The Implementation of Benjamin Graham Criteria and *K*-Means Clustering on Stock Selection and Computational Methods for Portfolio Optimization

Bryan Florentino Leo, Sarah Zhafira Fitriyani, and Veren Oktavia Halim
Study Program of Bachelor in Statistics
Department of Mathematics
Faculty of Mathematics and Natural Sciences
Gadjah Mada University

ABSTRACT

This paper explores the integration of Benjamin Graham's investment criteria and *K*-Means Clustering techniques as computational methods for stock selection and portfolio optimization. Benjamin Graham, a renowned figure in the world of finance, is widely recognized for his value investing principles, which emphasize fundamental analysis and the identification of undervalued stocks. *K*-Means Clustering, on the other hand, is a powerful data-driven technique used for grouping similar assets based on their characteristics. The research presented in this paper combines these two distinct approaches to create a sophisticated stock selection strategy. By leveraging the strengths of value investing and data-driven clustering, the objective is to enhance portfolio diversification and improve risk management. Furthermore, advanced computational methods are explored to optimize investment portfolios, aligning them with specific risk-return profiles. The primary goal of this research is to provide investors with a structured and measurable framework for portfolio management. Moreover, this research contributes valuable insights to the field of quantitative finance, offering a promising avenue for investors seeking data-driven approaches to enhance their investment decision-making processes to inform and advance smarter and more effective investment practices, aligning portfolios with investor objectives and market conditions.

Keywords: Benjamin Graham Criteria, K-Means Clustering, Stock Selection, Portfolio

INTRODUCTION

In the ever-evolving world of finance and investment, the quest for intelligent and efficient portfolio management remains an ongoing challenge. Investors across the globe find themselves navigating through increasingly complex and diverse financial markets, marked by high volatility, rapid technological advancements, and a myriad of investment options. As a result, the search for effective methods to manage investment portfolios has never been more critical.

One prominent approach in the realm of investment is the "Graham criteria" developed by Benjamin Graham, a legendary investment expert. These criteria have served as essential guidelines for investors seeking to identify stocks with strong growth potential. Emphasizing fundamental analysis and the identification of undervalued stocks, Graham's approach has withstood the test of time, offering valuable insights to investors.

On the other hand, data analysis techniques, such as *K*-Means Clustering, have gained prominence in various domains. Clustering allows investors to group similar assets based on shared characteristics, facilitating the identification of patterns and relationships among these assets. This approach has proven to be instrumental in portfolio management.

In light of these distinct yet valuable methodologies, this research paper delves into the fusion of Benjamin Graham's investment criteria with *K*-Means Clustering techniques. By integrating these two diverse approaches, the aim is to develop a more sophisticated stock selection strategy, capable of enhancing portfolio diversification and risk management. Moreover, this study explores advanced computational methods to optimize investment portfolios, aligning them with specific risk-return profiles.

The primary objective of this research is to provide investors with a structured and measurable framework for portfolio management. By harnessing the enduring wisdom of Benjamin Graham and incorporating modern data analytics, this paper endeavours to offer deeper insights into the formulation of more intelligent and efficient investment strategies. It is anticipated that the utilization of advanced computational techniques will yield more accurate and tailored solutions for optimizing investment portfolios.

This research contributes invaluable insights to the field of quantitative finance and offers a promising avenue for investors seeking data-driven approaches to enhance their investment decision-making processes. In today's dynamic financial landscape, characterized by high volatility and rapid technological advancements, efficient and informed portfolio management is imperative. The findings of this research have the potential to inform and advance more intelligent and high-performing investment practices, aligning portfolios with investor objectives and market conditions.

METHODS

Literature Review

Recently this year, the research by Putrihadiningrum and Violita (2023) has founded that in investing, portfolios composed of stocks which were precedingly selected by Benjamin Graham criteria implicate to the significance of dividend yield towards the stock return. In the same period, Manurung, Machdar, Foeh, and Sinaga (2023) did a research skewness as another stock selection criterion. The authors stated that most investors would choose stocks with positively skewed returns since extreme right-tail events are more likely than left-tail events. Sortino and Van Der Meer (1991) had precedingly proven the advantages, which some amongst are a stronger Sortino ratio and lower semi-deviation. Chunhachinda, Dandapani, Hamid, and Prakash (1997) also agreed about how portfolio decision involving skewness had caused a major change in the optimality of a portfolio.

Besides, there are plenty of researches which utilized various methods of clustering in stock selection. To begin with, Craighead and Klemesrud (2002) combined cluster and outlier analyses. In their research, clustering was stated to be effective to specify nonuniform stock series. The combination with outlier analysis gave a portfolio a 20.2% annual return. Wider to South America, Da Costa, Jr., Cunha, and Da Silva (2005) founded that clustering could help an investor to detect stocks with similar returns, but different risks. In another words, clustering could assist an investor to choose stocks with a certain rate of return with lower risks, or stocks with a certain level of risk with higher rates. Later research included a publication by Wang (2011), where cluster analysis in selecting stocks improved the success rate and yield of a portfolio, an article by Fadilah and Witiastuti (2018), which explored financial characteristics of companies by clustering, and a practice by Siregar and Pangruruk (2021) of clustering in the LQ45 stock index.

Constructing an optimal portfolio means to find a combination of weights implicating high returns and low risks among many possibilities. There are numerous approaches to optimize a portfolio. One of the simplest is the minimum variance method, which visualizes a curve called the efficient frontier to help investors detect a pair of portfolio return and risk. This method is based on the principle of maximizing return and minimizing risk. Another computational approach is by simulation, where a computer program is constructed to simulate values of portfolio stock weights such that they suit a certain objective function along with its constraints. Wieczynski (2012) shared a practice of stock portfolio simulation using R programming language.

A more modern and currently popular technique to construct an optimal portfolio is machine learning. Machine learning usually appears as programmed models which are trained and tested onto a specified dataset, hoping that the models have sufficient accuracy, precision, and other scores to measure their performance when predicting future datasets. Long-Short Term Memory (LSTM) is one of the highly implemented algorithms, at least from half a decade ago. Ta, Liu, and Tadesse (2019) compared LSTM with multiple portfolio optimization techniques. LSTM outperformed prediction models, such as linear regression and Support Vector Machine (SVM) in terms of accuracy and showed a significant improvement in the return and Sharpe ratio of the respective constructed portfolio. A fine-tuned LSTM also informed investors about stocks with higher returns, both actual and predicted, and precision, as researched by Sen, Dutta, and Mehtab (2021). Finally, a specific implementation of LSTM on LQ45 index was done by Salsabila, Saepudin, and Rohmawati (2023). The outcomes showed that LSTM prediction and Genetic Algorithm could produce an optimal portfolio with high Sharpe ratio.

Benjamin Graham Stock Selection Criteria

Benjamin Graham is an American economist who is often dubbed “the father of value investing". In 1949, Graham authorized his well-known book titled "The Intelligent Investor", which was then considered as the bible of value investing. Warren Buffett, one of the most successful global investors, is one of Graham's disciples who now presents as a brilliant legacy of his.

A great start to achieve a powerful asset portfolio is by utilizing a selection checklist. In terms of stocks, Graham provides ten stock selection criteria. The first five criteria measure return and are focused on stock price, earnings, and dividends, while the second five criteria are related closer to risk and are weighted on the financial soundness of the respective company. The goal of Graham's ten stock selection criteria is to obtain the maximum ratio of return to risk.

Benjamin Graham's stock selection criteria are listed as follows.

1. An earnings-to-price yield at least twice the AAA bond rate.
2. Price-to-earnings (P/E) ratio is less than 40% of the highest P/E ratio the stock had over the past five years.
3. Dividend yield of at least two-thirds the AAA bond yield.
4. Stock price is below two-thirds of its tangible book value per share.
5. Stock price is below two-thirds of the net current asset value (NCAV).
6. The total debt of the company is less than its book value.
7. The current ratio is less than two.
8. The total debt of the company is less than twice the net current asset value (NCAV).
9. The earnings growth of the company during the last ten years is at least at a 7% annual compound rate.
10. The earnings growth stability, that is, there are no more than two declines of 5% or higher in year-end earnings in the last ten years.

Stocks which verify all criteria had been found to be highly rare. Therefore, while exercising these criteria, it is recommended to practice possible combinations along with stepwise verification.

K-Means Clustering

Clustering or cluster analysis is a multivariate statistical method which involves grouping observations into groups called *clusters*. An optimal clustering is defined such that observations within each cluster are as similar as possible, while the clusters are as different as possible to each other.

There are numerous techniques to state the index of similarity or dissimilarity between pairs of observations, but a convenient measure is distance. A well-known distance utilized as an index of dissimilarity in many algorithms of clustering is Euclidean distance. If given two observations, where each observation is represented as a vector of p variables, which are $\boldsymbol{x} = (x_1, x_2, \dots, x_p)'$ and $\boldsymbol{y} = (y_1, y_2, \dots, y_p)'$, the Euclidean distance can be computed as:

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})'(\boldsymbol{x} - \boldsymbol{y})} = \sqrt{\sum_{j=1}^{p} (x_j - y_j)^2}.$$

K-means clustering is a clustering method which is approached by partitioning. That is, the observations are divided into several clusters by initially deciding cluster centers and then reallocating the observations with respect to some optimality criterion.

The algorithm behind a *K*-means clustering is given as follows.

1. Select *k* observations randomly which are at least *r* distance apart. These observations are called seeds.
2. Assign each observation to the nearest seed based on the Euclidean distance.
3. Calculate the centroid of each cluster, which is equivalent to the mean of observations in each cluster.
4. Examine whether each observation is closer to the centroid of another cluster. If so, assign the observation to the new cluster.
5. Repeat steps 3 and 4 until there are no further improvements.

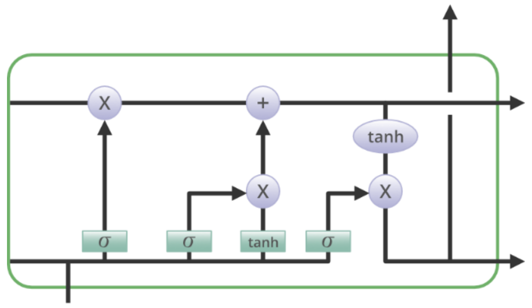
K-means clustering is considered as one of the easiest algorithm of partition-based clustering. However, it is sensitive to the initial choice of seeds. Different initialization leads to different final clusters. Furthermore, a mean value depends on the value of numbers around, so in cases where outliers are present, there might be significant updates on the centroids of some clusters.

Long-Short Term Memory

Long-Short Term Memory (often abbreviated as LSTM) is a modification of Recurrent Neural Network (RNN). While RNN is not capable to predict a sequence (such as series, speech, and text) stored in a long-term memory, but is accurate to do from a recent information, LSTM could retain long-term dependencies by introducing a memory cell, which is a container to hold information for an extended period of time.

LSTM can be stacked into layers of a deep LSTM network to read more complex patterns in a sequential data. LSTM could also be combined with other types of neural network architectures, such as Convolutional Neural Networks (CNNs) to analyse images and/or videos.

LSTM has a chain structure of four neural networks and different memory cells. The structure is illustrated by the image below.



From right to left, the elements labelled by X are the forget, input, and output gates. Consecutively, these gates hold roles to control, remove, and compose outputs from information. The memory cell allows LSTM to discard unrequired information as it passes through the cell, which allows it to learn long-term dependencies.

Starting from the bottom left of the structure, two inputs, x_t and h_{t-1}, consecutively representing input at a particular time *t* and the output from the previous memory cell, are fed to the gate while being multiplied by weights and added by biases. The result passes through the sigmoid activation function,

$$\sigma(x) = \frac{1}{1 + e^{-x}}$$

which gives an output of 0 or 1. If the output equals to 0, the information is forgotten, while if the output is 1, the information is retained. The equation of the forget gate is given as

$$f_t = \sigma\left(\boldsymbol{W}_f \cdot (\boldsymbol{h}_{t-1}, \boldsymbol{x}_t) + \boldsymbol{b}_f\right).$$

In the input gate, useful information is regulated by the sigmoid function. At the same time, the *tanh* activation function, generates a vector with elements between -1 and 1. Both vectors are then multiplied to obtain useful information. The equations of the input gate are:

$$i_t = \sigma(\boldsymbol{W}_i \cdot (\boldsymbol{h}_{t-1}, \boldsymbol{x}_t) + \boldsymbol{b}_i)$$

$$\hat{C}_t = \tanh(\boldsymbol{W}_c \cdot (\boldsymbol{h}_{t-1}, \boldsymbol{x}_t) + \boldsymbol{b}_c)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \hat{C}_t$$

where $\odot$ denotes element-wise multiplication.

Last but not least, the output gate works by initially generating a vector by a *tanh* function. Then, the information is regulated by a sigmoid activation function. The vector and regulated information are finally multiplied and sent as an output or an input to the next memory cell. The equation of the output gate is

$$o_t = \sigma(\boldsymbol{W}_o \cdot (\boldsymbol{h}_{t-1}, \boldsymbol{x}_t) + \boldsymbol{b}_o).$$

Other than handling long-term dependencies, LSTM could also deal with vanishing and exploding gradients, which is a disability in a classic RNN. However, LSTM networks are computationally more expensive and time-consuming. Plus, it is difficult to parallelize the work of processes.

Stock Portfolio

Portfolio is defined as a collection of investments assembled to meet one or more investment goals. Statistically, portfolio is represented as a vector $(w_1, \dots, w_p)$, where the sum of the elements in the vector is 1.

A stock is the share or division of the entire ownership stake of a corporation. A corporation itself is a legally defined, artificial firm, separate from its owners, where there is no limit on the number of owners. Stock is a risky asset, which means it gives a higher return for a higher risk. Portfolios are designed to weigh a set of stocks in order to achieve maximum expected returns (the expected value of portfolio returns) and minimum risk (equivalent to the standard deviation of a portfolio return).

A portfolio is called optimal if it provides the highest return for a given rate of risk, or the lowest risk for a given rate of return. The set of optimal portfolios is represented as the efficient frontier, which could be visualized as a curve on a plot of return to risk. One of the methods to achieve an optimal portfolio is by diversifying the assets involved in it.

Portfolio Simulations

Modern Portfolio Theory (MPT), introduced by Harry Markowitz, forms the basis of portfolio simulations. MPT emphasizes diversification as a means to optimize risk-adjusted returns. MPT works under the assumption that investors are risk-averse, preferring a portfolio with less risk for a given level of return. Under this assumption, investors will only take on high-risk investments if they can expect a larger reward.

The construction of portfolios comprising N assets, each characterized by expected returns (μ) and a covariance matrix ($\boldsymbol{\Sigma}$). Portfolios are defined by weight vectors ($\boldsymbol{\omega}$) that sum up to 1. The expected return (μ_R) and variance (σ^2_R) of a portfolio are expressed as weighted combinations of individual asset returns and covariances, respectively.

In practice, Portfolio Simulations involve the generation of a large number of random portfolios (e.g. 10,000) for analysis. The term "optimal" refers to portfolios with specific characteristics:

1. Portfolios with Minimum Variance
   These portfolios aim to minimize risk and often consist of lower-risk assets, such as treasury inflation-protected bonds.
2. Portfolios with Maximum Sharpe Ratio
   The Sharpe ratio (Sharpe) is defined as the excess return over the risk-free rate (μF) divided by the portfolio's risk (σR). Portfolios with the highest Sharpe ratios offer the best risk-adjusted returns, signifying effective diversification.

The Sharpe ratio is expressed as $Sharpe = \frac{\mu_R - \mu_F}{\sigma_R}$. Importantly, lower correlations between asset returns result in higher Sharpe ratios, reinforcing the significance of diversification in portfolio optimization.

RESULTS AND DISCUSSION

Stock Selection

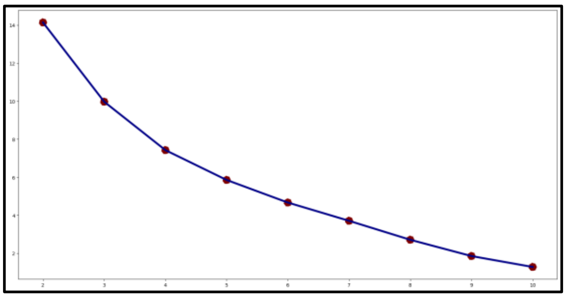
For the selection of the best stocks, a robust selection feasibility analysis is employed by combining the Benjamin Graham stock selection criteria and *K*-Means Clustering. The value of variables required for Graham criteria were imported from multiple financial and investing websites. From the LQ45 stocks, we have obtained the top 15 stocks that meet at least five criteria in the Benjamin Graham theorem. These fifteen stocks are ACES, ADRO, AKRA, ASII, HRUM, ICBP, INDF, INKP, INTP, ITMG, KLBF, SCMA, SIDO, TINS, and UNTR. Below are provided the mean, standard deviation, and skewness values of the selected stocks based on Benjamin Graham's criteria.

Stock	Mean	SD	Skewness
ACES	−0.00306	0.06041	0.27434
ADRO	0.00644	0.06155	0.00039
AKRA	0.00663	0.05023	0.37163
ASII	0.00177	0.03499	0.22344
HRUM	0.00294	0.06830	0.36408
ICBP	0.00312	0.03004	0.51503
INDF	0.00102	0.02386	0.44295
INKP	0.00371	0.04213	0.22850
INTP	0.00669	0.08034	0.84200
ITMG	0.00448	0.05827	−0.15582
KLBF	0.00264	0.03161	0.72842
SCMA	−0.00699	0.05499	−0.07468
SIDO	−0.00194	0.03648	−0.61034
TINS	−0.00447	0.05153	−0.00079

Table 1. Mean, standard deviation, and skewness values of the fifteen selected stocks.

The "Mean" column represents the average returns of each stock. It tells us the typical return an investor can expect from holding that stock over a given period. For instance, for ACES, the mean return is -0.00306, indicating that, on average, this stock has experienced a slight negative return. The "SD" column provides the standard deviation of the stock returns. Standard deviation measures the dispersion or volatility of returns. A higher standard deviation implies greater price volatility or risk. For example, ICBP has a low standard deviation of 0.03004, suggesting relatively lower price volatility compared to other stocks. The "Skewness" column reflects the skewness of the distribution of returns. Skewness measures the asymmetry in the distribution of returns. A positive skewness (greater than zero) indicates a distribution with a longer right tail, meaning the possibility of higher positive returns. Conversely, a negative skewness (less than zero) suggests a longer left tail with a higher chance of negative returns. For instance, INTP has a skewness of 0.84200, indicating a positively skewed distribution.

After obtaining the selected stocks based on Benjamin Graham's criteria, the *K*-Means Clustering method was employed for implementation in order to identify the best stocks using a robust selection method. Attributes involved in clustering are the values of Graham stock selection criteria, along with the expected return, risk, and skewness of the rate of return from each stock. This research utilizes the elbow curve as a visual tool to understand to what extent increasing the number of clusters will provide a significant advantage in data separation.



According to the elbow curve, the most optimal number of clusters is three. The result indicates that there are seven top-performing stocks which cluster fulfils the most Graham criteria, while the expected return and risk are on a mid-range between all clusters. The average skewness in the cluster is also positively valued, which indicates that the cluster of stocks holds a preferable characteristic to investors. These stocks are ACES, ADRO, HRUM, INTP, ITMG, SIDO, and TINS.

Portfolio Weights Calculation

The computation performed to determine portfolio weights using three computational methods, namely Minimum Variance, simulation, and machine learning, have yielded results comparing the performance of these three portfolios at the time of formation by examining the Sharpe ratio in the following table.

Method		Sharpe Ratio				
		First Simulation	Second Simulation	Third Simulation	Fourth Simulation	Fifth Simulation
Minimum Variance		0.1315015	0.1270041	0.2217588	0.2291188	0.1676936
Simulation	Minimum Variance	−0.1835393	−1.1461530	−1.0584464	−1.0206010	−0.8689011
	Maximum Sharpe Ratio	1.7282269	0.2277073	0.0578812	0.0303895	0.0544872
Long-Short Term Memory	Maximum Return	0.4969645	0.7318907	0.8758054	0.8261709	0.7326569
	Minimum Volatility	−0.3722693	0.3847386	−0.1555419	0.9485296	0.6831946

Table 2. Sharpe ratios obtained from five simulations of three methods.

In Table 2, we can observe the Sharpe ratios associated with each method. These ratios serve as indicators of the performance and risk-adjusted returns of the respective portfolios, with higher values generally indicating better performance.

Minimum Variance Method

Based on the weighting above, the weights and the amount of money to be invested by the investor with 1 billion IDR are as follows.

Stock	Weight				
	First Simulation	Second Simulation	Third Simulation	Fourth Simulation	Fifth Simulation
ACES	0.1105	0.1116	0.1121	0.1123	0.1096
ADRO	0.0490	0.0480	0.0540	0.0524	0.0514
HRUM	0.0456	0.0451	0.0191	0.0180	0.0197
INTP	0.3023	0.3001	0.3006	0.2991	0.2991
ITMG	0.1226	0.1236	0.1299	0.1310	0.1310
SIDO	0.3284	0.3293	0.3275	0.3275	0.3289
TINS	0.0415	0.0423	0.0568	0.0598	0.6030

Table 3. Sets of weights from five simulations by Minimum Variance method.

Simulation Method

From the weighting using R, in could be inferred that the Min-Max algorithm in the simulation method results in two sets of weights as follows.

Stock	Weight				
	First Simulation	Second Simulation	Third Simulation	Fourth Simulation	Fifth Simulation
ACES	0.17704630	0.07161494	0.07161494	0.07161494	0.07161494
ADRO	0.04883583	0.03417148	0.03417148	0.03417148	0.03417148
HRUM	0.04713576	0.07688875	0.07688875	0.07688875	0.07688875
INTP	0.21352989	0.25404973	0.25404973	0.25404973	0.25404973
ITMG	0.14178680	0.08795271	0.08795271	0.08795271	0.08795271
SIDO	0.33535388	0.25365313	0.25365313	0.25365313	0.25365313
TINS	0.03631155	0.22166926	0.22166926	0.22166926	0.22166926

Table 4. Sets of weights from five simulations by the simulation method which yield maximum return.

Meanwhile, the weights yielding a maximum Sharpe ratio are given on the table below.

Stock	Weight				
	First Simulation	Second Simulation	Third Simulation	Fourth Simulation	Fifth Simulation
ACES	0.02425792	0.47759236	0.47759236	0.47759236	0.05089807
ADRO	0.38654515	0.01287043	0.01287043	0.01287043	0.01387659
HRUM	0.00594539	0.29041689	0.29041689	0.29041689	0.35812862
INTP	0.35350089	0.19241368	0.19241368	0.19241368	0.47553540
ITMG	0.16284233	0.01378216	0.01378216	0.01378216	0.00320024
SIDO	0.02647821	0.01182240	0.01182240	0.01182240	0.07930282

TINS	0.04043010	0.00110208	0.00110208	0.00110208	0.01905826

Table 5. Sets of weights from five simulations by the simulation method which yield maximum Sharpe ratio.

Machine Learning: Long-Short Term Memory

By training a Long-Short Term Memory (LSTM) model for every simulation, we obtain two sets of weights, which are the weights yielding a maximum return and a minimum volatility. The weights yielding a maximum return from five simulations are given on the table below.

Stock	Weight				
	First Simulation	Second Simulation	Third Simulation	Fourth Simulation	Fifth Simulation
ACES	0.00692496	0.00578116	0.23979003	0.09441550	0.02295950
ADRO	0.33291285	0.58578426	0.06598553	0.14047385	0.41317110
HRUM	0.07377323	0.16305982	0.13687468	0.26502682	0.13180856
INTP	0.00115149	0.07729007	0.22950683	0.10876953	0.03409998
ITMG	0.54982394	0.14810961	0.07933459	0.24817899	0.01388027
SIDO	0.03002099	0.00398191	0.18132838	0.07250189	0.20549174
TINS	0.00539253	0.01599317	0.06717996	0.07063343	0.17858885

Table 6. Sets of weights from five simulations by the Long-Short Term Memory method which yield maximum return.

Meanwhile, the weights yielding a minimum volatility are given on the table below.

Stock	Weight				
	First Simulation	Second Simulation	Third Simulation	Fourth Simulation	Fifth Simulation
ACES	0.07544433	0.19556648	0.12627042	0.12303095	0.11646811
ADRO	0.02498744	0.19249173	0.02291872	0.02408930	0.06186528
HRUM	0.06803585	0.19367959	0.01655343	0.04424168	0.01249609
INTP	0.29533228	0.17726421	0.25449203	0.29721830	0.31144550
ITMG	0.13220024	0.15073583	0.18645415	0.15355577	0.15665780
SIDO	0.36648279	0.04948771	0.35038532	0.31782141	0.29984900
TINS	0.03751708	0.04077445	0.04292594	0.04004259	0.04121822

Table 7. Sets of weights from five simulations by the Long-Short Term Memory method which yield minimum volatility.

Portfolio Performance

A mock-up portfolio is constructed for each set of weights with an initial deposit of Rp1,000,000,000.00. Shares of the selected stocks were purchased on 27 September 2023 according to the nominals of investment respective to each set of weights. The performance and price changes were observed for five days.

The close prices per share for each selected stock on 27 September 2023 and 5 October 2023 are given on the table below.

Stock	Price on 27 September 2023 (Rp)	Price on 5 October 2023 (Rp)
ACES	755	760
ADRO	2,870	2,610
HRUM	1,870	1,635
INTP	10,075	10,125
ITMG	28,375	26,325
SIDO	585	605
TINS	800	770

Table 8. One-week price change of the seven selected stocks.

It could be interpreted that four stocks (ADRO, HRUM, ITMG, and TINS) showed a price decline, while the prices of the other three stocks (ACES, INTP, and SIDO) increased. Hence, after five days of investment, the total returns per stock by each method of portfolio weighing are as follows.

Stock	Return (Rp)				
	First Simulation	Second Simulation	Third Simulation	Fourth Simulation	Fifth Simulation
ACES	146,357,616	141,265,823	153,561,644	155,972,222	148,108,108
ADRO	170,731,171	16,271,186	19,014,085	18,131,488	19,179,104
HRUM	24,385,027	24,780,220	10,552,486	10,810,811	12,507,937
INTP	30,004,963	29,712,871	27,962,791	27,630,485	28,216,981
ITMG	4,320,705	4,276,817	4,479,310	4,260,163	4,486,301
SIDO	561,367,521	548,833,333	555,084,746	545,833,333	534,796,748
TINS	51,875,000	52,222,222	71,000,000	67,570,621	67,752,809
Total Return	**835,384,002**	**817,362,472**	**841,655,061**	**830,209,123**	**815,047,989**

Table 9. Return of the portfolio generated by Minimum Variance method.

Stock	Return (Rp)				
	First Simulation	Second Simulation	Third Simulation	Fourth Simulation	Fifth Simulation
ACES	1,172,492.05	474,271.13	474,271.13	474,271.13	474,271.13
ADRO	-4,424,151.85	-3,095,674.15	-3,095,674.15	-3,095,674.15	-3,095,674.15
HRUM	-5,923,477.86	-9,662,489.97	-9,662,489.97	-9,662,489.97	-9,662,489.97
INTP	1,059,701.69	1,260,792.70	1,260,792.70	1,260,792.70	1,260,792.70
ITMG	-10,243,627.84	-6,354,292.70	-6,354,292.70	-6,354,292.70	-6,354,292.70
SIDO	11,465,089.91	8,671,901.88	8,671,901.88	8,671,901.88	8,671,901.88
TINS	-1,361,683.13	-8,312,597.25	-8,312,597.25	-8,312,597.25	-8,312,597.25

Total Return	**-8,255,657.02**	**-17,018,088.36**	**-17,018,088.36**	**-17,018,088.36**	**-17,018,088.36**

Table 10. Return of the portfolio generated by simulation method and associated with maximum return.

Stock	**Return (Rp)**				
	First Simulation	**Second Simulation**	**Third Simulation**	**Fourth Simulation**	**Fifth Simulation**
ACES	160,648.50	3,162,863.34	3,162,863.34	3,162,863.34	337,073.34
ADRO	-35,018,027.89	-1,165,962.57	-1,165,962.57	-1,165,962.57	-1,257,112.59
HRUM	-747,148.44	-36,496,240.56	-36,496,240.56	-36,496,240.56	-45,005,468.67
INTP	1,754,346.85	954,906.60	954,906.60	954,906.60	2,359,977.17
ITMG	-11,764,820.39	-995,715.24	-995,715.24	-995,715.24	-231,206.41
SIDO	905,237.88	404,184.48	404,184.48	404,184.48	2,711,207.42
TINS	-1,516,128.75	-41,327.96	-41,327.96	-41,327.96	-714,684.90
Total Return	**-46,225,892.25**	**-34,177,291.92**	**-34,177,291.92**	**-34,177,291.92**	**-41,800,214.64**

Table 11. Return of the portfolio generated by simulation method and associated with maximum Sharpe ratio.

Stock	**Return (Rp)**				
	First Simulation	**Second Simulation**	**Third Simulation**	**Fourth Simulation**	**Fifth Simulation**
ACES	45,860.66	38,285.83	1,588,013.44	625,268.21	152,049.67
ADRO	−30,159,352.26	−53,067,563.62	−5,977,783.21	−12,725,854.01	−37,430,134.49
HRUM	−9,270,967.41	−20,491,474.71	−17,200,828.77	−33,305,509.47	−16,564,177.33
INTP	5,714.59	383,573.55	1,138,991.71	539,799.16	169,230.67
ITMG	−39,722,963.07	−10,700,429.97	−5,731,662.01	−17,930,112.05	−1,002,803.65
SIDO	1,026,358.63	136,133.68	6,199,260.85	2,478,697.09	7,025,358.63
TINS	−202,219.88	−599,743.88	−2,519,248.50	−2,648,753.63	−6,697,081.88
Total Return	**−78,277,568.73**	**−84,301,219.13**	**−22,503,256.47**	**−62,966,464.69**	**−54,347,558.37**

Table 12. Return of the portfolio generated by LSTM method and associated with maximum return.

Stock	**Return (Rp)**				
	First Simulation	**Second Simulation**	**Third Simulation**	**Fourth Simulation**	**Fifth Simulation**
ACES	499,631.32	1,295,142.25	836,227.95	814,774.50	771,311.99
ADRO	−2,263,670.52	−17,438,275.19	−2,076,260.35	−2,182,305.92	−5,604,520.14
HRUM	−8,549,959.76	−24,339,413.72	−2,080,243.88	−5,559,783.32	−1,570,364.25
INTP	1,465,668.88	879,723.13	1,262,987.74	1,475,028.78	1,545,635.24
ITMG	−9,551,030.56	−10,890,165.69	−13,470,696.30	−11,093,897.04	−11,318,008.46
SIDO	12,529,326.15	1,691,887.52	11,978,985.30	10,865,689.23	10,251,247.86
TINS	−1,406,890.50	−1,529,041.88	−1,609,722.75	−1,501,597.13	−1,545,683.25
Total Return	**−7,276,924.98**	**−50,330,143.58**	**−5,158,722.29**	**−7,182,090.89**	**−7,470,381.01**

Table 13. Return of the portfolio generated by LSTM method and associated with minimum volatility.

In conclusion, among five existing sets of weights produced from three methods, the Long-Short Term Memory (LSTM) method associated with minimum volatility is considered the best method for this case since it yields the least loss.

CONCLUSION

Out of the latest 45 stocks involved in LQ45 stock index, fifteen stocks are selected based on Benjamin Graham stock selection criteria, where those stocks fulfil at least five out of ten criteria. Moreover, three clusters are formed based on *K*-means algorithm such that at the end of the stock selection process, there are seven stocks considered best for a portfolio.

To look for the most optimal portfolio, three computational methods are compared: minimum variance, simulation, and Long-Short Term Memory as a machine learning approach. By computation, five sets of portfolio weights are obtained, where Min-Max simulation method yields sets of weights associated with minimum variance and maximum Sharpe Ratio, and Long-Short Term Memory gives out sets of weights producing maximum return and minimum volatility.

The mock-up investment began on 27 September 2023. Every set of weights are implemented into distinct portfolios with initial deposit of Rp1,000,000,000.00. The performance of portfolios were observed for five work days. Four out of seven selected stocks showed a significant decline in price, while the other three stock prices slightly increased. Therefore, unfortunately, given period of investment, all five sets of weights would yield loss to the investor, where weights generated by Long-Short Term Memory (LSTM) method associated with minimum volatility would bring the least loss, while the weights produced by minimum variance approach would implicate the greatest loss.

REFERENCES

Chunhachinda, P., Dandapani, K., Hamid, S., & Prakash, A. J. (1997). Portfolio selection and skewness: Evidence from international stock markets. *Journal of Banking & Finance, 21*(2), 143-167. doi:10.1016/S0378-4266(96)00032-5

Craighead, S., & Klemesrud, B. (2002). Stock Selection Based on Cluster and Outlier Analysis. *IMA Conference*, -. Retrieved from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=139cc954610236f86b1dbe1db1667c56b0d1105f

Da Costa Jr., N., Cunha, J., & Da Silva, S. (2005). Stock selection based on cluster analysis. *Economics Bulletin, 13*(1), 1-9. Retrieved from http://www.accessecon.com/pubs/eb/2005/volume13/eb-05m20001a.pdf

Fadilah, I., & Witiastuti, R. S. (2018). A Clustering Method Approach for Portfolio Optimization. *Management Analysis Journal, 7*(4), 436-447. doi:10.15294/maj.v7i4.23378

Ferretti, S. (2023). On the Modeling and Simulation of Portfolio Allocation

Schemes: an Approach Based on Network Community Detection. *Computer Economics, 62*, 969-1005. doi:10.1007/s10614-022-10288-w

Li, P., & Xu, J. (2022). A Study of Different Existing Methods for the Stock Selection in the Field of Quantitative Investment. (M. F. Hashmi, Ed.) *Wireless Communication and Mobile Computing, 2022*, -. doi:10.1155/2022/2695099

Manurung, A. H., Machdar, N. M., Foeh, J. E., & Sinaga, J. (2023). Stock Selection Using Skewness to Construct a Portfolio and the Effects of Variables on Portfolio Return. *Open Journal of Business and Management, 11*, 1000-1012. doi:10.4236/ojbm.2023.113055

Peachavanish, R. (2016). Stock Selection and Trading Based on Cluster Analysis of Trend and Momentum Indicators. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2016*, -. Retrieved from https://www.iaeng.org/publication/IMECS2016/IMECS2016_pp317-321.pdf

Putrihadiningrum, D. C., & Violita, C. E. (2023). THE SMART INVESTOR: USING BENJAMIN GRAHAM SELECTION CRITERIA ON STOCK RETURN. *TALIJAGAD, 1*(5), 26-31. Retrieved from https://journal.unusida.ac.id/index.php/tali-jagad/article/view/856/568

Salsabila, T., Saepudin, D., & Rohmawati, A. A. (2023). Optimizing LQ45 Stock Portfolio To Maximize Sharpe Ratio Value Using LSTM. *Journal of Computer System and Informatics (JOSYC), 4*(2), 339-347. doi:10.47065/josyc.v4i2.3065

Sen, J., Dutta, A., & Mehtab, S. (2021). Stock Portfolio Optimization Using a Deep Learning LSTM Model. *IEEE Mysurucon 2021*, -. doi:10.48550/arXiv.2111.04709

Siregar, B., & Pangruruk, F. A. (2021). Portfolio Optimization Based on Clustering of Indonesia Stock Exchange: A Case Study of Index LQ45. *Indonesian Journal of Business Analytics, 1*(1), 59-70. doi:10.55927/ijba.v1i1.5

Ta, V.-D., Liu, C.-M., & Tadesse, D. A. (2020). Portfolio Optimization-Based Stock Prediction Using Long-Short Term Memory Network in Quantitative Trading. *Applied Sciences, 10*(2), 437. doi:10.3390/app10020437

Wang, R. (2012). Stock Selection Based on Data Clustering Method. *tock Selection Based on Data Clustering Method*, 1542-1545. doi:https://doi.org/10.1109/CIS.2011.345

Wieczynski, P. (2022, - -). *Portfolio Simulations*. Retrieved from RPubs by RStudio: https://rpubs.com/pawel-wieczynski/874454