**Stock Price Prediction with LSTM and Portfolio Optimization using Markowitz's Portfolio Optimization Model**

A picture containing graphical user interface

Description automatically generated

**Stock Price Prediction with LSTM and Portfolio Optimization using Markowitz's Portfolio Optimization Model**



**University of Westminster**

Business School

A thesis submitted in fulfilment of the requirements

for the degree of Master of Business Analytics with Fintech

Author: Ms. Nishat Zereen

Supervisor: Dr. Jason Cheng

Module Leader: Mrs. Ann Thapar

**September 1, 2022**

# Acknowledgement

I'm relieved to have finished the master thesis, which was the most challenging component of my academic program. First, I want to express my gratitude to Dr. Jason Cheng for giving me the responsibility of implementing this to support a specific thesis and to motivate me to focus my research on recently from my thoughts. His input was crucial because he was always available to educate me anytime I had a problem.

I also want to thank my friends and family members for their support and encouragement throughout my academic journey.

# Declaration

I completed this dissertation as part of the University of Westminster MSc. Fintech with Business Analytics, I confirm that, except for some areas where I needed ideas or words from other sources and referenced in the text, everything else is the outcome of my research, and the research data was collected from secondary sources.

**Table of Contents**

[Acknowledgment 3](#_Toc112831797)

[Declaration 3](#_Toc112831798)

[Abstract 6](#_Toc112831799)

[1.1 Introduction 7](#_Toc112831800)

[1.2 Research Objective 7](#_Toc112831801)

[1.3 Research Question and Hypothesis 7](#_Toc112831802)

[1.3.1 Research Sub-questions 8](#_Toc112831804)

[1.3.2 Hypothesis 8](#_Toc112831805)

[Chapter # 2: Literature Review 9](#_Toc112831806)

[2.1 Financial Science 9](#_Toc112831807)

[2.2 Stock Market 9](#_Toc112831808)

[2.3 Available Financial Data 10](#_Toc112831809)

[2.4 Machine Learning And LSTM Model 10](#_Toc112831810)

[2.5 Markowitz's Portfolio Optimisation Model 12](#_Toc112831811)

[2.6 Stock Prediction using Deep Learning 15](#_Toc112831812)

[2.7 Portfolio Construction Model 16](#_Toc112831813)

[2.8 Performance of Combined Model 18](#_Toc112831814)

[Chapter # 3 Research Design 20](#_Toc112831815)

[3.1 Research Philosophy 20](#_Toc112831816)

[3.2 Research Design 20](#_Toc112831817)

[3.3 Research Approach 20](#_Toc112831818)

[Step # 1: Use of LSTM to determine upraising stocks 20](#_Toc112831819)

[Step # 2: Creating an Optimal Portfolio using Marowiz's Algorithm 20](#_Toc112831820)

[Chapter # 4 Research Methodology 23](#_Toc112831821)

[4.1 Sampling and Data Collection 23](#_Toc112831822)

[4.1.1 Industry and Stock Selection 23](#_Toc112831823)

[4.1.2 Data Acquisition 23](#_Toc112831824)

[4.1.3 Application of the LSTM Model 24](#_Toc112831825)

[4.2 Study Variables 24](#_Toc112831826)

[4.3 Ethical Considerations 24](#_Toc112831827)

[Chapter # 5 Data Analysis and Calculations 25](#_Toc112831828)

[5.1 Stock Selection for Portfolio 25](#_Toc112831829)

[5.1.1 Deep Learning (LSTM) Modeling 25](#_Toc112831830)

[5.1.2 RMSE Calculation 26](#_Toc112831831)

[5.2 Constructing Optimal Portfolio with Markowiz's Algorithm 27](#_Toc112831832)

[Chapter # 6 Conclusion & Discussion 30](#_Toc112831833)

[6.1 Study Findings and Result 30](#_Toc112831834)

[6.2 Study Limitations and Recommendations 32](#_Toc112831835)

[6.3 Conclusion 33](#_Toc112831836)

[References 34](#_Toc112831837)

**Table of Figures**

[Figure 1LSTM Architecture (Anita Yadava, 2019) 10](file:///E:\Desktop\Nishat%20Dissertation.docx#_Toc112841699)

[Figure 2 LSTM Prediction Procedure 11](file:///E:\Desktop\Nishat%20Dissertation.docx#_Toc112841700)

[Figure 3 Efficient Frontier 13](file:///E:\Desktop\Nishat%20Dissertation.docx#_Toc112841701)

[Figure 4 Flowchart of Research Approach 20](#_Toc112841702)

[Figure 5 Heatmap of the six stocks 29](file:///E:\Desktop\Nishat%20Dissertation.docx#_Toc112841703)

[Figure 6 Minimal portfolio 30](file:///E:\Desktop\Nishat%20Dissertation.docx#_Toc112841704)

[Figure 7 Optimal Portfolio 30](file:///E:\Desktop\Nishat%20Dissertation.docx#_Toc112841705)

[Figure 8 Pie Chart of Investment Ratio 31](#_Toc112841706)

**List of Tables**

[Table 1 Summary of Acquired Data 23](#_Toc112797131)

[Table 2 LSTM Outcome of Final Sample Data 26](#_Toc112797132)

[Table 3 Daily return and log-returns of the stock…………………………………………………......27](#_Toc112797133)

[Table 4 Covariance & Correlation Matrix*………….………………………………………………………...*28](#_Toc112797134)

[Table 5 Comparison of Sharpe Ratio**s** 30](#_Toc112797135)

[Table 5 Portfolio Output 31](#_Toc112797135)

# Abstract

Intelligent investors want to know about the stocks and their movements before inserting their money into any funds. However, Predicting future stock prices and their movement patterns is difficult. This situation led people to study predicting stock prices and constructing an ideal portfolio for a considerable time. Several techniques have been studied for this purpose, but LSTM and Markowitz are the most widely used Machine Learning techniques. Therefore, it is significantly more challenging to construct a portfolio of capital assets using expected prices to accomplish the optimal between return and risk. The time series of the historical prices of the top five stocks from the seven distinct sectors of the Canadian stock market from July 1, 2017, to June 30, 2022, have been analysed in this study. Selecting stocks from seven sectors is to diversify the investment and reduce risk. From each of these sectors, the best possible combination is created. A long-and short-term memory (LSTM) model is also created and improved for forecasting future stock values. Using the LSTM model results, the researcher applied Markowiz's algorithm to construct an optimal portfolio for investors. Finally, researchers have observed that LSTM And Markowitz's Model is positively related to predicting stock prices and creating an optimal portfolio.

**Chapter # 1 Introduction**

## **1.1 Introduction**

Financial markets can be challenging to analyse due to their complexity and volatility. Consequently, creating a portfolio is a sophisticated optimisation issue involving allocating a given amount of capital to a set of stocks or assets to maximise the investment's return and risk. The problem relates to the class of computations known as NP-hard problems, which no algorithm can solve in polynomial time. The difficulty of the challenge increases even further if the optimisation methods need a solid prediction of the returns and risks associated with each portfolio stock, as accurately predicting the future values of stock prices is also exceedingly challenging. Even though Markowitz's seminal work (Markowitz, 1952) created the mathematical framework for portfolio construction, the rapid development of technology and artificial intelligence has ushered in a new age for this concept. Academics' interest in using machine learning techniques in financial applications has increased over the past decade. It offers numerous opportunities, such as predicting future stock movement, selecting the best investments, and engaging in algorithmic trading.

This paper offers a way to generate an optimal and efficient portfolio by selecting vital stocks from seven essential areas of the Canadian economy. After identifying the seven industries for the study, random selection is used to determine the four to five most important stocks for each sector. Using their ticker symbols, the APIs of the Python programming language retrieve the historical prices of these 30 equities listed on the Canadian stock market from the Yahoo Finance website. APIs of the Python programming language using their ticker names. By using the daily historical data of the past five years, an optimal portfolio has been designed from the combination of seven different industry stocks. In addition, an LSTM-based deep learning regression model is designed to predict future stock values effectively. A correlation matrix has been constructed to determine the relationship and interdependence between the stocks. Then, by computing the sharp Ratio and comparing it to an actual market index, this study proved effective in predicting stock prices and constructing an ideal portfolio using LSTM and Markowitz's Model.

The structure of the paper is as follows: The second section provides an overview of various related works, discussing both the individual model and their combined work. The third section provides the details of the data used and outlines the methodology followed. Section four presents the architecture and the top-level design of the LSTM model and the application of Markowiz's Algorithm. Section five presents the results of the portfolio and their actual and predicted returns and risks. Finally, the paper is concluded in section six.

## **1.2 Research Objective**

The primary objective of this study is to propose a framework for portfolio construction that incorporates stock price forecasts into calculations for the optimal portfolio. Then, the best-performing Canadian stocks from the various stock exchanges were chosen from seven different sectors. After that, a "long–short-term-memory" (LSTM) network to forecast the future returns of each stock was trained. Then, only the companies with positive price movements were included in Markowitz's portfolio computations approach.

## **1.3 Research Question and Hypothesis**

### Under this study, the researcher aims to determine if it is possible to predict the stock prices using the deep learning method (LSTM Model) and use that outcome to create an optimum portfolio following Markowiz's algorithm**.**

### **1.3.1 Research Sub-questions**

1. Can LSTM predict stock prices and Markowitz's model create an optimum portfolio?
2. Is LSTM and Markowitz's model for forecasting and building an optimum portfolio a good combination that can outperform the market benchmark?

### **1.3.2 Hypothesis**

**Hypothesis Development:** In this research, the study aims to find out if deep learning methods, such as LSTM Model, can predict the value of the stock currently to find out which stock value will rise. Later, with the predicted stocks, it aims to develop a portfolio using Markowiz's algorithm. This research is an outcome of those two combined models working together. The research hypothesis is as follows:

***Null hypothesis (H0):*** *LSTM model And Markowitz's algorithm can predict the stocks correctly and prepare an optimal portfolio.*

***Alternate Hypothesis (H1):*** *LSTM model And Markowitz's algorithm cannot predict the stocks correctly and prepare an optimal portfolio.*

# Chapter # 2: Literature Review

## **2.1 Financial Science**

Budgeting, saving, investing, lending, and forecasting are all included in the finance definition, which is money management. The three primary categories are personal, business, and public finance.

Personal finance is the study and management of one's financial affairs, including income generating, spending, saving, and investing. Corporate finance plans and implements management resources while balancing risk and profitability to optimise a company's worth. Public finances refer to the administration of a nation's taxation, spending, and debt by various governmental agencies. In addition, when referring to finance, "art" seems complementary to "science" because, aside from arithmetic and statistics, human emotions may also influence important financial decisions. (J. Sen, 2021)

Analysing the link between people and businesses is required to teach the fundamental financial principles. This money cycle depends on the market's ability to distribute products, concentrate, and redistribute the available capital resources to function. The stock markets and financial institutions make up the last one. The primary credit institutions are banks, which collect money from people and companies and lend it back with interest to people and businesses who need it. In addition, shares of publicly traded corporations can be exchanged on the stock market. (Onur Sen, 2021)

## **2.2 Stock Market**

Consumers can trade shares in securities markets, which are regulated locations. The focus is on setting up the prerequisites for fair commerce, which provides the following advantages. First, give the area and equipment required for transactions. Second, through planning and managing transactions, preventing speculative activities against investors. Third, to investors, provide information about transactions, such as daily quotations, transaction prices, and notifications of listed firms. Finally, the firms, people, stocks, and the venue where trades are conducted make up the stock market.

The Businesses A pool of capital initially owned by the co-founders—investors with rights and duties towards the firm—is necessary for the creation and growth of a company. Like its shareholders, this business is privately held. However, through the initial public offering, which facilitates trading on the stock market, the corporation is transformed into the public when investors elect to sell their interest or liquidate the funds they have invested. Or the point at which the corporation issues shares for the first time to the general public. The following are some of the factors that may influence a company's decision to go public. (Onur Sen, 2021)

First, if the business succeeds and shareholders have to liquidate their cash, selling their shares will result in a profit relative to the amount they initially invested. Secondly, if the business desires to expand, it will need additional funding from current investors. Third, the company wishes to raise money from the larger investing "public" to generate interest or establish itself. The primary market is located here.

Those Involved Participants include customers, realtors, brokerage firms, portfolio managers, administrators, and depot service providers. Investors are the people who purchase or sell stocks; they favour businesses with comparatively better prospects and give more emphasis to those that run more effectively. However, this preference makes it relatively simple for companies in the group mentioned above to get the cash they require by issuing additional shares; in other words, the stock market supports well-organised commercial endeavours and growth in thriving industries. (Onur Sen, 2021)

A company's development also influences the stock's current value since its owners are reluctant to sell when it is growing. After all, the possibilities of doing so at a greater price in the future are higher because everyone wants to invest in a successful business. Therefore, a high price results from strong demand and a short supply. The price of such shares is lower because supply is more than demand, in contrast to those who wish to sell share of a struggling firm because they are expected to sustain further losses over time. As a result, shareholders own, trade, and purchase the shares that firms have already issued. The name of this market is Secondary.

## **2.3 Available Financial Data**

Machine learning is built on data, but there have been multiple phases before the data can be fed into decision-making algorithms. Each stage is significant and covers a wide range of topics for examination, even if it goes beyond the scope of the postgraduate thesis. Nevertheless, certain crucial phases will be briefly discussed since they may contain the solutions to why the algorithm did not provide the expected results. The fundamental kind of data, known as raw data, is directly derived from selecting organs like a supermarket counter. The information is then stored in databases. (Perez, 2021)

Cleaning is crucial before usage, including eliminating duplicates, detecting missing data, and categorising information to make it simple to access, edit, and administer. Databases may be classified into various groups depending on how they are organised, including relational, graph, distributed, and cloud databases.

Regarding customers, relational databases are often organised in rows and columns to make them accessible through queries like SQL. A different type of database is a graph database, where nodes and edges define the relationships between the data. Specific semantic search terminology, such as SPARQL, is required for graph databases. Finally, in a distributed database, different parts of the database are physically kept in different places, and processing is spread or duplicated among various points in the network. (Onur Sen, 2021)

## **2.4 Machine Learning And LSTM Model**

Machine Learning (ML) is a form of Artificial Intelligence (A.I.) that uses statistical techniques to provide computer models with the ability to learn from a dataset and carry out specified tasks without the need for explicit programming. To uncover better investment opportunities, many investors use various algorithms, which has led to the development of machine learning applications for return forecasting, portfolio construction, and risk modelling. In recent years, the use of predictive models based on machine learning algorithms and deep neural network topologies for predicting stock prices has become increasingly widespread. (Jaydip Sen, 2021). The most important machine learning techniques in quantitative finance are MLPs, SVMs, Random Forest, KNN, RNN, and LSTM.

LSTM, on the other hand, is a variant of RNN that shares most of the RNN model's properties and has connected memory blocks (cells) in the hidden layer. Two states are passed from one cell to the next: the cell state and the hidden state. Memory blocks are responsible for remembering things, and three major gate mechanisms are used to manipulate this memory. Nonetheless, the key principle of the LSTM model is to implement a cell state link that will preserve the necessary memory information. Similarly, its internal entry controls structure changes and outputs control data. RNNs are potent artificial neural networks that can store input memory internally, making them excellent for solving sequential data issues, such as time series data. LSTMs have longer memories and can learn from inputs with lengthy time delays. (Weidman, 2019)

An LSTM consists of three gates: an input gate that selects whether or not to accept new input, a forget gate that deletes irrelevant information and an output gate that determines what information to output. These three analogue gates are based on the sigmoid function of zero(0) to one(1). These three sigmoid gates are depicted in Figure. A horizontal line that can be seen running through the cell represents the cells. (Anita Yadava, 2019)

A picture containing text, clock

Description automatically generated

Figure 1LSTM Architecture (Anita Yadava, 2019)

The LSTM model has some specific ways of calculating, and the equations are given below -

=

=

=

=

= =

Here, the σ is the logistic function, and 'tanh' is the tangent function of hyperbolic. The three gates mentioned above open and close according to 'ft' which is the gate controller. 'it' and 'ot' are also the controllers of the gates. They use the logistic function for activation, and the out range is [0 and 1]. Given LSTM cells, it is common to stack multiple layers of the cells to make the model deeper to capture the data's nonlinearity. An efficient prediction model is needed that can predict based on the previous data generated from the stock market. (Fischer & Krauss, 2018) The following diagram shows the working steps of LSTM for predicting the stock price trend with the help of historical data.

Diagram

Description automatically generated

Figure 2 LSTM Prediction Procedure

## **2.5 Markowitz's Portfolio Optimisation Model**

Markowitz's MV model serves as the basis for selecting an investment portfolio. Only this model can create various portfolios based on investor risk and return preferences, as different individuals have risk tolerance levels and have their own asset choice. This model assesses investment return and risk in proportion to expected return and variation. Rational investors want the lowest risk with a given projected return or the highest return with a shared risk to maximise expected utility. Using a multi-objective optimisation technique, the MV model attempts to compromise between maximising profits and minimising risks. MVO (Mean Variance Optimizer) requires estimates of the expected returns of all included assets, their standard deviations, and the correlation matrix to determine the optimal asset allocations and a set of efficient portfolios. (Samer Obeidat D. S., 2018)

Usually, rate of returns are expressed as a percentage expected by an investor, can be calculated from historical data and future investors' expectations. For example, the rate of return "rt" on investments between time (t) and (t − 1) can be calculated by a specific formula. However, the expected return rate can be calculated using a different equation. (G. Chacko, 2005). The equations are given below;

Rate of Return =

Expected Rate of Return =

Calculating Variance and the Standard Deviation is another crucial step in the portfolio construction procedure. Standard deviation is often used to measure the risk of their assets by the investors. The standard deviation is the most common statistical indicator of an asset's risk, which measures the dispersion around the expected value. The higher the standard deviation means greater risk and the opposite is also true. (L.J. Gitman, 1988) The formula of Variance and Standard Deviation is as follows –

Variance =

Standard deviation = =

The next step is calculating the Correlation and Covariance because correlations are used in advanced portfolio management, which measures the degree to which two variables move relative to their mean values over time. This step is computed as the correlation coefficient, which has a value that must fall between -1.0 and +1.0. A positive covariance says that the rates of return for multiple investments tend to move in the same direction during a particular time. Conversely, a negative covariance means that rates of returns for those assets move into opposite directions. (W.P. Chen, 2010) For calculating covariance, the following formula can be used -

Covariance =

After constructing the portfolio, the most crucial step is to measure the sharp Ratio, a measure of portfolio performance that provides the risk premium per unit of total risk, as assessed by the standard deviation of return for the portfolio. A portfolio's risk premium equals the entire portfolio return less the risk-free rate. In other words, Sharpe's measure divides the average excess return of a portfolio by its standard deviation during the same period. Sharpe's measurement can be represented using the formula below:

Sharp Ratio,

In risk-return space, every potential asset combination may be plotted, and the collection of all such portfolios constitutes an area in this space. The line along this region's upper edge is the efficient frontier. This efficient frontier is also known as the "Markowitz bullet," representing portfolio combinations with the lowest risk for a given level of return. In contrast, for a given level of risk, the portfolio on the efficient frontier represents the optimal return combination. The efficient frontier is the mathematical intersection of the set of portfolios with the lowest risk and the collection of portfolios with the highest return. (M. Ivanova L. D., 2017)

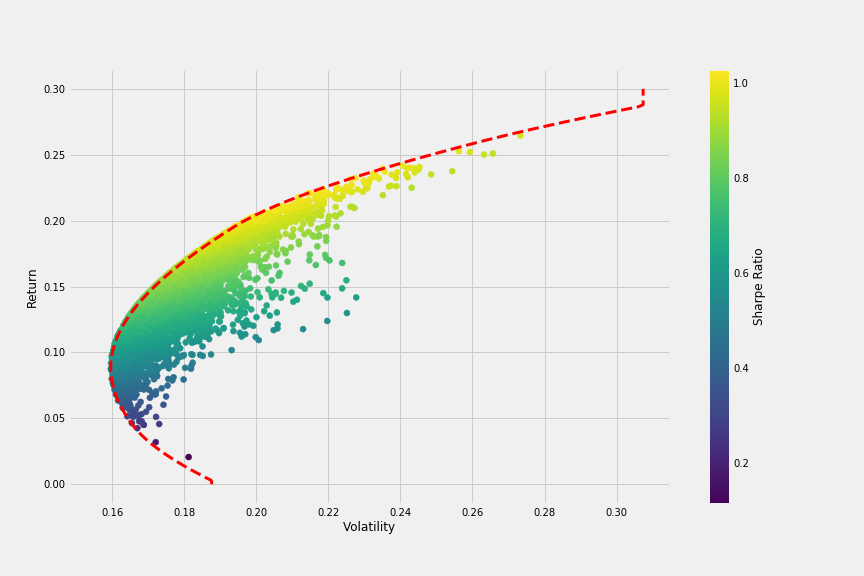


Figure 3 Efficient Frontier

The above Figure shows the entire investment opportunity set of all possible combinations of risk and return. Portfolios offer this set, and assets form it in differing proportions. (M. Ivanova L. D., 2017)

## **2.6 Stock Prediction using Deep Learning**

Machine learning methods are widely employed in stock market direction, price, or return prediction from historical data (Chong, 2017). In this research, Stock market prediction is the initial stage; therefore, this portion of the report summarises the related work in this field. Due to the complexity of the challenges and their impact on real-world applications, there are several models for stock price forecasting and robust portfolio-building to optimise returns and risk in a portfolio. In recent years, the use of predictive models based on machine learning algorithms and deep neural network topologies for predicting stock prices has become increasingly widespread. (Jaydip Sen, 2021) ANN, KNN, and SVM, all these approaches can help investors in making investment decisions with the help of their advanced calculations. In addition, Some scholars have also recommended the use of multi-objective optimisation and Eigen portfolios using principal component analysis in portfolio creation.

However, a recurrent neural network (RNN) has been designed to work with sequence prediction. Long Short-Term Memory (LSTM) is an RNN architecture that addresses the vanishing gradient problem. The key to the LSTM solution to the technical difficulties was the specific internal structure of the units used in the model. Today's markets are immensely complex and chaotic, which contains a dynamic environment making forecasting the stock market difficult. However, the constructed portfolios based on the LSTM prediction model outperformed other constructed portfolios-based prediction models. Many researchers have applied LSTM for predicting stock prices and market movement and found the LSTM model is the most accurate and Robust. Therefore, this report will entirely focus on LSTM, a special kind of RNN for forecasting stock prices with the help of historical data, as it is proven to be the most accurate and robust by many researchers.

One researcher, Murtaza, used the LSTM model to predict the stock market indices and proved it a better forecasting method for investment decisions. (Murtaza Roondiwala, 2015) In his research, he used historical data from NIFTY 50 and got RMSE results for training RMSE of 0.00983 and testing RMSE of 0.00859.

Another researcher, Nelson and his research group (Nelson, Pereira, & Oliveira, 2017), also examined the application of LSTM networks to forecast future stock price patterns based on price history and technical analysis data. The researcher developed a prediction model and conducted a series of tests to achieve this objective. First, they examined against several criteria to determine if this algorithm offers any advantages over existing Machine Learning approaches and investment strategies. The results gained are encouraging, with an average accuracy of 55.9% when forecasting whether the price of a particular stock will rise or fall in the future.

On the other hand, Achyut Ghosh and his fellow researchers considered physical or physiological factors, rational or irrational behaviour, investor sentiment, market rumours, etc., influencing while predicting stock prices (Achyut Ghosh, 2019). Then, they applied the machine learning method in the Indian stock market to predict future prices by using the historical stock price as the sum of all contributing elements. First, they infer the future trend using Machine Learning (ML) techniques to uncover patterns and insights they overlooked before for precise forecasting. Then, using the LSTM model and the businesses' net growth calculation technique, they predicted a company's future growth.

Some researchers compared methods and their accuracy rates in forecasting stock prices. In addition, some researchers applied multiple models on the same data and environment to test the models' validity and authenticity. Finally, they waited to get the actual result and see whether this Machine learning technique works in the real market. For example, Pushpendu Ghosha (Pushpendu Ghosha, 2020) compared the accuracy rate of Random Forest and LSTM; the multi-feature configuration yields a daily return of 0.64% with LSTM networks and 0.54% with random forests. This research also outperforms the single-feature structure consisting of just daily closing price returns, with 0.41% and 0.39% daily returns for LSTM and random forests, respectively.

Another researcher, Jaydip Sen (Jaydip Sen S. M., 2021), anticipated and actual returns demonstrate that the LSTM model is highly accurate. He proposes an algorithmic method for creating optimal risk and Eigen portfolios for five thematic sectors of India's National Stock Exchange (NSE). The reason behind this research is optimising portfolios has been a topic of intense attention among quantitative and statistical finance academics and financial analysts.

On the other hand, Qihang Ma (Ma, 2020) compared ARIMA, LSTM, and ANN methods for forecasting stock prices and believed that the LSTM model might have the best predictive ability. According to him, the ANN and ARIMA model have their advantages and disadvantages, but the LSTM model has the best predictive ability. Still, it is greatly affected by data processing. However, he believes the ANN method worked better than the ARIMA model, and LSTM is on top of them.

This study found another research about the comparison between LSTM and another model, where Krishna (Sai Krishna Lakshminarayanan, 2019) separately implemented SVM and LSTM using moving averages and concluded that LSTM gives far more accurate results while using the models separately for forecasting. However, they also used a hybrid method for predicting and analysing the data on stock purchases and sales to identify how fluctuations in stock price and other factors have affected the purchasing and selling pattern, contributing to the development of more precise results. In addition, the models can predict other obstacles, such as the weather, diseases, housing prices, etc.

Lastly, the study found another research work on LSTM, done by Jaydip Sen, where he examined the historical price series of the top five stocks in each of the Indian stock market's nine sectors. (Jaydip Sen A. D., 2021). That research constructed optimal portfolios for each of these businesses. A long-and-short-term memory (LSTM) model is also built and optimised for predicting future stock prices. After five months of portfolio construction, each portfolio's actual and predicted returns and risks are computed. As a result, each portfolio's forecast and substantial returns are high, supporting the LSTM model's high precision.

## **2.7 Portfolio Construction Model**

The building of a portfolio is the second stage of this research. This study examined numerous literature on this topic to determine which portfolio construction model is superior. According to Jaydip (J. Sen, 2021) and Erwin (K. Erwin, 2020), several propositions exist in different kinds of literature for portfolio optimisation using strategies such as fuzzy logic, swarm intelligence, genetic algorithms, and principal component analysis. During that period, the study found Several studies approaching different methods for building a profitable portfolio. To illustrate, Heaton's work (J. B. Heaton, 2016) used auto-encoders. It proposed deep portfolio theory, which presents a framework for creating portfolios that outperform a benchmark approach by a specific amount. Jaydip Sen, generalised autoregressive conditional heteroscedasticity (GARCH) is widely used in estimating portfolio risks. (J. Sen S. M., 2021). Lee and Zhang used the LSTM model and evaluated their portfolios. Lee, (Lee, 2018) examined the impact of various threshold values on the risk of portfolios formed using LSTM predictions, and Zhang (Zhang R. H., 2018) employed LSTM stock return forecasts during the monthly portfolio weight computation with a predefined threshold value. Nevertheless, many researchers have also proposed multi-objective optimisation approaches for robust portfolio design using metaheuristics and constraint-imposed heuristics. (P. Zhao, 2020) and (M. Corazza, 2021).

On the other hand, some authors present constraint-based strategies to overcome those shortcomings of Markowitz's model. For example, Zhang and his research mates (Zhang, Leung, & Aravkin, 2019) used the model and tried to overcome the limitation and build a profitable portfolio. Finally, even though Markowitz's original mean-variance optimisation approach has several practical constraints, including purchase limit and cardinality, Samer (Samer Obeidat, 2018) used Markowitz's model to construct a and proved that the model could generate a better portfolio.

Markowitz's methodology may design portfolios based on investors' risk tolerance and asset preferences. Given the priorities, the portfolio consists of various securities, not just one or two stocks. In addition, Markowitz asserts that the quality of a portfolio is contingent upon its constituent assets. Therefore, the risks associated with two investments are distinct when viewed separately and jointly. All investors must limit risks and maximise profits, as investors worldwide extensively employ the Markowitz model. After considering all the matters, this study used Markowitz's Model to construct an optimum portfolio and reviewed more research works on this model.

Ivanova worked between 2013 and 2016 (M. Ivanova, 2017), analysing Bulgarian stock market performance where he used Markowitz's model. Based on the weekly closing prices of 50 companies listed on the Bulgarian Stock Exchange from January 2013 to December 2016, the theory produced efficient borders and optimal portfolios. Then, Bulgarian investors can choose an appropriate portfolio based on risk tolerance. During the study period, the report demonstrated that the efficient portfolios created by the Markowitz model outperformed any individual domestic security on the market.

Busetti, (Busetti, 2006), found that Markowitz's method can construct better portfolios. This researcher constructed a realistic model, analysed the effectiveness of its solution using two heuristic methods, genetic algorithms and tabu search, and then examined the insights supplied by optimising actual portfolios. This study applies a model based on the classical mean-variance approach, along with floor and ceiling restrictions, cardinality constraints, and nonlinear transaction costs. Altogether they include a significant illiquidity premium to an extensive portfolio of 100 stocks. According to the researcher, the performance of genetic algorithms is superior to that of tabu search for large portfolios.

Another researcher applied this model to hedge funds. Ejara (Ejara, 2016) compared the performance of long-short equity hedge funds to the market index using mean-variance criteria, including more significant moments. Most long-short equity hedge funds outperform the market index according to the mean-variance criterion. In contrast, when higher moments are utilised to evaluate performance, more hedge funds underperform the market index, indicating the necessity of incorporating higher moments into portfolio optimisation.

Establishing a portfolio in Ghana was challenging, but Logubayom, (Anuwoje Ida Logubayom, 2019), applied the model and successfully established an efficient portfolio. To the investor's relief, this study demonstrated how to utilise the Markowitz model on the Ghana Stock Exchange and revealed the most efficient portfolio among selected stocks. The study indicated that GCB Bank limited had the highest returns (4.2%) and the lowest risk (13.1%), followed by CAL (3.5%) and  (11.7%). At the same time, UGL presented the lowest risk and returns (6.8%) and (2.1%). Risk-seeking investors like UGL, whereas risk-seeking investors favour GCB and CAL. GCB and CAL are the most productive investment portfolios. Risk-tolerant investors can invest their entire portfolio in GCB, while risk-averse investors can invest 39.21% in GCB and 60.79% in CAL. CAL and GCB bank limited anticipate 3.9% returns with 10.6% risk, followed by TOTAL and GCB with 3.4% and 12.3% risk, respectively. The portfolio's high risk precluded the Sharpe ratio from reflecting the high expected return of TOTAL and GCB. The high Sharpe ratio of CAL and GCB shows that this portfolio is the most efficient.

## **2.8 Performance of Combined Model**

Over time, investing in the stock market has grown in popularity and complexity, making the current market more competitive. Nonetheless, machine learning has simplified this work for investors. Although, as mentioned above, this study reviewed other research papers and models and uncovered several studies on ML approaches for stock price forecasting for investment decisions and building portfolios of their own choices. It found that people are utilising ANN, KNN, SVM, RNN, etc., for stock price forecasting, and many can accurately anticipate the stock price trend. During this search, numerous publications about the design of portfolios and their influence on investors were located. Investing in portfolios can help investors lower their risk, and there are now many indices on the market. There are various methods for generating portfolios, including fuzzy logic, swarm intelligence, genetic algorithms, GMV model, MV model, etc., which some investors prefer.

Moreover, some researcher has used both techniques in one research to construct the most efficient portfolio. For example, Irma, (Irma Fitria, 2016), used ARIMA-Kalman to forecast the stock prices and MPC (Model Predictive Control) to build an optimum portfolio. Perez, (Perez, 2021), used Markowitz's MV model and Random Forest forecasting in his research. Whereas Van-Dai, (Van-Dai Ta, 2019), used the LSTM model for predicting prices and multiple models for portfolio construction, including Monte Carlo Simulation,  equal-weighted modelling (E.Q.), and mean variant optimisation (MVO). Etc. After reviewing all the relevant literature, this study adopted Markowitiz's MV model for portfolio development. Before that, this study chose specific equities by predicting their future trend using LSTM, which numerous researchers have previously employed. This study has reviewed their research papers to get better ideas for completing this work; some of them are discussed in the following paragraphs.

Hasan Sami and his fellow researchers, (Hasan M Sami, 2021), studied the Bangladeshi stock market. In that study, they used machine learning techniques to forecast the stock prices before constructing an optimal portfolio for a risk-averse investor. According to Portfolio Optimization principles, the study analysed two alternative portfolios based on LSTM's future forecasting capabilities for projecting the maximum feasible output. Technical identifiers and financial benchmarks characterise global financial decision-making. Technical indicators such as MACD, RSI, and financial ratios (EPS and P.E.) helped identify DSE assets. Due to its Gated Structure, LSTM has achieved significant success in human behaviour-based nonlinear forecasting. The Gated Structure retains important information and eliminates unneeded information by decreasing and extending the gradient. At the end of the research, they found that these two methods can work perfectly for a risk-averse investor.

Samer (Samer Obeidat, 2018) proposed a framework for determining the asset weights in a diverse portfolio consisting of different types of assets. At first, the study predicted the prices by performing principle components analysis (PCA) dimensionality reduction on the features using an LSTM network. After that, it applied a mean-variance portfolio optimisation. Next, the model estimated the selected assets' expected return, volatility, and correlation. These neural network outputs were then turned into action recommendations using a Mean-Variance Optimisation framework augmented using a forward-looking rolling window technique. Finally, testing was performed on a dataset with a 7.66-year duration. This work concludes that a Long Short-Term Memory model can generate better risk-adjusted returns than conventional strategic passive portfolio management.

However, Zeynep, (Zeynep Cipiloglu Yildiz, 2020), proposed a novel paradigm to integrate future return forecasts into portfolio creation processes. This paper used BIST30 Turkey index stocks for this research. The solutions generate a mean annualised return of 25 percent, about 50 percent higher than the returns caused by the baseline approaches. At first, a LSTM model is initially trained to learn the monthly closing prices of the stocks. After that, five portfolio construction methodologies were introduced. Moreover, the average Sharpe ratio of the techniques is 0.57, which is higher than the sharp ratios (0.29 and 0.32) for the methodologies compared with this study. A comprehensive analysis of the results reveals that including expected returns into portfolio construction significantly improves portfolio performance.

Another article used the LSTM model for forecasting and created an optimum portfolio using Markowitz's method for a risk-averse investor. In this research, Jaydip Sen, (Jaydip Sen A. D., 2021), analyses the historical prices of five well known stocks from July 1, 2017, to June 30, 2022, and then built Optimal portfolios for each of the nine sectors. Both the predicted and actual returns of each portfolio are high, indicating the high precision of the LSTM model.

On the other hand, Yilin, (YILIN MA, 2020), showed how DNNs can develop prediction-based portfolio models. In her research, she employed three deep neural networks (DNNs) to develop prediction-based portfolio optimisation models that combine deep learning and current portfolio theory. First, these models forecast stock returns using DNNs, and then predictive mistakes are used to quantify stock risk. Afterward, portfolio optimisation models were created, and these models were evaluated with three equal-weighted portfolios selected by DMLP, LSTM, and CNN. This article used China Securities 100 index as experimental data. This report concluded that the prediction-based portfolio model based on DMLP performs best among all models under varied intended portfolio returns. Furthermore, a high desired portfolio return can further improve this model's performance.

# 

# Chapter # 3 Research Design

## **3.1 Research Philosophy**

The research project structure and the method are that, it will try to answer the research questions, whether they are crucial to the research philosophy or not. The researcher frequently embraces positivist, interpretivist, and pragmatic research philosophies; the third is optional. One of them is positivism, which has its roots in the natural sciences and focuses on the scientific testing of hypotheses and the identification of logical or mathematical evidence that may be derived from data analysis (Collis, 2014). Because of this, positivists seek to use large sample sizes and offer precise, impartial, and quantitative data. (Collis, 2014) This study planned to use a pragmatic approach to the inquiry because of the nature and construction.

## **3.2 Research Design**

This study chooses its methodology at the end of the research design phase. Both qualitative and quantitative approaches are employed, depending on the nature of the study issue and the desired result. The primary facts used in quantitative strategy to produce research outcomes are figures and facts. In qualitative research, non-numerical data is gathered and analysed better to understand concepts, viewpoints, or individual experiences.

Research is essential to learn more about a topic, develop fresh research ideas, and understand it better. But for the sake of the study, the researcher used secondary data sources to compile the study data. Consequently, it adopted a quantitative strategy for this study, pertinent to the research project's methodology and philosophy.

## **3.3 Research Approach**

### **Step # 1: Use of LSTM to determine upraising stocks**

Python programming language was used in this study to forecast the trend of stock prices and to generate stock portfolios. Using the rich libraries of the **'sklearn'** and **'Keras'** frameworks, the LSTM regression model was constructed. The model separated the data into training and test sets. This study projected the future trend of the stocks and then computed the RMSE of the train and test scores to determine the model's accuracy.

### **Step # 2**: **Creating an Optimal Portfolio using Marowiz's Algorithm**

This study then generated an optimal portfolio with the chosen stocks. Before constructing the portfolio, this study analysed the stocks' risks, returns, correlation, and covariance. This study utilised Markowitz's Algorithm and operated one million times to get the optimal portfolio composition. Finally, the procedure is completed by calculating this portfolio's sharp Ratio for evaluating the portfolio.

Figure 4 Flowchart of Research Approach

Diagram

Description automatically generated

The above diagram is to visualise the whole procedure of this study. At first, the seven industries were chosen, and after that, 30 stocks were considered for building the portfolio, where the LSTM algorithm was used for the stock selection procedure. This study used the authentic Yahoo Finance website to collect historical data. This study found that specific equities lacked five years of data, so it removed them and began searching for other stocks in the same industry. After that, the LSTM model was applied to each stock using jupyter notebook, and the price trend of each stock was determined. As a result, only 19 of 30 equities with an upward tendency were included in the portfolio construction exercise.

Then, utilising Markowitz's Model, this study began constructing its portfolio with 19 stocks. After importing the \*.csv file with the adjusted closing prices of 19 stocks into the Jupyter notebook, this study estimated the return of stocks and, subsequently, the portfolio return of 19 stocks. The portfolio's variance and standard deviation were then calculated. With the aid of the risk and return diagram, the researcher determined that if a person accepts 3.2% of the risk, they may anticipate an approximate 11% return.

Then the correlation and covariance were determined in this study. Here, this study computed the correlation between the stocks to decide whether they are favourably or negatively associated, as investing in companies with a negative correlation will reduce losses in the event of a price decline. (By Tze Leung Lai, 2011) This report then generated a heatmap to visualise the stocks correlations and estimated the covariance by multiplying the correlation between two random stocks by the standard deviation of each stock. In contrast, covariance assesses the degree to which two variables deviate from their respective means across time. A positive covariance indicates that the returns on two assets tend to move in the same direction over time. Negative covariance signifies that the returns on two different investments move in opposite directions. (By Tze Leung Lai, 2011)

Finally, this study created a portfolio by weighting the 19 stocks equally and calculating risk and return by multiplying them by the weight (1/19 = 0.053).

While doing the portfolio construction procedure, this study executed a series of loops to generate slots for the risk, return, and weights. First, this study utilises the Markowitz model and asks the machine learning algorithm to generate one million portfolios using the 19 stocks provided. Here, the asset length corresponds to the number of rows, and there is currently no weight parity among the funds. In this phase, this study provided Machine Learning with blank areas to experiment with various combinations of funds. This loop was executed 100,000 times, with each iteration considering different asset weights and calculating the return and risk of each portfolio combination. After that, this study used the **"np.random.random()"** method to generate random numbers for consequences and then split the results by their cumulative sum because the sum of importance must equal 1. (By Tze Leung Lai, 2011). Then it designed a simple portfolio. Next, the study computed the optimal risky portfolio using a risk-free rate of 0.05. This investment portfolio boasts the greatest Sharpe Ratio. The balance is the difference between the average return earned and the risk-free rate per unit of volatility or absolute risk. Volatility is a measure of an asset's or portfolio's price changes. The risk-free rate of return is the return on investment with no risk, i.e., the rate of return that investors can anticipate if they take no risks. This one was visualised in a diagram with a green star by Machine Learning.

Finally, after repeatedly conducting the procedure, this study found an optimal portfolio with a particular risk and return. This procedure ended by evaluating the portfolio with the help of the sharpe ratio.

# Chapter # 4 Research Methodology

## **4.1 Sampling and Data Collection**

### **4.1.1 Industry and Stock Selection**

This study has used historical prices of the stocks existing in the Canadian Stock Exchange Market from 01/07/2017 to 30/06/2022, as collected from Yahoo Finance (https://ca.finance.yahoo.com). Firstly, it selected 30 stocks from 7 different industries (the telecommunication sector, health care sector, financial sector, Gaming sector, Real Estate sector, Energy sector, and Technology sector).

### **4.1.2 Data Acquisition**

The historical stock values for each industry were collected from the Yahoo Finance website. This report preferred to download the data and convert them into a CSV file instated of using the DataReader method available in the data sub-module in the **pandas\_datareader** module of the Python language and importing the CSV in the notebook. The period for which the stock prices are used for portfolio design spans over five years starting from July 1, 2017, and ending on June 30, 2022.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sl.** | **Industry** | **# of Companies** | **# of Data Point** |
| 1 | Telecommunication | 5 | 6,300 |
| 2 | Health Care | 4 | 5,040 |
| 3 | Financial | 4 | 5,040 |
| 4 | Gaming | 4 | 5,040 |
| 5 | Real estate | 3 | 3,780 |
| 6 | Energy | 5 | 6,300 |
| 7 | I.T. Sector | 5 | 6,300 |
| **Total** | | **30** | **37,800** |

***Table # 1: Summary of Acquired Data***

### **4.1.3 Application of the LSTM Model**

After the initial sampling selection, I will clean up the data and prepare it for LSTM Model Application. Again, I will use Python for this step. The primary objective is to check the accuracy of the LSTM Model and to find out which price is in an upwards trend.

## **4.2 Study Variables**

Web-extracted raw data have six attributes: open, high, low, close, volume, and adjusted close. Since the current work is based on univariate analysis, only the variable near the specified period was preserved for portfolio design and further study; the other factors were disregarded.

## **4.3 Ethical Considerations**

When using secondary data, it is commonly considered that the researcher is relieved of the responsibility of gaining ethical authorisation and, in some situations, is relieved of the obligation to consider ethics. However, whether or not primary data collection is involved, ethical considerations must be felt throughout the entire research process. The procedure begins with the original design of the research, which should be geared toward the public good. It continues by disseminating the study's findings, which should ensure transparency, publicness, and reproducibility of results. Therefore, this study has taken the following ethical considerations for the study:

* Data were de-identified before public release or review, and the workings were kept out of anyone's reach and encrypted.
* The analysis outcomes would not harm the goodwill of the sampled businesses.
* The result would not harm or insult any organisation or would not be the reason for any damage to their goodwill.

Also following safeguards were taken for confidentiality, security, and safeguard:

* The data was collected by using a secured and private internet connection.
* This study has collected all the data from verified stock exchange websites (ca.finance.yahoo.com) and later confirmed through a second source, e.g., Bloomberg.
* It maintained the exported data with password-protected M.S. Excel Files.
* This study also took regular backups of the data and works.

# Chapter # 5 Data Analysis and Calculations

## **5.1 Stock Selection for Portfolio**

### **5.1.1 Deep Learning (LSTM) Modeling**

The data was exported into a \*.csv file from Yahoo Finance. Using Jupyter Notebook, this data was later fed into LSTM Script for train and testing. Firstly, the researcher has performed some data cleanup for calculation. After the learning, the outcome was plotted into chat to show the trend. Later, the stocks with upward trends would be used for portfolio optimisation. Table # 2 shows the result of this step.

# create and fit the LSTM network

model = Sequential()

model.add(LSTM(4, input\_shape=(1, look\_back)))

model.add(Dense(1))

model.compile(loss='mean\_squared\_error', optimizer='adam')

model.fit(trainX, trainY, epochs=100, batch\_size=1, verbose=2)

# make predictions

trainPredict = model.predict(trainX)

testPredict = model.predict(testX)

# invert predictions

trainPredict = scaler.inverse\_transform(trainPredict)

trainY = scaler.inverse\_transform([trainY])

testPredict = scaler.inverse\_transform(testPredict)

testY = scaler.inverse\_transform([testY])

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sl.** | **Company\_Index** | **Score\_Train** | **Score\_Test** | **Acc\_Train** | **Acc\_Test** | **LSTM** |
| 1 | APA.TO | 0.98 | 1.09 | 99.02% | 98.91% | Upwards |
| 2 | AND.TO | 1.25 | 1.34 | 98.75% | 98.66% | Upwards |
| 3 | AT.TO | 0.09 | 1.77 | 99.91% | 98.23% | Downwards |
| 4 | BHC.TO | 0.79 | 0.95 | 99.21% | 99.05% | Downwards |
| 5 | BMO.TO | 1.08 | 2.65 | 98.95% | 96.34% | Upwards |
| 6 | BTE.TO | 0.11 | 0.21 | 99.89% | 99.79% | Upwards |
| 7 | CHP-UN.TO | 0.17 | 0.16 | 99.83% | 99.84% | Upwards |
| 8 | CM.TO | 0.61 | 1.58 | 99.39% | 99.82% | Upwards |
| 9 | CURLF | 0.45 | 0.46 | 99.55% | 99.54% | Downwards |
| 10 | CWB-PB.TO | 0.22 | 0.23 | 99.78% | 99.77% | Upwards |
| 11 | FANG | 2.58 | 3.35 | 97.42% | 96.65% | Upwards |
| 12 | GH.TO | 0.17 | 0.17 | 99.83% | 99.83% | Upwards |
| 13 | GTII.CN | 1.2 | 3.4 | 98.80% | 96.60% | Downwards |
| 14 | LWRK.TO | 0.38 | 0.77 | 99.62% | 99.63% | Upwards |
| 15 | MAGT.TO | 1.71 | 1.23 | 98.29% | 98.77% | Downwards |
| 16 | MPC | 1.46 | 1.65 | 98.54% | 98.35% | Upwards |
| 17 | NVA.TO | 0.11 | 0.21 | 99.89% | 99.79% | Upwards |
| 18 | PBL.TO | 0.54 | 4.81 | 99.48% | 95.36% | Downwards |
| 19 | REI-UN.TO | 0.32 | 0.36 | 99.68% | 99.64% | Upwards |
| 20 | SFTC.TO | 1.1 | 0.83 | 98.90% | 99.87% | Downwards |
| 21 | SRU-UN.TO | 0.47 | 0.37 | 99.53% | 99.63% | Upwards |
| 22 | TCS.TO | 0.65 | 1.87 | 99.35% | 98.13% | Upwards |
| 23 | TOI.V | 3.25 | 3.03 | 96.75% | 96.97% | Upwards |
| 24 | T.TO | 1.81 | 1.22 | 98.90% | 98.80% | Downwards |
| 25 | RCI-B.TO | 0.65 | 1.88 | 99.10% | 98.90% | Upwards |
| 26 | ENB.TO | 3.25 | 3.04 | 97.01% | 97.80% | Upwards |
| 27 | RY.TO | 2.57 | 3.2 | 97.10% | 96.70% | Downwards |
| 28 | MI-UN.TO | 0.32 | 0.36 | 99.68% | 99.64% | Upwards |
| 29 | WELL.TO | 1.1 | 0.83 | 98.90% | 99.87% | Downwards |
| 30 | SHOP.TO | 0.65 | 1.87 | 99.35% | 98.13% | Downwards |

***Table # 2: LSTM Outcome of Final Sample Data***

**LSTM Modeling Outcome:** Out of the initial 30 selections of stock, 19 stock prices went up, and 11 went down. Stocks with only upwards trends would be considered for the next step of the calculation.

### **5.1.2 RMSE Calculation**

This study has also performed a Root Mean Squared Error (RMSE) calculation as it is a method that can calculate the error or accuracy in the prediction. RMSE is highly prevalent and is a great error metric for general numerical predictions. If a comparison has been made to the Mean Absolute Error, the RMSE magnifies and severely penalises significant errors. In addition, the error or the difference between the target and the output value is minimised using the RMSE value. The root square of the mean/average square of all errors is the RMSE. (Murtaza Roondiwala, 2015) This study's outcome was satisfactory, with 98% confidence (Average, Train: 98.7% Test 98.9%). Moreover, Along with this calculation, the study also used historical data and observed whether the predicted and real trends match each other. (WAR AHMED, 2018)

The Formula for RMSE is –

RMSE =

# calculate root mean squared error

trainScore = math.sqrt(mean\_squared\_error(trainY[0], trainPredict[:,0]))

print('Train Score: %.2f RMSE' % (trainScore))

testScore = math.sqrt(mean\_squared\_error(testY[0], testPredict[:,0]))

print('Test Score: %.2f RMSE' % (testScore))

RMSE = \*\*.\*\*

accuracy = 100 - np.mean(RMSE)

print('Accuracy:', round(accuracy, 2), '%.')

RMSE = \*\*.\*\*

accuracy = 100 - np.mean(RMSE)

print('Accuracy:', round(accuracy, 2), '%.')

## **5.2 Constructing Optimal Portfolio with Markowiz's Algorithm**

1. **Deriving Portfolio Return and Volatility:** Every day, each stock's return and log return values are calculated based on its historical closing values. The daily return and the log return for a stock, calculated in percentage values, represent the change in the closing values over successive days and their respective logarithms. The daily return and log returns are calculated using Python's pct change library function. The values of each stock's daily and annual volatility are then calculated. The standard deviation of a stock's daily returns determines its daily fluctuations. The volatility numbers are computed using the Python function std.

# Log of percentage change of all stocks in the list

Ret= df.pct\_change().apply(lambda x: np.log(1+x))

Ret.head()

Ret.plot(title='Stocks Daily returns',figsize=(15,10))

The output of Covariance and Correlation Matrices

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Date** | **PBL.TO** | **BMO.TO** | **TSC.TO** | **MPC** | **BTE.TO** | **LWRK.TO** |
| 7/4/2017 | NaN | NaN | NaN | NaN | NaN | NaN |
| 7/5/2017 | 0.055444 | -0.00063 | -0.03806 | -0.00377 | -0.07632 | 0.004437 |
| 7/6/2017 | 0.00353 | 0 | 0.012848 | -0.00095 | 0.037229 | -0.00444 |
| 7/7/2017 | 0.012259 | 0.001148 | 0.027284 | 0.022822 | -0.03379 | 0.012155 |
| 7/10/2017 | -0.00786 | 0.002397 | 0.031253 | 0.000185 | 0.013652 | -0.0033 |

***Table # 3: Daily return and log returns of the stock***

1. **Covariance and Correlation Matrices of Stock Returns:** The covariance and correlation matrices of the stocks are constructed based on their return values in the training dataset after each stock's return, and volatility values have been obtained. These matrices provide crucial information for the creation of a portfolio by assisting in comprehending the patterns of linkage among the stocks in a certain sector. Python routines called 'cov' and 'corr' are used to compute the covariance and correlation matrices. One of the main optimisation goals for portfolio design work is the minimisation and optimisation of risk. The algorithm tries to distribute the money among stocks with little to no association with one another in a diversified portfolio that reduces risk. It is feasible to identify these stocks by examining their covariance or correlation matrices.

# Covariance and Correlation Matrix Calculation

Ret\_Corr = Ret.corr()

print(Ret\_Corr)

import matplotlib.pyplot as plt

import seaborn as sns

sns.heatmap(Ret\_Corr)

Ret\_Covar = Ret.cov()

print(Ret\_Covar)

The output of Covariance and Correlation Matrices

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **PBL.TO** | **BMO.TO** | **TSC.TO** | **MPC** | **BTE.TO** | **LWRK.TO** |
| **PBL.TO** | 1 | 0.24799 | 0.148425 | 0.06926 | 0.155093 | 0.158996 |
| **BMO.TO** | 0.247997 | 1 | 0.193 | 0.058845 | 0.462104 | 0.231469 |
| **TSC.TO** | 0.148425 | 0.193 | 1 | -0.010873 | 0.149386 | 0.071545 |
| **MPC** | 0.148425 | 0.058845 | -0.010873 | 1 | 0.061215 | 0.022047 |
| **BTE.TO** | 0.155093 | 0.462104 | 0.149386 | 0.061215 | 1 | 0.082253 |
| **LWRK.TO** | 0.155093 | 0.231469 | 0.071545 | 0.022047 | 0.082253 | 1 |

***Table # 4: Covariance & Correlation Matrix***

1. **Estimation of Portfolio Return and Risk:** The initial set of portfolios is built at this stage using each sector's covariance and correlation matrices. The portfolios are first created by giving each of the five stocks in a particular industry the same weight. For example, a sector's stocks are each given a weight of 0.2 to make their aggregate equal to 1. Next, each of the nine sectors' annual return and volatility numbers is calculated for the equal-weight portfolio. Finally, the expected return of a portfolio made up of n capital assets (i.e., stocks) indicated as C1, C2,...., Cn, with the weight wi allocated to the i-th asset is computed using (1), which is based on the historical return values of the portfolio.

() = +)+ ) …………(1)

The annual returns and volatility of the equal-weighted portfolio for each sector are calculated using the stock's annual return and volatility numbers. For calculating the mean yearly returns for each component stock in a portfolio, the Python function resample uses the argument 'Y'. By dividing the daily volatility values by a factor equal to the square root of 252, on the other hand, the annual volatility values for the equal-weight portfolios are obtained. The equal-weight portfolios offer a starting point for return and risk associated with the sectors over the training period. They may be used as a benchmark for assessing other portfolios' performance. However, the return and risk projections using equal-weighted portfolios are incredibly inaccurate predictors of future returns and dangers. Therefore, more accurate projections of the potential return and hazards are required. Therefore, creating portfolios with the lowest and highest levels of risk is necessary.

1. **Designing Optimum Portfolio:** Due to low returns, stock market investors seldom adopt the risk-reduction technique recommended by the minimum-risk portfolio. (Onur Sen, 2021) Most typically, when larger risks are linked with better rewards, investors are willing to take them. A statistic known as the Sharpe Ratio is used to build an optimal-risk portfolio by maximising the risk and return in a portfolio (2). The Sharpe Ratio of a portfolio is a statistic used to determine the ideal risk portfolio. A portfolio's Sharpe Ratio is determined by (2).

… (2)

Equation # 2 denotes the current portfolio return, the risk-free portfolio return, and the current portfolio risk, respectively, as Ret(curr), Ret (risk-free), and Ris (curr); its yearly standard deviation measures it. It is assumed that the portfolio with a risk value of 1% is risk-free. The optimal-risk portfolio maximises the Sharpe Ratio's value. If the optimum-risk portfolio is compared to the minimum-risk portfolio, the optimum-risk portfolio achieves a very high return with a negligibly increased level of risk. The Python 'idmax'function determines which candidate portfolio on the **E.F.** has the greatest Sharpe Ratio.

1. **Comparison of study result with the actual result:**

Portfolio performance evaluation is taken to test the notion of market efficiency. The evaluation process has been conducted to get three benefits: increase efficiency, monitor risk, and analyse returns. Several methods evaluate portfolio efficiency: conventional, risk-adjustment, etc. Here, this study used the conventional method and compared the Sharp Ratios of the portfolios to see whether the constructed portfolio could beat the benchmark or not.

|  |  |  |  |
| --- | --- | --- | --- |
| Sharpe Ratio | Existing Portfolio (TSX) | Study Portfolio | Variance |
| Ratio | 0.57 | 1.66 | 1.09 |

***Table # 5: Comparison of Sharpe Ratios***

The Sharp Ratio of the study portfolio is higher than the market Index, and this study established that an investment in this constructed portfolio would be profitable.

# 

# Chapter # 6 Conclusion & Discussion

## **6.1 Study Findings and Result**

In this section, the performance results of the constructed optimum portfolio have been presented. At the same time, the study commenced with 30 stocks from Canadian Market. The LSTM Deep Learning architecture showed the upraising stocks with 98% (average) accuracy for 19 stocks. Later these data were analysed with Markowiz's Algorithm, and an optimum portfolio was constructed after 100,000 times of computation with 6 stocks combination. Heatmap created by Markowiz's Algorithm is as follows:

A picture containing checker

Description automatically generated

Figure 5 Heatmap of the six stocks

Chart, scatter chart

Description automatically generated

Figure 6 Minimal portfolio

Chart, scatter chart

Description automatically generated

Figure 7 Optimal Portfolio

The final portfolio constructed using the algorithm is as follows:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sl.** | **Industry** | **Company Name** | **Weight** | **% of investment** | **Risk (Avg.)** | **Return (Avg.)** |
| 1 | Gaming | PBL.TO | 0.200564 | 20.06 | 13% | 23% |
| 2 | Financial | BMO.TO | 0.00985 | 0.99 |
| 3 | Technology | TSC.TO | 0.352957 | 35.3 |
| 4 | Energy | BTE.TO | 0.043242 | 4.32 |
| 5 | Energy | MPC | 0.107602 | 10.76 |
| 6 | Health | LWRK.TO | 0.285785 | 28.58 |

***Table # 6: Portfolio Output***

Chart, pie chart

Description automatically generated

Figure 8 Pie Chart of Investment Ratio

This paper applied a neural network-based strategy for predicting stocks and constructing an optimum portfolio for investors. In combined LSTM with Markowiz's Algorithm to predict and construct a portfolio at 98% confidence level with a Sharp ratio of 1.66189794. Referring to Chapter # 1.3.2 Hypothesis, the **Alternate Hypothesis** is supported in this study.

|  |  |  |
| --- | --- | --- |
| **Ref #** | **Hypothesis** | **Result** |
| H0 | Null Hypotheis | Refused |
| H1 | Alternate Hypothesis | Supported |

## **6.2 Study Limitations and Recommendations**

There were some inherent limitations to this study. First, the study was conducted using historical secondary data rather than current primary data. The primary focus was on Canadian Market. Given the complexity of stock price prediction, there are no limits to the potential future projects.

First, it would be worthwhile to conduct a further in-depth study on the extent to which informative features enhance LSTM performance and noisy data reduce generalisation error. Additionally, it may be integrated with research on clustering algorithms and their capacity to classify stocks in a way that offers valuable data on one another's stock values.

Investigating the link between categorisation performance and actual profitability is another helpful strategy. To determine if there is a relationship between categorisation metrics and profitability, combine numerous models with varying degrees of accuracy, precision, and recall.

## **6.3 Conclusion**

The major objective of this portfolio construction was to offer a risk-averse client a better investment opportunity. Initially, seven unique and well-known industries were chosen to construct this portfolio. Then, with the primary goal of generating a return on this investment, this study allocated funds from various areas and then applied LSTM model to each stock to identify the price trend of each stock. This study also calculated the RMSE value and presented the accuracy rates of test and train scores to show the correctness of the predictions. Thus, the study started ensuring the effectiveness of the portfolio.

Then, utilising Markowitz's Model, this study began constructing the portfolio with 19 stocks out of 30. This study estimated the return of stocks and, subsequently, the portfolio return of those stocks. The portfolio's variance and standard deviation were then calculated. With the aid of the risk and return diagram, the researcher determined that if a person accepts 3.2% of the risk, they may anticipate an approximate 11% return. Then, this study computed the correlation between the stocks to decide whether or not they are favourably or negatively associated, as investing in companies with a negative correlation will reduce losses in the event of a price decline. (By Tze Leung Lai, 2011) Finally, this study created a portfolio by weighting the 19 stocks equally and calculating risk and return by multiplying them by the weight.  While doing the portfolio construction procedure, this study executed a series of loops to generate slots for the risk, return, and weights. This loop was executed 100,000 times, with each iteration considering different asset weights and calculating the return and risk of each portfolio combination.

In this study, the procedure was conducted numerous times as it removed the stocks with the lowest Ratio and waited for a perfect combination to produce a greater correlation between the stocks. This study added and deducted stocks to create a perfect correlation. Finally, this study created a portfolio of six (6) stocks with a Sharpe ratio of 1.66, indicating that the portfolio will yield a 23% return if the investor can afford to take a 13% risk.

# References

Achyut Ghosh, S. B. G. M., 2019. *Stock Price Prediction Using LSTM on Indian,* s.l.: Proceedings of 32nd International Conference on Computer Applications in Industry and Engineering.

Anuwoje Ida Logubayom, T. A. V., 2019. *Portfolio Optimization of Some Stocks on the Ghana Stock Exchange Using the Markowitz Mean-Variance Approach,* s.l.: s.n.

Busetti, F., 2006. *Heuristic approaches to realistic,* s.l.: s.n.

By Tze Leung Lai, H. X. A. Z. C., 2011. *MEAN–VARIANCE PORTFOLIO OPTIMIZATION WHEN MEANS AND COVARIANCES ARE UNKNOWN,* s.l.: Institute of Mathematical Statistics.

Chong, E. H. C. P. F. C., 2017. *Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies.,* s.l.: https://doi.org/10.1016/j.eswa.2017.04.030.

Collis, J. a. H. R., 2014. *Business research. 4th ed. Basingstoke: Palgrave Macmillan.,* s.l.: s.n.

Ejara, D. D., 2016. *Evaluating Investments Using Higher Moments,* s.l.: : Ejara, D.D. (2016) Evaluating Investments Using Higher Moments. Modern Economy, 7, 320-326..

Hasan M Sami, L. F. a. D. S. R., 2021. *Portfolio Optimization in DSE Using Financial Indicators, LSTM &,* s.l.: International Journal of Material and Mathematical.

Irma Fitria, E. A. a. E. R. M. P., 2016. *Investment Management Using Portfolio Optimization with Stock Price Forecasting,* s.l.: s.n.

J. B. Heaton, N. G. P. J. H. W., 2016. *Deep Learning in Finance,* s.l.: s.n.

J. Sen, S. M., 2021. *A comparative study of optimum risk portfolio and eigne portfolio on the Indian stock market,* s.l.: s.n.

J. Sen, S. M. a. A. D., 2021. *Volatility modeling of stocks from selected sectors of the Indian economy using GARCH,* s.l.: 10.1109/ASIANCON51346.2021.9544977.

Jaydip Sen, A. D. S. M., 2021. *Stock Portfolio Optimization Using a Deep Learning LSTM Model,* s.l.: s.n.

Jaydip Sen, S. M., 2021. Portfolio Optimization on NIFTY Thematic Sector.

Jaydip Sen, S. M. S. M., 2021. *Portfolio Optimization on NIFTY Thematic Sector Stocks Using an LSTM Model,* s.l.: s.n.

K. Erwin and A. Engelbrecht, 2020. *improved set based particle swarm optimization for portfolio optimisation,* s.l.: s.n.

K. Erwin, A. E., 2020. *Improved set-based particle swarm optimization for portfolio optimization,* s.l.: s.n.

Lee, S. I. a. Y. S. J., 2018. *A new method for portfolio construc-tion using a deep predictive model,* s.l.: Paper presented at the Pro-ceedings of the Seventh International Conference on EmergingDatabases, 260–266..

M. Corazza, G. d. T. G. F. a. R. P., 2021. *a nobel hybrid PSO-based metaheuristic for costly portfolio selection problem,* s.l.: s.n.

M. Ivanova, L. D., 2017. *APPLICATION OF MARKOWITZ PORTFOLIO OPTIMIZATION ON BULGARIAN STOCK MARKET FROM 2013 TO 2016,* s.l.: International Journal of Pure and Applied Mathematics.

Ma, Q., 11 December 2020. *Comparison of ARIMA, ANN and LSTM for Stock Price Prediction,* s.l.: https://doi.org/10.1051/e3sconf/202021801026.

Ma, Q., 2020. *Comparison of ARIMA, ANN and LSTM for Stock Price Prediction,* s.l.: s.n.

Murtaza Roondiwala, H. P. a. S. V., 2015. *Predicting Stock Prices Using LSTM,* s.l.: International Journal of Science and Research (IJSR).

Nelson, D. M. Q., Pereira, A. C. M. & Oliveira, R. A. d., 2017. *Stock market's price movement prediction with LSTM neural networks,* s.l.: IEEE.

Onur Sen, J. U. N. J. A. D., 2021. *Kinetochore life histories reveal an Aurora-B-dependent error correction mechanism in anaphase,* s.l.: s.n.

P. Zhao, S. G. a. N. Y., 2020. *Solving multi-objective portfolio optimization problem based on MOEA/D,* s.l.: s.n.

Perez, T. K. a. K., 2021. *Building portfolios based on machine learning.* [Online].

Pushpendu Ghosha, A. N. J. K. S., 2020. *Forecasting directional movements of stock prices for intraday trading using LSTM and random forests,* s.l.: ScienceDirect.

Sai Krishna Lakshminarayanan, J. M., 2019. *A Comparative Study of SVM and LSTM Deep Learning,* s.l.: AICS.

Samer Obeidat, D. S. M. L., 2018. *Adaptive Portfolio Asset Allocation Optimization with Deep Learning,* s.l.: s.n.

Tun-JenChang, S.-C. K.-J., 2009. *Portfolio optimization problems in different risk measures using genetic algorithm,* s.l.: s.n.

Van-Dai Ta, C.-M. L. D. A. T., 2019. *Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading,* s.l.: s.n.

WAR AHMED, M. B., 2018. *The accuracy of the LSTM model for predicting the S&P 500 index and the difference between prediction and backtesting,* s.l.: s.n.

Wei Chen, H. Z. M. K. M. L. J., 2021. *Mean–variance portfolio optimization using machine learning-based stock price prediction,* s.l.: s.n.

YILIN MA, R. H. ,. A. W. W., 2020. *Prediction-Based Portfolio Optimization Models,* s.l.: s.n.

Zeynep Cipiloglu Yildiz, S. B. Y., 2020. *A portfolio construction framework using LSTM-based stock markets forecasting,* s.l.: s.n.

Zhang, J., Leung, T. & Aravkin, A., 2019. *A Relaxed Optimization Approach for Cardinality-Constrained Portfolios,* s.l.: s.n.

Zhang, R. H. C. Z. W. &. C. S., 2018. *Multi factorstock selection model based on LSTM.,* s.l.: International Journal ofEconomics and Finance,10(8), 1–36.