Ajay Byanjankar

# Predicting Risk and Return in Peer-to-Peer Lending with Machine Learning

A Decision Making Approach

Åbo Akademi University

# Ajay Byanjankar

Born 1985

## Previous studies and degrees

Master of Science (Economics and Business Administration), Åbo Akademi University, 2015

Bachelor of Business Administration, Kemi-Tornio University of Applied Sciences, 2012


# Predicting Risk and Return in Peer-to-Peer Lending with Machine Learning

A Decision Making Approach

Ajay Byanjankar

Information Systems
Faculty of Social Sciences, Business and Economics
Åbo Akademi University, 2021

## Supervisors


| Docent Markku Heikkilä | Docent Jozsef Mezei | Anna Sell |
|---|---|---|
| University Teacher | Associate Professor | University Lecturer |

Faculty of Social Sciences, Business and Economics, Information Systems
Åbo Akademi University
ASA,Vänrikinkatu 3 B, 20500, Turku
Finland


## Reviewers


Professor Pasi Luukka
School of Business
LUT University
Yliopistonkatu 34, 53850 Lappeenranta
Finland

Docent Tomas Eklund
Senior lecturer, Associate Professor
Department of Informatics and Media
Uppsala University
Box 513, SE-751 20 Uppsala
Sweden


## Opponent


Professor Pasi Luukka
School of Business
LUT University
Yliopistonkatu 34, 53850 Lappeenranta
Finland



The printed version: ISBN 978-952-12-4128-4
The digital version: ISBN 978-952-12-4129-1
Painosalama Oy, Turku, Finlnad, 2021

# Acknowledgments

I am very grateful for all the support and guidance I have received throughout the writing of my dissertation. It is my great pleasure to thank all the people that have helped me in this process.

I would first like to thank my supervisors Markku Heikkilä, Jozsef Mezei, and Anna Sell for all the support and guidance that they have provided. I am sincerely thankful for the motivation that they provided me to pursue my doctoral studies after working with them in my Master's thesis. They have been a great mentor in guiding me to perform proper research and have given me the freedom to perform independent research. Besides the academics, they have been available all the time for any kind of discussions to keep me motivated and provide support. I would also like to especially thank Jozsef Mezei for helping me out with the technical needs in completing my articles, and for always providing new ideas.

I am very thankful that I got to know and work with Markus Viljanen in one of my articles. I have enjoyed the research related discussions that we had and they have been very productive for my PhD studies. He has helped me with the ideas and guided me to apply the 'Survival Analysis' method for the article. I would like to thank our Department head Prof. Anssi Öörni for providing an excellent research environment and being supportive(financial support for attending the conferences). I would also like to thank Xiaolu Wang for being there to lend suggestions and have discussions at many times.

I am also thankful to all the teaching members of the Information Systems Department for being supportive throughout my PhD studies. I would like to extend my gratitude to Susann Brännkärr and Tina Sigfridsson for helping me with the administrative works related to travel and course purchases.

I am extremely grateful to my family and friends, who have provided me with great motivation and been supportive in my PhD career. They have been there to listen to all my success stories and most importantly my frustrations and struggle.

Finally, I would like to sincerely thank the Faculty of Social Science, Business and Economics for selecting me for the fully-funded doctoral position. Also, I am thankful for the grants I received from Liikesivistysrahasto, Sparsbanken, and Turun Kauppaopetussäätiö.

# Abstract

Credit risk is one of the prominent risk types in the financial industry. It is the risk associated with lending money to borrowers, where the risk is the likelihood of not receiving back the money as a result of borrowers defaulting. Therefore, credit risk is directly related to the loss of investment, which makes credit risk a significant risk in terms of potential losses. While the loss from credit risk can be extreme, assessing the credit risk is a difficult task, and it receives high importance in decision making. With the growth of the credit industry, it also increases the likelihood of increased defaulted loans that further increases the need for careful selection of borrowers in providing credits. To prevent the loss from providing credits to risky borrowers, credit risk evaluation plays a critical role in differentiating between 'low-risk' and 'high-risk' borrowers.

An alternative financial market with easy and quick access to loans has been emerging as a popular alternative to the traditional market. Peer-to-peer lending is one such popular market that connects borrowers directly to lenders through an online platform for loan transactions. With the absence of collateral and financial intermediaries, peer-to-peer lending provides an easy access to credits at a lower cost. Additionally, its online and automated operation allows for quick access to credits. However, the absence of collateral also becomes the source of credit risk. The risk is further increased due to most lenders being non-professional investors and thus, lack analytical skills. Therefore, it requires careful selection of borrowers to prevent loss in presence of high risk in peer-to-peer lending.

The objective of this thesis is to study credit risk evaluation in peer-to-peer lending for supporting lending decisions. With credit scoring as the statistical tool for evaluating credit risk, the primary aim of the thesis is to apply predictive analytics for estimating credit risk. Machine learning algorithms are implemented to create more accurate credit scoring models with high predictive performance.

Implementing multiple approaches to analyzing credit risk in peer-to-peer lending, this thesis attempts to generate solutions for better risk identification to support lending decisions. For a more realistic estimation of return from peer-to-peer loans, return is estimated by accounting risk that ensures for profitable investments in the presence of risk to the lenders. Due to the presence of different risk levels in peer-to-peer lending, credit risk modeling is performed to create risk-specific decisions. The risk evaluation performed at a group level contributes to more accurate risk identi-

fication and lower misclassification costs in differentiation between 'low-risk' and 'high-risk' borrowers. With portfolio optimization, lenders' need for budget allocation in ensuring overall profit is achieved. Portfolio optimization supports lenders in loan selection with budget allocation to achieve a high return by accepting a certain level of risk.

# Sammanfattning

Kreditrisk är en framträdande risktyp i den finansiella industrin och beskriver risken associerad till utlåning där det finns en viss sannolikhet att de lånade medlen inte fås tillbaka efter att låntagaren har försummat sina avbetalningar. Eftersom kreditrisken är direkt anknuten till långivares investering, blir den en signifikant risk med hänsyn till potentiella förluster. Förlusten orsakad av kreditrisk kan bli ytterst stor vilket gör riskbedömningen ett svårt men synnerligen betydelsefullt steg i beslutsfattandet. I och med att kreditindustrin växer, växer också sannolikheten av försummade lån och låntagarnas behov av att kunna bättre välja till vilka låntagare de kan bevilja kredit. Evalueringen av kreditrisk spelar en kritisk roll i att skilja åt de potentiella låntagarna till ”lågrisk” och ”högrisk” för att kunna förebygga förluster. Den alternativa finansiella marknaden har blivit ett populärt alternativ till den traditionella finansiella marknaden genom att erbjuda lätt och snabb tillgång till lån. Här är P2P-utlåning, d.v.s. utlåning från långivare direkt till låntagare utan mellanhand (en. peer-to-peer, P2P) genom en webbservice, en viktig nyare serviceform. I och med att man inte kräver realsäkerhet eller finansiella mellanhänder erbjuder P2P-utlåning en lätt tillgång till krediter med en lägre kostnad. Utöver detta erbjuder automatiserade webbtjänsterna även snabbare tillgång till kredit. Å andra sidan ökar frånvaro av säkerheten kreditrisken och risken blir ännu större i och med att de flesta långivarna inte är professionella investerare och ofta saknar analytiska färdigheter. Den höga risken i P2P-utlåning betyder att långivare har ett stort behov av att kunna välja omsorgsfullt mellan de potentiella låntagarna. Avsikten i denna avhandling är att forska i evalueringen av kreditrisk och erbjuda stöd till beslutet att bevilja P2P-lån. Kreditbedömning (en. credit scoring) är metoden som används i den statistiska evalueringen av kreditrisk och det primära målet i avhandlingen är att tillämpa prediktiva analytiska metoder i estimeringen av kreditrisk. Maskininlärningsmetoder tillämpas med avsikten att skapa modeller med hög prediktiv prestanda för mer exakt kreditbedömning. Efter att ha tillämpat diverse metoder att analysera kreditrisk erbjuder avhandlingen lösningar som möjliggör bättre identifiering av risker och bättre stöd till lånebeslut. För att man skall kunna estimera en realistisk avkastning för investeringar i P2P-lån ska man beakta långivarnas kreditrisk på den risknivå som garanterar lönsamma investeringar. Man kan hitta olika risknivåer i P2P-lån och kreditriskmodeller utvecklas för att stödja riskspecifika beslut. När riskevaluering tillämpas skilt i grupper av lånekontrakt med olika nivåer av risk

får man som resultat exaktare identifiering av riskerna och mindre felklassificeringskostnader när man skiljer mellan lågrisk och högrisk låntagare. Med optimeringen av låneportföljer åstadkommer man en lönsam allokering av medlen i långivarnas investeringsbudget. Med portföljoptimeringen erhåller långivarna ett urval av lånekontrakt som de kan finansiera med sin investeringsbudget och som uppnår hög avkastningsnivå med en viss nivå av risk.

# List of Original Publications

**Paper 1**

Byanjankar A., Viljanen M. (2020) Predicting Expected Profit in Ongoing Peer-to-Peer Loans with Survival Analysis-Based Profit Scoring. In: Czarnowski I., Howlett R., Jain L. (eds) Intelligent Decision Technologies 2019. Smart Innovation, Systems and Technologies, vol 142. Springer, Singapore.
https://doi.org/10.1007/978-981-13-8311-3_2

**Paper 2**

Byanjankar, Ajay. "Improving Credit Risk Analysis with Cluster Based Modeling and Threshold Selection." Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020.

**Paper 3**

Byanjankar A., Mezei J., Wang X. (2020) Analyzing Peer-to-Peer Lending Secondary Market: What Determines the Successful Trade of a Loan Note?. In: Rocha Á., Adeli H., Reis L., Costanzo S., Orovic I., Moreira F. (eds) Trends and Innovations in Information Systems and Technologies. WorldCIST 2020. Advances in Intelligent Systems and Computing, vol 1160. Springer, Cham.

**Paper 4**

Byanjankar, A, Mezei, J, Heikkilä, M. Data-driven optimization of peer-to-peer lending portfolios based on the expected value framework. Intell Sys Acc Fin Mgmt. 2021; 1– 11.
https://doi.org/10.1002/isaf.1490

# Contents

# List of Figures

# List of Tables

# 1. Introduction

## 1.1. Credit Risk

Credit risk has been prevalent in banking history as a principal and a critical risk type. It remains the single most significant risk in terms of potential losses that is difficult to manage, and as a result, has been a widely studied topic in the financial industry [3, 97, 121]. The term 'credit' is referred to as the money being lent by a financial institution that needs to be repaid along with interest in a given time frame, in regular installments [43]. Everyone who borrows credit may not necessarily have the capability to repay the credit; this generates a risk for financial institutions associated with lending [88]. Credit risk implies the risk that a borrower will default due to failure to fulfill the debt's obligations. It usually arises when a counterpart is unable to pay the debt on time [121]. Some common reasons for default include a weak financial situation of the borrower, high debt burden, low and unstable income, fraudulent cases, and flaws in the information system and technology resulting in 'technical defaults' [121]. The most common practice to define a default event is a payment delay of three consecutive payments. The loss from the defaults of a small number of customers may be significant for financial institutions [121, 117].

The three major risks identified in credit risk are default risk,loss risk and exposure risk [121].

- **Default Risk** : It is the risk that describes the probability that a default event will occur, and is therefore termed as probability of default ($PD$). The probability is expressed as a value between 0 and 1 [121].
- **Loss Risk**: In the case of a default event, loss risk determines the loss as a portion of exposure that is unrecovered. It is commonly termed as loss given default ($LGD$). The value of LGD is zero, when there is no loss and the value is 100%, when the full exposure amount is lost. The LGD value is not a fixed parameter and varies for default events [121].
- **Exposure Risk** : Exposure risk is the uncertainty concerning the amount that is at risk at the time of a future default event. It is termed as the exposure at default ($EAD$) [121].

While the later two risks, loss and exposure risk are derived in the case of default event, default risk is estimated as the probability that a default

event occurs. Default risk is very frequently used as interchangeable with credit risk in describing the risk associated with granting a credit loan [121, 60].

The increase in the number of credit loans most likely increases the number of defaulted loans. Thus, it becomes essential for financial institutions to differentiate between 'good' and 'bad' applicants before granting credit [124]. The loss from credit risk probing to be extreme, its assessment becomes critical, and therefore the need for credit risk management arises. Credit risk management is a technique that is applied in administering the risks related to credit. It involves sequential steps that include recognizing possible risks, assessing the risks, the appropriate treatment, and deploying the risk models.

Credit risk evaluation is the key to financial success in the lending industry, where selecting the right customers becomes important. Failing to evaluate the credit risk and making wrong decisions increases the likelihood of heavy financial loss to the financial institutions [12]. The potential default applicants can be identified by estimating the probability that an applicant will default using the information received at the time of application, which can be used as a basis for accepting or rejecting the loan application [43]. To avoid payment defaults, the development of credit risk decision support model is necessary to enhance the evaluation process with fast and accurate decisions. Successful credit risk management requires efficient tools and techniques for risk measurement [12, 63, 121].

## 1.2. Peer to Peer Lending as the Credit Market

Alternative lending, also known as marketplace lending, is an alternate lending mechanism through online platforms [1]. The emergence of Web 2.0 has paved the way for online market creation with convenient accessibility, and strong collaboration [32]. The focus of alternative lending is on providing access to credit to small businesses and borrowers that struggle to gain credit from traditional financial services[2]. When it comes to 'micro' credit, banks do not find it profitable to offer them as part of their service portfolio. In addition, small borrowers lack collateral and good credit history, which makes it difficult to gain credit from banks [4]. Hence, alternative lending offers an alternate solution to small borrowers that connects them

[1] https://www.morganstanley.com/im/en-us/financial-advisor/insights/investment-insights/an-introduction-to-alternative-lending.html

[2] https://www.businessinsider.com/alternative-lending-nonbank-industry?r=US&IR=T

to investors through an online platform.

As an online marketplace, Peer-to-Peer (P2P) lending creates an environment for lenders and borrowers to meet virtually to conduct loan transactions [32]. It provides micro-finance services by helping to match lenders and borrowers through an online platform, facilitated by a P2P lending service provider [6]. Compared to traditional financial services, there are some significant differences in P2P lending from both lender's and borrower's perspectives. The main highlights of P2P lending, in contrast to traditional financial services, are the absence of expensive financial intermediaries and collateral. In addition, all the services are operated online, which allows for rapid automated processing that ensures easy and quick access to loans at a lower cost [56, 39]. These attributes are the primary source of attraction to borrowers who face difficulties in accessing loans from traditional financial service providers [109, 50].

Loan transactions in P2P lending generally include small to medium loan amounts for a short time period, where the lenders are usually private individuals [39]. P2P has introduced new dynamics in the microfinance industry, with the focus being on small-scale borrowers, such as individual borrowers and small firms [50]. However, the use of P2P loans is mainly seen to be as complementary to credit cards and not as replacement for bank debts [39]. With quick and easy access to loans being main benefits to borrowers, lenders are attracted to P2P lending due to higher return advertised to them compared to similar traditional investments, such as bank deposits [86, 81]. Hence, considering the benefits to both parties, P2P lending has seen increasing popularity and growth in recent years [62].

The fascinating growth of P2P lending, however, is characterized by the presence of high credit risk. Lack of collateral is the primary source of credit risk in P2P lending. The lenders in P2P, being individual investors, who are mostly non-professionals, has an adverse effect on credit evaluation, leading to credit risk [56, 67, 86]. Information asymmetry is seen as another source of credit risk in P2P lending, as it is typically sharper than in case of traditional financial services. As a result of information asymmetry, the risk of being a fraud victim is higher [133]. Information asymmetry as a term describes the situation where there is unequal information distribution during a transaction. In P2P lending, the situation arises when borrowers do not accurately provide the information[3]. Hence, high growth and high credit risk make P2P lending an attractive credit market for studying credit risk. In addition, credit risk evaluation can be seen as an effective

[3] https://corporatefinanceinstitute.com/resources/knowledge/finance/asymmetric-information/

tool for guiding unprofessional investors in P2P lending, allowing them to make more informed and rational decisions [42].

## 1.3. Credit Risk Evaluation

Credit risk evaluation is one of the key processes in the financial industry for credit management decisions. It involves collecting, analyzing, and classifying different credit elements and variables as the basis of credit decisions [1]. Credit risk is evaluated and estimated based on the borrowers' capability of paying back the credit [60]. The general idea of credit risk evaluation includes applying a classification technique that analyzes the relation between characteristics of a customer and likelihood of failure/default [71]. Utilizing these techniques involves the use of past customer data, related to both successful and default loans. Intuitively, when comparing a new customer's characteristics with past customer records, if those characteristics are similar to customers who defaulted after receiving the credit, the credit application may be justified to be rejected. In the case when the characteristics match customers who successfully paid back the credit, the new customer may be granted the credit [1, 71]. Credit risk evaluation is mostly conducted in two ways [40, 82]:

- Human (Expert) Judgment: Mostly qualitative analysis
- Credit Risk Model: Mostly quantitative analysis

Human judgment, which is a traditional method of evaluating credit risk, involves evaluating the credit risk of each credit applicant separately by an expert, based on the experience from previous decisions [43]. When applying human judgment on consumer credit loans, investors examine some of the following information (termed as 3C's, 4C's or 5 C's) [117, 121]:

- Credit History: Customer's credit history.
- Capital: Loan amount being applied for.
- Collateral: Applicant's resources as security for a loan.
- Capability: Applicant's ability to repay.
- Condition: Credit conditions in the market.

Human judgment is subjective, as an analyst reviews each credit application manually applying own experience and consequently the decision is

based on personal insights, knowledge, and intuition of the analyst, which can create bias in decision making. The decisions can be inconsistent as they are to a large extent motivated by individual preferences [12].

The continuous growth in scale and complexity of financial institutions with increased demand for credit has generated the need for more sophisticated techniques to manage and monitor credit risk. Human judgment-based evaluation alone is insufficient to process a large volume of credit applicants. Henceforth, it is necessary to implement sophisticated statistical models in credit granting decision [43, 54].

Credit scoring models serve as credit risk management techniques in evaluating a borrower's credit risk [121]. It supports in credit decisions on granting credit to credit applicants [118]. More formally, credit scoring can be defined as a statistical method used for predicting the likelihood of a credit applicant or an existing borrower defaulting [85]. In general, the main purpose of a credit scoring model is to classify credit applicants to be "accepted" or "rejected" for granting the credit [82]. Credit scoring models are developed using statistical techniques based on borrowers historical credit records for estimating the creditworthiness of current borrowers [118]. Credit scoring models use information from application forms, and other sources as predictor variables for estimating the default probability [43]. They are used as automated decision support tools to handle a large volume of credit applicants [118]. Various credit scoring models have proven to be effective tools in handling increased credit defaults in the financial industry [130, 70].

## 1.4. Machine Learning for Credit Risk Evaluation

Machine learning(ML) refers to a set of algorithms, which are designed for tackling and automating computationally intensive problems of pattern-recognition in large databases [54, 14]. The objective of machine learning is to facilitate the knowledge engineering process with increased automation that replaces time-consuming human activities [64]. It aims at discovering patterns and exploiting regularities in data, which can then be used for making accurate predictions on the behavior of new examples [64, 14, 41].

ML algorithms have the ability to learn from data and make predictions by applying not only statistics but also mathematical optimization. They focus on making accurate predictions based on the generalizing patterns and can dynamically adapt to changes in the data [41]. In comparison to purely statistical modeling approaches, ML approaches are imposed with fewer restrictions for pattern recognition in terms of rigid distribution cri-

teria of the data set, contributing to increased popularity in data mining applications [14]. Furthermore, ML models can improve the learning process with additional data that allows for improved predictive power over time [71].

ML techniques are in extensive use in the field of data mining applications. The main business sectors where machine learning techniques are extensively used include finance, marketing, telecommunication, web analysis, insurance, and many others [14]. ML is identified as one of the important technologies for risk management in building more accurate risk models. ML methods are very well suited for creating credit scoring models because of the presence of large data samples and the ability of the methods to identify the complex and nonlinear patterns by exploring the relationship between transactions and consumer characteristics [31, 71].

ML is widely seen to have the potential to fulfill the analytical capability required in financial services [71]. Credit risk evaluation has been one important area where ML methods are extensively used. ML provides higher accuracy and predictive power; it has been widely applied to construct credit scoring models to predict the credit risk of loan applicants. Furthermore, several studies have shown that ML techniques are superior to traditional statistical methods for credit scoring, where they outperform using nonlinear pattern classification [119].

## 1.5. Research Objectives

Peer-to-peer lending's growth and popularity have established it as a significant part of alternative lending, emerging as a new alternative to financial services. However, credit risk is the prime concern for investments in P2P lending that arises from the lack of collateral and analytical skills. Therefore, with the primary objective of evaluating credit risk in P2P lending, the study aims at applying multiple approaches in identifying credit risk in P2P lending for supporting lending decisions with predictive analytics and machine learning. Credit scoring will be applied as the statistical tool for estimating new applicants' credit risk in P2P lending to support lenders in screening borrowers.

While credit risk evaluation and prediction of credit risk for new applicants can help lenders, the profitability of the investments is also a significant concern to the lenders. Hence, this study also focuses on estimating returns from P2P loans while considering credit risk. Focusing on analyzing credit risk in P2P lending the study aims to answer the following research questions:

1. What is the current state-of-the-art in modeling credit risk in P2P lending?

   Investigating credit risk in P2P lending has been a major study topic with the growth of P2P lending. Since P2P lending behaves differently from traditional financial services, it may require different approaches to understanding credit risk. As part of understanding the approaches to credit risk analysis in P2P lending, one of this thesis' main objectives is to understand and critically evaluate the current state-of-the-art modeling approaches being applied in credit risk evaluation of P2P lending.

2. How can an estimation of return be made from P2P loans to ensure profitable investments in the presence of credit risk?

   While most studies have focused on predicting credit risk in P2P lending with credit scoring, it may not accomplish lenders' requirements to achieve higher return. While predicting credit risk provides an estimation of the risk associated with a loan that helps lenders screening borrowers, high return is the main attraction to lenders in P2P lending. Risky loans are typically the ones that yield higher return. However, in the presence of high risk, higher return cannot be guaranteed, and the return (interest rate) associated with the risk may not be fully compensated for. Therefore, estimating the return by accounting for the credit risk can help determine a loan's profitability. The estimation of return in the presence of the risk will assure lenders that the risk they take in investing will (most likely) result in higher profits.

3. How can different risk groups in P2P lending be accommodated in modeling credit risk for precise loan selection decisions?

   Borrowers in P2P lending are placed in different credit risk groups based on their credit history. The risk groups show borrowers with a diverging level of risk, resulting from a large number of loan applicants with varying credit histories. With such diversity in the credit risk level, it can be challenging to make a loan selection decision based on a single predictive model as it may not provide accurate decisions for all levels of risk. Considering the similarities of borrowers in terms of risk, segmented modeling can be performed that targets specific groups. Segmented modeling will allow the predictive models to be

more focused on specific behavior patterns in a group of similar borrowers that can yield specific decision criteria contributing to more accurate decisions.

4. What strategies and modeling approaches can be applied to create a profitable portfolio of P2P loans with rational allocation of investments?

   Lenders in P2P lending have the opportunity to select borrowers to invest in from a large pool of applications. The possibility to partially fund a borrower allows them to spread their investments to multiple borrowers and diversify risk. With the possibility to partially fund a borrower, the primary decision to be made is the amount to invest in a borrower. While the risk and return estimation for a single loan can help in loan selection, it does not offer any indication on the amount to invest. Lenders who mostly have a fixed amount to invest would typically prefer to obtain an overall estimation of return from their entire investment. This requires a careful selection of loans and allocation of the investment to the loans while considering maximising the return and keeping the risk at an acceptable low level. The need for overall estimation of portfolio risk and return can be addressed with the typical approach of portfolio optimization that considers the risk and returns and provides support for allocating the investments on the loans.

## 1.6. Overview of the Thesis

The remaining part of the thesis is organized as follow:
Chapter 2 presents the overview of P2P lending as an alternative market that highlights its growth and popularity. It describes the general process of loan transactions and finally discusses the credit risk in P2P lending. Research methodologies used in the thesis is presented in chapter 3. In chapter 4, credit scoring as a tool for credit risk evaluation is presented in detail. The chapter introduces machine learning for credit scoring and discusses other issues related to credit scoring. Chapter 5 is dedicated to presenting the machine learning and statistical models used for constructing credit scoring models.

Chapter 6 presents the evaluation metrics used for evaluating performance of machine learning models and discusses their choice of selection for credit scoring models. In chapter 7, some of the most related studies

in P2P credit risk evaluation are presented. Chapter 8 introduces the data sources and briefly describes the data used in the experimental research. Chapter 9 is the presentation of the experimental researches performed for answering the research questions of the thesis. Following the research experiments, chapter 10 is the discussion on the results obtained with the research experiments. It also presents the discussion to justify the selection of the research experiments and its contribution to the literature. Finally, chapter 11 includes the conclusion of the thesis that summarises the main concepts and results of the thesis and presents some limitations.

# 2. Peer to Peer Lending

In this chapter, peer-to-peer lending as an alternative market for credit loans is described to present the detail overview of its purpose and operation. This chapter introduces P2P lending for credit loans and its major characteristics, and highlights the general operational process of P2P lending. Additionally, the presence of high credit risk in P2P lending is described to emphasize on the relevance of performing credit risk analysis on P2P loans.

## 2.1. Introduction to Peer to Peer Lending

Peer-to-peer lending as a microfinancing service started in 2005 from the UK with 'Zopa' as the first service provider. P2P lending saw rapid growth, becoming an excellent alternative to the traditional financial markets, where the financial crisis in 2008 contributed to a 'quasi-explosive' growth in the industry [39]. P2P lending aims to provide an online platform that connects borrowers and lenders directly, without any financial intermediaries [56]. The direct connection to the lenders allows borrowers to obtain a loan that is free from the involvement of financial intermediaries in the decision-making process and any unnecessary or unwanted services coupled with traditional intermediaries [6, 122].

P2P lending provides mutual benefits to both borrowers and lenders, where borrowers benefit from low-cost loans and lenders have the possibility of gaining higher return compared to similar investments in traditional financial services [81]. The absence of financial intermediaries and collateral in P2P lending allows access to loans at a low cost. The cost is further reduced due to low operating costs of the online platforms, and the processes are automatized using a simple business model. The online operations in P2P lending connect borrowers and lenders instantly over the Internet, which makes the lending process easy and quick [109, 29, 81].

Due to the low cost and lack of collateral, P2P lending attracts small borrowers placed at the long tail of credit, who are not seen beneficial by financial institutions, such as banks [109, 50]. In addition, the higher return advertised on investments compared to traditional investments, such as bank deposits, attracts lenders to the service. A loan request by a borrower in P2P lending is typically funded by multiple lenders, which results in lenders holding interests in several borrowers [122]. Since P2P lending allows

lenders to fund a minimum amount (e.g. $25 on LendingClub[4]), lenders have the opportunity to spread their investments to hundreds of borrowers to diversify the risk. Additionally, P2P platforms provide lenders with an abundance of potential borrowers' choices to lend [122, 81].

The information listed on loan applications, which includes demographic, financial, and credit history of borrowers, is the primary source of lenders' investment decisions [56]. In addition, lenders rely on credit ratings assigned to the loans by the lending platform and external credit agencies [6]. Borrowers and lenders can also exchange messages for developing trust. Some P2P lending platforms provide an additional platform for social networking that provides an opportunity for lenders to share their experience and make a collaborative effort in making investment decisions [44, 39, 76]

P2P lending has seen rapid growth in recent years. The growth can be observed in the number of P2P lending platforms and loan volumes [72]. Figure 1 shows the growth in P2P loan volumes globally from 2013 to 2018 [5]. However, lenders in P2P lending are also exposed to high credit risk as P2P lending platforms simply act as intermediaries without any credit services. The lack of collateral is the primary source of credit risk in P2P lending that leads to loss of investment as the major problem in P2P lending [72, 129]. Hence, in the absence of security, lenders may lose a large portion of their investment if a borrower fails to pay the credit. With the sector's growth, P2P platforms attempt to impose regulations in securing loans and collecting loan payments. However, this still does not guarantee a full recovery of investments to lenders.

## 2.2. Peer to Peer Lending Process

P2P lending processes are seen to vary depending on the lending platform and also the region of operations. Therefore, this section aims to present a general overview of the P2P lending process described in Figure 2.

The P2P lending process begins with a borrower making a loan application to a P2P lending platform by providing information such as loan amount, loan period, the purpose of the loan, income, and open credit lines. Additionally, the application includes the demographic and financial information of the borrower and some credit history. The application then goes through underwriting activities to validate the information and evaluate the borrower's ability to repay to decide whether the borrower is eligible

[4] https://www.lendingclub.com/

[5] https://p2pmarketdata.com/p2p-lending-explained/

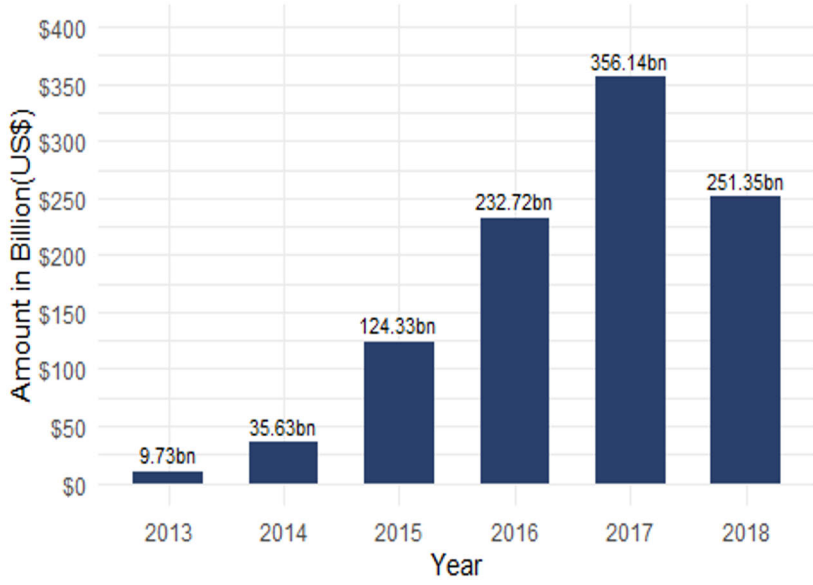


**Figure 1:** *Global P2P Lending Market(2013-2018)*

for the loan. [6, 122]. Based on the evaluation of the borrower's credit history, the application is validated and evaluated and is assigned to a credit risk group. Each group reflects a level of credit risk, and an interest rate is assigned accordingly. Finally, the loan is placed in the market for bidding [122].

From the list of approved loan applications, investors can now make their selection for investments. After an application obtains the required amount of bidding to fund the requested loan amount, the application then becomes a loan. In the process of bidding, investors can partially bid on an application allowing them to spread their investment to multiple applications for risk diversification [44, 122]. The credit risk groups assigned to the loans by the P2P lending platform serve as a primary evaluation metric to investors. Investors can freely bid on loans from different risk groups, according to their risk-return requirements [42].

Borrowers make the loan payments as monthly installments within a time period determined after receiving the loan amount. For late payments, there are additional penalty charges on the payments [44]. When a borrower fails three consecutive payments in the payment process, the borrower is given the label Default. The credit risk on the loan is high at this stage, and the borrower is regularly sent notifications for the payments. In

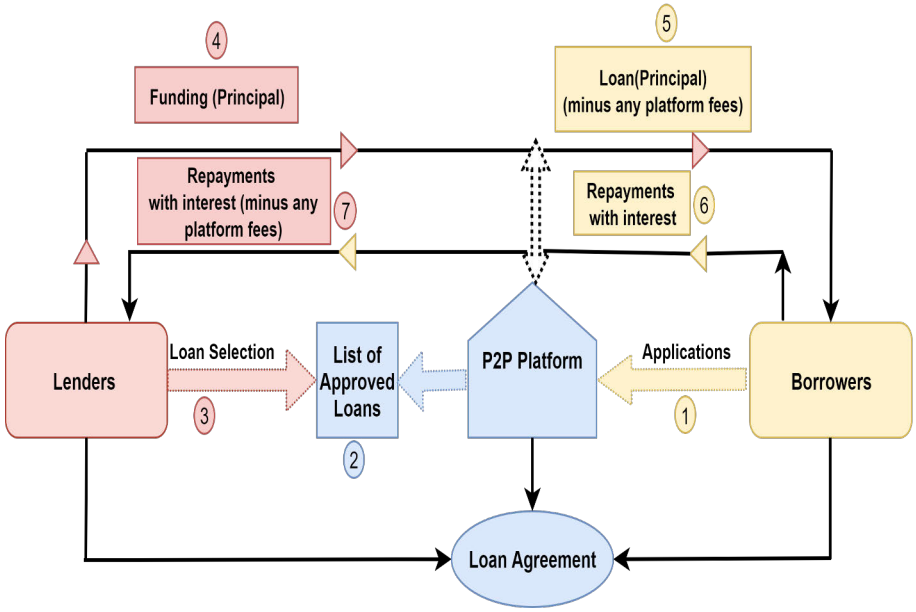


**Figure 2:** *P2P Lending Process*

some cases, borrowers continue to make regular payments after the default period with penalties incurred. However, most borrowers after the default period fail to make regular payments. Hence, the lending platform's recovery measures try to recover as much amount as possible before finally labelling the loan as charged off, meaning no additional amount could be recovered. Some P2P platforms also sell such loans to credit recovery agencies and distribute the amount to the lenders [50].

Some P2P lending platforms also provide an additional secondary market service, where transactions on already existing loans can be performed. The secondary market's primary purpose is to provide the opportunity to lenders to resell their loan holdings. The secondary market allows lenders to put their loan holdings on the market any time before the maturity of the loan. Lenders can generate multiple loan notes by splitting the original loan into small loans and place it on the market for sale. The generated loan notes can be listed at discount or premium depending on the loan's current performance. After the listing, it follows a similar bidding process as on the primary market. Typically, lenders list their loan holdings in the secondary market when they see a discrepancy in loan payments from the borrowers. Hence, most of the secondary market loans have high credit risk, and they are listed with high discount rates.

## 2.3. Credit Risk in Peer to Peer Lending

Lenders benefit from a comparatively higher return to traditional financial services but also face high credit risk due to the lack of collateral. In the absence of collateral, lenders are faced with the challenging task of the proper credit risk evaluation. In addition, lenders suffer from various biases that may lead to failure in making intelligent investments based on the available information [86]. The assessment of default risk in P2P lending becomes increasingly prominent due to the unsecured nature of loans which makes the accurate prediction of default risk essential criteria in identifying credit risk for avoiding losses [109, 73].

In P2P lending, traditional financial metrics may not capture the non-conventional dynamics present in the field which makes credit risk identification in P2P lending challenging. Loans with higher credit risk are charged high interest rates allowing lenders to earn a higher return for the risk they take [81]. However, the higher interest rates charged for high-risk loans may not be sufficient to compensate for the credit risk [32].

Traditional financial and baking services implement high levels of credit risk management measures to safeguard their investments. A similar level of credit risk management measures is difficult to be applied in P2P lending due to the high associated costs. Furthermore, applying such measures becomes complicated because P2P lending is operated online and there is no physical meeting between lenders and borrowers [77, 32]. Besides, the industry's lack of clear and rigid rules and regulations adds complications and uncertainty to safeguarding investments.

The unsecured nature of loans in P2P lending contributes to a high credit risk, due to which the full recovery of the investments is not guaranteed. To overcome the issue of unsecured loans, some P2P lending platforms implement protection mechanisms, such as capital protection and recovery of arrears. They intend to secure investments with a low credit risk [72]. Information asymmetry is another source of credit risk in P2P lending. Since all the operations take place online, there is a possibility of wrong information being provided by borrowers that may lead to misinterpreting credit risk by lenders [133]. However, the risk from information asymmetry is typically low as the P2P platform validates information before loan approval.

The risk of investment loss also comes from the lack of analytical skills of lenders. Most lenders are individual and non-professional investors, who are not trained to evaluate investment risk, and therefore have problems selecting loans to invest on [56]. Lenders in P2P lending, being private investors, follow different strategies of investment. Hence, it is difficult to

create any explicit rules to guide lenders in investment decisions. Lenders tend to primarily rely on verified financial information for investment decisions. Therefore, lenders' skills in evaluating credit risk from the available financial information becomes pivotal [44, 50].

Furthermore, with extensive applications, manual assessment of the risk becomes difficult and may require a high level of expertise. The lack of expertise in risk assessment leads to most lenders showing a herding behavior, where they simply bid on a loan that has a high number of bids without any further analysis [67]. The lack of analytical skills is a major drawback and concern to lenders; most P2P lending platforms nowadays provide an automation system for investments. According to lenders' requirements, all the investment decisions are automatically handled with the automation system through the platform's system.

# 3. Research Methodology

This chapter presents some of the research methodologies in information systems that apply to this dissertation. We identify three commonly used research methods: Positivist Studies, Design Science and Predictive Analytics, in information systems to be aligned with the dissertation.

## 3.1. Positivist Studies

Positivism has a long history in the field of science and is a prominent paradigm in academic research [90, 114]. The concept of paradigm as defined by Kuhn [59] is:

> "universally recognized scientific achievements that, for a time, provide model problems and solutions for a community of practitioners"

According to Hughes [46], positivist paradigm perceives that the world is based on unchanging universal laws and there is an explanation to everything that occurs based on the knowledge of these universal laws. In order to understand the universal laws, a systematic approach can be implemented to observe and record events around us to find the underlying principle that has 'caused' the event to occur.

Positivism follows 'hypothetico-deductive' model of science, where it aims at verifying priori hypotheses that are often stated quantitatively. It attempts at deriving functional relationships that establish explanatory associations between explanatory factors(independent variables) and outcomes (dependent variables) [100]. Hence, a primary goal of positivist studies is to derive explanatory associations that help to make predictions and control the phenomena in question [106].

Positivist studies can be related to different research methods that include scientific, quantitative, experimental, and correlational methods [90]. This dissertation mostly falls in the quantitative method of positivist studies. The quantitative methodology can be seen as a structured approach to research, where the detailed procedure of the research is determined before even the data collection is performed. In quantitative methodology, the focus is on the task of measuring, quantifying or finding the extent of a phenomenon [61].

## 3.2. Design Science

The majority of research in Information Systems can be seen to follow behavioral science and design science paradigms. In behavioral science,

research attempts in developing and verifying theories that explain or predict human or organizational behavior. Design science in research seeks to extend human knowledge through the creation of innovative artifacts that provides a solution to business needs. Design science consists of two primary activities, 'build' and 'evaluate'. The build activity includes the process of constructing artifacts for a purpose, while in the evaluate phase, the performance of the artifact is tested. [45, 83].

The build and evaluate activities in design science is an iterative process, where it goes through multiple iterations. The build phase in design science involves a sequence of expert activities that generates a design artifact as an innovative product. The evaluation phase is responsible for providing feedback on the artifact and a better understanding of the problem that helps to improve the quality of the product and the design process [45, 84]. The evaluation of the quality and effectiveness of the artifact is primarily performed with computational and mathematical methods [45].

Hevner et al.[45] describe design science as a problem-solving process and provide guidelines for implementing design science research framework. Following the guidelines by Hevner et al. [45] for design science research, the research process for this dissertation can be described as follow:

- **Guideline 1: Design as an Artifact**
  Creating credit scoring models with multiple approaches applying machine learning.

- **Guideline 2: Problem Relevance**
  The growth in P2P lending and high credit risk to lenders signifies the study of credit risk evaluation in P2P lending for better investment decisions.

- **Guideline 3: Design Evaluation**
  Multiple standard evaluation metrics(ROC-AUC curve, PR-AUC curve, FScore) were applied for evaluating machine learning models to demonstrate the utility and quality of performance.

- **Guideline 4: Research Contributions**
  Implementation of new approaches to estimate return from P2P loans in addition to credit risk.

- **Guideline 5: Research Rigor**
  Selection of appropriate methods in studying credit risk evaluation.

Study of literature to select methods for training and evaluating machine learning(classification/predictive) models.

- **Guideline 6: Design as a Search Process**
  Iteratively training machine learning models for best performance and testing multiple approaches to obtain the optimal solution.

- **Guideline 7: Communication of Research**
  The research results are presented fully describing the technical aspects of the research process. The results are presented to address the need of P2P lenders that are easy to understand by the lenders.

## 3.3. Predictive Analytics

Predictive analytics comprises statistical models and empirical methods, such as data mining algorithms which are focused on generating empirical predictions. It also includes the methods that are used for evaluating the quality of the predictions generated which describes the predictive power. Besides generating predictions for practical usefulness, predictive analytics can also contribute to theory building, theory testing and relevance assessment [30, 2, 110].

Predictive analytics helps to answer questions based on the past data pattern on what would be the outcome, given different conditions. It provides a quick and inexpensive approximation of relationships between variables [126]. Predictive analytics as a subfield of Data Science, combines statistical modeling, data mining techniques and machine learning to analyze historical data and detect a pattern to make predictions. Predictive models are built relying on variable association and not causation. Predictive models require the population on which the prediction is to be made to be similar to the sample used for training and evaluating the model. The change in data distribution will not guarantee the predictive accuracy of the model [110].

The predictive power or predictive accuracy of a model shows the ability of the model to generate accurate predictions. Predictions are made on new observations, which are observations in future time periods or the observations can be the ones that were not included in the original sample used for training the model [110].

Predictive analytics constitute an important research method for this dissertation, where it is applied in generating predictions related to credit risk. Applying machine learning algorithms for risk modeling based on historical P2P credit loan data, the aim is to predict the credit risk for similar

borrowers and estimate return. This dissertation follows the steps mentioned by Shmueli & Koppius [110] to build a predictive model as shown in Figure 3.

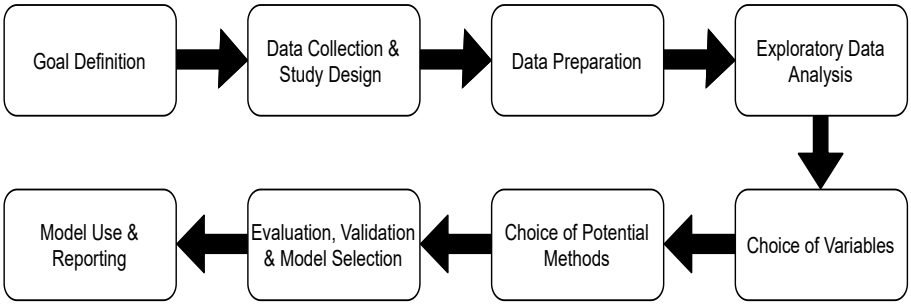


**Figure 3:** *Predictive Analytics Process*

The predictive analytics process in Figure 3 implemented for this dissertation can be summarised as follow, following the description by Shmueli & Koppius [110]:

- Goal Definition: Predicting Credit risk and estimating return in P2P lending.
- Data Collection & Study Design: Collection of data through online P2P platforms and plan the research design to utilize the data.
- Data preparation: Data cleaning steps, including imputing missing values and removing or replacing the outliers. Data partitioning for modeling (train, validation and test sets).
- Exploratory Data Analysis: Perform exploratory data analysis to understand the relation between variables through visualizations.
- Choice of Variables: Selection of variables for modeling through feature selection methods and understanding from exploratory data analysis.
- Choice of Potential methods: Listing out the potential machine learning models, including both supervised and unsupervised methods and other data science methods.
- Evaluation, Validation & Model Selection: Performing model evaluation and validation with evaluation metrics and select the best performing model.

- Model Use and Reporting: Perform predictions with the model and report the results using appropriate metrics(ROC-AUC, Precision, Recall) and visualizations.

# 4. Credit Scoring

This chapter is a detailed presentation of credit scoring for credit risk analysis. Along with the introduction and benefits of credit scoring, this chapter describes the application of Machine Learning for credit scoring. Furthermore, this chapter presents the overview of implementing credit scoring for classification of credit loan applicants.

## 4.1. Introduction to Credit Scoring

Credit scoring is a general term used for various statistical methods applied in evaluating the credit risk of credit applicants. Credit scoring models are multivariate statistical models that use economic and financial indicators to predict the probability that a credit applicant or existing borrower will default [85, 71]. Credit scoring is the most well-known technique applied in evaluating the creditworthiness of credit applicants [63]. The objective of credit scoring is to evaluate credit risk accurately and quickly to determine credit applicants ability to repay their debt [130]. This is achieved through predictive models that assign a score for each credit applicant, and, based on a cutoff value applied to the score, the decision is made to grant or decline the credit [8]. Credit scoring models use historical data of borrowers for whom the credit has been granted to study the effect of applicants' various characteristics on the likelihood of default. Furthermore, it helps to identify the characteristics that are useful in estimating the default risk [43, 85].

Credit scoring models help in making consistent credit decisions with automatic processing, allowing for handling a large volume of credit applications. The objective of a credit scoring model is to quantify and manage credit risk for better lending decisions in less time and more objectively [118, 57]. Credit scoring models help in achieving increased accuracy in credit risk assessment which results in more profit by granting credit to more creditworthy applicants and by denying credit to risky applicants [63]. Credit scoring models' importance in reducing risk and generating profits have made them a widely studied topic in accounting and finance [119].

Credit scoring is sometimes called 'application scoring' as the objective is to support decision-making related to granting credit to new applications [118]. As an extension to application scoring, a related model is 'behavioral scoring.' Behavioral scoring deals with existing customers and is used for predicting their future credit status [118, 117]. The purpose of behav-

ioral scoring is to make decisions for existing customers by assigning credit limits, marketing new products, and identifying recovery strategy if the account turns bad. In constructing behavioral scoring, recent past information on customers' payment and purchase behavior and the application details are used [117, 118].

With the credit industry's growth, the traditional way of evaluating credits with expert judgments becomes impossible as it requires a high amount of manual work and economic costs. Therefore, credit scoring models are in extensive use because of their capability of processing a large volume of credit applications in a short time with minimal labour reducing the cost of operations [48, 54]. The advancement in information technology has made it possible for the development of sophisticated credit scoring models that can reduce the cost and effort in credit granting decisions [43, 3].

## 4.2. Benefits of Credit Scoring

Credit scoring for evaluating credit risk has several benefits that contribute to an enhanced credit risk management system. Credit scoring adds benefit to credit risk evaluation with improved objectivity in the loan approval process. The objectivity helps in ensuring that all applicants are reviewed under the same underwriting criteria, regardless of race, gender, or any other factors that are prohibited by law to be used in credit decisions [85]. This eliminates the discrimination as credit providers focus only on the information related to credit risk, and personal subjectivity of credit analysts is excluded [24].

Credit scoring is a statistical methodology that can take into consideration a large number of customer's characteristics simultaneously and their interactions for credit evaluation, which is too challenging and complex to perform manually [27]. In addition, credit scoring models are built upon much more extensive data samples compared to what loan analysts can consider in human evaluation [1]. The automation of the evaluation process enabled through credit scoring allows for a more consistent and faster loan evaluation process [118]. Furthermore, it can handle a large volume of credit applications, resulting in a reduced need for human labour [48]. Hence, credit scoring can reduce the cost of processing credit applications and risk associated with bad credit, contributing to enhanced credit decisions and less time and effort [70]. Credit scoring helps in determining credit risk, pricing the loan by setting a suitable interest rate according to the risk and determining the credit limits to be set for a loan applicant [37, 24].

As credit scoring helps in the quick lending decisions for 'accepting' or 'rejecting' applications, it allows loan analysts to consider more time for the cases that the scoring system might have issues handling [37, 24]. In addition to cost savings, credit scoring provides benefits to customers with quick application processing. Customers need to provide only the information used in the scoring system, making the application shorter [85]. Furthermore, credit scoring models can learn over time as the statistical models can be continuously re-estimated with broader data for improving model performance [37].

## 4.3. Machine Learning for Credit Scoring

Credit scoring models are built using predictive models, and the approaches to the predictive models can be categorized into two main groups: statistical and artificial intelligence (machine learning) [131]. Some standard and widely used statistical methods for developing credit scoring models include linear discriminant analysis, linear regression, and logistic regression. Their popularity are mostly justified by the simplicity of implementation and ease of interpretation [43, 48, 12]. However, they have their limitations related to handling high-dimensional data, and they rely on linear separability and normality assumptions [48]. There has been active development of credit scoring models with various methods and techniques to overcome the issues with traditional statistical methods, as even a minimal increase in credit scoring accuracy can bring significant profits [130]. Machine learning algorithms in recent times have become a popular choice over traditional statistical methods to improve performance.

Machine learning is a field of artificial intelligence, which involves using multiple statistical, probabilistic, and optimization techniques, allowing computers to learn from previous examples [92]. Machine learning follows an iterative process to learn from data and precisely detect hidden patterns in noisy and complex datasets [94]. A general layout of a machine learning process can be seen in Figure 4. With complex models and effective algorithms, machine learning helps in gaining greater predictive accuracy. Machine learning models have been shown to be effective in handling predictive tasks and, in addition, identifying behaviors that are most likely to derive some preferred outcomes [92, 94].

Machine learning algorithms can be broadly categorized into supervised, unsupervised, semi-supervised, and reinforcement learning [49, 68]. Out of the categories, supervised and unsupervised learning are the most applicable to credit scoring. Supervised learning is the training of machine

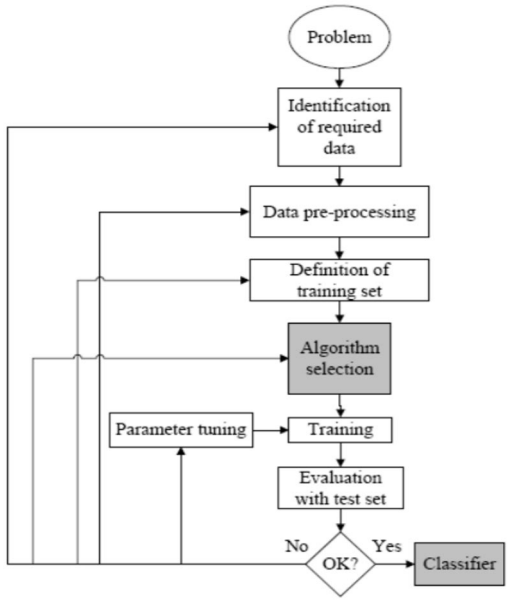


**Figure 4:** *Machine Learning Process [96, page. 2]*

learning algorithms, where the algorithms are fed with a dataset consisting of some input features and the known corresponding output label. In supervised learning, the algorithms find a mapping function that captures the relation between the input features and output label. The learned mapping function is then capable of estimating the output label for a new example [5, 49, 68].

Classification is a typical form of supervised learning used to classify instances in data to a predefined set of classes. It is used in the cases where a prediction is to be made to assign a target label to an instance [53]. The simplest form of classification problem is a binary classification, where the classifier is trained to predict target labels with only two possible values (e.g.: 'good' and 'bad' loans). At the same time, it can be extended to train for multiple target labels ('good', 'acceptable', 'bad' loans) [53]. Credit scoring models are classification models that classify loan applicants to one of the predefined classes of 'good' or 'bad'.

Wide ranges of machine learning algorithms have been applied to perform the classification task for credit scoring in search of better performance. Some commonly used machine learning algorithms for credit scoring include Decision Tree, Random Forest, Gradient Boosting, Neural Network, Naive Bayes, and Support Vector Machines. Many studies have per-

formed comparative analysis to explore machine learning models' performance over traditional statistical models, such as linear discriminant analysis and logistic regression, showing that machine learning models have better performance in credit scoring problems [82, 130, 12, 48, 9].

In unsupervised machine learning models are trained on a dataset with only input features. With unknown output labels for the input data, unsupervised learning attempts to capture the relationship among the input data and group them to provide meaningful insights [89, 49, 68]. One of the standard application tasks of unsupervised learning is clustering. Clustering aims at discovering a new set of categories in the data [103]. The objective of clustering is to organize a large set of patterns in data into clusters or groups based on similarity [51, 125].

Clustering groups data into disjoint and homogeneous clusters, where similar instances are grouped, and instances that are different are organized into separate groups [51, 103]. Clustering in credit scoring is mainly applied for segmented modeling, which allows for generating specific classification rules for each segment of borrowers. A single classification model may not capture the behavior pattern of varying nature of borrowers [75]. Therefore, with clustering, borrowers with similar behavior and risk can be segmented into separate groups, and for each group, a separate credit scoring model is developed. A separate credit scoring model allows for generating classification rules that are more specific to a segment, contributing to increased accuracy of risk identification [26, 107].

## 4.4. Credit Scoring as a Binary Classification Model

Credit scoring is a binary classification problem because it aims at classifying credit applicants as "good" or "bad" based on their likelihood of repayment [119]. The 'good' applicants are referred to the ones that are more likely to repay the credit and hence, they are granted the credit. The 'bad' applicants are the ones that have a high risk of defaulting on the credit and therefore are denied the credit [63, 55, 9].

As a classification model, credit scoring takes the input characteristics as the details from the application form and information from credit bureaus on the applicant to estimate the output as "good", or "bad" [117]. The selection of the right characteristics or variables is important for developing a credit scoring model [31]. The variables can be categorized based on the strength and reliability to predict default probability and the explanatory power in analyzing credit applications [124]. The variables used in credit scoring can further be segmented into four categories [124, 88]:

- Demographic Indicators: Applicant's age, sex, marital status, number of dependents, home status, and district of address.
- Financial Indicators: Applicant's total assets, gross income, gross income of the household, the monthly cost of the household.
- Employment Indicators: Applicant's employment type, length of current employment, number of employments over last $n$ years
- Behavioral Indicators: Applicant's Checking Account(CA), the average balance on CA, loans outstanding, loans defaulted or delinquent, number of payments per year, collateral/guarantee.

Creating a credit scoring model makes use of the applicants' sample of historical credit records over a fixed period when the credit has been issued. The samples are classified into the classes 'good' or 'bad' according to their repayment performance [82, 24]. With the available data sample, statistical and analytical methods are applied to train a credit scoring model that can discover the relationship between historical information and future credit performance [24, 63]. The formulation of the credit scoring model can be represented mathematically as [63]:

$$f(x_1, x_2, .....x_m) = y \tag{1}$$

where, $x_1$, $x_1$,.... $x_m$ are the attributes describing the credit applicant and $y$ is the outcome of the model that identifies the applicant as "good" or "bad". The function or credit scoring model is represented by $f$ that determines the relation between the applicant's attributes and the credit risk.

The model specifies the weights associated with the input variables or attributes, and the weighted sum of attribute values for a new applicant gives a credit score. Furthermore, the model has a cut-off point that is used as a baseline to classify an applicant as 'good' or 'bad' [24, 71]. With statistical methods, a credit score can be represented as a probability score, which entails the probability of a customer defaulting on the loan. Formally, the probability of default for a customer $i$ with features $X_i$ can be represented as [8]:

$$p_i = P(y_i = 1|X_i) \tag{2}$$

The estimated probability score is compared against a threshold value $t$ for classifying the customer $i$ to a class: if $p_i$ < t, the customer is classified as 'good' and if $p_i$ > t, the customer is classified as 'bad' [8].

A good credit scoring model should differentiate between good and bad credit applicants based on their characteristics. While it is unlikely to obtain a perfect model, performance can be increased with sufficient historical data representing the loan performances during both bad and good economic periods [85].

## 4.5. Threshold Selection for Classification

Credit scoring models produce ”score” as a primary output used as the base for classifying the applicants. A good credit scoring model should possess a high discriminating capability, assigning high scores to applicants that are likely to perform well, and low scores to applicants who are more likely to perform poorly on their payments [85, 118].In the credit scoring model scenario as a classification model, the ”score” is obtained as the default probability that ranges between values 0 and 1.

The decision on what percentage of applicants to accept for granting the credit is dependent on managerial preferences related to business measures and risk. Depending on the amount of risk the lender is willing to accept and other business measures, a ”threshold” or a ”cut-off” score is determined. With the threshold score being stated and considering the default probability as the score, the lender rejects the applicant having default probability above the threshold and accepts if it is lower than the threshold as shown in Table 1. For the cases where the score is very close to the threshold, additional supervision, in the form of human judgment, may be required [85, 118].

| Default Probability | Thresholds | | |
|---|---|---|---|
| | 0.2 | 0.5 | 0.7 |
| 0.6 | Bad | Bad | Good |
| 0.4 | Bad | Good | Good |
| 0.1 | Good | Good | Good |
| 0.8 | Bad | Bad | Bad |
| 0.3 | Bad | Good | Good |
| 0.05 | Good | Good | Good |
| 0.35 | Bad | Good | Good |
| 0.55 | Bad | Bad | Good |

**Table 1:** *Threshold Selection*

There are multiple factors that lenders consider while selecting a thresh-

old for the classification of the applicants. The decision can be subjective or more based on the empirical evidence. In order to make use of the data in threshold selection, calculations, such as marginal good: bad odds at a threshold and change in the good to the bad ratio by varying the threshold is performed [118]. With machine learning being applied for credit scoring and multiple metrics available for evaluating classification models' performance, the threshold can be selected by evaluating the model's performance at a threshold against an evaluation metric.

Most of the metrics for evaluating classification models' performance give importance only to misclassification error, i.e. minimizing the number of credit applicants (or an evaluation metric) that the model incorrectly classifies. However, credit scoring can be considered a special case of classification, where not only misclassification error but the cost of misclassification error also needs to be emphasized. The cost of misclassifying bad applicants as good can be more costly than misclassifying good applicants as bad. Hence, these costs can be considered in selecting a threshold to achieve low cost of misclassification [95, 8].

## 4.6. Problem of Imbalanced Data

The class imbalance problem is often present in supervised classification tasks, where the data is imbalanced and contains more instances of a particular class compared to other classes. More often, in imbalanced data, the class of interest occurs very rarely compared to the other classes [21, 78]. In the presence of imbalanced data, standard machine learning classifiers, which are designed for overall accuracy and assume balanced class distribution, tend to classify more accurately the larger classes, while they ignore the smaller class [21, 115, 33]. This bias behavior of classifiers towards the majority class results in a significantly higher misclassification rate for the minority class [78]. However, in some applications, the correct classification of samples in the minority class has higher importance than the majority class [115]. In addition, the cost associated with misclassification of the minority class is higher compared to the majority class [78].

Credit scoring as a classification problem for default prediction is one such application characterized by imbalanced data, where the number of default examples is typically very low compared to non-defaults. The low number of defaults is seen as the result of rejecting bad applicants during the loan approval process [95]. In imbalanced data, it becomes more important to correctly recognize the minority class as they represent the larger loss when misclassified [78]. The minority class in credit scoring is

'bad' loans: approving a 'bad' applicant can result in the total loss of the loan amount. In contrast, the loss is only the opportunity cost when rejecting a 'good' applicant. Therefore, the cost of misclassifying an applicant that subsequently defaults is significantly different from the cost of misclassifying a 'good' client [95]. There are several methods in practice to handle the issue of imbalanced data. The methods used for solving imbalanced data problems can be mainly categorized into three groups: Data Sampling, Algorithmic Modification, and Cost-Sensitive Learning [95, 78, 33].

Data sampling is the most common approach used for handling imbalanced data. The training data is modified to balance the number of instances in the majority and minority classes by resampling the data [91, 36]. The balancing allows the classifier to learn similar to a standard classification process by reducing the effect from imbalanced data [78, 36]. Data sampling methods are applied as a preprocessing step and, therefore, are applicable to any problem, independent of the classifier used [36, 95]. Some of the most commonly used data sampling methods include:

- Oversampling: the aim is to balance the data by replicating the instances from the minority class. The selection of instances to replicate can be either random or from the areas close to decision boundaries [11, 36, 91]. Oversampling is prone to overfitting as it makes exact copies of the minority instances and, therefore, fails to provide any new information to the classifier for learning [11, 95].

- Undersampling: The balancing of the data is performed by removing instances from the majority class. The selected instances can be at random or from area far from the decision boundary [11, 36, 91]. The major drawback of Undersampling is that it could potentially remove useful data [11, 36].

- Synthetic minority oversampling technique (SMOTE): It is an oversampling method that generates new synthetic examples of the minority class by interpolating the closest existing minority examples. It selects the $k$ nearest neighbors for each minority class example and generates synthetic examples along the direction of the all or some of the $k$ neighbors depending on the size of oversampling specified [20, 36].

Algorithm modification approaches attempt to handle imbalanced data issues by making changes to the base classifier to make it more attentive to the minority class [95, 78]. Some common modifications that can be made

are to assign unequal class weights to penalize more errors made in minority class, make changes in classification threshold, and assign different costs to the errors made for the classes [95].

Cost-sensitive learning incorporates both the approaches of data sampling and algorithm modification or the combination of both approaches, and assign a higher cost of misclassification to the minority class [78, 36]. Taking into consideration a cost matrix (an example of cost matrix for default prediction is shown in Table 2), which holds the misclassification cost for the classes, cost-sensitive learning tries to minimize the total misclassification cost [8]. With higher misclassification costs assigned to the minority class, the classifier tends to be more sensitive towards making incorrect classification within the minority class [36]. The misclassification costs are usually not available in the data, and defining the cost matrix can be difficult, and usually performed by domain experts and other heuristics approaches [78, 36, 95].

| | | Actual | |
|---|---|---|---|
| | | Good | Bad |
| Predicted | Good | 30 | -100 |
| | Bad | -50 | 30 |

**Table 2:** *Cost Matrix*

# 5. Machine Learning and Statistical Methods for Credit Scoring

In this chapter, some of the most commonly used statistical and machine learning models for credit scoring are discussed. The models includes supervised and unsupervised algorithms that are applied for credit scoring. This chapter introduces the models to understand the general overview of their implementation.

## 5.1. Classification Models

### 5.1.1 Logistic Regression

Logistic Regression is an example of the statistical model family termed "Generalized Linear Models", which is an extension of Linear Regression. Logistic Regression determines the relation between a set of potential independent variables or predictors and a dichotomous dependent variable. With the dependent variable being dichotomous, Logistic Regression is applied as a binary classification method to classify input vectors into one of the two classes. Logistic regression estimates the probability of a binary outcome using the 'logit function', which predicts the logarithm of the odds of belonging to a class as a linear combination of the input features. Given a set of input features $x_1, x_2, x_3.....x_n$, the logit model can be represented as [70, 58, 120]:

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3.....\beta_n x_n \tag{3}$$

where $p$ is the probability of belonging to the class of interest, and $\alpha$ and $\beta_1, \beta_2...\beta_n$ are the regression coefficients. The coefficients are estimated by applying the 'maximum likelihood' method. A 'logistic' function, also called 'sigmoid' is applied to the outcome from the logit model to obtain the the probability of belonging to a class, a value in the range between 0 and 1, which is represented as [58]:

$$P = \frac{1}{1 + \exp^{-}(\alpha + \beta_1 x_1 + \beta_2 x_2...\beta_n x_n)} \tag{4}$$

Logistic Regression is widely used as the first choice in developing credit risk models in predicting the probability of a customer defaulting in the loan payment. Although many modern machine learning models have the

ability to provide better predictive power, Logistic Regression is still popular due to its simplicity in model development and ease in interpreting the results [28, 26]. The popularity of Logistic Regression for credit scoring is visible from its extensive use in developing different models [28, 70, 12, 26, 40, 48].

### 5.1.2 Decision Tree

A decision tree is a sequential model, that recursively partition an input instance, following a series of tests with respect to a target variable. In essence, it generates a set of rules for classifying an input instance to a target class [104, 128]. Decision trees are composed of nodes, where the starting node is called a 'root' node. From the 'root' node it is directed towards other nodes, where a node with a further outgoing edge is called a 'internal' or 'test' node. A node, that does not have any outgoing edges is called a 'terminal' or 'leaf' node. When a leaf node is encountered, it is assigned a class label representing the most likely target value [104, 99].

At each internal node, including the root node, an input feature is selected, where a test is performed and a decision is made for the given problem, splitting the instance space into two or more sub-spaces. The selection of a feature at a node for decision making is based on a splitting criterion, where 'Information Gain' and 'Gini Index' are the two most used criteria [104]. The split of a node takes place according to the input feature's value for a categorical feature, while a threshold value is used for numerical features. Instances in decision trees are classified by passing them from the root of the tree to the leafs, performing the tests along the path [99, 128]. A simplistic example of a decision tree is presented in Figure 5.

### 5.1.3 Random Forest

Random Forest is an ensemble method that combines the predictions from multiple decision trees by aggregation. Random forest applies the bagging method, where each tree is trained on a bootstrapped sample from the original data that creates randomness and increases the diversity of the trees [16]. In addition, Random Forest introduces more randomness by allowing the trees to train only on a set of randomly selected features. It selects the best feature to split a node in the tree from a random subset of input features rather than all the input features. By using the bagging approach to train multiple decision trees on different training data subsets with a random subset of input features, Random Forest brings diversity in training, which reduces the variance and improves the generalization error

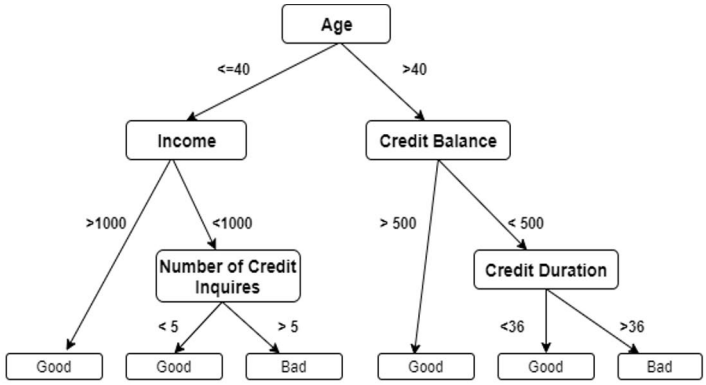


**Figure 5:** *Decision Tree*

[16, 74, 98].

In a Random Forest classifier, the trees' results are combined by unweighted voting (majority voting) for a new prediction. When training a Random Forest classifier, it consists of $n$ trees as defined by the user, and the number of random features to use in the tree splitting is also assigned by the user. Random Forest training process can be summarised as [16, 74] :

- Take $n$ (number of trees) bootstrap samples from the original data.
- For each of the bootstrap samples, an unpruned decision tree is trained. For each decision tree, the best split is chosen from a random subset of features rather than from all the features.
- The prediction for a new data is made by aggregating the predictions from the $n$ trees (majority voting for classification and average for regression)

Random Forest evaluates the model training based on the predictions made on the samples not included in the bootstrap sample, called "out-of-bag"(OOB) samples. For each bootstrap samples created to train a decision tree, it creates an OOB subset that is not considered for training the tree. At each bootstrap sample, predictions can be made for the OOB elements using the decision tree trained with the bootstrap sample. The final error

rate of the classifier can be calculated as the aggregate of the OOB predictions (proportion of the misclassification over the total number of OOB elements), which is termed as the "OOB error" [74, 98].

### 5.1.4 Gradient Boosting Model

Gradient Boosting is an ensemble algorithm that implements the combination of bagging and boosting. Gradient Boosting combines the results of multiple base learners (usually a Decision Tree) by sequentially fitting them at each iteration [35]. A base learner is sequentially built on the "pseudo"-residuals of the previous iteration, where pseudo-residuals are the gradient of the loss function [35, 87]. In an iteration, a base learner is trained with a random sub-sample (without replacement) of training data used for computing model update for the current iteration. The model update is performed by assigning weights to the data samples, where the misclassified samples are assigned higher weights compared to the correctly classified samples. These weights force the base learner to emphasize incorrectly classified samples during the next iteration [66, 87]. Instead of full classification trees, relatively small depth classification trees are created at each iteration, which helps solving the problem of over-fitting [66].

### 5.1.5 Artificial Neural Network

Artificial neural network (ANN) algorithm is based on the concept of how the human brain works. ANN can be viewed as an advanced soft statistical computing tool for information processing capable of learning for generalization and is adaptive [55, 9]. ANN is comprised of three different layers: input, hidden(can have multiple hidden layers), and output, and hence, ANN is also referred to as Multi-Layer Perceptron (MLP) [130]. An example of a single hidden layer MLP is shown in Figure 6. Each layer consists of many processing units called neurons or nodes that are highly interconnected [69]. The nodes are responsible for processing the information, where the input layer is first fed with the input data (raw data, corresponds to features). The input layer results are then passed to the hidden layer, and with further processing, is passed to the output layer that outputs the result [55, 69].

The flow of information from input to hidden and then to the output layer is called the feed-forward network structure (connections). The output obtained is compared against the actual value (using a loss function), and the results are backpropagated (know as back-propagation network) [130, 69]. The back-propagation, following gradient descent algorithm, up-

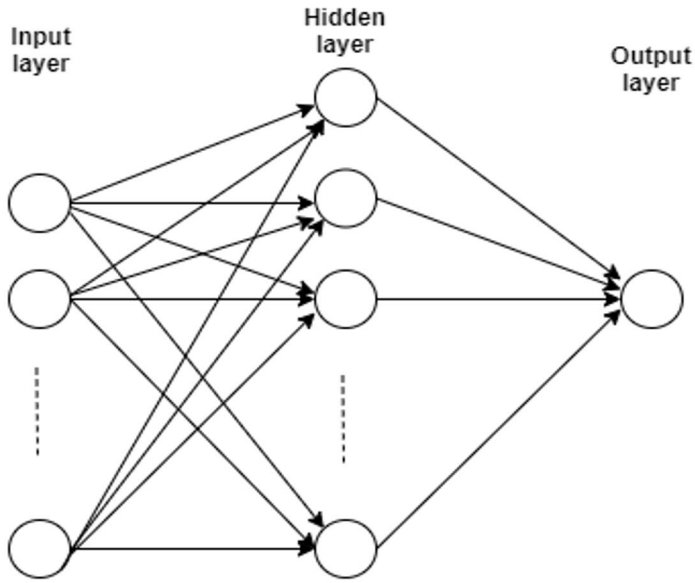


**Figure 6:** *Artificial Neural Network*

dates the nodes (weights) in the layers. This process of feed-forward and back-propagation is continued for a number of iteration until the desired result is obtained [69, 55]. ANN is shown to have performed better in learning non-linear and complex relationships between input and output features [55].

## 5.2. K-Means Clustering

K-Means is an unsupervised algorithm introduced by MacQueen [80] for generating clusters of similar instances from a data set. The simplicity of the algorithm and its fast computation capability for efficiently partitioning a huge amount of data are the reasons for the popularity of K-Means [134]. K-Means algorithm follows an iterative process in generating $k$ mutually exclusive clusters, where $k$ is the desired number of clusters specified by the user. The iterative process of the algorithm can be summarised as [134, 18]:

1. A centroid for each of the $k$ clusters is chosen randomly in the data space.

2. Distances of each data points from the cluster centroids are calcu-

lated (the distance measure is typically the Euclidean distance).

3. The data points are then assigned to a cluster according to the smallest distance to the cluster centroids.

4. After the data points have been assigned to a cluster, the centroids of the clusters are updated in accordance with their surrounding data points.

5. The distances of the data points from each of the updated centroids are recalculated, and the data points are assigned to the same or new cluster following step 3.

Iterating over the steps from 3 to 5, the process continues to the point where no data points are required to be updated to a new cluster. During the iterative process, the algorithm attempts to minimize the sum of the squared distances of the data point to its cluster centroid, which allows for creating homogeneous clusters [18].

## 5.3. Survival Analysis

Survival analysis can be seen as a collection of statistical procedures that study the occurrence and timing of events in making predictions. Its objective is to model the time for an event to occur. It performs the modeling by studying the time to an event for a population over a given time horizon, called observation period [13, 79]. Survival analysis can be applied in credit risk analysis to model the distribution of time to default for a loan, where the event of interest is a loan being the default. With survival analysis, the objective is to predict the default probability of a loan in a given time horizon of choice [79, 47].

Survival analysis can incorporate censored data into the modeling, which is one of the major strengths. A data point is said to be censored if the event of interest has not occurred during the data collection time [113]. When collecting data for credit scoring, if a loan has been fully paid or is still in the payment process, it is marked as censored. In creating credit scoring models, usually, only loan data for which the outcome is known (either default or non-default) are included and excludes information on loans that are still ongoing. With survival analysis, it can also perform the modeling by including the ongoing loans, which allows for more accurate predictions [79, 13].

Survival analysis mainly relies on two time-dependent probabilities for modeling: survival function and hazard function. The survival function

represents the probability that the observation will survive in a specified future time. The hazard function, on the contrary, represents the probability that the event will occur for an observation within the time frame. Given an observation time frame of $T$, the survival probability at a given time of $t$ is represented by the survival function $S(t)$ as [25]:

$$S(t) = P(T > t) = 1 - F(t) \tag{5}$$

where, $F(t)$ is the cumulative distribution function up to time $t$.

The hazard function gives the probability that the event will occur for an observation at a time interval $t$ and $\Delta t$, given that the observation has survived until time $t$. The hazard function $h(t)$ is represented as [7, 25]:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T + \frac{\Delta t}{T} \geq t)}{\Delta t} \tag{6}$$

# 6. Evaluation Metrics

The evaluation metrics used for evaluating the performance of credit scoring models are discussed in the chapter. The evaluation metrics discussed in this chapter includes most commonly used metrics for evaluating classification models. This chapter also tries to give a justification of selection of the metrics for evaluating credit scoring models.

## 6.1. Confusion Matrix

Confusion matrix is a standard measure used for evaluating the performance of a classification model. It summarizes a classification model's predictions using a contingency table, which displays the comparison between the actual classes and the predicted classes from the classification model. The contingency table cells are represented with the raw counts that show the association of a predicted class to the actual class [101, 123]. The contingency table is formulated as a matrix of size $n \times n$, where n is the number of classes in the classification model. With the confusion matrix, we can get a holistic overview of the performance of the classification model to identify the model's strength and weaknesses, and errors made by the model. An example of a confusion matrix for a binary classification model with two classes that can be represented as a $2 \times 2$ matrix is shown in Table3.

**Table 3:** *Confusion Matrix*

| | | **Prediction** | |
|---|---|---|---|
| | | **Positive** | **Negative** |
| **Actual value** | **Positive** | True Positive (TP) | False Negative (FN) |
| | **Negative** | False Positive (FP) | True Negative (TN) |

In Table3 we have two classes from a binary classification model:'Positive' and 'Negative'. General practice is to consider 'Positive' as a more important class to correctly predict due to a higher cost associated with misclassifying it. Since credit scoring is primarily a binary classification problem and with

loan applicants divided into classes 'Bad' and 'Good', the confusion matrix for a credit scoring model can be represented as in Table 4

**Table 4:** *Confusion Matrix for Credit Scoring model*

| | | **Prediction** | |
|---|---|---|---|
| | | **Bad** | **Good** |
| **Actual value** | **Bad** | True Positive (TP) | False Negative (FN) |
| | **Good** | False Positive (FP) | True Negative (TN) |

The 'Bad' class in Table 4 is treated as the positive class and 'Good' as the negative class. Here, the 'Bad' is considered the positive class because when the model classifies the possible Bad loans as Good, there is a high risk of losing the investment. The loss is comparably lower when the model classifies Good loans as Bad, as in this case, we are bound to lose only the opportunity to earn the possible interest from the loan.

A credit scoring model as a classifier maps each instance of loan applications to one of the classes Bad or Good. This process of mapping the instances to the classes can result in four possible outcomes [34].

- True Positive(TP): When the model correctly predicts a Default loan to be Default.
- False Negative(FN): When the model incorrectly predicts Default as Good.
- False Positive(FP): When the model incorrectly predicts Good as Default.
- True Negative(TN): When the model correctly predicts Good as Good.

Several other common metrics can be derived from the confusion matrix. Depending on the data distribution and context of the classification problem, multiple metrics are applied in evaluating a classification model.

### 6.1.1 Accuracy

Accuracy is the most commonly used metric to evaluate the performance of a classification model. It simply states how accurate the model is in correctly predicting the class labels, i.e. to what extent the predicted class labels match the actual class labels. Mathematically, it can be represented as the ratio of correctly classified samples to the total data samples [78, 116]. Referring to the confusion matrix in Table 4, accuracy can be calculated as:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{7}$$

From equation 7, we can observe that accuracy is simply the number of correct predictions the model has made out of the total observations. Accuracy treats all the classes equally and does not distinguish between the number of correct predictions for different classes [111]. Therefore, in the presence of imbalanced data, accuracy as an evaluation metric is not a proper measure and might provide misleading results [78, 11]. In applications such as credit scoring, where data imbalance can be severe, e.g. with 99% of examples belonging to the majority class and only 1% to the minority class, an accuracy of 99% can be achieved by simply classifying all the examples as the majority class. The result gives a wrong impression and also does not correctly classify any of the examples in the minority class that are more important and have higher error costs [78, 11]. Hence, considering the class distribution in the data, other performance metrics need to be applied that measure the classification performance of the classes independently [11].

### 6.1.2 Precision and Recall

Precision is the ratio of the number of correctly predicted samples from the positive class to the total number of samples predicted as positive. Precision, therefore, is the metric that quantifies the accuracy for the positive class predictions [112].

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Recall represents the proportion of positive class samples that the model correctly predicted. Recall helps identifying the number of missed positive predictions, which shows the effectiveness of the model in identifying the positive class [112, 116].

$$Recall = \frac{TP}{TP + FN} \tag{9}$$

Precision and Recall summarizes the classification performance on the class of interest. There is a trade-off in selecting the two metrics, where maximizing the Precision tends to decrease Recall, and vice versa [17]. The business problem directs the importance of one of the two metrics. Maximizing the Precision is appropriate when the focus is on minimizing the False Positives, i.e., reducing incorrect predictions of the Negative class as Positive. Similarly, maximizing Recall is appropriate when the focus is on minimizing False Negatives, i.e., reducing incorrect predictions of the Positive class as Negative.

### 6.1.3 F Score

F measure, also called F score, is defined as the harmonic mean of Precision and Recall. As opposed to the arithmetic mean of two numbers, the harmonic mean tends to be closer to the smaller of the two numbers. Therefore, for a high F score, both Precision and Recall are required to be reasonably high. When one of them is considerably lower than the other, its effect can be seen as a lowered F score [115].

$$Fscore = \frac{2 * Precision * Recall}{Precision + Recall} \tag{10}$$

The value of the F score ranges between 0 and 1, and higher values of the F score indicates higher classification performance of the model. When the classification model fails to correctly predict any of the Positive class samples(True Positive = 0), the F score's value is the minimum, i.e. 0. Its value is maximum, i.e 1, when the model perfectly classifies the samples to its respective classes without any errors(False Negative = 0, False Positive = 0) [116, 23].

The F score, which is also commonly referred to as the F1 score, assigns equal weights to both Precision and Recall. An abstraction of F score, Fbeta score controls the balance of Precision and Recall in equation 10 with a parameter $\beta$. Hence, equation 10 can be generalized as [105]:

$$Fbeta = \frac{(1 + \beta^2) * Precision * Recall}{\beta^2 Precision + Recall} \tag{11}$$

A value of $\beta$ higher than 1 indicates that higher importance is given to Recall during the F score calculation, and a value of $\beta$ lower than 1 is assigned to give more importance to Precision.

## 6.2. Receiver Operating Curve (ROC)

Receiver Operating Curve (ROC) is a two-dimensional graphical representation of the classification results, where the X-axis represents False positive rates (FPR) and Y-axis the True positive rate (TPR). ROC curve helps combining the individual measures of both the positive and negative classes, which allows understanding of how good the classification performance for both the classes is [78]. It depicts the relative trade-off between the FPR and TPR [34, 116]. FPR and TPR can be calculated from the elements of the confusion matrix as:

$$FPR = \frac{FP}{FP + TN}$$

$$TPR = \frac{TP}{TP + FN}$$

With classifiers resulting in a probabilistic score for a sample belonging to a class, the class for the sample can be determined by comparing the score with a threshold value. If the probability score is above the threshold value, it can be classified as Positive and as Negative otherwise. By varying the threshold value, the class prediction can be changed, resulting in a new confusion matrix. Therefore, each threshold value can generate its own set of TPR and FPR, which can be linked to creating a ROC curve. Conceptually, we can experiment by varying the threshold value from $-\infty$ to $+\infty$ [34, 115]. In practice, the threshold value is varied from the highest to the lowest probabilistic score obtained for the Positive class [38]. An example of a ROC curve can be seen in Figure 1.

For an ideal classification model, one would expect to have TPR = 1 and FPR = 0, which would push the ROC curve to be more on the left corner in Figure 7. A model that is not better than a random guess would have the ROC curve residing along the main diagonal, which connects the point (TPR = 0, FPR = 0), where all the samples are predicted as Negative class, with (TPR = 1, FPR = 1), where all the samples are predicted as Positive class. A model with a ROC curve residing along the main diagonal will always have TPR = FPR for every threshold value. Any model for which the ROC curve lies below the diagonal is considered to be worse than random guessing [34, 116, 115].

The ROC curve provides a quick summary of the performance of a classification model. However, it can be difficult to compare multiple classification models based on the ROC curve. ROC curves for different models can overlap, and unless there is one curve above all the other curves over the

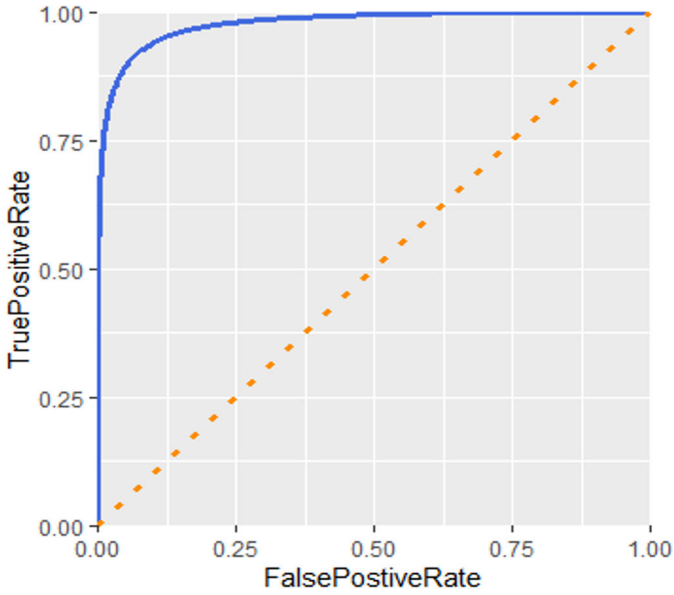


**Figure 7:** *ROC Curve*

entire space, it is hard to claim one model as the best. Therefore, to represent the ROC curve as a single scalar value, the area under the ROC curve (AUC) can be calculated, which estimates the model's performance. With AUC score, performance of multiple classification models can be compared across a range of thresholds [38, 115, 116, 34].

AUC score is calculated by adding the areas of trapezoids below the ROC curve. The AUC value always lies between 0 and 1, with an ideal model having an AUC score of 1, and a model equal to random guessing having an AUC score of 0.5. Hence, a good model should have an AUC score above 0.5 and closer to 1, and any model having an AUC score below 0.5 is considered unusable [115, 116]. AUC score can also be interpreted as the model's ability to rank Positive examples before Negative examples. A higher AUC score indicates that the model assigns a higher probabilistic score to Positive examples compared to the Negative examples [34, 38].

## 6.3. Precision Recall(PR) Curve

Precision-Recall(PR) curve follows a similar approach to the ROC curve for evaluating binary classification models' performance. PR curve captures the model's performance at a range of thresholds, which shows the

trade-off between Precision against Recall. Changes in the class predictions can be achieved by varying threshold values, resulting in different Precision and Recall values for each threshold. A PR curve is created by plotting Recall on the X-axis against the Precision on the Y-axis for the corresponding threshold values [15, 34, 116]. An example of a PR curve can be seen in Figure 8.

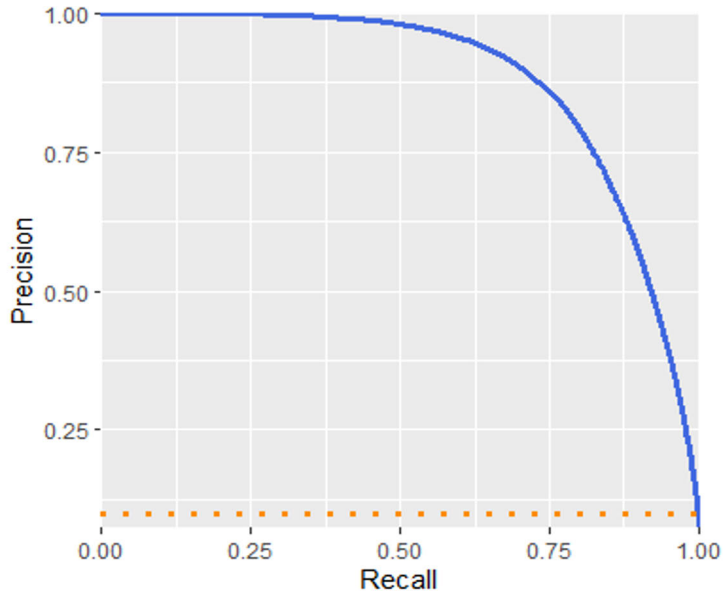


**Figure 8:** *PR Curve*

An ideal classification model with respect to Precision and Recall would have $Precision = Recall = 1$, where the model can completely separate positive classes from Negative classes. Therefore, in Figure 8, the ideal point on the curve would be at the point (1,1), and when the curve is more to the right corner, it indicates the better performance of the model [116]. Since the PR curve deals with Precision and Recall, and they are metrics related to summarising the model's performance for the Positive class, the PR curve neglects the model's capability of handling the Negative class [101]. PR curves are a better alternative over ROC curves in the presence of imbalanced class distribution, where PR curves account for the imbalanced class distribution, which is neglected when using the ROC curve [38].

Similar to the ROC curve, a PR curve can be summarised into a single scalar value by calculating the area under the PR curve (PR-AUC). Its value also ranges between 0 and 1, where 1 is the optimal value. However, unlike

the AUC score, where for a random model, the AUC score is 0.5, the PR-AUC score for a random model is the ratio of Positive examples in the data [38]. The PR-AUC score for a relatively good model should be above the ratio of Positive examples in the data. For example, if the proportion of Positive examples in the data is 0.1, a good model should have a PR-AUC score greater than the value 0.1. The dotted line in Figure 8 represents the model with no skill, a model that classifies every sample to be in the Positive class.

# 7. Previous Studies in Peer to Peer Lending

Following the rapid growth of P2P lending, it has gained a lot of attention as an alternative market. With easy access to the loan transaction data, the number of studies focusing on analyzing credit risk in P2P lending is abundant. This growth across the globe has resulted in an increasing number of P2P lending and credit risk studies. In this section, some of the studies are discussed to present the general overview of P2P lending credit risk analysis studies.

## 7.1. Exploratory Analysis Approaches

Initial studies in P2P lending for credit risk analysis were performed as exploratory analysis to understand the behavior patterns of borrowers and lenders and their effects on the loan transactions. Most of these studies applied basic statistical approaches in performing the analysis.

Klafft (2008) [56] conducted a study on data from a U.S P2P lending platform, Prosper[6]. Using exploratory analysis on the loan performance data, Klafft demonstrated that lenders could benefit from their investments with some simple selection criteria based on the financial information of borrowers. A study performed by Iyer et al.(2009) [50] discover that lenders are partly able to infer borrowers' creditworthiness from the information provided, mainly relying on standard financial information. In addition, for high-risk groups, lenders also make use of non-financial information for risk evaluation.

Puro et al. (2011) [102] studied the investment behavior of lenders in the lending platform 'Prosper', where they found that lenders follow different bidding strategies. They identified nine different strategies being practised by lenders, out of which three of them were most commonly used, namely evaluator strategy, late-bidding, and multi-bidding. Analyzing the lenders' behavior from a Korean P2P lending platform with multinomial logit market-share model, Lee and Lee (2012) [67] observed the presence of strong herding behavior in lenders, where most lenders tend to bid on loans with a high number of bids. They further discovered that the herding behavior had a diminishing marginal effect on participation, where lenders would be attracted more to bids with high participation rates in the early stage compared to the later stage.

From the study of text description used by borrowers to persuade lenders,

[6] https://www.prosper.com/

Larrimore et al. (2011) [65] identified that the use of extended narratives with quantitative words describing financial situation were more likely to secure loans. Applying Linguistic Inquiry and Word Count (LIWC) software, they further discover that irrelevant information in displaying the potential to repay and justifying the current financial situation is negatively related to funding success.

Lin et al. (2011) [76] studied the impact of social networks on a loan being funded and default rates, where social networks were seen to positively affect the funding of a loan and low default rates. Furthermore, they identify social networks in P2P lending to be an additional source of soft information for risk evaluation in the absence of hard information. Chen et al. (2014) [22], based on trust theories, created a trust model for P2P lending to understand the factors building the lenders' trust in borrowers. They applied structural equation modeling with the data from a Chinese P2P lending platform to evaluate the trust model. The trust model reveals that trust in borrowers and lending platforms significantly affects the lenders' decision to lend, where trust in borrowers is more crucial. In addition, the quality of borrowers' information is the most driving factor for gaining lenders' trust. The service quality and credit security from the lending platform are influential in building trust on the platform.

## 7.2. Statistical and Machine Learning Approaches

With many borrowers selecting from a large number of investments, the main task for lenders in P2P lending is precisely selecting the borrowers with low credit risk. The selection of borrowers becomes very critical given the high credit risk in P2P lending. Hence, in recent times, many studies have focused on developing credit risk models that accurately evaluate the default risk of borrowers in P2P lending, applying credit scoring. Compared to the traditional credit scoring of banks, credit scoring in P2P lending can be challenging due to the high-dimensional data, the diversified form of data and large amount of data [73]. There can be seen a wide variety of approaches to credit risk modeling in P2P lending, from standard statistical methods for credit scoring to machine learning methods, and recently deep learning.

Traditional statistical models for credit scoring, including linear regression, linear discriminant analysis, and logistic regression, which are commonly used in P2P lending credit risk evaluation. Linear regression was used for instance in the study by Mild et al.(2015) [86] for developing a decision support tool to estimate the repayment ratio of a loan. With the

tool, lenders can select a subset of the most profitable loans based on the estimated repayment ratio. In a study by Emekter et al.(2015) [32], logistic regression is used to predict the default risk of borrowers, using data from the P2P platform 'LendingClub'. The logistic regression model and some non-parametric tests identified important features influencing borrowers' credit risk, including credit grade, FICO score, debt-to-income ratio, and revolving line utilization. In addition, the Cox Proportional Hazard model shows that with an increase in the maturity period of loans, credit risk also increases.

Machine learning models in recent times have been a popular choice for credit scoring in P2P lending, with wide ranges of algorithms being tested to produce optimal results. Multiple studies have performed a comparative analysis of machine learning algorithms, including traditional statistical models. Random forest was seen to outperform other models in predicting borrowers' risk in a study by Malekipirbazari and Aksakalli (2015) [81]. Random forest had better performance than logistic regression, support vector machines, and K-nearest neighbor. It also performed better than FICO scores and LendingClub's credit grade in credit risk prediction. Comparison study of machine learning models by Chang et al. (2016) [19] identified Naive Bayes with Gaussian distribution as the winner. It had better performance in predicting default risk compared to logistic regression and support vector machines.

To address the problem of imbalanced data and misclassification costs in credit scoring, Xia et al. (2017) [132] applied cost-sensitive learning. A cost-sensitive boosted tree model using extreme gradient boosting (XGBoost) was applied to predict the risk and return from P2P loans. The constructed model was able to outperform other methods, such as logistic regression and Random forest, when tested on two P2P loan datasets.

While most studies only used hard financial information for credit risk evaluation, Jian et al. (2018) [52] combined both soft and hard information for credit risk evaluation of P2P loans in China: abstracting loan-related features with topic modeling from text description provided by borrowers combined with available features for creating a default prediction model. Logistic Regression, Naive Bayes, Support Vector Machine, and Random Forest were used in creating the models. With multiple combinations of feature sets for modeling, the authors observed that the combination of soft and hard features increased the performance for default prediction, where Random Forest performed the best. In a study by Niu et al. (2019) [93], they combined borrowers' social network information with regular features for creating a credit scoring model with Random Forest, AdaBoost,

and LightGBM. They extracted social network information from mobile phone data of borrowers in P2P lending in China. With the combined features, the results indicated that including social network information improved the results.

Ensemble approaches have been regularly applied to boost the prediction performance. A heterogeneous ensemble framework was applied by Li et al. (2018) [73] for default risk prediction of P2P borrowers in China. Their approach included a linear weighted combination of Extreme gradient boosting model, Deep Neural Network, and Logistic Regression. The results show that the model is able to handle high-dimensional and imbalanced data problems. Zhou et al. (2019) [135] applied a linear weighted ensemble of decision tree-based models to predict the default probability of individual loans in P2P lending. The ensemble model consisted of Gradient boosting decision trees, Extreme gradient boosting, and light gradient boosting machines. Their method outperformed other existing machine learning methods on the same dataset.

Neural networks have been one of the popular models for credit scoring in P2P lending, with the standard multilayer perceptron form of Deep learning widely applied. Wang et al. (2018) [127] implemented LSTM (Long term short-term memory) Deep learning framework to develop a credit scoring model for P2P lending. Analyzing the sequence of events of borrowers in online operation, the authors construct an Event2Vec Model to represent the events in a vector space. Finally, the LSTM network is used for predicting the probability of default from the extracted features.

In addition to credit scoring for default prediction in P2P lending, profit scoring is another common approach used for evaluating credit risk in P2P lending. Using the internal rate of return (IRR) as the measure of profit, Cinca and Nieto (2016) [108] applied a Decision tree model to create a profit scoring model. They compared the results with the credit scoring model and observed it to perform better in risk assessment. Following a similar approach, Bastani et al. (2019) [10] used Deep learning for profit scoring. They implemented two-stage modeling, where a credit scoring model was used in the first stage to predict default probability. The loans that passed the first stage were then sent through a profit scoring model.

# 8. Data Collection

In this chapter, the data used in the experimental research are briefly presented. The data used in the research were accessed through the publicly available sources of the P2P platforms. Three different P2P platforms were selected based on ease of access to the data and their popularity.

## 8.1. Bondora

Bondora[7] is one of the leading P2P platforms in Europe. Established in 2009, Bondora currently operates in Estonia, Finland and Spain. With 845,139 customers, as of May 24, 2021, Bondara has issued P2P loans that totals to about 435.145 million euros [8]. Two different data sets, from the primary market and secondary market, were collected from Bondora platform[9].

Data from the primary market was used in the experimental research for estimating returns from P2P lending, presented in Paper 1. The data consisted of loan records between 28 February 2009 and 4 October 2018 with 65,675 total loan records. For the analysis purpose, only the loan records from 2013 onward were utilised as there were no ratings assigned to loans by Bondora before 2013. The data includes 112 features that provide demographic and financial information on loans and their current performance. A sample of the features is shown in Table 5

The 'Status' feature describes the current status of a loan that indicates whether the loan is either defaulted, currently in the payment process or repaid. Out of the total loan records, 36.5% were defaulted, 41% were current loans, and 21.7% were repaid, which shows a high default rate. A summary of the loan status for the loans issued across the years is shown in Table 6. Table 6 also shows the base interest rate (taken as average for A-rated loans) to be decreasing in recent years.

The secondary market data includes the information on loan notes available from March 2013 to July 2019. The secondary market data set was used in Paper 3 to study the investment decisions in P2P secondary market. There are approx. 7.3 million records of loan notes in the secondary market data set. The large number of loan records is due to loan splittings into multiple small loan notes when the loans are put in the secondary market. The loan notes have a unique loan id that connects them to the loan infor-

[7] https://www.bondora.com/en

[8] https://www.bondora.fi/en/about-bondora/

[9] https://www.bondora.com/en/public-reports.

| Features | Description |
|---|---|
| LoanID | Unique loan identifier |
| NewCreditCustomer | Does customer have prior credit history with Bondora |
| VerificationType | Method used for loan application data verification |
| AppliedAmount | Total loan amount applied |
| MonthlyPayment | Monthly payments to be made |
| LoanDuration | Loan duration in months |
| Interest | Interest rate on loan |
| UseOfLoan | Purpose of the loan |
| MaritalStatus | Marital status of borrower |
| EmploymentStatus | Current employment status of borrower |
| HomeOwnershipType | Type of home ownership pf borrower |
| IncomeTotal | Total income of borrower |
| Rating | Rating level assigned by Bondora |
| DebtToIncome | Ratio of borrower's monthly gross income that goes toward paying loans |
| LiabilitiesTotal | Total monthly liabilities of borrower |
| Status | Current status of the loan |

**Table 5:** *Bondora Feature Sample for Primary Market*

| | Year | | | | | |
|---|---|---|---|---|---|---|
| | 2013 | 2014 | 2015 | 2016 | 2017 | 2018 |
| **Current** | 182 | 775 | 1436 | 2533 | 8199 | 14200 |
| **Repaid** | 1447 | 2416 | 2139 | 2539 | 2568 | 1081 |
| **Default** | 846 | 3942 | 4471 | 5441 | 7166 | 1209 |
| **Interest(%)** | 24.95 | 24.77 | 16.22 | 12.40 | 11.84 | 11.76 |

**Table 6:** *Bondora Loan Summary Statistics*

mation from the primary market. In addition, some other features provide additional information on loan notes, which are shown in Table 7.

| Features | Description |
|---|---|
| LoanID | Unique loan identifier |
| StartDate | Time when the investment was added to Secondary Market |
| EndDate | Time when the investment was sold or removed from the Secondary Market |
| DiscountRate | The discount/premium set by the seller |
| DebtDaysAtStart | The number of days the loan had been in debt at StartDate |
| DebtDaysAtEnd | The number of days the loan had been in debt at EndDate |
| PrincipalAtStart | Outstanding principal at StartDate |
| PrincipalAtEnd | Outstanding principal at EndDate |
| LoanDuration | Loan duration in months |
| Interest | Interest rate on loan |
| UseOfLoan | Purpose of the loan |
| Result | If the loan was traded in the secondary market or not |

**Table 7:** *Bondora Feature Sample for Secondary Market*

## 8.2. Prosper

Prosper is the first P2P lending platform in the United States that was established in 2005. It has facilitated more than $18 billion in loan amounts to more than 1,100,000 borrowers[10]. Data accessed from Prosper was used in Paper 2, where segmented modeling was applied to generate risk-specific decisions for credit risk evaluation. The data includes loans issued between 2005 and 2014: 113937 loans with 81 features. For the analysis purposes, the data was processed with multiple data processing steps to obtain the final data size of 55084 loans and 21 features. Some of the features used in the analysis are presented in Table 8.

The default rate in the data was 30.8%, with 69.2% non-default loans. The loans are assigned a rating class, with AA is the best rating and HR is the worst(risky) rating. A summary of loan statistics in accordance with the ratings class is shown in Table 9

[10] `https://www.prosper.com/about`

| Features | Description |
|---|---|
| LoanStatus | Status of the loan |
| BorrowerRate | Interest rate on the loan |
| ProsperRating | Rating assigned to the loan |
| OpenRevolvingAccounts | Number of open revolving accounts |
| InquiriesLast6Months | Number of inquiries in past 6 months |
| PublicRecordsLast10Years | Number of public records last 10 years |
| DebtToIncomeRatio | Debt to income ratio of borrower |
| creditscore average | Average of lower and upper credit scores |
| Investors | Number of investors that funded the loan |
| AvailableBankcardCredit | Total available credit via bank card |
| ListingCategory | Purpose of loan |
| CurrentCreditLines | Number of credit lines |

**Table 8:** *Prosper Feature Sample*

| Prosper Rating | Loan Counts | Default Rate(%) | Average Interest(%) | Average Credit Score |
|---|---|---|---|---|
| AA | 6922 | 18 | 12 | 746 |
| A | 5292 | 12 | 9 | 796 |
| B | 7762 | 26 | 16 | 710 |
| C | 9465 | 30 | 19 | 678 |
| D | 11022 | 34 | 24 | 660 |
| E | 7199 | 41 | 28 | 627 |
| HR | 7230 | 49 | 29 | 607 |

**Table 9:** *Prosper Data Summary*

## 8.3. LendingClub

LendigClub established in 2007, is a leading P2P platform in the United States. The data accessed from LendingClub includes loans issued in the year 2015. Including only either charged-off or fully paid loans, the data set contains 383,735 loan transactions, where 80% were fully paid and 20% were charged off. The data from LendingClub was utilised in Paper 4, where portfolio optimization is performed with P2P loans for investment decisions. Two different sources of information on the loans were utilised.

The first source provided information on loan and borrower characteristics used in constructing the credit scoring model for default prediction. A sample of the feature set is shown in Table 10 and a summary statistics of the loans is presented in Table 11. The second source of information provided data on monthly payments made by borrowers on the loans, which includes information such as monthly principal, interest, and late fee payments. Using the monthly payments information, Internal Rate of Return on the loans were calculated.

| **Features** | **Description** |
|---|---|
| funded_amnt_inv | Amount funded by investors |
| int_rate | Interest rate on the loan |
| annual_inc | Annual income of borrower |
| dti | Debt to income ratio of borrower |
| delinq_2yrs | Number of account deliquent in past 2 years |
| revol_bal | Amount of Revolving balance |
| term | Loan period in months |
| grade | Rating assigned to loan |
| emp_length | Length of current employment |
| fico_score | FICO score |
| pub_rec_bankruptcies | Any record of bankruptcies |
| loan_stauts | Status of the loan |

**Table 10:** *Lendingclub Feature Sample*

| Grade | Count | Default Rate(%) | Interest Rate(%) | FICO Score | Debt to Income(%) |
|---|---|---|---|---|---|
| A | 72133 | 6 | 6.93 | 721 | 16.28 |
| B | 108765 | 13 | 10.04 | 696 | 17.98 |
| C | 107410 | 22 | 13.29 | 688 | 19.62 |
| D | 55178 | 32 | 16.72 | 685 | 21.30 |
| E | 29762 | 41 | 19.29 | 684 | 21.56 |
| F | 8530 | 50 | 23.63 | 683 | 21.47 |
| G | 1957 | 54 | 26.84 | 681 | 20.14 |
| **Total/Avg** | **383735** | **20.21** | **12.43** | **695** | **18.96** |

**Table 11:** *LendingClub Data Summary Statistics*

# 9. Research Methods and Approaches

This chapter presents the research approaches implemented in accomplishing the research objectives with the focus on credit risk evaluation in P2P lending and loan selections for investment. The detailed implementation of the research methodologies and the experiment results are presented and discussed in the corresponding published research papers. This chapter summarises the research papers and briefly discusses the results.

## 9.1. Estimating Returns from Peer-to-Peer Loans

Paper 1: "Predicting Expected Profit in Ongoing Peer-to-Peer Loans with Survival Analysis-Based Profit Scoring"

Lenders in P2P lending are on the higher end of risking their investments due to the presence of high credit risk. The high credit risk in P2P loans is compensated by allocating higher interest rates to the loans in high-risk categories. However, the higher interest rates may not be sufficient to compensate for the loans' default risk fully. Hence, the combination of interest rate and default risk can be used to calculate the expected return or profit from a loan. As for lenders, their ultimate goal is to receive higher returns from investments. Therefore, estimating the profit in the presence of default risk gives a broader perspective of return on investments to lenders.

Few studies have applied profit scoring to predict the profit on a loan in P2P lending, with IRR (internal rate of return) used to measure profit. The major drawback of these studies has been the use of selective data sets to include only completed loans in the modeling. While the approach helps estimating the profit, however, there is a high possibility of introducing bias in the analysis as they may leave out loans that are more likely to survive with a good return. In addition, with P2P lending experiencing significant changes in interest rates and credit ratings, analysis performed on only a subset of historical data may not provide accurate predictions. For example, when we take only completed loans having a loan period of 4 years, we would be using the information on interest rates and risk from 4 years back, which may not be relevant at present due to rapid changes. Hence, to overcome this issue, survival analysis can be introduced in the modeling, which can include both repaid and ongoing loans in the analysis. Survival analysis allows for modeling with more recent information and thus, provides accurate predictions.

The primary usage of survival analysis in the modeling is to predict an

event's occurrence in a given time frame. Hence, it is applied to predict the event of default for loans at every month in the payment period. This monthly default probability explains the probability that a loan will default in an interval between loan payments, where the loan has survived up to the interval. With the features of a loan, the default probability is calculated using Cox proportional hazards model. Similar to default probability, profit at each monthly payment is estimated with a simple estimation formula:

$$i = (1–h)I + hD \tag{12}$$

where $h$ is the monthly default probability, $I$ is the interest rate and $D$ is the loss given default.

At every monthly payment time, a loan either survives or defaults (the outcome is given as 0 or 1), as modeled by the monthly default probability $1 - h$ and $h$. If a loan survives, the investors make a monthly profit of interest $I$ on the remaining principal. If the loan defaults, the investor's loss is quantified by the loss given default value $D$. Hence, with this formulation of default risk and profit calculation, the profit is estimated at each monthly period to give an unbiased estimation in the presence of censoring.

The proposed model is experimented using a dataset from the P2P platform 'Bondora'. The predicted default probabilities and the profit are evaluated with mean squared error (MSE) metric. In addition, the AUC score is calculated for the default probabilities. The results obtained show the model's good performance with low MSE values and a good AUC score, as seen in Table 12.

| | MSE Default | MSE Profit | AUC Default |
|---|---|---|---|
| Train Set | 0.0288 | 0.0170 | 0.7114 |
| Train Set | 0.0292 | 0.0167 | 0.7098 |

**Table 12:** *Evaluation Results*

Furthermore, comparing the results of expected profits with the interest rates shows a significant difference in the profits. As promised to the lenders, interest rates would be the profits if the loans did not default. However, in the presence of risk, the expected profits are very different compared to the original interest rates, as it can be observed in Figure 9. The difference can also be explained by the profit formula used in equation 12, which shows a direct relationship between default risk and interest rate. Thus, the results clarify that the return in P2P lending is not as high as advertised and careful selection of loans is necessary to avoid any loss.

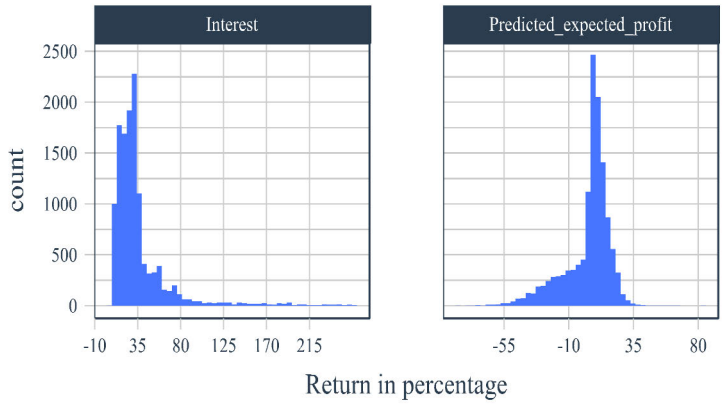


**Figure 9:** *Histogram of interest and estimated profit rates on test set*

To analyze the effect of a proposed model on the portfolio return of the lenders, a comparison is performed on different strategies for loan selection. The results are presented in Figure 10. As it can be seen from Figure 10, selecting loans based on the proposed profit model offers a higher return compared to selecting loans based on credit ratings and default probability. Therefore, these results justify the proposed profit model's effectiveness to guide P2P lenders in investment decisions for better loan selection.

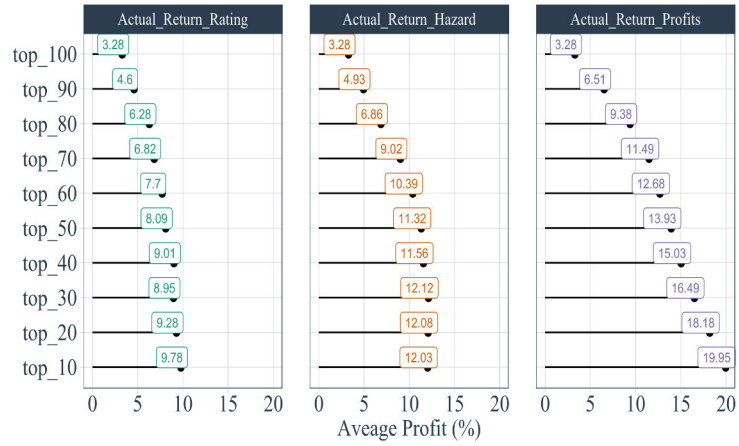


**Figure 10:** *Average Portfolio Return*

## 9.2. Loan Selection Decision in P2P Lending with Segmented Modeling

Paper 2: "Improving Credit Risk Analysis with Cluster Based Modeling and Threshold Selection."

While credit scoring is widely used as the standard tool to evaluate the credit risk of borrowers, it is also a common practice to deploy a single credit scoring model to cover all applicants. However, with the growth in credit borrowers, their characteristics are diversified. In the presence of varying nature of the borrowers, a single credit scoring model may not accurately capture such diverse feature sets. This issue can be handled with segmented modeling, where borrowers are segmented to multiple segments based on their similarities, and a new credit scoring model is trained for each segment. The procedure allows for granular risk evaluation as credit scoring models can be more specific, where borrowers show similar behavior. This approach has been successfully applied in the studies for improving results on credit evaluation.

Credit scoring as a binary classification applies a threshold or a cutoff score to classify borrowers as 'good' or 'bad' depending on the likelihood of belonging to the classes. As with a single credit scoring model, a single threshold may not cover the diversified risk of applicants. Borrowers in low-risk groups have a lower default probability when compared to higher-risk groups. The difference in the risks creates a problem in selecting a perfect threshold to classify accurately. Hence, selecting a separate threshold for each segment, considering the risk levels, can elevate the classification accuracy. Furthermore, different strategies can be applied for threshold selection. In these types of cost-sensitive problems, misclassification cost is a suitable measure to decide on the threshold. It accounts for different costs related to False Negative and False Positive errors, with False Negative errors having higher costs.

The segmented modeling with threshold selection for improved results in credit risk identification was implemented on a P2P lending dataset from 'Prosper'. Unsupervised learning with KMeans was first used for segmenting data into segments of similar borrower characteristics, displaying similar risk levels. Applying three different machine learning models (Logistic Regression, Random Forest, Gradient Boosting Model), credit scoring models were trained for the entire dataset and separate models for each of the segments. With models being trained for the full dataset and segments, threshold selection is performed with the relative cost of misclassification

specified by the formula:

$$RC = \alpha(P_I C_I) + (1 - \alpha)(P_{II} C_{II}) \tag{13}$$

The probability of belonging to the bad class is given by $\alpha$. $P_I$ is the probability of being False Negative and $C_I$ is the relative cost of a False Negative. Similarly, $P_{II}$ is the probability of being a False Positive and $C_{II}$ is the relative cost of a False Positive.

When measuring the relative cost of misclassification with multiple values of threshold for the segments, a clear distinction of the best threshold value was observed for the segments as seen in Figure 11.

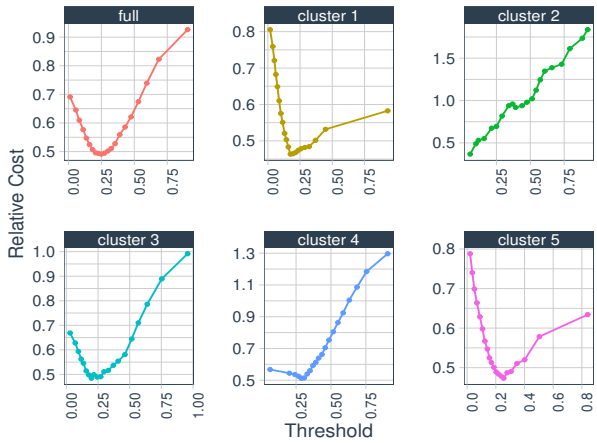


**Figure 11:** *Threshold Selection*

The threshold (which here is referred to as probability of belonging to a bad class or 'default probability' cutoff) at which the model achieves the lowest relative cost of misclassification is selected to be the cut-off point for classifying borrowers. In addition, to see the effect of relative costs on threshold selection, multiple relative costs values were tested, with the results presented in Figure 12.

In Figure 12, the summary of the threshold optimization process with multiple cost ratios for the segments and full data is presented. The figure shows the best performing model for the segments with a corresponding threshold value and the relative cost of misclassification at different cost ratios. The results show that the optimal threshold for the segments is different at varying cost scenarios. Segment 2 has a very low relative cost compared to other segments, and this observation can contribute to overall cost reduction. In addition, the threshold value is seen to decrease with

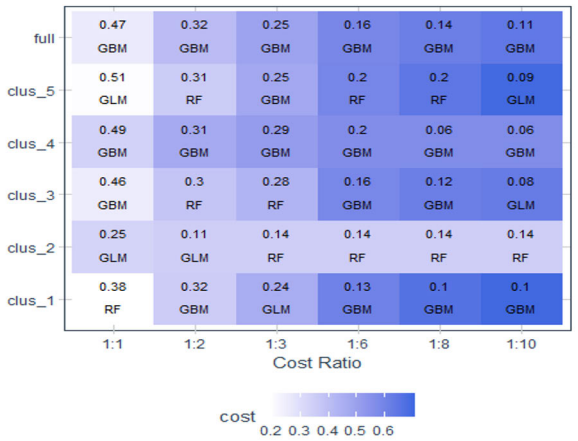


**Figure 12:** *Threshold Optimization Summary*

an increase in the cost for False Negatives, and segment 4 can be seen to have a very low threshold at higher costs scenario, which suggests it constitutes high-risk loans.

Finally, when comparing the average relative cost across the segments at the best thresholds with the full dataset, there is an improvement in reduced relative cost as presented in Table 13. The reduction in the relative cost shows a clear improvement of risk analysis with segmented modeling and threshold selection.

| Cost Ratios | Full Data | Segment Average | Reduction (%) |
|---|---|---|---|
| 1:1 | 0.256 | 0.261 | -1.7 |
| 1:2 | 0.406 | 0.379 | 6.51 |
| 1:3 | 0.491 | 0.445 | 9.46 |
| 1:6 | 0.587 | 0.539 | 8.2 |
| 1:8 | 0.617 | 0.568 | 7.96 |
| 1:10 | 0.639 | 0.587 | 8.19 |

**Table 13:** *Relative cost comparison*

## 9.3. Analyzing P2P Lending Secondary Market to study Investment Decisions

Paper 3: "Analyzing Peer-to-Peer Lending Secondary Market: What Determines the Successful Trade of a Loan Note?"

Many P2P lending platforms provide a secondary market, where investors can resell their loan holdings. In doing so, investors can split the loan holdings into multiple smaller loan notes and can place them in the secondary market with discounts or premiums. The main reason investors resell their loan holdings is the problem of recovery from borrowers. This makes trading loans in the secondary market riskier compared to the primary market. However, splitting loans into multiple smaller loan notes generates abundant data to study the investors' behavior, further providing plenty of options for investments and making it computationally challenging to apply machine learning. While there have been many studies performed with primary market data from P2P lending platforms, there are very few studies utilizing secondary market data. Therefore, with loan transactions data from the secondary market in the P2P platform 'Bondora', we study how to improve the selection of borrowers and loan characteristics for investment in the secondary market.

By performing an extensive exploratory analysis on the data, two features, 'Discount Rate' (discount or premium offered in the loan note) and 'Days in Debt' (number of days the loan has been in debt), were identified to have a higher impact on a loan note being successfully traded in the secondary market. Further exploring the impact of the two features, it is observed that when both features have a value of 0, the loan notes are likely to be sold in almost all (97%) cases. However, the success rate is very low when both the features are different from 0. The low success rate signifies investors' priority on loan notes that behave as a new loan from the primary market and a low-risk strategy. Keeping aside all the loan notes having both discount and debt days as 0 (as a straightforward rule can be stated to determine their results), further modeling approaches are performed on the remaining data with machine learning.

With the available features on loan notes and extracting additional features about the loan notes from the primary market, machine learning models were trained to create classification models to predict the success of a loan note in the secondary market. The trained classification models had a very good performance with Random Forest, showing the best performance as shown in Table14

| Models | Accuracy | AUC | Logloss | F1 |
|---|---|---|---|---|
| LR | 0.745 | 0.800 | 0.446 | 0.571 |
| **RF** | **0.925** | **0.969** | **0.189** | **0.840** |
| GBM | 0.886 | 0.931 | 0.268 | 0.756 |
| NN | 0.894 | 0.940 | 0.252 | 0.774 |

**Table 14:** *Classification Results*

The two features, discount rate and debt days, were the top two best features for all the trained models when observing the variable importance of the trained models. The two features had very high importance over other features for the best performing Random Forest model. The modeling results matched the results from the exploratory analysis results, identifying discount rate and debt days to be deciding factors to investment decisions in the secondary market. This result was further illustrated with a good performance of the classification models with just two of the features used in the model.

## 9.4. P2P Loan Selection with Portfolio Optimization

Paper 4: "Data-driven optimization of peer-to-peer lending portfolios based on the expected value framework."

While selecting investments in P2P lending, lenders do not need to fully fund a single borrower, they can partially fund a borrower. The partial funding creates a situation where borrowers have multiple investors and lenders having their investments spread over multiple borrowers, diversifying their portfolios. A simple strategy for lenders to create a portfolio would be to select loans from different credit ratings following the desired risk and return. However, the main problem remains as the amount to invest in each of the selected loans to ensure a maximum return while accepting a certain amount of risk. This problem can be addressed by traditional approaches of portfolio optimization, which would help lenders to select loans for creating an optimal portfolio according to their choices of risk and return.

There is an abundance of studies on predicting default probabilities of P2P loans that estimate the credit risk. In addition, few studies attempt to estimate the return from the loans with profit scoring. The downside of these studies has been that they either isolate the risk and return in profit estimation or only estimate profit in the presence of risk for a single loan at

a time. Both approaches may not serve the requirement of lenders, where they seek the overall return from their portfolio with a certain amount of risk. The complete solution would be to treat loan selection as a portfolio optimization process that provides lenders with an estimation of their overall return from investments in the presence of risk.

There is no sufficient historical data on individual borrowers to perform portfolio optimization. Therefore, their similarities with past loans can be used to estimate their performance. This is termed as the instance-based approach. Probability of default has been used as the measure for similarity in past studies. We extend this approach using an untapped method,' Expected Value Framework', as the measure for similarity. It considers potential losses and gains in the presence of risk, offering a broader estimation of return.

The portfolio optimization process is summarized in Figure 13. We first determine the risk (default probability) with a machine learning model applying cost-sensitive approaches. The risk is passed through the expected value framework that gives an estimation of return considering all possible outcomes that can take place. With the estimated return as the similarity measure, a kernel weights approach with Euclidean distance is applied to determine the weights to quantify the similarity of a new loan with historical loans. The risk and return for a new loan are then estimated as the weighted average of risk and return on similar historical loans. The return is represented by IRR (internal rate of return).

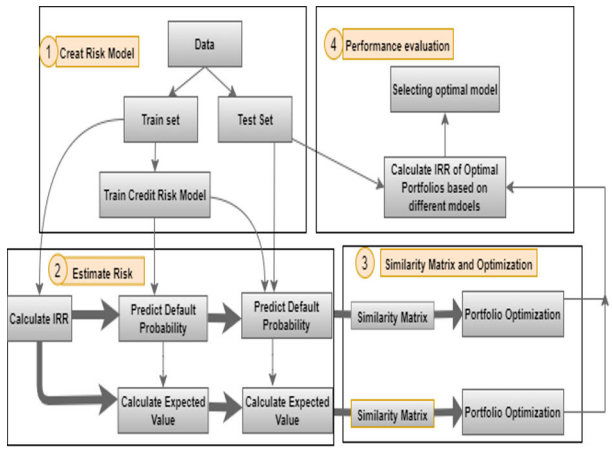


**Figure 13:** *Portfolio Optimization Process*

The portfolio optimization is performed by minimizing the risk with a

required return rate, and also considering the minimum amount of investment required. The portfolio optimization results obtained with the expected value framework for similarity measure are compared to the default probability approach. The results obtained show improvement and better performance. Furthermore, to verify the improvement, sensitivity analysis is performed with varying loan amounts, required return rate, and weight constraints. The results from the sensitivity analysis further highlight the higher performance of using the expected value framework over default probability as a measure of similarity used in the instance-based approach for portfolio optimization. A comparison results from the sensitivity analysis with budget size and weight constraints is presented in Figure 14. The results show consistently high performance using the 'Expected Value Framework' approach over the 'Default Probability'.

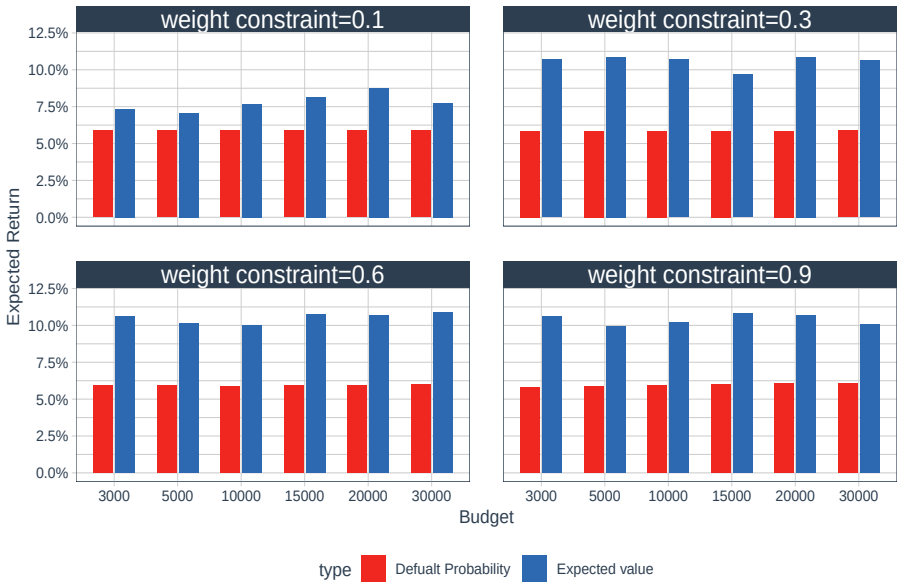


**Figure 14:** *Sensitivity Analysis Results*

# 10. Results and Discussion

This study has focused on credit risk evaluation in peer-to-peer lending to support the lenders in their lending decisions. Credit scoring and machine learning algorithms are applied as the primary methods to apply predictive analytics for estimating default probabilities in P2P loans. Machine learning algorithms have shown consistently higher performance over traditional statistical methods, such as logistic regression for default predictions. Besides predicting default probabilities for risk estimation, return estimation was another main focus of the study. The return estimation is accomplished by incorporating the risk on a loan to give a better and realistic estimation of investment return. In this section, the answers to the research questions proposed in the study are presented.

- **What is the current state-of-the-art in modeling credit risk in P2P lending?**

In addressing the first research question proposed in the study to understand the current state-of-art in P2P credit risk evaluation, an extensive literature review is performed. The summary of the related major findings from the literature review is presented in Chapter 7. The early stages of research studies in P2P lending have mainly focused on exploratory analysis with few statistical methods being applied. The studies primarily explored the behavioral patterns of borrowers and lenders in loan transactions. They relied on simple statistical methods such as linear regression and logistic regression to determine the features that affected funding decisions on the loans. While the early studies in P2P lending did not use predictive analytics and machine learning, they successfully presented the presence of high credit risk in P2P lending. Furthermore, these studies presented a good overview of features and behavior patterns that are most likely to affect a loan funding decision and a loan failure.

The recent studies on credit risk evaluation in P2P lending focus on predicting credit risk, which is mainly based on credit scoring and machine learning algorithms. Multiple variants of machine learning algorithms have been applied to boost the prediction accuracy of the credit scoring models, which also includes ensemble modeling. With the growing popularity of Deep Learning, there has been increasing use of Deep Learning models for credit scoring. Similarly, different approaches to data collection and

feature generation are applied for machine learning models. Some major problem with credit scoring, such as imbalanced data and cost-sensitive issues, have also been addressed in the studies. Profit estimation for P2P loans have appeared in some recent studies, but the contribution is very sparse. From the perspective of supporting lending decisions, profit estimation is a crucial lending criterion. Thus, the lack of substantial studies in estimating the profit provides a significant opportunity for contribution to the field.

- **How can an estimation of return be made from P2P loans to ensure profitable investments in the presence of the credit risk?**

There are only a few studies that have studied profit estimation in P2P lending. The studies primarily use a simple regression approach for profit scoring, considering the Internal Rate of Return as the measure of profit. These approaches do not consider the associated credit risk in a loan during profit estimation. Hence, this study extends the studies for profit estimation in P2P lending by incorporating the credit risk as presented in Paper 1:*"Predicting Expected Profit in Ongoing Peer-to-Peer Loans with Survival Analysis-Based Profit Scoring"*.

A clear and significant difference between the promised return (interest rates) and the estimated return was observed from the study results. The results obtained saw the estimated returns calculated integrating the credit risk to be lower than the return proposed on the loans. The lower estimated returns show that high interest rates promised are not always obtained, and thus, careful consideration needs to be made in investing in such loans. The high interest rates imposed on the high-risk loans do not always compensate for the risk. But when we estimate the return on such high-risk loans by integrating the risk, we get a more realistic estimation of the return. With the more realistic estimation of the return, lenders now can be more confident in loan selection. Furthermore, it can support lenders who seek higher returns from high-risk loans. All high-risk loans do not necessarily end in loss, and careful selection of such loans can result in higher profits.

As presented in the paper, the loan selection made based on the profit estimation proposed provides the highest overall return in comparison to the credit ratings and default probability-based selection. The credit ratings and default probability-based selection only consider the risk, where it

prioritizes low-risk loans and return for low-risk loans are low. The profit estimation-based selection prioritizes both the risk and return, which allows for selecting confidently high-risk loans that assure a higher overall return.

- **How can different risk groups in P2P lending be accommodated in modeling credit risk for precise loan selection decisions?**

With many credit applicants in P2P lending, there is a list of applicants of diverse behavior, contributing to different risk levels. When we deploy a single credit scoring model to estimate risk on such diversified risk behavior of loans, it may fail to give an accurate estimation. The threshold value used to classify loans as 'good' or 'bad' may not be appropriate to classify loans at all levels of risk. Segmentation can be applied to segment borrowers into groups that allow for clustering borrowers with similar risks. An independent credit scoring model can be deployed for each cluster to give classification decisions specific to the risk levels, contributing to overall high-performance classification results. In paper 2: *"Improving Credit Risk Analysis with Cluster-Based Modeling and Threshold Selection"*, segmentation modeling is applied with the cost-sensitive approach for threshold selection to classify loans of different risk levels.

Applying unsupervised machine learning with the KMeans algorithm, the borrowers were segmented into groups. The risk levels in each of the groups were seen to be different. The threshold selection for the classification was applied with a cost-sensitive approach that minimized the relative cost of classification. The segmentation modeling and the threshold selection provided the results with different best threshold values for the segments. The difference in the best thresholds signifies the applicability of segmentation modeling for better loan selection. In addition, the threshold selection from segmented modeling achieved overall improvement in the classification results that reduced the misclassification costs.

Using the segment-specific threshold for the classification resulted in lower misclassification costs than using a single threshold for each segment. Furthermore, it significantly improved the performance for some segments that contributed to the overall performance enhancement. Therefore, with specific decision criteria for each risk group, the loan selection becomes fitting in selecting good loans. In addition, with a cost-sensitive

approach applied for the threshold selection, the loan selection ensures the low risk of investment loss as it is trained to penalize misclassification with higher costs.

- **What strategies and modeling approaches can be applied to create a profitable portfolio of P2P loans with rational allocation of investments?**

Individual loan selection criteria can guide lenders in P2P lending for making rational investment decisions with the estimation of risk and return. However, as lenders in P2P lending make partial funding on the loans and invest on multiple loans, the amount to invest on a loan also becomes equally important along with the selection of the loan. Therefore, combining both the objectives of loan selection and investment amount can give better decision support to lenders in estimating their overall return from their investments. This objective of creating a profitable portfolio of loans with rational allocation of investments is performed with the paper 4: *"Data-driven optimization of peer-to-peer lending portfolios based on the expected value framework"*.

Paper 4 extends the previous studies to portfolio optimization for P2P loans that use an 'instance-based' approach. The previous studies' extension is performed by applying the 'Expected Value Framework' to measure the similarities of loans with the past loan performances, where default probability was used in the previous studies. The use of the 'Expected Value Framework' over the default probability for the similarity measure resulted in better portfolio optimization performance. For similar requirements of risk and return, the results with the Expected Value Framework were better than the default probability approach. The improvement in the results was further tested and confirmed with extensive sensitivity analyses.

The portfolio selection of the loans provides the lenders in P2P lending a better selection strategy compared to the individual loan selection. It reduces the burden on the lenders in putting extra effort into analyzing a large number of loans individually for investments. The less effort in risk analysis can be more effective and convenient to lenders who are primarily non-professionals. The portfolio selection also considers the lenders' risk and return requirements that allow for flexibility in making investment decisions. Furthermore, the portfolio selection most importantly guides the

lenders to make rational decisions on spreading their investments. It allows the lenders to rationally divide their investment on multiple borrowers to provide back the desired return of their choice.

# 11. Conclusion

Credit risk has been an imperative part of the financial industry that significantly affects financial success. Credit risk implies the probability of a financial loss as a result of failure to recover the debts. The financial loss can be significant resulting from failed debts, and therefore, credit risk receives very high importance in the decision making process. Besides being highly important, it is also a risk type that is difficult to manage. The growth in the credit applicants also increases the number of credit defaults, resulting in high amounts of losses. Hence, to prevent the loss, it requires careful selection of the loan applicants through the credit risk evaluation process. Credit risk management is applied by financial organizations to tackle the problem of credit risk that identifies the risk and the necessary treatments to eliminate the loss from the risk. Credit risk management provides decision support to making quick and accurate credit decisions with the help of efficient tools and techniques for risk measurement.

Alternative markets have been established as a good source of credits to small borrowers with Peer-to-Peer lending being one of the popular alternative markets. P2P lending creates an online market that connects borrowers and lenders directly through an online platform. The significant differences between P2P lending and traditional financial services are the absence of collateral and costly financial intermediaries. Following online and automated processes, P2P lending enables small borrowers to access credits at less time and low cost. P2P lending favours small borrowers that are placed at the lower end of the credits. Lenders are attracted to P2P lending due to the high return advertised compared to similar investments. Lenders can partially fund a borrower that allows them to spread their investments and diversify their portfolio. In recent time, P2P lending has been growing rapidly and gaining popularity among small borrowers and investors.

P2P lending also experiences high credit risk, which is primarily due to the lack of collateral. P2P lending platform acts as a simple intermediary for loan transactions, and as such, the credit risk is mainly imposed on the lenders. Information asymmetry and lack of analytical skills of lenders are also the main reasons for credit risk in P2P lending. P2P lenders, who are primarily non-professional individuals, face difficulties in loan selection due to low skills in risk evaluation. The high credit risk in P2P lending creates the necessity to perform robust credit risk evaluation for loan selection. With the objective of studying credit risk evaluation in P2P lending, this study focused on implementing multiple approaches to risk identifica-

tion and loan selection for profitable investments in P2P lending.

Credit scoring has been established as a prominent tool for credit risk evaluation in the financial industry. The primary use of credit scoring is to differentiate between the risky and non-risky loan applicants to help in loan granting decisions. While statistical methods are commonly used for creating credit scoring models, machine learning algorithms are being preferred over the traditional statistical methods due to the high performance of machine learning algorithms. Multiple studies show the successful application of credit risk evaluation in P2P lending with machine learning and credit scoring. The studies have primarily focused on the default predictions of the loan applicants in P2P lending. Multiple machine learning modeling and data preparation approaches have been applied to create a credit scoring model for predicting the default probability.

Besides the credit risk, lenders in P2P lending are concerned about profits as high return is one of the major factors that attract the lenders. There have been very few contributions to profit scoring or profit estimation in P2P lending. Therefore, this study extends the studies in profit estimation for P2P loans for better loan selection decisions. By incorporating the credit risk for estimating the profit, a more informative estimation of profit is achieved for P2P loans, highlighting the difference between promised and the estimated return. The profit estimation in the presence of the credit risk was successful in making loan selections for higher portfolio return.

The study also addressed the diverse behavior of P2P lending borrowers and divergent credit risk levels with segmentation modeling for better loan selection decisions. The segmentation modeling helped select the risk-level specific decision boundary in the form of thresholds for credit scoring models. With risk-level thresholds for loan selections, improved overall performance was obtained as reduced cost of misclassification. In addition to the primary market of P2P lending, this study evaluated the secondary market in P2P lending. The secondary market in P2P lending provides a vast amount of data for multiple analysis purpose. In this study, with the data from the secondary market, the influencing features for the loan transactions are analyzed.

Finally, for a concrete risk and return evaluation of P2P loans, the study proposed a portfolio optimization approach that fulfils the need of lenders to make rational loan selection and investments. The portfolio selection with the portfolio optimization provides better decision support for lenders, where the lenders can make their investments according to their desired risk and return. The proposed portfolio optimization approach in the study provides better performance results compared to the existing approaches

for P2P lending.

The results obtained through the experimental researches are seen to be effective in supporting lending decisions in P2P lending. However, one of the limitation of the research is the problem of generalization of the research results. As P2P lending platform operating in a region can have different cultural environment, financial systems and business model from others, it can affect outcomes of the research results. Behavior pattern of borrowers can differ significantly from one P2P lending platform to another that can have big change in data distribution, effecting in the modeling process. Therefore, future research may be performed in comparing the modeling results with respect to different P2P lending data samples.

# References


[1] Hussein A Abdou and John Pointon. Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent systems in accounting, finance and management*, 18(2-3):59–88, 2011.

[2] Kaplan Abraham. The conduct of inquiry: Methodology for behavioral science, 1964.

[3] Eliana Angelini, Giacomo di Tollo, and Andrea Roli. A neural network approach for credit risk evaluation. *The quarterly review of economics and finance*, 48(4):733–755, 2008.

[4] Denise L Anthony. Micro-lending institutions: Using social networks to create productive capabilities. *international Journal of sociology and social Policy*, 1997.

[5] Taiwo Oladipupo Ayodele. Types of machine learning algorithms. *New advances in machine learning*, 3:19–48, 2010.

[6] Alexander Bachmann, Alexander Becker, Daniel Buerckner, Michel Hilker, Frank Kock, Mark Lehmann, Phillip Tiburtius, and Burkhardt Funk. Online peer-to-peer lending-a literature review. *Journal of Internet Banking and Commerce*, 16(2):1, 2011.

[7] Bart Baesens, Tony Van Gestel, Maria Stepanova, Dirk Van den Poel, and Jan Vanthienen. Neural network survival analysis for personal loan data. *Journal of the Operational Research Society*, 56(9):1089–1098, 2005.

[8] Alejandro Correa Bahnsen, Djamia Aouada, and Björn Ottersten. Example-dependent cost-sensitive logistic regression for credit scoring. In *2014 13th International conference on machine learning and applications*, pages 263–269. IEEE, 2014.

[9] Arash Bahrammirzaee. A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, 19(8):1165–1195, 2010.

[10] Kaveh Bastani, Elham Asgari, and Hamed Namavari. Wide and deep learning for peer-to-peer lending. *Expert Systems with Applications*, 134:209–224, 2019.

[11] Gustavo EAPA Batista, Ronaldo C Prati, and Maria Carolina Monard. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD explorations newsletter*, 6(1):20–29, 2004.

[12] Hussain Ali Bekhet and Shorouq Fathi Kamel Eletter. Credit risk assessment model for jordanian commercial banks: Neural scoring approach. *Review of Development Finance*, 4(1):20–28, 2014.

[13] Tony Bellotti and Jonathan Crook. Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60(12):1699–1707, 2009.

[14] Indranil Bose and Radha K Mahapatra. Business data mining—a machine learning perspective. *Information & management*, 39(3):211–225, 2001.

[15] Kendrick Boyd, Kevin H Eng, and C David Page. Area under the precision-recall curve: point estimates and confidence intervals. In *Joint European conference on machine learning and knowledge discovery in databases*, pages 451–466. Springer, 2013.

[16] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[17] Michael Buckland and Fredric Gey. The relationship between recall and precision. *Journal of the American society for information science*, 45(1):12–19, 1994.

[18] Zeynel Cebeci and Figen Yildiz. Comparison of k-means and fuzzy c-means algorithms on different cluster structures. *Agrárinformatika/journal of agricultural informatics*, 6(3):13–23, 2015.

[19] Shunpo Chang, Simon Dae-oong Kim, and Genki Kondo. Predicting default risk of lending club loans. *Machine Learning*, pages 1–5, 2016.

[20] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[21] Nitesh V Chawla, Nathalie Japkowicz, and Aleksander Kotcz. Special issue on learning from imbalanced data sets. *ACM SIGKDD explorations newsletter*, 6(1):1–6, 2004.

[22] Dongyu Chen, Fujun Lai, and Zhangxi Lin. A trust model for online peer-to-peer lending: a lender's perspective. *Information Technology and Management*, 15(4):239–254, 2014.

[23] Davide Chicco and Giuseppe Jurman. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1):6, 2020.

[24] Koh Hian Chye, Tan Wei Chin, and Goh Chwee Peng. Credit scoring using data mining techniques. *Singapore Management Review*, 26(2):25–48, 2004.

[25] Taane G Clark, Michael J Bradburn, Sharon B Love, and Douglas G Altman. Survival analysis part i: basic concepts and first analyses. *British journal of cancer*, 89(2):232–238, 2003.

[26] Alejandro Correa, Andres Gonzalez, Catherine Nieto, and Darwin Amezquita. Constructing a credit risk scorecard using predictive clusters. In *SAS Global Forum*, volume 128, 2012.

[27] JN Crook. Credit scoring: An overview (working paper series no. 96/13). *British Association, Festival of Science, University of Birmingham and the University of Edinburgh*, 1996.

[28] Jonathan N Crook, David B Edelman, and Lyn C Thomas. Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183(3):1447–1465, 2007.

[29] Yuliya Demyanyk and Daniel Kolliner. Peer-to-peer lending is poised to grow. *Economic Trends*, 2014.

[30] R Dubin. Building theory, 1969.

[31] Yosi L Eddy and Engku Muhammad Nazri Engku Abu Bakar. Credit scoring models: Techniques and issues. *Journal of Advanced Research in Business and Management Studies*, 7(2):29–41, 2017.

[32] Riza Emekter, Yanbin Tu, Benjamas Jirasakuldech, and Min Lu. Evaluating credit risk and loan performance in online peer-to-peer (p2p) lending. *Applied Economics*, 47(1):54–70, 2015.

[33] Andrew Estabrooks, Taeho Jo, and Nathalie Japkowicz. A multiple resampling method for learning from imbalanced data sets. *Computational intelligence*, 20(1):18–36, 2004.

[34] Tom Fawcett. An introduction to roc analysis. *Pattern recognition letters*, 27(8):861–874, 2006.

[35] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

[36] Mikel Galar, Alberto Fernandez, Edurne Barrenechea, Humberto Bustince, and Francisco Herrera. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(4):463–484, 2011.

[37] Cynthia A Glassman and Howard M Wilkins. Credit scoring: probabilities and pitfalls. *Journal of Retail Banking Services*, 19:53–56, 1997.

[38] Mark Goadrich, Louis Oliphant, and Jude Shavlik. Gleaner: Creating ensembles of first-order clauses to improve recall-precision curves. *Machine Learning*, 64(1-3):231–261, 2006.

[39] Laura Gonzalez and Yuliya Komarova Loureiro. When can a photo increase credit? the impact of lender and borrower profiles on online peer-to-peer loans. *Journal of Behavioral and Experimental Finance*, 2:44–58, 2014.

[40] Maria Aparecida Gouvêa and Eric Bacconi Gonçalves. Credit risk analysis applying logistic regression, neural networks and genetic algorithms models. In *POMS 18th Annual Conference*, 2007.

[41] Dominique Guegan and Bertrand Hassani. Regulatory learning: how to supervise machine learning models? an application to credit scoring. *The Journal of Finance and Data Science*, 4(3):157–171, 2018.

[42] Yanhong Guo, Wenjun Zhou, Chunyu Luo, Chuanren Liu, and Hui Xiong. Instance-based credit risk assessment for investment decisions in p2p lending. *European Journal of Operational Research*, 249(2):417–426, 2016.

[43] David J Hand and William E Henley. Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3):523–541, 1997.

[44] Michal Herzenstein, Rick L Andrews, Utpal M Dholakia, and Evgeny Lyandres. The democratization of personal consumer loans? determinants of success in online peer-to-peer lending communities. *Boston University School of Management Research Paper*, 14(6):1–36, 2008.

[45] Alan R Hevner, Salvatore T March, Jinsoo Park, and Sudha Ram. Design science in information systems research. *MIS quarterly*, pages 75–105, 2004.

[46] Patrick Hughes. Paradigms, methods and knowledge. In *Doing early childhood research : international perspectives on theory and practice*. Routledge, 2010.

[47] Joon-Ku Im, Daniel W Apley, C Qi, and X Shan. A time-dependent proportional hazards survival model for credit risk analysis. *Journal of the Operational Research Society*, 63(3):306–321, 2012.

[48] Huseyin Ince and Bora Aktan. A comparison of data mining techniques for credit scoring in banking: A managerial perspective. *Journal of Business Economics and Management*, 10(3):233–240, 2009.

[49] Mirjana Ivanović and Miloš Radovanović. Modern machine learning techniques and their applications. In *International Conference on Electronic, Communication, and Network*, 2014.

[50] Rajkamal Iyer, Asim Ijaz Khwaja, Erzo FP Luttmer, and Kelly Shue. Screening in new credit markets: Can individual lenders infer borrower creditworthiness in peer-to-peer lending? In *AFA 2011 Denver meetings paper*, 2009.

[51] Anil K Jain, M Narasimha Murty, and Patrick J Flynn. Data clustering: a review. *ACM computing surveys (CSUR)*, 31(3):264–323, 1999.

[52] Cuiqing Jiang, Zhao Wang, Ruiya Wang, and Yong Ding. Loan default prediction by combining soft information extracted from descriptive text in online peer-to-peer lending. *Annals of Operations Research*, 266(1-2):511–529, 2018.

[53] Gopalan Kesavaraj and Sreekumar Sukumaran. A study on classification techniques in data mining. In *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, pages 1–7. IEEE, 2013.

[54] Amir E Khandani, Adlar J Kim, and Andrew W Lo. Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11):2767–2787, 2010.

[55] Adnan Khashman. A neural network model for credit risk evaluation. *International Journal of Neural Systems*, 19(04):285–294, 2009.

[56] Michael Klafft. Online peer-to-peer lending: a lenders' perspective. In *Proceedings of the international conference on E-learning, E-business, enterprise information systems, and E-government, EEE*, pages 371–375, 2008.

[57] Hian Chye Koh, Wei Chin Tan, and Chwee Peng Goh. A two-step method to construct credit scoring models with data mining techniques. *International Journal of Business and Information*, 1(1):96–118, 2006.

[58] Münevver Köküer, Raouf NG Naguib, Peter Jančovič, H Banfield Younghusband, and Roger Green. Towards automatic risk analysis for hereditary non-polyposis colorectal cancer based on pedigree data. In *Outcome Prediction in Cancer*, pages 319–337. Elsevier, 2007.

[59] TS Kuhn. The structure of scientific revolutions, 1996.

[60] Ashish Kumar, Roheet Bhatnagar, and Sumit Srivastava. Analysis of credit risk prediction using arsknn. In *International Conference on Advanced Machine Learning Technologies and Applications*, pages 644–652. Springer, 2018.

[61] Ranjit Kumar. *Research methodology: A step-by-step guide for beginners*. Sage, 2018.

[62] Vinod Kumar, S Natarajan, S Keerthana, KM Chinmayi, and N Lakshmi. Credit risk analysis in peer-to-peer lending system. In *2016 IEEE International Conference on Knowledge Engineering and Applications (ICKEA)*, pages 193–196. IEEE, 2016.

[63] Adel Lahsasna, Raja Noor Ainon, and Ying Wah Teh. Credit scoring models using soft computing methods: A survey. *Int. Arab J. Inf. Technol.*, 7(2):115–123, 2010.

[64] Pat Langley and Herbert A Simon. Applications of machine learning and rule induction. *Communications of the ACM*, 38(11):54–64, 1995.

[65] Laura Larrimore, Li Jiang, Jeff Larrimore, David Markowitz, and Scott Gorski. Peer to peer lending: The relationship between language features, trustworthiness, and persuasion success. *Journal of Applied Communication Research*, 39(1):19–37, 2011.

[66] Rick Lawrence, Andrew Bunn, Scott Powell, and Michael Zambon. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. *Remote sensing of environment*, 90(3):331–336, 2004.

[67] Eunkyoung Lee and Byungtae Lee. Herding behavior in online p2p lending: An empirical investigation. *Electronic Commerce Research and Applications*, 11(5):495–503, 2012.

[68] In Lee and Yong Jae Shin. Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*, 63(2):157–170, 2020.

[69] Tian-Shyug Lee and I-Fei Chen. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28(4):743–752, 2005.

[70] Tian-Shyug Lee, Chih-Chou Chiu, Chi-Jie Lu, and I-Fei Chen. Credit scoring using the hybrid neural discriminant technique. *Expert Systems with applications*, 23(3):245–254, 2002.

[71] Martin Leo, Suneel Sharma, and Koilakuntla Maddulety. Machine learning in banking risk management: A literature review. *Risks*, 7(1):29, 2019.

[72] Jianjun Li, Sara Hsu, Zhang Chen, and Yang Chen. Risks of p2p lending platforms in china: Modeling failure using a cox hazard model. *The Chinese Economy*, 49(3):161–172, 2016.

[73] Wei Li, Shuai Ding, Yi Chen, and Shanlin Yang. Heterogeneous ensemble for default prediction of peer-to-peer lending in china. *IEEE Access*, 6:54396–54406, 2018.

[74] Andy Liaw, Matthew Wiener, et al. Classification and regression by randomforest. *R news*, 2(3):18–22, 2002.

[75] Michael K Lim and So Young Sohn. Cluster-based dynamic scoring model. *Expert Systems with Applications*, 32(2):427–431, 2007.

[76] Mingfeng Lin, Nagpurnanand R Prabhala, and Siva Viswanathan. Judging borrowers by the company they keep: Friendship networks and information asymmetry in online peer-to-peer lending. *Management Science*, 59(1):17–35, 2013.

[77] Xuchen Lin, Xiaolong Li, and Zhong Zheng. Evaluating borrower's default risk in peer-to-peer lending: evidence from a lending platform in china. *Applied Economics*, 49(35):3538–3545, 2017.

[78] Victoria López, Alberto Fernández, Salvador García, Vasile Palade, and Francisco Herrera. An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics. *Information sciences*, 250:113–141, 2013.

[79] Francisco Louzada, Vicente Cancho, Mauro de Oliveira Jr, and Yiqi Bao. Modeling time to default on a personal loan portfolio in presence of disproportionate hazard rates. *Journal of Statistics Applications & Probability, An International Journal. J. Stat. Appl. Pro*, 3:1–11, 2014.

[80] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[81] Milad Malekipirbazari and Vural Aksakalli. Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42(10):4621–4631, 2015.

[82] Rashmi Malhotra and Davinder K Malhotra. Evaluating consumer loans using neural networks. *Omega*, 31(2):83–96, 2003.

[83] Salvatore T March and Gerald F Smith. Design and natural science research on information technology. *Decision support systems*, 15(4):251–266, 1995.

[84] M Lynne Markus, Ann Majchrzak, and Les Gasser. A design theory for systems that support emergent knowledge processes. *MIS quarterly*, pages 179–212, 2002.

[85] Loretta J Mester et al. What's the point of credit scoring? *Business review*, 3(Sep/Oct):3–16, 1997.

[86] Andreas Mild, Martin Waitz, and Jürgen Wöckl. How low can you go?—overcoming the inability of lenders to set proper interest rates on unsecured peer-to-peer lending markets. *Journal of Business Research*, 68(6):1291–1305, 2015.

[87] Gretchen G Moisen, Elizabeth A Freeman, Jock A Blackard, Tracey S Frescino, Niklaus E Zimmermann, and Thomas C Edwards Jr. Predicting tree species presence and basal area in utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods. *Ecological modelling*, 199(2):176–187, 2006.

[88] Thabiso Peter Mpofu and Macdonald Mukosera. Credit scoring techniques: a survey. *International Journal of Science and Research (IJSR)*, pages 2319–7064, 2014.

[89] Iqbal Muhammad and Zhu Yan. Supervised machine learning approaches: A survey. *ICTACT Journal on Soft Computing*, 5(3), 2015.

[90] Penny Mukherji and Deborah Albon. *Research methods in early childhood: An introductory guide*. Sage, 2018.

[91] Satuluri Naganjaneyulu and Mrithyumjaya Rao Kuppa. A novel framework for class imbalance learning using intelligent undersampling. *Progress in artificial intelligence*, 2(1):73–84, 2013.

[92] B Nithya and V Ilango. Predictive analytics in health care using machine learning tools and techniques. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 492–499. IEEE, 2017.

[93] Beibei Niu, Jinzheng Ren, and Xiaotao Li. Credit scoring using machine learning by combing social network information: Evidence from peer-to-peer lending. *Information*, 10(12):397, 2019.

[94] Pariwat Ongsulee, Veena Chotchaung, Eak Bamrungsi, and Thanaporn Rodcheewit. Big data, predictive analytics and machine learning. In *2018 16th International Conference on ICT and Knowledge Engineering (ICT&KE)*, pages 1–6. IEEE, 2018.

[95] Stjepan Oreški and Goran Oreški. Cost-sensitive learning from imbalanced datasets for retail credit risk assessment. *TEM JOURNAL-Technology, Education, Management, Informatics*, 7(1):59–73, 2018.

[96] FY Osisanwo, JET Akinsola, O Awodele, JO Hinmikaiye, O Olakanmi, and J Akinjobi. Supervised machine learning algorithms: classification and comparison. *International Journal of Computer Trends and Technology (IJCTT)*, 48(3):128–138, 2017.

[97] Vincenzo Pacelli, Michele Azzollini, et al. An artificial neural network approach for credit risk management. *Journal of Intelligent Learning Systems and Applications*, 3(02):103, 2011.

[98] Mahesh Pal. Random forest classifier for remote sensing classification. *International journal of remote sensing*, 26(1):217–222, 2005.

[99] Vili Podgorelec, Peter Kokol, Bruno Stiglic, and Ivan Rozman. Decision trees: an overview and their use in medicine. *Journal of medical systems*, 26(5):445–463, 2002.

[100] Joseph G Ponterotto. Qualitative research in counseling psychology: A primer on research paradigms and philosophy of science. *Journal of counseling psychology*, 52(2):126, 2005.

[101] David Martin Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. 2011.

[102] Lauri Puro, Jeffrey E Teich, Hannele Wallenius, and Jyrki Wallenius. Bidding strategies for real-life small loan auctions. *Decision Support Systems*, 51(1):31–41, 2011.

[103] Lior Rokach and Oded Maimon. Clustering methods. In *Data mining and knowledge discovery handbook*, pages 321–352. Springer, 2005.

[104] Lior Rokach and Oded Maimon. Decision trees. In *Data mining and knowledge discovery handbook*, pages 165–192. Springer, 2005.

[105] Yutaka Sasaki. The truth of the f-measure. *Teach Tutor Mater*, 01 2007.

[106] Daniel Sciarra. The role of the qualitative researcher. *Using qualitative methods in psychology*, 37, 1999.

[107] Sanja Scitovski and Nataša Šarlija. Cluster analysis in retail segmentation for credit scoring. *Croatian Operational Research Review*, pages 235–245, 2014.

[108] Carlos Serrano-Cinca and Begoña Gutiérrez-Nieto. The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (p2p) lending. *Decision Support Systems*, 89:113–122, 2016.

[109] Carlos Serrano-Cinca, Begoña Gutiérrez-Nieto, and Luz López-Palacios. Determinants of default in p2p lending. *PloS one*, 10(10):e0139427, 2015.

[110] Galit Shmueli and Otto R Koppius. Predictive analytics in information systems research. *MIS quarterly*, pages 553–572, 2011.

[111] Marina Sokolova, Nathalie Japkowicz, and Stan Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In *Australasian joint conference on artificial intelligence*, pages 1015–1021. Springer, 2006.

[112] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437, 2009.

[113] Maria Stepanova and Lyn Thomas. Survival analysis methods for personal loan data. *Operations Research*, 50(2):277–289, 2002.

[114] Ning Su. Positivist qualitative methods. In *The SAGE handbook of qualitative business and management research methods*, pages 17–31. SAGE Publications Ltd, 2018.

[115] Yanmin Sun, Andrew KC Wong, and Mohamed S Kamel. Classification of imbalanced data: A review. *International journal of pattern recognition and artificial intelligence*, 23(04):687–719, 2009.

[116] Alaa Tharwat. Classification assessment methods. *Applied Computing and Informatics*, 2020.

[117] Lyn C Thomas. A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International journal of forecasting*, 16(2):149–172, 2000.

[118] Lyn C Thomas. *Consumer credit models: pricing, profit and portfolios: pricing, profit and portfolios*. OUP Oxford, 2009.

[119] Chih-Fong Tsai and Jhen-Wei Wu. Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert systems with applications*, 34(4):2639–2649, 2008.

[120] Alfonso Urso, Antonino Fiannaca, Massimo La Rosa, Valentina Ravì, and Riccardo Rizzo. Data mining: Prediction methods. *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, page 413, 2018.

[121] Tony Van Gestel and Bart Baesens. *Credit Risk Management: Basic concepts: Financial risk components, Rating analysis, models, economic and regulatory capital*. OUP Oxford, 2008.

[122] Andrew Verstein. The misregulation of person-to-person lending. *UCDL Rev.*, 45:445, 2011.

[123] Sofia Visa, Brian Ramsay, Anca L Ralescu, and Esther Van Der Knaap. Confusion matrix-based feature selection. *MAICS*, 710:120–127, 2011.

[124] Martin Vojtek, Evžen Koèenda, et al. Credit-scoring methods. *Czech Journal of Economics and Finance (Finance a uver)*, 56(3-4):152–167, 2006.

[125] Michael N Vrahatis, Basilis Boutsinas, Panagiotis Alevizos, and Georgios Pavlides. The new k-windows algorithm for improving thek-means clustering algorithm. *journal of complexity*, 18(1):375–391, 2002.

[126] Matthew A Waller and Stanley E Fawcett. Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management, 2013.

[127] Chongren Wang, Dongmei Han, Qigang Liu, and Suyuan Luo. A deep learning approach for credit scoring of peer-to-peer lending using attention mechanism lstm. *IEEE Access*, 7:2161–2168, 2018.

[128] Gang Wang, Jinxing Hao, Jian Ma, and Hongbing Jiang. A comparative assessment of ensemble learning for credit scoring. *Expert systems with applications*, 38(1):223–230, 2011.

[129] Hui Wang and Martina E Greiner. Prosper—the ebay for money in lending 2.0. *Communications of the Association for Information Systems*, 29(1):13, 2011.

[130] David West. Neural network credit scoring models. *Computers & Operations Research*, 27(11-12):1131–1152, 2000.

[131] Yufei Xia, Chuanzhe Liu, YuYing Li, and Nana Liu. A boosted decision tree approach using bayesian hyper-parameter optimization for credit scoring. *Expert Systems with Applications*, 78:225–241, 2017.

[132] Yufei Xia, Chuanzhe Liu, and Nana Liu. Cost-sensitive boosted tree for loan evaluation in peer-to-peer lending. *Electronic Commerce Research and Applications*, 24:30–49, 2017.

[133] Haewon Yum, Byungtae Lee, and Myungsin Chae. From the wisdom of crowds to my own judgment in microfinance through online peer-to-peer lending platforms. *Electronic Commerce Research and Applications*, 11(5):469–483, 2012.

[134] Krista Rizman Žalik. An efficient k′-means clustering algorithm. *Pattern Recognition Letters*, 29(9):1385–1391, 2008.

[135] Jing Zhou, Wei Li, Jiaxin Wang, Shuai Ding, and Chengyi Xia. Default prediction in p2p lending from high-dimensional data based on machine learning. *Physica A: Statistical Mechanics and its Applications*, 534:122370, 2019.

Ajay Byanjankar

# Predicting Risk and Return in Peer-to-Peer Lending with Machine Learning

## A Decision Making Approach

This dissertation performs the study of credit risk evaluation in Peer-to-Peer Lending. The dissertation attempts to evaluate the risk and estimate the risk and return of loans in peer-to-peer lending that supports the lending decisions.

The evaluation of the credit risk is performed with credit scoring Predictive analytics is applied with machine learning algorithms to create more accurate credit scoring models with higher predictive performance.

ISBN 978-952-12-4129-1