# An Evaluation of Automated Credit Scoring System for Financial Services in Developing Countries

**REBEKA SULTANA**
**Student Id: 012142005**

A Thesis

in

The Department

of

Computer Science and Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science in Computer Science and Engineering

United International University

Dhaka, Bangladesh

January, 2018

# Approval Certificate

This thesis titled " **An Evaluation of Automated Credit Scoring System for Financial Services in Developing Countries**” submitted by **Rebeka Sultana**, Student ID: **012142005**, has been accepted as Satisfactory in fulfillment of the requirement for the degree of Master of Science in Computer Science and Engineering on [Date].

**Board of Examiners**

1.

______________________________ Supervisor

Dr. Khondaker Abdullah Al Mamun,
Associate Professor, CSE.

2.

______________________________ Head Examiner

Mohammad Nurul Huda,
Professor & Coordinator – MSCSE

3.

______________________________ Examiner-I

Suman Ahmmed,
Asst. Professor, CSE

4.

______________________________ Examiner-II

Swakkhar Shatabda,
Associate Professor & Undergraduate Program Coordinator, CSE

5.

______________________________ Ex-Officio

Salekul Islam,
Associate Professor & Head of the Dept. CSE

# Declaration

This is to certify that the work entitled "**An Evaluation of Automated Credit Scoring System for Financial Services in Developing Countries**" is the outcome of the research carried out by me under the supervision of Dr. Khondaker Abdullah Al Mamun, Associate Professor, CSE.

Rebeka Sultana,
Student ID: 012142005,
The Department of Computer Science and Engineering

In my capacity as supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

Dr. Khondaker Abdullah Al Mamun,
Associate Professor, CSE.

## Abstract

As the world is dependent on monetary stuff, credit has become consequential in our way of life. Therefore, credit scoring has become a comprehensively practiced strategy that helps bank and other financial organizations to evaluate creditworthiness of a client who applied for loans. The core purpose of this research is to carry out a comprehensible assessment of automated credit scoring framework for monetary service applications. The study tries to recognize the major deciding factors for developing an automated system for credit scoring purposes. It suggested a set of features which will be better to use in constructing scoring model for our country. We gave priority on employment, applicant's salary, previous loan history, purpose of loan, requested loan amount in the optimal feature set. It presents a comparative assessment identified with Statistical and Artificial Intelligence (AI) methods that are utilized for automated credit scoring system. Moreover, it helps us to be aware of most accepted and effective methods practiced in credit scoring system. After comparing different methods, we found Neural Networks and Genetic Programming has higher predictive ability. This analysis notified that there is no supreme statistical approach employed to construct credit scoring framework which works on all circumstances. This study revealed that enhancements are necessary (in the current credit scoring framework) to successfully address every single financial environment. Despite the fact that credit scoring is greatly in practice in developed countries, nevertheless in developing countries it is not executed in numerous financial administrations. By using credit scoring system in our country we can facilitate loans specially micro credits to people who are applied for loan. In this paper, few recommendations are provided for microfinance and micro-lender of developing countries. To actualize better credit scoring framework, few conceivable methodologies were suggested as well. Although, it has great prospect of determining reliability, however credit scoring management is due for a noteworthy overhaul.

## Acknowledgement

Some people's guidance and assistance was needed for the final and successful outcome of this thesis and I am extremely fortunate to have got this all along the completion of my thesis. I would not forget to thank them as all my success is only due to their careful supervision and a lot of assistance.

I respect and thank Dr. Khondaker Abdullah Al Mamun, Associate Professor, CSE, for providing me an opportunity to do the thesis work in United International University and giving me all support and guidance which made me complete this thesis duly. I am extremely thankful to him for providing such a nice support and guidance, although he had a busy schedule.

I am also thankful and fortunate enough to get constant support, guidance and encouragement from all the Teachers of Computer Science and Engineering Department of United International University which helped me in successfully completing my thesis.

# Publication

Rebeka Sultana, Samira Muntaha, D. M. Anisuzzaman, Farhana Sarker, and Khondaker A. Mamun, “Automated Credit Scoring System for Financial Services in Developing Countries,” International Conference on advanced information and communication Technology 2016.

# Table of Contents

# LIST OF TABLES

## LIST OF FIGURES

# Chapter 1

# Introduction

## 1.1 Introduction

The event of loaning and borrowing was part of human conduct from the very beginning of human life. Along these lines, credit has turned out to be significant in our way of life and it is a topic as aged as trade and commerce. By and large, credit alludes to credit card, mortgage, rent, contracts, financing trade, bond and so forth. Credit evaluation is one of the most troublesome exertions for bank and financial establishments. The customary procedure to allow a credit card was set up by credit specialists utilizing past understanding and different rationales. Extreme planning costs, precarious and wrong choices for the comparable applications are some consistent characteristics of this technique, which may come about choosing an inaccurate or broken choice while allowing credit. Therefore, financial establishments may drop reliable shoppers or bear huge capital wreck because of the ensuing default by the customer [1]. Once in a while these money related troubles can lead to bankruptcy. Because of the uncertainty and competitions in the business environment, any financial organizations may face bankruptcy. These lapses have prompted development of more methodical and correct strategies to evaluate the credit danger.

Now, computerized credit scoring has turned into an urgent apparatus for bank and other monetary establishments to assess reliability, decrease plausibility of threat, settle on regulated decisions and enhance the viability as well as the financial solidness. Bad loans can be detected very easily by a good credit scoring model. The center thought behind credit-scoring involves the grouping of potential clients into great quality candidates, who has the opportunity to reimburse the loan and poor quality candidates who has the likelihood of inability to reimburse the loan [2],[3]. Credit scoring has become much more instinctive than before because of the high efficiencies of computing technology.

## 1.2 Motivation

From the very beginning, I wanted to do research with something which will be beneficial to my country. Loan is very much integrated in our way of life. Now a days almost everyone lends or borrows from each other always be that how small contribution that is. In developed country, people take loan for almost all the personal work to professional project which help them to fund and finish whichever personal work or professional project they want e.g. education, treatment, business, wedding, buying house or car etc. In developing countries like ours, we also require a lot of money to do these. But, we do not get loan very easily as there are no worth mentioning loan assessment system or we are not habituated of using credit scoring. This is why I wanted to get in depth knowledge in credit scoring by reading various research papers which will help to create a credit scoring model. If we can get loan as much easily as any other developed countries, it will help us develop ourselves as well as our country. Thinking of these, I decided to study as much research paper I could on credit scoring and do a review paper on that.

## 1.3 Contribution

In recent years, different credit scoring methods in particular fields have been discussed by many international journals or papers. Synthesizing and analyzing various methods of credit scoring proposed in those articles is one of major purposes of this thesis paper. On the basis of the published papers or reports, I tried to analyze the predictive ability and successfulness of several methods applied in credit scoring framework for instance Linear Discriminant Analysis, Logistic Regression, K-Nearest Neighbor, Decision Tree along with the Artificial Intelligence approaches like Expert System, Support Vector Machine, Fuzzy Logic, Neural Network and last one is Genetic Programming. Furthermore, this study attempts to focus the key ideas of the above mentioned methods. Although, the selection of variables is varying from situation to situation, this study tries to find out all possible and most common variables which used to build the model. This thesis likewise brings up several experimentations accomplished in the sector of credit scoring in creating nations. Moreover, it gives a few suggestions alongside the conceivable methodologies that can be assumed to actualize potent scoring system in developing nations. This study will certainly helps many researchers or investigators to practice their study in the credit scoring field and also support to abstain overlapping endeavors.

## 1.4 Organization of thesis book

In this paper, chapter 2 describes credit rating, credit scoring system and judgmental system.

Chapter 3 provides the amenities and limitations of credit scoring system. The applications or purposes of credit scoring are also described here. Moreover, one of the crucial parts of credit scoring system, the major deciding factor or variables that are used for constructing the model is provided here. In this section, we tried to summarize the optimal feature set of credit scoring system which may suitable for our country.

Chapter 4 describes the Statistical and Artificial Intelligence (AI) approaches of credit scoring system. In statistical approaches, there are four particular methods are described. These are Discriminant Analysis (DA), K-Nearest Neighbor (KNN), Logistic Regression (LR) and Decision Tree (DT). Moreover, in Artificial Intelligence (AI) technique five different methods are provided. These are Expert System (ES), Support Vector Machine (SVM), Fuzzy Logic, Neural Network (NN) and Genetic Programming (GP). Comparative analysis of these methods is also provided in this section.

In chapter 5 we tried to provide an idea about practicing credit scoring on developing countries along with prospect of credit scoring. This section imparts few recommendations and methodologies which will help working on credit scoring in our country in future.

Chapter 6 presents a brief review of this thesis work. All the problems that are confronted during this thesis are also provided in this section. We have some future plan regards this thesis on automated credit scoring system, which are included here.

All the references that are used in this paper are provided in chapter 7.

# Chapter 2

# Credit Rating, Credit Scoring and Judgmental System

## 2.1 Credit Rating

Credit rating is an evaluation of the creditworthiness of a scrounger in regular terms or in respect of a specific debt or monetary obligation. It can also be defined as a method that aids the lenders to take their decision about granting credit to the applicants with respect to the applicant's nature [3]. It consists of a set of decision models and methods to help the bestowers or granters [4]. They can grant customer credit by evaluating the risk of lending to particular users. Credit rating can be ascribed to any organizations – a particular corporation, state or local authority, sovereign management that needs to borrow money. To analyze the credit risk, the common way is to utilize a characterization procedure over comparative features of past clients who are dependable along with who are faulty considering the end goal to discover a connection between the invariant and possible failure. Precise classifier ought to be settled in favor of assort new clients or current clients as trusty or faulty [2]. A company's credit rating plays a very significant foreword to the business itself, its investors, stakeholders, suppliers, and debtors. A company will be assigned a good credit rating, when it is in a good financial position with growth prospects. To see this good credit rating, shareholders will be more confident with their investment and the reliance of the business partners will raise on the company [5]. Generally, a credit rating agency assesses and evaluates the credit for organizations and governments. These rating agencies are remunerated by the organization that is trying to obtain a credit rating for itself or for its any due case. For individuals, credit ratings are obtained from their credit background provided by credit reporting firms [6].

## 2.2 Credit scoring

To assess the credit rating of counterparties, credit scoring model or credit scoring is one of the most frequently used methods. The objective of scoring system is to quantify fiscal hazard of the credit; therefore, the investors can take loan granting agreement instantly. Generally, scoring system is developed based on the credit information gathered from credit authorities. This credit score is used by lenders, such as banks or credit card communities to evaluate the future risk posed by providing loan to clients and to reduce the losses due to faulty investment. It is also used by the granters to determine who are able to grant a loan with credit limits and interest rate. They also practice credit score to determine which clients are likely to fetch the greatest profit and the likelihood that one will become delinquent on an account at some point in the near future. Credit scoring is not only used by the banks, but also, the other organizations such as the insurance companies, mobile phone communities, property owners and government departments. Credit scoring also has enough overlap with data mining, which uses numerous similar methods. These methods incorporate thousands of identical or similar factors [7].

## 2.3 Credit scoring system versus judgmental system

### 2.3.1 Credit scoring system

In a credit scoring model, to separate the acceptable and unacceptable credit application, analyst most often utilize their past knowledge with borrower in order to develop a quantitative model. Most of these models are made up of financial history and typical knowledge taken from a sample of authentic organizations. By using this scoring, investigators are able to assess the creditworthiness instantly; therefore, the customer services are improved. Furthermore, credit scoring has been scrutinized through measurable issues with the data applied to infer the model. The presumptions of the specific statistical method considered to obtain fair score are also a problem. In spite of having few drawbacks of credit scoring approaches, those can be considered as most fruitful models utilized as a part of business and fund segment.

### 2.3.2 Judgmental system

In comparison with quantitative credit scoring systems, judgmental credit scoring system deals with unwritten rules, particular combinations of commercial and qualitative information, ambiguous and missing information. Comparing the particularity or the characteristics of a client with the previous clients is the overall idea of credit evaluation in judgmental system. Previous clients are those who have been granted credit against their loans but failed to repay or have consequently defaulted. Therefore, the application of those clients will typically be dropped. The application will habitually be granted if the client's characteristics are sufficiently similar to those, who have not defaulted.

The inconvenience of this system is cannot be used to figure out an unexpected loss, as they do not render a default probability by themselves.

## 2.4 Conclusion

Credit score is really a mathematical phrase depending on an even examination of a client's credit rating data, to be able to signify the reliability of this client properly. Whereas, the progresses of a judgmental system rely upon the previous knowledge along with the superior judgment of the credit analyst. Hence, judgmental approaches are affiliated with individuality, disparity as well as particular inclination inspiring decision. Automated credit scoring systems are not only slows down human participation on credit evaluation but also saves both expenditure and time of the bank and clients. Credit scoring systems are unerring the inclination when the result considering only the recognized applications. They perform this through expecting the execution of rejected application in case they had been recognized. The execution of credit scoring scheme might be watched, obliged, likewise adjusted at whatever point.

# Chapter 3

# Amenities, Challenges and Applications of Credit Scoring

## 3.1 Conveniences of credit scoring system

Scoring system is practiced successively over credit assessment by means of few certain advantages it obtains. Credit scoring paradigms are designed over simple delivery architecture. As they are based on expert system and different artificial techniques, also required less information, it can obtain a decision very quickly. It includes only those variables which are correlated with payoff execution, therefore, reduces unwanted variables [8]. The excellence of these models made them proficient for internet deployment; as a result they are extensively available through the web channels [9]. Several credit analysts or statisticians can simply and evidently evaluate the same data given same weights, which is crucial for credit scoring. It also makes the loan approval procedure much quicker. With the assistance of credit scoring, financial organizations can measure the dangers connected with allowing credit to a specific candidate in a shorter time.

Automated credit scoring has numerous advantages that gather to the moneylenders as well as to the recipients. To give a verifiable analysis of a client's reliability utilizing credit scoring scheme, discrimination can be reduced. This engages credit providers to fix attention on details that concerns to investment risk and sidestep the individualism of loan analyst. Such judgment helps loan specialists guarantee for applying the equivalent endorsing criteria to each and every asker giving slight concentration on race, gender, or particular features varied through demand being applied for credit selections [10]. Through urging to raise the speed, exactness and consistency of the loan application process, credit score confirms the computerization of the lending procedure. It can likewise enhance the portion of assets towards the "first best equilibrium" [11]. By utilizing credit score, money related organizations can repair their loan fee that they ought to charge their purchasers [12]. Most extreme danger customers are charged a higher loan fee. In light of the customer's credit score, the financial organizations are likewise ready to decide as far as possible to be set for the shoppers. These help financial organizations to lead their records all the more successfully and productively.

## 3.2 Limitations of credit scoring system

Though scoring system for credit has intellectual benefits, some of its flaws ought to likewise be noted. For the reason that, credit scoring is an automatic schema for evaluating the earlier clients, as a result there is a chance to collapse along with estimate a few facts inexactly. The creditors assess many applications perfectly, decently and quickly in a legitimately planned credit scoring frame work. In the event that, this is genuine, then the converse should likewise be valid. A large number of candidates assessed by a weakly outlined credit scoring framework and wrong suggestions are made inevitably. It is never possible to measure the credit danger precisely, and any model that claims the other way is incorrect. It might change overnight additionally. For instance if the business heritor passes on and no one is present to supplant him, a vast problem may occur. One major problem can come out, while building up a scoring framework applying an uneven case of purchasers and clients who selected for credit [13]. It may occur in light of the fact that simply the great clients are spoken to as the specimen is uneven along with refused clients won't be consolidated within the data while building up the scoring structure. The scoring structure developed using such illustration may not perform sufficiently in all people whereas the information applied to produce such model is not same as the information the structure will be associated with. As the qualities of earlier candidates delegated "good" or "faulty" creditors can be utilized the creditworthiness of new candidate, so the change of patterns after some time is one of the most serious issue that can emerge when developing credit scoring models [14]. In a factual credit scoring, it requires a huge amount of data on every loan. Moreover, requires a specialist to direct and watch everything. Some of the time professionally outlined models just furnish the acknowledge administrator for a numerical score. With this constrained data, it can be very troublesome for the credit administrator to clarify a negative credit choice to a furious credit candidate, or to an active customer. Statistical credit scoring can dismiss broken applications, yet it cannot change them. It is additionally vulnerable to abuse [15]. The measurable precision of the model is suspicious when the variable does not fulfill certain required presumptions. Either the weights or the variable, or both are thought to be altered after some time, which cause the model barely precise. Subsequently the model might not able to give satisfactory results after some time. For this situation, the model won't be predictable. For instance, if the bank is resolved to allow credit cards to students, utilizing credit scoring model that is planned on tests that do not contain the student population, then the model cannot separate good from bad clients [16]. Drives a bilateral result, for example, the debtors are whether faulty or not is one more objection of this system. Furthermore, these scoring models are much expensive to procure. This system may drive out reliable clients when anyone switching one’s job or place of residence or business where one can be contacted [8]. Despite above restrictions, we have no suspense that automated scoring system bestowal an imperative tool in predicting monetary danger over customer loaning.

## 3.3 Applications or purposes of credit scoring

Usually, applications of automated credit scoring have been employed on individual areas, consecutively assimilation over varieties of certain processes. These applications can be classified into finance and book keeping, developing, constructing and manufacturing, health and direction, general purposes and so on. Primarily financial industries practiced credit score assessment that makes a decision applying credit card option for furtherance usage. However, the utilization of automated scoring system has broadened over numerous sectors such as habitation, security, constitutive service applications, enrollment or recruitment and so on. In the accounting and economic sector, basically monetary organizations implement credit score assessment to decide upon credit plan for client’s applications who asked for loan. In view of this, automated credit scoring system is utilized in consideration of various directions, for example, bankruptcy forecasting along with liquidation systematization [17], [18], monetary desolation [19], scoring purposes [20] etc. Credit scoring exercise at sparing cash portions have enlarged amidst the most recent few decades [21], [22]. Remarkably, in light of the enormous quantity of credit relevance for several bank items, giving an extensive variety of latest approaches that may be utilized by them. Credit scoring system has distinctive usage over bank items, for example, customer loans, which are a standout amongst the most critical and vital applications generally utilized as a part of the field [23], [24]. Customer credit has turned into a colossal industry, and the quantity of uses has expanded. Likewise the people on the east side of Europe and Chinese have begun to mark the helpfulness of purchaser or customer credit, on the ground that a crucial job of purchaser credit is to compose credit broadly accessible as well as profitable too. The assessment of current client is a standout amongst the most significant employments of scoring structure in the last few decades [20], [25].In late years , automated credit scoring has additionally been utilized as a feature of the choice procedure for giving credit to little businesses [26], [27]. Credit scoring system also expanded over microfinance, small and medium enterprises (SME) where micro credits are provided to the people who applied for loan [28], [29]. In consideration of selecting a particular application of a loan receiver, financial institutions promptly practice credit scoring to set maximum or minimum loan amounts, supervise data management and measure the possible profit of consumers and clients. For instance, the Australia and New Zealand Banking Group rehearses credit scoring to recognize candidates who ought to give credit, decide the measure of credit that the candidates ought to get, and the strides that ought to be taken should there be a breakdown in loan payment (see http://www.sas.com/success/anzcredit.html ). Besides, on the accommodation loan purposes, automated scoring system is extensively employed [30], [31].

## 3.4 Major deciding factor of credit scoring

In credit scoring framework, the general intention of variable selection or feature selection is to get a decent example including less directivity. Independents variables have been utilized for a large portion of the smart credit scoring models. These variables are given by the banks without changes. The precision of the system may enhance with the assistance of approved method for finding the most perfect client features as well as many - sided characteristics of the framework may decrease through allocating the unimportant variables or features. Therefore, we can say variable choice might influence the execution of the model. Fundamentally, credit scoring was committed to figuring people who were allowed loans, both current and new clients. Credit experts are worked with preplanned score, investigate client's monetary information and creditworthiness in order to reduce the chance of defraud or failure. The crucial significance and the target of a credit scoring model is the arrangement of good and bad credit [41], [42]. The need of a proper grouping method is in this way self-evident. Another candidate is controlled by a few attributes, for example, gender, marital status, age, wards, having a telephone, occupation, informational level length of remaining at current place of residence as well as holding a bank card or debit card. These attributes are universally applied as part of constructing scoring models [3], [21], [24], [42]–[44]. The functioning of scoring models also used the duration of staying at present occupation, loan amount, loan length of time, house proprietor, month to month pay, bank accounts, having an auto, contract, motivation behind credit, certifications and others [24],[43]. Now and again better-half's personal information, for instance, age, remuneration, financial, balances and others has been consolidated into the summary of factors. More factors, for instance, most exceedingly awful record condition, recruitment duration, duration accompanying bank or other financial organizations and others are less from time to time used as a piece of building scoring models [45],[46]. A credit scoring utilizes the data as a part of credit report as opposed to utilizing the scores from pop tests and papers. It is much like an instructive establishment, where an educator figures a last grade by taking every one of the scores from tests, homework, participation and whatever else they need to utilize, measuring every one as per significance. To determine individual credit scoring payment or credit details over certain period of times or debt, new credit or types of credit are used. A hypothetical motivation behind why such variables have been picked is not set up in any of the examination evaluated in this paper. By and large, creators have expressed that a specific establishment gave a specific arrangement of information. Along these lines, the determination of the variables relies on upon information suppliers and the information accessibility as expressed by those creators. There is no optimal feature set that can be employed to construct effective credit scoring framework. Complete variable selection depending through the medium of the information. For instance, in [44], they utilized a set of 41 features whereas a set of 29 features have been employed in [47]. Moreover, the determination of the specimen size is another issue. It is trusted that the scoring model's precision is better when the specimen size is bigger. These determinations mostly hand-off o the way of the business sector, the information accessibility and to what degree this specific information set will speak to the community. "Validation method" or the order of the specimen has been generally utilized as a part of financial credit scoring applications. While a few analysts have connected an acceptance strategy by isolating the specimen into preparing and approval and testing sub-testes [47], different scientists have used a basic approval system by separating the example into preparing and testing sub-samples [48], [49]. In short, there is no perfect credit scoring technique that can be connected to various banks in various situations.

Table I provides the sorts of variables that are practiced in constructing credit scoring framework in different literature. Generally, the selection of variables will differ from situation to situation and country to country. The information which may be utilized in a credit scoring system is subject to changing legislation. For example, in the UK, it is not permissible to discriminate on gender grounds [69]. Although different kinds of variables were used for different methods there are some common variables among them, regardless the situation or country. Employment, Age and Marital status are the most commonly used variables. So we can see that whatever the platform of modeling, situation or country is these variables are most certainly used. There are some other variables that are not as popular as these three (Employment, Age and Marital status), but we can say them popular. These are – applicant's salary, home status, time at present address, purpose of loan, time with employer, the number of dependents and the requested credit amount. Last of all there are some uncommon variables which are only used depending on the situation or country and these are - postcode, telephone no, credit card, type of bank account, education, applicants monthly /yearly expenditure, gender, county court judgments, time with bank, insurance required, no of car, behavior, credit history, worst account status (0–99), time since most delinquent account, checking account status, the duration of the credit period in months, assets owned, other debt and history of past payment.

Table 1: Summary of Variables That Are Practiced In Constructing Credit Scoring Framework In Different Literature.

| | Hand & Henley 1997 [69] | West 2000 [106] | Bellotti 2009 [45] | Tsai2009 [18] | Yeh linen 2010 | Yap 2011 [54] | Harris2015 | |
|---|---|---|---|---|---|---|---|---|
| | UK database | German dataset | Credit card openers in the same 3 month period's dataset from a major financial institution | Dataset of customer of unsecured consumer loan from a certain financial institution in Taiwan | Credit card holder's dataset from bank of Taiwan | Dataset of payment history of members from a recreational club | German dataset | Barbados dataset |
| **Time at present address** | ✓ | | | | | ✓ | | ✓ |
| **Home status** | ✓ | ✓ | ✓ | | | | ✓ | |
| **Postcode** | ✓ | | ✓ | | | | | |
| **Telephone** | ✓ | | | | | | ✓ | |
| **Applicant's salary** | ✓ | | | ✓ | | | | ✓ |
| **Credit card** | ✓ | | ✓ | | | | | |
| **Bank account type** | ✓ | | | | | | ✓ | |
| **Employment** | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ |
| **Education** | | | | ✓ | ✓ | | | |
| **Applicants monthly /yearly expenditure** | | | | | | | | ✓ |
| **Age** | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| **Gender** | | | | | ✓ | ✓ | | |
| **County Court judgments** | ✓ | | | | | | | |
| **Purpose of loan** | ✓ | ✓ | | | | | ✓ | ✓ |
| **Marital status** | ✓ | | | | ✓ | ✓ | ✓ | ✓ |
| **Time with bank** | ✓ | | ✓ | | | | | |
| **Time with employer** | ✓ | ✓ | | | | | | ✓ |
| **Insurance required** | | | ✓ | | | | | |
| **No of car** | | | ✓ | | | | | |
| **Behavior** | | | ✓ | ✓ | | | | |
| **Credit history** | | ✓ | ✓ | | | | | |
| **Worst account status (0–99)** | | | ✓ | | | | | |
| **Time since most delinquent Account** | | | ✓ | | | | | |
| **Checking account status** | | ✓ | | | | | ✓ | |
| **Duration of the credit period in months** | | | | | | | ✓ | |
| **The number of dependents** | | | | | | ✓ | ✓ | ✓ |
| **The credit amount requested** | | ✓ | | | ✓ | | ✓ | ✓ |
| **Assets owned** | | ✓ | | | | | | |
| **Other debt** | | ✓ | | | | | | |
| **History of past payment** | | | | | ✓ | ✓ | | |

### 3.4.1 Optimal feature set

In forming credit scoring framework, perfect variable selection is a very important aspect. There is not any fixed number of feature to construct the scoring model as it varies from situation to situations and country to country. Even though I did not find any example of credit rating in Bangladesh, we can use some variables for our country which I found after studying a lot of research papers which will help to construct an effective scoring model. These are: employment, applicant's salary, previous credit history, purpose of loan, requested loan amount. Moreover, there are other important variables which can be used as per the need of a situation such as gender, age, education, number of children, phone no, credit card type if available, insurance if available, applicant's monthly expense, property, bank balance, assets owned, other debt, etc.

## 3.5 Conclusion

A credit scoring model makes it intuitive for a moneylender to archive the business explanation behind utilizing a component that may have an excessively negative result on specific gatherings of clients shielded by law from segregation. The measurement of the respective quality of each attribute's connection with credit accomplishment is provided by the counterweights of the model. Moreover, to develop the credit scoring model we need to practice new security approaches along with re-establishing the current approaches. In feature selection, if a scoring framework does not included all conceivable features in it as well does not redesigned, consistently it will misclassify a couple of group of peoples or not ready to give sufficient outcome. Our analysis notified that there is no supreme or best scoring model holding pointed variables that can be utilized as a part of various markets. Consequently, the danger system and the credit industries need to be modified.

# Chapter 4

# The Techniques of Credit Scoring Scheme

These days credit scoring enhancement is a growing point whereabouts diverse analysts are utilizing distinctive system in consideration of picking correct candidate as long as diminish loan default. Particularly for credit allotting institutes, similar to commercial banks and a few retailers, it is critical to choose reliable clients from reprobate clients. To get an acceptable credit scoring framework, various strategies have been suggested. In view of this, we talked about four statistical systems: decision tree, discriminant analysis, logistic regression and k-nearest neighbor. We also talked about five artificial Intelligence (AI) strategies: expert system, support vector machine, fuzzy logic, neural network and genetic programming. A short portrayal of these strategies is talked about in this segment.

## 4.1 Statistical and Maximization approaches

### 4.1.1 Discriminat Analysis (DA)

In modeling classification tasks discriminant analysis (DA) has been reported as the most generally discussed and practiced statistical methods. It aims to recognize which variables are the most excellent predictors for accepting or rejecting new applicants. It immensely enhances an executive's choice making process as it considers and assesses the pre-noteworthy information which has as of now been built up. This would be for all intents and purposes unthinkable for a person to physically do as it would be excessively mind boggling and greatly tedious. Fisher proposed discriminate analysis as a separation and indexing tool. It was one of the main strategies connected to building credit scoring models by separating between those candidates who in the past had reimbursed financial commitment and the individuals who had defaulted [50]. DA depends on the presumption that, for every given class of active variable, the illustrative factors are circulated as a multivariate ordinary dissemination with a typical change covariance network. So we can say that DA is an alternate to logistic regression. The purpose of Fisher’s rule is to reduce the gap within each group and to maximize the gap between distinct groups. In [51] Durand inspected car loan applications and demonstrated that DA could deliver great forecasts of credit reimbursement.

It was the initially distributed record of the utilization of discriminant investigation to deliver a scoring framework. Altman is the primary scientist who predicts the failure of firms from various commercial ventures by utilizing DA [52]. He suggested that organizations with certain budgetary arrangements have massive chance of failure inside of the following period than firms with inverse attributes. To measure the well known Z-score, Altman constructed a predictive formula based on five key financial ratios by employing multivariate discriminant analysis (MDA) [53]. MDA is a statistical technique focused with the classification of different sets of observations and it attempts to discover the combination of variables that presumes the group to which an observation belongs. MDA's essential target is to order and/or make forecasts in issues where the dependent variable shows up in subjective structure, e.g. bankrupt or non-bankrupt. The investigator's space dimensionality can be diminishing from the quantity of various autonomous variables to K–1 dimension(s) by utilizing MDA; where K measures up to the quantity of unique from the earlier gatherings. MDA's essential leverage in characterization issues is the capability of breaking down the whole variable profile of the article all the while as opposed to consecutively analyzing its individual attributes. In bankruptcy prediction models, specialists are worried with two gatherings, comprising of bankrupt firms and non-bankrupt firms. In this way, the investigation is changed into one measurement. Some application and comparison of discriminant analysis in credit scoring were explained by [42], [54]–[57].Few researchers have revealed explicit critical remarks of practicing DA in credit scoring [58].Despite of having some criticisms, discriminant analysis is as yet a standout amongst the most usually rehearsed statistical procedures in credit scoring [59].

### 4.1.2 K-Nearest Neighbor (KNN)

K-nearest neighbor (KNN) is an entirely different, nonparametric and similarity based classification approach. KNN classifiers are established on learning by analogy. It is one sort of case-based realizing where all calculation is stopped until characterization finished and the capacity is just possible locally. In a K-nearest neighbor, attaining instances is collected. The KNN classifier initially characterized the element vectors or separation capacity and names of the training examples, eventually the similar components will be figured for the new example whose classes are obscure. To select the K nearest samples, distances between data points are calculated. At the point when given another specimen, a KNN classifier hunts the example space down the KNN which are nearest to the new specimen in term of separation between the elucidatory variables. The new sample is assigned to the class, which its maximum KNN belong to the same class. Therefore, the most class label inside a neighborhood of the training or developing instances gives the prediction for each new sample. KNN method has been used in numerous classification problems. It is also used in fraud detection and to recognize flaws of credit card clients [70]. Predictive accuracy of the KNN is to a great degree influenced through the measure of separation along with the cardinality of the neighborhood. It is an inconvenience of the KNN. Additionally, a basic classification formula is not exuded by KNN [54].Some studies have described the applications of K- nearest neighbor method and some are compared this method to other methods [13], [61], [68], [69], [72], [73].

#### 4.1.3 Logistic Regression (LR)

Logistic regression is a summarized form of linear regression. This summed up linear model is practiced for binomial regression. Logistic regression is developed to predict the likelihood of happening an occasion by fitting information to a strategic capacity from an arrangement of factors that might be persistent, dichotomous, discrete or a blend of any of these. LR model is more acceptable for the fraud detection problems. Not in any manner like other quantifiable mechanisms, it can fit a couple sorts of scattering limits, for instance, normal distributions, poison and gamble. Another kindness of this procedure is that it doesn't require regular dispersion variables and does not expect linearity of relationship between the free and reliant variables. LR also allows the interpretation of the output variable as a probabilistic binary value. In this way, in the event that we forecast whether another customer will be a defaulter, the expectation does not just turn out just yes/no answer, it accompanies a refined evaluated likelihood that the occasion will occur. This prediction output is continuous, that is, a relevant threshold value (cutoff point) must be determined for classification [61], [62].The disability of Logistic regression is, it can't legitimately resolve the issues of non-linear and intelligent impacts of informative variables. Some researchers described logistic regression and logit analysis very evidently [53]–[55], [63]–[67]. Some of credit scoring models practicing logistic regression have been explored [20], [68].

#### 4.1.4 Decision Tree (DT)

Decision tree is a categorization method employed in the framework of automated credit score assessment. With a view to deal with the categorization problems, a tree of decisions and their probable outcome is utilized. The root node always deals with the decisions whereas the inner nodes carry out the test. A lot of algorithms are developed for constructing decision tree including CHAID (Chi squared automatic interaction detection), CART, QUEST and C5.0 [54]. In consideration of taking care of regression issues, calculations like CART presented by Breiman in 1984, can be utilized as well. CART is a solitary methodology that can be utilized to break down either categorical or persistent data utilizing the similar approach [74]. "Divide and conquer", recursive partitioning heuristic search procedure is followed by decision tree strategies and generate tree-like successive decision models. By means of a sequence of tiny tests or decision, the general predictive decision can be made. Each of these usually includes a solitary property, which is the assumption of most decision tree inducers. Distinct decision tree inducers generally differ in the goodness evaluate used to determine the splitting attribute at each internal tree node. For instance, a famous decision tree inducers named ID3 determines the attribute that causes in the maximum information gain is introduced by Quinlan [75]. C4.5 is a very popular decision tree classifier which is a successor of ID3. It builds decision trees utilizing recursive partitioning[68], [72]. Over fitting can be a problem of using this method. In order to obtain decisive predictions, they also need a large number of data samples [76]. James Ang, Jess Chua and Clinton Bowling were amongst the first to build a non-parametric credit scoring system. They applied the decision tree technique to a credit scoring related problems [77]. Some employments of decision trees in credit scoring framework were sated by [63], [68], [78]–[80].

## 4.2 Artificial Intelligence (AI) technologies

### 4.2.1 Expert System (ES)

Expert system is classified among the conventional approaches in evaluating credit score. It is where programming intended to mimic the state of mind of human specialists. An outline of utilization and potential effect of expert framework in finance was presented in [81]. In a pro structure the loan determination is depend on the branch experts. The expert's aptitude, personal decision and influence of definite key features act a vital part in the judgment procedure. The experts evaluate the five factors and come to decision based on the intuitive balance between the five Cs. The five Cs are character, capital, cash-flow, collateral and condition. [71] stated that the ability to explain outcomes is one of the experts system's privileges which is discussed furthermore in [8], which justify reasons for denying credit applicant. The advantages of utilizing expert system for credit investigations are pace and exactness, both, which far exceed human limit. A few different credit evaluation ES have been produced [82] to enhance the throughput and exactness of loans allowed and to protect more noteworthy coherence of loan assessment. The vast majority of developed ES have been contrasted with customary strategies and existing techniques in financial area. For instance [83] contrasted his proposed ES and veritable appraisal of five credit officers in two distinct organizations. He reasoned that ES performs better than these five credit officers. These frameworks face two principal issues: (i) human specialists might be conflicting and subjective in their assessments; (ii) customary expert systems determine no weighting scheme that would reliably arrange the '5 Cs' as far as their relative significance in anticipating the default: what are the critical regular variables to examine crosswise over various sorts of borrower?[84].

### 4.2.2 Support Vector Machine (SVM)

Support vector machine (SVM) is an effective and promising information characterization and capacity estimation apparatus. The support vector machine is first proposed by Vapnik. SVM have recently been used in a range of problems including bioinformatics, pattern recognition, and text categorization and face or fingerprint identification and can perform both characterization and relapse. SVM is learning frameworks that utilize a speculation space of linear functions in a high-dimensional space, prepared with a learning algorithm from optimization hypothesis that actualizes a learning inclination got from measurable learning hypothesis. SVM uses a linear model to separate the feature space from the input features, at this point a kernel is used for this purpose. The delineation underneath demonstrates the essential thought behind SVM. In the left half of the schematic mapped we can see the original objects, i.e., modified, utilizing an arrangement of scientific capacities, known as kernels. The procedure of revising the items is known as mapping (transformation). Note that in this new setting, the right side of the schematic that is the mapped objects are straightly detachable and, hence, instead of building up the brain boggling twist (left schematic), we ought to just find a perfect line that can detach the green and the red things.

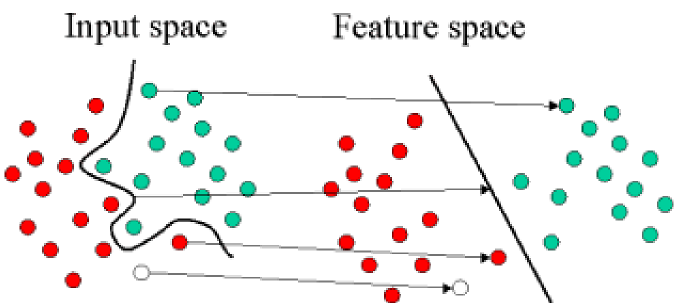


Figure 1: Support Vector Machine

SVM goes for minimizing an upper bound of the speculation blunder by boosting the edge between the isolating hyper plane and the data [2]. Examples from the training that are near the most extreme margin hyperplane be named support vector.

The support vectors are then used to develop an ideal linear separating hyperplane or a linear regression function over the feature space [85]. There are several types of kernels that can be utilized as a part of SVM models. These incorporate linear, polynomial, radial basis function (RBF) as well as sigmoid. A credit scoring model was explored in [86], which utilizes a machine learning philosophy that joins accounting data with the option-based methodology of Black, Scholes, and Merton for listed organizations in the Greek stock exchange. [87] Proposed another methodology taking into account direct search and components positioning innovation to upgrade features selections and parameter setting for 1-norm and least-squares SVM models for bankruptcy prediction. [88] Pursue a methodology for machine learning driven credit risk assessment utilizing linear Support Vector Machines, consolidated with sliding window approach for testing.

SVM can be utilized effectively as an element choice strategy to decide those application factors that can be utilized to most considerably show the probability of default [4].The disability of SVM is that, standard formulation has no specification of business constraints. Moreover, standard support vector machines are impressionable to outliers, which is a great flaw of its [2]. When using SVM, selection of a subset containing most favorable input items and most fitting kernel range settings problems may have confronted [89].

### 4.2.3 Fuzzy logic

Fuzzy logic is a logical arrangement, which is an expansion of ambiguous logic. It is a hypothesis in where it identifies with categories of objects with blunt or unsharp limits in which participation is a matter of degree. Numerous parameters, which are generally unclear, hard to characterize, and notwithstanding clashing, are utilized as a part of deciding the credit scoring. The methodologies, which depend on the customary or “crisp” ideas, don't have the adaptability for treating linguistic expressions. These semantic expressions are dubious as well as subjective. Besides, though lexical variables are utilized, these degrees might be overseen by particular capacities something that people have been overseeing for quite a while. Therefore, fuzzy set theory was practiced to manage this type of doubtful and linguistic situation, and developing the effectiveness of credit scoring. Fuzzy logic is a system of multiple valued logic that deals with approximate, on behalf of fixed and accurate interpretation.

Fuzzy inference process is described below in a short while.

Five steps:

- Fuzzified the input data (converting crisp to fuzzy logic)
- Applied the fuzzy operators (AND or OR) to the antecedent
- Connected the antecedent to the consequent
- Aggregated the consequents beyond the rules
- Defuzzified the data (converting fuzzy logic to crisp logic)

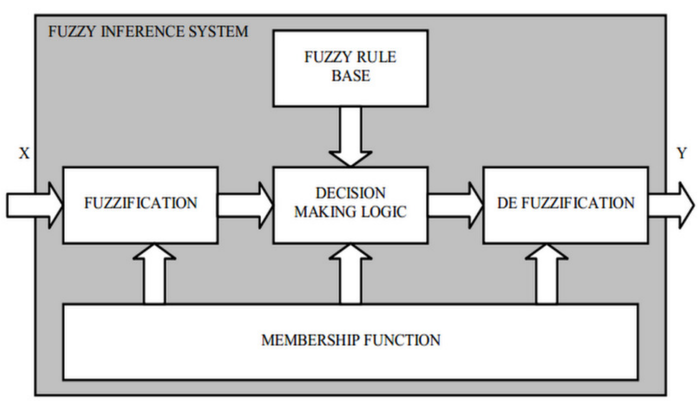


Figure 2: Fuzzy logic system

Fuzzy rule based system helps the creditors to propose rules that precisely derive the credit score with clarification, while most of credits scoring models focus on estimating a score without clarification how the results acquire. In [90], they proposed a fuzzy classification system for credit scoring named IFAIS, which combines fuzzy logic and AIS concepts. To judge the risk several AI systems are being used which has both advantages and disadvantages. Fuzzy logic is one of them. It assigns scores on each input variable for evaluation. The evaluation value should be varied according to economic status. It forms a less biased system and avoids queues along with increases efficiency and time effectiveness. Another advantage of fuzzy logic is the ability of managing both the qualitative and quantitative factors. Therefore, if the set of inputs are large, scoring outcomes will be lower sensitive to small computation errors [80]. Another advantage is that Fuzzy set represents the vague or linguistic system as it is, it is neither overly accurate nor overly simplified representation [91].The limitations of fuzzy logics are that, for fuzzy rules, different expert used to give different rules which let us lose their consistency. Another disadvantage of using this method is that the variable may not be correctly chosen.

### 4.2.4 Neural Network

Neural networks (NNs) have been strongly adapted in credit scoring and corporate distress prediction to different types of actual classification tasks in science and business industry. Neural networks are the mathematical representations stimulated by the performance of human brain. It can distinguish the various patterns amongst input and output dataset, at that point anticipate the aftereffect of current absolute input details.

It comprises a vast amount of nodes known as units or neurons, which are connected by links. To build a model that maps the input to the output properly is the objective of NN. Figure 3 construct a model of neural network with an output neuron and one hidden layer.

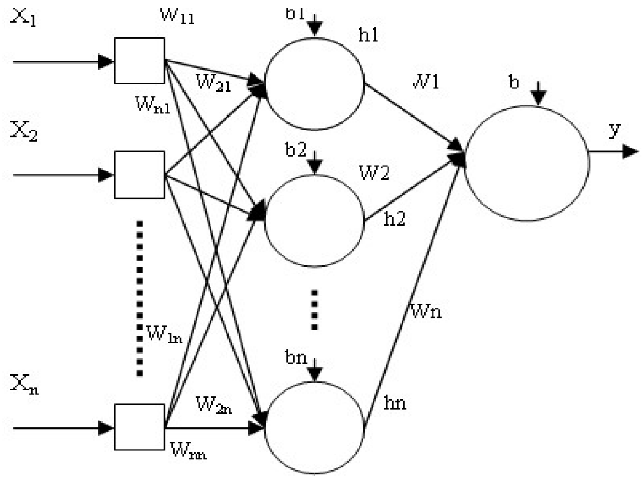


Figure 3: Neural Network construction

A classic framework of a feed-forward network contains input tier, output tier and hidden tier. The nodes of input layer obtain attribute values of every training example and send out the weighted result towards hidden tier. The weighted results of the concealed or hidden tier are contribution to unit making up the output tier, which transmits the expectation for certain examples [18], [24], [92]. Neural network has heavy learning capacity, which is one of the prime conveniences of neural network. ANN has a better accurateness than LR and DA. It had improved performance for appropriately classifying faulty loans than LR model. This model also represents data superior than LR and CART as per the claim of [93]. Some researchers have utilized neural networks for credit scoring and reported its precision as better than that of conventional statistical strategies in managing credit scoring issues [93]–[96]. As ANN can deal with composite data sets and don't need different assumptions like normality and linearity, [97] predicated that the ANN model broadly outperformed the LDA model in credit scoring. Two different neural network systems were proposed for evaluating credit hazard [84]. They experimented their suggested algorithms on a real-world data, and the experimental outcome reports that the neural network is a favorable algorithm for evaluating loan risk. A disadvantage of NN is that a number of parameters like the network topology must be defined analytically. Another major drawback of neural networks can be found in their weak understandability. It is difficult to make knowledge representation for ANN due to the nature of the "black box". Designing and optimizing the network topology is another issue. Moreover, the experiment process is very complex. Even though, the final classification result can be affected by the measure of units and layers in hidden layer part, the leading weight values and distinctive activation function. ANN needs countless specimens and long learning time. Hypothetically, different classifiers ought to perform superior to single classifiers. However, different neural system classifiers don't beat a solitary best neural network classifier in many cases.

### 4.2.5 Genetic Programming (GP)

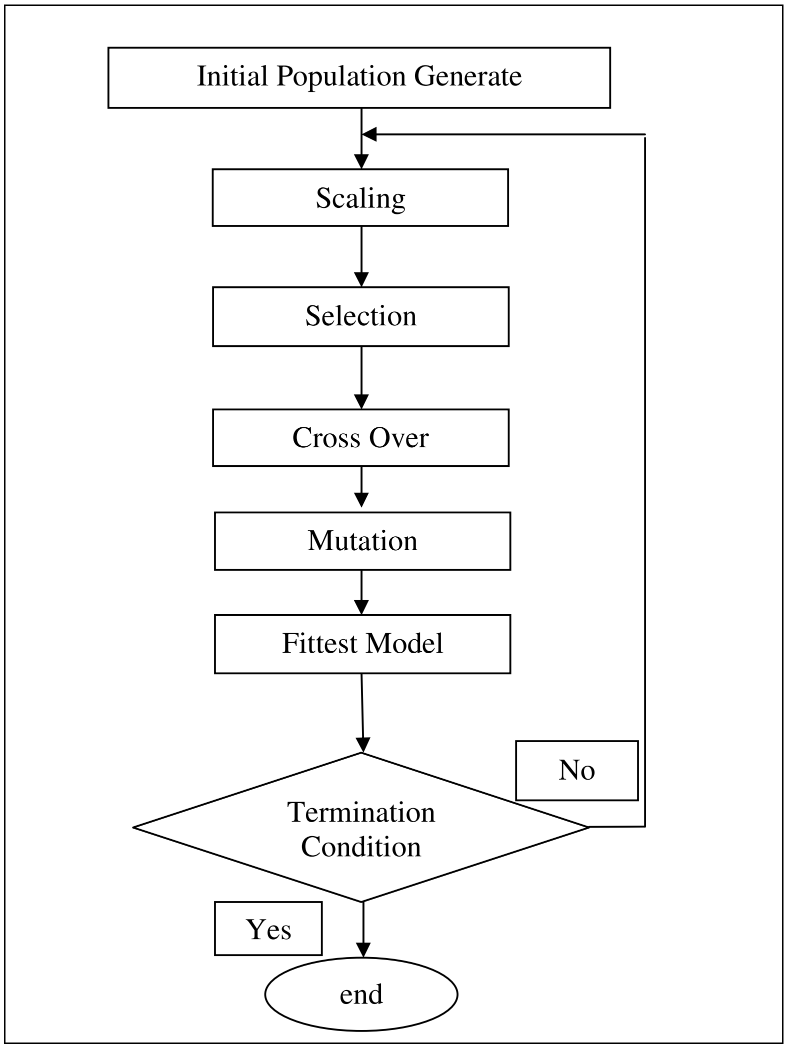


Figure 4: Steps of genetic algorithm (adopted from [2])

To automatically extract hypothetical relationship in an arrangement, genetic programming was first recommended by Koza [98], which has been applied in many applications, for example: classification and symbolic regression. It is a variant of the genetic algorithm (GA) realized by Holland in1970, which is based on the concept of adaptive survival in natural organisms.

GA is a search heuristic that mirrors the procedure of characteristics advancement. It changes a dataset as per fitness esteem by utilizing genetic operations. Having a place GA, the result is like a pattern of string. Each string obtains encrypted binary, real etc form of a nominee solution. A gauge work relates a wellness measure to each string speaking to its wellness for the case.

Generic operators are applied by Standard GA such as selection, crossover and mutation on an initially arbitrary populace keeping in mind the end goal to figure an entire era of new strings. Figure 5 illustrates the step of GA. A three step hybrid algorithm including reduction in search space, refinement in reduced feature subset, increment in design stages are proposed in [99]. Hybrid genetic algorithm and neural network (HGA-NN) is combined there, which is used in credit risk assessment to recognize most favorable feature subset and to raise the classification efficiency and scalability. Accuracy and coverage are two distinct conditions that fulfills the wellness capacity of the GA which is inferred in [100] as the composite measure of GA to find decision rules. The utilization of genetic programming applications is a quickly developing territory, and the quantity of uses has expanded last few years, such as bankruptcy forecasting and credit scoring [17], [43], [100]–[105].

## 4.3 Comparative Analysis

Table 2 gives a detail of various statistical methodologies alongside AI advancements utilized as a part of different articles by the researchers. This table incorporates the investigations, references and essential components of each one of those given strategies.

Table 2: Comparative Analysis of Various Techniques for Automated Credit Scoring from Distributed Research

| | METHODS | COMMENTS |
|---|---|---|
| Statistical approaches (SA) | LDA | Until now it is a standout amongst the most extensively settled strategies to arrange client's credit rating as good or awful by utilizing linear functions. On the other hand, DA can't appropriately manage non-linear issues. |
| | LR | It performs well on large dataset. In any case, this technique can be connected on a little dataset or an information set with a short reimbursement history, yet the nature of the scoring model can diminish. |
| | KNN | It empowers the displaying of anomalies in the function over the component space and a genuinely natural technique and can be utilized progressively however its prescient precision is to a great degree influenced by the measure of division and the cardinality of the zone. |
| | DT | Classification and regression problems are solved by this. As like LR, it requires large dataset keeping in mind the end goal to obtain reliable predictions. |
| AI technologies | Expert System | It doesn't wind up with a score card which offers weights to every reply rather it characterizes the customers into class, every class being homogeneous in its default risk. |
| | SVM | It renders worldwide ideal arrangement which can function admirably with few examples, yet choosing kernel and its specification is a precarious issue. |
| | Fuzzy Logic | It can construe human conspicuous lead and has low computational need, however, arbitrary decision of participation function can predisposition the outcome. |
| | NN | It is great at capacity estimate, predicting, characterization, grouping and improvement assignments, however, asks for a huge amount of training data and planning cycles. |
| | GP | It can perform superior to conventional procedures, for example, C4.5, MLP, CART, rough set and so on. After all, it is hard to turn out with a non specific structure for all type issues. Moreover, genetic programming requires great handling power. |

In light of the writing, Table 3 compares the adequacy of various strategies utilized as a part of credit scoring framework. It analyzes the accuracy (rate of effectively classified instances) of the strategies. By far maximum studies that concentrated on correlation among various methods of credit scoring have procured that AI technique, for instance, neural networks and genetic programming are better than the traditional ones considering the average correct classification rate criterion. Be that as it may, the other unambiguous categorization approaches such as, linear discriminant analysis along with logistic regression has a better accomplishment in this circumstance.

Table 3: An Examination of Various Strategies (SA and AI) Practiced in Credit Scoring Framework (In light of the literature)

| | | LR | LDA | DT | NN | CART | KNN | GP | SVM |
|---|---|---|---|---|---|---|---|---|---|
| **[106]** | **West(2000)** | 81.8 | 79.3 | 77.0 | **82.6** | 76.9 | 76.7 | | |
| **[42]** | **Lee et al. (2002)** | 73.5 | 71.4 | | 73.7 (**77.0**) hybrid LDA and NN | | | | |
| **[107]** | **Baesens(2003)** Avg of 8 dataset | 79.3 | 79.3 | 77.0 | 79.4 | | 78.2 | | **79.7** |
| **[43]** | **Ong et al. (2005)** | | 80.8 | 78.4 | 81.7 | | | **82.8** | |
| **[108]** | **Yu et al. (2008)** | 73.2 | | | 77.2 | | | | **78.8** |
| **[109][105]** | **Etemadi(2009) Shiri(2012)** | 75.0 | 71.3 | | 79.62 | | | **92.0** | 76.9 |
| **[110]** | **Tsai(2009)** | 84.7(avg) | 76.8 | | **92.7** | | | | |
| **[111]** | **Chuang(2009)** | 76.5 | 76.0 | | **79.5** | 77.5 | | | |
| **[112]** | **Peng(2011)** Avg of 7 dataset | **86.2** | | | | | 82.2 | | 85.9 |
| **[113]** | **Wang(2012)** | 71.6 | | 69.0 | 71.5 | | | | **72.4**(avg) |
| **[72]** | **Brown(2012)** Avg of 5 dataset | | | | **77.4** | | 74.5 | | 73.24 |

### 4.3.1 Analysis

Different approaches have several advantages and disadvantages according to their characteristics when using in credit scoring framework. One of the biggest comforts of KNN methodology is that it is not advantageous to form predictive model before classification. In credit scoring applications, it has some striking features. For instance, it is attainable to surpass the issue of population drift by utilizing KNN as it is dynamic to upgrade insistently by adding new contender to the outline and dropping previous cases. Regardless of having this quality KNN strategies have not been widely used in the credit scoring field, in light of the noticed computational need [71]. The Quadratic Discriminant Analysis (QDA) should be utilized if the covariance matrices of the given populations are not equal, which implies that, the partition surface of the discriminant function is quadratic. Some research emerged that in credit scoring LR exhibits better performance than LDA [60]. But using of logistic regression has a problem that it can't legitimately analyze the issues of non-linear and intelligent impacts on informative features. When using SVM, selection of a subset containing most favorable input items and most fitting kernel range settings problems may have confronted. Fuzzy logic has the ability of managing both the qualitative and quantitative factors. To construct credit scoring framework neural network with back propagation is applied comprehensively, in which neurons get signal from previous layers as well as send output to the next layers. But when the data sets are small in size, its performance is relatively poor, they are only suitable for vast feature sets. Whereas, genetic programming be able to accomplish well even on tiny feature sets [43].

### 4.3.2 Critical Analysis

After comparing genetic programming(GP) to decision trees, logistic regression(LR), rough sets, and neural network (NN) through two-true informational indexes, we can infer that GP can furnish preferred execution over different models with precision 82.8 % based on the outcomes [43]. The first data set incorporates Australian acknowledge scoring information for 307 instances of credit commendable clients and 383 instances for credit unworthy clients. It comprises fourteen attributes, where 6 are constant properties and 8 are absolute attributes. The German Credit Data Set is the second informational index that was given by Prof. Hofmann in Hamburg where 700 records considered to be credit respectable along with 300 records considered to be credit blamable. It incorporates client credit scoring information for twenty features, for example, age, sexual orientation, conjugal status, purpose of loan, previous credit history, occupation, account, other personal data, and so on. Genetic programming (GP) is more appropriate for the credit scoring issues for the accompanying reasons. Not at all like the standard quantifiable methods require the presumptions of the informational gathering and the characteristics, GP is a nonparametric mechanism that sensible for any conditions as well as informational collections. In comparison to artificial neural networks (ANNs), GP can decide the satisfactory discriminant work naturally as opposed to appointed the transfer work by leaders. Furthermore, GP can likewise choose the considerable variable consequently. At long last, the discriminant work which is determined by GP can give the preferred predicting execution over the enlistment based methods. Furthermore, not at all like ANNs which are acceptable for huge informational collections, GP can execute well even in limited informational collections. NN has most noteworthy precision around 79.5% in comparison with other classification strategies such as LR, LDA, CART and so on [111]. Here, a reassign organize is attempting to diminish the Type I blunder (a client with positive credit is wrongly classified as a client with terrible credit) by reassigning the rejected great credit candidates to the restrictive acknowledged class by utilizing the CBR-based grouping procedure. This may be a reason to get better predictive accuracy. Likewise MARS is utilized to acquire noteworthy input factors of the ANNs model to lessen the quantity of input hubs, improve the system structure and decrease the model constructing period. This is another motivation to acquire better exactness for NN. A scholastic dataset acquired from UCI repository of machine learning databases is received in this to assess the forecasting exactness where the information comprise of a collections of credit given to thousand credit card candidates. With a specific end goal to develop the scoring model twenty independent factors are utilized as a part of the dataset, for example, the candidate's age, previous credit history, loan amount, job, accommodation and so on. In most of the model, the dependent factor is the credit status of the client, that is, great or awful credit. By investigating the execution of credit scoring by incorporating the back-propagation neural networks together with conventional discriminant analysis (DA) technique, it can be watched that cross breed form has the most noteworthy accuracy around 77.00%, compared to NN (73.70%), DA (71.4%) and LR (73.5%) methods[42]. While there is no hypothetical strategy to decide the perfect input factors of a neural network system, DA methodology is executed as a for the most part acknowledged technique for deciding a decent subset of input factors. With a specific end goal to confirm the attainability and adequacy of the suggested hybrid system, data collections of credit card applicants got from a provincial bank in Taipei, Taiwan is utilized as a part of this investigation. Individual bank client in the data file includes 9 indicator factors, to be specific, sexual orientation, occupation, age, conjugal status, academic level, work position, yearly salary, home status as well as credit limits. Here also the response factor is the credit condition of the client great or awful credit. 6000 datasets concerning the proportion of good and awful credits were arbitrarily chosen and afterward used to assemble the credit scoring framework. Between them, 4000 samples will be utilized for training purpose as long as the 2000 will be held for testing purpose. Discriminant technique is embraced for variable choice method keeping in mind the end goal to construct the scoring model as well as obtain around 71.40% exactness. 6 imperative pointer factors are picked in the conclusive discriminant work, in particular sexual orientation, age, occupation, credit amount, yearly salary along with home status. At the point when LR strategy is utilized as a part of constructing credit scoring framework, 4 important factors, sex, age, credit cutoff points, and yearly pay were incorporated into the terminal regression model where the average accuracy is around 73.45%. In [72], Iain Brown chose 10 classifiers which give a harmony between entrenched scoring strategies, for example, LR, decision trees along with NN and recently created machine learning procedures, for example, random forests, gradient boosting as well as least square SVM. Each of these strategies assessed regarding their area under the receiver operating characteristic curve (AUC). This is a scale of the separation energy of a classifier without respect to class circulation or misclassification require. Here, we usually notice that the gradient boosting and random forest classifiers done well in managing tests where a substantial class irregularity was available. In addition, here NN act out superior to SVM with exactness 77.4 % while SVM has the precision around 73.24%. Five true credit scoring informational collections are utilized to fabricate classifiers and analysis their execution. This investigation contained German credit and Australian credit informational indexes , two informational indexes from Benelux (Belgium, Netherlands and Luxembourg) organizations where, an awful client was characterized as somebody who had missed 3 sequential months of installments , and the last informational index was a behavioral scoring informational collection, which was likewise acquired from a Benelux foundation. To improve the predictive accuracy a little bit, David west analyze the possibility of achievement by investigating 5 NN models [106]. German datasets and Australian datasets for credit scoring are utilized to test the forecasting precision of the credit scoring models. German credit scoring dataset comprises of 700 cases of financially sound candidates along with 300 illustrations where credit was not be expanded.

In German credit scoring dataset, twelve most imperative input features are - durability of account, work arrangement, financial records status, resources claimed including home and auto, years in living arrangement, other existing advances, lodging order, savings account condition, loan purpose, previous credit history, loan amount as well as years employed. Whereas, the Australian dataset is adjusted with 307 and 383 cases of every result. The outcome of this examination propose that NN is marginally more exact than the other scoring models with exactness of 82.6% where LR has the accuracy about 81.1% .The nonparametric models like KNN and CART did not deliver empowering results but rather may exhibit enhanced exactness for expansive data collections. In [109] the investigation test comprises of 144companies recorded in Tehran stock trade. Here, the datasets utilize money related proportions (e.g. liquidity, dissolvability, asset, gainfulness, resource organization, firm size, development), income data, and different factors of intrigue that incorporate data on macroeconomic, firm particular factors and so on. A few factors of the recorded organizations' money related status have been chosen to examine the budgetary misery. Five pre-planned prediction factors are chosen to exhaustively determine the effective forecast factors of final models. Then the informational indexes are actualized on different data mining calculations, for example, NN, LR, SVM, Bayesian system along with Decision trees (CART, QUEST, CHAID, and C5) and notified that the CART method has higher predictive accuracy about 94.93%.

## 4.4 Research gaps in this area

Apart from various Artificial Intelligence and statistical technique which had been already used in credit risk management mechanism, we can also use Gray theory. There is not much credit risk management system built on Gray theory. By using Gray theory on developing credit risk management system, it will be more systematic and scientific. There are few other works which has not been used yet, such as: customer retention, customer profit analysis, market basket analysis. If we can build a credit scoring system using these, we will be able to get better predictive accuracy. Variable selection is very important to build the scoring framework. If it is not done correctly, whole credit scoring structure can get destroyed. We can use the above-mentioned procedures for variable selection. Apart from that, we did not find much study where money attitude was utilized for scoring model. The money or finances attitude that impacts all parts of human life, and incorporated the spending conduct which identified with the loan behavior. As the customer credit default behavior will be perhaps impacted by ones attitude about money, it should be estimated by utilizing the poll, yet the banking industry may in any case need to consider with the effects on different angles when basically directing such survey study. In this way, this examination recommended the future researchers and scientists who might direct the exploration investigation on going for the possibility of all intents and purposes executing such poll overview to give related ventures easily advancing the arrangement of their survey about money attitude. Additionally, we cannot get best effectiveness if scoring model is constructed for longer time. Client's credit status should be checked at a regular interval to get an idea about faulty customers. Look into need to build up time-series credit scoring models that incorporate the difference in credit condition in each time interval. By utilizing a strategy with time-series credit scoring model, credit candidates can be classified into more subgroups because of new factors. Additionally, more developed administration systems for the clients in the subgroups can be set down. Another gap in this area is that, big data or cloud computing was also not used for constructing scoring model. Furthermore, with the development of big data as well as cloud computing, many monetary institutions may change their credit scoring framework. Data mining is a very effective technology to solve big data.

## 4.5 Conclusion

After comparing Statistical and Artificial Intelligence (AI) methods that are utilized for automated credit scoring system, we found Neural networks (NN) and genetic programming (GP) are better for prediction purposes. Time to time different researchers had applied different methods of credit scoring framework on same dataset, sometimes they applied them to other datasets and discovered the result. After investigating several result, we found that NN and GP has better predictive ability as their accuracy is higher than other methods. Moreover, be that as it may, the other unambiguous categorization approaches such as, linear discriminant analysis along with logistic regression has a better accomplishment capability. This analysis notified that there is no supreme statistical approach employed to construct credit scoring framework which works on all circumstances.

# Chapter 5

# Credit Scoring in Developing Countries

## 5.1 Developing countries and credit scoring

The purpose of credit scoring is to quantify the budgetary danger of the advance, so that the advance supplier can settle using a loan loaning choices rapidly and impartially. As credit scoring empowers advancers to survey the reliability rapidly, banks in created countries, for example, the US, UK and Europe, have effectively moved far from utilizing human judgment and dominatingly utilize acknowledge scoring procedures for sensible achievement [114] . To diminish the cost for recognizing reliable candidates, there are expansive credit scoring firms like Equifax, Transunion and so on. Despite the fact that in creating nations credit scoring can be awkward without solid data about the credit or budgetary history of potential bank customers for quite a long while. Presently many creating nations have thought about attempting to locate a mechanized credit scoring framework that would work for them. In some of these nations, as of now have utilized credit scoring framework which has been planned by created countries and huge numbers of them are attempting to make their own particular credit scoring frameworks to give advance in the modern area[105]. Be that as it may, credit scoring has not been honed viably in little money related ranges like home Icons, charge cards or individual advances. It is normal that coordination of computerized credit scoring framework in creating nation could convey advantage to the money related part and economy. As the money related associations can figure out if there is a hazard or not to allow the advance to the client. In addition, in creating nations little and medium ventures (SMEs) are thought to be a critical wellspring of advancement and business as a result of their adaptability in reacting to new market opinion and business as a result of their adaptability in reacting to new market opening and their potential for development. Many creating nations have as of now initiated to rehearse robotized acknowledge scoring as an apparatus for their financial improvement.

Table 4 gives a thought of adjustment of credit scoring techniques and advancement being utilized for creating nations.

Table 4: Summary of Research Work That Consider Developing Countries For Practicing Automated Credit Scoring Framework

| Country | Reference | Data set | Using method or Domain | Adaptation of credit scoring |
|---|---|---|---|---|
| **Bangladesh** | [115] | Information were gathered from 156 towns in three locale in northern Bangladesh (Nilphamary, Kurigram and Rangpur) as a feature of the standard study for BRAC | OLS regression | Researches the impact of MFI program intercession on moneylender premium costs in northern side of Bangladesh and gain that moneylender funding expenditure increment with the percentage of families getting from MFIs in the city. |
| **India** | [116] | 5 years information for 9 organizations were separated from the Indian manufacturing division which had documented for bankruptcy | Altman Z-score and KMV Merton Distance | They tried to determine how far off these prediction models can expect that the arrangements would be in monetary trouble. |
| | [117] | Information sets are taken from three sorts of banks with the assistance of Center for Monitoring Indian Economy Pvt Ltd as effective banks, recouped banks and backslide bank. | GA, Fuzzy c-means Clustering and MARS | Conceptualizes the cross breed simulation of genetic algorithm, Fuzzy c-means calculation, MARS for expectation of bankruptcy and demonstrated that this model improvise all around contrasted with static Bankruptcy Models. |
| **Pakistan** | [118] | Karachi Stock Exchange recorded 50 organizations, which are taken. | Abbas model and Altman's model | Demonstrated that Abbas simulation and Altman's Z score scheme was a powerful instrument for checking the fiscal strength of nonfinancial organizations recorded at Karachi stock trade. |
| | [119] | Karachi Stock Exchange recorded 52 organizations, which are taken with some conditions | MDA | Recognized the fiscal proportions that are most noteworthy in foreseeing bankruptcy in Pakistan. |
| **Malaysia** | [120] | Bursa Malaysia recorded 30 troubled and untroubled organizations that are taken | LR | Analyze the elements of credit risk along with showing that the cash out proportion was extremely large in determining credit risk. |
| **Iran** | [109] | The information gathered from 144 organizations situated at the Tehran stock exchange | CART,C5, QUEST,CHAID,SVM,NN, BN,LR,DA | Created Information mining models form financial data of an association to arrange the non-bankrupt and bankrupt firms. |
| | [105] | Information sets are comprises of 144 Iranian organizations recorded on the Tehran Stock Exchange (TSE) | GP and MDA | Utilized GP and MDA to see if it is conceivable to anticipate the survival or disappointment of Iranian partnerships in view of financial rations. |
| **Ghana** | [121] | The experimentation was done utilizing Kwikplus Micro Finance as a contextual investigation. Information gathered was helped out through meeting and perceptions of the operations the firm | Fuzzy logic | Given a Fuzzy approach to deal with credit scoring keeping in mind the end goal to lessen the loan default among the Micro-Finance Organizations. |

| | | | | |
|---|---|---|---|---|
| **Nepal** | [122] | 31 banks from Nepal provided this data | Econometric model | Showed such loan hazard supervision is a crucial pointer of bank's monetary performance. Thereby, acquirement of bank achievement depends on risk authority. |
| **Sudan** | [123] | 45 banks from Sudan helped providing this data | LR and DA | Proposed to foresee bank's disappointment. |
| **Vietnam** | [5] | Vietnam's stock exchange provided data of 643 organizations using their reported economical valuation | Fuzzy logic | Applied a technique to solve the current flaws in credit assessment and turn over a casual formulation of credit scoring value. |
| **Turkey** | [124] | Around 100 firms of Turkey provided this data | DEA,RA and DA | Improved the determinable investigation used in the context of financial accomplishment unit of credit scoring framework. |

## 5.2 Aspect of Credit Scoring in Developing Countries

In growing countries, monetary system is for the most part micro-finance. Thusly, likelihood of credit scoring in those countries is related to alteration in microfinance of credit scoring.

A couple of suggestions for microfinance associations and micro-lenders in progressing countries.

I. The amount of dispersed credit scoring ponders for microfinance is compelled. There is a need to expand the topographical scope of credit scoring looks towards Eastern Europe-Central Asia and Middle East-North Africa as meager amounts of studies have been appropriated in these geographical regions.

II. The onerous power execution of credit scoring structures for microfinance remains exorbitantly weak, making it difficult to legitimize an entire reversal of the traditional credit process towards scoring. Be that as it may, credit scoring should wind up a refinement instrument in the present methodology as it has successfully wound up being relentless, easy to use, besides to have a particular oppressive power. Upgrades of the unfair power by methods for show blends dismiss enlistment examination and more realistic affirmation, may a tiny bit at a time fabricate the piece of credit scoring in the credit procedure.

As there are very little satisfying credit scoring arrangement in progressing countries, we ought to consider proficient ways to deal with actualize better credit scoring framework. Conceivable methodologies that can be assumed to execute effective credit scoring in those nations:

- Pointers of lead collected from phone trade records can be farsighted of credit portion [48]. To gage this wise method for the procedure, examine have been made where combined bank data from credits have been done with account holder's phone records. It make sense of who among these people groups wound up repaying their credit, in light of how they used their phone before taking a progress. Moreover, the investigation found that the judicious exactness of the method approaches that of credit scoring methodologies utilizing conventional information as a part of more created settings.

- Involving web-based social networking in credit scoring can be productive route in developing nation. An immense number of individuals run facebook and following their web based social networking action can help loan specialists compute the hazard variables. Moreover, there are contradictory sides of this too. The practice of "enormous information" in promoting to target specific customer social occasions is starting now a faulty exercise, generally since couple of clients ever recognizes they are being followed.

- Microfinance industry faces a big challenge on constructing long -term relationship with their customers. Generally, first credits are quantity and short-haul loans. Therefore, to get in the market, a distinctive arrangement is required. If all the conditions are accomplished by the customer, financial organizations can provide them higher amounts. It gives the customer a motive to stay with the organization.

- It is also very important for banks to concentrate on security. It is important so that, guarantee credits depend on the customer's ability to reimburse and in addition execution in business. Operational expenses remain an industrious issue even if the default rate is low. Lenders need to discover approaches to decrease operational expenses along with diminishing lose. This is the place credit scoring becomes an integral factor. It can help not just in the underlying choice of potential microfinance customers, additionally in the recognizable proof of the best customers for new sorts of services.

## 5.3 Applications

In developing countries like Bangladesh, we can use credit scoring in various areas such as bankruptcy prediction and liquidation classification. Liquidity problem is a big problem in our country now a day. We can get rid of these problems if we can properly practice credit scoring. Banks provide loans to various organizations. By using credit scoring bank will be able to predict reliability of their customer better. Moreover, we will be able to use credit scoring on Small and Medium Enterprises (SME) and micro-finance. In our country, there is a tendency of taking small loans, though our banks provide very less opportunities for it. For many years, Grameen Bank had been providing such small loans. They even provide small loans to buy a cow or goat or even sewing machine so that one can easily earn their livelihood or start a small business. These small loans changed the life of many people. We all should think about how we can facilitate these small loans to the poor people. By properly using credit scoring, we will be able to encourage other small to large scaled banks to provide such loan.

## 5.4 Conclusion

The amount of credit scoring research for microfinance ought to be broadened. At this point some researchers need to be assigned in credit scoring discipline. Whereas researchers in one field has no time and prominence to pay attention to research in other fields. Moreover, they need to be inventive in implementing conceivably important features as general and economic change in scoring procedures. Microfinance industry needs to avoid long term relationship with their clients. First time they generally grant small and short-haul loans. After accomplishing all the terms and conditions, clients may get higher amounts according to their requirements. In the case of involving web-based social media in credit scoring system, we must be aware that, all the information which are provided by the consumers are authentic. Because these informations play a very important role in constructing effective scoring model.

# Chapter 6

# Conclusion

## 6.1 Discussion

At different time of doing this thesis, I studied 124 papers published on credit scoring and tried to compare different type of information, result which I gathered. I was able to gather specific findings which will help the next generation to work on credit scoring. In this paper, I made an effort to provide basic idea regarding credit rating, credit scoring system and judgmental system. Moreover, this study attempted to explain why we should be using automated credit scoring system instead of judgmental system which were used previously. Apart from having so many advantages, credit scoring system has few limitations as well, which are discussed in this paper. We also discussed about using various variables on different situations of creating credit scoring framework. We discussed about areas we should use credit scoring to get an effective result in this paper. We contended different statistical and AI techniques which can be used to construct credit scoring framework and also did a comparative analysis on these. We tried to provide an idea about practicing credit scoring on developing countries along with prospect of credit scoring. Our analysis has provided few recommendations and methodologies which will help working on credit scoring in our country in future.

## 6.2 Limitations

In recent time, credit scoring is a rising topic, but I had to face various difficulties on different time while working with it. Like, I did not find any worth mentioning paper for credit scoring on developing countries like ours. Due to the lack of available research paper on credit scoring of developing countries, it was very difficult for me to do analysis and compare as the informations are not sufficient enough. I had to face problem for creating optimal feature set as well where it was difficult for me to find which variable is suitable for our country at what situation. From these, we can realize how far we are to practice automated credit scoring in our country.

## 6.3 Conclusion

Credit scoring is a comprehensively used strategy that helps bank and other monetary institutions to decide whether to permit credit to clients who applied for loans. Nonetheless, with movement of innovation, the method for credit assessment ought to be updated. Using colossal information to decide lower, customized rates on credit cards and loans will advantage fiscally tried and true people in such a way that current system does not, allowing dependable debtor to reimburse little and avoid commitment speedier. Also, framing a customer base of financially mindful people will advantage banks – decreasing the danger of deception, and moreover default and economical institutions cash as time goes on. Putting assets into the fate of stable individuals is like placed assets into the fate of stable fiscal association and national economy. It is the incomparable shot for the monetary associations to snatch a greater degree of markers to decide budgetary concern; fundamentally the past strategies for assessment can't remain mindful of better systems for living and utilizing. This begins with party a better perception of purchasers than make money related game plans that meet their individual needs. Instruments are being made that will make advancing more gainful to both the banks and the account holders. Equipped with cable programming and refined data science, the inevitable destiny of the financial associations lays on the capacity to carry taught singular back into the present time span.

## 6.4 Future Work

Working on automated credit scoring can go a long way. I would try to work more on this in future. I want to collect information from various banks and other financial organizations regarding the criteria they choose before providing microcredit or small loans. Gathering these information, I would like to analyze and construct an effective scoring model which will be suitable for our country. Moreover, I will try to use AI approach such as Neural Network, Genetic Programming, or Fuzzy logic in order to create scoring model as using these will make the model more effective.

## References


[1] T. Van Gestel, B. Baesens, J. a. K. Suykens, D. Van den Poel, D.-E. Baestaens, and M. Willekens, "Bayesian kernel based classification for financial distress detection," *Eur. J. Oper. Res.*, vol. 172, no. 3, pp. 979–1003, 2006.

[2] Y. Wang, S. Wang, and K. Lai, "A new fuzzy support vector machine to evaluate credit risk," *Fuzzy Syst. IEEE Trans. …*, vol. 13, no. 6, pp. 820–831, 2005.

[3] M. C. Chen and S. H. Huang, "Credit scoring and rejected instances reassigning through evolutionary computation techniques," *Expert Syst. Appl.*, vol. 24, no. 4, pp. 433–441, 2003.

[4] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 3302–3308, 2009.

[5] V. O. Hong and N. Dinh, "A new approach to determining credit rating & its applications to vietnam 's listed firms," *42nd Aust. Conf. Econ.*, no. April 2013, pp. 1–26, 2013.

[6] Root, "Credit Rating," Investopedia, LLC 2015.

[7] "Credit score," Wikipedia, 2015 .

[8] H. a Abdou and J. Pointon, "Credit Scoring, Statistical Techniques and Evaluation Criteria: A Review of the Literature," *Intell. Syst. Accounting, Financ. Manag.*, vol. 18, no. 2–3, pp. 59–88, 2011.

[9] R. Systems, "Crafting a Market Landscape," no. February, 2000.

[10] L. J. Mester, "What's the Point of Credit Scoring ?," *Bus. Rev.*, vol. 3, pp. 3–16, 1997.

[11] T. Jacobson and K. Roszbach, "Bank lending policy, credit scoring and value-at-risk," *J. Bank. Financ.*, vol. 27, no. 4, pp. 615–633, 2003.

[12] "avery2000.pdf." .

[13] D. J. Hand, "Modelling consumer credit risk," *IMA J. Manag. Math.*, pp. 139–155, 2001.

[14] H. Koh, W. Tan, and C. Goh, "A two-step method to construct credit scoring models with data mining techniques," *Int. J. Bus. …*, vol. 1, no. 1, pp. 96–118, 2006.

[15] M. Schreiner, "Benefits and Pitfalls of Statistical Credit Scoring for Microfinance Author ' s Note Benefits and Pitfalls of Statistical Credit Scoring for Microfinance 1 . What is scoring ?," vol. 28, no. 1, pp. 63–86, 2004.

[16] M. Janeska, K. Sotiroski, and S. Taleska, "Application of the Scoring Model for Assessing the Credit Rating of Principals," vol. 3, no. 1, pp. 1–5, 2014.

[17] J. H. Min and C. Jeong, "A binary classification method for bankruptcy prediction," *Expert Syst. Appl.*, vol. 36, no. 3 PART 1, pp. 5256–5263, 2009.

[18] C. Tsai and J. Wu, "Using neural network ensembles for bankruptcy prediction and credit scoring," vol. 34, pp. 2639–2649, 2008.

[19] Y.-C. Hu, "Incorporating a non-additive decision making method into multi-layer neural networks and its application to financial distress analysis," *Knowledge-Based Syst.*, vol. 21, pp. 383–390, 2008.

[20] J. N. Crook, D. B. Edelman, and L. C. Thomas, “Recent developments in consumer credit risk assessment,” *Eur. J. Oper. Res.*, vol. 183, no. 3, pp. 1447–1465, 2007.

[21] D. Hand, S. Sohn, and Y. Kim, “Optimal bipartite scorecards,” *Expert Syst. Appl.*, vol. 29, no. 3, pp. 684–690, 2005.

[22] T. H. T. Dinh and S. Kleimeier, “A credit scoring model for Vietnam’s retail banking market,” *Int. Rev. Financ. Anal.*, vol. 16, no. 5, pp. 471–495, 2007.

[23] M. Sustersic, D. Mramor, and J. Zupan, “Consumer credit scoring models with limited data,” *Expert Syst. Appl.*, vol. 36, no. 3, pp. 4736–4744, 2009.

[24] T. Lee and I. Chen, “A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines,” vol. 28, pp. 743–752, 2005.

[25] B. Baesens, T. Van Gestel, M. Stepanova, D. Van den Poel, and J. Vanthienen, “Neural network survival analysis for personal loan data,” *J. Oper. Res. Soc.*, vol. 56, no. 9, pp. 1089–1098, 2005.

[26] M. Bensic, N. Sarlija, and M. Zekic-Susac, “Modelling small-business credit scoring by using logistic regression, neural networks and decision trees,” *Intell. Syst. Accounting, Financ. Manag.*, vol. 13, no. 3, pp. 133–150, 2005.

[27] W. S. Frame, A. Srinivasan, L. Woosley, J. Money, and N. Aug, “The Effect of Credit Scoring on Small-Business Lending,” *J. Money*, vol. 33, no. 3, pp. 813–825, 2007.

[28] A. Blanco, R. Pino-Mejías, J. Lara, and S. Rayo, “Credit scoring models for the microfinance industry using neural networks: Evidence from Peru,” *Expert Syst. Appl.*, vol. 40, no. 1, pp. 356–364, 2013.

[29] J. Gupta, N. Wilson, A. Gregoriou, and J. Healy, “The effect of internationalisation on modelling credit risk for SMEs: Evidence from UK market,” *J. Int. Financ. Mark. Institutions Money*, vol. 31, no. October, pp. 397–413, 2014.

[30] A. Haughwout, R. Peach, and J. Tracy, “Juvenile delinquent mortgages: Bad credit or bad economy?,” *J. Urban Econ.*, vol. 64, no. 2, pp. 246–257, 2008.

[31] a Heuson, W. Passmore, and R. Sparks, “Credit scoring and mortgage securitization: Implications for mortgage rates and credit availability,” *J. Real Estate Financ. Econ.*, vol. 23, no. 3, pp. 337–363, 2001.

[32] C. a E. Goodhart, “The background to the 2007 financial crisis,” *Int. Econ. Econ. Policy*, vol. 4, no. 4, pp. 331–346, 2008.

[33] J. Y. Lin, “The Impact of the Financial Crisis on Developing Countries,” 2008.

[34] D. Platz, “Infrastructure finance in developing countries — the potential of sub-sovereign bonds,” *DESA Work. Pap.*, no. 76, 2009.

[35] D. Gültekin-Karakaş, M. Hisarcıklılar, and H. Öztürk, “Sovereign Risk Ratings: Biased Toward Developed Countries?,” *Emerg. Mark. Financ. Trade*, vol. 47, no. 0, pp. 69–87, 2011.

[36] V. V Acharya, I. Drechsler, and P. Schnabl, “A Pyrrhic Victory ? Bank Bailouts and Sovereign Credit Risk,” *Money*, vol. LXIX, no. 6, pp. 1–40, 2011.

[37] R. Al-Sakka and O. ap Gwilym, “Split sovereign ratings and rating migrations in emerging economies,” *Emerg. Mark. Rev.*, vol. 11, no. 2, pp. 79–97, 2010.

[38] S. Brana and D. Lahet, "Determinants of capital inflows into Asia: The relevance of contagion effects as push factors," *Emerg. Mark. Rev.*, vol. 11, no. 3, pp. 273–284, 2010.

[39] "Credit rating agencies and their potential impact on developing countries," no. 186, 2008.

[40] D. Ratha, P. K. De, and S. Mohapatra, "Shadow Sovereign Ratings for Unrated Developing Countries," *World Dev.*, vol. 39, no. 3, pp. 295–307, 2011.

[41] M. K. Lim and S. Y. Sohn, "Cluster-based dynamic scoring model," *Expert Syst. Appl.*, vol. 32, no. 2, pp. 427–431, 2007.

[42] T.-S. Lee, C.-C. Chiu, C.-J. Lu, and I.-F. Chen, "Credit scoring using the hybrid neural discriminant technique," *Expert Syst. Appl.*, vol. 23, no. 3, pp. 245–254, 2002.

[43] C. Ong, J. Huang, and G. Tzeng, "Building credit scoring models using genetic programming," *Expert Syst. Appl.*, vol. 29, no. 1, pp. 41–47, 2005.

[44] M. Leshno and Y. Spector, "Neural network prediction analysis: The bankruptcy case," *Neurocomputing*, vol. 10, no. 2, pp. 125–147, 1996.

[45] T. Bellotti and J. Crook, "Support vector machines for credit scoring and discovery of significant features," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3302–3308, 2009.

[46] J. Banasik and J. Crook, "Reject inference, augmentation, and sample selection," *Eur. J. Oper. Res.*, vol. 183, no. 3, pp. 1582–1594, 2007.

[47] L. M. Salchenberger, E. M. Cinar, and N. a. Lash, "Neural Networks: A New Tool for Predicting Thrift Failures," *Decis. Sci.*, vol. 23, no. 4, pp. 899–916, 1992.

[48] V. Limsombunchai, C. Gan, M. Lee, C. Division, P. O. Box, and N. Zealand, "An analysis of credit scoring for agricultural loans in Thailand," *Am. J. Appl. Sci.*, vol. 2, no. 8, p. 1198, 2005.

[49] M. Landajo, J. de Andrés, and P. Lorca, "Robust neural modeling for the cross-sectional analysis of accounting information," *Eur. J. Oper. Res.*, vol. 177, no. 2, pp. 1232–1252, 2007.

[50] B. Soediono, "Risk Elements in Intensity, Energy and Acceleration," *J. Chem. Inf. Model.*, vol. 53, p. 160, 1989.

[51] David Durand, "Risk Elements in Consumer," vol. I, 1941.

[52] E. I. Altman, "Financial ratios, discriminant analysis and the prediction of corporate bankruptcy," *J. Finance*, vol. XXIII, no. September, 1968.

[53] G. Lanine and R. Vander Vennet, "Failure prediction in the Russian bank sector with logit and trait recognition models," *Expert Syst. Appl.*, vol. 30, no. 3, pp. 463–478, 2006.

[54] I.-C. Yeh and C. Lien, "The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2473–2480, 2009.

[55] T.-H. Lin, "A cross model study of corporate financial distress prediction in Taiwan: Multiple discriminant analysis, logit, probit and neural networks models," *Neurocomputing*, vol. 72, no. 16–18, pp. 3507–3516, 2009.

[56] X. Xu and Y. Wang, "Financial failure prediction using efficiency as a predictor," *Expert Syst. Appl.*, vol. 36, no. 1, pp. 366–373, 2009.

[57] M. A. Boyacioglu, Y. Kara, and Ö. K. Baykan, "Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 3355–3366, 2009.

[58] R. a. Eisenbeis, "Problems in applying discriminant analysis in credit scoring models," *J. Bank. Financ.*, vol. 2, pp. 205–219, 1978.

[59] H. a Abdou, "An evaluation of alternative scoring models in private banking," *J. Risk Financ.*, vol. 10, no. 1, pp. 38–53, 2012.

[60] V. B. Kronborg Dorte, Tjur Tue, *Credit scoring: Discussion of methods and a case study.* .

[61] Y. Liu, "A framework of data mining application process for credit scoring," no. 01, 2002.

[62] P. Taylor, S. J. Press, and S. Wilson, "Journal of the American Statistical Association Choosing between Logistic Regression and Discriminant Analysis Choosing Between logistic Regression and Discriminant Analysis," no. October 2013, pp. 37–41, 2012.

[63] H. Zhao, A. P. Sinha, and W. Ge, "Effects of feature construction on classification performance: An empirical study in bank failure prediction," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2633–2644, 2009.

[64] L. C. Thomas, K. M. Jung, S. D. Thomas, and Y. Wu, "Modeling consumer acceptance probabilities," *Expert Syst. Appl.*, vol. 30, no. 3, pp. 499–506, 2006.

[65] L.-C. Chi and T.-C. Tang, "Bankruptcy Prediction: Application of Logit Analysis in Export Credit Risks," *Aust. J. Manag.*, vol. 31, no. 1, pp. 17–27, 2006.

[66] Melek Acar Boyacioglu, Yakup Kara, and Omer Kaan Baykan, "Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods: A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey," *Expert Syst. Appl.*, vol. 36, no. 2, pp. 2 3355–3366, 2009. .

[67] S. Cabras and J. Morales, "Extreme Value Analysis within a Parametric Outlier Detection Framework," *Appl. Stoch. Model. Bus. Ind.*, vol. 23, no. 0316, pp. 157–164, 2007.

[68] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," *J. Oper. Res. Soc.*, vol. 54, no. 6, pp. 627–635, 2003.

[69] W. E. HENLEY, D. J.HAND, and KeynesMilton, "Construction of a k-nearest-neigbbour credit-scoring systemf," pp. 305–321, 1997.

[70] N. B, "Computer Science and Management Studies A Review on Data Mining and its Techniques in Finance Sector," *Int. J. Adv. Res.*, vol. 3, no. 3, pp. 29–37, 2015.

[71] A. D. V Lindley, S. Journal, R. Statistical, and S. Series, "Statistical Classffication Methods in Consumer Credit Scoring: a Review," vol. 135, no. 3, pp. 462–463, 2014.

[72] I. Brown and C. Mues, "An experimental comparison of classification algorithms for imbalanced credit scoring data sets," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3446–3453, 2012.

[73] R. S. Society, "for assessing consumer credit risk classifier A k-nearest-neighbour," vol. 45, no. 1, pp. 77–95, 2014.

[74] T.-S. Lee, C.-C. Chiu, Y.-C. Chou, and C.-J. Lu, “Mining the customer credit using classification and regression tree and multivariate adaptive regression splines,” *Comput. Stat. Data Anal.*, vol. 50, no. 4, pp. 1113–1130, 2006.

[75] M. Learning, M. Learning, K. A. Publishers, K. A. Publishers, A. C. Sciences, A. C. Sciences, R. August, and R. August, “Induction of Decision Trees,” *Expert Syst.*, pp. 81–106, 2007.

[76] P. Ravi Kumar and V. Ravi, “Bankruptcy prediction in banks and firms via statistical and intelligent techniques - A review,” *Eur. J. Oper. Res.*, vol. 180, no. 1, pp. 1–28, 2007.

[77] J. S. Ang, J. H. Chua, and C. H. Bowling, “The Profiles of Late-Paying Consumer Loan Borrowers: An Exploratory Study: Note,” *J. Money, Credit Bank.*, vol. 11, no. 2, p. 222, 1979.

[78] S. Piramuthu, “On preprocessing data for financial credit risk evaluation,” *Expert Syst. Appl.*, vol. 30, no. 3, pp. 489–497, 2006.

[79] J. Stefanowski and S. Wilk, “Evaluating business credit risk by means of approach-integrating decision rules and case-based learning,” *Int. J. Intell. Syst. Accounting, Financ. Manag.*, vol. 10, no. 2, pp. 97–114, 2001.

[80] A. Keramati and N. Yousefi, “A Proposed Classification of Data Mining Techniques in Credit Scoring,” pp. 416–424, 2011.

[81] C. W. Holsapple, K. Y. Tam, A. B. Whinston, and Y. Tam, “Adapting Expert System Technology to Financial Management,” vol. 17, no. 3, pp. 12–22, 1988.

[82] A. Bahrammirzaee, “A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems,” *Neural Comput. Appl.*, vol. 19, no. 8, pp. 1165–1195, 2010.

[83] K. Bryant, “ALEES: an agricultural loan evaluation expert system,” *Expert Syst. Appl.*, vol. 21, no. 2, pp. 75–85, 2001.

[84] E. Angelini, G. di Tollo, and A. Roli, “A neural network approach for credit risk evaluation,” *Q. Rev. Econ. Financ.*, vol. 48, no. 4, pp. 733–755, 2008.

[85] M.-C. Lee, “Enterprise Credit Risk Evaluation models: A Review of Current Research Trends,” *Int. J. Comput. Appl.*, vol. 44, no. 11, pp. 37–44, 2012.

[86] D. Niklis, M. Doumpos, and C. Zopounidis, “Combining market and accounting-based models for credit scoring using a classification scheme based on support vector machines,” *Appl. Math. Comput.*, vol. 234, pp. 69–81, 2014.

[87] L. Zhou, K. K. Lai, and J. Yen, “Bankruptcy prediction using SVM models with a new approach to combine features selection and parameter optimisation,” *Int. J. Syst. Sci.*, vol. 45, no. July 2015, pp. 1–13, 2012.

[88] P. Danenas and G. Garsva, “Selection of Support Vector Machines based classifiers for credit risk domain,” *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3194–3204, 2015.

[89] C.-L. Huang, M.-C. Chen, and C.-J. Wang, “Credit scoring with a data mining approach based on support vector machines,” *Expert Syst. Appl.*, vol. 33, no. 4, pp. 847–856, 2007.

[90] E. Kamalloo and M. S. Abadeh, “Credit Risk Prediction Using Fuzzy Immune Learning,” vol. 2014, 2014.

[91] Y.-R. Syau, H.-T. Hsieh, and E. S. Lee, "Fuzzy Numbers in the Credit Rating of Enterprise Financial Condition," *Rev. Quant. Financ. Account.*, vol. 17, no. 4, pp. 351–360, 2001.

[92] X. Li and Y. Zhong, "An Overview of Personal Credit Scoring : Techniques and Future Work," vol. 2012, no. October, pp. 181–189, 2012.

[93] M. Zekic-Susac, N. Sarlija, and M. Bensic, "Small business credit scoring: a comparison of logistic regression, neural network, and decision tree models," *26th Int. Conf. Inf. Technol. Interfaces*, vol. 1, pp. 265 – 270, 2004.

[94] R. Malhotra and D. K. Malhotra, "Evaluating consumer loans using neural networks," *Omega*, vol. 31, no. 2, pp. 83–96, 2003.

[95] and Z. H. Maher Alaraj, Maysam Abbod, "Evaluating Consumer Loans Using Neural Networks Ensembles," *Int. Conf. Mach. Learn. Electr. Mech. Eng.*, vol. 12, no. 10, pp. 993–1001, 2014.

[96] V. S. Desai, J. N. Crook, and G. a. Overstreet, "A comparison of neural networks and linear scoring models in the credit union environment," *Eur. J. Oper. Res.*, vol. 95, no. 1, pp. 24–37, 1996.

[97] K. Kumar and S. Bhattacharya, "Artificial neural network vs linear discriminant analysis in credit ratings forecast: A comparative study of prediction performances," *Rev. Account. Financ.*, vol. 5, no. 3, pp. 216–227, 2006.

[98] P. J. Angeline, *Genetic programming: On the programming of computers by means of natural selection,* vol. 33, no. 1. 1994.

[99] S. Oreski and G. Oreski, "Genetic algorithm-based heuristic for feature selection in credit risk assessment," *Expert Syst. Appl.*, vol. 41, no. 4, pp. 2052–2064, 2014.

[100] M.-J. Kim and I. Han, "The discovery of experts' decision rules from qualitative bankruptcy data using genetic algorithms," *Expert Syst. Appl.*, vol. 25, no. 4, pp. 637–646, 2003.

[101] S. Finlay, "Credit scoring for profitability objectives," *Eur. J. Oper. Res.*, vol. 202, no. 2, pp. 528–537, 2010.

[102] P. Ravisankar, V. Ravi, and I. Bose, "Failure prediction of dotcom companies using neural network–genetic programming hybrids," *Inf. Sci. (Ny).*, vol. 180, no. 8, pp. 1257–1267, 2010.

[103] T. Lensberg, A. Eilifsen, and T. E. McKee, "Bankruptcy theory development and classification via genetic programming," *Eur. J. Oper. Res.*, vol. 169, no. 2, pp. 677–697, 2006.

[104] J.-J. Huang, G.-H. Tzeng, and C.-S. Ong, "Two-stage genetic programming (2SGP) for the credit scoring model," *Appl. Math. Comput.*, vol. 174, no. 2, pp. 1039–1053, 2006.

[105] H. Etemadi, A. A. Anvary Rostamy, and H. F. Dehkordi, "A genetic programming model for bankruptcy prediction: Empirical evidence from Iran," *Expert Syst. Appl.*, vol. 36, no. 2 PART 2, pp. 3199–3207, 2009.

[106] D. West, "Neural network credit scoring models," vol. 27, 2000.

[107] B. Baesens, T. Van Gestel, S. Viaene, M. Stepanova, J. Suykens, and J. Vanthienen, "Benchmarking state-of-the-art classification algorithms for credit scoring," pp. 627–635, 2003.

[108] L. Yu, S. Wang, and K. K. Lai, “Credit risk assessment with a multistage neural network ensemble learning approach,” vol. 34, pp. 1434–1444, 2008.

[109] M. M. Shiri and M. Ahangary, “Corporate Bankruptcy Prediction using Data Mining Techniques: Evidence from Iran,” *African J. Sci. Res. Vol*, vol. 8, no. 1, 2012.

[110] M. Tsai, S. Lin, C. Cheng, and Y. Lin, “Expert Systems with Applications The consumer loan default predicting model – An application of DEA – DA and neural network,” *Expert Syst. Appl.*, vol. 36, no. 9, pp. 11682–11690, 2009.

[111] C. Chuang and R. Lin, “Constructing a reassigning credit scoring model,” *Expert Syst. Appl.*, vol. 36, no. 2, pp. 1685–1694, 2009.

[112] Y. Peng, G. Wang, G. Kou, and Y. Shi, “An empirical study of classification algorithm evaluation for financial risk,” *Appl. Soft Comput. J.*, vol. 11, no. 2, pp. 2906–2915, 2011.

[113] G. Wang and J. Ma, “Expert Systems with Applications A hybrid ensemble approach for enterprise credit risk assessment based on Support Vector Machine,” *Expert Syst. Appl.*, vol. 39, no. 5, pp. 5325–5331, 2012.

[114] V. Tech, “Credit-scoring models in the credit-onion environment using neural networks and genetic algorithms,” pp. 323–346, 1997.

[115] D. Mallick, “Microfinance and Moneylender Interest Rate: Evidence from Bangladesh,” *World Dev.*, vol. 40, no. 6, pp. 1181–1189, 2012.

[116] R. V. Narendar and G. Atmanathan, “ANALYSIS OF BANKRUPTCY PREDICTION MODELS AND THEIR EFFECTIVENESS: AN INDIAN PERSPECTIVE,” *Igarss 2014*, vol. 7, no. 1, pp. 1–5, 2014.

[117] M. a, V. Gayathri, and E. Al, “a Hybrid Model for Bankruptcy Prediction Using Genetic,” *Int. J. Soft Comp*, vol. 2uting (IJ, no. 1, pp. 12–24, 2011.

[118] M. S. Roomi, W. Ahmad, M. Ramzan, and M. Zia-ur-Rehman, “Bankruptcy Prediction for Non-Financial Firms of Pakistan,” *Int. J. Account. Financ. Report.*, vol. 5, no. 2, p. 26, 2015.

[119] A. Rashid and Q. Abbas, “Predicting Bankruptcy in Pakistan,” *Theor. Appl. Econ.*, vol. 18, no. 9, pp. 103–128, 2011.

[120] N. A. Manab, N. Y. Theng, and R. Md-Rus, “The Determinants of Credit Risk in Malaysia,” *Procedia - Soc. Behav. Sci.*, vol. 172, pp. 301–308, 2015.

[121] U. F. Ibn, J. K. Panford, J. Ben Hayfron-, and R. Base, “Fuzzy Logic Approach to Credit Scoring for Micro Finances in Ghana ( A Case Study of KWIQPLUS Money Lending ),” vol. 94, no. 8, 2014.

[122] R. Prakash and S. Poudel, “The impact of credit risk management on financial performance of commercial banks in Nepal,” *Int. J. Arts Commer.*, vol. 1, no. 5, pp. 9–15, 2012.

[123] Khalafalla Ahmed Mohamed Arabi, “Predicting Banks’ Failure: The Case of Banking Sector in Sudan for the Period (2002-2009),” *J. Bus. Stud. Q.*, vol. 4, no. 3, p. 160, 2013.

[124] A. B. Emel, M. Oral, A. Reisman, and R. Yolalan, “A credit scoring approach for the commercial banking sector,” *Socioecon. Plann. Sci.*, vol. 37, pp. 103–123, 2003.