Multi-objective Loan Portfolio Optimization in Peer-to-Peer Lending Markets using Machine-Learning Techniques

Maakgetlwa Seleme Shoky
Student No: 15003124

Supervisor: Dr S Moyo
Co-supervisor: Mr N Mphephu

Research Project for the Master of Science Degree in Applied Mathematics

School of Mathematical and Natural Sciences
Department of Mathematics and Applied Mathematics
University of Venda
Thohoyandou, Limpopo
South Africa

30 March 2022

Declaration

I, ***Maakgetlwa Seleme Shoky****,* ***Student No: 15003124****, declare that this research project is my original work and has not been submitted for any degree at any other university or institution. The project does not contain any other persons' writing unless specifically acknowledged and referenced accordingly.*

30/03/2022

Miss SS. Maakgetlwa

(Student)

Date

Abstract

Portfolio optimization problems in the Peer-to-Peer lending Platforms involve selecting good loan applications (less risky) from various potential borrowers. Such loans have lower level of risk in terms of funding and earning higher returns. The aim of this study is to find ways to maximize returns and minimize the risks associated with the investment. It becomes more complicated to optimally allocate weights to the loan application when there is an increased number of applications for funding. This study focused on devising techniques which can be used to optimally select portfolios of loan applications for funding with desired returns on the investment. Harry Markowitz pioneered the Modern Portfolio theory also known as Mean-variance theory to construct a portfolio but the theory failed since it was built on unrealistic assumptions in terms of real life situations. This study explored and compared the mean-variance theory and other machine learning methods to construct a portfolio of loans from peer-to-peer lending market in order to be able to recommend the best approach to achieving high returns with minimum risk. The study employed the evolutionary algorithms (Particle Swarm Optimization and Genetic Algorithm) and the Reinforcement learning algorithm.

Keywords: *Calibration, Genetic Algorithm, Machine Learning, Markowitz's mean-variance, Reinforcement Learning, Optimization, Portfolio Optimization, Particle Swarm Optimization, Peer-to-Peer Lending Market.*

Acknowledgments

My appreciation goes to the following people:

- I would like to begin by thanking the Lord almighty God for the wisdom, knowledge and strength which made the completion of this study possible. Had He not been at the center of my life all this would not have possible.
- Mr N Mphephu and Dr S Moyo for their great supervision, intellectual support and encouragement when things got hard and felt like giving up.
- My friends, family and fellow classmates for making it fun and easier to study.
- This work is dedicated to both my parents who are deceased for even in their absence they have been role models and have been the reason I pushed towards achieving what they hoped and wished for my life.
- The support of the DSI-NICIS National e-Science Postgraduate Teaching and Training Platform (NEPTTP) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NEPTTP.
- My heartfelt appreciation goes to Mr N Mphephu who aside from being my co-supervisor, been a mentor, a brother and a friend. Without him the process would have been so much harder.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The rapid growth of information technology has led to the birth of electronic market platforms whereby some traditional lending intermediaries have become less important or even redundant for the economic interactions of market participants (Berger and Gleisner, 2009). As a result these developments have led to the creation of online platforms that aim to use the old ideas of credit decision processes to develop their new ways of disbursing loans. Peer-to-Peer lending platform is one of those online platforms that has come along with the evolution of technology. Peer-to-Peer implies that, there are two role players involved in the process. These role players are borrowers and lenders in an investment venture without a middle man nor supervisor. Given this setting, the Peer-to-Peer lending platform can be defined as an online-based platform that allows borrowers and lenders to come together to apply for loans and, respectively, to lend money. This type of borrowing and lending has grown exponentially attractive to many investors and borrowers in recent years. The new market forces have significantly contributed to the improvement of the efficiency of the financial markets in many ways (Helder and José, 2011). Apart from being online-based platforms that facilitate borrowing and lending between individuals, they have grown to be very complex ecosystems of technologies, institutions and auxiliary start-ups (Mateescu, 2015). Studies have proved that within the next few years Peer-to-Peer lending platforms will occupy about 10% of the market share worldwide for retail lending and financial planning (Mateescu, 2015).

The online nature of the Peer-to-Peer lending platform implies that there is very little involvement of the middle man in the loan application and approval processes. That is, there are no traditional bank processes involved in the loan application and approval. Therefore, the borrower does not have to endure a lengthy process to get the loan application approved inimical to traditional banking systems. The Peer-to-Peer platforms offer an overall cost reduction advantage because the transaction costs are much lower compared to conventional banks. The elimination of the lengthy application process and transaction costs makes these Peer-to-Peer platforms more attractive to borrowers and lenders (Chi et al., 2019). However in these platforms there are no traditional financial intermediaries implying that there is very little information about the borrowers' creditworthiness, which in turn makes investing in such loans very risky. The uncertainty around borrowers' creditworthiness is one of the disadvantages of these platforms because there is no assurance of whether the borrower will default or not. Against this backdrop, these lending platforms carry high investment risk than the traditional banks. Another factor that makes these platforms more risky is the unsecured nature of the loans issued. Almost all of these loans are not protected by a guarantor or collateralised on specific assets of the borrower in case of default.

After the global financial crisis of 2008, whereby financial institutions and banks failed the public which resulted in the loss of people's huge investments, there was a substantial loss of trust in these traditional institutions Mateescu (2015). The regulated institutions lost public trust on their ability to put enough security around their investments and as a result this gave birth to the idea of disintermediation or simply put, the elimination of traditional intermediaries in the financial system. (Mateescu, 2015) noted that "Not only did the mainstream financial system implode, leaving millions of borrowers baring an extraordinary debt burden, but, the contraction that followed left individuals and small businesses cut-off from fresh sources of credit". The loss of public trust in the regulated banks, resulted in the establishment of a Basel committee with an aim of setting stringent rules for the banks to avoid similar stressful experiences. The committee was formed and declared between July 2009 and September 2010 as reform programme meant to address the lessons from the crisis which among other regulations had the below mentioned regulations (Basel Committee, 2010)

- Raising the cost of resources to insure that banks are best prepared to withstand risks on the grounds of all existing issues and worries
- Growing requirements for the supervisory oversight mechanism (Pillar 2) and public reporting (Pillar 3) along with appropriate advice in the areas of sound accounting procedures, stress monitoring, liquidity risk control, corporate governance and compensation.
- Promoting the build-up of capital reserves in good times that can be drawn up during cycles of stress, including both the capital retention buffer and the countercyclical buffer to shield the banking system from periods of excess credit production.

- Increase the risk coverage of the financial system, in particular for investment practices, securitization, off-balance-sheet asset exposures and counterparty collateral risks resulting from derivatives;

These regulations were designed to protect both the banks and the public. However, since it had been proved that the regulated banks were at fault in that crisis, the public had already lost trust in them and the new regulations made things worse since it had become difficult for the public to get loans from banks due to fact that they did not qualify (according to new regulations) or they were still skeptical on trusting the banks. This meant that regardless of the riskiness of Peer-to-Peer lending platforms, currently a large population of people approximately 50%, are choosing to engage in this line of business. Digital evolution which is causing a lot of disruption in many sectors of the economy and most notably the financial sector is giving these online lending platforms a massive advantage over the traditional banks because these lending platforms are technologically driven. The upcoming of the 4^{th} Industrial Revolution (4IR) which aims at digitizing everything puts these platforms at an advantage through technology.

Dominated by United States(U.S.) based LendingClub[a] and Prosper[b] together with United Kingdom (U.K.) based Zopa[c], these platforms have succeeded because of their ability to provide credit at lower rates than borrowers would normally receive and through introducing a lucrative and alternative asset class for investors(Kunal, 2016). An example of how these platforms operate can be summarized as demonstrated in Figure 1.1 below. This description is based on the Lending Club platform.

[a]https://www.lendingclub.com/
[b]https://www.prosper.com/
[c]https://www.Zopa.com/

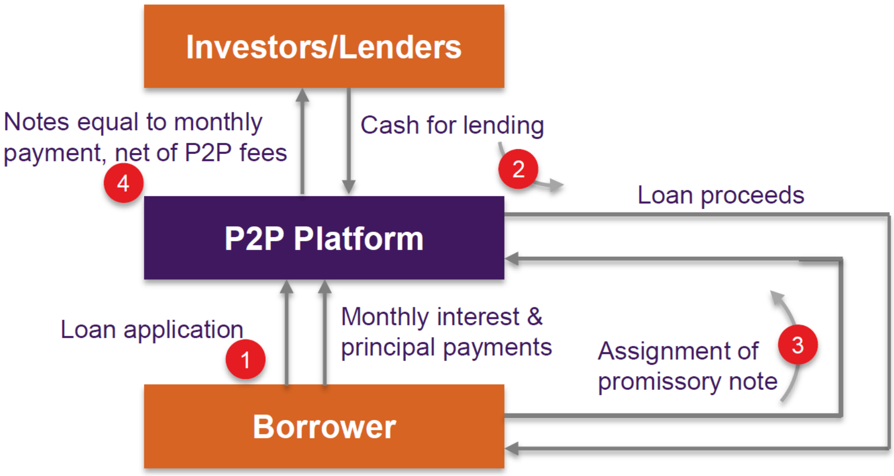


Figure 1.1: P2P Lending Mechanism
(Chi et al., 2019)

This process is initiated when a borrower applies for a loan called lists. The Peer-to-Peer lending platform then goes through the borrower's credentials to check if they meet their minimum requirements. If the borrower passes the background check stage, the loan request is then listed on the platform for investors to place a bid to fund the loan. Typically, lenders prefer to spread their money across many different loans to help reduce risk. As a result, a loan may end up being funded by many investors (Luo et al., 2011). Once the investors pledge to fund the loan request, the funds are transferred and a certain percentage is deducted by the platform in the form of transaction cost, and the remaining funds are then transferred to the borrower. Once the borrower receives the funds, they send a promissory note to the platform. The platform's promissory notes are then used to distribute the interests and principal payments among the investors.

Ideally, a rational investor would want to invest his/her money where there is prospect of getting higher returns. Since investment decision is entirely upon the investor and not the platform, the investor needs to spend more time evaluating different types of loans and associated risks. This goes with the famous rule of thumbs for investment that advises on not putting all eggs in one basket. This then shows the importance of funding multiple different loans with different levels of risk. As indicated by Chi et al. (2019), an efficient and fair investment in peer-to-peer loans must be based on a realistic credit risk distribution estimate in order to be effective and reasonable. Because it is difficult to get historical returns (or losses) data on similar past loan applications, estimating the credit risk distribution of Peer-to-Peer loans can be particularly difficult to assess. In other words, historical return data on a borrower who is identical to the current borrower is rarely accessible. When there is minimal data or specialist knowledge available, it is common to make educated guesses about the distribution of loan returns (and losses). In most cases, these estimates are not considerably precise, and the problem is referred to as distribution ambiguity (also known as probability measure or uncertainty) (Chi et al., 2019).

The purpose of this study is to find ways to ease the amount of pressure from the investors in conducting investigation which has to be done before any investment decision is made and to assist the investors in making more data-driven investment decisions. It is not about how they feel about the need to invest or to advance a loan but to make an informed choice with high chances of high returns. Deep learning models will be utilized in the diversification of reducing risk while increasing returns from the investments. In this way, the investor will be able to select a diversified portfolio of loans which depend on the level of risk as well as ensuring the desired return. This will take into consideration the constraint parameters put in place by the investor for regulation of loss levels. One could either go for aggressive investments (Very risky/volatile) or medium/low-risk investments.

1.1 Problem Description

As a rule of thumb in investment management, it is important to diversify investments to avoid financial loss. Thus, the investor is faced with the problem of choosing the best loan proposals out of many offers which customarily are associated with various risks(i.e., the default probability of each one of those proposals), trading efficiency(i.e., winning-bid probability and fully-funded probability) and the whole performance(i.e., optimal portfolio selection) (Xing and Marwala, 2018). From the performance attribute, the goal was to hold the most desirable group of loans that yielded higher returns. This task of selecting the most profitable group of loans led to the following research question:

- *How to automatically assess the loan requests and optimally select the best loans to create a portfolio of loans which is based on investment objectives?*

In answering this research question, the study made use of Data Science, that is Artificial Intelligence and Machine Learning tools, to automatically assess the loan requests based on the investor's objectives and select the best loans for creating optimal loan portfolios. The study also made use of the lending club data and the evolutionary algorithms (Particle Swarm Optimisation and Genetic Algorithm), which had their foundation from the traditional mean-variance algorithm, adapted from the portfolio optimization used in financial institutions. While considering the use of these two optimization algorithms, the researcher endeavoured to use the reinforcement learning algorithm. All these algorithms are discussed below.

1.2 Motivation for this Study

Portfolio optimization problems in the Peer-to-Peer Lending Platform involves selecting good loans (less risky) from different loans with different levels and kind of risks for funding with the goal of reaping high returns from an investment.The aim is to minimize the risks associated with the investment in order to realise high returns. For this study it becomes more complicated to optimally allocate weights to these loans when the number of loans available to select from for investing is large (Thomas et al., 2017). The South African President, Cyril Ramaphosa, in August 2019 signed the bill of national credit amendment, which meant that banks would have to price with higher risk and reduce/cease lending to low-income customers (Geldddenhuys, 2019; South African Goverment, 2019). This means that most people are pushed towards the Peer-to-Peer Lending Platforms for loans due to failure to qualify for funding or to avoid high-interest rates charged by the banks. This also opens a huge amount of lending opportunities for lenders/investors in the Peer-to-Peer lending platforms.

Investors in Peer-to-Peer lending platforms, are presented with a large number of loans with different risks and have a challenge of optimally selecting the portfolio of loans they can fund. The study is as a result of the need to be able to optimally select a portfolio of loans for investors. The main question is; how does one automatically assess the loan requests and optimally select the appropriate loans to create a portfolio of loans based on investment objectives?.

1.3 Aim of the study

This study investigates the effectiveness of machine learning models in selecting an optimal portfolio from groups of loans from a Peer-to-Peer lending platform, and this entails implementing the traditional Mean-Variance methods on portfolio optimisation and comparing those results with those obtained through machine learning algorithms.

1.4 Objectives of the study

The primary objective of this study is to optimize the selection of the best loan portfolio based on the Peer-to-Peer lending platform data. The selection of a group of different loans is done by using heuristic methods namely particle swarm optimization, genetic algorithm, and reinforcement learning. The use of these heuristic methods provides the basis for:

1. Efficient ways of creating a portfolio of loans that will yield maximum returns and
2. Adjusting constraint parameters to optimize the performance of the loans through minimisation of risks resulting in maximized returns.

1.5 Expected contribution to knowledge

The Peer-to-Peer lending market is gaining approval, yet there is little literature available on these developments in the industry. Unlike the traditional banks, where the investments are made by qualified investors who are knowledgeable about risk management, the Peer-to-Peer lending market is open to everyone who wishes to invest in those loans. It is, therefore, essential for an investor to be knowledgeable in portfolio and risk management issues to avoid making costly uninformed investment decisions. In general the platform does not advise on how the investors should select their portfolios. The study seeks to, provide investment information to Peer-to-Peer investors deduced from data-driven techniques that will enable investors to optimally invest their funds. Results and recommendations from this study will then help investors to make more data-driven decisions when investing in the portfolios of their choice.

1.6 Organization of study

This study comprises of 6 chapters. Chapter 1 presents the introduction of the Peer-to-Peer lending platforms, the motive behind its creation and the processes involved in running it. It also discusses the problem layout as a result of the gaps identified in the reviewed literature and the need to fill in those gaps and contribution of this work to the body of knowledge. The aim and objectives of the study is to epitomize the significance of the study. Chapter 2 discusses relevant earlier investigations that in the area and in the process one identifies gaps occurring in the field which need to be studied further to provide answers to current teething lending problems. Chapter 3 presents data collection and methods applied in the study.This chapter also covers ethical issues ethics followed. In Chapter 4 discusses the exploratory data analysis conducted to the study. Chapter 5 presents the results obtained thereafter analysis of those outcomes. Chapter 6 provides the conclusions of the study, recommendations with regards to future work and the potential limitations of the study.

Chapter 2

Literature Review

2.1 Introduction

Due to the complexity of investment strategies specifically in the Peer-to-Peer lending platforms, a lot of research has been done to assist investors/lenders to choose investments (loans) that will yield desired returns depending on the risk they are willing to absorb. This helps investors to make informed decision before choosing investment options. One needs to assess the risk involved in each investment and the possible returns associated with it. That involves the creation of portfolios that would result in the desired returns. Below are a few studies that deal with the problem of portfolio optimization and the methods applied for this study.

2.2 Mean-Variance

To select an optimal portfolio, Chi et al. (2019) made use of an instance-based credit risk assessment method together with relative entropy constraints that are structured so that with the use of kernel regression, an optimal weighted average is used to predict the expected return. With the help of Chi et al. (2019), a resilient portfolio optimization model was developed using the Mean-Variance, which reduces the risk of losses due to loan distribution uncertainty and can provide an optimal portfolio even in worst-case circumstances. The use of the instance-based framework and Mean-Variance model in credit risk assessment and portfolio optimization was first applied by Guo et al. (2016) in the context of Peer-to-Peer lending platform loans.

2.3 Evolutionary Algorithms

In the multi-objective loan portfolio optimization problem, automatic assessment and selection of loans for one's portfolio is critical as well as challenging. A study by Zhao et al. (2016) addresses such a problem through using decision trees. To analyze loans from multi-objective methods, they employed a gradient boosting decision tree that combines static and dynamic characteristics. They used two methodologies to choose the best loan portfolios (weighted objective optimization strategy and multi-objective optimization strategy). The multi-objective optimization technique delivers a Pareto-optimal portfolio set, whereas the weighted objective optimization strategy seeks to produce one optimum portfolio. To solve the optimization issue efficiently, these techniques are combined into the algorithms DPA and EVA.

Evolutionary algorithms are well known to have been applied in optimization problems. With the use of these evolutionary algorithms specifically Nondominated Sorting Genetic Algorithm (NSGA II), Pareto Envelope-based Selection Algorithm (PESA) and Strength Pareto Evolutionary Algorithm (SPEA 2), a multi-objective approach to solving the bi-objective portfolio optimization problem was designed. (Diosan, 2005) states that the approach is used in such a way that the objectives of the optimization problem are initialized with the Markovitz mean-variance. The data used in their study is of the daily rate of exchange for a set of assets quoted to Euronext Stock during June to December, 2002, 6 months worth of data. Results have shown that PESA substantially surpasses the competitive algorithms in all experiments. Mishra et al. (2009) study also applies to bi-objective portfolio optimization problem, the four well-known multiobjective evolutionary algorithms which are i.e. Parallel Single Front Genetic Algorithm (PSFGA), Strength Pareto Evolutionary Algorithm 2(SPEA2), Nondominated Sorting Genetic Algorithm II(NSGA II) and Multi Objective Particle Swarm Optimization (MOPSO). MOPSO proves to significantly outperform the other comparative algorithm and the data used was from OR library maintained by Prof Beasley as a public benchmark data set and is derived from Heng Seng data set with 31 assets.

Creating an investment portfolio is one of the most critical financial considerations made by people and companies. Conventional portfolio strategy focuses on around a reasonable investor determining the proportion of assets in the portfolio to reduce risk and optimize anticipated returns. The constrained problem of portfolio selection is discussed and a heuristic algorithm based on particle swarm optimization (PSO) is developed to solve the constrained optimization problem (Wei et al., 2006). They developed a new portfolio selection concept in view of certain dynamic rational constraints and a PSO algorithm is available to solve this new model (Wei et al., 2006). The Markowitz's mean-variance model was used as a benchmark in their study due to the fact that the model was inefficient in solving constrained portfolio selection problems. When there are additional constraints the PSO algorithm is successful in solving the problem of portfolio optimisation and has the ability to tackle portfolio management real-time problems.

Just as investors need to be able to select a portfolio of loans in which they wish to invest, it is also crucial that a system which creates segments for these investors depended on their risk preference and how they have been investing previously. Luo et al. (2011) developed a data-driven investment decision-making framework that takes use of the investor composition of each investment to improve decision-making in Peer-to-Peer lending. They created investor profiles based on quantitative research of historical performance, risk preferences, investing experiences, and the trustworthiness of investors. They then created an investor composition analysis model based on investor profiles, which may be utilized to choose attractive assets and enhance investment selections.

2.4 Reinforcement Learning

Saud and Yang (2017) used the recurrent reinforcement learning (RRL) method to create both buy/sell signals and optimal asset allocation weights concurrently in a dynamic optimization problem with a statistically coherent downside risk adjusted performance objective function. They demonstrate that the expected maximum drawdown risk-based objective function outperforms previously proposed RRL objective functions (such as the Sharpe ratio and the Sterling ratio) in terms of return performance, and that variable weight RRL long/short portfolios outperform equal weight RRL long/short portfolios under various transaction cost scenarios.

To automate financial investment strategies efficiently, Nhi et al. (2019) suggest a Deep Responsible Investment Portfolio model that involves a neural network with Multivariate Bidirectional Long Short Term Memory to forecast market returns for creating a socially responsible investment portfolio. Their deep reinforcement learning techniques have been adapted to retrain neural networks and re-balance portfolios on a periodic basis whereby the empirical data used showed that the DRIP framework could achieve competitive financial performance and better social impact than traditional portfolio models, sustainable indexes, and funds.

Given that the Peer-to-Peer lending market is new, there is little literature with regards to portfolio optimization. Thus, this study contributes to the literature in that sense. The models proposed in this study are used for portfolio optimization in the traditional financial institutions. The study took that as an advantage to investigate how these models performed given such a platform. In relation to Peer-to-Peer lending markets, the work that is done with regards to machine learning techniques, uses algorithms such as decision trees, neural networks, random forest to mention but a few. The gaps identified from literature are that although there is a rapid growth in the Peer-to-Peer lending market, there is minimum work done on its data or on optimizing the manner in which investors interact with the platform and when done it is with basic machine learning techniques. Hence, this study aims to assist interested investors with or without any prior knowledge of portfolio theory making use of the evolutionary algorithms proposed for this study.

Chapter 3

Research Methods

3.1 Introduction

This chapter presents an overview of loan portfolio optimization techniques employed in this study. The chapter provides a brief discussion about the dataset at hand together with the parameters used in the algorithms for the determination of optimum solutions for the problem. The challenges and limitations of the algorithms and data used are also discussed. It is also important to state that there are no ethical considerations to be taken into account in this problem, since all the data is obtained from open sources.

3.2 Data description

The data used in this study is collected from the LendingClub platform. The platform is a leading global Peer-to-Peer lending company. It was deemed suitable for the study because of the publicly available historical loan data. The LendingClub data dates back from 2007 up to 2018 with $2\ 260\ 701$ accepted and $27\ 648\ 741$ rejected applications for loans. For the purpose of the study the focus is only on the accepted loans although a brief description of the nature of rejected loans is given.

3.3 Methods

3.3.1 Mean Variance Portfolio

Markowitz's mean-variance portfolio theory, is used to model the random return rate on assets/loans and choose portfolio weight factors optimally. The principle is that the ideal weight set is that which is associated with an acceptable portfolio baseline rate of return with minimal volatility. There is a certain expected return and an investor wants to choose a strategy with the expected return provided to minimize his or her risk (variance). Since the expected return given may not be maximum, an optimal strategy may not be optimal in the traditional sense of variance minimization problems in Markowitz's mean-variance portfolio problems (Xianping et al., 2012). Below we present a mathematical mean-variance optimization theory (Gerard and Reha, 2007; Markowitz, 1952).

Itis important to begin with defining some important parameters as follows

μ_i to be the expected return of loan i

σ_i to be the standard deviation of the return of loan i

ρ_{ij} to be the correlation coefficient of the returns of loans i and j for $i \neq j$

where $\mu = [\mu_1, \ldots, \mu_n]^T$, $\Sigma = (\sigma_{ij})$ is a symmetric covariance matrix, $\sigma_{ij} = \sigma^2$ and $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$

The study optimised the return of the portfolio by maximizing the objective function, thus

$$\max_{w} R_p = \sum_{i=1}^{N} \mu_i w_i = \boldsymbol{w}^T \boldsymbol{\mu} \tag{3.3.1}$$

where w_i is the weight of loan i. The variance of the portfolio was minimised by considering

$$\min_{w} \sigma_p^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ij} w_i w_j = \frac{1}{2} \boldsymbol{w}^T \Sigma \boldsymbol{w}, \tag{3.3.2}$$

subject to the constraint

$$\sum_{i}^{N} w_i = 1,\ 0 \leq w_i \leq 1, i = 1, \ldots N \tag{3.3.3}$$

where $\rho_{ii} \equiv 1$, N is the number of loans available and w_i are the decision variables giving the composition of the loan portfolio; the weight rating per loan i. This is a multi-objective management problem with two opposing objectives. The first is to , simultaneously, optimize the return of the loan portfolio and reduce the volatility (risk) of the loan portfolio. Equation (5.2.3) sets the limits for this loan portfolio optimization problem. That is, equation (5.2.3) implies that all available loans are assigned to a portfolio, and the amount of the various percentages of funds in the portfolio is 1, thus $0 \leq w_i \leq 1$ guarantees that the weights of the return rates of the loans in the portfolio are non-negative and reduces the risk of short selling.

For the purpose of finding a vector $\boldsymbol{w}$ that minimizes the variance $\boldsymbol{w}^T \Sigma \boldsymbol{w}$, a Lagrangian function is formulated as follows:

$$\boldsymbol{L}(\boldsymbol{w}, \lambda_1, \lambda_2) = \frac{1}{2}\boldsymbol{w}^T \Sigma \boldsymbol{w} + \lambda_1(\boldsymbol{w}^T \boldsymbol{\mu} - R_p) + \lambda_2(\boldsymbol{w}^T \mathbf{1} - 1) \tag{3.3.4}$$

To find the optimal parameters derivatives were equated to the Lagrangian, with respect to $\boldsymbol{w}, \lambda_1$ and λ_2 to zero and solve the resulting equations for values of the parameters at an optimum value of the variance. Thus we have:

$$\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{w}} = \Sigma \boldsymbol{w} + \lambda_1 \boldsymbol{\mu} + \lambda_2 \mathbf{1} = 0, \tag{3.3.5}$$

$$\frac{\partial \boldsymbol{L}}{\partial \lambda_1} = \boldsymbol{w}^T \boldsymbol{\mu} = 0, \tag{3.3.6}$$

$$\frac{\partial \boldsymbol{L}}{\partial \lambda_2} = \boldsymbol{w}^T \mathbf{1} = 0 \tag{3.3.7}$$

From equation 3.3.5, there is

$$\Sigma \boldsymbol{w} = -(\lambda_1 \boldsymbol{\mu} + \lambda_2 \mathbf{1})$$

Hence there is

$$\boldsymbol{w} = -\Sigma^{-1}(\lambda_1 \boldsymbol{\mu} + \lambda_2 \mathbf{1}) \tag{3.3.8}$$

Multiplying equation 3.3.8 by $\boldsymbol{\mu}^T$ we obtain

$$\mu^T \boldsymbol{w} = -\lambda_1(\Sigma^{-1}\boldsymbol{\mu})\boldsymbol{\mu}^T - \lambda_2(\Sigma^{-1}\mathbf{1})\boldsymbol{\mu}^T \tag{3.3.9}$$

Since $\boldsymbol{\mu}^T \boldsymbol{w} = R_p$, it follows that

$$-\lambda_1 \boldsymbol{\mu}^T \Sigma^{-1} \boldsymbol{\mu} - \lambda_2 \boldsymbol{\mu}^T \Sigma^{-1} \mathbf{1} = R_p \tag{3.3.10}$$

Similarly, multiplying equation 3.3.8 by $\mathbf{1}^T$ obtained

$$\mathbf{1}^T \boldsymbol{w} = \lambda_1(\Sigma^{-1}\boldsymbol{\mu})\mathbf{1}^T - \lambda_2(\Sigma^{-1}\mathbf{1})\mathbf{1}^T \tag{3.3.11}$$

Since $\mathbf{1}^T \boldsymbol{w} = 1$, follows that

$$-\lambda_1 \mathbf{1}^T \Sigma^{-1}\boldsymbol{\mu} - \lambda_2 \mathbf{1}^T \Sigma^{-1}\mathbf{1} = 1 \tag{3.3.12}$$

By letting

$$A = \mathbf{1}^T\Sigma^{-1}\boldsymbol{\mu}, \quad B = \boldsymbol{\mu}^T\Sigma^{-1}\boldsymbol{\mu}, \quad \text{and} \quad C = \mathbf{1}^T\Sigma^{-1}\mathbf{1}$$

Equation 3.3.10 and equation 3.3.12 can be written as

$$\begin{pmatrix} -B & -A \\ -A & -C \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} R_p \\ 1 \end{pmatrix}$$

The solution to this system is given by

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \frac{1}{BC - A^2} \begin{pmatrix} -C & -A \\ -A & -B \end{pmatrix} \begin{pmatrix} R_p \\ 1 \end{pmatrix}$$

Thus there is

$$\begin{aligned} \lambda_1 &= \frac{-CR_p + R_p}{BC - A^2} = -\left(\frac{CR_p - A}{BC - A^2}\right) \\ \lambda_2 &= \frac{R_pA - B}{BC - A^2} = -\left(\frac{B - R_pA}{BC - A^2}\right) \end{aligned}$$

Hence, the optimal weights given is

$$\boldsymbol{w}^* = \left(\frac{CR_p - A}{BC - A^2}\right)\Sigma^{-1}\boldsymbol{\mu} + \left(\frac{B - R_pA}{BC - A^2}\right)\Sigma^{-1}\mathbf{1}$$

Simplifying there is

$$\boldsymbol{w}^* = \frac{1}{BC - A^2}\left[B\Sigma^{-1}\mathbf{1} - A\Sigma^{-1}\boldsymbol{\mu}\right] + \frac{1}{BC - A^2}\left[C\Sigma^{-1}\boldsymbol{\mu} - A\Sigma^{-1}\mathbf{1}\right] R_p \tag{3.3.13}$$

This is the expression for optimal portfolio weights that minimises the variance for an expected return which this study directly implements to obtain optimal portfolio weights. In the building up of Mean-Variance algorithm one begins by preparing the data to be suitable for the algorithm. While the interest rates charged for these loan requests imply that the borrower makes monthly repayments, which take into account chargeable interest rate according to the level of risk associated with that individual. Those amounts are not the monthly returns on a loan advanced to the borrower. The actual return is obtained through amortisation of the loans making use of the features ”funded amount”, ”interest rate ” and ”loan term”. Below is the mathematical expression of the amortization of the loans.

After a borrower receives a loan for the ”funded amount” $\mathrm{p_i}$, they are expected to make t equal payments, $\mathrm{a_i}$, where

$$\mathrm{a_i} = \mathrm{p_i}\left[\frac{\frac{i_i}{n}(1+\frac{i_i}{n})^t}{(1+\frac{i_i}{n})^t - 1}\right]. \tag{3.3.14}$$

where n corresponds to the number of payments in a year ($n = 12$) and i_i is the interest rate of the loan i. The total return rate then becomes

$$r_{1i} = \frac{t\mathrm{a_i}}{\mathrm{p_i}} - 1. \tag{3.3.15}$$

This is under the assumption that there are full repayments and no prepayments. One then applies the lagrangian method of obtaining weights to the amortized data in order to obtain optimal weights for portfolios.

3.3.2 Particle Swarm Optimization

The Particle Swarm Optimization (PSO) method/algorithm is a meta-heuristic strategy founded by Kennedy and Eberhart (1995). There is little literature on the use of PSO on portfolio optimization from the Peer-to-Peer lending platforms. Loans Portfolio optimization is concerned with the problem of funds allocation, in which one considers the right option to spend given the sum of money in a given group of loans. While the challenge of providing low risk and maximum return appears straightforward, there is more than one way of creating an optimal loan portfolio. This study utilized the theory of PSO to optimally select portfolio of loans for investors. Prior to showing how the algorithm was used, a background of the PSO is given below.

Zhan et al. (2009) observes that the PSO uses a simple mechanism that mimics swarm behavior in birds flocking and fish schooling to guide the particles to search for globally optimal solutions. Initial simulations were modified to incorporate nearest-neighbor velocity matching, eliminate ancillary variables, and incorporate multidimensional search and acceleration by distance (Kennedy and Eberhart, 1995; Eberhart and Shi, 2001).

The evolutionary cycle is initiated by a set of random particles (solutions). The i_{th} particle (solution) is defined by its location as a point in the D-dimensional space where D is the number of variables. During the current lifecycle, each particle tracks three values: The present (current) location $X_i = x_{i1}, x_{i2}, ..., x_{iD}$, the maximum (best) location reached in the previous cycles $(P_{loc(i)} = p_{i1}, p_{i2}, ..., p_{iD})$, and its traveling velocity $(V_i = v_{i1}, v_{i2}, ..., v_{iD})$.The $P_{glo(g)}$ location of the best particle $(glo(g))$ is determined as the best fitness of all particles in each period. Consequently, each particle updates its velocity V_i in order to keep up with the strongest particle g, as follows (Zhan et al., 2009):

$$V_{id}^{k+1} = \omega V_{id}^{k} + c_1 r_1^k (P_{loc(id)}^k - X_{id}^k) + c_2 r_2^k (P_{glo(gd)}^k - X_{id}^k) \tag{3.3.16}$$

$$X_{id}^{k+1} = X_{id}^{k} + V_{id}^{k+1},\ V_{max} \geq V_{id} \geq -V_{max} \tag{3.3.17}$$

where $d = 1, \dots D$, $i = 1, \dots N$ and N is the size of the swarm, c_1 and c_2 are two positive constants known as learning variables, r_1 and r_2 are two random parameters in the range $[0, 1]$, and V_{max} is the upper limit for the maximum increase in particle velocity. The parameter ω is an inertia weight that is used to monitor the effect of the past background of velocity on new velocity and plays a part in integrating global and local searches (Elbeltagi, 2013; Eberhart and Shi, 2001).

PSO conducts a search using a population of random solutions, corresponding to an individual. Each potential solution called particle is also assigned a randomized velocity. Each particle in PSO flies in the hyperspace with a velocity which is dynamically adjusted its position according to their own and their neighboring-particles experience, moving toward two points: the best position so far by itself is called Pbest and by its neighbor is called Gbest at every iteration. The particle swarm optimization concept consists of, at each time step, changing each particle's velocity toward its Pbest and Gbest.

In this case the researcher implemented PSO for the loan portfolio optimization by initially assuming that data (The amortized return rates mentioned above) represents initial particles (solutions). Each particle is a vector whose components are the number of grades for each index. Each grade has weights which specify its location/position by the return rate it possesses. The weights for each index must collectively sum up a unit. Each particle retains the memory of its best prior location as well as the best previous position visited by every other particle in the population throughout its lifetime. That is, a particle travels in the direction of its best prior position and the direction of the best particle. With regard to its previous position (weights combination) where it has met the greatest fitness value and the neighbor's previous position where the neighbor has reached the best fitness value. A particle travels in solution space with respect to the neighbor's previous position (weights combination). Particle motions (changes in position) and velocity are influenced by the weights in the sense that the solution space travels about, choosing the locations based on their relative importance in the solution space. In this instance, fitness function was specified as the trade-off between risk and return (That is the minimum volatility and Sharpe ratio). If an improvement in any of the best fitness values is detected during any of the iterations, the particle's personal best position and the particle's best neighbor in the population are updated throughout each iteration.

3.3.3 Genetic Algorithm

Genetic Algorithm (GA) is a stochastic search strategy focused on a framework for natural selection and natural genetics. The critical focus of GA studies is to maintain a balance between exploitation and exploration in the search for an optimal solution for survival in different environments. The algorithm has proven to provide a comprehensive search in complex search spaces (Lin and Gen, 2007). Different from traditional search methods, the genetic algorithm begins with an initial collection of random solutions (population). Every individual in the population is called a chromosome, which is a solution to the problem. The chromosomes, called generations, evolve through successive iterations. The chromosomes are determined by taking certain fitness measurements for each generation. To establish new chromosomes with the next generation, called offspring. The offspring is formed by combining the two current-generation chromosomes using the crossover operator and modifying a chromosome using the mutation operator. A new generation is selected according to the fitness values of parents and offspring and then filters out weak chromosomes to maintain the scale of the population stable. The algorithms merge into the most influential chromosome and are ideally the optimal or sub-optimal approach to the problem (Lin and Gen, 2007).

Through the use of a GA the researcher began by choosing a population from the amortized return rates at random and defining a fitness function that was comparable to the fitness function in the PSO, which was then tested. The algorithm chose an elite population, which became the new population, based on the fitness function, in order to allow for the formation of generations. Prior to deciding on this new population, the researcher changed it using the mutation operator and then applied cross over to it in order to produce an elite next generation of individuals. The study made use of two different crossover techniques, the Arithmetic Crossover and the Heuristic Crossover, to get results. That is, each index of weights was mutated and then crossed over the other index of weights to produce a new generation of weights. This technique is then used to produce the next generation of elites, and so on.

3.4 Implementation

Data visualisations are done using Python, MATLAB, tableau and qliksense while the implementations of the Mean-Variance, PSO, GA and RL algorithms are done through the python packages.

3.5 Evaluation Parameters

The Sharpe Ratio is the difference between the expected return and the risk-free return on a portfolio divided by the standard deviation of the excess return on a portfolio (total risk). This ratio is large if the difference is small. It is a metric for the risk-adjusted return measurement developed by Nobel Laureate William F. Sharpe and widely used in industry. The ratio allows one to quantify the relationship with the average expected return obtained in excess of the risk-free cost per unit of uncertainty or total risk. It is calculated by using the formula

$$SharpeRatio = \frac{R_p - R_f}{\sigma_p}$$

where, R_p is the expected return from a portfolio, R_f the free rate and σ_p is the standard deviation of the portfolio's mean return.

The efficient frontier is an increasing curve, which represents the best trade-off between expected return and variance (risk). Portfolios associated with the lowest risk for a given amount of expected return form what is called the Efficient Frontier. For any level of the desired expected return, this efficient frontier represents the best way for capital investment Fernández and Gómez (2007); Cura (2009); Chang et al. (2009).

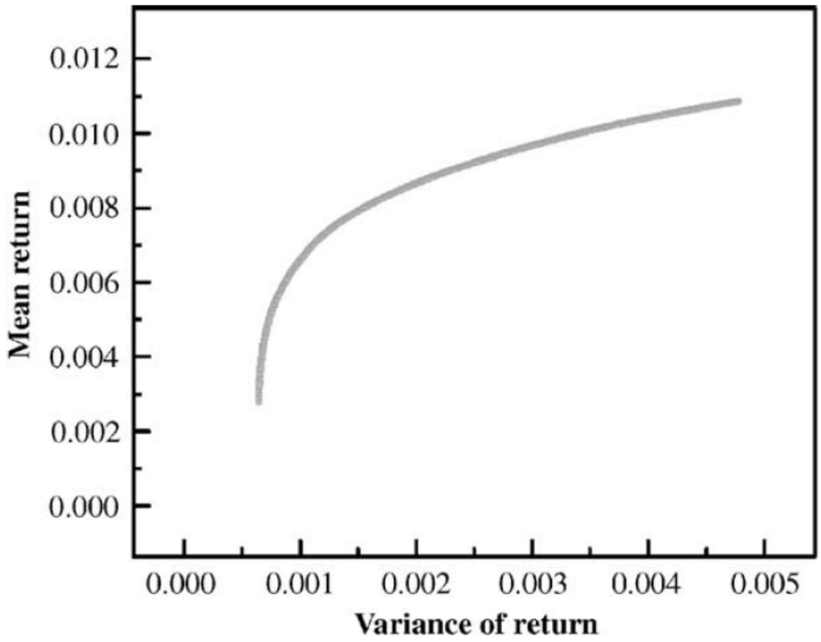


Figure 3.1: Efficient Frontier
Fernández and Gómez (2007)

3.6 Ethical considerations

The study used data from the LendingClub which is publicly available. This implies that there was no need to sign any non-disclosure agreement forms from the LendingClub since the data is available for everyone to use. Thus, there is no need for consents. The data does not contain any personal information of the borrowers and the lenders. It is structured in a way that it is accessible to everyone anonymously.

Chapter 4

Exploratory Data Analysis

This chapter focused on the exploration and understanding of the data. It lays out the manner in which the data was prepared for modelling purposes and the preprocessing.

4.1 Data description and preprocessing

The tables below highlight the dataset of the accepted loans, provide a preview of some of the variables which are worth noting in this study. Table 4.1 shows grades in their different levels focusing on loan amount ranges on each grade, the interest rate given to that loan, installments required from the borrower, the annual income of the borrower and the count (frequency) of loans in each grade. Thus, the higher the risk the higher the reward. The grades are graded in descending order in terms of high to low chances of defaulting, where grades A and G, respectively, represent lowest and highest risks. The interest rates in the case of high risk borrowers, and subsequently the installments, are also high since they are correspondingly associated with high chance of defaulting on repayment.The unit currency for this data is in US Dollars ($)

Grade	Loan_Amnt ($)	int_rate	installment ($)	annual_inc ($)	Frequency
A(lowest risk)	$500 - 40000$	$5.31 - 9,63$	$14.77 - 1268.46$	$0 - 9573072$	433027
B	$500 - 40000$	$6.00 - 14.09$	$15.91 - 1347.38$	$0 - 110000000$	663557
C	$500 - 40000$	$6.00 - 17.27$	$16.47 - 1424.32$	$0 - 9930475$	650053
D	$500 - 40000$	$6.00 - 22.35$	$7.61 - 1534.88$	$0 - 10999200$	324424
E	$600 - 40000$	$6.00 - 27.27$	$4.93 - 1628.08$	$0 - 8500000$	135639
F	$1000 - 40000$	$6.00 - 30.75$	$27.82 - 1714.54$	$0 - 2500000$	41800
G(highest risk)	$600 - 40000$	$6.00 - 30.99$	$21.59 - 1719.83$	$0 - 980000$	12168

Table 4.1: Grades Distribution

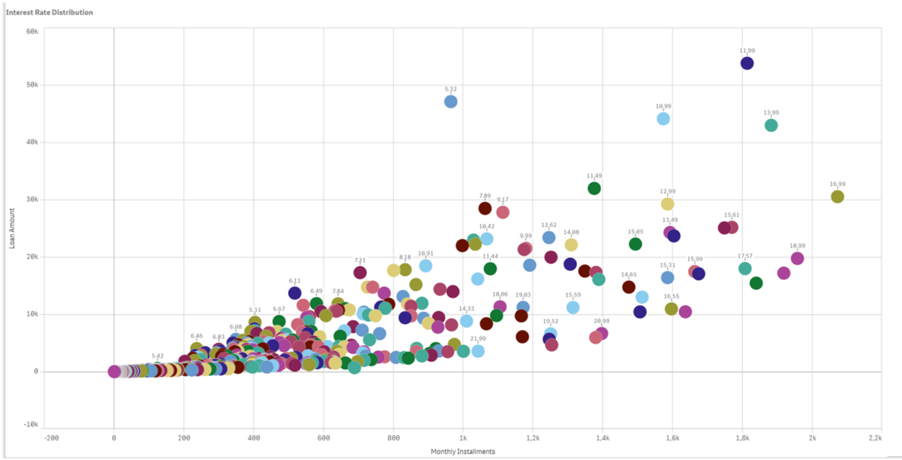
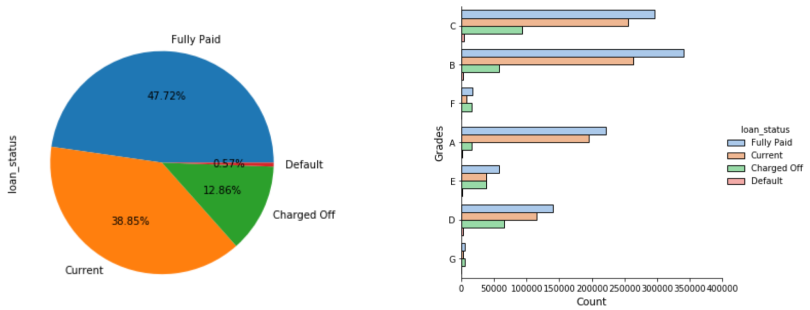


Figure 4.1: Interest Rate Distribution

Furthermore, within the grades the interest rates are charged based on personal information such as credit score, annual income to mention but a few. They range from 5.31% to 30.99%, having also considered loan term as well as the default rate in some cases of high risk borrowers. The study further investigated the status of the loan filtering with grades shown in figure 4.2.



(a) Loan Status Percentages

(b) Grades distribution by Loan Status

Figure 4.2: Loan status and grades

The loans are categorized in different phases which represent different statuses of the loans. **Current** implies the loan repayment is still running, **Late** repayment is when the period of payment lies between $31 - 120$ days which would then fall into **charged off** loan status when the platform decides the loan will not be paid. The **Late** repayment of $16 - 30$ days implies that the loan is delayed for that period and is automatically **in grace** period and will then fall into **default** given the borrower does not make the payment. The loans that are late by $(16 - 30)$ days, in grace period and default fall under default and those that are late by $(31 - 120)$ days and charged off fall under charged off and there is current and fully paid cases which make the four categories of the loan statuses. Obviously, loans with a high risk have a low percentage of fully paying back borrowers as compared to those with low risk and this is also true in the case of the default rate.

The borrower's credit rating is calculated by FICO (Fair, Isaac, and Company) ranking agencies, and ranges from 610 to 850 in this dataset for the approved loans. This company also helps to grade the rating of the borrowers. The minimum FICO score for a borrower to be considered for a loan is 640 (between 300 and 850 is the generic or classic FICO score). The FICO score is determined from numerous pieces of an individual's credit details obtained from national credit bureaus in the US. The score helps to calculate an individual's probability to default based on personal financial history which comprises of measurement and a combination of many different variables. Exact formulas are not made public, but payment history contributes 35% of the score, 30% goes towards debt burden and the balance on accounts, 15% length of credit history, 10% credit mix (loan types used) and the last 10% goes towards recent new loan searches. The other feature worth considering is the purpose of the loan which is shown by the figure below.

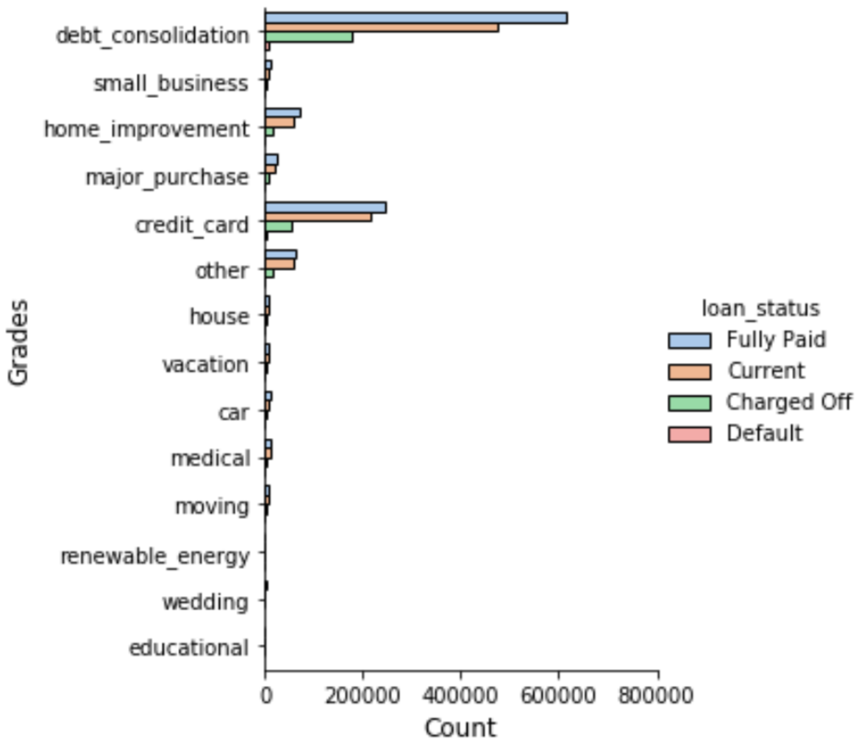


Figure 4.3: Purpose of the loan vs loan status

The bar chart above, Figure 4.3, represents the borrowers' needs for a loan. The study also considered the Debt-to-Income Ratio which also contributes towards the credit rating of the borrower.

Table 4.2 presents a preview of the rejected data which explains why some loans could be rejected. The researcher noticed that the individual's risk score plays a huge role in the approval of the loans.Thus, most of the borrowers whose applications for loans are rejected fall under the category of high risk scores and a high debt-to-ratio income.

Amount Requested ($)	Risk_Score	Debt-To-Income Ratio	Employment Length
6000	698	38.64%	< 1 year
8000	708	10%	< 1 year
2500	573	11.76%	4 years
6100	684	24.69%	2 years

Table 4.2: Rejected Loans

There are many factors which play a role in the approval or rejection of an individual loan application in the Peer-to-Peer lending markets. The unavailability of the financial intermediaries does not mean that anyone is just given loans regardless of their creditworthiness. From the given data the researcher observes that they also consider the financial history of the borrowers.

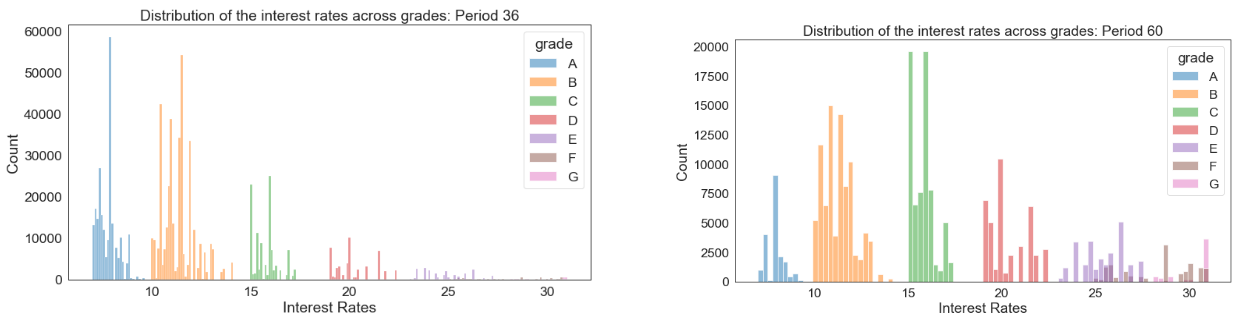
4.1.1 Data Preprocessing and Cleaning

As previously mentioned, the data includes over 150 features and millions of entries/loan re-quests. There are entries with missing details in some of the features, and this messiness had to be cleaned in order for the models to learn from the data. Initially there was a determined percentage of information missing in each column/feature, and if the missing information is greater than 90% $(\geq 90\%)$ those features are excluded because they are insignificant. Those represented have percentage below 90% with the mean or median or mode in these features, depending on the type of the feature, whether it is quantitative or categorical.

The loan demands in the lending club data are only valid for 36 and 60 months, making it necessary to divide the data into two parts based on the term so that the algorithms are implemented with the knowledge of the loan period. In that way the machine can conveniently pick the best portfolios for the investor's preferred time or life of the loan. This means that all algorithms will be run/tested on data-frames with separate loan periods: one for requests made for a loan period of 36 months, and another for requests made for 60 months loan. This procedure allows the algorithm to choose the best portfolio for investors based on the time span in which they are prepared to invest.

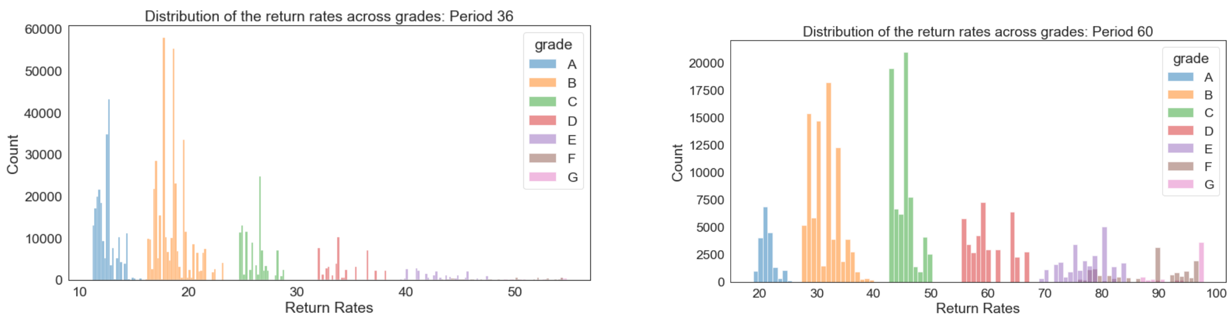
The grades into which the loan requests are classified are critical in the return/loss on a transaction due to the principle of diversification of the loans requests from which the lenders must pick. This relates to the allocation of loans among the various levels of grades (A-G) and their relationship to the other variables that might affect the loan application selection. The loan requests are allocated corresponding to their grades and the associated interest rates for the sake of the models' efficiency as grades, interest rates, and loan terms are the most important features of the selection methods. As a result, it is critical to ensure that the grade allocation is structured for portfolio selection to be as effective as possible. A study could either use stratified sampling, under-sampling, over-sampling (SMOTE), or the researcher can easily determine the number of loans needed in each grade and that number should be the same throughout all the grades. The balancing of grades is for the purpose of the performance of the machine learning algorithms.

The amortization of the interest rates was discussed above and thus below is a representation of both the interest rates and the return rates after the amortization for comparison purposes.



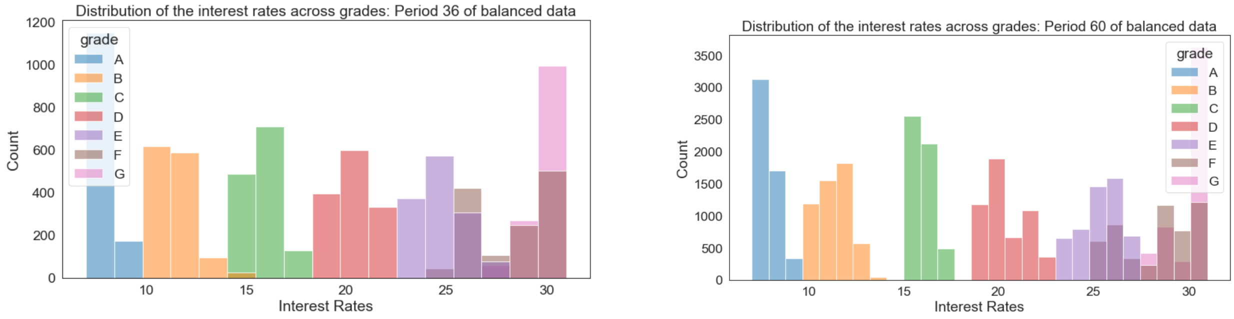
(a) **(b)**

Figure 4.4: Distribution of the interest rates across grades



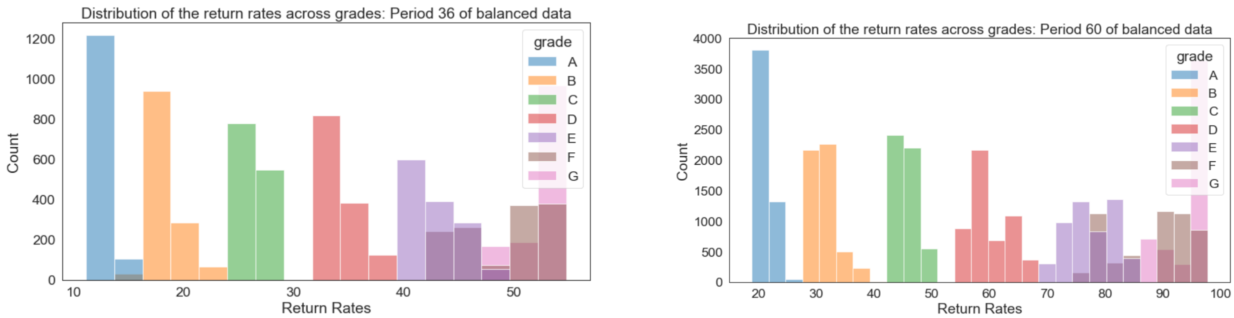
(a) **(b)**

Figure 4.5: Distribution of the return rates across grades



(a) **(b)**

Figure 4.6: Distribution of the interest rates across grades for balanced data



(a) **(b)**

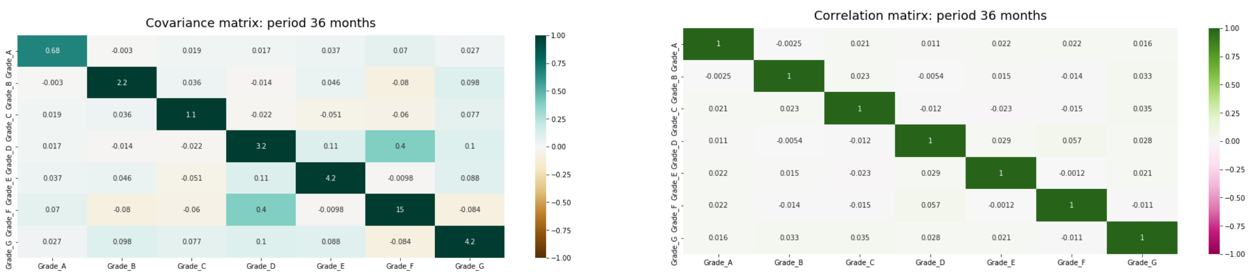
Figure 4.7: Distribution of the return rates across grades for balanced data

It was observed that the higher the risk level a borrower is classified into the higher is the return for the investors. This then concurs with the fact that those in the higher Grade, say A, are of lower risk and therefore charged lower interest rate thus the investors would make a low return if they choose to invest in these individuals compared to those in Grade say G only. If those in G do not default otherwise this assertion would not be true.

These are also divided into categories based on the length of the loan. It can be shown that the return rates for the balanced data in both the 36-month and 60-month periods are evenly spread across the grades when compared to the returns rates for the non-balanced data in both periods. Despite the fact that the random sampling technique that was used to balance the data took all of the loan return rates belonging to Grade_G, which is mainly because it is the Grade with the fewest loan requests, the technique randomly balanced the loan return rates of all the grades; with the grade with the fewest loan return rates serving as the benchmark for all of the grades. That is, the amount of loan return rates in Grade_G was applied as a benchmark to all of the grades at this point. Within the non-balanced data set, it was observed that the levels of loan return rates in Grade_A and Grade_B are high in the 36-month term compared to the 36-month term return rates within the balanced data set. However only a handful of loans were available in the remaining of the loans within the 36-month term balanced data set throughout the same period as a result of the balancing technique that was utilised.

This then shows that for this particular term most of the loan requests are from individuals with a lower risk while those with a higher risk usually are those with little resources and would normally prefer the life of the loan to be longer which would give them more time for repayments. Grade_C are individuals with a moderate level of risk and seem to be inclined towards loan requests of longer periods compared to those in higher grades.That could be attributed to the fact that they are flexible, they could choose to put more resources into loan repayment or put less and spend the money on other needs. From these distribution figures one is able to determine an individual's choice of a period as well as their level of risk.

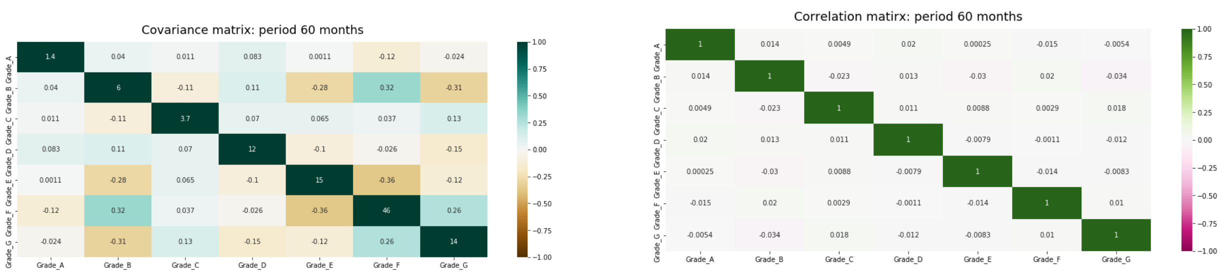
It is also important to investigate the correlation and covariance of the grades of the loan return rates. This enables the machine learning models to perform in an optimal manner when the correlation and covariance of the grades are managed correctly. That is ensuring that the features are not all strongly correlated to one another, their relationship should be diversified. Below are the figures detailing the relationships between the grades of these loan return rates.



(a) Covariance

(b) Correlation

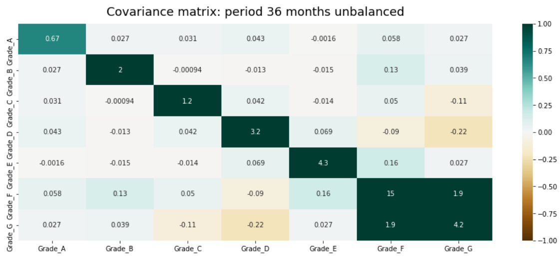
Figure 4.8: Covariance and Correlation of the balanced loan grades in the 36 months period



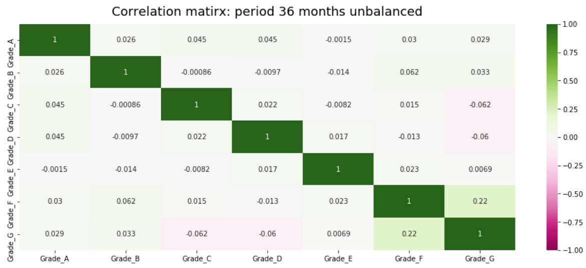
(a) Covariance

(b) Correlation

Figure 4.9: Covariance and Correlation of the balanced loan grades in the 60 months period

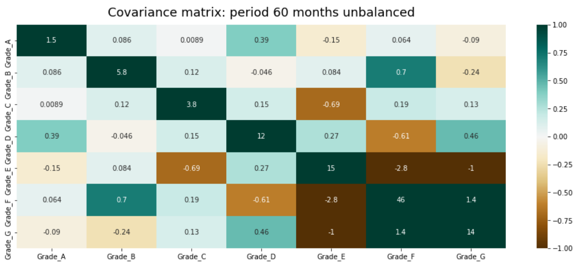


(a) Covariance

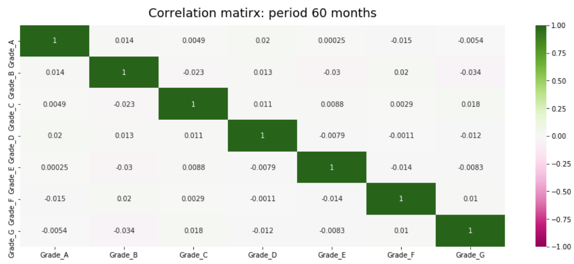


(b) Correlation

Figure 4.10: Covariance and Correlation of the not balanced loan grades in the 36 months period



(a) Covariance



(b) Correlation

Figure 4.11: Covariance and Correlation of the not balanced loan grades in the 60 months period

From these relationships especially from the correlation plots, none of the grades are strongly correlated which makes perfect sense since these grades are more like brackets/categories of individuals who are completely different judging from their risk levels. In the machine learning context it is beneficial to the models when the features are not highly correlated implying that our models could learn effectively from such data. In terms of the covariance, it was observed that there was very low and negative co-variances between these grades. This means that putting together a portfolio would reduce the risk and volatility of an investment. Thus one can conclude that optimization models stand a chance of performing efficiently given the nature of the data set and the relationships among the features.

Chapter 5

Results and Discussion

5.1 Introduction

The purpose of this chapter is to present and explain the findings obtained through the loan portfolio optimization algorithms that were utilized in this research. This research, made use of the highest Sharpe ratio and the minimum amount of volatility as indicators of success. As previously stated, the study made use of four distinct kinds of data sets, each of which is shown in the table below, to complete analysis.

Period	Data-set	Shape
36 months	Balanced	$(1321, 7)$
36 months	Not Balanced	$(359257, 7)$
60 months	Balanced	$(5152, 7)$
60 months	Not Balanced	$(87118, 7)$

Table 5.1: Data Sets shape

These data sets, which are listed in the preceding table, are utilized as input when the codes based on the algorithms discussed above are executed.

5.2 Analysis

Before starting the details of the findings, it is important to review the major elements that contributed to this project. Basically, the aim was to help investors in peer-to-peer lending platforms to make data-driven choices throughout the development of their loan portfolios. Investment opportunities in loans are varied depending on credit quality/grades, resulting in a total of seven "assets" that are made available to potential investors. Using the credit grade score system, each credit grade represents its anticipated value of return (μ_i) and variance (σ_i), with the covariance between the credit grades i and j represented by σ_{ij}. The weight assigned to each grade is denoted by the symbol w_i. In order to choose the best loans for portfolios, the researcher had to use the modern portfolio theory of optimization, which requires a definition of the following variables

$$\text{Expected return: } r_p = \sum_{i=1}^{N} \mu_i w_i = \boldsymbol{w}^T \boldsymbol{\mu} \tag{5.2.1}$$

$$\text{Variance: } \sigma_p^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} \sigma_{ij} w_i w_j = \frac{1}{2} \boldsymbol{w}^T \Sigma \boldsymbol{w} \tag{5.2.2}$$

Investors aim to invest in the most attractive portfolios in which higher returns cannot be achieved at the same level of variation, and lower variance cannot be obtained at the same level of expected returns, as defined by the portfolio's efficiency. This is subject to the restriction outlined below, which states that investors are not permitted to short sell loans.

$$\sum_{i}^{N} w_i = 1,\ 0 \leq w_i \leq 1, i = 1, \ldots N \tag{5.2.3}$$

In order to optimize their profits, investors may either limit their volatility or maximize their Sharpe ratio for a given level of risk exposure.

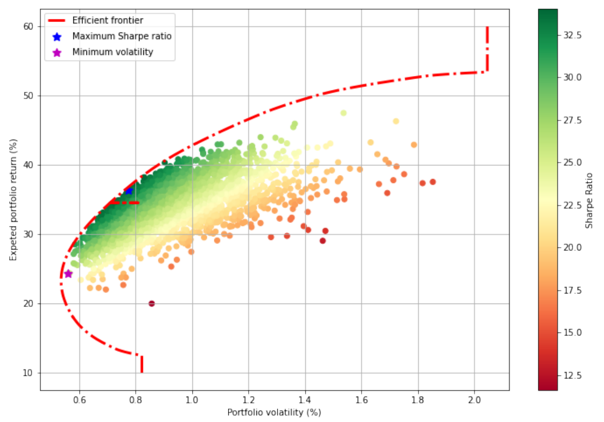
5.2.1 Results from Loans of 36 Months Period

The following is a table which shows a comparison of results obtained by algorithms with the weights of how the funds were be distributed, the expected returns, the volatility and the Sharpe ratio for each loan portfolio. There are several ways to judge these algorithms. For example, one may assess them based on the levels of Sharpe ratios achieved or by evaluating how volatile the investment returns are depending on algorithms used. If they were to be compared based on the highest Sharpe ratios, the GA would have outperformed all of the other algorithms (MVO and PSO), with the highest Sharpe ratio of 48.79%. Within the same loan term and data set, the PSO is the algorithm with the lowest volatility, with a volatility of 0.005506, and in terms of expected returns, MVO has the highest return when compared with the other algorithms. There is a mapping between portfolio returns and volatility, with their respective Sharpe ratios, in the following graphs, see Figures 5.1.

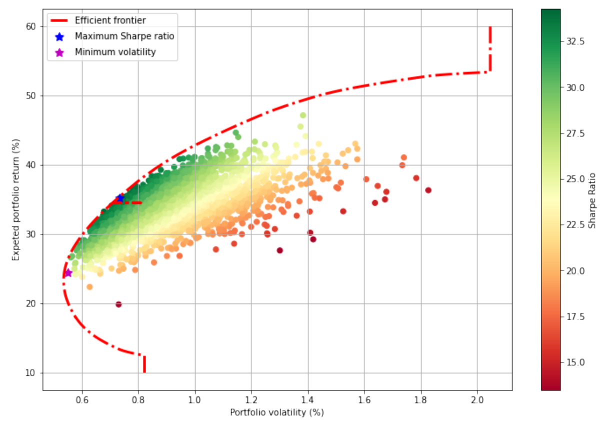
The risks and levels of volatility in each loan portfolio provide investors with the ability to choose an investment portfolio depending on whatever aspect of the portfolio is important to them. Choosing investments that are safe and secure is desirable, because this has higher chances of rewarding the investor with higher expected returns for a defined expected level of risk. By analyzing these efficient frontiers found, the GA algorithm yields portfolios that are volatile with high expected returns, which is good for risk-loving investors but bad for risk-averse ones. The low volatility and expected returns are estimated between 20% to 34%, with the MVO and PSO performing comparably.

Balanced data: Period 36		Return	Volatility	Sharpe - Ratio	Grade_A	Grade_B	Grade_C	Grade_D	Grade_E	Grade_F	Grade_G
Maximum Sharpe Ratio	MVO	0.363633	0.007746	34.035801	0.017179	0.130765	0.232532	0.184882	0.157837	0.050500	0.226305
	PSO	0.351657	0.007342	34.276945	0.118160	0.057570	0.244149	0.152824	0.150771	0.047839	0.228687
	GA	0.320540	0.006569	48.792511	0.180038	0.116171	0.199836	0.145463	0.141811	0.043455	0.173227
Minimum Volatility	MVO	0.243370	0.005614	25.536416	0.387547	0.164831	0.155703	0.127149	0.072021	0.000262	0.092488
	PSO	0.244362	0.005506	26.219255	0.337183	0.170187	0.235524	0.092700	0.095545	0.002448	0.066415
	GA	0.246311	0.005526	44.570974	0.342577	0.129974	0.236313	0.146330	0.083156	0.021139	0.040511

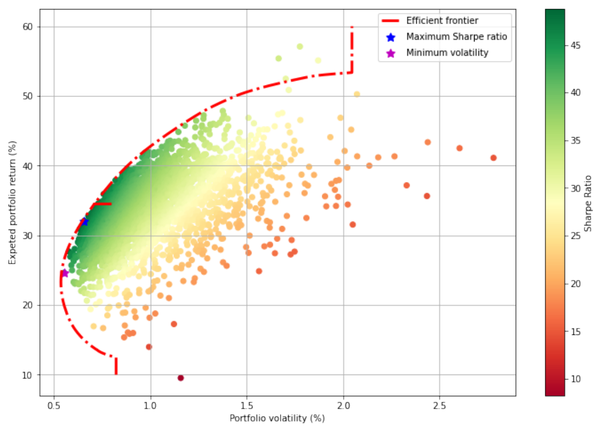
Table 5.2: Minimum Volatility and Maximum Sharpe Ratio of the Algorithms



(a) Mean Variance Optimization



(b) Particle Swarm Optimization



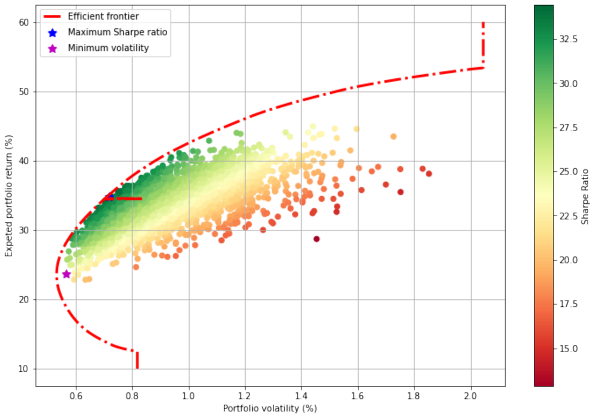
(c) Genetic Algorithm Optimization

Figure 5.1: Efficient frontiers for the 36 months term with balanced data sets

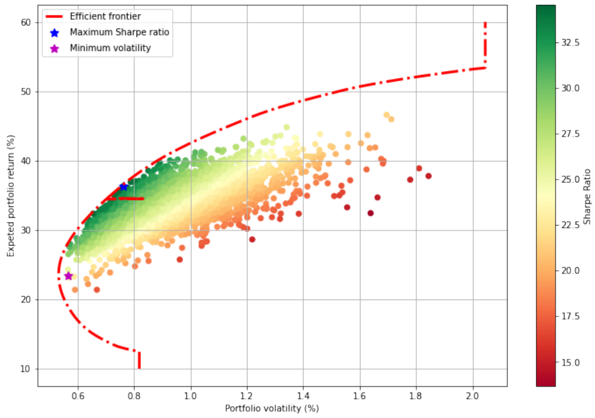
Using unbalanced data set, one looks at a set of results that came from the same term (36 months) as before. The GA had the highest Sharpe ratio of 49.70%, followed by the PSO at 34.90% and lastly the MVO with 34.45%. When it comes to volatility, the PSO has the lowest rate at 0.28%. It is followed by the GA, which has a rate of 0.56%, and the MVO with a rate of 0.56%, all of which are nearly identical in terms of volatility. When it comes to expected portfolio returns, there is always a trade-off between an investor's desire for higher returns, at any level of risk, and their desired targeted return. Some portfolios with high risk do not always have good returns. The worst scenario observed is that of the GA efficient frontier figure, where examples of portfolios with extremely high risk with corresponding low returns, as well as lower Sharpe ratios were observed. This can be linked to bad investment strategy. Furthermore, from the table the GA results indicated lower returns, smaller volatility and larger Sharpe rate while the PSO results are associated with higher returns and volatility, lower Sharpe ratios and finally the MVO results seemed to resemble those of PSO. This is according to the table with the maximum Sharpe ratio. With regard to minimum volatility, PSO results indicated the lowest returns, the lowest risk and medium-class Sharpe ratios, while the GA yields high returns, middle class risk and high Sharpe ratio, and the MVO results are in a lower Sharpe ratio, high volatility and middle class returns.

Unbalanced data: Period 36		**Return**	**Volatility**	**Sharpe - Ratio**	**Grade_A**	**Grade_B**	**Grade_C**	**Grade_D**	**Grade_E**	**Grade_F**	**Grade_G**
Maximum Sharpe Ratio	MVO	0.347918	0.007197	34.446112	0.100259	0.068863	0.239294	0.189817	0.175599	0.014668	0.2115011
	PSO	0.399568	0.008585	34.895039	0.030177	0.085594	0.306060	0.170948	0.156375	0.025790	0.225055
	GA	0.299735	0.006031	49.697442	0.167382	0.160700	0.279082	0.115184	0.101909	0.015065	0.160678
Minimum Volatility	MVO	0.236956	0.005647	24.251289	0.339130	0.251249	0.187573	0.072577	0.048223	0.007472	0.093776
	PSO	0.182032	0.002748	29.849018	0.202316	0.197953	0.213982	0.131247	0.090478	0.053470	0.110554
	GA	0.251860	0.005580	45.135284	0.311945	0.206240	0.190280	0.129084	0.049060	0.014521	0.098869

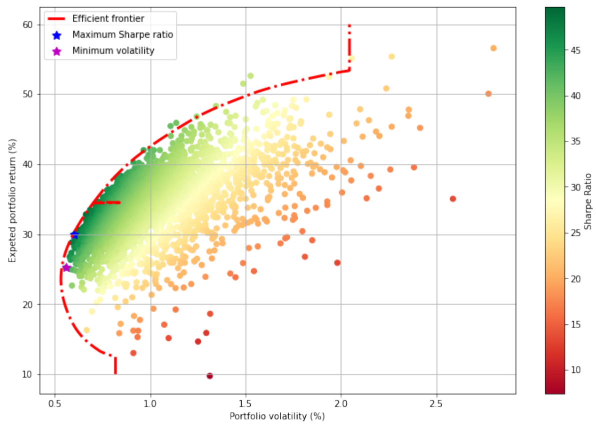
Table 5.3: Minimum Volatility and Maximum Sharpe Ratio of the Algorithms



(a) Mean Variance Optimization



(b) Particle Swarm Optimization



(c) Genetic Algorithm Optimization

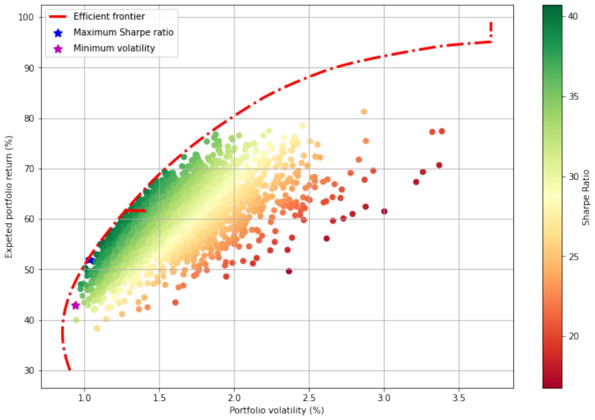
Figure 5.2: Efficient frontiers for the 36 months term with unbalanced data sets

5.2.2 Results from Loans of 60 Months Period

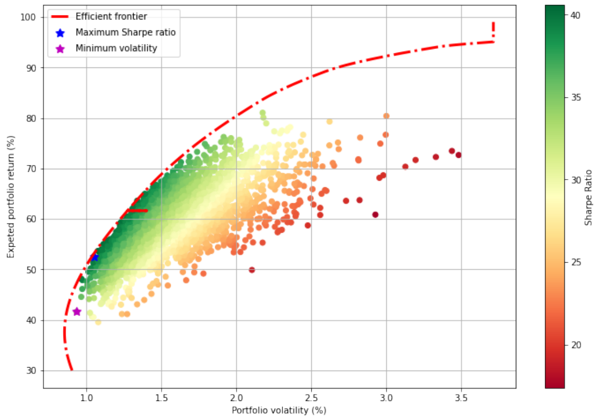
The previous section stated that the loans had two lifespans: 36 and 60 months. The results for the 36-month period are shown above, and in the section below, shows how algorithms perform when choosing loan portfolios with a 60-month lifespan. In fact, loans with longer repayment terms provide greater returns than loans with shorter repayment periods, owing in part to cumulative total interest over a longer period. According to the table below, the GA generated the best possible Sharpe ratio of 49.99% while the PSO generated the lowest possible volatility of 0.33% and the MVO generated the highest returns of $51,97\%$. Taking a look at the weights assigned to these grades, loans in grade A and grade C seem to have been allocated the largest proportion of the money available for investment in these portfolios while the loans in grade F receive the smallest proportion. This is reasonable, considering the lower risk associated with loans in grade A compared to loans in grades G and F, it does align with the goal of loan diversity. This means, portfolios of loans that are much less risky are not just built, but also include loans that are significantly riskier in order to strike a balance between the concepts of high returns and low volatility. These findings may also be seen in the figures 5.3 below. As a result, there are portfolios that are extremely volatile and have greater anticipated returns with lower Sharpe ratios, there are also portfolios that are highly volatile and have higher projected portfolio returns but have a higher Sharpe ratio. Consequently, these algorithms help investors in determining which portfolios to invest in, even when the markets are extremely volatile. When the idea of the Sharpe ratio is incorporated, an investor is able to determine portfolios that are less risky and those that are highly risky.

Balanced data: Period 60		**Return**	**Volatility**	**Sharpe - Ratio**	**Grade_A**	**Grade_B**	**Grade_C**	**Grade_D**	**Grade_E**	**Grade_F**	**Grade_G**
Maximum Sharpe Ratio	MVO	0.519672	0.010304	40.727555	0.258666	0.112619	0.220270	0.110708	0.102941	0.027150	0.167646
	PSO	0.413158	0.007605	41.176020	0.202898	0.107928	0.243870	0.113063	0.123534	0.044397	0.164309
	GA	0.470220	0.009405	49.997773	0.291903	0.160621	0.232796	0.078119	0.101976	0.023286	0.111299
Minimum Volatility	MVO	0.428757	0.009368	35.093581	0.308208	0.175235	0.271746	0.084611	0.085357	0.042866	0.031977
	PSO	0.235158	0.003346	40.396461	0.226773	0.129116	0.219879	0.114830	0.113972	0.042083	0.153347
	GA	0.354619	0.009302	38.121537	0.595806	0.145189	0.082212	0.011486	0.107591	0.007808	0.049908

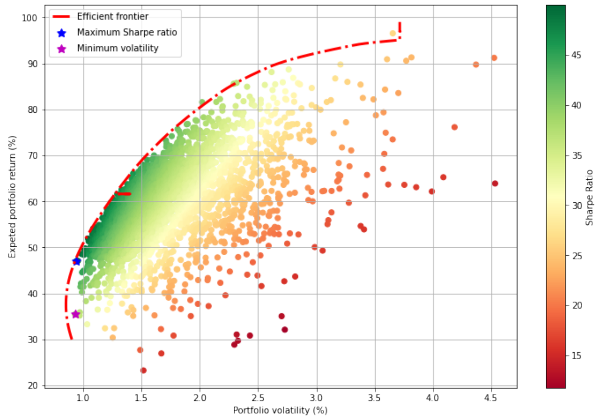
Table 5.4: Minimum Volatility and Maximum Sharpe Ratio of the Algorithms



(a) Mean Variance Optimization



(b) Particle Swarm Optimization



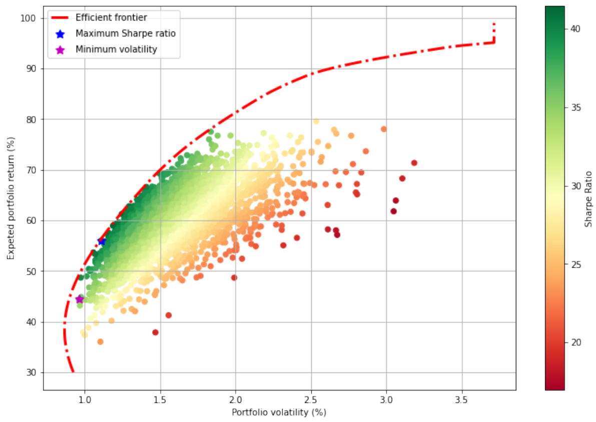
(c) Genetic Algorithm Optimization

Figure 5.3: Efficient frontiers for the 60 months term with balanced data sets

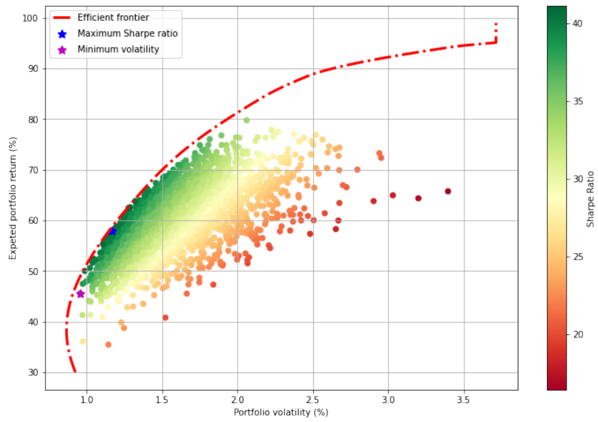
The observations stated above were made in relation to a balanced data set. The following section examines an unbalanced data set and how the algorithms perform in that situation, among other things. According to the table below, GA produced the greatest possible Sharpe ratio of 50.14%, PSO provided the lowest possible volatility of 0.42%, and GA also generated the highest returns of 56.12%. Weights distribution across the grades, reflect a pattern that is similar to the previous results in that the loans in grades A and C get a larger part of the money while the loans in the other grades receive lower amounts of the funds. It is important to note that loans in grades B and D are included in the group of loans that get fewer funds for investment, despite the fact that they are not deemed to be of particularly high risk. This illustrates how the weights assigned to each loan are somewhat evenly spread within the imbalanced data set in question. The findings stated in the table are supported by the figures in 5.4.

Unbalanced data: Period 60		Return	Volatility	Sharpe - Ratio	Grade_A	Grade_B	Grade_C	Grade_D	Grade_E	Grade_F	Grade_G
Maximum Sharpe Ratio	MVO	0.560121	0.011094	41.475088	0.215355	0.081104	0.214400	0.116017	0.163405	0.053902	0.155816
	PSO	0.474570	0.008964	41.785954	0.187464	0.088367	0.259016	0.097164	0.153919	0.048053	0.166016
	GA	0.561253	0.011193	50.142529	0.208761	0.126093	0.193062	0.076472	0.163507	0.047534	0.184570
Minimum Volatility	MVO	0.444130	0.009625	35.752254	0.376132	0.185789	0.108651	0.108051	0.104878	0.009054	0.107447
	PSO	0.273497	0.004263	40.696424	0.223626	0.118873	0.222340	0.091816	0.147592	0.049892	0.145860
	GA	0.407977	0.009400	43.400440	0.531643	0.060187	0.117543	0.105423	0.078042	0.046499	0.060663

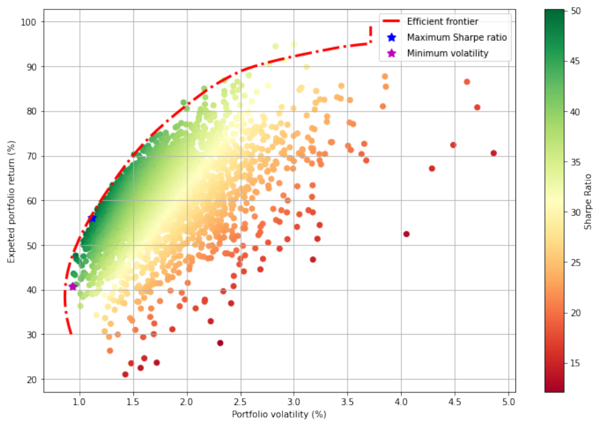
Table 5.5: Minimum Volatility and Maximum Sharpe Ratio of the Algorithms



(a) Mean Variance Optimization



(b) Particle Swarm Optimization



(c) Genetic Algorithm Optimization

Figure 5.4: Efficient frontiers for the term 60 months with unbalanced data sets

The weightings within loan portfolios are completely explicable in terms of the diversification theory. Loan portfolios consisting of groups of loans in grades A, C, E and G have slightly higher weightings than the rest of the portfolios, owing to the fact that they are mixtures of highly risky and less risky loans. At the same time, grades B, D, and F havd their weights lowered somewhat, in keeping with the concept of not putting all eggs in one basket (as in the case of gambling). These results are consistent across all trials, suggesting that the algorithms used in these studies do in fact create loan portfolios that aim to maximize profits while reducing volatility.

5.3 Summary

This section highlights overview of how the algorithms performed when implemented on the four different data sets.

MVO		Return	Volatility	Sharpe - Ratio
Maximum Sharpe Ratio	Balanced_36	0.363633	0.007746	34.035801
	Unbalanced_36	0.347918	0.007197	34.446112
	Balanced_60	0.519672	0.010304	40.727555
	Unbalanced_60	0.560121	0.011094	41.475088
Minimum Volatility	Balanced_36	0.243370	0.005614	25.536416
	Unbalanced_36	0.236956	0.005647	24.251289
	Balanced_60	0.428757	0.009368	35.093581
	Unbalanced_60	0.444130	0.009625	35.752254
PSO		**Return**	**Volatility**	**Sharpe - Ratio**
Maximum Sharpe Ratio	Balanced_36	0.351657	0.007342	34.276945
	Unbalanced_36	0.399568	0.008585	34.895039
	Balanced_60	0.413158	0.007605	41.176020
	Unbalanced_60	0.474570	0.008964	41.785954
Minimum Volatility	Balanced_36	0.244362	0.005506	26.219255
	Unbalanced_36	0.182032	0.002748	29.849018
	Balanced_60	0.235158	0.003346	40.396461
	Unbalanced_60	0.273497	0.004263	40.696424
GA		**Return**	**Volatility**	**Sharpe - Ratio**
Maximum Sharpe Ratio	Balanced_36	0.320540	0.006569	48.792511
	Unbalanced_36	0.299735	0.006031	49.697442
	Balanced_60	0.470220	0.009405	49.997773
	Unbalanced_60	0.561253	0.011193	50.142529
Minimum Volatility	Balanced_36	0.246311	0.005526	44.570974
	Unbalanced_36	0.251860	0.005580	45.135284
	Balanced_60	0.354619	0.009302	38.121537
	Unbalanced_60	0.407977	0.009400	43.400440

Table 5.6: Minimum Volatility and Maximum Sharpe Ratio of the Algorithms

In overall, the returns on loans with a term of 36 months seem to be lower than those on loans with a term of 60 months, and this appears to be due mostly on the lower credit ratings assigned to loans with a term of 36 months as opposed to 60 months. Similar to this, portfolios with a duration of 36 months have lower volatility than portfolios with a period of 60 months, and the similar trend is also seen in Sharpe ratios. As a result, portfolios with a holding time of 60 months would generate greater returns while also experiencing more volatility as compared to portfolios with a holding period of 36 months. This observation supports the concept in finance with regards to the longer the period the higher the returns due to higher interest rates charged on those. It is probable that if one chose diversity by choosing portfolios from both the 36-month and 60-month time periods one would achieve higher portfolio yields. But that was not the goal or aim of this study.

Data was divided into two groups: the balanced data set and the unbalanced data set. According to the Sharpe ratios, the unbalanced data seems to contain a greater quantity of ratios than the balanced data. The same is true for volatility, where the portfolios of the unbalanced data set are more risky than the portfolios of the balanced data set. A similar pattern may be seen in the returns on the portfolio. This indicates that the unbalanced data set is extremely volatile while at the same time yielding greater returns on the investment. For the sake of future planning, balancing the data does not always suit the goals of risk-loving investors who want high returns.

Observing how the algorithms have performed in comparison to one another, the information provided on the table above considered. According to the findings, the GA has produced better portfolio compared to the other two algorithms being implemented in this study. It also produces higher Sharpe ratios while maintaining a moderate and or high degree of risk. The MVO produced portfolios with a high level of risk, low Sharpe ratios, and moderate returns, while the PSO produced portfolios with a very low level of risk and return, as well as intermediate Sharpe ratios. Clearly, the other two algorithms are also capable of meeting certain objectives of investors depending on the goals. Investment in portfolios from the PSO would be appropriate for investors who do not want to expose themselves to very high risk, while those who seek greater returns at any level of risk would choose the GA and MVO.

Chapter 6

Conclusion

6.1 Potential limitation

One of the challenges in loan portfolio optimization of the Peer-to-Peer lending platforms, in South Africa, is the lack of actual data or that Peer-to-Peer lending platforms are not popular. Due to unavailability of reliable Peer-to-Peer lending platform data sets the portfolio optimization cannot be easily implemented. In South Africa the information on borrowers is confidential thus rendering the operations of Peer-to-Peer lending markets to be constrained. This is because the market is fairly new in this country. Therefore, due to the unavailability of reliable Peer-to-Peer lending platform data sets the portfolio optimisation techniques cannot be easily implemented.

6.2 Conclusions

Investors on a Peer-to-Peer platform may or may not have quality information about how to invest and what it takes to build a portfolio. This is where this research comes in, to enable those investors who have little understanding of investments or portfolio management to engage in these investments with through the recommendations from this study. Clearly, from the above explanations about the quality of information produced by different types of algorithms with different capabilities one can observe that everything is dependent on the goals of each investor in as far as choosing investment portfolios is concerned. That is, investors have the option of either limiting their volatility or increasing their Sharpe ratio for a given amount of risk. All of this is dependent on the investor's risk appetite as well as the amount of accessible money. The following is the overall ranking of the performance of various algorithms: Compared to other strategies, the GA generates better portfolio returns of about 56.13% and higher Sharpe ratios of approximately 50.15%, while maintaining a moderate risk level of approximately 0.94%. The PSO generates lower returns of about 27.35% and moderate Sharpe ratios of approximately 41.79%, while also having the lowest volatility of approximately 0.55%. The MVO produces portfolios with very high risk (about 1.11%) and high return (approximately 56.01%), as well as very low Sharpe ratios (approximately 34.45%).

The most important finding from this study is that investors must use the information derived from this study to decide on the optimal distribution of investments based on available information on credit grades. The investors must also take into account that higher risk loans may sometimes fail to generate sufficient returns to cover the risks taken since the loans with the highest expected returns are those with high risk. As a result, investors should diversify their loan portfolios, and the number of loans should increase with the increase in risk associated with a particular credit grade.

Further research could be a continuation of this work, for example to concentrating on the implementation of the Reinforcement learning algorithm in comparison to the already used methods. According to the literature review as indicated in chapter 2, study revealed that the Reinforcement algorithm performs better for such portfolios. It would be interesting to consider diversification of loan portfolios that combines loans in 36 and 60 term periods as, probably, that could be interpreted to mean diversification of loan portfolio and a diversified environment. It would also be beneficial to include other variables associated with the borrower and the investor to the asset allocation so as to observe how the algorithms performance will improve.

For further references about the algorithms please see Algorithms or go to the next url: https://github.com/Maakgetlwa-Shoky/MSc-PO-Project/tree/master

References

S. Almahdi and S. Y. Yang. A Constrained Portfolio Trading System Using Particle Swarm Algorithm and Recurrent Reinforcement Learning. *Expert Systems with Applications*, 4 2019.

A. S. Arnold, J. S. Wilson, M. G. Boshier, and J. Smith. A Simple Extended-Cavity Diode Laser. *Review of Scientific Instruments*, 69(3):1236–1239, 3 1998.

G. Babaei and S. Bamdad. A multi-objective instance-based decision support system for investment recommendation in Peer-to-Peer lending. *Expert Systems with Applications*, 2 2020.

Basel Committee. The Basel Committees Response to the Financial Crisis Report to the G20. Technical Report 4002, Basel Committee on Banking Supervision, Oct. 2010.

S. C. Berger and F. Gleisner. Emergence of Financial Intermediaries in Electronic Markets: The Case of Online P2P Lending. *Business Research*, 2(1):39–65, May 2009. ISSN 2198-2627.

T.-J. Chang, S.-C. Yang, and K.-J. Chang. Portfolio optimization problems in different risk measures using genetic algorithm. *Expert Syst. Appl.*, 36:10529–10537, 2009.

G. Chi, S. Ding, and X. Peng. Data-Driven Robust Credit Portfolio Optimization for Investment Decisions in P2P Lending. *Mathematical Problems in Engineering*, pages 1–10, 2019.

M. C. Cohen, C. D. Guetta, K. Jiao, and F. Provost. Data-Driven Investment Strategies for Peer-to-Peer Lending: A Case Study for Teaching Data Science. *Big Data*, 6, 09 2018.

T. Cura. Particle swarm optimization approach to portfolio optimization. *Nonlinear Analysis: Real World Applications*, 10:2396–2406, 08 2009. doi: 10.1016/j.nonrwa.2008.04.023.

L. Diosan. A multi-objective evolutionary approach to the portfolio optimization problem. volume 2, 2005.

G. Dounias, P. Xia, Z. Ni, X. Zhu, and L. Ni. A Novel Key Influencing Factors Selection Approach of P2P Lending Investment Risk. *Mathematical Problems in Engineering*, 2019: 6086089, 2019.

R. Eberhart and Y. Shi. Particle swarm optimization: developments, applications and resources. In *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, volume 1, pages 81–86, 2001.

E. E. Elbeltagi. *Metaheuristic Applications in Structures and Infrastructures: Swarm Intelligence for Large-Scale Optimization in Construction Management*. 2013.

I. Erlich, G. Venayagamoorthy, and N. Worawat. A mean-variance optimization algorithm. In *IEEE Congress on Evolutionary Computation*, pages 1–6, 2010.

A. Fernández and S. Gómez. Portfolio selection using neural networks. *Computers Operations Research*, 34(4):1177–1191, 2007. ISSN 0305-0548. doi: https://doi.org/10.1016/j.cor.2005.06.017. URL https://www.sciencedirect.com/science/article/pii/S0305054805002042.

A. Filos. Reinforcement learning for portfolio management, 2019.

J. Geldddenhuys. National credit amendment bill - why banks are concerned. *Moonstone*, Aug. 2019.

C. Gerard and T. Reha. *Optimization Methods in Finance*. Mathematics, finance, and risk. Cambridge University Press, 2007.

Y. Guo, W. Zhou, C. Luo, C. Liu, and H. Xiong. Instance-based credit risk assessment for investment decisions in P2P lending. *European Journal of Operational Research*, 249(2): 417–426, March 2016.

C. J. Hawthorn, K. P. Weber, and R. E. Scholten. Littrow Configuration Tunable External Cavity Diode Laser with Fixed Direction Output Beam. *Review of Scientific Instruments*, 72(12):4477–4479, 12 2001.

F. d. M. Helder and S. F. José. Central Bank Transparency and Financial Market: Evidence for the Brazilian. *Brazilian Society of Finance*, 9(1):51–67, 2011.

S. Herrero-Lopez. Social Interactions in P2P Lending. In *Proceedings of the 3rd Workshop on Social Network Mining and Analysis*, SNA-KDD '09, pages 1–8, 2009.

J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948, 1995.

M. Kunal. Beyond FICO: Default Prediction and Optimal Lending Strategies in Online P2P Investing. candthesis, Harvard College, 2016.

C.-M. Lin and M. Gen. An effective decision-based genetic algorithm approach to multiobjective portfolio optimization problem. 2007.

C. Luo, H. Xiong, W. Zhou, Y. Guo, and G. Deng. Enhancing Investment Decisions in P2P Lending: An Investor Composition Perspective. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 292–300, 08 2011.

H. Markowitz. Portfolio Selection. *Journal of Finance*, 7:77–9, 1952.

A. Mateescu. Peer to peer lending. *Data and Society Research Institute*, 2, July 2015.

S. K. Mishra, G. Panda, and S. Meher. Multi-objective particle swarm optimization approach to portfolio optimization. pages 1612–1615, 2009.

E. Mollick. The dynamics of crowdfunding: An exploratory study. *Journal of Business Venturing*, 29(1):1–16, 2014.

N. Y. Nhi, X. He, L. Shaowu, and X. Guandong. Deep Learning for Decision Making and the Optimization of Socially Responsible Investments and Portfolio. *Decision Support Systems*, 7 2019.

L. V. C. Oswaldo and V. A. Michael. A generalized multi-period mean–variance portfolio optimization with Markov switching parameters. *Automatica*, 44, 2008.

A. Saud and S. Y. Yang. An adaptive portfolio trading system: A risk-return portfolio optimization using recurrent reinforcement learning with expected maximum drawdown. *Expert Systems with Applications*, 87, 11 2017.

South African Goverment. National Credit Amendment Act: Act No. 7 of 2019. Technical report, Parliametary Monitoring Commitee, Aug. 2019.

C. Szepesvari. *Algorithms for Reinforcement Learning*. Morgan and Claypool Publishers, 2010.

H. Thomas, B. Y. Qu, Q. Zhou, J. M. Xiao, J. J. Liang, and P. N. Suganthan. Large-Scale Portfolio Optimization Using Multiobjective Evolutionary Algorithms and Preselection Methods. *Mathematical Problems in Engineering*, 2017:1–14, 2017.

B. Tomas, M. Agatha, and Y. Z. Xun. Mean-Variance Portfolio Optimization with State-Dependent Risk Aversion. *Mathematical Finance*, aop, 2012.

C. Wei, Z. Run-tong, C. Yong-ming, and F.-s. Xu. Particle Swarm Optimization for Constrained Portfolio Selection Problems. 2006.

C. E. Wieman and L. Hollberg. Using Diode Lasers for Atomic Physics. *Review of Scientific Instruments*, 62(1):1–20, 1 1991.

G. Xianping, Y. Liuer, and Y. George. A mean–variance optimization problem for discounted Markov decision processes. *European Journal of Operational Research*, 220, 2012.

B. Xing and T. Marwala. *Smart Computing Applications in Crowdfunding*. CRC Press, 2018.

Z. Zhan, J. Zhang, Y. Li, and H. S. Chung. Adaptive particle swarm optimization. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 39(6):1362–1381, 2009.

H. Zhao, Q. Liu, G. Wang, Y. Ge, and E. Chen. Portfolio Selections in P2P Lending: A Multi-Objective Perspective. In *Proceedings of the 22nd ACM SIGKDD: International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 2075–2084, 2016.

X.-l. Zheng, M.-y. Zhu, Q.-b. Li, C.-c. Chen, and Y.-c. Tan. Finbrain: When Finance Meets AI 2.0. *Frontiers of Information Technology and Electronic Engineering*, 20, 7 2019.