Development of a data-driven decision support system for credit risk assessments in Indian Small and medium enterprises

Kalyan Nagaraj[1]

[1]RVCE

November 29, 2021

Abstract

Credit rating is performed to estimate the performance of companies in SME sector. Banks and Financial institutions use process driven protocols and documentary evidence for giving credit rating to SMEs, which is often a time-consuming and costly affair. As an alternative, strategy soft computing techniques have found preference in many financial sectors as approach of credit risk assessment in cases where time is constrained. In this study, the SME's credit rating data was collected from CMIE Prowess database and Ace Analyzer web portal between the financial years 2013 to 2021. The dataset was initially subjected for pre-processing to remove the outliers. Further various classification and clustering techniques were tested for predicting the nature of credit risk. On the basis of accuracy, the top two ranked clustering algorithms (i.e. k-means and hierarchical) were deployed as a predictive tool for the credit rating of SME dataset in R programming language and python.

1. Introduction

According to the Micro, Small and Medium Enterprises Development (MSMED) Act of 2006, SMEs are defined as small and medium enterprises whose investment in plant and machinery does not exceed Rs. 10 crore. These SMEs are further sub categorized as micro (investment upto 25 lakhs), small (25 lakhs to 5 crore) and medium (5 crore to10 crore) enterprises based on the investment in plant and machinery (Micro, Small and Medium Enterprises). In India, there are about 48 million SME's which contributes to about 40% of the total economy, thereby generating about 1.3 million jobs every year. It is expected that about 12 million people would join the sector thus increasing the growth of SMEs by 8% in the coming years (Malini Goyal, 2013).

The SMEs today are widespread in various sectors including, food engineering, agriculture, chemicals, textiles, leather and goods, polymers and plastics, pharmaceuticals and paper mills to name a few. The SME sector enhances the capital growth by increased elasticity of manpower. The expansion of SMEs on a global scale favors the entrepreneurs to adopt the ever growing innovation and technology. A major hindrance to the expansion of the SMEs is due to the lack of adequate and well-timed capital to support their plans. In this scenario, the banks and financial organizations are striving hard to extend the funds for the growth of the SMEs. Often the capital granted to SMEs is not well managed and controlled leading to risk in the credit offered. According to the reports of MSME, only 16% of the SMEs adopts for financial support from banking sector to organize their business needs, while most others are left unaware (S. Nehru et al, 2012). Thereby most of the SMEs today are facing shortage of financial assets, failing to generate revenue on a timely basis. In order to improve and facilitate the growth of SMEs sector the government has initiated the credit rating system for SMEs. The financial performance of the SMEs is adjudged based on these ratings. Some of the prominent agencies that provide SME ratings are SMERA (http://www.smera.in/), CRISIL (http://www.crisil.com/index.jsp) and ICRA (http://www.icra.in/). There is no formal theory or principle

incorporated by these agencies that defines the procedure to derive credit ratings. The ratings are manifested based on financial and non-financial data for each firm. The ratings declared are dynamic with respect to time which may be upgraded or either downgraded (Michael O' Neill et al, 2005).

Statistical and mathematical models are being developed using data mining and artificial intelligence techniques to predict credit scoring. Credit scoring is a typical multi-classification problem in data mining where the output class is having more than two instances like A, B, C and D. Different data mining algorithms have been employed for multi-classification in past (Yoram Singer et al, 2000; Mohamed Aly, 2005). The major objective of credit scoring is to develop a robust model for evaluation of credit risk in SMEs. Risk evaluation can be used to analyze the position of companies in the current financial year.

Several data mining algorithms have been implemented for predicting credit rating of SMEs in the recent past. In this study, a robust model is developed for prediction of credit ratings using data mining techniques. The model is further incorporated as a simple user friendly tool to help users perform prediction of SME credit rating. Major Headings

2. Literature Review

2.1. Credit scoring- An Introduction

The term credit defines the phenomenon of borrowing and lending resources like finance and consumer goods (Sheffrin et al, 2003; Y Yang, 2007). It is important to interpret the nature of risk associated during the transaction of credit. Credit risk is formulated by developing a scorecard based on data collected from banks and financial organizations. Consequently these scorecards can be used to adjudge the probability of a transaction being default. Credit scoring is associated with several benefits as discussed by Crook et.al (Crook J N et al, 1996). Credit scoring is defined as the process of developing models for evaluating credit worthiness by Jacka and Hand in 1998 (Jacka S. D et al, 1998). Over the years several other definitions have been framed for credit scoring. Credit scoring was defined by Thomas et.al as decision models which aid lenders prior to granting consumer credit (Thomas L C, 2000). The scoring models based on statistically significant parameters are better than judgmental models. Correlation analysis is well performed in the credit scoring models rather than judgmental models. Credit scoring models are having enormous applications is the field of finance and accounting. Consequently due to large number of credit applications in the recent years, scorecards are used effectively in modeling consumer loans and mortgages (Orgler Y E, 1971; Steenackers A et al, 1989; Chen I et al, 2005). Credit worthiness of new customers can be evaluated using the credit scoring models enabling a better decision making process (Zupan J et al, 2009; Mark Jenkins et al, 2013).

2.2. Data mining techniques in credit scoring

Several studies have been performed for prediction of credit scoring using data mining and soft computing techniques in the recent past. Initial studies for prediction of credit score was performed using statistical techniques like multiple discriminant analysis (G. E. Pinches et al, 1973; A. Belkaoui, 1980) logistic regression (L. H. Ederington, 1985) and Naïve Bayes (AC Antonakis et al, 2009). In further years, artificial intelligence techniques like neural network and its variants were used for credit rating (S. Shekhar et al, 1988; RT Redmond et al, 1993).It was followed by implementation of other computing techniques like support vector machine (Z. Huang et al, 2004;Cheng-Lung Huang et al, 2006) decision trees (Yi Ziang et al, 2008; Salih Gunes et al, 2009) random forest (D Poel et al, 2008). Different technologies like One-against-one, One-against-all, directed acyclic graph SVM, constraint classification were used for automatic classification for prediction of S&P bond ratings along with Gaussian RBF optimization (Z. Jingqing et al, 2006). The accuracy of these

techniques were better compared to other data mining techniques like neural network, case based reasoning and multiple discriminant analysis. An empirical study for credit scoring was performed in a study using data mining techniques to identify non-defaulting applicants (Wei Li et al, 2011). In another study, credit rating was performed using adaptive fuzzy-rule based systems to classify US companies and municipalities. The results can be classified to increase performance of different classifiers (Petr Hájek, 2011). Different data mining techniques and applications were discussed in brief to highlight their importance in different domains (Shu-Hsien Liao et al, 2012). In another study, different data mining techniques were implemented to predict credit scoring in banking sector (Shin-Chen Huang et al, 2013). Credit scoring was performed in a study combining performance and interpretation in kernel discriminant analysis. Italian bank case study was utilized to test the performance of the strategy (Caterina Liberati et al, 2015).

2.3. Clustering approaches in data mining

Clustering is an unsupervised learning approach that groups unlabeled data instances (M. R. Anderberg, 1973). A cluster is defined as a subgroup of similar entities. The distance between any two entities in the subgroup is lesser than the distance between any entity enclosed in the cluster and an entity outside the subgroup. Clustering is widely used in pattern recognition, classification, image analysis and machine learning tasks. Different clustering techniques include connectivity based clustering (hierarchical), centroid based clustering (k-means), distributive based clustering (expectation maximization), and density based clustering (R. C. Dubes et al, 1988). In a study, support vector machine (SVM) and k-means clustering techniques were implemented to form a hybrid model for real time business intelligence applications (Jiaqi Wang et al, 2005).

Clustering has been implemented for prediction of credit rating in previous studies. In a study, clustering and association rule method was used for prediction of customer behavior (Shahriar Mohammadi et al, 2010). In another study, clusters were developed for prediction of credit risk initially using logistic regression and multi-layer perceptron. The results were compared with the prediction from K-means and Kohonen self-organization maps. Further the credit score was generated based on the clustered results. The approach generated better performance compared to individual approaches (Alenjandro Correa et al, 2012).

In another study, credit score was predicted for bank customers using semi-supervised constrained clustering. The accuracy of classification was improved for prediction of default customers using this technique (Seyed Sadatrasoul et al, 2013).The fuzzy profiles were identified by c-means clustering, depending on hardship of population. The performances were adjudged based on datasets compared (Silvestro Montrone et al, 2015).

Based on literature review, it was observed that clustering techniques have been adopted for credit scoring applications. The previous studies have not attempted to develop a graphical user interface for credit scoring to detect the performance of SMEs. This study is an attempt to create a graphical user interface for prediction of credit rating of SMEs in India.

3. Methodology

The methodology of this study is represented in Figure 1.

3.1. Retrieving SME Credit Rating Dataset

The credit ratings data for Indian Small and Medium Enterprise (SME) companies was collected from CMIE Prowess Database (https://prowess.cmie.com/) and ACE Analyser (http://www.aceanalyser.com/). Annual data was collected from the financial reports of companies for the financial year 2018-2021. Lot of key variables in a financial statement can affect a credit worthiness of company, in this study only a limited

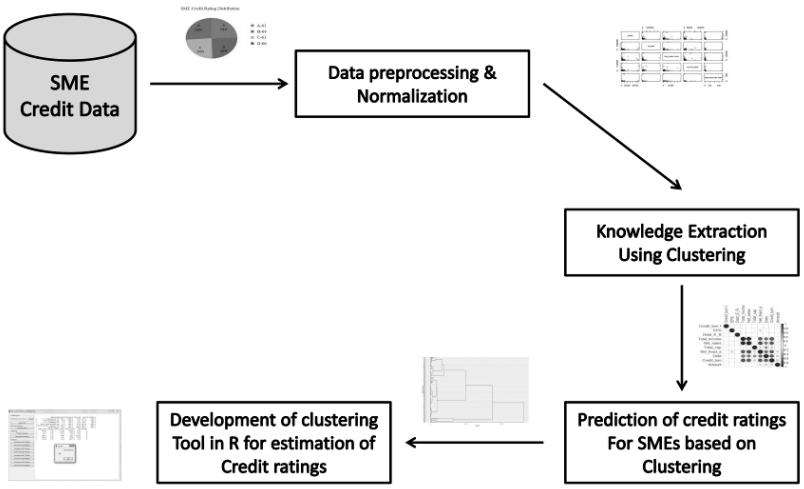

Figure 1: The overview of credit rating process

number of parameters has been selected based on their proven significance in past studies (Ward et al, 1963; M Jayadev, 2006; Kamaleshkumar Patel, 2011).

Fourteen parameters were filtered out from the data sources for the purpose of building the predictive model. These attributes forms the independent variables for the model. The credit ratings (which form the dependent variable in the model) comprised of the credit assessment report provided by CRISIL for the financial year 2013-2014. The description of the 14 financial attributes for SME companies is given in Table 1.

Sl. No	Financial Parameter	Description
1	Amount	Numerical: Credit amount
2	Net sales	Numerical: Sales generated after deduction of allowances
3	Total capital	Numerical: Financial estimate based on risk
4	Debt	Numerical: Allowance borrowed for financial purposes
5	Net fixed assets	Numerical: Price estimate for fixed assets
6	Current assets	Numerical: Estimate of all assets to be converted to cash
7	Debt to equity ratio	Numerical: Measure of ratio of debt to total capital
8	Creditors turnover	Numerical: Time period to pay the creditors
9	Total income	Numerical: Sum of revenue generated
10	EPS (Earnings per share)	Numerical: Estimate of profit for an organization
11	Loans and advances	Numerical: Savings to earn interest for lending purpose
12	Investment	Numerical: An act of purchase to generate income in future
13	Dividend rate	Numerical: Estimate of dividend payments
14	Equity face value	Numerical: Amount to be paid after maturity of transaction
15	Credit rating	Categorical: {A, B, C, D}

Table 1: Description of the financial attributes selected for the year 2013-2014 retrieved from CMIE Prowess, Ace Analyser and CRISIL databases

3.2. Data Pre-processing

The financial descriptors identified from the dataset were subjected to correlation analysis. The inter relationship and distribution of variables were analyzed using Pearson correlation coefficient. Based on the correlation matrix, redundant features and outliers were identified. A threshold score of 0.7 was set to remove the attributes with similar behavior from the dataset. The correlated features from dataset were further standardized using min-max normalization technique. Normalization was performed to scale the range of data from zero to one. Once the dataset was normalized, it was partitioned in training and test set respectively. The training set comprised on 2/3rd of the dataset to build the model, while the test set comprised of 1/3rd of the dataset to predict the performance of the models.

3.3. Model building using soft computing techniques

Two different approaches have been employed in this study for building credit rating models which are as follows:

3.3.1. Model building using classification techniques

The pre-processed dataset was subjected to develop scoring models by using different classification techniques. The algorithms used for developing the model are logistic regression, Bayesian classifier, adaboost, bagging, support vector machine, neural network. These algorithms were implemented in R by invoking packages including glmnet (logistic regression), e1071 (Naïve Bayes, SVM), adabag (AdaBoost and Bagging), neuralnet (Artificial neural network). **3.3.2. Model building using clustering technique**

The pre-processed dataset was subjected to unsupervised clustering techniques. Clustering technique was chosen for the dataset as it is a multi-class classification problem, making most of the binary classifiers unsuitable for the prediction. The different clustering techniques employed are discussed below:

3.3.2.1. Hierarchical clustering: This method is based on hierarchical organization of similar clusters. It may be either bottom up (Agglomerative) or top down (Divisive) approach. Dissimilarity is computed among the objects based on different metrics like Euclidean distance, Manhattan distance and linkage gauge (Sathya Varathan et al, 2012). A dendrogram is used to visualize the clusters. The technique is best employed for small datasets as it is time consuming.

***3.3.2.2. K-means clustering*:** This algorithm employs to partition the objects into k clusters based on the mean value. The initial value of k is randomly chosen. Based on k value the clusters are formed by assigning the objects with the nearest mean. The centroid of these clusters is evaluated which is assigned as the mean value for next iteration (Wong et al, 1979).

***3.3.2.3. Fuzzy C-means clustering*:** This algorithm is similar to that of k-means. In fuzzy c-means each data point is assigned to a cluster with fractional membership. The membership function that computes a minimal distance to its centroid to considered to be optimal (Zhong-dong Wu et al, 2003). It is also referred as soft clustering technique due to the association between elements at membership level.

***3.3.2.4. Density based clustering*:** This technique defines a cluster based on density distribution metric. The algorithm employed for the application is DBscan. Initially, the algorithm estimates the distance for a point along its neighbourhood. The radius of neighbourhood objects are computed and called as 'eps'. The number of

neighbours along the radius are computed and referred as 'minpts'. The algorithm is associated with certain extent of outliers from the data (Arthur et.al, 2011).

The above algorithms were implemented in statistical programming language R. R programming language is an open-source data mining framework which consists of enormous packages to perform common data mining tasks (www.r-project.org). The packages utilized in this study are represented in Table 3 respectively.

4. Deployment of clustering technique as a GUI in R

The clustering techniques namely k-means and hierarchical were deployed as a Graphical User Interface in R. The GUI was implemented using 'gWidgets' (http://cran.r-project.org/web/packages/gWidgets/gWidgets.pdf) and 'RGtk2' (http://cran.r-project.org/web/packages/RGtk2/RGtk2.pdf) packages. The GUI was implemented to predict the credit ratings of SMEs in India based on clustering techniques.

5. RESULTS AND DISCUSSION

5.1. Indian SME Dataset

The financial data of the SMEs and their respective credit ratings were retrieved from various sources including CMIE prowess and Ace Analyser Database for the financial year 2013-2014. The dataset comprised of fourteen different financial ratios along with the credit rating of companies. It is noted that CRISIL estimated the SME ratings for about 1452 companies for the financial year 2013-14.

The sample SME dataset was selected from the total population of companies rated by CRISIL for the financial year 2013-14, by using stratified random sampling technique (Tyrrell Sidney et al, 2001). The population was divided into different strata (a stratum refers to the Credit Class here e.g. A, B, C, D) and random subset was sampled from the entire population based on the characteristics of the dataset.

The SME dataset encompassed of 252 companies from different sectors including paper mills, cotton mills, food sectors, chemical industries and pharmaceutical firms. The dataset is described in Table 1. The distribution of credit ratings for Indian SME sector is shown in Figure 2.

5.2. Preprocessing the dataset

The SME credit rating dataset was subjected initial pre-processing. Correlation analysis was performed to analyze the relationship between the different financial estimates. The Pearson correlation coefficient was computed by using an in-built function in R namely cor ( ) and the plot ( ) function in R was used to plot the correlation matrix. The features having correlation greater than 70% were ignored for the subsequent steps. The correlation plot (as shown in Figure 3) estimates the relationship between the attributes and here each attribute is represented as a circle in the plot. The plot represents only ten features out of the fourteen features in the dataset in this view. The observation demonstrates a negative correlation between the attributes. Based on the ranking by correlation filters the redundant features were eliminated from the dataset. The attributes selected out of initial fourteen estimates are shown in Table 2. These attributes were subjected to normalization in the subsequent steps. Normalization was performed to obtain normal distribution of data that is scaled between zero and one. Min-max method was employed for normalization purpose. This method using the following formula:

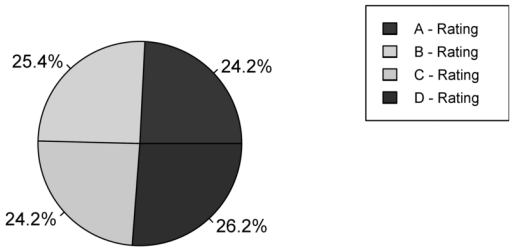

Figure 2: The distribution of SME Credit ratings

$$Normalization\,(e_i) = \frac{e_i - E_{\min}}{E_{\max} - E_{\min}}$$

From Equation 1,

e_i = the normalized value for each financial estimate

E_{min} = minimum range for the estimate (taken as zero)

E_{max} = maximum range for the estimate (taken as one)

Normalization was performed in R based on the above formulae resulting in normalized values for each attribute in the dataset. The pre-processed dataset with the reduced number of attributes was used for building the predictive model in further steps.

Sl. No	Selected financial attributes
1	Amount
2	Total capital
3	Debt
4	Total income
5	Debt equity ratio
6	EPS
7	Credit turnover
8	Net fixed assets

Table 2: Selected financial parameters after correlation

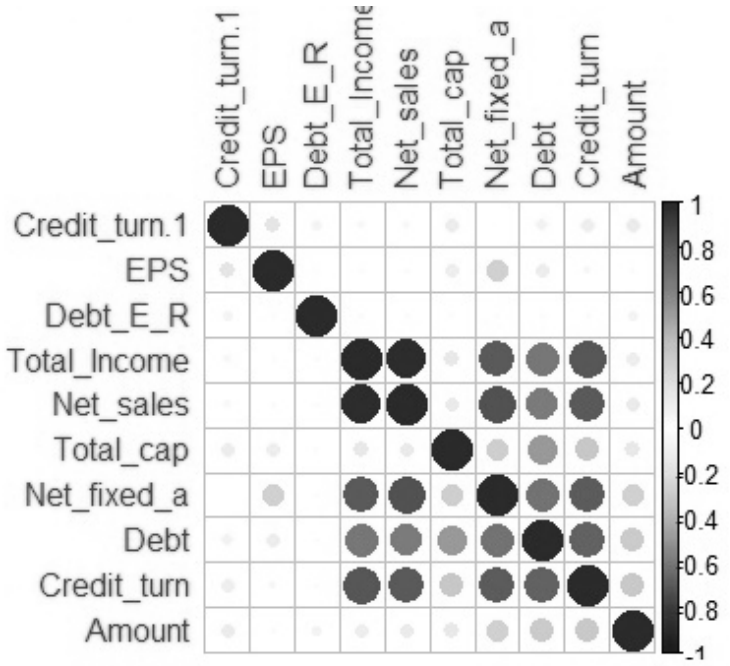

Figure 3: The correlation plot depicting the importance of features in SME dataset.

5.3. Performance of the algorithms

Ten different classification and clustering techniques were implemented for building an accurate model for credit rating of the Indian SME dataset. The accuracy of these models is described in Table 3.

Sl. No	Algorithm	Package used in R	Accuracy of prediction (%)
1	Logistic regression	glmnet	53.68
2	Bayesian classifier	e1071	54.99
3	AdaBoost	adabag	55.1
4	Bagging	adabag	55.96
5	Support vector machine	e1071	64.93
6	Neural network	neuralnet	66.67
7	k-means clustering	cluster	77.14
8	Hierarchical clustering	cluster	77.86
9	Fuzzy c means clustering	e1071	58.95
10	Density-based clustering	fpc	56.54

Table 3: Accuracy of algorithms for SME dataset

5.4. Hierarchical and k-means clustering Analysis

Based on the accuracy (Table 3), it is observed that two clustering techniques namely, k-means and hierarchical clustering have outperformed the other data mining techniques. These two techniques are further explored in this study. The companies were clustered into four groups based on their credit ratings A, B, C and D in these techniques. The distribution of financial assets among the clusters is shown in Figure 4. In

this figure, each row represents one feature. The range is defined for each feature based on the distribution of data. The dependent variable (i.e. credit ratings) is represented as circles for each feature. The figure demonstrates four different colored circles for each rating category (i.e. A, B, C and D).

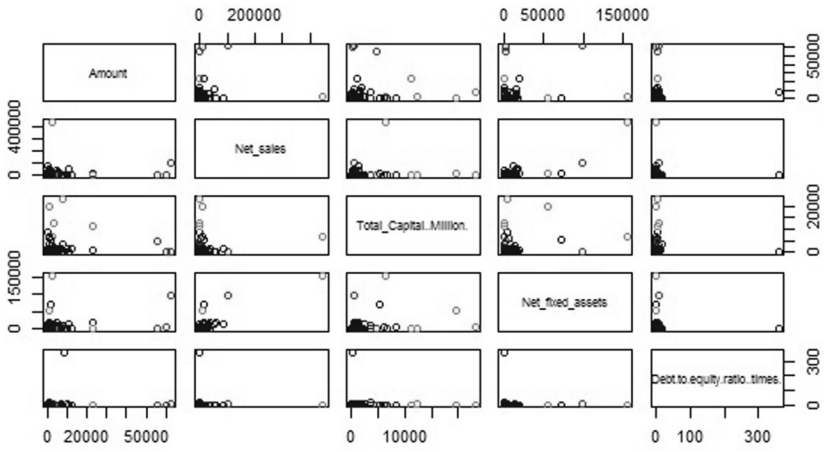

Figure 4: 5.4.1. Hierarchical clustering analysis: The technique was employed using 'cluster' package in R. The "Ward's agglomerative clustering technique"

5.4.1. ***Hierarchical clustering analysis:*** The technique was employed using 'cluster' package in R. The "Ward's agglomerative clustering technique" was applied for hierarchical clustering method (Fionn Murtagh, 2010). The clusters are visualized using a dendrogram which is a tree like diagram representing the predicted credit ratings along its height. The dendrogram is shown in Figure 5. The pair wise dissimilarity between features is computed based on Euclidean distance. The terminals formed after computing dissimilarity are termed as leafs. Each leaf represents the estimate of credit rating for a company. Each branch generated from the leaf is termed as a clade. The clades represent the estimate of similarity between different features. The height of each clade represents the measure of dissimilar or similar characteristics between the features. The dendrogram represents credit rating for each SME company based on the features explored.

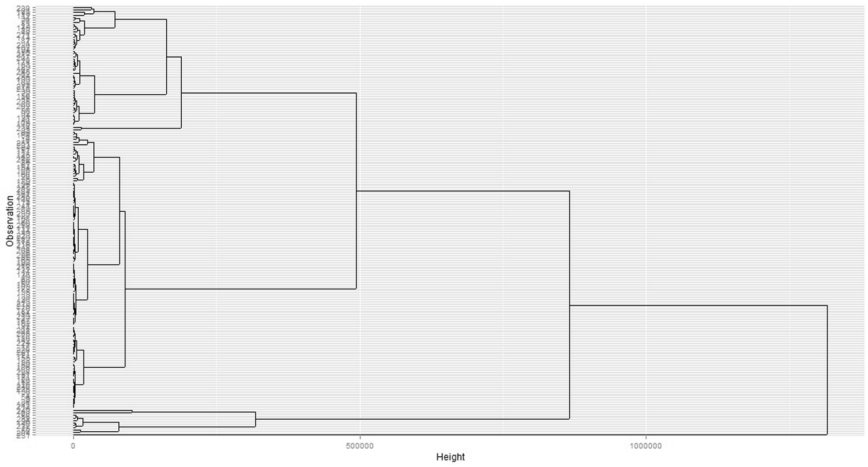

Figure 5: The predicted dendrogram for the SME's credit rating

5.4.2. K-means clustering analysis: K-means clustering approach was implemented using 'cluster' package in R. The K-means clustering was executed by setting the value of k to be four since the data has four categories. The clusters so formed represent the predicted credit ratings for each company based on the financial attributes. The discriminant plot representing the clusters obtained from hierarchical and k-means clustering is depicted in Figure 6. The plot reveals the pattern of clustering for the credit rating dataset. Each plot consists of four clusters representing the four classes of credit rating respectively. The clustering pattern differed for both the techniques. The credit ratings are represented in numeric values which are interpreted as seen in Figure 6.

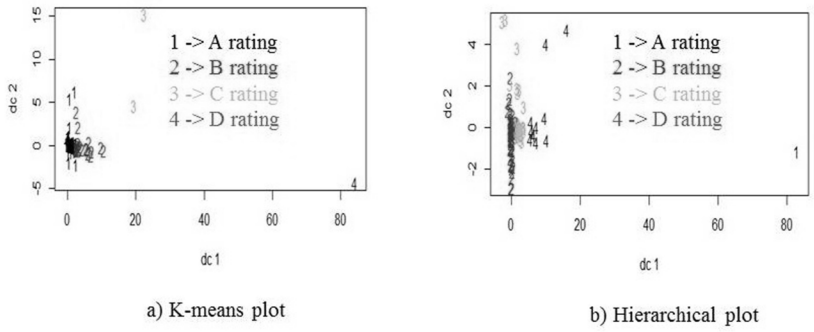

Figure 6: Discriminant plot of clustering techniques. 6(a): The discriminant plot of k-means clustering. 6(b): The discriminant plot of Hierarchical clustering technique

6. Deployment of clustering algorithms in a GUI

The Graphical User Interface (GUI) was developed by deploying k-means and hierarchical clustering algorithms. The tool facilitates quantification of the credit worthiness of a company based on selected financial variables. The step-wise procedure to use the predictive tool is depicted in Figure 7. The implementation

of the clustering tool for prediction of credit rating of SME companies is shown in Figure 8. The tool is designed to be user-friendly which can be effectively utilized by SME governing agencies to gauge the financial performance of the companies.

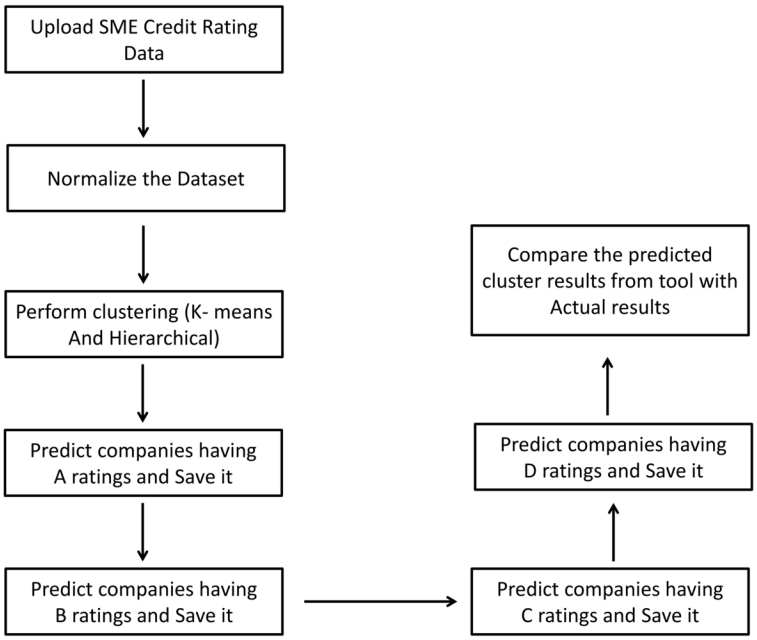

Figure 7: The step wise procedure representing the development of clustering tool

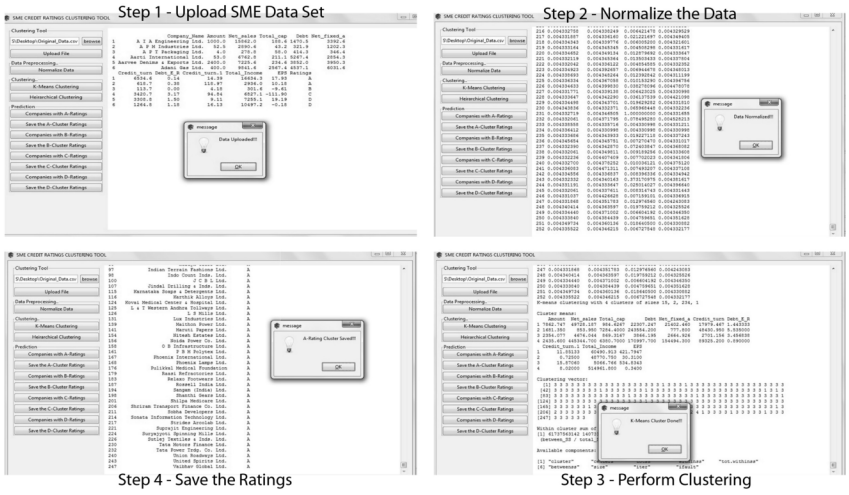

Figure 8: The step wise operations performed by the clustering tool for the prediction of credit ratings

7. CONCLUSION

The results suggest that clustering techniques used for SME dataset seems to be an effective approach to predict the credit ratings. To further serve the rating agencies and councils to estimate credit ratings for SMEs, a tool is developed based on clustering techniques. The tool can predict credit rating for a new SME company based on its financial attributes using clustering techniques. The tool can also be used in mobile platform like tablet devices which requires preloaded R environment. The managerial personnel in financial organizations such as Banks, Industrial Development Financial Corporations, and Payment Banks, Non-Banking Financial Companies can utilize this predictive decision-support system as a screening tool for filtering customers. The predictive system thereby reduces the time and cost for the managers by providing a swift platform of assessment of SME credit rating. However, the proper usage of the results of the tool lies solely in the hands of the user, and he can use his discretion to either approve or reject the assessment of this predictive system.

8. LIMITATIONS AND FUTURE SCOPES

The credit rating models normally employed by credit rating agencies and financial institutions essentially takes into account many key factors that are grouped as industry risk, business risk, financial risk and management risk. Each type of risk is given a specific weightage and financial factors are normally given an estimate weightage of about 40% in the total credit rating model. The current study is solely focused on building a credit rating models only based on financial risk factor, while other risk based parameters are not considered in this study. This is one of the limitations of the current study.

It is beneficial to select data pertaining to this other risk measures from reports of credit rating agencies to develop an automated credit rating model. The problem in collectingdata which includes all the risk based parameters from banking or other financial institutions is that it breaches the privacy of the individuals and there is an inherent risk of exploiting such data. To overcome this issue it is suggested to collect sanitized data which has eliminated details revealing the identity of the customer before given to third party researchers for model building.

The prediction tool can be further used in future as mobile based decision support system for real time SME credit rating datasets deployed in cloud architecture. Big data technologies is another promising approach for predicting dynamic response on massive credit rating datasets using the advantage of distributed computing.

References

1. A. Belkaoui., (1980). Industrial bond ratings: a new look, Financial Management, Volume 9, Issue 3, pages: 44–51.
2. A. K. Jain., R. C. Dubes., (1988). Algorithms for Clustering Data, Printice Hall.
3. AC Antonakis., ME Sfakianakis., (2009). Assessing naïve Bayes as a method for screening credit applicants, Journal of Applied Statistics, Volume 36, Issue 5, pages: 537-545.
4. Alenjandro Correa et.al. (2012). Constructing a credit risk score card using predictive clusters, SAS Global Forum: Data mining and Text analytics, pages: 1-13.
5. Alfredo Cuzzocrea., Jerome Darmont., Hadj Mahboubi., (2009). Fragmenting very large XML data warehouses via K-means clustering algorithm, Int. J. of Business Intelligence and Data Mining, Volume 4, Issue 3/4, pages: 301-328.
6. Anita Prinze., D Poel., (2008). Random forests for multiclass classification: Random Multinomial Logit, Expert System with Applications: An International Journal, Volume 34, Issue 3, pages: 1721-1732.
7. Anthony Brabazon., Michael O Neill., (2005). Biologically inspired algorithm for Financial Modelling, Natural Computing Series, pages: 230-247, Springer Verlag Berlin Heidelberg.

8. Arthur O Sullivan., Steven M Sheffrin., (2003). Economics: Principles in action, page:1-512, Pearson.
9. Caterina Liberati., Furio Camillo., Gilbert Saporta., (2015). Advances in credit scoring: combining performance and interpretation in kernel discriminant analysis, Advances in Data Analysis and Classification.
10. Cheng-Lung Huang, Mu-Chen Chen., Chieh-Jen Wang., (2006). Credit scoring with a data mining approach based on support vector machines, Expert Systems with Applications, Volume 33, Issue 4, pages: 847-856.
11. CRISIL: http://www.crisil.com/index.jsp (Accessed on June, 2016).
12. Crook J N., (1996), Credit scoring: An overview, The University of Edinburgh.
13. Dr. Kamaleshkumar Patel., (2011). Development of credit risk model for Bank Loan ratings, International Journal of Research in Commerce, Economics and Management, Volume 1, Issue 6, pages: 112-116.
14. Erin Allwein., Robert Shapire., Yoram Singer., (2000). Reducing multiclass to binary: A unifying approach for margin classifiers, Journal of Machine Learning Research, Volume 1, pages: 113–141.
15. Fionn Murtagh., (2014). Wards Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Wards Criterion?, Journal of Classification, Volume 31, Issue 3, pages: 274-295.
16. G. E. Pinches., KAA Mingo., (1973). Multivariate analysis of industrial bond ratings, The Journal of Finance, Volume 28, Issue 1, pages: 1–18.
17. G. Vairava Subramanian., S. Nehru., (2012). Implementation of credit risk for SME-How is beneficial to Indian SMEs?, International Journal of Scientific and Research Publications, Volume 2, Issue 4, pages: 1-7.
18. Hand D. J., Jacka S. D., (1998). Statistics in Finance, Arnold, Applications of Statistics.
19. Hartigan J A., Wong M. A., (1979). Algorithm AS 136: A K-Means Clustering Algorithm, Journal of Royal Statistical Society, Volume 28, Issue 1, pages: 100–108.
20. Hunt Neville., Tyrrell Sidney., (2001). Stratified Sampling, Webpage at Coventry University. (Accessed on 20 March 2015).
21. ICRA: http://www.icra.in/ (Accessed on June, 2016).
22. Jiaqi Wang., Xindong Wu., Chengqi Zhang., (2005). Support vector machines based on K-means clustering for real-time business intelligence systems, Int. J. of Business Intelligence and Data Mining, Volume 1, Issue 1, pages: 54-64.
23. JW Kin., HR Weistroffer., RT Redmond., (1993). Expert systems for bond ratings: a comparative analysis of statistical, rule-based and neural network systems, Expert Systems, Volume 10, Issue 3, pages: 167-172.
24. K Polat., Salih Gunes., (2009). A hybrid intelligent method based on C4.5 decision tree and one-against-all approach for multi-class classification problem, Expert System with Applications: An International Journal, pages: 1587-1592.
25. Kriegel., Hans-Peter., Arthur et.al. (2011). Density-based Clustering, WIREs Data Mining and Knowledge Discovery, Volume 1, Issue 3, pages: 231–240.
26. L. Cao., L. K. Guan., Z. Jingqing., (2006). Bond rating using support vector machine, Intelligent Data Analysis, Volume 10, Issue 3, pages: 285–296.
27. L. H. Ederington. (1985), Classification models and bond ratings, *The Financial Review*, Volume. 20, Issue 4, pages: 237–262.
28. Lee T., Chen I., (2005). A Two-Stage Hybrid Credit Scoring Model Using Artificial Neural Networks and Multivariate Adaptive Regression Splines, Expert Systems with Applications, Volume 28, Issue 4, pages: 743-752.
29. Ling Tan., David Taniar., Kate A. Smith (2005). A clustering algorithm based on an estimated distribution model, Int. J. of Business Intelligence and Data Mining, Volume 1, Issue 2, pages: 229-245.
30. Liran Einav., Mark Jenkins., Jonathan Levin., (2013). The impact of credit scoring on consumer lending, RAND Journal of Economics, Volume 44, Issue 2, pages: 249-274.
31. M Jayadev., (2006). Predictive Power of Financial Risk Factors: An Empirical Analysis of Default

Companies, Research, Volume 31, Issue 3, pages: 45-56.
32. M. R. Anderberg., (1973). Cluster Analysis for Applications, Academic Press.
33. Malini Goyal.(2013), SMEs employ close to 40% of Indias workforce, but contribute only 17% to GDP, Economic Times: http://articles.economictimes.indiatimes.com/2013-06-09/news/39834857_1_smes-workforce-small-and-medium-enterprises(Accessed at March, 2016).
34. Micro, Small and Medium Enterprises (2016), Reserve Bank of India [online], https://rbi.org.in/Scripts/FAQView.aspx?Id=84 (Accessed on feb, 2016).
35. Mohamed Aly., (2005). Survey on Multiclass Classification Methods, pages: 1-9, Available at: http://vision.caltech.edu/malaa/publications/aly05multiclass.pdf (Accessed on March, 2016).
36. Mohammad Farajian., Shahriar Mohammadi., (2010). Mining the banking customer behavior using clustering and association rule methods, International Journal of Industrial Engineering and Production Engineering, Volume 21, Issue 4, pages: 239-245.
37. Mohammad Gholamian., Saber Jahanpour., Seyed Sadatrasoul., (2013). A New Method for Clustering in Credit Scoring Problem, Journal of Mathematics and Computer Science, Volume 6, pages: 97-106
38. Orgler Y E., (1971). Evaluation of Bank Consumer Loans with Credit Scoring Models, Journal of Bank Research, Volume 2, Issue 1, pages: 31-37.
39. Package RGtk2. Available at:http://cran.r-project.org/web/packages/RGtk2/RGtk2.pdf (Accessed on June, 2016)
40. Package gWidgets. Available at:http://cran.r-project.org/web/packages/gWidgets/gWidgets.pdf (Accessed on June, 2016)
41. Petr Hájek., (2011). Credit rating analysis using adaptive fuzzy rule-based systems: an industry-specific approach, Central European Journal of Operations Research, Volume. 20, Issue 3, pages: 421-434.
42. Prowess Database: https://prowess.cmie.com/ (Accessed on March, 2016)
43. R Programming: http://www.r-project.org (Accessed on March, 2016)
44. S. Dutta., S. Shekhar., (1988). Bond rating: a non-conservative application of neural networks, Proceedings of the IEEE International Conference on Neural Networks, Volume 2, pages: 443–450.
45. Sathya Varathan., Priya Kalyanasundaram., S. Tamilenthi. (2012). Credit policy and credit appraisal of canara bank using ratio analysis, International Multidisciplinary Research Journal, Volume 2, Issue 7, pages: 19-28.
46. Shin-Chen Huang., Min-Yuh Day., (2013). A comparative study of data mining techniques for credit scoring in banking, International Conference on Information Reuse and Integration, pages: 684-691.
47. Shu-Hsien Liao., Pei-Hui Chu., Pei-Yuan Hsiao., (2012). Data mining techniques and applications – A decade review from 2000 to 2011, Expert Systems with Applications, Volume 39, pages: 11303-11311.
48. Silvestro Montrone., Paola Perchinun., (2015). The identification of fuzzy profiles through the c-means clustering, Int. J. of Business Intelligence and Data Mining, Volume. 10, Issue 1, pages: 62-72.
49. SMERA: http://www.smera.in/ (Accessed on June, 2016).
50. Steenackers A., Goovaerts M J., (1989). A Credit Scoring Model for Personal Loans, Insurance, Mathematics and Economics, Volume 8, Issue 8, pages: 31-34.
51. Subhagata Chattopadhyay., Dilip Kumar Pratihar., S.C. De Sarkar (2007). Some studies on fuzzy clustering of psychosis data, Int. J. of Business Intelligence and Data Mining, Volume 2, Issue 2, pages: 143-159.
52. Sustersic M., Mramor D., Zupan J., (2009). Consumer credit scoring models with limited data, Expert Systems with Applications, Volume 36, Issue 3, pages: 4736-4744.
53. Thomas L C., (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers, International Journal of Forecasting, Volume 16, Issue 2, pages: 149-172.
54. Ward., Joe H., (1963). Hierarchical Grouping to Optimize an Objective Function, Journal of the American Statistical Association, Volume 58, Issue 301, pages: 236–244.
55. Wei Li., Jibiao Liao., (2011). An Empirical Study on Credit Scoring Model for Credit Card by Using Data Mining Technology, Seventh International Conference of Computational Intelligence and Security, pages: 1279-1282.
56. Xi Yue Zhou., DeFu Zhang., Yi Ziang., (2008). A new credit scoring method based on rough sets and

decision tree. Advances in knowledge discovery and data mining, pages: 1081-1089.
57. Y Yang., (2007). Adaptive credit scoring with kernel learning methods, European Journal of Operational Research, Volume 183, Issue 3, pages: 1521–1536.
58. Yannis Marinakis., Magdalene Marinaki., Nikolaos Matsatsinis., (2008). A stochastic nature inspired metaheuristic for clustering analysis. Int. J. of Business Intelligence and Data Mining, Volume 3, Issue 1, pages: 30-44.
59. Z Huang., H Chen., C J Hsu.. (2004). Credit rating analysis with support vector machines and neural networks: a market comparative study. Decision Support Systems. Volume 37, Issue 4, pages: 543–558.
60. Zhong-dong Wu., Wei-xinXie., Jian-ping Yu., (2003). Fuzzy C-means clustering algorithm based on kernel method. Proceedings in Fifth International Conference on Computational Intelligence and Multimedia Applications, pages: 49-54.