Project Dissertation Report on

LOGISTIC REGRESSION BASED FRAMEWORK FOR CREDIT RISK DECISION

Submitted By:

AMARJEET KUMAR

(2K17/MBA/10)

Under the Guidance of:

Dr. G.C. Maheshwari

Professor

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road, Delhi 110042

CERTIFICATE FROM THE INSTITUTE

This is to certify that the project "Logistic Regression Based Framework For Credit Risk Decision", to the best of my knowledge is a bonafide work carried out by Mr. Amarjeet Kumar of MBA 2017-2019 batch and has been submitted to Delhi School of Management, Delhi Technological University, Bawana Road, Delhi 11oo42 in partial fulfilment of the requirement for the award of the degree of Masters of Business Administration.

Signature of Guide

Dr. G.C. Maheshwari

Place:

Date:

Signature of HOD(DSM)

Seal of HOD(DSM)

DECLARATION

I, Amarjeet Kumar, student of MBA 2017-2019 batch of Delhi School of Management, Delhi Technological University, Bawana Road, Delhi 110042 declare that the project dissertation report on “Logistic Regression Based Framework For Credit Risk Decision” has been submitted in partial fulfilment of degree of Masters of Business Administration is the original work conducted by me.

The information and data given in the report is authentic to best of my knowledge.

This report is not being submitted to any other university for award other degree diploma or fellowship.

Place: Amarjeet Kumar

Date (2K17/MBA/10)

ACKNOWLEDGEMENT

At the outset, I express my heartfelt thanks & gratitude to those who sincerely helped and supported me throughout the project I have undertaken to do & without whose active support & help it would not have been possible for me to complete the venture. As such, I once again extend my sincere thanks & gratitude to all of them.

I gratefully acknowledge my profound indebtedness towards my esteemed guide, Dr. G. C Maheshwari, Professor, Delhi School of Management, Delhi Technological University for his invaluable guidance, excellent supervision and constant encouragement during the entire duration of the project work. This project would never have been possible without his guidance and supervision.

Moreover, I would also like to express my heartfelt gratitude to Dr. Archana Singh, Assistant Professor, DSM(DTU) for her kind co-operation, guidance and encouragement which also helped me a lot in completing the project, the constant support from all the others members of the office has really helped me finish the project.

I am also thankful to Dr. Rajan Yadav, Head of Department and all the faculty members of Delhi School of Management, DTU, Delhi.

Finally, I would also like to express my earnest gratitude to my friends and family members for their constant support & encouragement without which the assignment would not have been completed, besides the constant blessings of Almighty.

EXECUTIVE SUMMARY

Risk is said to be an uncertainty of occurrence of economic loss. Credit risk is one of the most important topics in risk management, it is the risk of default on a debt that may arise from a borrower failing to make the required payments. In this project, I focus on the credit risk problem at the firm level. I try to identify key financial ratios that would help to distinguish between credit worthy companies which are unlikely to default and less credit worthy companies which are more likely to default in India based on the credit ratings given by various credit rating agencies. I consider the organizations with credit ratings of "Baa2" or higher to be stable and the organizations with credit ratings lower than "Baa2" to be unstable.
A framework of multinomial logistic regression is used to identify the key financial factors from a pool of 33 financial ratios.

This model will help the organizations to mitigate the losses by giving loans to the companies which are less likely to default.

CONTENTS

1. INTRODUCTION

Bonds provide a critical mechanism for companies to raise funds to finance new and continuing activities and projects. Corporations raise substantially more capital in the bond market each year than they do in the equity market. In 1993 companies raised $444 billion in the bond market compared to $102 billion in the equity market (*Investment Dealers' Digest,* 1994). The process of raising new cash for the corporation to utilize in its operations initiates within the management of the organization, but must then proceed through various outside parties.

The typical process begins with the corporation obtaining a necessary bond rating for the issue from a major rating agency, such as Moody's or Standard & Poor's, and ends with an investment banker bringing the issue to market. An important part of raising money in a free market is the assessment of the organization by independent parties. The independent bond rating agency examines the financial outlook of the company and the characteristics o f the issue, and assigns a rating that indicates an assessment of the degree of default risk associated with the firm's bonds. Essentially, the bond rating is an attempt to inform the public (with no guarantees) of the likelihood of an investor receiving, as scheduled, all the promised interest and principal payments associated with the bond issue. The issue is assigned to a rating class that indicates the perceived quality (or riskiness) +of the bonds. These categories begin at the highest quality issues (e.g. Aaa) and proceed down to lower-quality issues (e.g. Aa, A, Baa, Ba, etc.). The company obtains the actual rating by contacting the bond rating agency prior to issuance and requesting a rating be assigned to their new issue. The rating agency then assigns an analytical team to conduct basic research about the company and individual issue characteristics. It meets with the issuer to obtain any additional information it may deem pertinent. Finally, there is a rating committee meeting which results in an issued rating. This rating is monitored for the life of the issue and may be upgraded or downgraded at any time by the rating agency. Although the bond rating agency receives a fee for the service, the actual rating received is not dependent upon the amount of the fee. The rating agency's very existence depends upon being independent, along with the associated credibility the public attributes to the ratings it issues.

Risk measures future uncertainty about deviation from expected earnings or expected outcome. Risk implies the uncertainty that an investor is willing to take to realize a gain from an investment. Risks are of different types and originate from different situations such as, liquidity risk, sovereign risk, insurance risk, business risk, credit risk, etc. In this paper, we focus on credit problem at the firm level. Credit risk is defined as the risk of loss of principal or loss of a pecuniary reward stemming from a borrower's failure in repaying a loan or else wise to meet a contractual debt. Credit risk arises when a borrower is looking ahead to use future cash flows through the payment of a current obligation. The investors are rewarded for presuming credit risk through the way of interest payments from the issuer or from the borrower of a debt contract. Credit risk occurs due to: volatility in the difference between investment's interest rates and the risk-free return rate; borrowers are not able to make contractual payments; resulting from the downgrades in rating the risk of an issuer.

Unlike the bankruptcy prediction, this project emphasizes on the probability of default of organizations applying for loans. We try to identify key financial factors that helps to distinguish between creditworthy companies (CWCs) which are unlikely to default and less creditworthy companies (LCWCs) which are more likely to default in India based on the long-term credit ratings provided by various credit rating firms e.g. Moody's, ICRA, CARE etc. A framework of factor analysis and multinomial logistic regression is used to identify key financial factors from the pool of 33 financial ratios of 15 Indian companies.

Logistic regression has been a reliable tool in many Statisticians/Economists toolkit for many years when dealing with binary problems where the output is o/1, True/False, or any variation of a dichotomous problem. But the reality is that Multinomial Logistic regression is a very important 'algorithm' in the machine learning sphere.

Multinomial logistic regression is an extension of the binary logistic regression which allows for more than two categories of the dependent or outcome variable. While Logistic regression is commonly used for discrete binary problems, Multinomial Logistic regression is built with an eye towards multi-class classification or regression problems.

A Logistic classifier uses either a Sigmoid or Softmax function (both are variations of the commonly known Logistic function):

- **Sigmoid function:** binary classification or regression using logistic regression model.

$$Sigmoid\ Function = f(x_i) = \frac{1}{1 + e^{-x_i}}$$

- **Softmax function:** multi-classification or multinomial regression using multinomial logistic regression model.

$$Softmax\ Function = f(x_i) = \frac{e^{x_i}}{\sum_{i=0} e^{x_i}}$$

Multinomial Logistic regression models are ideal for forecasting credit migration matrices. The model can use effect of independent variables and predict the probabilities of different possible outcomes.

Disadvantages:

- Logistic Regressions do not perform well when feature space is too large
- Doesn't handle large number of categorical features/variables well
- Relies on transformations for non-linear features

Credit rating models can be used as a guideline when evaluating unrated firms. One of the best-known models in this area is applied by E. I. Altman (1968), whose default model is often used as a tool in financial analysis of a company. This model has the ability to identify companies with the possible financial problems and was proposed on the basis of multivariate discriminant analysis. The other research in this area is primarily focused on bond rating and bond rating models.

The proposed model will help the companies to easily identify to which organizations they should lend their money. Better credit risk management presents an opportunity for organizations to improve overall performance and secure competitive advantage.

2. LITERATURE REVIEW

Credit risk is one of the most important topics in the management of Risk. As claimed by the Basel capital accord it is the major risk of banks and financial institutions encountered (Stephanou and Mendoza, 2oo5). The bankruptcy prediction is a well-known estimation method for measure credit risk. Many studies in the literature reported models that can predict the failure of the firm or if the firm is going to bankrupt in near future with some key financial ratios selected from a number of candidate financial ratios (Jun Huang and Haibo Wang, 2o17). Much effort has gone while we are measuring the various parameters of credit risk. Various Institutes dealing with credit risk, a serious concern arises from the fact that credit risk has both an unusual feature and a systematic component. Various counter-party default may various factors which are unique to the borrower, such as poor management and bad luck. It also may arise in the wider contexts of political turmoil, financial market crashes and economic recessions.

Banks always implement a credit risk analysis before making new loans to the potential client (Inderst, Mueller, 2008). A potential client's credit risk level in banks can be evaluated by many of the internal credit risk assessment models. The main aim of these models is to determine whether the potential client has the capacity to repay the loan or not. This is normally done using historical data and various other statistical techniques (Emel, Oral, Reisman, Yolalan, 2003). The primary issue of the credit risk research is to determine what variables significantly influences the probability of default. A second and one of the main important issue is the construction of credit scoring model (Marshall, Tang, Milne, 2010).

(The seminal study by Altman 1968). Altman adopted the discriminant analysis (DA) to select financial ratios which then were used for firm-level bankruptcy prediction. The 22 potential ratios in Altman's paper were grouped into five financial factors which were as follows: - profitability, liquidity, solvency, activity and leverage ratios. Altman (1968) selected five features (financial ratios) with multiple discriminant analysis (MDA) and then they predicted firm-level bankruptcy with the selected ratios. (Leshno and Spector 1996) They selected 29 financial ratios out of 7o ratios and evaluated the prediction capability of various neural network (NN) models which were differed in terms of data span, number of iterations and neural network architecture.

(Frydman et al 1985) analyzed financial distress based on 200 firms with 20 financial ratios, and they introduced recursive partitioning algorithm (RPA), a non-parametric technique based on pattern recognition, to improve the classification accuracy. Shin et al. (2005) selected 52 ratios out of more than 25o financial ratios using independent-samples t-test in the first stage. They further selected 1o ratios by MDA stepwise method and evaluated the predictive performance of bankruptcy based on the selected ratios with support vector machines (SVM). Ryu and Yue (2005) developed simple feature reduction techniques such as stepwise discriminant analysis (SDA), sequential elimination and mutual information-based feature selection to choose ratios from 23 financial ratios. They introduced a linear programming technique called isotonic separation (IS) to classify bankrupt and non-bankrupt firms. McKee and Lensberg (2002) used rough sets model to identify variables that are important for the prediction and construct a bankruptcy prediction model with a genetic programming (GA) algorithm.

Etemadi et al. (2009) selected 5 financial ratios out of the 43 candidate ratios with the DA. Prediction of corporate bankruptcy was then conducted by using a GA model. Min and Lee (2008) used data envelopment analysis (DEA) method for bankruptcy prediction. In their study, 57 financial ratios were classified into factors of profitability, growth, productivity, liquidity, activity and cost structure. The final six financial ratios were chosen based on judgment of the experts along with factor analysis (FA). Min and Jeong (2009) identified 9 ratios from 27 financial ratios based on various feature selection methods such as independent-samples t-test, DA, logistic regression (LR) and decision trees (DT). They proposed a binary classification method, solved with genetic approach, to classify firms into bankrupt and non-bankrupt according to the distance between a representative firm and an observation in data set, implying the similarity or non-similarity between them. Fedorova et al. (2013) first selected 75 financial ratios from 98 ratios with ANOVA test and then applied combinations of learning algorithms, including MDA, LR, classification and regression trees, to identify final financial ratios. These ratios were evaluated by two types of artificial neural networks (ANN) to derive the classification accuracy rate for the bankruptcy prediction.

2.1 STATISTICAL APPROACHES FOR CREDIT RATING PREDICTION

The three following methods are generally used, multinomial logistic regression, discriminant analysis and decision trees. All techniques are suitable for the problem of credit rating prediction, where there are more than two categories of dependent variable (such as five rating categories)

2.1.1. Multivariate discriminant analysis

Discriminant analysis is a common statistical method used for classification tasks a suitable method for credit rating modelling. The analysis can be used for two major objectives: i) description of group separation and ii) prediction or allocation of observations to groups. Discriminant functions are linear combinations of variables that best separate groups, for example the k groups of multivariate observations. For the following part of this paragraph, the explanation and definitions were taken from Rencher (2002) and Huberty and Olejnik (2006).

2.1.2. Multinomial logistic regression

The multinomial logistic regression is a modification of binary logistic, where only two possible outcomes can occur. The definitions and derivations used in this chapter were extracted from Hosmer and Lemeshow (2000). The model for dichotomous outcome variable is based on logistic distribution and we use the quantity $\pi(x) = E(Y|x)$ to represent the conditional mean of Y given x when the logistic distribution is used,

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}.$$

2.1.3. Decision trees

Using decision trees enables to create a tree-based classification model and the rules can be used for prediction purposes. Decision trees can graphically represent alternative choices that can be made and enable the decision maker to identify the most suitable option in a particular circumstance. A decision problem can be presented in the form of a matrix (table) or a tree. Decision trees conventions are described for example in Mian (2011, pp. 168). Rokach and Maimon (2008) states that a decision tree is a predictive model and can be used both for decision and classification problems. Classification trees can be used to classify an object or an instance (such as companies) to a predefined set of classes (rating groups). The companies are firstly classified according to the most relevant variable, then into subgroups according to other variable, and so on (Witzany, pp. 46, 2010). As Witzany (2010) says, the decision rules should maximize a divergence measure of the difference in default risk between the two subsets. The splitting is repeated until no group can be split into two subgroups which are statistically different. According to Wei-Yin (2008), there are three major tasks of a classification tree: (i) how to partition the data at each step, (ii) when to stop partitioning and (iii) how to predict the value of y for each x in partition. Common algorithms for decision tree induction include ID3, C4.5, CART, CHAID and QUEST (Rokach and Maimon, 2oo8).

CART (or CRT) refers to classification and regression trees. This algorithm splits the data into segments that are as homogenous as possible with respect to the dependent variable. A terminal node in which all cases have the same value for the dependent variable is a homogenous (pure) node. The extent to which a node does not represent a homogenous subset of cases is an indication of impurity. For categorical dependent variables such as rating, the Gini index, twoing or ordered twoing can be used as impurity measures.

CHAID refers to chi-squared automatic interaction detection. At each step, CHAID selects the independent (predictor) variable that has the strongest interaction with the dependent variable. If categories of each predictor are not significantly different with respect to the dependent variable, they are merged4. For each input attribute ai, CHAID finds the pair of values in Vi that is least significantly different with respect to the target attribute. The significant difference is measured by the p value obtained from a statistical test. The statistical test used depends on the type of target attribute. An F test is used if the target attribute is continuous; a Pearson chi-squared test if it is nominal; and a likelihood ratio test if it is ordinal (Rokach and Maimon, 2008). For each selected pair of values, the p value obtained is compared with a certain merge threshold. If it is greater, it merges the values and searches for an additional potential pair to be merged. It is repeated until no significant pairs are found. The best input attribute to be used for splitting the current node is then selected, such that each child node is made of a group of homogeneous values of the selected attribute. This procedure stops also when one of the following conditions is fulfilled: (i) maximum tree depth is reached; (ii) minimum number of cases in a node for being a parent is reached, so it cannot be split any further; (iii) minimum number of cases in a node for being a child node is reached (Rokach and Maimon, 2008).

3. DATA

In this study, we initiate with 33 financial ratios of 15 Indian companies. These financial ratios are calculated using annual reports published by the company. These ratios are grouped into eight financial factors, namely, profitability factor, management efficiency factor, solvency factor, investment valuation factor, cash flow indicator factor, debt coverage factor, liquidity factor and operating factor. A brief description about all the 8 factors are given as follows, while 33 financial ratios are shown in table below:

The factor on profitability explains how well companies used their existing resources to generate profit and value for shareholders.

The factor on management efficiency explains how well companies use its assets and manage liabilities effectively.

The factor on solvency assesses a company's ability to repay the debt and the interest.

The factor on investment valuation is used to compare and determine a better investment.

The factor on cash flow indicator explains the company's ability to generate cash and how well current liabilities are envelop by the generated cash flows.

The factor on debt coverage explains the generated revenue is enough to cover debt payments.

The factor on liquidity explains a company's ability to repay the short-term obligations.

The factor on operating explains how efficiently companies use their capital.

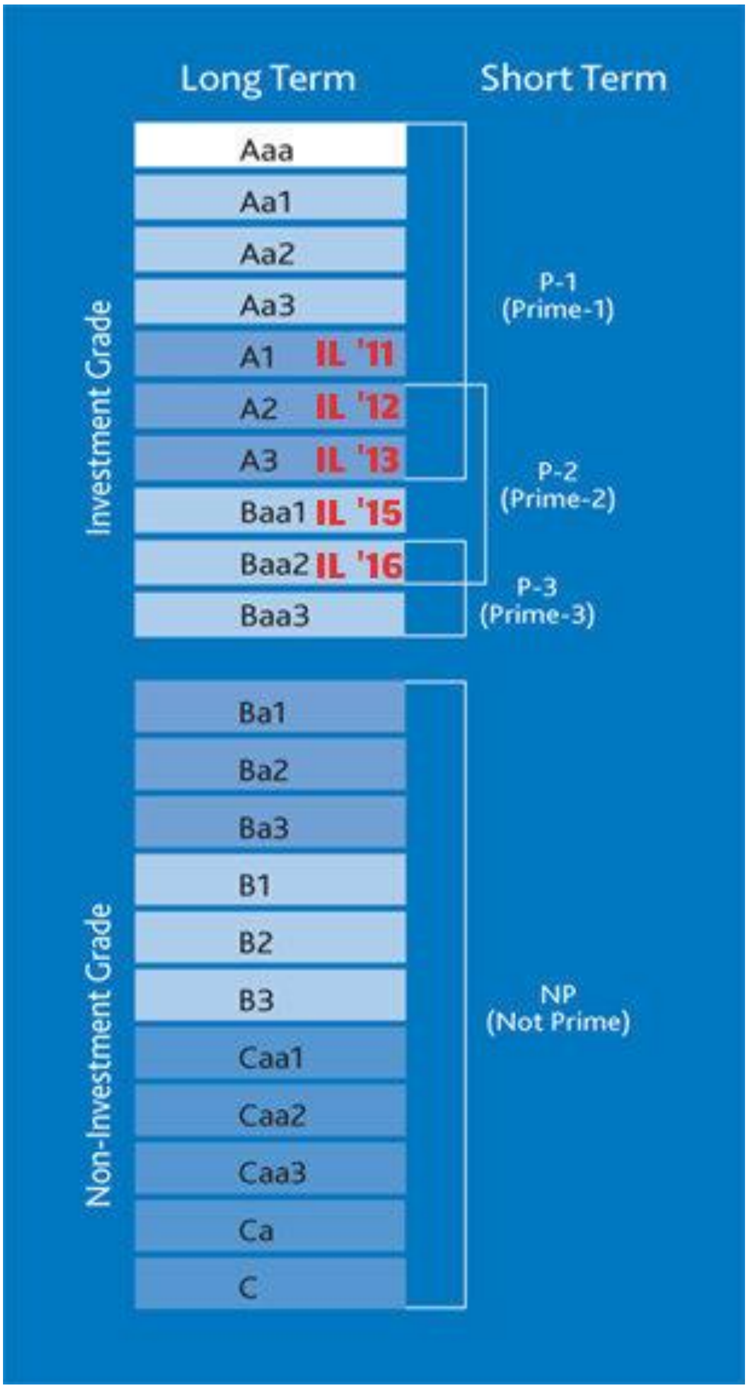
Table 1 List of Financial Factors Considered for Evaluation of Credit-worthiness

Financial Factor	Financial Ratios
Probability Factor	Gross Profit Margin
	Operating Profit Margin

	EBIT Margin
	Cash Profit Margin
	Net Profit Margin
	Adjusted Net Profit Margin
	Adjusted Cash Margin
Management efficiency factors	return on capital employed
	Return on Net-worth
	Adjusted Return on Net-worth
	Return on long term funds
	Fixed asset turnover ratio
	Total asset turnover ratio
	Asset turnover ratio
	Dividend per share
Solvency Factors	Debt-Equity Ratio
	Long Term Debt-Equity Ratio
	Total Debt to Owner's Fund
	Inventory Turnover ratio
	Investment turnover ratio
Investment Valuation Factor	Operating Profit Per Share
	Return on assets excluding Revaluations
	Return on assets including Revaluations
Cash flow indicator factors	dividend pay-out ratio net profit
	dividend pay-out ratio cash profit
	Earning Retention Ratio
	Cash Earning Retention Ratio

Debt coverage factors	Interest Coverage
	Financial Charges Coverage Ratio
	Financial Charges Coverage Ratio Post Tax
Liquidity factors	Current Ratio
	Quick Ratio
Operating Factors	Number of days in working capital

All the companies are classified into credit worthy companies and less credit worthy companies. The classification is done based on Moody's credit ratings, one of the top three credit rating agencies. These ratings are used by the company to identify the ability of the issuers to meet the financial obligation on time.



SOURCE:

https://www.google.com/search?q=moody%27s+credit+rating+scale&rlz=1C1RLNS_enIN831IN831&source=lnms&tbm=isch&sa=X&ved=0ahUKEwjT6o6__PTgAhXBbX0KHZR9Cv0Q_AUIDigB&biw=1366&bih=657#imgrc=crWY0_X2BEMa0M

The above mentioned figure is a 21 point rating scale used by Moody’s to classify the company on the basis of their ability to pay the due amount in full and on time. According to Moody’s credit ratings, the companies with credit rating higher or equal to "Baa3" are considered to credit worthy companies and the companies with credit rating lower or equal to "Ba1"are considered less credit worthy companies.

Ratings assigned on Moody's global long-term and short-term rating scales are forward-looking opinions of the relative credit risks of financial obligations issued by non-financial corporates, financial institutions, structured finance vehicles, project finance vehicles, and public sector entities. Moody's defines credit risk as the risk that an entity may not meet its contractual financial obligations as they come due and any estimated financial loss in the event of default or impairment. The contractual financial obligations addressed by Moody's ratings are those that call for, without regard to enforceability, the payment of an ascertainable amount, which may vary based upon standard sources of variation (e.g., floating interest rates), by an ascertainable date. Moody's rating addresses the issuer's ability to obtain cash sufficient to service the obligation, and its willingness to pay. Moody's ratings do not address non-standard sources of variation in the amount of the principal obligation (e.g., equity indexed), absent an express statement to the contrary in a press release accompanying an initial rating. Long-term ratings are assigned to issuers or obligations with an original maturity of one year or more and reflect both on the likelihood of a default or impairment on contractual financial obligations and the expected financial loss suffered in the event of default or impairment. Short-term ratings are assigned to obligations with an original maturity of thirteen months or less and reflect both on the likelihood of a default or impairment on contractual financial obligations and the expected financial loss suffered in the event of default or impairment. Moody's issues ratings at the issuer level and instrument level on both the long-term scale and the short-term scale. Typically, ratings are made publicly available although private and unpublished ratings may also be assigned.

3.1 Moody's investors service/rating symbols

Long-Term Rating Scale

Aaa-Obligations rated Aaa are judged to be of the highest quality, subject to the lowest level of credit risk.

Aa-Obligations rated Aa are judged to be of high quality and are subject to very low credit risk.

A-Obligations rated A are judged to be upper-medium grade and are subject to low credit risk.

Baa-Obligations rated Baa are judged to be medium-grade and subject to moderate credit risk and as such may possess certain speculative characteristics.

Ba-Obligations rated Ba are judged to be speculative and are subject to substantial credit risk.

B-Obligations rated B are considered speculative and are subject to high credit risk.

Caa-Obligations rated Caa are judged to be speculative of poor standing and are subject to very high credit risk.

Ca-Obligations rated Ca are highly speculative and are likely in, or very near, default, with some prospect of recovery of principal and interest.

C-Obligations rated C are the lowest rated and are typically in default, with little prospect for recovery of principal or inter

We have taken a 4-point scale, companies with rating “Baa3” or “Aaa” are classified as credit worthy group and companies with credit rating “Ba1” or “Ba2” are classified as less credit worthy group.

Aaa
Baa3
Ba1
Ba2

The firms in the study are selected from various sectors of the economy. The data set includes 33 financial ratios of 15 Indian companies from 2o14 to 2o18 with 1o CWCs and 5 LCWCs. The ratio of CWCs to LCWCs is set to 2:1.

4. METHODOLOGY

In this study, a model is developed to predict the classification of CWCs and LCWCs based on selected financial ratios. We have adopted the technique of Factor analysis to reduce a large number of variables into fewer number of factors. Under factor analysis, principal component analysis method was used to solve the problem of multicollinearity among the independent variables. In a good model, the correlations etween dependent variable *y* and independent variables x_i should be high and those between the x_i variables should be low (Eksioglu *et al.*, 2oo5). With the help of factor analysis 33 financial ratios were reduced to 8 factors.

In order to find the most effective financial ratios in the credit risk model multinomial logistic regression method was used. Value labels were assigned to the dependent variable i.e. credit ratings, “Aaa” was coded as 1, “Baa2” as 2, “Ba1” as 3 and “Ba2” as 4.

Finally, multinomial logistic regression method was adopted to develop a model that would help to classify between CWCs and LCWCs. The significance level was checked at 95% confidence interval. The eight financial factors were treated as independent variables and the credit ratings of the companies as dependent variables. The overall accuracy of the model was checked with the help of Cox and Snell R square.

5. RESULTS

The results from the multinomial logistic regression are as follows:

Likelihood Ratio Tests				
Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	64.838	42.485	3	.000
BART factor score 1 for analysis 1	31.728	9.375	3	.025
BART factor score 2 for analysis 1	68.042	45.690	3	.000
BART factor score 3 for analysis 1	67.226	44.873	3	.000
BART factor score 4 for analysis 1	43.320	20.967	3	.000
BART factor score 5 for analysis 1	22.560	.207	3	.977
BART factor score 6 for analysis 1	63.149	40.796	3	.000
BART factor score 7 for analysis 1	34.136	11.783	3	.008
BART factor score 8 for analysis 1	64.151	41.798	3	.000

Pseudo R-Square	
Cox and Snell	.908
Nagelkerke	.974
McFadden	.889

Classification					
Observed	Predicted				
	1	3	4	2	Percent Correct
1	20	0	0	0	100.0%
3	0	10	0	0	100.0%
4	0	0	18	2	90.0%
2	0	0	3	22	88.0%
Overall Percentage	26.7%	13.3%	28.0%	32.0%	93.3%

From the above table, we can identify significant and insignificant financial factors while evaluating credit ratings of a company. Factors with p value less than .05 are treated as significant and factors with p value greater than .05 are treated as insignificant.

Financial factors which affect the credit ratings of a company are:

- Profitability factor
- Management efficiency factor
- Solvency factor
- Investment valuation factor

- Debt coverage factor
- Liquidity factor
- Operating factor

The value of Cox and Snell R-square is .908, which explains that 90.8% variability in credit ratings of a company is explained by seven key financial factors. The classification table shows the overall prediction accuracy of the model, which is 93.3%

6. CONCLUSION

The credit risk assessment model developed allowed to compile set of informative financial ratios for the estimation of credit risk. The total accuracy reached by above model is 93.3%, and the value of R square reached by above model is 90.8%.

From the initial 33 financial ratios which are grouped into eight financial factors, we found out that seven out of these eight financial factors help to predict the credit ratings of a company.

The financial ratios described in the research allow to assess credit risk of companies successfully. The results can help the developers of credit risk assessment models to compose the initial set of financial ratios.

With the help of proposed framework, financial institutions and companies can gain a better understanding of the risk associated with the applicants and can mitigate losses by giving loans to the companies which are less likely to default.

A limitation of this study is that we have only used financial ratios to distinguish between CWCs and LCWCs. In future study, we can include more features like main activity of the organization, age of business, future prospects of the business, etc.

6.1. FUTURE DIRECTION

I am planning to develop a framework of machine learning and neural network for credit risk modelling using python. The output of the neural network can be used as a classifier that will help the organizations to identify whether the borrower will default or not default.

7. REFRENCES

1. Alessandri, P., & Drehmann, M. (2010). An economic capital model integrating credit and interest rate risk in the banking book. Journal of Banking & Finance, 34, 730–742.
2. Angelini, E., Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. The Quarterly Review of Economics and Finance, 48, 733–755.
3. Bonfim, D. (2009). Credit risk drivers: Evaluating the contribution of firm level information and of macroeconomic dynamics. Journal of Banking & Finance, 33, 281–299.
4. Bystrom, H., & Kwon, O. K. (2007). A simple continuous measure of credit risk. International Review of Financial Analysis, 16, 508–523.
5. Chen, W. H., & Shih, J. Y. (2006). A study of Taiwan's issuer credit rating systems using support vector machines. Expert Systems with Applications, 30, 427–435.
6. Lin, S. L. (2009). A new two-stage hybrid approach of credit risk in banking industry. Expert Systems with Applications, 36, 8333–8341.
7. Marshall, A., Tang, L., & Milne, A. (2010). Variable reduction, sample selection bias and bank retail credit scoring. Journal of Empirical Finance, 17, 501–512.
8. Min, J. H., & Lee, Y. C. (2008). A practical approach to credit scoring. Expert Systems with Applications, 35, 1762–1770.
9. Paleologo, G., Elisseeff, A., & Antonini, G. (2010). Subagging for credit scoring models. European Journal of Operational Research, 201, 490–499.
10. Pang, S. L., & Wang, Y. M. (2008). Credit decision model and mechanism with default risk parameter. Systems Engineering - Theory & Practice, 28(8), 81–88.
11. Piramuthu, S. (2006). On preprocessing data for financial credit risk evaluation. Expert Systems with Applications, 30, 489–497.
12. Stefanescu, C., Tunaru, R., & Turnbull, S. (2009). The credit rating process and estimation of transition probabilities: A Bayesian approach. Journal of Empirical Finance, 16, 216–234.
13. Xu, X., Zhou, C., & Wang, Z. (2009). Credit scoring algorithm based on link analysis ranking with support vector machine. Expert Systems with Applications, 36, 2625–2632.
14. Zhang, J. L., & Hardle, W. K. (2010). The Bayesian additive classification tree applied to credit risk modelling. Computational Statistics and Data Analysis, 54, 1197-1205.
15. Jun Huang, Haibo Wang, (2017) "A data analytics framework for key financial factors", Journal of Modelling in Management, Vol. 12 Issue: 2, pp.178-189.
16. Stephanou, C. and Mendoza, J.C. (2005), Credit Risk Measurement Under Basel II: An Overview and Implementation Issues for Developing Countries, World Bank Policy Research.

17. Tian, S., Yu, Y. and Guo, H. (2015), “Variable selection and corporate bankruptcy forecasts”, Journal of Banking & Finance, Vol. 52, pp. 89-100.
18. Wang, Y., Lü, Z., Glover, F. and Hao, J.-K. (2012), “Path relinking for unconstrained binary quadratic programming”, European Journal of Operational Research, Vol. 223 No. 3, pp. 595-604.
19. Wei, H. and Billings, S.A. (2007), “Feature subset selection and ranking for data dimensionality reduction”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 29 No. 1, pp. 162-166.
20. Yang, J. and Li, Y.-P. (2006), “Orthogonal relief algorithm for feature selection”, in Huang, D.-S., Li, K.and Irwin, G. (Eds), Intelligent Computing, Springer, Berlin, Heidelberg, pp. 227-234.
21. Yu, L. and Liu, H. (2004), “Efficient feature selection via analysis of relevance and redundancy”, Journal of Machine Learning Research, Vol. 5, pp. 1205-1224.
22. Zuccaro, C. (2010), “Classification and prediction in customer scoring”, Journal of Modelling in Management, Vol. 5 No. 1, pp. 38-53.