs) for building

credit scoring models in Peer to Peer Lending (P2P), to gain market share in the financial industry. This research used the neural network credit scoring model. The data have been divided in the following manner: 70% of the observations have been used for training and 30% of observations have been used for testing. The neural network credit scoring model has shown a promising result in classifying credit applications to allow the lenders taking a smart decision in selecting a loan application and predicting the credit risk. In 2015, A. Gepp and K. Kumar³⁵ proposed a semi-parametric survival analysis model consisting in Cox, Discriminate Analysis (DA), Logistic Regression (LR) and a non-parametric CART decision tree; the above models have been applied and compared to financial distress prediction. Regarding classification accuracy, the CART model had led to the lowest error of classification and regarding performance analysis of prediction accuracy, the Cox model had the lowest weighted error in 40% of the cases, while DA and the CART model had the lowest error in about 60% of the cases. The overall result provided empirical evidence which supports the use of survival analysis and decision tree techniques for financial distress.

In 2016, R. G. Lopes et al.²⁵ developed three predictive models: Generalized Linear Modeling (GLM), Gradient Boosting Method (GBM) and Distributed Random Forest (DRF) by using the R language, to predict the recovery of credit operations in a Brazilian bank. All models have been built in 10-folds cross validation and there has been obtained a high evaluation result of the ROC curve. The GBM model has shown a better performance; this model has been used to help identify customers having the best potential with an 85.5% accuracy rate. This research has also identified the delinquent clients that had the highest probability of short-term recovery, in order to support the activities of account managers.

4. Machine learning models proposed for financial analysis

4.1 Classification techniques

Support Vector Machine (SVM) is a tool to find the hyper-plane that can be used for classification; it is based on kernel functions^{17,34,36}. The Gaussian kernel is the most versatile kernels^{47,37}. By the width parameter σ of the Gaussian kernel function, one can control the flexibility of

SVM classifier results. The Gaussian function can be used not only as a kernel for SVM, but also for some exciting neuro-fuzzy classifiers³⁸.

Decision trees are classifiers expressed as a recursive part of the instance space. *Classification and Regression Trees (CART)* model is a flexible method to describe how the variable Y distributes after assigning the forecast vector X of the metric. The CART model uses a binary tree to divide the forecast space into certain subsets on which Y distribution is assumed continuously^{39,40}.

Artificial Neural Networks (ANNs) constitute a non-linear statistic model based on the function of the human brain⁴¹. ANNs provide powerful tools of data mining techniques for data analyst relationship modeling. ANNs are able to recognize the complex patterns in input data and also they can predict the outcome of the new independent input data precisely⁴². ANNs have the remarkable ability to derive meaning from complicated data or imprecise data. It can be used to extract patterns and detect trends using specific techniques⁴³. ANNs are very suitable for identifying patterns, and they are also very well suited for prediction or forecasting data⁴⁴. One of the most well-known ANNs is the Multi-layer Perceptron (MLP)⁴⁵ named also as the Back-Propagation Neural Network (BPN); its algorithm is based on the computation of the errors of each output neuron after processing an input data³⁵. It is a general technique called automatic differentiation. BPN is characterized by backward propagation of output errors, namely these errors are computed at the output layer and the training is distributed back to the weights of the previous layers to reduce the output errors⁴⁶.

Survival Analysis method is a new technique of credit scoring model. A common way that banks can differentiate customer information when they apply for a loan from the bank. Banks can separate the good information from the bad information regarding the loan application. The system can calculate the profitability of customers and also it can evaluate the profit scoring from the customers³⁵. Worthwhile Survival analysis can predict the duration of the event will take place in advance and forecast the probability of occurrence of an event to occur⁴³.

The H2O team discovered these famous data mining techniques to analyze the group datasets. These techniques are Generalized Linear Models (GLM), Gradient Boosting Method (GBM) and Distributed Random Forest (DRF). GLM are similar with the linear regression model. Data mining techniques are used for regression analysis and

data classification. GLM model is very popular because it is easy to be interpreted and it is also a high-speed processing stage when used for the large datasets²⁸. GBM model is a tool for prediction using regression or classification. It is an ensemble of the tree models and provides considerably accurate results. GBM model applies weak classification algorithms to incrementally change data, to create a series of decision trees²⁸. Finally, the DRF is an ensemble of tree models, where each tree is related from other trees. DRF is the most powerful technique for classification and regression. DRF can generate a big forest of classification or regression trees rather than a single classification or regression tree²⁸. In addition, DRF builds half as many trees for binomial problems with a single tree to estimate class 0 by probability(p_0) then computes the probability of another class 1 as (p_1). For multiclass problems, DRF is used to estimate the probability of each class separately⁴⁷.

4.2 Clustering technique

Cluster analysis groups are the data mining techniques used to classify as variable or split into small groups of two or more. The objects within a group are similar to one another and different from the objects in other groups. *K-means clustering* is a method of clustering the observations into a specific number of disjoint clusters^{15,17,33}. In principle, K-means clustering aims to partition a dataset as $\{X_1, X_2, \dots, X_N\}$ into K subsets to minimize the distortion measure defined by the function given below

$$\sum_{n=1}^n \sum_{k=1}^k r_{nk} \|X_n - \mu_k\|^2$$

where binary indicator $r_{nk}=1$, only if only if data point X_n is assigned to the k th cluster (for the other cases, $r_{nk}=0$) and μ_k denotes the mean of the k th cluster.

In Table 1 there are given a lot of papers for financial (banking) applications with the corresponding machine learning techniques and research results.

Table 1. List of research papers with corresponding data mining tasks, machine learning techniques and research results

Reference	Data mining tasks	Machine learning techniques	Research results
15	Clustering	K-means clustering	There can be found the panel data structure. There can reflect analysis for different periods.
29	Classification Clustering	SVM K-means clustering	There can be applied to panel data to find knowledge which is different from the regression knowledge discovered by the traditional linear regression.
34	Classification	SVM BPN	There are compared the results obtained by SVM and BPN for financial distress. The research results had shown that SVM leads to a lower error rate than BPN.
23	Classification	Decision trees – The CART model – MDT – RF	Decision trees techniques can reduce the manual errors, to obtain faster and saving time processing, they can reduce the misjudgments, can classify the customers directly and can reduce loss for the financial institutions.
24	Classification	Decision trees – PT BPR X, C.50	There is reduced the number of non-performing loans.
30	Classification	ANNS	There are classified the credit applications in order to allow the lenders taking a smart decision to select a loan application and to predict the credit risk.
35	Classification Prediction	Decision trees – The CART model A survival analysis – Cox model – DA model – LR model	The presented results provide empirical evidence to support decision trees and a survival analysis in banks for financial distress to compare the performance analysis. The CART model had obtained the best classification accuracy. In addition, the Cox, CART and DA model had led also to good prediction accuracy.
25	Classification	GLM Model GBM Model DRF Model	GBM model has shown a better performance. GBM had the highest probability of short-term recovery to support the activities of account managers and increase the efficiency of their approach with customers.

5. Conclusion

Data mining based on machine learning techniques is a technology that can be used to analyze existing data, applications and customer needs in order to build and maintain long-term customer relationships. It can build confidence for clients making customer satisfaction and business the longest. Using machine learning techniques for classification and clustering tasks is popular in the loan payment prediction and the customer credit policy analysis of the banking system. In this paper, we proposed data mining techniques which contain two main processing stages. Classification stage consists of several models including SVM, ANNs, Decision Trees and BPN. We found that SVM model and Decision Tree model are promising techniques for classification with financial applications. The above-mentioned techniques can reduce the manual errors, they can lead to faster and saving time processing, they reduce them is judgments for classifying the customers directly and thus they can reduce loss of the financial institutions. In clustering stage, K-means clustering is the best performing model for customer credit management of credit scoring model. The scoring methods are used to estimate the creditworthiness applicant. When credit loans and finances have the risk of being defaulted, credit managers have to develop and apply data mining methods to handle and analyze credit data in order to save time and reduce the errors. Data mining(implemented mainly using techniques of machine learning) will be a challenge for the future research in banking and financial areas.

6. References

1. Felista U, Sunday O, Rose-kezia C. Industry Wide Applications of Data Mining. *Int J Adv Stud Comput Sci Eng. IJASCSE*. 2014; 3(2):28–37.
2. Hu ZG, Ing Li J, Hu L, Yang Y. Research and Application of Data Warehouse and Data Mining Technology in Medical Field. 2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). 2015.
3. Mohammed M, Anad M, Mzher A, Hasson A. Meta-Data and Data Mart Solutions for Better Understanding for Data and Information in E-Government Monitoring. 2012; 9(6):78–83.
4. Raman VVR, Tewari V. Gaining Competitive Advantage on the Basis of Data Warehousing and Data Architecture. *Proc World Congr Eng*. 2011; 3:1–6. PMID:20056374
5. Computing M, Chhabra R, Pahwa P. Data Mart Designing and Integration Approaches. 2014; 3(4):74–9.
6. Seah BK, Selan NE. Design and Implementation of Data Warehouse with Data Model using Survey-Based Services Data. Luton, United Kindom, 2014. Crossref
7. Singh S, Malhotra S. Data Warehouse and ITs Methods. *J Glob Reasearch Comput Sci*. 2011; 2(5):113–5.
8. Han J, Kamber M. Data Mining: Concepts and Techniques Second Edition. San Francisco: an imprint of Elsevier, 2006.
9. Deshpande S, Thakare V. Data Mining System and Applications : A Review. *Int J Distrib Parallel Syst*. 2010; 1(1):32–44. Crossref
10. Wang HF, Chan ZW. A Pruning Approach to Pattern Discovery. *Int J Inf Technol Decis Mak*. 2008; 7(4):721–36. Crossref
11. Larose DT. Discovering Knowledge in Data: An Introduction to Data Mining, 2nd Edition. New Jersey: John Wiley and Sons, 2014. Crossref
12. Nie G, Zhang L, Liu Y, Zheng X, Shi Y. Decision Analysis of Data Mining Project Based on Bayesian Risk. *Expert Syst Appl*. 2009; 36(3) PART 1:4589–94.
13. Wikipedia. Data mining. 2017. [Online]. Available from: Crossref. [Accessed: 15-Nov-2017].
14. Pulakkazhy S, Balan RVS. Data Mining in Bangking and ITS Applicatons : A Review. *J Comput Sci*. 2013; 9(10):1252–9. Crossref
15. Li X, Zhu Z, Pan X. Credit Card Customer Analysis Based on Panel Data Clustering. *Procedia Comput Sci*. 2010; 1(1):2489–97. Crossref
16. Romero C, Ventura S, Pechenizkiy M, Baker RSJD. Handbook of Educational of Data Mining. Florida: Chemical Rubber Company, 2010. Crossref
17. Bishop C. Pattern Recognition and Machine Learning. New York: Springer, 2006.
18. Wikipedia. Machine Learning. 2017. Date Accessed: 15-Nov-2017. Available from: https://en.wikipedia.org/wiki/Machine_learning.
19. Jayasree V, Balan RVS. A Review on Data Mining in Banking Sector. *Am J Appl Sci*. 2013 10(10):1160–5. Crossref
20. Lin WY, Hu YH, Tsai CF. Machine Learning in Financial Crisis Prediction: A Survey. 42(4):421–36.
21. Rose N. Powers of Freedom : Reframing Political Thought. United Kingdom: Cambridge University Press, 1999. Crossref
22. Foust D, Pressman A. Credit Scores : Not-So-Magic Numbers. *Bus Week*. 2008; 4071:38.
23. Chopde K, Gosar P, Kapadia P, Maheshwari N, Chawan PM. A Study of Classification Based Credit Risk Analysis Algorithm. *Int J Eng Adv Technol*. 2012; 1(4):142–144.

24. Ngurah IG, Mandala N, Badra C. Assessing Credit Risk : An Application of Data Mining in a Rural Bank. *Int Conf Small Mediu Enterp. Dev with a Theme innovation Sustain SME Dev (ICSMED 2012)*, vol. 2012; 4:406–12.
25. Lopes RG, Ladeira M, Carvalho RN. Use of Machine Learning Techniques in the Prediction of Credit Recovery. *Adv Sci Technol Eng Syst J*. 2017; 2(3):1432–42. Crossref
26. Apostolik R, Donohue C, Went P. *Foundations of Banking Risk : An Overview of Banking, Banking Risks, and Risk-Based Banking Regulation*. Hoboken, New Jersey: John Wiley & Sons, 2009.
27. Van Gestel T, Baesens B. *Credit Risk Management Basic Concepts : Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital*. United Kingdom: Oxford University Press, 2008. Crossref
28. Guo YJ, Wu C. Research on Credit Risk Assessment in Commercial Bank Based on Information Integration. *Proc 2009 Int Conf Manag Sci Eng*. 2009. p. 515–25.
29. Chen W, Xiang G, Liu Y, Wang K. Credit Risk Evaluation by Hybrid Data Mining Technique. in *Systems Engineering Procedia*. 2012; 3:194–200. Crossref
30. Byanjankar A, Heikkila M, Mezei J. Predicting Credit Risk in Peer-to-Peer Lending : A Neural Network Approach. *Proc - 2015 IEEE Symp Ser Comput Intell*. 2015; 719–25. Crossref
31. Ramageri BM, Desai BL. Role of Data Mining in Retail Sector. *Int J Comput Sci Eng*. 2013; 5(1):47–50.
32. Moradi M, Salehi M, Ghorgani ME, Yazdi HS. Financial Distress Prediction of Iranian Companies Using Data Mining Techniques. *Organizacija*. 2013; 46(1):20–7. Crossref
33. Janakiraman S, Umamaheswari K. A Survey on Data Mining Techniques for Customer Relationship Management. *Int J Eng Bus Enterp Appl (IJEBA)*. 2014; 55–61.
34. Lee MC, To C. Comparison of Support Vector Machine and Back Propagation Neural Network in Evaluating the Enterprise Financial Distress. *Int J Artif Intell Appl*. 2010; 1(3):31–43. Crossref
35. Gepp A, Kumar K. Predicting Financial Distress: A Comparison of Survival Analysis and Decision Tree Techniques. in *Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015)*. 2015. 54. p. 396–404.
36. Vapnik V. *The Nature of Statistical Learning Theory*. New York: Springer, 1995. Crossref PMid:8555380
37. Wu DD, Zheng L, Olson DL. A Decision Support Approach for Online Stock Forum Sentiment Analysis. *IEEE Trans Syst MAN, Cybern Syst*. 2014; 44(8):1077–87. Crossref
38. Neagoe VE, Iatan R, Iatan I. A Nonlinear Neuro-Fuzzy Model for Prediction of Daily Exchange Rates. in *International Symposium on Soft Computing for Industry held at the 6th World Automation Congress WAC'04*. 2004; 573–8.
39. Neelamegam S, Ramaraj E. Classification Algorithm in Data Mining: An Overview. *Int J P2P Netw Trends Technol*. 2013; 4(8):369–74.
40. Sindhuja D, Priyadarsini R. A Survey on Classification Techniques in Data Mining for Analyzing Liver Disease Disorder. *Int J Comput Sci Mob Comput*. 2016; 5(5):483–8.
41. Bahrammirzaee A. A Comparative Survey of Artificial Intelligence Applications in Finance : Artificial Neural Networks, Expert System and Hybrid Intelligent Systems. *J Neural Comput Appl*. 2010; 19(8):1165–95. Crossref
42. Alaraj M, Abbod M, Alhnaity B. Evaluation of Consumer Credit in Jordanian Banks : A Credit Scoring Approach. 2015 17th UKSIM-AMSS in International Conference on Modeling and Simulation Evaluation. 2015. p. 125–30.
43. Wekesa OA, Samuel M, Peter M. Modelling Credit Risk for Personal Loans using Product-Limit Estimator. *Int J Financ Res*. v 2012; 3(1):22–32.
44. Tamilselvi R, Kalaiselvi S. An Overview of Data Mining Techniques and Applications. *Int J Sci Eng. Res*. 2013; 1(1–3): 506–9.
45. Tsai C, Wu J. Using Neural Network Ensembles for Bankruptcy Prediction and Credit Scoring. *Expert Syst Appl*. 2008; 34(4):2639–49. Crossref
46. Wikipedia. Backpropagation. Data Mining, Machine Learning [Online]. Available from: Crossref [Accessed: 12-Sep-2017].
47. Team THA. Distributed Random Forest. [Online]. Available from: Crossref. [Accessed: 12-Sep-2017].