MASARYK UNIVERSITY

FACULTY OF ECONOMICS AND ADMINISTRATION

Explainable Credit Risk Scoring in P2P Lending

Master's thesis

YADAV, RISHEE

Supervisor: Oleg Deev, Ph.D.

Department: Department of Finance

Programme Finance

Brno 2022

MUNI
ECON

MUNI
ECON

MASARYKOVA UNIVERZITA Faculty of Economics and Administration
Lipová 41a, 602 00 Brno
IČ: 00216224
DIČ: CZ00216224

Master's thesis description

Academic year: 2021/2022

Student:	Rishee Yadav
Programme:	Finance
Title of the thesis/dissertation:	Explainable Credit Risk Scoring in P2P Lending
Title of the thesis in English:	Explainable Credit Risk Scoring in P2P Lending
Thesis objective, procedure and methods used:	The thesis aims to develop a transparent credit scoring model in P2P lending using explainable machine learning methods. Plan: 1. Introduction. Problem statement and definitions. 2. Literature review on P2P lending, credit scoring and explai-nable machine learning (model-agnostic) methods. 3. Data collection. Data description. Methodology description. 4. Analysis of the performance and explainability of chosen machine learning methods applied to credit scoring in P2P lending. 5. Conclusions. Discussion of the results. Methodology: analysis, comparison, description, statistical modelling
Extent of graphics-related work:	According to thesis supervisor's instructions
Extent of thesis without supplements:	60 – 80 pages
Literature:	MOSCATO, Vincenzo, Antonio PICARIELLO and Giancarlo SPERLÍ. A benchmark of machine learning approaches for credit score prediction. *Expert Systems with Applications*. Elsevier, 2021, roč. 165, s. 113986. ISSN 0957-4174. ARIZA-GARZÓN, M J, J ARROYO, A CAPARRINI and M SEGOVIA-VARGAS. Explainability of a Machine Learning Granting Scoring Model in Peer-to-Peer Lending. *IEEE Access*. IEEE Xplore Digital Library, 2020, roč. 8, s. 64873-64890. ISSN 2169-3536. doi:10.1109/ACCESS.2020.2984412. HADJI MISHEVA, Branka, Ali HIRSA, Joerg OSTERRIEDER, Onkar KULKARNI and Stephen FUNG LIN. Explainable AI in Credit Risk Management. *arXiv preprint*. 2021, 2103.00949. BUSSMANN, Niklas, Paolo GIUDICI, Dimitri MARINELLI and Jochen PAPENBROCK. Explainable machine learning in credit risk management. *Computational Economics*. Springer, 2021, roč. 57, č. 1, s. 203-216. ISSN 0927-7099.

BAZARBASH, Majid. FinTech in Financial Inclusion: Machine Learning Applications in Assessing Credit Risk. *IMF Working Papers*. 2019, roč. 19, č. 109.

MOLNAR, Christoph. *Interpretable Machine Learning*. Leanpub, 2020.

DASTILE, X, T CELIK and M POTSANE. Statistical and machine learning models in credit scoring: A systematic literature survey. *Applied Soft Computing*. 2020, s. 106263.

HASTIE, Trevor, Robert TIBSHIRANI and J. H. FRIEDMAN. *The elements of statistical learning : data mining, inference, and prediction*. 2nd ed. New York, N.Y.: Springer, 2009. xxii, 745. ISBN 9780387848570.

Thesis supervisor:	Oleg Deev, Ph.D.
Thesis supervisor's department:	Department of Finance

Thesis assignment date: 2021/03/26

The deadline for the submission of Master's thesis and uploading it into IS can be found in the academic year calendar.

In Brno, date: 2022/05/12

Bibliographic record

Author: YADAV, RISHEE
Faculty of Economics and Administration
Masaryk University
Department of Finance

Title of Thesis: Explainable Credit Risk Scoring in P2P Lending

Degree Programme: Master's

Field of Study: Finance

Supervisor: Oleg Deev, Ph.D.

Year: 2022

Number of Pages: 90

Keywords: credit risk, credit risk modelling, peer to peer lending (P2P)

Abstract

Credit risk arises when a lender or client in a debt agreement fails to make the required instalment. Major advancements in credit risk intelligence have been made in the last decade to mitigate these risks. Despite that, there are biases toward these models for not being transparent enough. This thesis focuses on the proposed interpretation of models to explain them in-depth and to assess the models' performance based on comparative analyses to determine their potential and limitations. Not only can these models' black-box natures be explained, but they can also be used to make some powerful predictions.

Keywords: Credit risk modelling, credit scoring, black-box, machine learning, peer to peer lending.

Declaration

I certify that I have written the Master's Thesis Explainable Credit Risk Scoring in P2P Lending by myself under the supervision of Oleg Deev, Ph.D. and I have listed all the literature and other sources in accordance with legal regulations, Masaryk University internal regulations, and the internal procedural deeds of Masaryk University and the Faculty of Economics and Administration.

Brno, ...

YADAV, RISHEE

Acknowledgement

I would like to express my deep gratitude to my thesis supervisor Oleg Deev, Ph.D., for his continuous support throughout this thesis and his quality lectures during my master's study.
I would also like to thank my family and girlfriend for their unconditional support and encouragement during my studies.

Table of Content

List of Figures Error! Bookmark not defined.

List of Tables Error! Bookmark not defined.

List of Terms and Abbreviations 13

Introduction 17

1.1 The Aim of Present Work 18

1.2 Research Problem and Objective 19

2 Basic Concepts 20

2.1 Credit Risk Modelling 20

2.2 Peer to Peer (P2P) Lending 25

2.3 Machine Learning Credit Risk Modelling in P2P Lending 28

3 Literature Review 32

3.1 Literature Review on Credit Risk Models 32

3.2 Artificial Intelligence Based Credit Risk Models 38

3.3 Artificial Intelligence Based Credit Risk Models in Peer-To-Peer Lending 42

4 Data and Methodology 45

4.1 Data 45

4.2 Methodology 52

4.3 Performance Measurements 59

4.4 Comparative Analysis 61

5 Results and Discussion 63

5.1 Logistic Regression 63

5.2 Decision Tree 66

5.3 XGBoost 70

5.4 Performance Comparison 74

Conclusion 78

References **79**

Appendix **90**

List of Figures

Figure 1: Proposed Research Methodology 46
Figure 2: Working-year distribution of users 48
Figure 3: Customers' home ownership 49
Figure 4: Loan purposes distribution 50
Figure 5: Loan status by loan grade 51
Figure 6: Loan status for customers 52
Figure 7: Percentage of loan status 52
Figure 8: Logistic regression's confusion matrix 64
Figure 9: Impact of all features 66
Figure 10: Decision Tree of Loan default 67
Figure 11: Decision tree's confusion matrix 69
Figure 12: XGBoost classifier 71
Figure 13: Comparative analysis based on validation accuracy 76
Figure 14: Comparative analysis based on confusion matrix parameter 76

List of Tables

Table 1: Detailed description of each attribute of the dataset 47
Table 2: Accuracy, precision, F1 score and recall result of logistic regression 65
Table 3: Accuracy, precision, F1 score and recall result of decision tree 69
Table 4: Accuracy, precision, F1 score, and recall result of XGBoost classifier 74

List of Terms and Abbreviations

ML	–	Machine learning
LR	–	Logistic Regression
DT	–	Decision Tree
EAD	–	Exposure at Default
LGD	–	Loss given default
LPM	–	Liquidation Process Model
FPS	–	First Passage Model
AI	–	Artificial Intelligence
SVM	–	Support Vector Machine
CID	–	Conditionally Independent Default

Introduction

Generally, credit risk arises when a lender or a client in a debt agreement fails or delays in repaying the obligation in whole or in instalments. Hence, it is a result of debt borrowers and creditors in derivatives agreements defaulting. Today, the relevance of credit risk is widely acknowledged among academics across various disciplines, and credit risk assessments have received a lot of attention, particularly in the sector of financial services. Major advancements in credit risk intelligence have been leading toward the development of a new generation of advanced credit risk modelling techniques. Various models that have been proposed are proprietary apps produced for internal usage by cutting-edge financial organizations and third-party programmers designed for sale or distribution as software. But with the arrival of these advanced models, the situation is also going to be more critical in parallel in this domain, as new credit products have been introduced, such as peer-to-peer (P2P) lending. In P2P lending, loans are made between individuals, borrowers, and investors who are linked only via technological platforms. This scenario led to the obviation of the need for established institutions to act as intermediaries and resulted in discrepancies in information, which makes it more difficult to quantify credit risk in P2P lending than it does in conventional products. Hence, more competent and intelligible credit risk models are required for peer-to-peer (P2P) lending. For such complex problems, machine learning is capable of quantifying credit risk. In consideration of that issue, in the present work, the author has proposed a machine learning- interpretation model in order to quantify credit risk in peer-to-peer (P2P) lending. The proposed model is evaluated on a publicly available benchmark dataset.

Many researchers have defined credit risk in their own way. For example, as per Anderson's (2013) definition of credit risk, there is a likelihood that a legally binding agreement may become worthless because the opponent fails and gets busted. Saunders and Cornett (2011) put it this way: there could be a possible concern about promised cash flows through loans and bonds held by banks that might not have been paid back. Hence, credit risk arises as a result of debt borrowers and creditors in derivative agreements defaulting (Hull, 2012).

Numerous credit risk concepts, models, and explanations have been created in the past, which also brought new phenomena. Predictions of bankruptcy by Altman (1968) and Atiya (2001), distance-to-default demonstrated by Merton (1974), derivative pricing by Duffie and Huang (1996) and Jarrow and Turnbull (1995), default level of intensity developed and explained by Lando (1998) and Duffie and Lando (2001), swaps on credit defaults by Parlour and Winton (2013) and Longstaff et al. (2011), contingent claim by Jarrow, Lando, and Turnbull (1997) and Duffie and Singleton (1999), risk of a counterparty

given by Jarrow and Yu (2001) and rate of recovery explained by Altman et al. (2005) are a few examples of those theories and models.

Credit risk has long been an important part of the financial industry's history. The threat of borrower not returning money due has compelled lenders to devise ways to mitigate this risk in order to prevent losing their principal and interest. Although the principle of credit risk has stayed relatively unchanged over time, the way it is managed has changed dramatically. Over the previous few decades, technological advancements have resulted in far more effective risk management for financial institutions, allowing them to grow their businesses and increase their creditworthiness. This has been possible with the deployment of various credit risk models.

1.1 The Aim of Present Work

The present research work aims to predict loan defaults in peer-to-peer lending and conduct a performance assessment of the various credit risk models. This thesis started with a theoretical understanding of the P2P lending system, most specifically the introduction to credit risk, credit risk modelling, the introduction to P2P loans, the merit and demerit of P2P lending, and machine learning approaches toward P2P lending. Next to it, in the empirical section of the present work, machine learning algorithms have been simulated on a publicly available benchmark dataset. This work's focus is on proposed interpretation models for explaining them in depth and for their critical analysis. The critical analysis of proposed interpretation models includes tackling dataset imbalance in those kinds of situations, hyper parameter optimization for the model's configuration, evaluation of each class's classification effectiveness, and the assessment of the model's performance based on statistical comparative analyses to determine its potential and limitations. The research findings of the current work surely will make significant contributions toward filling a research gap that has not been previously identified, as well as will address the various issues in the field of credit risk management. It will also help researchers working in the field of credit risk assessment and modelling in P2P lending to understand which model to use for the best results.

In the present research work, the major contributions are:

1. A detailed study of credit risk modelling in P2P lending has been conducted.
2. The author has proposed and simulated a machine learning-interpretation model on a publicly available benchmark dataset in order to quantify credit risk in P2P lending.
3. The proposed machine learning-based interpretation models have been explained in depth for their critical analysis, which includes tackling dataset imbalance in

those kinds of situations, hyper parameter optimization for the model's configuration, and evaluation of each class's classification effectiveness.
4. An assessment of the model's performance based on statistical comparative analyses has been performed in order to determine its potential and limitations.
5. An analysis of the performance and explainability of selected credit risk detection models has been performed.
6. The credibility of the black box technique has been tested.

1.2 Research Problem and Objective

Although peer-to-peer lending is beneficial to numerous users as it provides a digital platform for money lending, with the exception that intermediaries are not always legally required, and financial institutions can use the platform to make loans to customers or small businesses. Furthermore, P2P loans also typically offer a higher return, but with a higher degree of risk in comparison to fixed-income investments. Hence, the default prediction in that case of P2P loans is a very crucial issue.

With that, the traditional methods based on statistical techniques do not give the desired results. On the other hand, results obtained by highly efficient artificial intelligence or machine learning-based models often lack transparency. Truly, the black-box nature of such a system enables powerful predictions, but they cannot be explained directly.

The research problem identified for the present work is to develop an explainable credit risk modelling in P2P lending based on machine learning approach.

For any type of research work, research objectives must be brief, concise, and expressive. The research study's objectives should be well written and self-explanatory to summaries what the study is supposed to accomplish.

The formulations of research objectives must be based on the scope of the research.

The research objectives set for the present research work are:

1. To simulate a machine learning-interpretation model on a publicly available benchmark dataset in order to quantify credit risk in peer-to-peer (P2P) lending.
2. To interpret the proposed machine learning-based models in depth for their critical analysis, which includes tackling dataset imbalance in those kinds of situations, hyper parameter optimization for the model's configuration, and evaluation of each class's classification effectiveness.
3. To make a comparative analysis of different models on a publicly available benchmark loan dataset with the goal of understanding their potential and limitations.

2 Basic Concepts

2.1 Credit Risk Modelling

Credit risk modelling is the process of using statistical or financial models to predict how much money a company will lose if a borrower defaults. To assess default risk, banks and financial institutions use these models that incorporate borrowers' credit histories, third-party data (such as credit ratings data), and information from their own economic turmoil events to formulate models. Furthermore, credit risk capital modelling is an extension to that. It uses such models to determine the minimum amount of money that should be reserved as a buffer against these damages. Banks that are allowed to employ this family of techniques must calculate two factors: a borrower's likelihood of default and the business's own loss in the event of failure. The findings aid banks in allocating loss provisions and determining regulatory capital requirements, among several other things. Credit risk modelling is the most effective approach for financiers to determine the likelihood of a loan being returned. In other words, it is a technique for determining a borrower's default risk. This is particularly significant since the credit risk profile changes over time and in response to changing conditions.

Credit risk modelling has been used for a long time. Typically, as credit risk is a significant sort of challenge in finance and it is frequently recognized as the earliest sort of risk in capital markets, dating back to 1800 BCE in the Egyptian Civilization (Caouette et al., 2008), the employment of credit risk modelling has been recognized as a possible solution for that. Until that time, the various evaluations have been done so far in credit risk modelling.

2.1.1 The Evolution of Risk Modelling

The very first contemporary quantitative credit risk model that was reported was "Altman's Z-score 1" (Benzschawel, 2012). This model is based on multivariate discriminant analysis of the five-accounting ratio. Although this model is almost five centuries old, the Z-score remains an important instrument to many market participants. However, because accounting measures are based on past data, the methodology of this model has been criticized for being backward-looking and sporadic. Additional credit risk models, such as structured and reduced-form models, have been created as a result of this.

The origins of structural credit risk models can always be attributed to Black, Fischer, and Scholes's (1972) work. Focusing on theories of capital structure (Modigliani and Miller, 1958), structural models believe that a default happens when a company's assets are

smaller than its debt. Black, Fischer, and Scholes (1973) utilize an alternative price structure to value debt and equity, proving that call options on equity can affect debt value. The challenge with Black, Fischer, and Scholes's approach has been that a firm's asset prices could be directly measured. Extensions upon their findings (Merton, 1974) demonstrated that the value of the asset can be determined under certain hypotheses which can then be used to assess the probability of default that they refer to as the "distance to default". Merton's model is currently the most dominant model in credit risk modelling, at least in terms of structural models.

Apart from structural models, reduced-form models can predict the risk of default without jumping to any conclusions about the source of credit risk premium (Benzschawel, 2012). Reduced-form models are based on risk-neutral pricing theory which states that the estimated cash flows discounted at the risk-free rate equal to the market value of a risky investment. Jarrow and Turnbull (1995) and Duffie and Singleton (1999) have applied risk-neutral pricing theories to credit risk modelling.

There are three main elements of credit risk: default risk, spread risk, and downgrade risk (Bielecki and Rutkowski, 2004). Default risk refers to the risk that the provider or counterparty may fail to fulfil the conditions of a financial contract's commitment. On the other hand, does credit spread risk refer to the failure or underachievement of an issue or problem caused by an increase in the credit spread? The credit spread illustrates how financial institutions respond to an issue's projected credit quality worsening. The possibility of credit ratings worsening is known as downgrade risk. When a credit agency assigns a poorer rating than the preceding grade, a borrower is at danger of being downgraded. The three categories of credit risk are inextricably linked.

When a firm wants debt funding, perhaps through a bond, a document is made accessible to investors so that they can learn enough about the business before purchasing the bonds. Investors typically do a credit investigation of the issuer prior to investing money, which involves name recognition and credit scores. Credit scores are official views expressed by a rating agency or departments relating the default risk of a particular debt issue. Name recognition is about the issuer's public image, whereas credit scores are official views expressed by ratings agencies or agencies regarding the default risk associated with a particular debt issue.

2.1.2 The Objective of Credit Risk Modelling

Credit risk modelling involves two basic objectives. The first goal is to examine all and every counterparty and transaction in a banking institution or individual's portfolio. It proceeds to a comprehensive review of the counterparty's creditworthiness as well as the transaction's framework. The second objective is to conduct a portfolio analysis of the

whole range of transactions and creditors. Any traditional loan items, along with all investment bank transactions and exchanged goods, should be included. The most important out of the two objectives is to determine whether the portfolio is in line with the bank's appetite for risk or not. The appraisal of actual transactions inside the context of the current portfolio and the determination about whether these fit into the bank's risk level is a crucial component of portfolio optimization.

2.1.3 The Most Basic Components of Credit Risk Modelling

A very important aspect of credit risk modelling is the examination of a single transaction and the theory of joint default. The various input and output parameters for that are explained below.

The exposure and the loss given default are two input components. The exposure is known for an outstanding amount that will be forfeited in the event of default (Engelmann and Rauhmeier, 2011). As a result, it is known as the Exposure-at-Default (EAD). While we evaluate the defaults that have accrued at time 1, we can assume that the quantity outstanding at time 1 is taken into account for the sake of clarity. Another perspective is to assume that defaults occur at moment 1. While the EAD is a constant, non-random quantity among most deployed models, it should be treated as a random vector in reality.

On the other hand, the loss given default (LGD) is the percent of the EAD that is lost following the work-out and recuperation phase (Johnston Ross and Shibut, 2022). Once more, it should be modelled as a random variable. To account for the LGD's unpredictability, beta distributions can sometimes be utilized. If the LGD is believed to be independent, the influence of stochastic LGD will have a minor impact on the investment measures in large portfolios because the law of higher numbers will apply. If someone wants to describe a randomized LGD, they should make the effort to include a relationship among LGDs as well as the other variables involved. According to empirical evidence, default rates and losses given a probability of default are positively connected.

Furthermore, the possibility of a counterparty default is measured by the default probability. A timeframe, such as one year, must be specified. The default probability, on the other hand, assesses the likelihood that the counterparties will default one year from now. This specification is the accepted standard by the industry, and it is what is used in the risk weighting functions.

The economic capital and the value-at-risk given are two output components. The credit risk models' collected output is dependent on the loss factor L, which is the total of all single transaction losses. The projected losses are not seen as a risk metric by most banking firms; instead, they are regarded as a cost which should be compensated by margin revenue rather than capital.

Economic capital (EC) is the level of investment needed as a buffer to survive even very significant losses, such as those produced by financial events (Engelmann and Rauhmeier, 2011). Similar to Regulatory Capital, EC is a mandated amount of cash. Severe losses imply that even higher losses are only possible at a very small chance. As a result, probability, referred to as a level of confidence in this context, is required as a parameter in Value-at-Risk. If the 99 percent Value-at-Risk turns out to be 10 million euro, the chances of accumulating portfolio losses exceeding 10 million euro are 1 percent. Other than this, expected shortfall, coherent risk measures, capital allocation, and contribution to expected shortfall are also important output components of credit risk modelling.

2.1.4 Types of Credit Risk Models

Credit risk assessment models can be divided into two categories: qualitative and quantitative models (Saunders and Cornett, 2011).

When assessing credit risk, qualitative elements such as borrower characteristics (e.g., reputation, leverage ratio, cash flow volatility, security, etc.) and market factors are taken into account (business, interest rates, etc.). Relying on such variables, a subjective decision has to be made – a decision whether to provide credit to a client or not. A quantitative model tries to generate a credit rating, which can be used to assess the likelihood of default, or to categorize debtors into different default risk classes.

Modelling credit risk quantitatively is however difficult due to the rarity of risk of default, which is an element of creditworthiness. Because of the idea of credit risk, a variety of quantitative credit risk models was proposed. These are generally built on theoretical foundations such as proportions, choice theory, econometrics, or intelligent agents, much like any other model. Frequently used instruments for financial models are econometrics, simulation, optimization, or a combination of these models. Credit risk models can be classified into three groups based on the methodology utilized, the application domain, and the assets involved.

On the other hand, various quantitative approaches that have been created are econometric approaches, neural networks, expert systems, hybrid models, structural and reduced-form models, first generation structural-form models, and reduced-form models.

Econometric approaches are statistical models based on dependent variables known as default probability or default premium. Linear probability, logit, probit model, discriminate analysis, and regression analysis are examples of a few of those kinds of modelling techniques (Saunders and Cornett., 2011). Both of the linear probability and logit models use historical data, including such financial ratios, to quantify the likelihood of default.

The distinction is that the latter relies on logistic regression, whereas the aforementioned is based on regression analysis. The discriminate function model utilizes historic

data as well, but it divides borrowers into high and low risk default groups. In general, discriminate and logistic models outperform the others (Altman and Saunders., 1997).

Neural networks are computer-based systems' networks. These types of networks use data over economic models for prediction, and they work based on a trial-and-error approach. Optimization models are purely statistical programming models that have been used to find the best weights for borrowers, reducing the lender's errors and increasing their revenues. These models are designed around a set of various decision rules, or "trees of decision," which are frequently developed using data from both the market and the personal scale.

On the other hand, hybrid models work on combination of both structural models and some other financial variables such as asset's book value, debts, net earnings, and return on capital etc.

Structural and reduced-form models are the most common quantitative credit risk models. These two models are current models and concerned with designing the default time and recovery rates. The invented structural models by Black, Fischer and Scholes (1973) and Merton (1974) stalled price organization lending risk by concentrating on the company's capital value. This leads to a new model known as firm-value models.

Among structural models, default is patterned as a choice, and investigators can price risky corporate liabilities using option pricing theory. The application of alternative pricing theory to default risk has greatly improved the conventional methods of charging companies for risky bonds. These structural models frequently focus solely on the default component of credit events. Default time is inherent in the sense that it is determined by the value of the firm trying to reach a certain default point. Whenever a bondholder defaults, other bondholders consider taking over the firm to defend their interests.

Another key tenet of the structural framework is that stockholders have control over the bankruptcy ruling. Stockholders can bring fresh capital into the company to enhance its leverage position, preventing it from going bankrupt. If they are unwilling to do so, debt holders will be able to take over the business in the case of default. Healing rates in structural models are also described endogenously inside the model.

Reduced-form models are a newer type of credit risk model (Anson et al., 2000). Reduced-form models, formed by Jarrow and Turnbull (1995) and Dougie and Singleton (1999), look outside of the business, as opposed to structural models. In specific, the company's value of the asset and working capital are not considered, and default time and healing rates are characterized exogenously inside the model. The models frequently concentrate on either the risk of default or the downgrade risk. As a result, reduced-form models can be divided into two types: intensity-based models and credit migration models proposed by Bielecki and Rutkowski (2004).

Major advances in credit risk modelling as a result of continued technological advancements have also resulted in the leading development of new credit products such as

peer-to-peer (P2P) lending. The following section of this chapter will discuss peer-to-peer (P2P) lending in this new age of society and its impressions over conventional credit methods.

2.2 Peer to Peer (P2P) Lending

2.2.1 Definition of P2P

The term "peer-to-peer" implies that the individuals are from both ends of the lending process, and the word "peer" is specifically defined as an individual having the same status. Somewhere, "peer to peer" also refers to the interactions between two parties that do not require the use of a central intermediary. Basically, this term refers to a network in which any machine or a computer can act as a client or a server to other machines or computers on the same network without the need to link to a central server. The internet is actually a peer-to-peer system in and of itself.

Further, peer-to-peer (P2P) lending is a recent approach to financing that was pioneered in 2005 by the British system Zopa. P2P lending has demonstrated that it can stay competitive with financial or banking institutions by offering lower-cost products over time. The proven economic success of P2P lending has resulted in a large number of economic studies focused on the same.

A large part of the research work looks at peer-to-peer lending from the point of view of the debtors or borrowers. For example, Livingston (2012) explored the feasibility of replacing payday loans with peer-to-peer lending. The author in this work also discovered that several customers might well gain from substitution, whereas others discover the qualities of P2P loans like financing speed, availability, requisite credit rating, etc. From the perspective of the lenders, Roure et al. (2016) used data from Auxmoney and banking by the Deutsche Bundesbank to investigate P2P lending and see if it is just a replacement or also a support to bank lending. The author concluded that, while rates of interest are nearly equivalent even before risk is monitored for, lending directed via P2P platforms is riskier when at a higher rate of interest than bank loans.

A further section concentrates on the investors' point of view. Several publications investigate the impact of soft character traits on borrower credibility. Herzenstein et al. (2011) are among those investigators. They discovered proof of corporate strategy herding behaviour and also proof of herding becoming beneficial to investors. Similarly, Duarte et al. (2012) used prosper data to assess the debtor's truthfulness predicated on one's picture. Lin et al. (2013) discovered that investors fairly use social connections as both a signal of creditworthiness and a source of research in this work. They used the same data as Prosper.

Even though many researchers intuitively know what the word means, peer-to-peer lending and its structures, there really is no widely accepted definition that encompasses all of its applications. From the point of view of financial intuition, a general definition of "peer to peer" is defined as "the lending system among private entities on digital platforms, with financial companies acting only as intermediaries" according to Bachmann et al. (2011), with the exception that intermediaries are not always legally required.

The peer-to-peer financial services organization, which was founded in 2011 as a spokesperson and self-regulation body for peer-to-peer lending in the U.K., describes P2P lending as follows: "Peer-to-Peer lending allows investors to loan to people directly through a digital platform." Financial institutions can use a platform to make loans to customers or small businesses. While platforms enable lending by performing credit checks as well as other risk assessment functions, they do not really act as a refutation to the line of credit, and agreements are made directly between both the investor and the lender.

It may appear that the interpretations of P2P are not considerably different. The aspect is that both the investor and the lender must be individual people. If the investor were a corporation, the essence would be essentially similar to that of traditional lenders. However, if the lender is a corporation, there really is no distinction between this and debt security.

P2P lending is a multifaceted concept. According to Kirby and Worner (2014), the business plans of P2P lending platforms vary based on the regulatory structure as well as the system's precepts. This might include the process to determine lending rates, the fees charged, the method of verifying clients' credibility, and the volume of creditor information given to investment firms.

According to Herrero-Lopez (2009), peer-to-peer lending is a concept that combines the ancient concept of personal credit with the world wide web. In other words, the online system of the P2P lending framework is novel (Chaffee & Rapp, 2012). This connection then considerably reduces costs as well as the overall cost of locating a suitable counterpart for the loan process.

A peer-to-peer lending system is a type of financial services provider platform. Other types of crowd funding include donation systems, reward platforms, and equity crowdfunding platform.

A P2P lending platform has been seriously influenced by the laws of the region in which individuals need to provide their own services, as well as domestic regulation. In other words, a financial model that works incredibly well in one region may be illegal in another, rendering it unusable. Kirby and Worner (2014) identify five concepts of P2P lending regulation in a situational setting. Most are either excluded or uncontrolled, leading to a shortage of descriptions. As an interpreter, it is controlled, financial services regulations apply, and it is a prohibited model. Platforms may use theories and models within

a given legal system. Three of those kinds of models are recognised by Kirby and Worner (2014).

The platform serves as a true mediator in the client-segregated account model (Fong, 2015). The lending in the notary model is not really made immediately available to the debtors by the investors. The system is linked to a financial institution, which would be the loan's provider. The bank will sell the loan to the channel, and the platform will issue a memo to the investor able to represent the made contribution value, with the refund contingent on the repayment of the loan amount. As a result, the investment is made in a note rather than in the loan itself. This model is widely used in the United States (Chaffee & Rapp, 2012) and China (Fong, 2015).

The guaranteed return model allows lending institutions to invest through the framework at an assured return on capital. The framework must either provide such a guaranteed return or cease operations. This model was mostly introduced in China (Kirby & Worner, 2014). The distinctions in legal models are significant because they influence the actual regulatory oversight influencing the specific system. In other words, as platform models differ, so do the legal requirements imposed on them. It further means that two P2P lending platforms operating in the same regulatory environment could be subjected to different regulation.

The platform may also differ in an almost unlimited number of statistical variables, like the necessary level of borrower credibility, the creditor details available to investors, the minimum quantity of funds decided to invest, other assistance offered by the platform, like insurance, or the level of charges levied by the platform for the various services. While platforms obtain fees premised on loan repayment, the platform's interest is most often allied with the borrower's interest, which is often correct.

P2P loans are really a kind of income investment that is fixed as it provides contractually binding payments. P2P loans typically offer a higher return at the cost of a higher degree of risk in comparison to fixed income investments. The cause of this is that the government is providing no assurance of return of payment. The distinction with other fixed income investments is primarily determined by particular attributes of the issued loans. The cash flow of P2P loans is heavily dependent on the P2P platform employed to release the loan as well as its rules. As a result, from a scientific perspective, P2P loans might be recognized as having features comparable to corporate bonds.

If P2P loans have similar attributes to corporate bonds, other investment opportunities must provide better returns at the expense of higher risk. As a result, P2P loans are presumed to be used in place of corporate bonds.

2.2.2 Significance of Peer-to-Peer Lending

The various benefits of peer-to-peer lending includes:

- Credit continues to flow, thereby assisting growth in the economy.
- P2P lending bridges a void left by banks.
- Enhanced cost effectiveness as a result of use of technology solutions.
- Because of the lower costs, it does provide a lower cost of capital and better returns.
- Diversification of investments.
- Investors enjoy the convenience.
- Growing competitive advantages debtors, lending institutions, and the overall economy.

Some risks are always associated with these various benefits of using peer-to-peer lending over conventional credit risk products. For example:

- Default danger.
- Risk associated with the platform – the possibility of the platform being completely or temporarily closed.
- The possibility of fraud.
- Performance and information discrepancies.
- Investor lack of experience poses a risk.
- The risk of liquidation.
- Cyber-attack danger etc.

However, peer-to-peer lending has numerous advantages, and it also results in greater progression in economic fields. Still, the necessity of affordable and effective credit risk models is the greatest challenge in that area.

Due to the advances in technology and the existence of various machine learning algorithms, it has become possible now to find an effective solution to the above-mentioned research problem. Machine learning is a subset of artificial intelligence that is gaining attention in various domains of knowledge already dealing with real-life problems. Similarly, machine learning is experiencing an increasing impact in business applications, with several solutions already in place and many other being investigated. Quintessential machine learning algorithms are providing excellent prediction performance in the area of credit risk because they are capable of detecting more complicated patterns in datasets. The next section of this chapter will discuss the machine learning approaches toward credit risk management.

2.3 Machine Learning Credit Risk Modelling in P2P Lending

For a long time, an accurate estimation of the risk of default in relation to lending was a critical subject for banks and all kinds of other small and large financial institutions. Nowadays, as huge datasets and accessible data sources are available and remarkable developments can be seen in computational and pattern recognition analytics, as a result of this, a

great progression has been seen in rekindled interest in risk prediction. In light of this, automation of the lending process has created new business opportunities for small enterprise owners and individuals. Previously, all of them faced restricted credit opportunities due to the high cost of human involvement in this process. Eventually, this automated lending process has the capability to minimize human bias and corrupt practices, making credit more equitable for all. Advances in financial technologies had been making significant impact on this quickly changing sector. The developed model introduced in the present research work is not limited to P2P lending, it can be also linked directly to microfinance in developing nation.

P2P lending is basically the practice of lending to people or companies via a digital platform that connects borrowers and lenders. Such platforms frequently strive to have fewer nodes than conventional financial institutions. This makes it possible for P2P lending to produce a higher rate of return to the investors than conventional saving methods, even if some fees have to be paid for that.

The present work focuses on helping researchers working in the field of credit risk assessment in P2P lending in order to gain an understanding of the benefits and drawbacks of automated lending technologies. Scientific articles examining the drivers of P2P lending (Möllenkamp, 2017; Emekter. et al., 2015) have only used simple machine learning models such as regression to achieve their goals and objectives. But the present work represents a significant advancement in the application of big data and machine learning techniques to peer-to-peer lending, incorporating two major disruptive fields.

2.3.1 Machine Learning Methods in P2P Lending

As described by Leo et al. (2019) and Serrano and Gutiérrez (2016), machine learning models that are widely used for credit risk analysis are: logistic regression, probit regression, discriminant analysis, etc. All such models perform well, are simple and easy to understand, and do not pose any computational challenges. In addition to that, machine learning solutions often provide better prediction accuracy as they can classify more complex risk behaviours. But almost all of these machine learning models are black boxes with little or no interpretability capability.

Numerous researchers have employed these machine learning models in their research work on P2P lending. Aksakalli (2015) for example compares the results of random forests with those of k-NN, SVM, and logistic regression models for evaluating risk assessment in social lending. Deep neural network-based models were used by researchers such as Zhang et al. (2017) and Yuan et al. (2017) to assess credit risk. In their research work, they also compared the deep neural networks' performance with other conventional methods of credit risk assessment like logistic regression, linear discriminant analysis, decision trees, and support vector machines. Some of the research work also reported survival

modelling techniques with machine learning models like Wang et al. (2018) and Jiang et al. (2019).

Furthermore, authors such as Ferreira et al. (2017) and Namvar et al. (2018) demonstrated how to use machine learning to resolve significant predictor's attributes such as the issue of imbalanced data, hyper-parameter optimization, meaningful feature extraction, or optimized bias replication in prediction.

In order to predict the likelihood of default, Jin and Zhu (2015) proposed SVM, decision trees and artificial neural networks. Furthermore, Li et al. (2018) employed ensemble algorithms like XGBoost, deep neural networks, and logistic regression to compare them with each other individually or in a group. The result of the reported work depicted the highest accuracy prediction technique which was the outcome of their categorizing one's proposals based on interpretability, and it has also been shown that the three methods of ensemble are the poorest at interpretation.

Through this study, it has been noticed that almost all of these machine learning models are black boxes with little or no interpretability capability. As a consequence, financial institutions are treading carefully when it comes to implementing machine learning for credit risk modelling. Besides that, regulatory enterprises' necessities are typically influenced by traditional models, which do not relate to the difficulties involved with all these alternative approaches. The use of machine learning models necessitates complementary validation components, as well as the study and identification of new biases or interpretability, among many other things, even when they are not explicitly specified. The various models of credit assessment mechanisms, including agreeing to give, behaviour and collection procedures or fraud prevention, require interpretability and transparency. They must ensure that borrowed products are delivered in an acceptable, controlled, and supervised manner. Enterprises and enterprise customers necessitate interpretable models, so models such as logistic regression, decision tree etc. are commonly used.

Numerous theoretical and empirical studies have already proposed interpretability and explainability as crucial components of machine learning techniques, as explained by Kim et al. (2016), Doshi and Kim (2017), and Miller (2019). From both an ontological and epistemological standpoint, Carvalho et al. (2019) examine interpretability. Through inspecting scholarly activities devoted solely to it, the authors in this work demonstrate the significant attention given to it. Similarly, Lipton (2018) considers the significance and complexities of the interpretation aspect in modelling, emphasizing the requirement and need for predictive and interpretable models, and as a result, refines the theory of interpretability in machine learning techniques. Molnar (2020) describes and develops a few components of interpretability and explainability, specifically how new techniques can be used for the above purpose. The author represents three characteristics to consider in interpretable modelling techniques: linearity, monotonicity, and interaction. Linearity is defined by the linear relationship among a variable and the reference variable or targeted

variable. Monotonicity denotes the correlation among a specified input and the intended results and often follows the same path across the entire domain of characteristics and interaction to anticipate the objective variable by naturally including relations among features. Several techniques can be employed to evaluate the aforementioned aspects as well as others.

The most comprehensive approach is provided by Ribeiro et al. (2016), known as the LIME method, which uses local surrogate models in order to explain each and every prediction made by machine learning models. This model seeks to comprehend why the machine learning model made a specific prediction based on the local linearity assumption. The LIME methodology employs data sample permutations across local linear regression approximations to examine the output impact.

Lundberg and Lee (2017) proposed another approach, known as SHAP values, to provide interpretation capability to complex models of various types. The SHAP values incorporate suggestions from the LIME methodology as well as the Shapley values that also describe the predictive value, supposing that every characteristic in a model is a game player. SHAP values also provide a good alternative to the computational issues posed by explanation models, taking into consideration all potential orderings of variables in the assessment of feature importance. Furthermore, they enable one to acquire a global interpretation based on Shapley values.

3 Literature Review

This chapter contains a detailed literature review on credit risk models and explainable artificial intelligence methods. It presents a detailed account of previous work that has been done for the purpose of designing and developing credit risk models and reported in the form of theses, dissertations, books, and articles. With that, a study has been conducted on various artificial intelligence methods proposed previously for credit risk modelling. The author has also conducted a literature review of the various studies that have been published previously on the use of these credit risk models in peer-to-peer lending. This chapter has also provided additional information regarding the various statistical techniques that were adopted by various researchers in order to quantify the credit risk of various financial institutions for money lending as well as in the peer-to-peer lending process.

The chapter is divided into three sections. The very first section is based on detailed studies reported in the past for designing and developing credit risk models. The second section is a survey of the use of various applicable artificial intelligence methods in credit risk models. And finally, the third section is devoted to the literature review of the deployment of these artificial intelligence-based credit risk models in peer-to-peer lending.

3.1 Literature Review on Credit Risk Models

Since the last decade, many of the world's top banks have attempted to develop efficient mechanisms for estimating credit risk arising from key components of their respective business lines. These models are designed to help banks measure, aggregate, and manage risk across geographies and lines of business. The resultant output from these models greatly contributes to the organizations' risk management and performance assessment operations like remuneration based on performance, client financial accounting, likelihood pricing, portfolio construction planning, and capital structure decisions. The working group acknowledges that credit risk modelling may certainly result in improved corporate risk management but can also create a set of capital criteria that is consistent across all institutions. Nevertheless, before such a portfolio modelling approach is used in the official charge of establishing capital requirements for credit risk, regulatory authorities must be self-assured that the designs are not only to be used to effectively manage risk, but that they are also abstractly sound, tested empirically, and create a set of capital criteria that is consistent across all institutions. At present, considerable obstacles must be overcome, namely in the areas of data accessibility and validation of the model, before all these goals

can be achieved, and the commission sees challenges in overcoming these obstacles within the timeframe set for modifying the capital accord.

In order to fully meet the above-mentioned criteria, various credit risk models have been proposed previously. Structural and reduced-form models are the two main types of models in the literature that aim to characterize default procedures for debt obligations and other default-able financial products. A detailed review of these models is presented in the proceeding section.

3.1.1 Structural Model

Structural models examine the evolution of structural variables such as asset and debt values in order to predict the timing of default. Merton's model was the very first structural model and also the first contemporary model of defaults (Merton, 1994). Hull et al. (2005) gives a detailed analysis of Merton's model, credit risk and volatility skews. In Merton's model, a company defaults if its assets are less than its unsecured balance at the moment of debt servicing. Next to it, Black and Cox (1976) presented a second strategy inside the structural framework. As per this technique, defaults commence as soon as the firm's asset value drops below a specified threshold. It is unlike Merton's strategy. Here, default can happen at any moment.

Reduced form models need not explicitly explore the relationship between default and company value. The most complicated reduced form model is the intensity model. Unlike structural models, the default time in intensity models is decided by the first leap of an exogenously imposed jump procedure, rather than the company's value. Market data is used to infer the parameters that influence the default level of risk.

Structural default models find the correlation between such a company's credit rating and its financial and economic circumstances. As a result, rather than being given exogenously as in the reduced approach, defaults are produced intracellular inside the model.

One other difference between these two techniques is how recovery rates are treated: while reduced models describe recovery rates exogenously, recovery rates in structural models are determined by the value of the firm's assets and obligations at default.

The structural credit risk model on default risk begins with Merton's (1974) work of study which was based on option pricing theory by Black, Fischer, and Scholes (1993) for modelling of a company's total. In Merton's model, the working capital of the business is supposed to be made up of stock and a zero-coupon bond with a maturity of T and a face value of D.

The company's equity is merely a European call option on the capital value with a T (a maturity) and a D (strike price), and the debt value is basically the asset value less the equity value. The whole design assumes a very basic and unrealistic capital base, implying that the zero-coupon bond can only default at maturity.

Further, the next model known as First Passage Models is basically the work performed by Black, Fischer, and Scholes (1993). The default threshold operates as a safeguard promise to secure bondholders when this is exogenously set, as proved by Black and Cox (1976) and mentioned in Longstaff and Schwartz (1995). It also can be endogenously set as a consequence of creditors attempting to determine the default threshold that maximizes the shareholders' value (Leland & Toft, 1996).

Interest rates have been treated as both no stochastic (Geske, 1977) and stochastic (Ronn & Verma, 1986; Nielsen et al., 1993; Briys & Varenne, 1997) phenomena in structural models.

Default occurs in First Passage Models when the asset value falls below a specific lower barrier for the first time, i.e., the company is liquidated instantly after the exception. In juxtaposition to First Passage Models, a set of new modelling techniques have been proposed, backed by the latest theories and empirical studies, wherein a default event does not really immediately result in liquidation but rather reflects the start of a process, the liquidation process, which may or may not result in liquidation after it is finished. Liquidation Process Models (LPM) is the name given to these models.

To comprehend LPM and its distinctions from FPM, the distinction between the terms "default event" and "liquidation" must be understood. A default event occurs whenever the company's asset value falls below the lower limit K (which is either exogenous, fixed, dependent on time, stochastic, or intracellular derived). A default event marks the start of a financially unstable time that may or may not end in liquidation. Liquidation occurs when a company is genuinely liquidated, its operations are halted, and the residual assets are split among its claimholders.

In the above-mentioned FPM, the default event happens at the same time as liquidation. Nevertheless, as Couderc and Renault (2005) note, most liquidations "do not happen spontaneously, but rather represent the culmination of a long-term process." As Moraux (2004) points out, empirical investigations in the United States have discovered that additional survival spans beyond the original default event can continue up to three years (Betker, 1995; Hotchkiss et al., 1995).

According to Helwege (1999), the prolonged default in the current US junk financial markets is 7 years. LPM extends FPM to deal with the fact that the liquidation time occurs even after the incidence of a default event (often a really long time after).

Various authors like Frangois and Morellec (2004), Moraux (2002), and Galai, Raviv, and Wiener (2005) suggest LPM based on theoretical framework.

Out of all these theories, Galai, Raviv, and Wiener (2005) took a major step ahead in the refining of LPM by providing a model that incorporates and extends the two preceding models. The author argues that, in comparison to other two theories, the single factor that determines whether or not a company is liquidated is the duration of time it has been in financial difficulties (successively or cumulatively), but they do not succeed in capture the

two most typical characteristics of bankruptcy operations: 1) recent crisis events may have a higher impact on a firm's choice to liquidate its assets than previous distress occurrences, and 2) the severe distress occurrences may well have a higher impact on a company's choice to close down than moderate distress occurrences. The author gives a structural model where a company is liquidated whenever a state parameter indicating the total weighted time spent in anguish by the business exceeds.

Galai, Raviv, and Wiener's (2004) model is a special example of Merton (1974), Black and Cox (1976), Leiand (1994), Fan and Sundaresan (2001), Frangois and Morellec (2004), and Moraux (2004). As a result, it now depicts the much broader LPM. To evaluate debt and equity, the researchers used Monte Carlo simulation, relying on Parisian options and Parisian contract approaches to resolve the model statistically.

State Dependent Models (SDM) and Liquidation Process Models (LPM) are two recent attempts to combine distinct affecting factors into structural models. Though they seem to make sense in a way, there is no empirical study to back them up.

SDM assumes that most of the characteristics affecting a company's power to develop working capital or its cost of capital are dependent on the state, with states representing the business cycle (crisis vs. growth) or the company's external credit rating.

SDM are structural models that are capable enough to eliminate the issue of predictability of defaults faced by general models because the organization is subject to exogenous alters in aspects that influence its capacity to generate working capital or its funding costs, which are the primary drivers of default likelihoods.

Hackbarth et al. (2004) and Elizalde (2005) proposed two frameworks that illustrated the preceding concepts. In both cases, the researchers offered closed form expressions for the value of equity and debt, the solutions to which require the solution of systems of ordinary differential. Cash flows and recovery rates, according to Hackbarth et al. (2004) are affected by the state of the business cycle.

On the other hand, Elizalde (2005) evolves a structural model that, while initially conceived for banks, could be implemented to any company. In comparison to prior models, the asset value of a firm is supposed to be unobservable by debt holders. Debt holders rely on financial institutions' ratings to determine the debt's coupon as a role of those rating system. As a result, the companies' funding costs are determined by their rating system.

Rating agencies' context of rapid audits of companies on a regular basis to identify one's risk and asset levels also evaluate the rating. Shifting from one rank to the other suggests changes associated with the cost of debt and, as a result, the company's creditworthiness.

State Dependent Models (SDIM), such as LPM, have been established theoretically, and their own empirical utilization and ability to reflect and anticipate equity markets and default likelihoods will determine their success in the future in credit risk modelling.

Other than this, estimation and calibration (Duan, 1994), cyclical default correlation, contagion default correlation (Giesecke, 2004), and factor models (Schonbucher, 2000), are other types of structural models.

3.1.2 Reduced-Form Model

This section examines reduced-form models for credit risk and the three major methods for incorporating credit risk correlation between many companies inside that framework of reduced models. In the literary works of intensity credit risk modelling, there are three different methods for modelling default correlation.

The very first method is conditionally independent defaults (CID) models. It incorporates credit risk reliance of default intensity mechanisms in companies on a basis of mutual state variables. The CID approach is extended by contagion models to account for default cluster formation. Eventually, copula functions can then be used to account for default dependencies. The copula strategy adopts the relatively modest probability of default of the different enterprises and put them to copula function, which gives the design with default structural relationship.

This process utilizes correlation into the firms' default levels of intensity, attempting to make them reliant on a set of common variables say Xt as well as a on firm-specific factor.

All of these kinds of models are known as conditionally independent defaults (CID) models as conditioned on the realization of the state variables Xt, the company's default intensities and default times are independent. The major downside of such models appears to be that they really do not create reasonably high default correlations. Moreover, Yu (2002) claims that it is not a concern with the model itself, but rather a result of a lack of sophistication in the selection of the system parameters.

Basically, two extensions of the CID method attempt to increase the models' default correlation. One of them is the possibility of combined jumps in the default levels of intensity (Duffle & Singleton, 1999), and the second is the default-event triggers that end up causing combined defaults (Duffle & Singleton, 1999; Kijima, 2000; Kijima & Muromachi, 2000).

Duffee and Singleton (1999), Zhang (2003), Driessen (2005), and Elizalde (2006) are some of those who suggest how to evaluate various CID models.

Another models are Contagion models that took CID models a step forward by incorporating two empirical facts into to the concept: that the default with one company can cause the default of many other companies, and also that default times tend to cluster under certain spans of time, increasing the probability of default of all companies.

The previous model is different from contagion mechanisms in that when an obligor does not really default, the intensity of the default of any linked obligor does not really change. There are two approaches to default contagion in literary works.

The infectious default model was developed by Davis and Lo (2001) and the propensity framework was developed by Jarrow and Yu (2001). The primary problems to be settled with these two models are related to challenges in calibrating them to market values.

The Davis and Lo model comes in two forms: a static variant, which only takes into account the number of defaults in a given period of time, and a dynamic variant, which also takes into account the timeframe of defaults. Jarrow and Yu (2001) broaden CID concepts to account for counterparty risk, i.e., the risk that a company's default may raise the probability of default of other companies with which it has connections through business interactions, in order for the model to correct for the cluster formation of default in particular periods. The above said enables them to account for default cluster formation in CID models by introducing extra-default reliance. In their first attempt, Jarrow and Yu believe that a company's default intensity is symmetrically dependent on the condition of the other companies. But even so, symmetric relying exposes circularity in the concept, which they relate to as "repeating defaults." It also makes constructing and deriving the joint probability distribution of default times extremely difficult and time-consuming.

This concept of designing the model presents a new origin of default correlation between secondary firms, as well as among the primary and secondary firms, but was unable to address the downsides of low correlation between primary firms that CID models appear to suggest, since the primary firm configuration is, after all, only a CID model.

The next approach that was introduced is the Copula approach. Copulas have been invented in 1959 and have now been widely used for model survival data in fields such as actuarial science.

The description of independent levels of intensity in CID and contagion model involves all the default dependence structure among businesses. The copula approach, on the other hand, differentiates individual probability of default from the credit risk dependence structure. The copula function is capable enough to take negligible probabilities as input and applies the dependence structure to generate mutual probabilities.

Because of the technological advancements and the emergence of various artificial intelligence and machine learning techniques, it is now possible to discover an efficient solution to the aforementioned issues faced in credit risk modelling. Machine learning is a subfield of artificial intelligence that has been gaining traction in a variety of cognitive domains and is already dealing with real-world problems. Likewise, artificial intelligence and machine learning are experiencing an increasing impact in business applications, with several solutions already in place and many more under consideration. Because they are capable of detecting more complex patterns in datasets, these methods provide excellent

prediction performance in the area of credit risk. The following section of this article will go over various applicable artificial intelligence methods in credit risk models.

3.2 Artificial Intelligence Based Credit Risk Models

Nowadays, financial credit risk evaluation is gaining popularity. Correct accounting of credit risk evaluation and business failure supposition are critical in both economics and society. As a result, in recent years, an increasing number of techniques and algorithms have been proposed. From this point forward, it is crucially necessary to review the current methods used to analyse financial credit risk. For this, in this section, the author tried to review and summarize cutting-edge intelligent methods based on artificial intelligence (AI) for predicting financial distress, with an emphasis on new accomplishments as a promising trend in this field.

3.2.1 General Way of Working of AI Based Credit Risk Models

Financial credit risk denotes the risk of lending the money, i.e., a borrower who is unable to pay the lenders or who will go into default. As per usual, financial credit risk evaluation seeks to address the following issue: Consider a dataset of financial variables that represent a firm's condition over just a given time period and a number of businesses labelled as wealthy and with poor credit or bankrupt and healthy. Determine the likelihood that the firm will fall into a high-risk category or go bankrupt in the coming years. The former is known as credit rating or scoring, while the latter is known as bankruptcy (failure) prediction or corporate financially distressed prediction. These issues are fixed in the same way by using the classification model. In this work, the two types of issues are considered financial credit risk assessment and evaluation. (Mantelaers, 2022)

For financial credit risk assessment, the executive route is commonly the same. The records or data under consideration are derived from company financial variables. Following data pre-processing (filtering, filling of incomplete data or missing values, normalization, and so on), attribute selection techniques are used to reduce dimensionality via transformation techniques that can be either linear or non-linear (Gao and Zhang, 2018). Following that, statistical, intelligent, or machine learning algorithms are used to investigate the correlations in the data sets. The domain-specific information is incorporated into the modelling algorithms to help with data exploration and comprehension. Users' cost preferences, the class allocation of examined data, the type of factors, the class label of findings,

and some essential private information are all included in the details. Eventually, the prediction results are analysed in terms of performance measures so that different classifiers can be compared.

Various studies have found convincing proof that intelligent techniques can significantly increase the precision of statistical techniques without relying on strict assumptions. Machine learning and deep learning methods like neural networks, fuzzy set, support vector machines, decision trees, rough set theory learning and ensemble learning are some of the approaches used for that. It is also evidenced through these theories that no method consistently outperforms all of the others across different data sets (Piramuthu, 1999).

The following subsection will discuss the latest work performed for credit risk assessment based on the above-mentioned intelligent techniques.

3.2.2 Artificial Neural Networks for Credit Risk Assessment

Artificial neural networks refer to a group of neural networks with different types of configurations. Some of the most popular types of artificial neural networks are multi-layer perceptions or neural networks with back-propagation, self-organizing maps, and learning, as well as radial basis function networks and probabilistic neural networks. As compared to other statistical techniques, all of these methods demonstrated impressive performance in the application of bankruptcy prediction or credit risk assessment. These methods' superior ability to handle both linear as well as nonlinear data is advantageous in identifying intrinsic patterns in complex financial data (Charalambous et al., 2000; Fuyuan, 2008).

Huang et al. (2004) applied the support vector machine and back-propagate neural networks to the Taiwan and United States markets. The results show that both of the intelligent methods obtained an accuracy of around 80% on the same data. The author has also compared the above-mentioned techniques with conventional statistical methods and demonstrated that the intelligent techniques outperform the conventional ones. A credit scoring model focused on the multilayer perceptron approach has been developed by Blanco et al. (2013). The proposed model has been applied to microfinance industry credit scoring. The experimentation performed resulted in the better performance of the multilayer perceptron as compared to other benchmark models like linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), and logistic regression (LR) etc. In a similar way, Rafiei et al. (2011) also proposed the multilayer perceptron for predicting the financial health of businesses in Iran. The proposed model has achieved an accuracy rate of around 98.6% and 96.3%, respectively, on training and holdout samples, which is higher than the accuracy rates achieved by multiple discriminant analysis, whose accuracy rate was likely to be only 80.6% and 79.9% on training and holdout samples. Authors such as

Chakraborty and Sharma. (2007) employed radial basis function networks for the prediction of corporate financial health. The results in this article show that RBFN received significant well classification performance in financial risk assessment due to its simple architecture and fast computation. Chen et al. (2013) employed DA, decision tree, and artificial neural network models and examined the differences in bankruptcy risk forecasting between Latin America and Central Europe. Dastile et al. (2020) conducted a survey on credit scoring models based on statistics and machine learning techniques. The author of this survey discovered that an ensemble of classifiers outperforms a single classifier. A credit card fraud detection system based on a deep learning model has been proposed by Abakarim et al. (2018). The author's proposed model is based on an auto-encoder that allows one to categorise credit card transactions as legitimate or fraudulent in real time. The author uses four distinct binary classifiers in order to compare the proposed model with them and the proposed model's effectiveness. The benchmark shows that the author's proposed model outperforms existing alternatives in terms of accuracy, recall, and precision. A new approach known as feedback mechanism for credit card fraud detection is developed by Jiang et al. (2018). The experimentation results show that author's approach is superior to others. Zhan and Yin (2018) have proposed automatic feature extraction method using neural network for Loan Application Fraud Detection of tech-driven internet finance company. In practice, this method has produced excellent results.

3.2.3 Support vector machines for Credit Risk Assessment

A support vector machine is a machine learning algorithm that uses kernel functions in order to transform the input vectors nonlinearly into a high-dimensional feature set, allowing the model to differentiate the data. The SVM's main concept is to find the optimal hyperplane that can classify the two classes and maximize the margin of split the data. SMVs have gained popularity due to their good generalization performance on high-dimensional data with a comparatively limited amount of data (Yang et al., 2011). Since bankruptcy prediction has sparked a great deal of research interest in the past, and new research has also shown that ML algorithms outperform conventional statistical methods. As such, Min & Lee (2005) applied SVMs to the bankruptcy forecasting model in an effort to formulate a new model with improved explanatory power and stability. A SVM and hybrid SVM are deployed by Jayanthi et al. (2011) for accurate prediction of bankruptcy data when fundamental network processes are nonlinear, non-stationary, and not defined. Bae (2012) demonstrates that the SVM model with radial basis kernel function (RSVM) performed exceptionally well in order to predict Korean manufacturing distress. Ribeiro et al. (2012) proposed a business distress prediction model centred on SVM+ that outperformed the benchmark SVM and multi-task learning with SVM by instinctually isolating the heterogeneous data into many structured groups depending on the size and annual revenue of

the businesses. For credit risk assessment, Zhou and Bai (2008) developed an effective support vector machine (SVM) using a genetic optimization algorithm (GA). The author in this work optimized the parameters of SVM classifiers using GA. The experimental results were compared with traditional discriminant analysis (DA) models, BP neural networks (BPN), and standard SVM and show the effectiveness of the proposed model. Zhen and Wenjuan (2016) have proposed a credit risk assessment technique for commercial banks based on improved SVM. A fuzzy integral-based SVM ensemble approach is presented in this paper. This method combines the output values of a wide range of aspects SVMs that have been granted special weights using a fuzzy integral. The results from the experiment demonstrate that the outcome of fuzzy integral support vector machines is acceptable. Mareeswari and Gunasekara (2016) have explored HSVM for credit card fraud detection mitigation. Using the proposed technique, various issues with existing systems, like scalability issues, extreme imbalanced classes, and time constraints, are eliminated.

3.2.4 Decision trees for Credit Risk Assessment

Decision trees are quite a traditional classification method that tends to indicate a decision framework in the form of a tree with leaf nodes and decision nodes. The intended attribute (class) of instances is indicated by the leaf nodes, and the decision nodes define the assessment on features, for each branch referring to a particular value. DTs have several benefits, including simple interpreting of the results, non-linear prediction, non-parametric form, good precision, relevance both to continuous and categorical factors, and identification of significant factors. DT classifiers include ID3, C45, CART, CHAID, MARS, and ADTree, among others. Li et al. (2010) demonstrated the utility of very well selected C4.5 and CART DTs in bankruptcy prediction tasks. In some other experiments, CART has been deployed to choose predictors, and then a classic LR model was built using the dummy variables predicated on the tree nodes (Brezigar-Masten & Masten, 2012). CHAID, C5.0, QUEST, and CART are used by Delen et al. (2013) for investigative purposes of the effect of financial ratios on corporate profitability. CHAID and C5.0 produced the greatest predictive performance of the four DT techniques. Dereliolu et al. (2008) have applied various machine learning models like multilayer perceptron (MLP) and a neural rule extractor (NRE) to real-life small and medium enterprises in Turkey. For that, feature selection has been performed by decision tree (DT), recursive feature extraction (RFE), factor analysis (FA), and principal component analysis (PCA) methods.

Some other machine learning models that are widely used for credit risk assessment are: logistic regression, probit regression or discriminant analysis as described by Leo et al. (2019) and Serrano and Gutiérrez (2016).

As per the analysis of the reported work, the author has discovered that there are numerous challenges in the field of credit risk assessment. Every model does have its own

set of risks and difficulties and cannot be entirely depended on for assessment. As per the well-known "no free lunch theorem," a single structural classifier is not an answer for credit ratings or perhaps even fraud detection. This is due to the various issues associated with this disease. Significant variation in the dataset results from various financial institutions from various geographic areas, and even from the same location, having different sets of rules. Therefore, numerous models are deployed by various researchers in different sets of problems to address the same issue. Hence, a proper investigation into the current issue needs to be conducted for the best possible result. Furthermore, to deeply understand the classification results, the interpretability of the models is another important research area.

3.3 Artificial Intelligence Based Credit Risk Models in Peer-To-Peer Lending

P2P lending, a new online unsecured credit model, has flourished nowadays and has become an effective complement to the conventional credit business. Credit risk, on the other hand, is unavoidable in this too. A critical barrier is developing a default prediction model for a P2P lending framework that can efficiently and reliably anticipate the default probability of each loan. Traditional statistical models cannot efficiently and reliably anticipate default probability due to the features of P2P lending credit information, such as higher dimensionality and class imbalance. As a result, an artificial intelligence-based credit risk model in peer-to-peer lending appears to be a viable solution. This section of the present work gives a literature review of various previously proposed artificial intelligence (AI) based credit risk models in peer-to-peer lending.

Jin and Zhu (2015) have proposed data mining (DM) methodology to anticipate the effectiveness of a P2P loan before it has been funded. Author has been used the random forest to extract the features in the modelling phase of the system. The whole approach has been implemented on data from the Lending Club. The empirical findings of this work indicates that term loan, yearly salary, loan balance, debt-to-income ratio, lending grade, and going to revolve line utilisation all play a significant role in loan defaults. Furthermore, the predictive accuracy of SVM, Classification and Regression Tree (CART), and Multi-layer Perceptron (MPL) is nearly identical.

Ha et al. (2019) worked on P2P lending in the United Kingdom and implied deep learning on it. The proposed design by the author is capable enough of evaluating the risks and returns of individual loans. The author has also applied the novel feature selection technique in order to eliminate irrelevant features and improve the efficiency of machine learning models. The exploratory experiments demonstrated that the developed model is capable of improving investment efficiency when compared to existing P2P lending methods.

For knowing the status of the borrower. Malekipirbazari and Aksakalli (2015) have used random forest (RF) based classification technique. Analysis of data from the popular online lending platform Lending Club (LC) has been performed in this work. The results of this work show that the RF-based model outperformed both FICO credit scores and LC grades in identifying good borrowers. In a similar way, Wang et al. (2018) have developed a behavioural scoring model which is based on a mixture survival analysis framework. In order to predict whether a borrower will default or not, a random survival forest has been used by the author, and a random survival forest is launched to design the time to default. The experiment has been performed on Chinese P2P loan dataset. The results of this work reveal that the developed ensemble mixture random forest (EMRF) outperforms the standard mixture cure model, Cox proportional hazards model, and logistic regression for prediction of the monthly dynamic probability of default. Furthermore, Cai et al. (2016) evaluated peer-to-peer lending behaviour online. For this work, the author chose the data from ppdai.com and implied logistic regression on it.

Credit scoring systems can be used to estimate the likelihood of loan default. Serrano-Cinca and Gutiérrez-Nieto (2016) have proposed the same thing. In this work, the author looks at 40,901 P2P loans in total. The factors determining loan profitability are examined, and it is discovered that all these factors differ from those that decide the likelihood of default. The findings indicate that peer-to-peer lending is not yet a fully efficient economy. The author discovered that in the analysed sample, a profit scoring system based on multivariate regression outperforms the results obtained using a traditional method based on logistic regression.

Li et al. (2018) have worked on real data that is available on online platforms in China (www.renrendai.com). The author in this work formulates the network's topology and systemic risk in the peer-to-peer lending market. The results indicate there is a meaningful correlation between the lending–borrowing topology of the network and defaulted loans, implying that all these factors have a positive bond to treat and prevent risk in the network.

A literature review on artificial intelligence-based credit risk models in peer-to-peer lending has shown that the research in this domain is very prominent. The existing study of the work on the same gives numerous solutions to the various challenges faced in peer-to-peer lending.

But with that, most of these machine learning models have been discovered to be black boxes with little to no interpretability capability. As a result, when it comes to implementing machine learning for credit risk modelling, financial institutions are treading carefully. Furthermore, the requirements of regulatory enterprises are typically influenced by traditional models, which do not address the difficulties associated with all of these alternative approaches. Even when they are not explicitly specified, the use of machine learning models necessitates complementary validation components, as well as the study and identification of new biases or interpretability, among other things. Interpretability

and transparency are required for the various models of credit assessment mechanisms, such as agreeing to give, behaviour and collection procedures, or fraud prevention. They must ensure that borrowed products are delivered in a safe, controlled, and monitored manner. Enterprises and enterprise customers necessitate interpretable models, which is also the focus of this present work.

This section's progression leads to a problem analysis, the formulation of a problem statement, and the proposed work research objective.

4 Data and Methodology

For experimental work in any kind of empirical research-based problem, the selection of a specific methodology is the most time-consuming and costly process. The present research problem of credit risk assessment in P2P lending based on artificial intelligence methods is basically a quantitative research. Quantitative research is one of the two different methods of empirical research. For this kind of research method, the appropriate selection of the dataset or material to be used and the selection of its attributes highly affect the experimental results. Hence, in the present work, the author has performed the initial screening of materials or dataset based on a literature review for selecting an appropriate dataset for the current research problem and that approach was selected only after alternatives have been recommended, the ideologies that are obviously inappropriate are omitted, and the attention has been focused on those that appear to be feasible.

In the same way, after a critical analysis of the literature review conducted in the previous chapter, the author has chosen the most accepted and preferred machine learning algorithm for the current research problem of quantifying credit risk in P2P lending and compare the proposed models' performances in terms of their effectiveness to determine their potential and cause of limitations.

The proposed methodology for the current research problem is shown in figure 1.

The following part of this chapter will be discussing the proposed methodology for credit risk assessment in P2P lending, as well as the data used to validate the proposed method in detail.

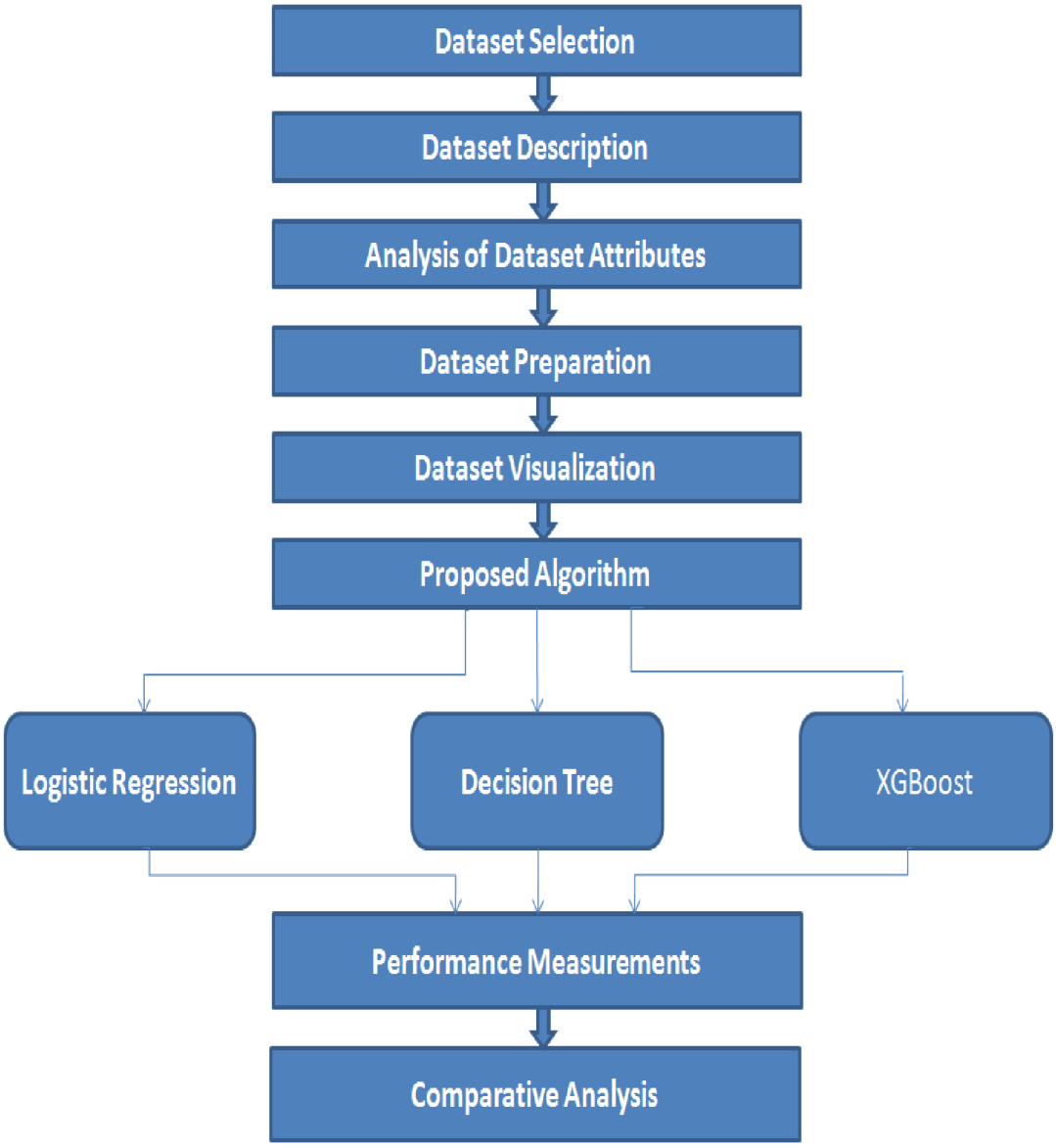
4.1 Data

From a lender's perspective, the success of a borrower's loan repayment is largely determined by the borrower's ability, willingness, and other socioeconomic characteristics. The goal of the proposed method is to distinguish prompt re-payers from those who do not repay. To validate the proposed method, the publicly available loan dataset has been chosen. The following sub section of this chapter will discuss all the phases from the dataset collection to dataset preparation before simulation of the proposed models on the same.

To validate the proposed method, the author chose the publicly available benchmark dataset after an initial screening through a literature review and after omitting another inappropriate method in order to select the most feasible approach in dataset selection.

The present research work is based on quantitative data. The public available benchmark dataset used for the research work is loan data of a financial institution called Lending Club.

Figure 1: Proposed Research Methodology



The author has been using a loan default dataset for this study. The dataset contains 32,581 observations (rows), and 12 features/attributes (columns). The dataset contains the loan information that is categorized by its six different attributes. The names of these six attributes with regard to loan information are "loan_intent", "loan_grade," "loan_amnt," "loan_int_rate," "loan_status" and "loan_percent_income." The first attribute, "loan_intent,"

is classified into six different categories. The names of the "loan_intent" categories are personal, education, medical, venture, home improvement, and debt consolidation. The second attribute, "loan_grade," is classified into seven different classes, from class A to class G. The third attribute is "loan_amnt," or loan amount, is in the range of 500 to 3500. The fourth attribute, "loan_int_rate," or loan interest rate, falls between 5.42% and 23.22%.The fifth attribute is "loan_status" which is either 0 or 1. The sixth or last attribute is "loan_percent_income," which is in the range of 0 to 0.83.

Other than this, six more attributes of the chosen loan default dataset belong to the information of the lender or client. These six attributes are "person_age," "person_income," "person_home_ownership," "person_emp_length," "cb_person_default_on_file," and "cb_person_cred_hist_length".

There is no dummy variable in the dataset and each attribute or dataset column has 32,581 entries and values.

Table 1 given below presents the detailed description of each attribute of the dataset.

Table 1: Detailed description of each attribute of the dataset

Attributes	Description
person_age	detail of the person or lender age varying from 20 to 144
person_income	lender annual income from 4000 to 6000000
person_home_ownership	home ownership of a lender classified into own, rent, mortgage and other
person_emp_length	the total length of lender employment from fresher to 123 year of experience
loan_intent	type of loan classified as personal, education, medical, venture, home improvement, and debt consolidation
loan_grade	credit grade classified from A to G
loan_amnt	credit amount ranging from 500 to 3500
loan_int_rate	loan interest rate, falls between 5.42% and 23.22%
loan_status	loan status which is either 0 or 1
loan_percent_income	total income generated in range of 0 to 0.83
cb_person_default_on_file	failure to repay a loan or repay the loan
cb_person_cred_hist_length	span of credit history in range of 2 to 30

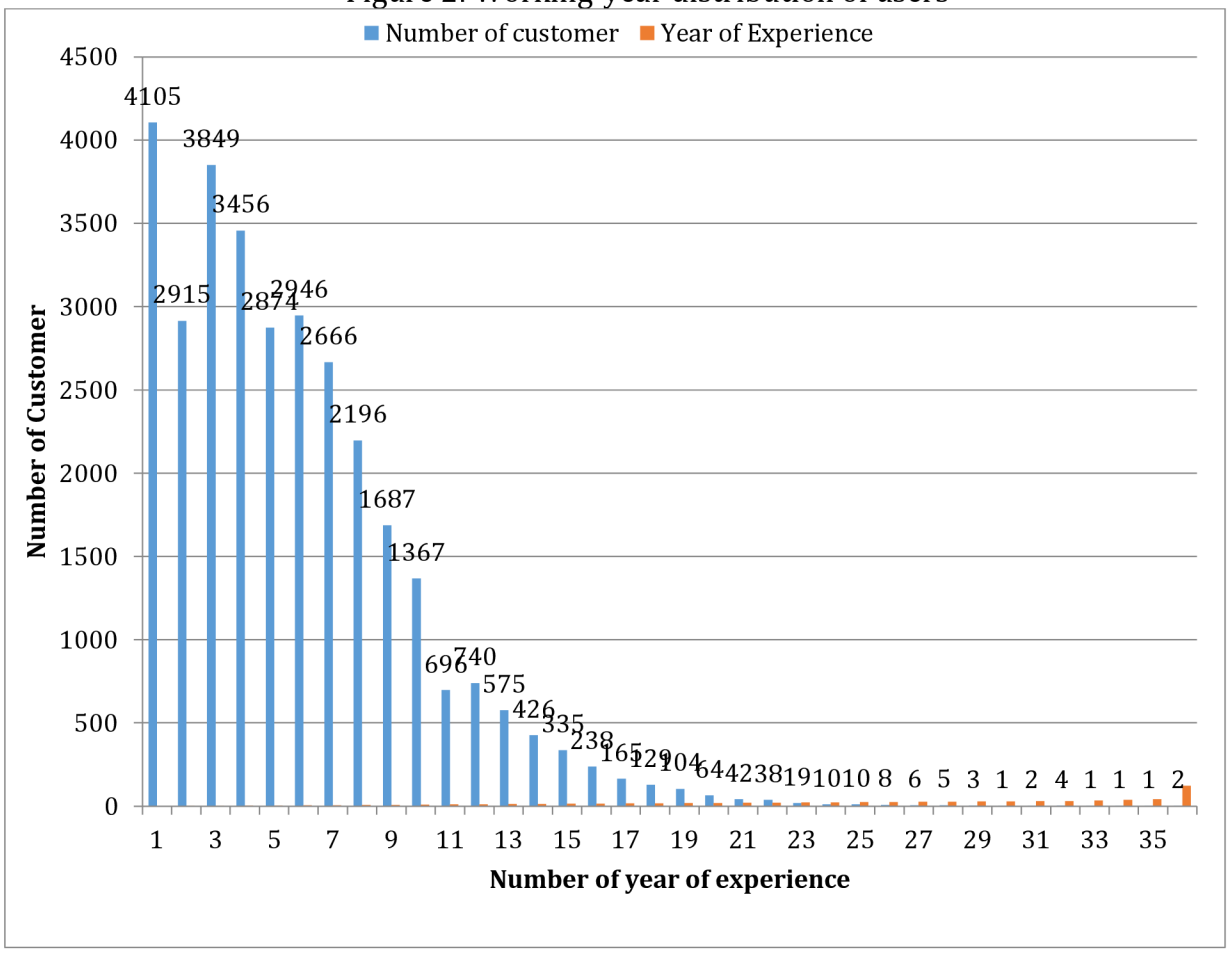
In the publicly available benchmark loan default dataset, some of the values against the attribute were missing. In order to tackle this issue, a conventional method was used,

and the mean of that particle column was used to replace the missing and NaN value in the dataset.

After dataset preparation, the proposed methodology moved toward dataset visualization or dataset analyzation in order to explore the characteristics of a dataset and to understand each dimension of the dataset before simulating the proposed model on the same. This step will be used to improve the accuracy and speed of the proposed models.

In figure 2, dataset visualization in term of working-year distribution of borrowers is given. As per figure 2, the minimum year of experience of the customers is 0 and the maximum is 123 years. It is also depicted by the figure that the majority of the customers have experience of less than 10.

Figure 2: Working-year distribution of users



Further, Figure 3, presented below, gives a detailed visualization of the customers' home ownership data. It is depicted by the figure that most of the customers have their

own home. The total count of the customers who have their own home is 16,446, while there are 13,444 customers who have a mortgage and 2,584 customers on rent.

Figure 3: Customers' home ownership

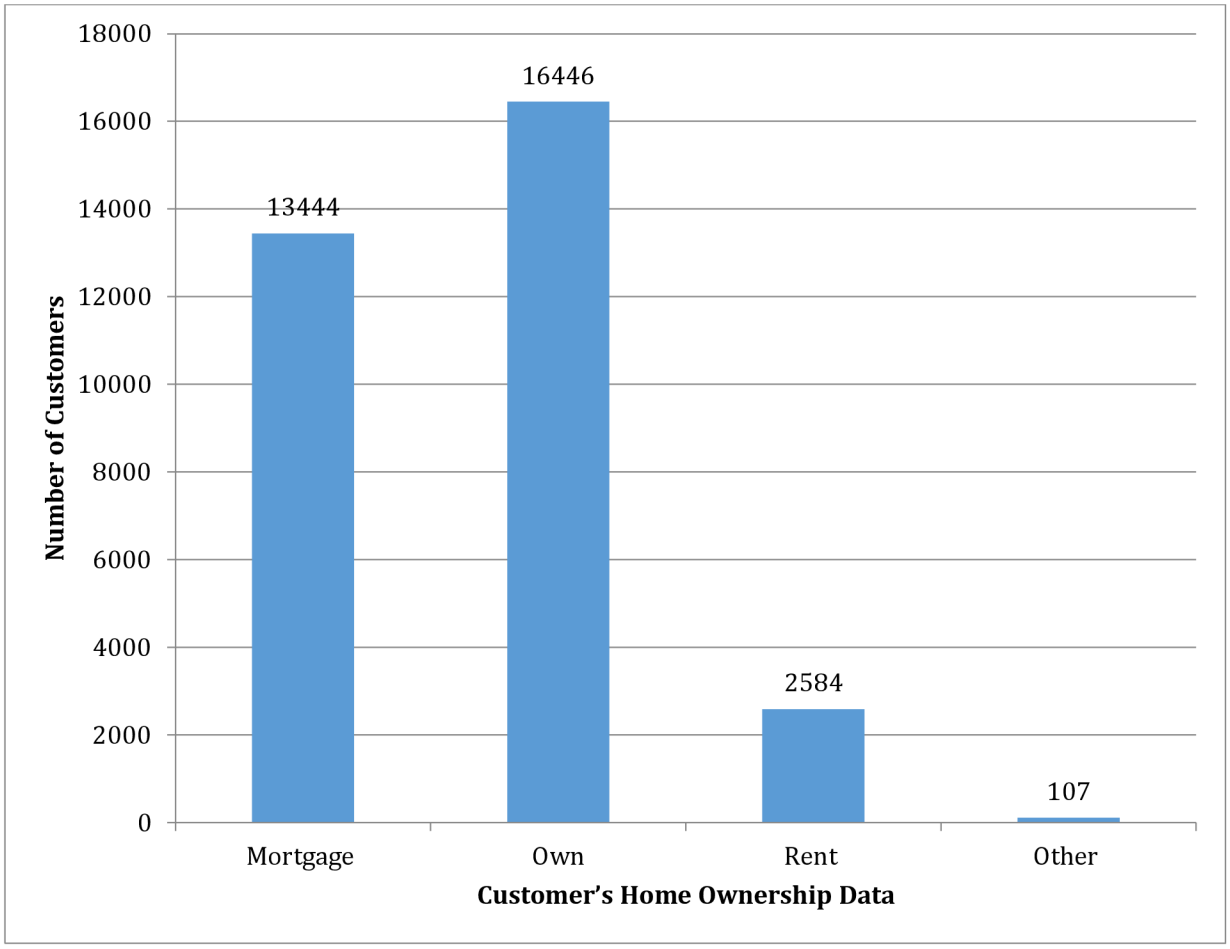


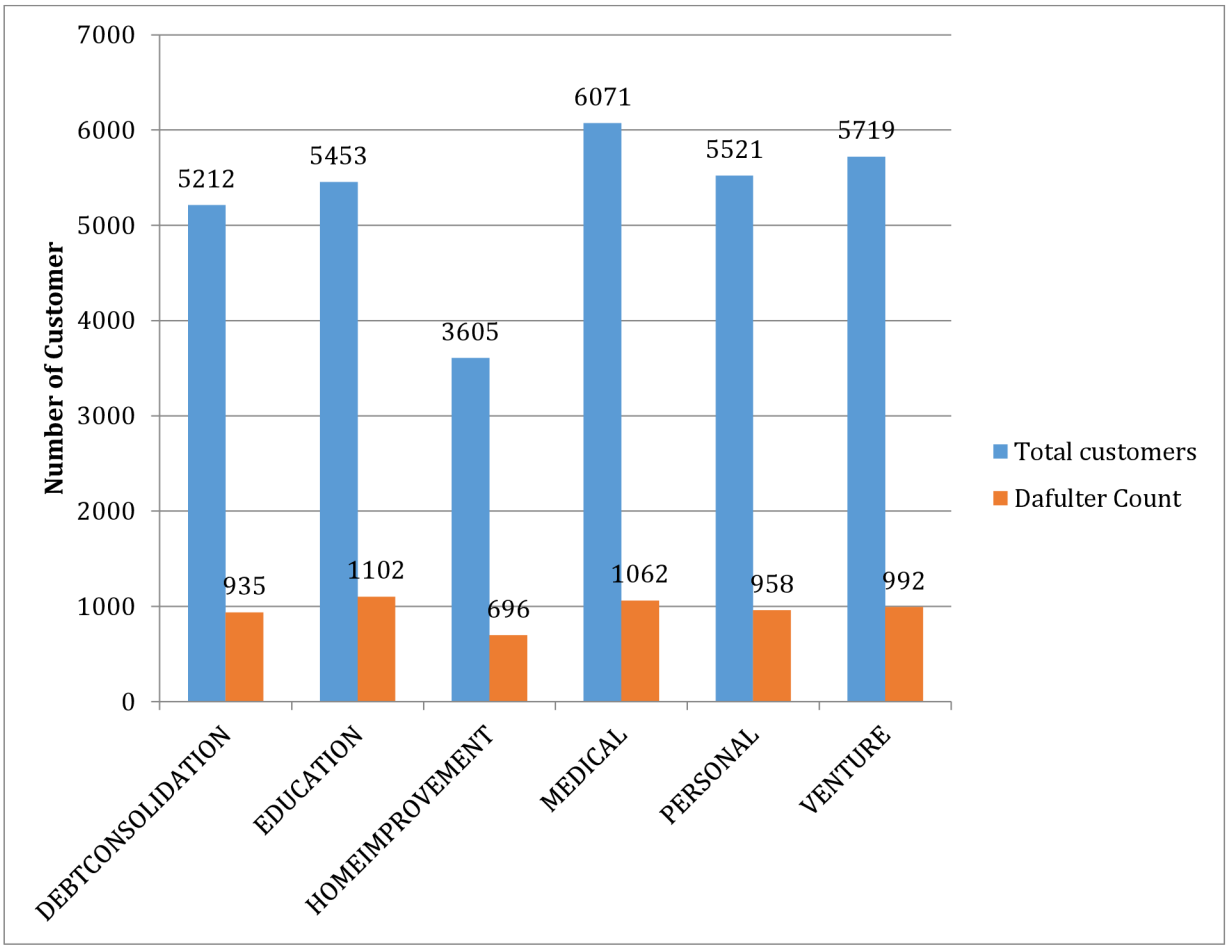
Figure 4 depicts the customers' reason for loan. It is clear that medical and education loans have a higher rate of application, whereas home-improvement loans have the lowest rate application.

A comparison has also been made between the total number of customers applying for loans for various purposes and the defaults for different loan purposes. It has been observed that medical and education loans have a higher rate of defaults, whereas home-improvement loans have the lowest rate of defaults.

Figure 5 presents loan status by loan grade for total count of customers and loan status by loan grade for total count of defaults. The loan grade ranges from A to G.

The maximum defaults rate is one grade D (3,626) and grade C (3,256), while minimum defaults rate is one grade E (465), grade F (112) and G (36). In short, the overall defaults rate is in between the C and D grades.

Figure 4: Loan purposes distribution



Finally, figure 6 represents the loan status of the customers. Number 0 represents customers who did not get the loan, and number 1 represents the customers who got the loan out of the total number of customers. It can be observed from the figure that the count for the customers who did not get the loan is greater than the customers who did get the loan.

In figure 7, the percentage value of the customers who get loan over the percentage value of the customers who do not get loan is given.

Figure 5: Loan status by loan grade

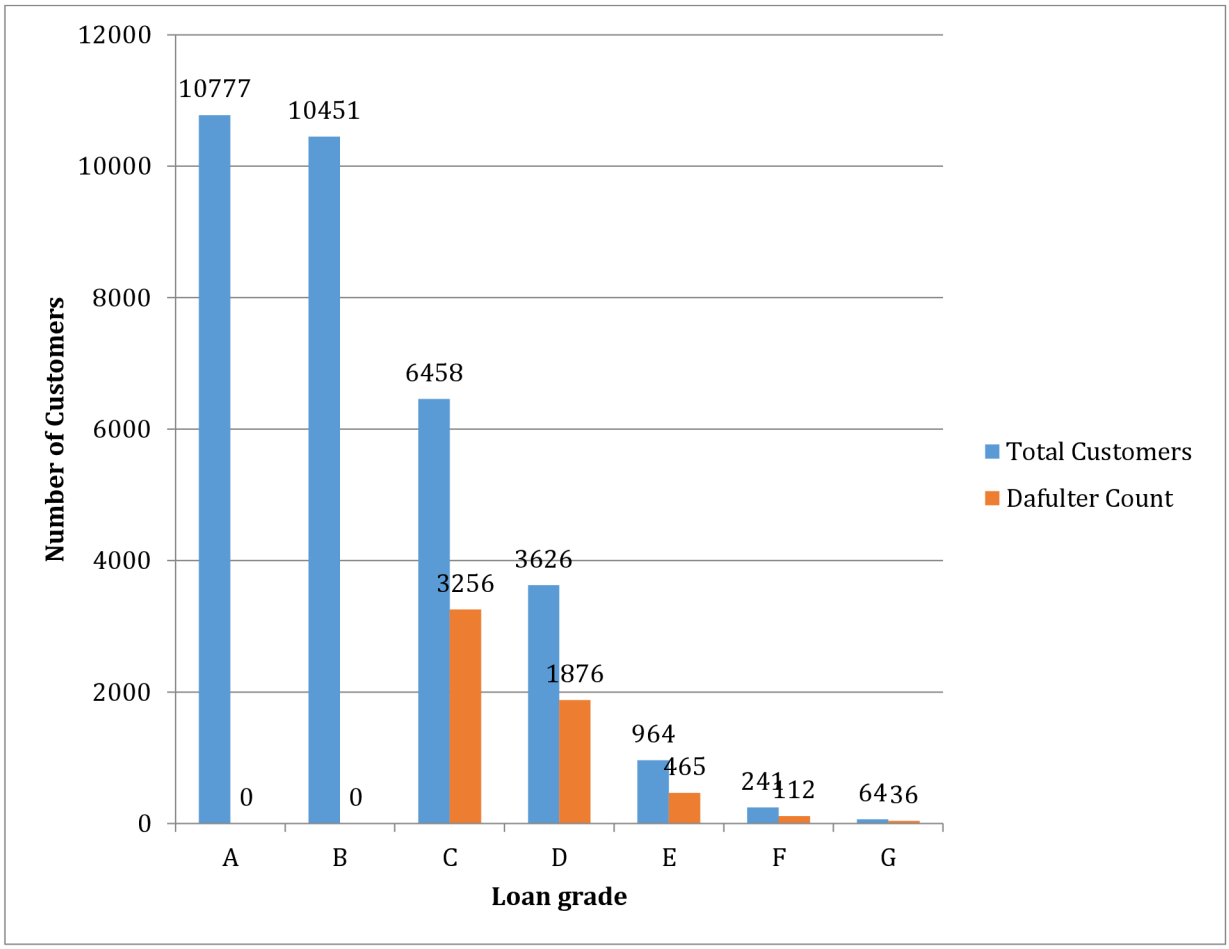


Figure 6: Loan status for customers

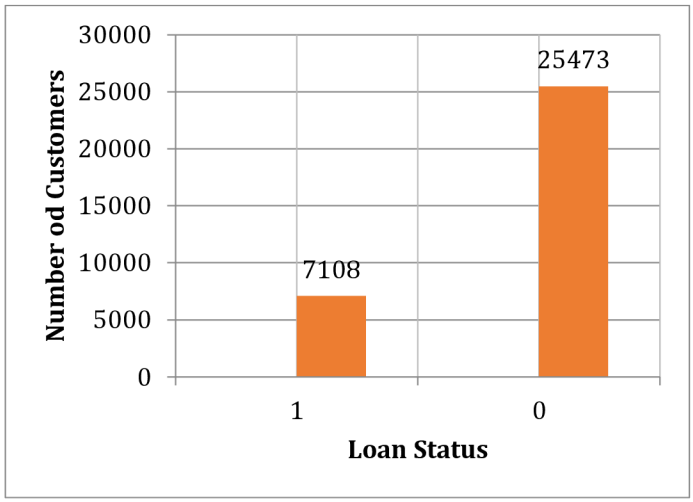
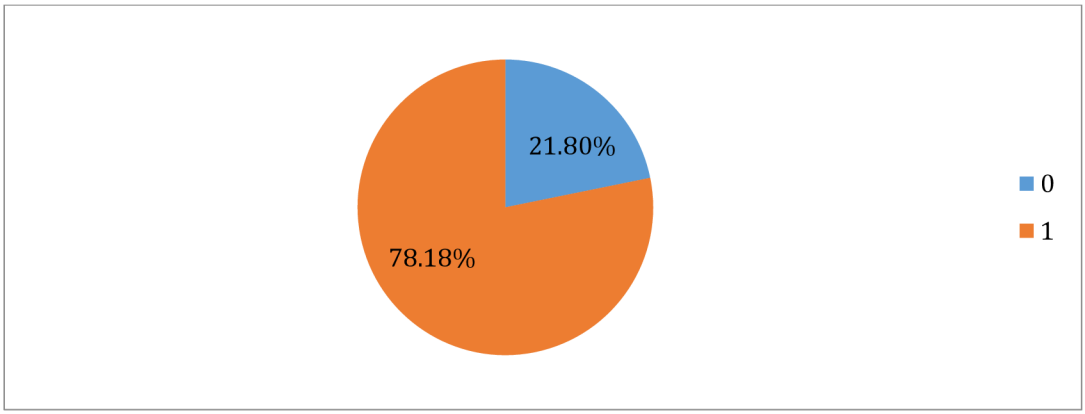


Figure 7: Percentage of loan status



4.2 Methodology

Nowadays, machine learning has led to advancements. The true strength of machine learning is yet to be realized, particularly in several fields of risk prediction associated with credit loans from financial institutions and peer-to-peer lending on online platforms. Various experiments have been carried out in the present research work using various machine learning algorithms. The names of these machine learning algorithms are logistic regression, decision tree, and XGBoost.

All these machine learning algorithms are selected based on their suitability and preference, which have been proved in past research work. Using these machine learning algorithms, the author performed the experiment in order to predict all conceivable factors influencing repayment. The following tests use various techniques to verify the impact of the factors on repayment default. All three algorithms were applied to a selected publicly available loan dataset to compare their potential, limitations, and drawbacks, and the default forecasting accuracy and performance of all three algorithms is also calculated and compared. In this section, author discusses the machine learning methods used in the present study, which will help to better illustrate the proposed research.

4.2.1 Logistic Regression

Logistic regression is a variant of ordinary regression in which the dependent variable is a binary variable (Jii Witzany, 2019). That means the dependent variable has only two values: zero and one (typically representing the occurrence or non-occurrence of certain events). On the other hand, independent variables in logistic regression are continuous, categorical, or can be both. For example, making predictions on whether a tumour is malignant or benign, or whether an e-mail is spam or not, are examples of binomial outcomes of logistic regression. Logistic regression can produce multinomial results, such as predicting a favoured cuisine that can be from Chinese cuisine, Italian cuisine, Mexican cuisine, etc. There could still be ordinary outcomes, such as product ratings ranging from 1 to 5, and so on. As a result, logistic regression is concerned with the prediction of categorical attribute values or variables. On the other hand, linear regression is concerned with the prediction of values of continuous variables, such as real estate prices over a three-year period. Furthermore, logistic regression, apart from ordinary linear regression, does not really make the assumption of having a linear relationship between independent variables and the dependent variable. It also does not make the assumption that the dependent variable or the errors are normally distributed (Jii Witzany, 2019).

Logistic regression does have the following clear benefits: easy implementation, computationally efficient, learning or training efficiency, and easiness of regularization and input features do not need to be scaled. This algorithm is primarily used to solve industrial-scale problems.

Because the outcome of logistic regression is a statistical likelihood score, in order to use it to solve business problems, customized performance measures must be specified in order to obtain a cut-off that can be used to classify the benchmark. Furthermore, slight noise in the records and multi-collinearity has no effect on logistic regression.

With numerous advantages of logistic regression, it also has various disadvantages such as the fact they are unable to solve non-linear problems because their decision surfaces are linear, prone to over-fitting, and ineffective unless all independent variables are

recognized (Akinyemi, 2009). A few practical applications of logistic regression include making predictions of the risk of developing a specified disease, cancer diagnosis, making predictions of the death rate of accident victims, and predicting the likelihood of failure of a particular process, framework, or item in designing.

Other than this, one of the biggest limitations of using logistic regression is that a sample of a large size is required for it to provide reasonable numbers in both response variable categories. With an increase in the number of explanatory variables, the sample size required is also large (Akinyemi, 2009).

With so many advantages, logistic regression's limitations can either be ignored or managed. Hence, logistic regression is by far the most frequently used approach in credit risk assessment, especially in customer classification. Among other uses, such model can also be used to evaluate the likelihood of timeous loan repayment and to decide if a client did belong with one of two groups – reliable or unreliable borrowers etc.

Consequently, according to Mestiri and Hamdi (2012) the primary objective of a logistic regression model in credit risk assessment is to decide the conditional likelihood that a particular observation belongs with one of two groups: the group of good or bad customers. Clients who have already defaulted on a loan are referred to as bad customers, while clients who have paid back their loan are referred to as good customers.

In our work, the performance assessment model is based on estimation using categorized variables that have been transformed using the WOE (Weight of Evidence) transformation. WOE enables continuous independent variables to be transformed into a collection of bins that measure the similarity of the dependent variable's distribution and denote the number of events (Anderson, 2007). That kind of technique also solves the outlier problem, permits the modelling of nonlinear relationships by employing linear models, and allows for a better interpretation of the data's relationships. WOE formulates the analysed group's in-between difference (for example, borrowers who repay their loans on time vs. late borrowers who miss instalments) within a given category of the selected variable.

Mathematically, logistic regression calculates the probability of a binary outcome as a function of a set of predictor variables and regression coefficients that are presented below:

$X = [X_0, X_1, X_2, X_3 \ldots\ldots X_n]^T$ Independent Variables
$B = [\beta_0 + \beta_1 + \beta_2 + \ldots\ldots \beta_n]^T$ Regression co-efficient
P=probability
β0—Constant

The logistic regression function output provides the probability and ensures that the probability is either 0 or 1. The overall function is divided with the 1+ function.

$$\text{In}\,\frac{P}{1+P} = \beta_0 + X_1\beta_1 + X_2\beta_2 \ldots\ldots X_n\beta_n \quad (1)$$

Forward method has been used to compute the outcomes. The forward method begins modelling with a step-by-step method that starts with one variable and continues to add relevant variables while eliminating irrelevant variables.

4.2.2 Decision Tree

A decision tree is a predictive model. It is also called a rule-based model. It attempts to map the data for an item toward a conclusion about the item's specified or targeted value. In a decision tree, every interior node corresponds to a variable is a path to a kid reflecting one of the variable's possible values. A leaf signifies the expected value of a target variable and provides the values of the variables depicted by the path from the root (Molnar, 2020).

A decision tree explains a tree-like structure in which leaves comprise classifications and branches reflect feature conjunctions that lead to all of those classifications.

The learning of a decision tree can be made possible by dividing the origin set or source set into subsets centred on a feature value test. The above process continues recursively on each sourced subset. Once splitting is not possible or if a single common classification has resulted towards each attribute of the sourced subset, the recursion is finished.

A decision tree is a mathematical model that uses probabilities to calculate the most likely outcomes. Nodes and branch-based decision trees can be used for both regression and classification problems (Kast and André Lapied, 2006). In classification problems, decision trees create sets of decision sequences at every step of data steaming from which the label of unlabelled data can be predicted. The decision variable in a classification tree is categorical (an outcome of Yes or No), whereas the decision variable in a regression tree is continuous. According to Kast and André Lapied (2006), decision tree has the following advantages:

- the decision tree algorithm is useful for regression and classification problems;
- it is easy to interpret;
- it is capable enough of filling null values in features with the most likely value;
- it has outstanding performance due to the efficiency of the tree traversal algorithm;

and disadvantages:

- decision tree may experience over-fitting;
- the choice of optimal number of decision trees is a major concern in various classification problems;
- it may be susceptible to sampling error;

- the size of the tree may be hard to control and unstable.

The primary difference among decision tree algorithms is how the features are chosen and the pruning mechanism used to create the tree. Quinlan created the ID3 decision tree algorithm based on the idea of learning systems. ID3 takes an optimization technique, building decision trees top-down recursively and using gain ratio as a feature extraction means to ensure a standard classification tree is achieved.

Mathematically D, the data partition, should be configured as a training dataset of class-labelled tuples. Assume the class label attribute has m distinct values to define m distinct classes $C_i(i = 1 \ldots . m)$. Further set C_iD as set of class tuple C_i in D . Modulus of D is tuples count in D. Modulus of C_iD is tuples count in C_iD. Keep set the N node to hold the splitting D tuples. As that of the splitting attribute of the N, the attribute with the highest information gain is selected.

To begin with, the data entropy (expected information) required to classify the tuple in D is defined by:

$$Info(D) = -\sum_{t=1}^{m} p_i \, log_2(p_i) \tag{2}$$

Assume it has applied some attribute A to the divided tuples in D. A has v unique values as per the training data$\{a_1, a_2, \ldots . a_v\}$. With using attribute A, tuples in D can be split into v subsets $\{D_1, D_2, \ldots . D_v\}$.

To determine the correct classification, the information must be evaluated by:

$$Info_A(D) = \sum_{j=1}^{v} \frac{modD_j}{modD} X \; Info(D_j) \tag{3}$$

where $\frac{modD_j}{modD}$ is weight of split jth

The information gain is then equivalent to the difference among the original and new information demand.

$$Gain(A) = Info(D) - Info_A(D) \tag{4}$$

The C4.5 and C5.0 algorithms have been enhanced to ID3, wherein the gain ratio has been used to resolve the characteristics that are biased toward test results with several outcome measures, so they can deal with continuous features. The gain ratio is defined as follows:

$$GainRate(A) = \frac{Gain\ (A)}{Splitinfo_A(D)} \tag{5}$$

C4.5 is an improved version of ID3, an earlier decision tree-based learner. The features are evaluated for selection based on many criteria such as min entropy, gain ratio, and so on, and it has become a renowned classifier in data mining due to its flexibility of learning and ability to comprehend results. The same method is used in the present research work.

4.2.3 XGBoost

XGBoost is a powerful machine learning algorithm for tree boosting that data analysts are using to accomplish cutting-edge outcomes for a variety of machine learning problems. Scalability is perhaps the most crucial component in XGBoost's success. XGBoost primarily broadens and enhances the GTB in three key areas. First, for dealing with data sparsity, an innovative tree training algorithm is used. Secondly, to resolve instances of weight training in approximate tree training, a conceivably legitimate weighted quantile sketch process is adopted. The last and third is that it is possible to support parallel and distributed computing. The XGBoost model, like GTB, is also learns in an additive fashion. (Wade, 2020)

The XGboost algorithm, which is centered on the GTB algorithm, introduces the regular function to regulate model over-fitting.

As a result, the target function of XGboost can be written as:

$$min\{obj\} = \min\{L(F_m(X), Y) + \Omega(f) + C\} \quad (6)$$

$$= \min\{\sum_i l(y_i, y'_i) + \sum_i \Omega(f_t) + C$$

Where:

- X = as per data input, all variables in use for prediction;
- Y = the outcome variables must be predicted;
- y_i = the true worth of the I[th] observation; the variable's final true worth must be predicted;
- y'_i = the predicted value of the I[th] observation;
- $F_m(X)$ = after m iterations, m weak classification techniques with a fixed weight were incorporated into the final predictable model;
- $l(y_i, y'_i)$ = a function of the difference between the actual truth value and the predicted value of the I[th] observation;
- $L(F_m(X), Y)$ = the error function; the discrepancy between the final predicted variables and the real ones after m iterations;
- $\Omega(f)$ = the final prediction model's regularization function after all iterations;
- $\Omega(f_t)$ = the regularization function in the t-wheel iterative process;

- Obj = the function is objective; the product of the loss and supplemental functions; each and every machine learning iterative process is a method that maximizes or minimizes the objective function. It must be ascertained if it is maximized or minimized based on real problem.

This is an alternative that is currently used in machine learning. The computer is given a specific iteration target. Then, depending on the particular problems being studied, the constant term C can be preferentially omitted. The routine item function, for example, is a function of "leaf numeric T" and "leaf node output result ω_j" in each sub-decision tree model.

$$\Omega(f_t) = yT + 1/2\tau \sum_{j=1}^{T} \omega_j^2 \quad (7)$$

These formulas could be used to define the algorithm's frequent term. This concept is a much more precise version acquired by the algorithm designer after such a huge series of experiments, and scientists can use it effectively. The emergence of standard items can actually prevent the occurrence of over-fitting. As long as the algorithms iterative process of learning keeps going, the error function decreases gradually, whereas the frequent terms keep expanding. A genuinely good model strikes a fair balance between algorithm accuracy and fairness. Using second derivative, that can locate the extreme value, it is possible to find the weak classifier that minimizes the objective function. Because the optimal solution can be found to minimize the target function.

According to Boschetti and Massaron (2018), XGboost algorithm has a number of outstanding characteristics. The following are the most significant advantages:

1. The regularization step is used to decrease the occurrence of over-fitting.
2. Parallel processing speeds up operations.
3. Users can determine user-defined optimal control goals and assessment methods in XGBoost, increasing flexibility.
4. XGBoost includes rules for dealing with incomplete data or missing values.
5. To regulate the complex nature of the decision tree, XGBoost employs special pruning actions.
6. XGBoost includes cross-validation to easily identify the optimum number of iterations.
7. XGBoost can keep training using existing models.

For the present research work, the XGBoost model has been developed using R software, and the same has been simulated using a publicly available benchmark loan dataset for credit risk assessment. This model serves as a solid foundation for default forecasting.

In this work, there have been three primary reasons for selecting the Logistic Regression, Decision Tree, and XGboost algorithms as research tools for credit risk assessment in

P2P loan lending. Firstly, P2P systems have a plethora of review content for projects and borrowers. In the presence of multiple variables, it is indeed hard to predict the default risk precisely using artificial processing. While models like XGboost can implement default forecasting by fully automated iteration without manual involvement of supervision and have deep theoretical and practical significance in the context of the P2P industry, automation is being pursued gradually. Secondly, logistic regression, decision tree, and XGboost algorithms are the most advanced machine learning techniques that have been developed recently. Such algorithms have been improved significantly on the basis of the previous machine learning techniques, which enhance predictive performance, are computationally efficient, and minimize the likelihood of over-fitting. The algorithm has received positive feedback from a variety of academics in a range of high-quality data analysis and forecasting projects. Thirdly, the dataset used for the present work on the default rate of P2P is massive and multidimensional. It is hard to obtain the desired research outcomes using a generalized method. As a result, the logistic regression, decision tree, and XGboost algorithms, which excel at dealing with high-dimensional data, are very consistent with the findings of the present work.

4.3 Performance Measurements

The performance measurement of the proposed models (logistic regression, decision tree, and XGboost) for credit risk assessment in P2P lending is the most important phase of the proposed research methodology. Because the actual state of the repayment has to be compared with the prediction produced by a proposed model, by comparing the model to real data, it is indeed possible to assess the model's performance and, as a result, its applicability to new cases (loans). The following is a brief description of how prediction models' performance and effectiveness can be evaluated.

For numerous cases, knowing how frequently the model under consideration is incorrect is insufficient. It will be more crucial to understand how frequently it fails to accurately predict a specific outcome. The confusion matrix gives valuable data about the model's ability to predict a specific result.

The confusion matrix comprises four cases that are possible in reality. It has assumed only two main categories for all four cases, and those are positive and negative. False Negatives, or FN, define the count of positive cases incorrectly predicted or classified as negative. False Positives, or FP, define the number of negative cases wrongfully predicted or classified as positive. True Negatives, or TN, are defined as the count of the number of negative cases correctly labeled into a negative class, and True Positives, or TP, are defined as the count of positive cases correctly labeled into the positive class (Batarseh, 2020).

Using these four cases of confusion matrix, it may lead to being used to deduce multitude of performance measures. Those are

1. accuracy
2. true-positive rate
3. false-positive rate
4. sensitivity
5. specificity
6. precision
7. recall
8. F1 score

Accuracy is defined as the ratio of correct predicted instances or samples to all instances or

$$\text{Accuracy} = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Postive + False\ Nagative} \quad (8)$$

Error rate is defined as the ratio of incorrectly predicted instances or samples to all instances.

True-positive rate or positive predictive value is defined as the ratio of instances predicted correctly as positive to all positive instances.

False-positive rate or negative predictive value is defined as the ratio of instances predicted incorrectly as positive to all negative instances.

Sensitivity is defined as the ratio of instances correctly predicted as positive to all instances predicted as positive.

Specificity is defined as the ratio of instances correctly predicted as negative to all instances predicted as negative.

Precision is defined as

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (9)$$

(In binary classification precision is specificity.)

Recall is defined as

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (10)$$

(Recall is also known as sensitivity in binary classification.)

F1 score is the final metric. The F1 score is calculated by taking the harmonic mean of precision and recall.

$$F1\ Score\ = \frac{2}{Recall^{-1} + Precision^{-1}} \tag{11}$$

Or

$$F1\ Score\ = 2\frac{Recall\ X\ Precision}{Recall + Precision} \tag{12}$$

Or

$$F1\ Score\ = \frac{True\ Positive}{True\ Positive + \frac{1}{2}(False\ Positive + False\ Negative)} \tag{13}$$

Two types of error rates can be faced in the credit risk assessment of loan datasets. The error rates are the type 1 and type 2 error rates. The percentage of defaulted or bad customers who are considered decent is defined as the type 1 error rate. The high percentage value of the type I error rate indicates that the financial institution is at credit loss risk.

On the other hand, the type II error rate defines the percentage value of good customers who are predicated as default. The high percentage value of the type II error rate indicates that the financial institution is subjected to significant business risk over a longer period, implying that the financial institution has restrictive policies for lending money and may lose business. The provided confusion matrix is generated for the specified cut-off point, which is a threshold that is used to decide if an observation relates to one of several classes.

4.4 Comparative Analysis

In order to handle the financial risk and protect small investors, financial institutions or organizations dealing with P2P lending should be able to make effective predictions of borrowers' defaults because numerous companies in the past have faced large-scale collapses due to a lack of the same. Borrowers who default on loans do not just harm lenders' interests and social trust but also pose financial risks. For a few years, academics have worked hard to build a machine learning approach that will assist regulators and P2P platforms to make accurate and efficient predictions of loan defaults. Due to the advances in machine learning technology, it has become easy to get an effective prediction of default in loan lending. However, each machine learning model faces many challenges in order to optimize the results. Of the numerous challenges faced by machine learning models, the most common are tackling dataset imbalance, hyper parameter optimization for the model's

configuration, evaluation of each class's classification effectiveness, and the true assessment of the model's performance, etc.

The performance of each type of machine learning model can vary due to the built-in constraints of the models. Hence, for any particular default prediction problem for any financial institution, it is necessary to simulate the available dataset on different machine learning models and also need to compare their performance for validating the results.

The same has been implemented and explored in the present research work in order to judge the effectiveness of the proposed models for predicting defaults in a given dataset and to improve the risk prediction capacity of the models.

5 Results and Discussion

In an effort to identify both credit-worthy and non-creditworthy clients, credit risk is assessed through the simulation of various classification models (logistic regression, decision tree, and XGBoost) on the publicly available benchmark loan dataset. As it has been well known, a very popular strategy for credit risk assessment is to apply some form of classification model to the customer's past data in order to discover a link between the client's attributes and loan failure. An accurate classifier that distinguishes between good and poor clients is a key element of modern credit risk rating approaches. Credit risk assessment using machine learning approaches is becoming more popular as a result of its significance and improved accuracy rates.

Hence, the proposed research methodology of the present work has been targeted toward performing credit risk assessment of loan default datasets under peer-to-peer lending in order to distinguish between good and poor clients. Various experiments have been carried out in the present research work using various machine learning algorithms. The names of these machine learning algorithms that are used for credit risk assessment of loan default datasets are logistic regression, decision tree, and XGBoost. All three algorithms have been applied to a selected publicly available loan dataset to compare their potential, limitations, and drawbacks, and the default forecasting accuracy and performance of all three algorithms is also calculated and compared.

The following section will discuss the results and findings of the chosen three machine learning algorithms (logistic regression, decision tree, and XGBoost) on loan default datasets.

5.1 Logistic Regression

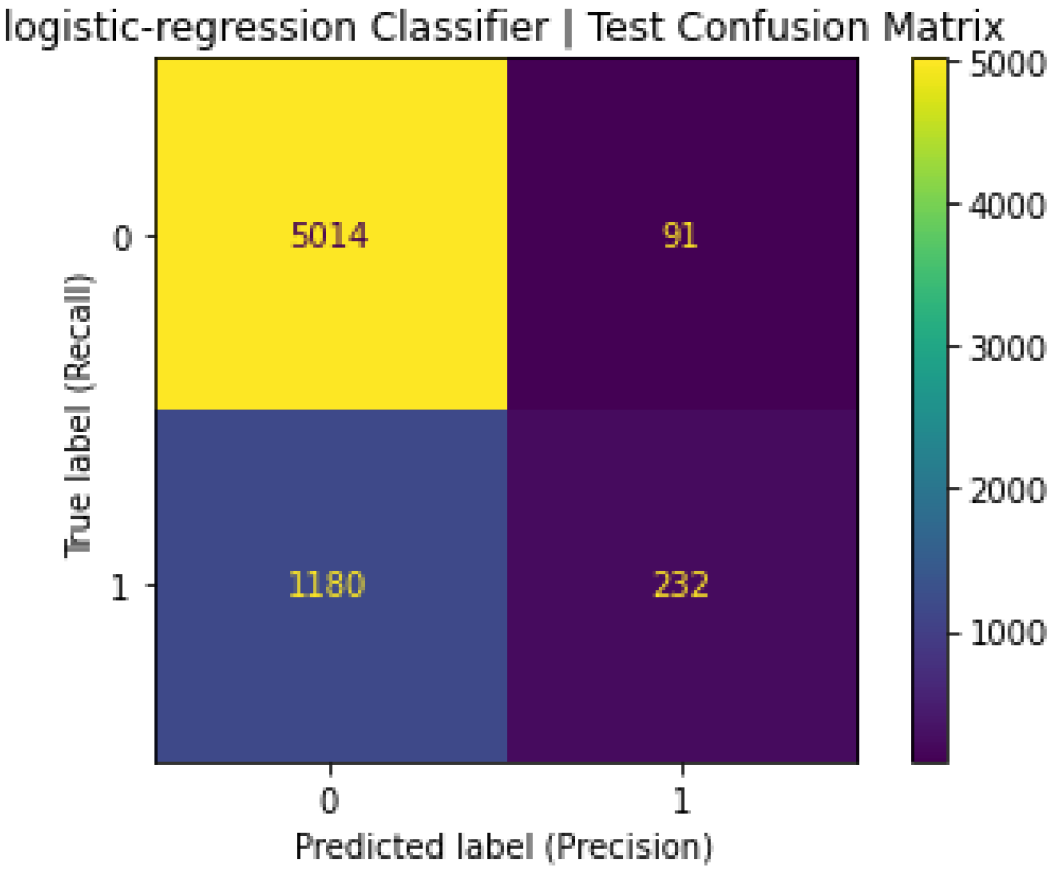
Machine learning and data mining applications rely heavily on classification algorithms. Classification challenges account for roughly 70% of Big-data issues. There are many various classification challenges to choose from, but logistic regression is a popular and effective regression method for addressing binary classification issues. Our current research problem of identifying both credit-worthy and non-creditworthy clients from loan dataset is also a binary research classification problem. Hence, logistic regression classifier has been simulated on the loan default dataset.

While simulating the logistic regression classifier on the loan default dataset, the dataset has been split into 80% for training and 20% for testing. The results obtained from the proposed experimentation have been elaborated in detail in the following sub-section.

A confusion matrix is kind of a table that is basically used to assess a classification model's performance. With that, a model's effectiveness can also be visualized. The ratio of right to wrong predictions is totalled class-by-class in a confusion matrix.

The confusion matrix obtained by simulating the logistic regression on loan default dataset is shown in figure 8.

Figure 8: Logistic regression's confusion matrix



Non-diagonal elements are inaccurate predictions, whereas diagonal values are accurate predictions. On the y axis of the confusion matrix, accurate or true labels are taken. On the other hand, on the x axis, predicated labels are considered. Actual forecasts are 5,014 and 232, whereas faulty guesses are 91 and 1,180.

As has been already mentioned, a confusion matrix is composed of four cases. The four cases are: False Negatives, False Positives, True Negatives, and True Positives.

Accordingly, the values of all four cases from the present confusion matrix (Figure 8) are:

False Negatives = 1180
False Positives = 91
True Negatives = 5014
True Positives = 232

Other than this, the results of other parameters of the confusion matrix known as accuracy, precision, F1 score and recall are shown in figure 9.

Table 2: Accuracy, precision, F1 score and recall result of logistic regression

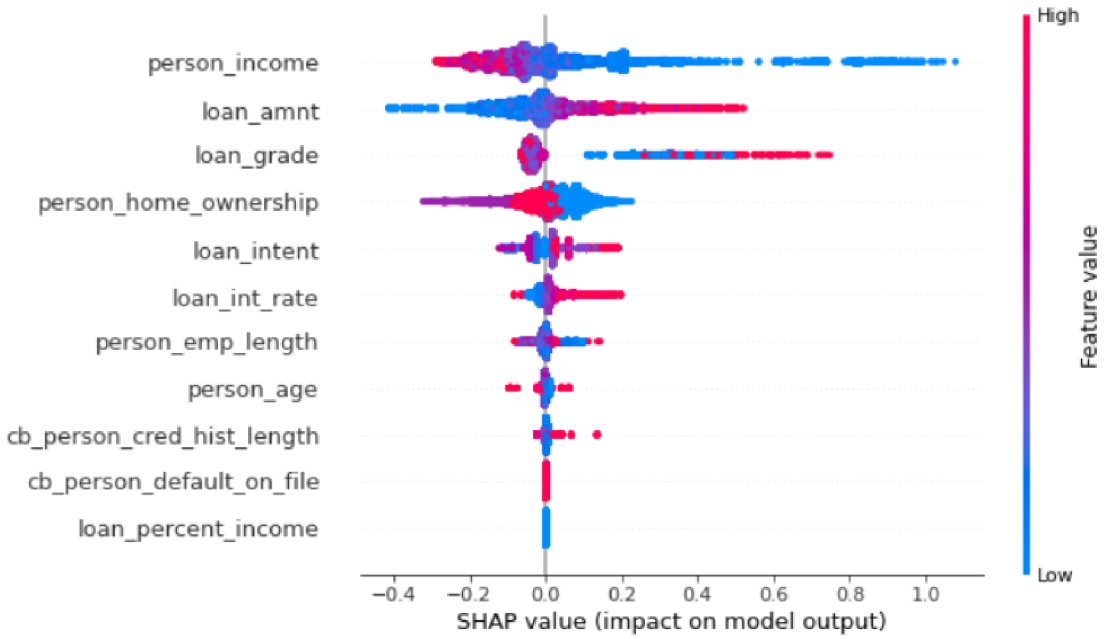
	Precision	Recall	F1 - score	Support
0	0.81	0.98	0.89	5105
1	0.72	0.16	0.27	1412
Accuracy			0.8	6517
Macro avg	0.76	0.57	0.58	6517
Weighted avg	0.79	0.8	0.75	6517

Table 2 shows that the validation accuracy of logistic regression on the loan default data set is 80%. That implies the ratio of correctly predicted instances or samples to all instances is 0.80. In the report, class 0 represents the non-default or good customer, and class 1 represents the default or bad customer. For class 0, the number of instances taken is 5,105 and for class 1, the number of instances taken is 1,412.

The resultant value of precision for class 0 is 0.81 and for class 1 is 0.72, as per the report shown in Table 2. Other than this, the result of the recall for class 0 is 0.98 and for class 1 is 0.16. Furthermore, the results of the F1 score are 0.89 for class 0 and 0.27 for class 1.

Figure 9 shows the distribution of each feature's impact on the performance of the model. The color reflects the meaning of the feature value - red meaning high, blue meaning low. While a high loan amount leads to an increase in the chances of default, a high income reduces the chances of default.

Figure 9: Impact of all features



5.2 Decision Tree

A decision tree basically looks like a structure of a tree or flowchart on which an internal node represents attributes or features of a dataset, the branches of the tree represent a rule for making a decision, and each leaf node indicates the obtained result. The root node is the uppermost node of a decision tree. It trains you to split data according to an attribute's value. Recursive partitioning is a way of recursively splitting the tree. This flowchart-like form aids in decision-making. It is a graphic in the style of a flowchart that strongly matches human thinking. As a result, decision tree algorithms are easy to understand and use.

Hence, a decision tree classifier has been deployed in our present research work to classify defaulters and non-defaulters of loans. A Python programming has been used for the implementation of the decision tree classifier on Google Colab by using the virtual hardware available with in Google Colab, which are 12.69GB of Random Access Memory and 107.72GB of Read-Only Memory. While simulating the decision tree classifier on the loan default dataset, the dataset has been split into 80% for training and 20% for testing.

The decision tree obtained while simulating on the loan default dataset is shown in figure 10.

Figure 10: Decision Tree of Loan default

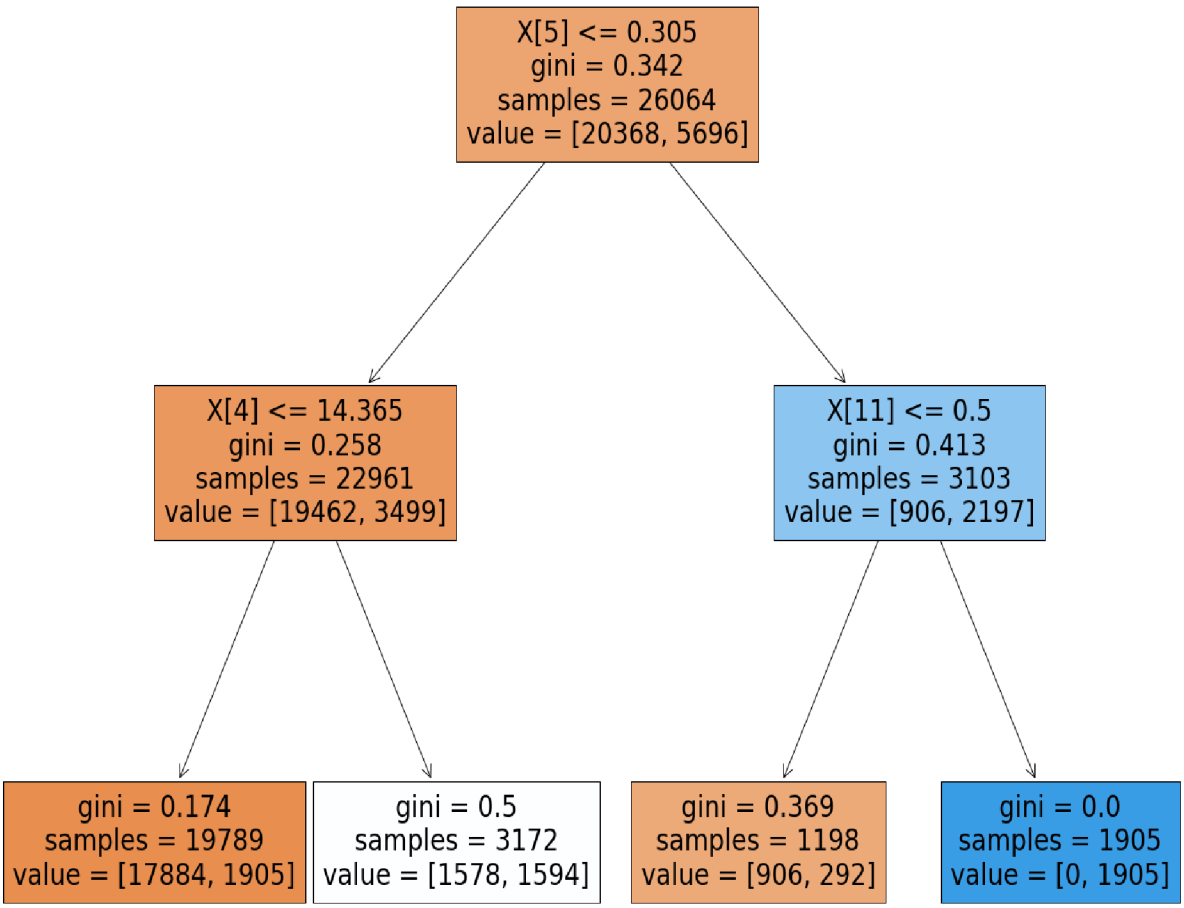


Figure 10 depicts that the decision tree has been created on three features: feature 5, feature 4 and feature 11. The root node of the decision tree has been created in feature 5. The root node starts splitting into sub-nodes at a threshold equal to or greater than 0.3. The total number of samples for both classes on the root node is 26,064, where 20,368 belong to one class and 5,696 belong to the other class. On features 4 and 11, the root node is split into two sub-nodes. The sub node on feature 4 has been created on 22,961 samples, which belong to two different classes, one class with 19,462 samples and the other with 3,499 classes. The threshold value for the feature 4 sub node is equal to or greater than 14.365. On the other hand, the feature 1's sub node has a total of 3,103 samples from two different classes, one class with 906 samples and the other with 2,197 classes. Furthermore, each sub-node has been split into two leaf nodes. The first leaf node of the feature 4's sub node has 19,789 samples from two different classes, the first class with 17,884

samples and the other with 1,905 classes. The second leaf node of the feature 4's sub node has a total of 3,172 samples from two different classes, with 1,578 samples from the first class and 1,594 samples from the second class. Similarly, the feature 11 sub-node has been divided into two leaf nodes. The leaf node of the feature 11 sub node has a total of 1,198 samples from two different classes, with 906 samples from the first class and 292 samples from the second class. The second leaf node of the feature 11's sub node has a total of 1,905 samples that belong only to the second class.

Gini refers to the Gini ratio, which is a measurement of the node's impurity. If all of the node's data belong to the same class, it is said to be the purest; such nodes are known as leaf nodes. The highest Gini ratio value (0.413) is on feature 11 sub node and the lowest (0) is on feature 11 sub node's second leaf node.

The other results obtained for decision tree from the proposed experimentation are presented in detail in the following sub-section.

The confusion matrix obtained by simulating the decision tree on loan default dataset is shown in figure 10.

Figure 11 depicts a confusion matrix in the form of an array. The dimension of this array is 2*2 because the current technique is based on binary classification. It has two classes 0 and 1.

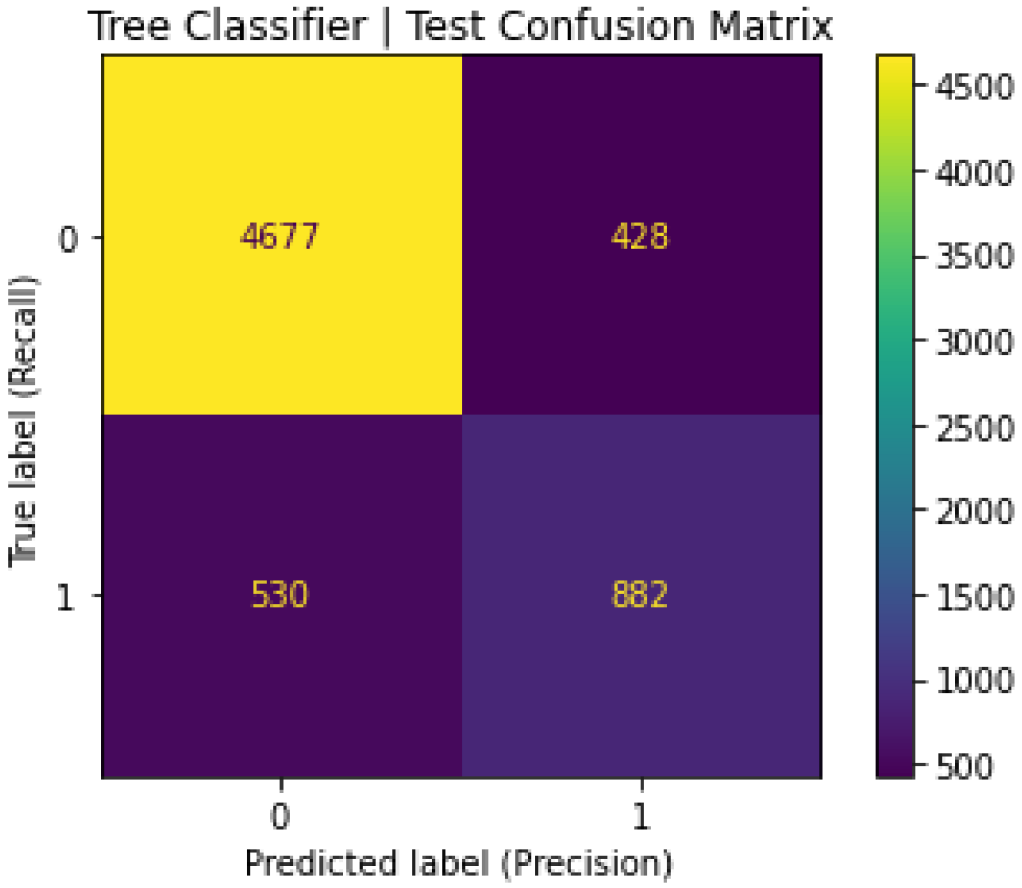
The y axis of the confusion matrix presents accurate or true labels and on the x axis presents predicated labels. All non-diagonal value of confusion matrix is inaccurate predictions from classification and diagonal values are accurate predictions. Actual forecasts are 4,677 and 882, whereas faulty predictions are 428 and 530.

As has been already mentioned, a confusion matrix is composed of four cases. The four cases are: False Negatives, False Positives, True Negatives, and True Positives.

Accordingly, the values of all four cases from the present confusion matrix (Figure 11) are:

False Negatives = 530
False Positives = 428
True Negatives = 4677
True Positives = 882

Figure 11: Decision tree's confusion matrix



The results of other parameters of the confusion matrix known as accuracy, precision, F1 score, and recall are shown in Table 3.

Table 3: Accuracy, precision, F1 score and recall result of decision tree

		Precision	Recall	F1 - score	Support
	0	0.9	0.92	0.91	5105
	1	0.67	0.62	0.65	1412
Accuracy				0.85	6517
Macro avg		0.79	0.77	0.78	6517
Weighted avg		0.85	0.85	0.85	6517

It has been shown in table 3 that the obtained validation accuracy of the decision tree on the loan default data set is 85%, which means the mean ratio of correctly predicted instances or samples to all instances is 0.85.

Class 0 in the report presents the non-default or good customer, and class 1 presents the defaulted or bad customer. As per the report, the number of instances that belong to class 0 is 5,105, and for class 1, the numbers of instances are 1,412.

The obtained precision value for class 0 is 0.90 and for class 1 is 0.67, as per the report shown in figure 10. Where the result of the recall for class 0 is is 0.92 and for class 1 is 0.62. Furthermore, the results of the f1 score are 0.91 for class 0 and 0.65 for class 1.

5.3 XGBoost

XGBoost has been credited as being the main driver for models that have previously won various machine learning challenges. Its performance and efficiency are unrivalled, and it outperforms all the other supervised learning techniques on a regular basis. It can be targeted at binary classification problems or multi-class classification problems. In the present work, it has been deployed to solve a binary classification problem.

In the simulation of the XGBoost classifier on a loan default dataset, the performance measure has been obtained using the model's loss and error. The obtained loss and error rate result of the XGBoost classifier while simulating on the loan default dataset is shown in figure 12 (a) and (b).

From figure 12 (a) and (b), it can be observed that the loss and error rate of XGBoost classifier have been decreased with learning of classifier. In between 0 to 20, both the error and the loss rate have been decreased sharply.

The confusion matrix for XGBoost is shown in figure 13.

In Figure 13, the confusion matrix of the XGBoost classifier has been presented in the form of a 2*2 array. The confusion matrix is based on binary classification and has two classes, 0 and 1.

Along the y axis of the confusion matrix, true labels are taken, and along the x axis, predicated labels are taken. All non-diagonal values of the confusion matrix are inaccurate predictions from classification, while diagonal values are accurate predictions. Actual forecasts are 5,077 and 978, whereas faulty predictions are 48 and 434.

As has been already mentioned, a confusion matrix is composed of four cases. The four cases are: False Negatives, False Positives, True Negatives, and True Positives.

Figure 12: XGBoost classifier

(a) Loss curve

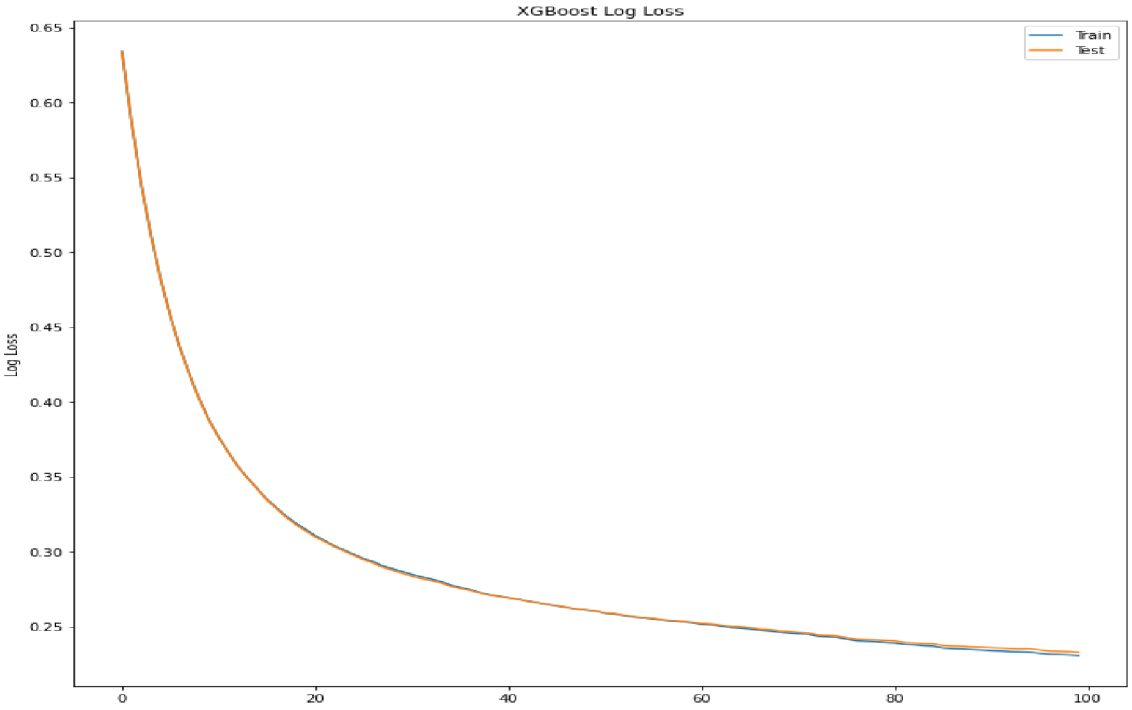


Figure 12: XGBoost classifier

(b) Error Curve

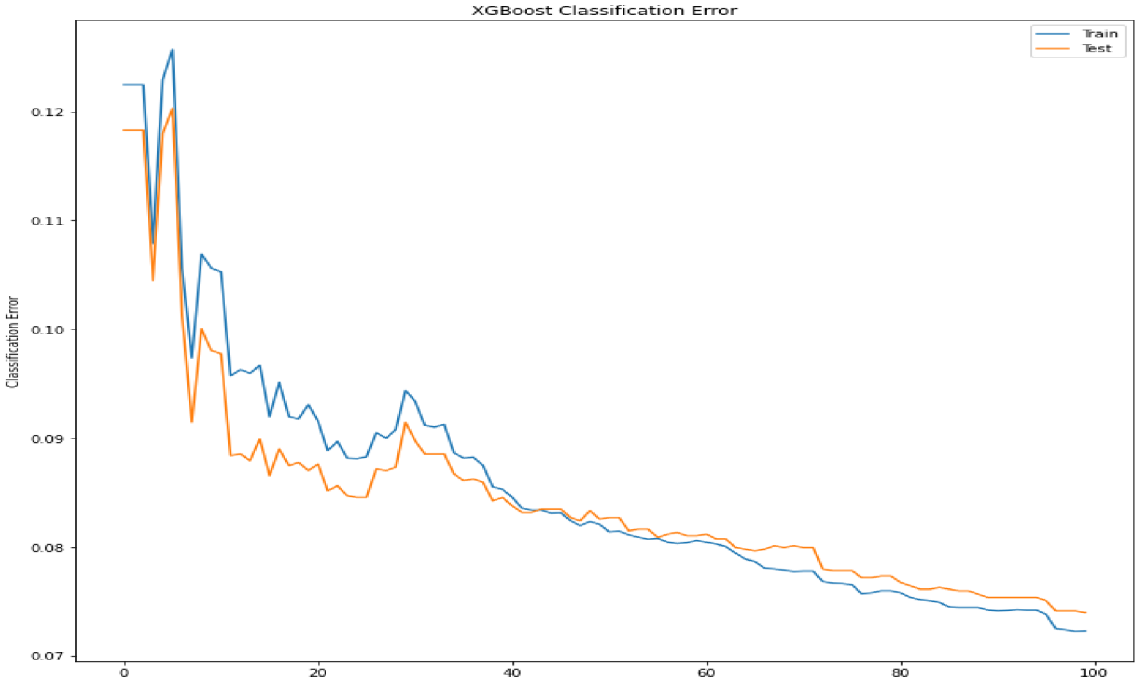
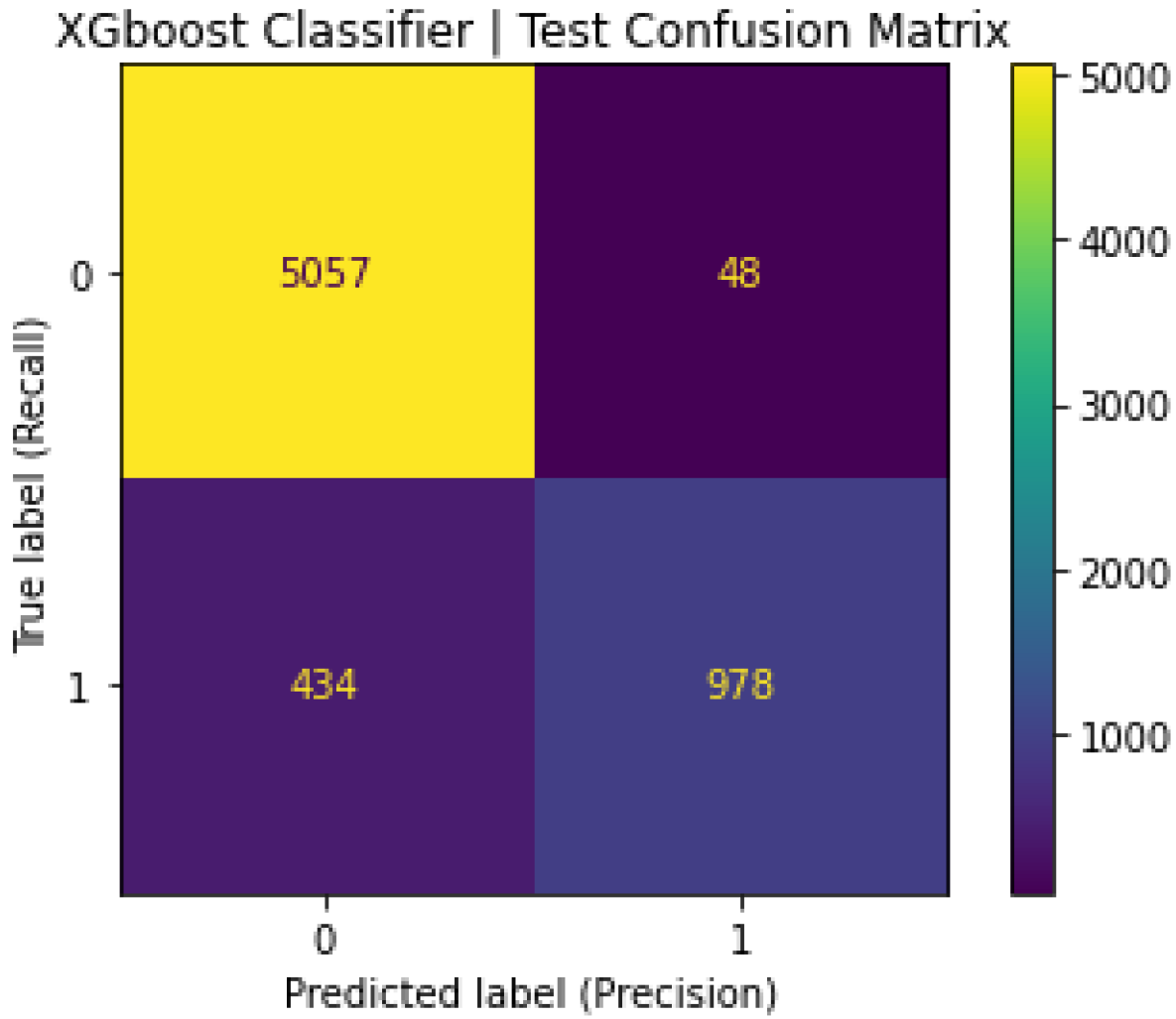


Figure 13: XGBoost's confusion matrix



Accordingly, the values of all four cases from the present confusion matrix (Figure 13) are:

False Negatives = 434
False Positives = 48
True Negatives = 5057
True Positives = 978

The results of other parameters of the confusion matrix known as accuracy, precision, F1 score, and recall are shown in Table 4.

Table 4: Accuracy, precision, F1 score, and recall result of XGBoost classifier

	Precision	Recall	F1 - score	Support
0	0.92	0.99	0.95	5105
1	0.95	0.69	0.8	1412
Accuracy			0.93	6517
Macro avg	0.94	0.84	0.88	6517
Weighted avg	0.93	0.93	0.92	6517

As per Table 4 the resultant validation accuracy of the XGBoost classifier is 93%, which means the mean ratio of correctly predicted instances or samples to all instances is 0.93.

In the report, Class 0 presents the non-default or good customer, and class 1 presents the defaulted or bad customer. As per the report, the number of instances that belong to class 0 is 5,105, and for class 1, the numbers of instances are 1,412.

From the report, it can be seen that precision value for class 0 is 0.92 and for class 1 is 0.95. Where the result of the recall for class 0 is is 0.99 and for class 1 is 0.69. Furthermore, the results of the F1 score are 0.95 for class 0 and 0.80 for class 1.

5.4 Performance Comparison

In this section, all three models (logistic regression, decision tree, and XGBoost) are compared for their potential, limitations, and drawbacks and the default forecasting accuracy and performance of all three models are also compared based on their results obtained on loan default datasets.

A performance comparison between three classification algorithms known as logistic regression, decision tree, and XGBoost has been performed based on their obtained validation accuracy in this section.

Figure 14 represents the comparative analysis between logistic regression, decision tree, and XGBoost based on validation accuracy.

Figure 13 shows that the XGBoost outperforms other classification models in terms of validation accuracy. The validation accuracy for XGBoost is 0.93, or 93%. On the second position, the decision tree gives an accuracy of 0.85 or 85% on the loan of default dataset. The least performance is for logistic regression. It has an overall validation accuracy of 0.8 or 80%.

Figure 13: Comparative analysis based on validation accuracy

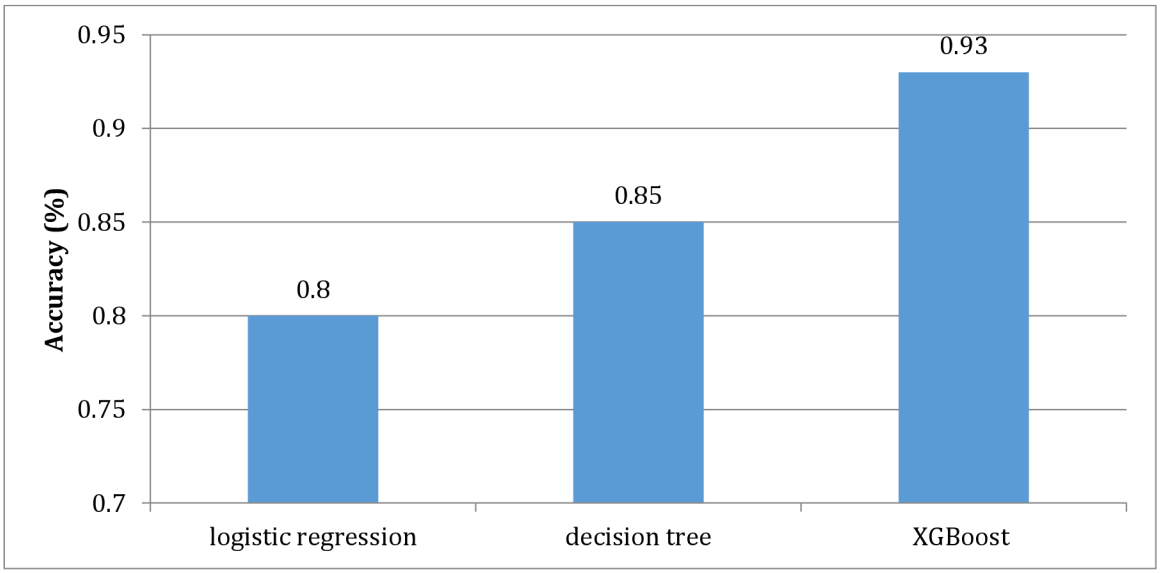
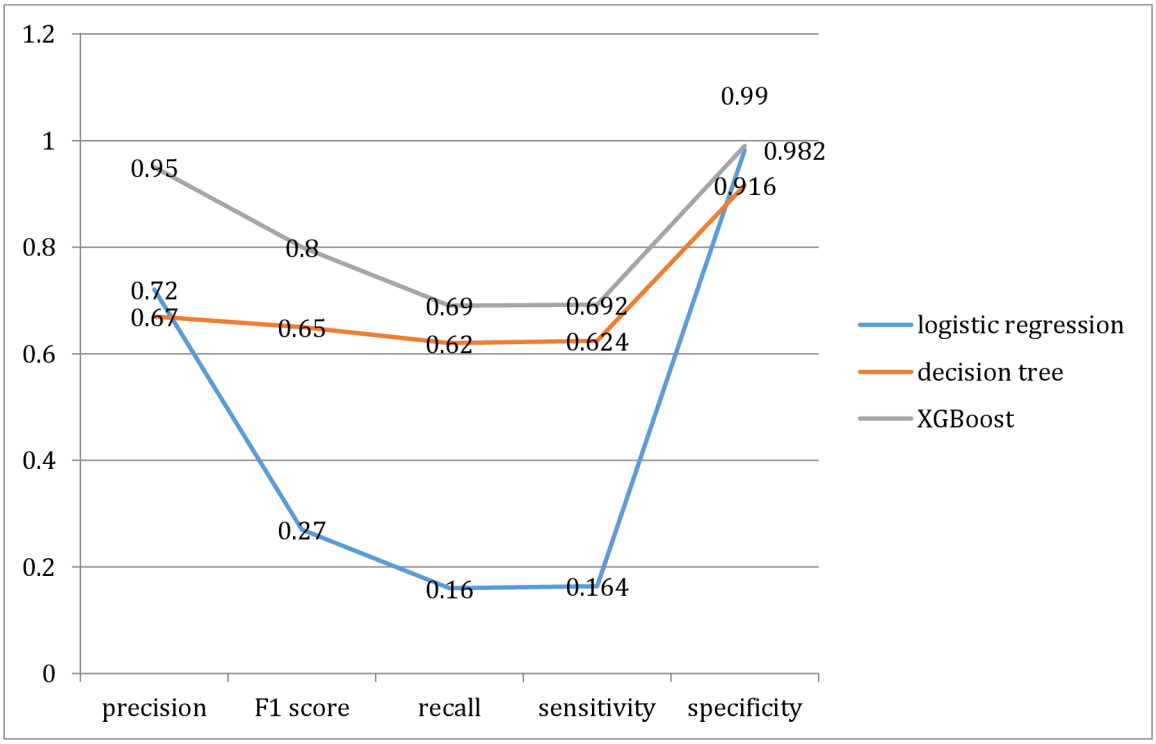


Figure 14: Comparative analysis based on confusion matrix parameter



In this section, performance comparisons have been obtained among three classification algorithms known as logistic regression, decision tree, and XGBoost, based on their confusion matrix parameters. The confusion matrix parameter belongs to precision, F1 score, recall, sensitivity, and specificity.

Figure 15 represents the comparative analysis between logistic regression, decision tree, and XGBoost based on confusion matrix parameter.

As per figure 15, the XGBoost classifier has outperformed the other two classifiers in terms of precision, F1 score, recall, sensitivity, and specificity with values of 0.95, 0.8, 0.66, 0.692, and 0.982, respectively. On the other hand, logistic regression performed the worst, as depicted by its confusion matrix parameters.

Conclusion

As per the obtained results and comparative analysis, out of the three classification models which are logistic regression, decision tree, and XGBoost, the XGBoost classifier has outperformed the other two classifiers on publicly available loan default datasets under peer-to-peer lending in the task of classifying defaulters and non-defaulters of loans. Other than this, the decision tree classifier has been performed, moderately for defaulter classification. The least performance is observed for logistic regression, as it has obtained the least accuracy and the lowest values of confusion matrix parameters like precision, F1 score, recall, sensitivity, and specificity.

For simulation of the three different machine learning classifiers, the loan defaults dataset has been split into a ratio of 80:20, where 80% is for training and 20% is for testing. The implementation of all three different machine learning classifiers has been done using Python programming on Google Colab. The Google Colab provides Python 3 environment for simulating the developed models.

The experimentation has been done successfully in classification of defaulters of loan using machine learning classification techniques. The results of experiment depicted that XGBoost has obtained maximum accuracy of 93% when simulated on loan defaulter dataset. Other than this, it has also obtained highest value of other confusion matrix parameter such as precision (0.95), F1 score (0.8), recall (0.66), sensitivity (0.692), and specificity (0.982).

As mentioned earlier, decision tree has been performed moderately, with accuracy of around 83% and other confusion matrix parameters for that obtained value of around 0.67 (precision), 0.65 (F1 score), 0.62 (recall), 0.624 (sensitivity), and 0.916 (specificity).

Finally, it has been observed that logistic regression has performed worst when compared to the other two classification models. The obtained validation accuracy for logistic regression is 80%, which is less than both the decision tree and the XGBoost classifiers. The values for precision, F1 score, recall, sensitivity, and specificity for logistic regression are 0.72, 0.27, 0.16, 0.164, and 0.982 respectively.

References

Abakarim, Y., Lahby, M. and Attioui, A., (2018) 'An efficient real time model for credit card fraud detection based on deep learning. In Proceedings of the 12th international conference on intelligent systems: theories and applications', pp. 1–7.

Akinyemi, I., (2009) 'Comparative analysis of logistic regression and neural network for credit risk assessment. International Journal of Natural and Applied Sciences', 3(4).

Altman, E. I. (1968) 'Financial Ratios, Discriminant Analysis and The Prediction of Corporate Bankruptcy.' *The Journal of Finance,* 23(4), pp. 589–609.

Altman, E. I., and A. Saunders. (1997) 'Credit risk measurement: Developments over the last 20 years.' *Journal of Banking & Finance,* 21(11-12), pp. 1721–1742.

Altman, E. I., B. Brady, A. Resti, and A. Sironi. (2005) 'The link between default and recovery rates: Theory, empirical evidence, and implications.' *Journal of Business,* 78(6), pp. 2203-2227.

Anderson, Edward J. 2013. 'Business Risk Management: Models and Analysis', Hoboken Wiley.

Anderson, R., (2007) 'The Credit Scoring Toolkit', New York: Oxford University Press Inc.

Anson, Mark J. P., Frank J. Fabozzi, and Moorad Choudhry (2000) 'Credit Derivatives : Instruments, Applications, and Pricing', Hoboken, New Jersey: Wiley.

Atiya, A. F. (2001) 'Bankruptcy prediction for credit risk using neural networks: A survey and new results.' *Ieee Transactions on Neural Networks,* 12(4), pp. 929-935.

Bachmann, A., Becker, A., Buerckner, D., Hilker, M., Kock, F., Lehmann, M., ... & Funk, B. (2011) 'Online peer-to-peer lending-a literature review', *Journal of Internet Banking and Commerce*, 16(2).

Bae, J.K., (2012) 'Predicting financial distress of the South Korean manufacturing industries', *Expert Systems with Applications*, *39*(10), pp. 9159-9165.

Barndorff-Nielsen, O.E. and Shephard, N., (2003) 'Realized power variation and stochastic volatility models', *Journal of Bernoulli*, *9*(2), pp. 243-265.

Batarseh, F., (2020) 'Data Democracy : At the Nexus of Artificial Intelligence, Software Development, and Knowledge Engineering', 1st ed, Elsevier Science, pp. 86-88.

Benzschawel, Terry. (2012) 'Credit Risk Modelling: Facts, Theory and Applications ', London: Risk Books.

Betker, B.L., (1995) 'An empirical examination of prepackaged bankruptcy' *Journal of Financial Management*, pp. 3-18.

Bielecki, Tomasz R, and Marek Rutkowski (2004) 'Credit risk: modelling, valuation and hedging'. Berlin, Germany: Springer.

Black, F. and Cox, J.C., (1976) 'Valuing corporate securities: Some effects of bond indenture provisions', *The Journal of Finance*, 31(2), pp. 351-367.

Black, Fischer, and Myron Scholes. (1972) 'The Valuation of Option Contracts and a test of Market Efficiency', *Journal of Finance*, 27(2), pp: 399-417.

Black, Fischer, and Myron Scholes. (1973) 'The pricing of options and corporate liabilities.' *The journal of political economy*, pp. 637-654.

Blanco, A., Pino-Mejías, R., Lara, J. and Rayo, S., (2013) 'Credit scoring models for the microfinance industry using neural networks: Evidence from Peru'. *Journal of Expert Systems with applications*, 40(1), pp. 356-364.

Boschetti, A. and Massaron, L., (2018) 'Python Data Science Essentials', 3rd ed. Birmingham: Packt Publishing Ltd, pp. 347-350.

Brezigar Masten, A. and Masten, I., (2007) 'Comparison of parametric, semi-parametric and non-parametric methods in bankruptcy prediction. Semi-Parametric and Non-Parametric Methods in Bankruptcy Prediction (March 1, 2007)'.

Briys, E. and De Varenne, F., (1997) 'Valuing risky fixed rate debt: An extension.', *Journal of Financial and Quantitative Analysis*, *32*(2), pp. 239-248.

Cai, S., Lin, X., Xu, D. and Fu, X., (2016) 'Judging online peer-to-peer lending behavior: A comparison of first-time and repeated borrowing requests'. *Journal of Information & Management*, *53*(7), pp. 857-867.

Caouette, John B., Edward I. Altman, and Paul Narayanan (2008) 'Managing Credit Risk : The Great Challenge for Global Financial Markets'. Hoboken, New Jersey: Wiley.

Chaffee, E. C., & Rapp, G. C. (2012) 'Regulating online peer-to-peer lending in the aftermath of Dodd-Frank: In search of an evolving regulatory regime for an evolving industry', *Wash. & Lee L. Rev.*, *69*, 485.

Chakraborty, S. and Sharma, S.K., (2007) 'Prediction of corporate financial health by artificial neural network'. *International Journal of Electronic Finance*, *1*(4), pp. 442-459.

Charalambous, C., Charitou, A. and Kaourou, F., (2000). 'Application of feature extractive algorithm to bankruptcy prediction'. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium* (Vol. 5, pp. 303-308). IEEE.

Chen, N., Ribeiro, B., Vieira, A. and Chen, A., (2013) 'Clustering and visualization of bankruptcy trajectory using self-organizing map'. *Journal of Expert systems with applications*, *40*(1), pp. 385-393.

Couderc, F. and Renault, O., (2005) 'Business and financial indicators: what are the determinants of default probability changes', *Available at SSRN 942236*.

Dastile, X., Celik, T. and Potsane, M., (2020) 'Statistical and machine learning models in credit scoring: A systematic literature survey', *Journal of Applied Soft Computing*, *91*, pp. 106-263.

Davis, M. and Lo, V., (2001) 'Infectious defaults'. *Journal of Quantitative Finance*, *1*(4), pp. 382.

Derelioğlu, G., Gürgen, F. and Okay, N., (2009) 'A neural approach for SME's credit risk analysis in Turkey', In *International Workshop on Machine Learning and Data Mining in Pattern Recognition* (pp. 749-759). Springer, Berlin, Heidelberg.

Doshi-Velez, F., & Kim, B. (2017) 'Towards a rigorous science of interpretable machine learning', *arXiv preprint arXiv:1702.08608*.

Driessen, J., (2005) 'Is default event risk priced in corporate bonds?'. *The Review of Financial Studies*, *18*(1), pp.165-195.

Duan, J.C., (1994) 'Maximum likelihood estimation using price data of the derivative contract. Mathematical Finance', 4(2), pp. 155-167.

Duarte, J., Siegel, S., & Young, L. (2012) 'Trust and credit: The role of appeerance in peer-to-peer lending. The Review of Financial Studies', 25(8),pp. 2455-2484.

Duffie, D. and Singleton, K.J., (1999) 'Modelling term structures of defaultable bonds', *Journal of The review of financial studies*, *12*(4), pp. 687-720.

Duffie, D., and M. Huang. (1996) 'Swap rates and credit quality', *Journal of Finance,* 51(3), pp. 921-949.

Duffie, Darrell. (1999) 'Credit swap valuation', *Journal of Financial Analysts Journal,* 55(1), pp. 73-87.

Elizalde, Abel. (2005) 'Credit risk models II: Structural models'.

Emekter, R., Tu, Y., Jirasakuldech, B., & Lu, M. (2015) 'Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending', *Journal of Applied Economics*, 47(1), pp. 54-70.

Engelmann, B. and Rauhmeier, R., (2011) 'The Basel II risk parameters', 2nd ed. Heidelberg: Springer, pp. 201.

Ferreira, L. E. B., Barddal, J. P., Gomes, H. M., & Enembreck, F. (2017) 'Improving credit risk prediction in online peer-to-peer (p2p) lending using imbalanced learning techniques', In *2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 175-181). IEEE.

Fong, A. (2015) 'Regulation of peer-to-peer lending in Hong Kong: state of play', *Journal of Law and Financial Markets Review*, *9*(4), pp. 251-259.

François, P. and Morellec, E., (2004) 'Capital structure and asset prices: Some effects of bankruptcy procedures', *The Journal of Business*, 77(2), pp. 387-411.

Galai, D., Raviv, A. and Wiener, Z., (2005) 'Liquidation triggers and the valuation of debt and equity', *Journal of Banking and Finance*.

Gao, Y. and Zhang, L., (2018) 'Research on Credit Risk Assessment of Small and Medium-Sized Enterprises in Commercial Banks' OALib, 05(11), pp. 1-11.

Giesecke, K., (2004) 'Credit risk modelling and valuation: An introduction', *Available at SSRN 479323*.

Ha, V.S., Lu, D.N., Choi, G.S., Nguyen, H.N. and Yoon, B., (2019) 'February. Improving credit risk prediction in online peer-to-peer (P2P) lending using feature selection with deep learning', In *2019 21st International Conference on Advanced Communication Technology (ICACT)* (pp. 511-515). IEEE.

Hackbarth, D., Miao, J. and Morellec, E., (2006) 'Capital structure, credit risk, and macroeconomic conditions', *Journal of financial economics*, *82*(3), pp. 519-550.

Helwege, J., (1999) 'How long do junk bonds spend in default?', *The Journal of Finance*, *54*(1), pp. 341-357.

Herrero-Lopez, S. (2009) 'Social interactions in P2P lending. In Proceedings of the 3rd Workshop on Social Network Mining and Analysis' (pp. 1-8).

Herzenstein, M., Dholakia, U. M., & Andrews, R. L. (2011) 'Strategic herding behavior in peer-to-peer loan auctions. Journal of Interactive Marketing', 25(1), pp. 27-36.

Ho, T.S., Stapleton, R.C. and Subrahmanyam, M.G., (1997) 'The Valuation of American Options with Stochastic Interest Rates: A Generalization of the Geske—Johnson Technique', *The Journal of Finance*, *52*(2), pp. 827-840.

Hotchkiss, R.S., Swanson, P.E., Knudson, C.M., Chang, K.C., Cobb, J.P., Osborne, D.F., Zollner, K.M., Buchman, T.G., Korsmeyer, S.J. and Karl, I.E., (1999) 'Overexpression of Bcl-2 in transgenic mice decreases apoptosis and improves survival in sepsis'. *The Journal of Immunology*, *162*(7), pp. 4148-4156.

Huang, F.Y., (2008) 'November. A genetic fuzzy neural network for bankruptcy prediction in chinese corporations'. In *2008 International Conference on Risk Management & Engineering Management* (pp. 542-546). IEEE.

Huang, Z., Chen, H., Hsu, C.J., Chen, W.H. and Wu, S., (2004) 'Credit rating analysis with support vector machines and neural networks: a market comparative study', *Journal of Decision support systems, 37*(4), pp. 543-558.

Hull, J., Nelken, I. and White, A., (2004) 'Merton's model, credit risk, and volatility skews', *Journal of Credit Risk Volume*, *1*(1), p. 05.

Hull, John. (2012) '*Risk Manangement and Financial Institutions*', 3 ed. New York Wiley.

Jarrow, R. A., and F. Yu. (2001) 'Counterparty risk and the pricing of defaultable securities', *Journal of Finance,* 56(5), pp: 1765-1799.

Jarrow, R. A., and S. M. Turnbull. (1995) 'Pricing derivatives on financial securities subject to credit risk', *Journal of Finance,* 50(1), pp: 53-85.

Jarrow, R. A., and S. M. Turnbull. (1995) 'Pricing derivatives on financial securities subject to credit risk', *Journal of Finance*, 50(1), pp: 53-85.

Jarrow, R. A., D. Lando, and S. M. Turnbull. (1997) 'A Markov model for the term structure of credit risk spreads.' *Journal of Review of Financial Studies*, 10(2), pp. 481-523.

Jarrow, R.A. and Yu, F., (2001) 'Counterparty risk and the pricing of defaultable securities', *the Journal of Finance*, *56*(5), pp. 1765-1799.

Jayanthi, J., Suresh Joseph, K. and Vaishnavi, J., (2011) 'Bankruptcy prediction using svm and hybrid svm survey'.

Jiang, C., Song, J., Liu, G., Zheng, L. and Luan, W., (2018) 'Credit card fraud detection: A novel approach using aggregation strategy and feedback mechanism', *IEEE Internet of Things Journal*, *5*(5), pp. 3637-3647.

Jiang, C., Wang, Z., & Zhao, H. (2019) 'A prediction-driven mixture cure model and its application in credit scoring', *European Journal of Operational Research*, *277*(1), pp: 20-31.

Jii Witzany (2019) 'CREDIT RISK MANAGEMENT : pricing, measurement, and modeling'.

Jin, Y. and Zhu, Y., (2015) ' A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending'. In *2015 Fifth International Conference on Communication Systems and Network Technologies* (pp. 609-613). IEEE.

Jin, Y., & Zhu, Y. (2015) 'A data-driven approach to predict default risk of loan for online peer-to-peer (P2P) lending'. In *2015 Fifth International Conference on Communication Systems and Network Technologies* (pp. 609-613). IEEE.

Johnston Ross, E. and Shibut, L., (2022) 'What Drives Loss Given Default? Evidence From Commercial Real Estate Loans at Failed Banks'.

Kast, R. and André Lapied (2006) 'Economics and finance of risk and of the future. Chichester, England', Hoboken, Nj: John Wiley & Sons.

Kijima, M. and Muromachi, Y., (2000) 'Credit events and the valuation of credit derivatives of basket type', *Review of Derivatives Research*, *4*(1), pp. 55-79.

Kijima, M. and Muromachi, Y., (2000) 'Evaluation of credit risk of a portfolio with stochastic interest rate and default processes', *Journal of Risk*, *3*, pp. 5-36.

Kim, B., Khanna, R., & Koyejo, O. O. (2016) 'Examples are not enough, learn to criticize! criticism for interpretability', *Advances in neural information processing systems*, *29*.

Kim, I.J., Ramaswamy, K. and Sundaresan, S., (1993) 'Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model', *Financial management*, pp. 117-131.

Kirby, E., & Worner, S. (2014) 'Crowd-funding: An infant industry growing fast. IOSCO Research Department', pp. 1-62.

Lando, David. (1998) 'On Cox processes and credit risky securities', *Review of Derivatives research* 2, (2-3), pp. 99-120.

Leland, H.E. and Toft, K.B., (1996) 'Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads', *The Journal of Finance*, 51(3), pp. 987-1019.

Leo, M., Sharma, S., & Maddulety, K. (2019) 'Machine learning in banking risk management: A literature review', *Risks*, *7*(1), pp. 29.

Li, H., Sun, J. and Wu, J., (2010) 'Predicting business failure using classification and regression tree: An empirical comparison with popular classical statistical methods and top classification mining methods', *Expert Systems with Applications*, *37*(8), pp. 5895-5904.

Li, W., Ding, S., Chen, Y., & Yang, S. (2018) 'Heterogeneous ensemble for default prediction of peer-to-peer lending in China', *Ieee Access*, *6*, pp. 54396-54406.

Li, Y., Hao, A., Zhang, X. and Xiong, X., (2018) 'Network topology and systemic risk in Peer-to-Peer lending market', *Physica A: Statistical Mechanics and its Applications*, *508*, pp. 118-130.

Lipton, Z. C. (2018) 'The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery', Queue, 16(3), pp. 31-57.

Livingston, L. (2012) 'Could Peer-to-Peer Loans Substitute for Payday Loans?', Accounting & Taxation, 4(2), pp. 77-94.

Longstaff, F. A., J. Pan, L. H. Pedersen, and K. J. Singleton. (2011) 'How Sovereign Is Sovereign Credit Risk?', *Journal of American Economic Journal-Macroeconomics,* 3(2), pp. 75-103. doi: 10.1257/mac.3.2.75.

Longstaff, F.A. and Schwartz, E.S., (1995) 'A simple approach to valuing risky fixed and floating rate debt', *The Journal of Finance*, 50(3), pp. 789-819.

Lundberg, S. M., & Lee, S. I. (2017) 'A unified approach to interpreting model predictions. In *Proceedings of the 31st international conference on neural information processing systems',* pp. 4768-4777.

Malekipirbazari, M. and Aksakalli, V., (2015) 'Risk assessment in social lending via random forests. *Expert Systems with Applications*', *42*(10), pp. 4621-4631.

Malekipirbazari, M., & Aksakalli, V. (2015) 'Risk assessment in social lending via random forests. *Expert Systems with Applications*', *42*(10), pp. 4621-4631.

Mantelaers, E., (2022) *The Financial Credit Risk Assessment Model: Three Perspectives*. [online] Academia.edu. Available at: <https://www.academia.edu/69592121/The_Financial_Credit_Risk_Assessment_Model_Three_Perspectives> [Accessed 6 May 2022].

Mareeswari, V. and Gunasekaran, G., (2016) 'February. Prevention of credit card fraud detection based on HSVM'. In 2016 International Conference on Information Communication and Embedded Systems (ICICES) (pp. 1-4). IEEE.

Merton, R.C., (1974) 'On the pricing of corporate debt: The risk structure of interest rates', *The Journal of finance*, 29(2), pp. 449-470.

Merton, Robert C. (1974) 'On the pricing of corporate debt: the risk structure of interest rates', *The Journal of Finance,* 29, pp. 449-470.

Mestiri, S. and Hamdi, M., (2012) 'Credit Risk Prediction: A comparative study between logistic regression and logistic regression with random effects. International Journal of Management Science and Engineering Management', 7(3), pp. 200-204.

Miller, T. (2019) 'Explanation in artificial intelligence: Insights from the social sciences', *Journal of Artificial intelligence*, *267*, pp. 1-38.

Min, J.H. and Lee, Y.C., (2005) 'Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters', *Journal of Expert systems with applications*, *28*(4), pp. 603-614.

Modigliani, Franco, and Merton H Miller. (1958) 'The cost of capital, corporation finance and the theory of investment', *The American economic review,* 48 (3), pp. 261-297.

Möllenkamp, N. (2017) '*Determinants of loan performance in P2P lending*', (Bachelor's thesis, University of Twente).

Molnar, C., Casalicchio, G., & Bischl, B. (2020) 'Interpretable machine learning–a brief history, state-of-the-art and challenges'. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (pp. 417-431). Springer, Cham.

Moraux, F. and Navatte, P., (2004) 'Extending the Maturity of a defaulting debt: when it is worthwhile!'. *Available at SSRN 492753*.

Moraux, F., (2002) 'Valuing corporate liabilities when the default threshold is not an absorbing barrier', *Available at SSRN 314404*.

Namvar, A., Siami, M., Rabhi, F., & Naderpour, M. (2018) 'Credit risk prediction in an imbalanced social lending environment', *arXiv preprint arXiv:1805.00801*.

Parlour, Christine A., and Andrew Winton. (2013) 'Laying off credit risk: Loan sales versus credit default swaps', *Journal of Financial Economics,* 107, pp. 25-45.

Pierrakis, Y., & Collins, L. (2013) 'Banking on Each Other: Peer-to-peer lending to business: evidence from funding circle'.

Piramuthu, S., (1999) 'Financial credit-risk evaluation with neural and neurofuzzy systems', *European Journal of Operational Research,* 112(2), pp. 310-321.

Rafiei, F.M., Manzari, S.M. and Bostanian, S., (2011) 'Financial health prediction models using artificial neural networks, genetic algorithm and multivariate discriminant analysis: Iranian evidence', *Journal of Expert Systems with Applications*, *38*(8), pp.10210-10217.

Ribeiro, B., Silva, C., Chen, N., Vieira, A. and das Neves, J.C., (2012) 'Enhanced default risk models with SVM+', *Journal of Expert Systems with Applications*, *39*(11), pp. 10140-10152.

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). ' Why should i trust you?' Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

Ronn, E.I. and Verma, A.K., (1986) 'Pricing risk-adjusted deposit insurance: An option-based model', *The Journal of Finance*, *41*(4), pp. 871-895.

Roure, Calebe de ; Pelizzon, Loriana ; Tasca, Paolo., (2006) 'How does P2P lending fit into the consumer credit market?', Deutsche Bundesbank.

Saunders, A. , and Marcia M. Cornett. (2011) 'Financial Institutions Management: A Risk Management Approach', 7 ed. New York: McGraw Hill.

Saunders, A., and Marcia M. Cornett. (2011) 'Financial Institutions Management: A Risk Management Approach', 7 ed. New York: McGraw Hill.

Schönbucher, P.J., (2000) 'Factor models for portofolio credit risk', (No. 16/2001). Bonn Econ Discussion Papers.

Serrano-Cinca, C. and Gutiérrez-Nieto, B., (2016) 'The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending', *Journal of Decision Support Systems*, *89*, pp. 113-122.

Serrano-Cinca, C., & Gutiérrez-Nieto, B. (2016) 'The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending', *Journal of Decision Support Systems*, *89*, pp. 113-122.

Wade, C., (2020) 'Hands-On Gradient Boosting with XGBoost and scikit-learn', Packt Publishing.

Wang, Z., Jiang, C., Ding, Y., Lyu, X. and Liu, Y., (2018) 'A novel behavioral scoring model for estimating probability of default over time in peer-to-peer lending', *Journal of Electronic Commerce Research and Applications*, *27*, pp. 74-82.

Wang, Z., Jiang, C., Ding, Y., Lyu, X., & Liu, Y. (2018) 'A novel behavioral scoring model for estimating probability of default over time in peer-to-peer lending', *Journal of Electronic Commerce Research and Applications*, *27*, pp. 74-82.

Yang, Z., You, W. and Ji, G., (2011) 'Using partial least squares and support vector machines for bankruptcy prediction', *Journal of Expert Systems with Applications*, *38*(7), pp. 8336-8342.

Yu, F., (2002) 'Correlated defaults in reduced-form models', *Workingpaper, University of California, Irvine*.

Yuan, Y. and Zhang, J.J., (2003) 'Towards an appropriate business model for m-commerce. *International journal of mobile communications*', *1*(1-2), pp. 35-56.

Yuan, Z., Wang, Z., & Xu, H. (2017) 'Credit Risk Assessment of Peer-to-Peer Lending Borrower Utilizing BP Neural Network', In *International Conference on Emerging Internetworking, Data & Web Technologies* (pp. 22-33). Springer, Cham.

Zhan, Q. and Yin, H., (2018) 'A loan application fraud detection method based on knowledge graph and neural network', In *Proceedings of the 2nd International Conference on Innovation in Artificial Intelligence* (pp. 111-115).

Zhang, Y., Wang, D., Chen, Y., Zhao, Y., Shao, P., & Meng, Q. (2017) 'Credit risk assessment based on flexible neural tree model', In *International Symposium on Neural Networks* (pp. 215-222). Springer, Cham.

Zhen, W. and Wenjuan, S., (2016) 'December. Commercial bank credit risk assessment method based on improved svm', In 2016 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS) (pp. 353-356). IEEE.

Zhou, J. and Bai, T., (2008) 'Credit risk assessment using rough set theory and GA-based SVM', In *2008 The 3rd International Conference on Grid and Pervasive Computing-Workshops* (pp. 320-325). IEEE.

Appendix

Appendix 1 – Codes for Python

```
import pandas as pd
df=pd.read_csv("loansData.csv")
df.shape
df.dtypes
df.isnull().sum()
meanvalue = df["person_emp_length"].mean()
abmean_value =int(meanvalue)
abmean_value
df["person_emp_length"].fillna(value=abmean_value, inplace=True)
mean_value = df["loan_int_rate"].mean()
ablmean_value =int(mean_value)
ablmean_value
df["loan_int_rate"].fillna(value=ablmean_value, inplace=True)
df.isnull().sum()
k=0;
for i in df['person_home_ownership'].unique():
    df=df.replace(to_replace =i,
                  value =k)
    k=k+1
k=0;
for i in df['loan_intent'].unique():
    df=df.replace(to_replace =i,
                  value =k)
    k=k+1
k=0;
for i in df['loan_grade'].unique():
    df=df.replace(to_replace =i,
                  value =k)
    k=k+1
k=0;
for i in df['cb_person_default_on_file'].unique():
    df=df.replace(to_replace =i,
                  value =k)
    k=k+1
df1=df.astype(int)
target=df1["loan_status"]
```

```
final_df=df1.drop(columns=["loan_status"])
final_df
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(final_df,target, tes
t_size = 0.2, random_state = 0)
```

Logistic Regression

```
from sklearn.metrics import classification_report
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score
from sklearn.metrics import confusion_matrix
import seaborn as sns
logreg = LogisticRegression()
logreg.fit(X_train, y_train)
y_pred = logreg.predict(X_test)
print(classification_report(y_test, y_pred))
na=accuracy_score(y_test, y_pred)
print("Test accuracy of Logistic regression:",na)
log= confusion_matrix(y_test, y_pred)
sns.heatmap(log, annot=True, cmap='Blues').set(title="Logistic regressio
n confusion matrix")
from sklearn import metrics
import matplotlib.pyplot as plt
fp, tp,_ = metrics.roc_curve(y_test, y_pred)
auc = metrics.roc_auc_score(y_test, y_pred)
plt.plot(fp,tp,label="AUC="+str(auc))
plt.ylabel('True Positive Rate')
plt.xlabel('False Positive Rate')
plt.title("Random Forest test ROC curve")
plt.legend(loc=4)
plt.show()
```

#Decision Tree

```
from sklearn import tree
dlf = tree.DecisionTreeClassifier()
dlf = dlf.fit(X_train, y_train)
predictions = dlf.predict(X_test)
print(classification_report(y_test, predictions))
nad=accuracy_score(y_test, predictions)
print("Test accuracy of Decision tree:",nad)
ranf= confusion_matrix(y_test, predictions)
```

```
sns.heatmap(ranf, annot=True, cmap='Blues').set(title="Decision tree con
fusion matrix")
fpr, tpr,_ = metrics.roc_curve(y_test,predictions)
auc = metrics.roc_auc_score(y_test, predictions)
plt.plot(fpr,tpr,label="AUC="+str(auc))
plt.ylabel('True Positive Rate')
plt.xlabel('False Positive Rate')
plt.title("Decision Tree ROC curve")
plt.legend(loc=4)
plt.show()
```

#XGboost

```
import xgboost as xgb
xg_reg = xgb.XGBRegressor(objective ='reg:linear', colsample_bytree = 0.
3, learning_rate = 0.1,
                max_depth = 5, alpha = 10, n_estimators = 10,early_stopp
ing=True)
eval_set = [(X_train, y_train), (X_test, y_test)]
xg_reg.fit(X_train,y_train,eval_metric=["error", "logloss"], eval_set=ev
al_set, verbose=False)
preds = xg_reg.predict(X_test)
print(classification_report(y_test, preds.round()))
xgd=accuracy_score(y_test, preds.round())
print("Test accuracy of XgBoost:",xgd)
xgc= confusion_matrix(y_test, preds.round())
sns.heatmap(xgc, annot=True, cmap='Blues').set(title="XgBoost confusion
matrix")
fprx, tprx,_ = metrics.roc_curve(y_test,preds.round())
auc = metrics.roc_auc_score(y_test, preds.round())
plt.plot(fprx,tprx,label="AUC="+str(auc))
plt.ylabel('True Positive Rate')
plt.xlabel('False Positive Rate')
plt.title("XgBoost ROC curve")
plt.legend(loc=4)
plt.show()
results = xg_reg.evals_result()
epochs = len(results['validation_0']['error'])
x_axis = range(0, epochs)
fig, ax = plt.subplots(figsize=(12,12))
ax.plot(x_axis, results['validation_0']['logloss'], label='Train')
ax.plot(x_axis, results['validation_1']['logloss'], label='Test')
```

```
ax.legend()
plt.ylabel('Log Loss')
plt.title('XGBoost Log Loss')
plt.show()
fig, ax = plt.subplots(figsize=(12,12))
ax.plot(x_axis, results['validation_0']['error'], label='Train')
ax.plot(x_axis, results['validation_1']['error'], label='Test')
ax.legend()
plt.ylabel('Classification Error')
plt.title('XGBoost Classification Error')
plt.show()
x={"Logistic Regression":0.80,"Decision Tree":0.85,"XgBoost":0.90}
names = list(x.keys())
values = list(x.values())
fig, ax = plt.subplots(figsize=(8, 5))

bars = ax.barh(names, values)
ax.bar_label(bars)
plt.xlabel("Classifiers")
plt.ylabel("Accuracy")
plt.title("Comparative analysis based on validation accuracy")
plt.show()
ax = df1.plot(x='parameters', y='logisticRegression', marker='o')
df2.plot(ax=ax, x='parameters1', y='Decisiontree', marker='o')
df3.plot(ax=ax, x='parameters2', y='XgBoost', marker='o')
plt.xlabel("Performance parameters")
plt.ylabel("value")
plt.title("Comparative analysis based on confusion matrix parameter")
plt.show()
```

#Shapley for Logistic regression

```
import shap
import sklearn
X100 = shap.utils.sample(final_df, 100)
model = sklearn.linear_model.LinearRegression()
model.fit(final_df,target)
print("Model coefficients:\n")
for i in range(final_df.shape[1]):
    print(final_df.columns[i], "=", model.coef_[i].round(4))
shap.plots.partial_dependence(
    "loan_int_rate", model.predict, X100, ice=False,
```

```
    model_expected_value=True, feature_expected_value=True
)
explainer = shap.Explainer(model.predict, X100)
shap_values = explainer(final_df)

# make a standard partial dependence plot
sample_ind = 18
shap.partial_dependence_plot(
    "loan_int_rate", model.predict, X100, model_expected_value=True,
    feature_expected_value=True, ice=False,
    shap_values=shap_values[sample_ind:sample_ind+1,:]
)
shap.plots.scatter(shap_values[:,"loan_int_rate"])
shap.plots.waterfall(shap_values[sample_ind], max_display=14)
shap.plots.beeswarm(shap_values_ebm, max_display=14)
```