BAREKENG: Journal of Mathematics and Its Applications
March 2026 Volume 20 Issue 1 Page 0381–0396
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss1pp0381-0396

COMPANY VALUATION AND PORTFOLIO ANALYSIS BASED ON K-MEANS CLUSTERING IN KOMPAS 100 STOCKS INDEX

Rohmatul Fajriyah [1*], **Yoel Christopher Tjen** [2]

[1,2] *Master Program in Statistics, Department of Statistics, Faculty of Mathematics and Natural Sciences, Universitas Islam Indonesia*
Jln. Kaliurang Km 14.4., Sleman, Yogyakarta 55584, Indonesia

*Corresponding author's e-mail: * **966110101@uii.ac.id***

Article Info

Article History:

Received: 27th March 2025
Revised: 24th May 2025
Accepted: 28th July 2025
Available online: 24th November 2025

Keywords:

K-means;
PBV;
PER;
Portfolio;
PSR;
Valuation.

ABSTRACT

The capital market plays a vital role in investment, providing a platform for trading long-term financial instruments. Indonesia's capital market has shown significant growth in recent years. This study aims not only to find stock clusters but to show that grouping stocks based on similar valuation characteristics can serve as a solid foundation for constructing superior-performing portfolios. The Kompas 100 index is used because it represents the most liquid and fundamentally stocks in Indonesia. The k-means clustering method is employed, and the number of clusters is determined using the elbow method. This approach resulted in four clusters, with the cluster identified as containing stocks with low PER, PBV, and PSR, representing the "best" portfolio each year based on valuation. Portfolios were formed from these clusters and compared to benchmark portfolios in Indonesia and globally. Global portfolios used as benchmarks include VSMPX, FXAIX, and SAM Equity. Over five years (2018–2022), the cluster-based portfolios outperformed Indonesian and global benchmarks in 2018, 2021, and 2022, while slightly underperforming global portfolios in 2019 and 2020 but still exceeding Indonesian benchmarks. This confirms that clustering techniques can deliver strong performance compared to conventional methods. A limitation of this study is that it focuses only on return performance without analyzing risk-adjusted returns, which future research should address.

How to cite this article:

R. Fajriyah and Y. C. Tjen., " COMPANY VALUATION AND PORTFOLIO ANALYSIS BASED ON K-MEANS CLUSTERING IN KOMPAS 100 STOCKS INDEX," *BAREKENG: J. Math. & App.*, vol. 20, iss. 1, pp. 0381-0396, Mar, 2026

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · Open Access

1. INTRODUCTION

The Indonesian capital market has experienced rapid growth in recent years. This is shown by the increase in the number of investors and transaction value on the Indonesian Stock Exchange (BEI). Based on data from the Indonesian Central Securities Depository (KSEI), in 2023 the number of capital market investors in Indonesia will reach 11.4 million, up 10.75% from the previous year.

One of the investment instruments that is popular is shares. Shares provide high profit opportunities for investors, but are also accompanied by high risks [2]. Shares can be measured by Stock Return, the level of profit from an investment. Stock Return can be calculated by dividing the current stock price by the previous price [3]. The risk of stock investment can be minimized by fundamental analysis [3]. Fundamental analysis seeks to assess the intrinsic value of a company using various financial indicators, [4] and [5]. One indicator that is often used is relative valuation.

According to [6], [7], [8], [9], [10], and [11], relative valuation is a method of assessing shares by comparing them with other similar companies. In this method, investors compare financial ratios such as PER, PBV, and PSR of a company with other companies in the same industry or with companies which have similar characteristics. According to Damodaran [7], the aim of the relative valuation method are finding undervalued stocks, by comparing a company's financial metrics to industry averages or competitors, investors can identify stocks that are undervalued relative to their potential. Undervalued stocks may have a higher chance of returning in the future. Next is comparing different companies, relative valuation allows investors to compare companies from different industries with different sizes and capital structures. Lastly, assessing company performance. Financial metrics used in relative valuation can help investors to monitor a company's performance compared to its competitors and identify trends.

Some researchers, such as [7], [8], [9], [10], and [11] explain that by using relative valuation, investors can find out whether a stock is undervalued (has a price that is cheaper than its intrinsic value) or overvalued (has a price that is more expensive than its intrinsic value). Although some sources as [6], [7], [8], [9], [10], and [11]) mentioned that the use of relative valuation has several problems, such as comparisons that are too simple so they do not take into account all existing factors, industry differences that will have different interpretations and financial metrics that can change significantly in the short term, Saputra et al. [12] has stated several advantages of the relative valuation method. The advantages are comparing different companies, finding undervalued/overvalued stocks, stock screening, monitoring company performance, and simplicity. Therefore, it can be seen that relative valuation is a useful tool for investors to compare companies, find undervalued shares, and monitor company performance.

Previous studies and practices in portfolio construction have predominantly relied on Modern Portfolio Theory based on Markowitz in 1952, which focuses on optimizing risk and return through diversification using historical variance and covariance of returns. However, this approach often overlooks the consistency of fundamental valuation metrics within a portfolio and faces practical limitations due to changing market dynamics and data instability. Recent studies have introduced clustering techniques as an alternative method for portfolio formation. For example, Issakainen [13] demonstrated that k-means clustering based on valuation metrics such as P/E, P/B, and P/CF on the Helsinki Stock Exchange produced portfolios with superior annual returns and risk-adjusted performance compared to traditional benchmarks. Similarly, Aslam et al. [14] showed that portfolios constructed through clustering based on three fundamental factors achieved higher Sharpe ratios, outperforming conventional optimization methods. In addition, Zhang [15] critically reviewed the weaknesses of Markowitz's model, especially in emerging markets, and highlighted the potential of machine learning-based clustering as a more adaptable and robust alternative for portfolio construction. However, research applying valuation-based clustering techniques remains scarce in Indonesia, particularly as a substitute for Markowitz’s approach. This study addresses this gap by constructing portfolios through clustering companies based on relative valuation metrics and testing their performance against benchmark portfolios both in Indonesia and globally.

This research focuses on unifying the value (valuation) of different companies through clustering techniques by grouping companies with similar valuation levels and profiling each group, using R/RStudio for computation. The research specifically addresses the gap where most investors and investment managers rely on Markowitz’s modern portfolio theory, while this study offers a new approach by constructing portfolios based on valuation clusters. The best valuation cluster will be used to form a portfolio and tested against real-world benchmark portfolios in Indonesia and globally to assess its performance.

2. RESEARCH METHODS

2.1 Research Design

This study is a quantitative research approach, utilizing the secondary dataset to perform valuation and portfolio analysis clustering-based. Initial exploration was performed using descriptive statistics to understand the general characteristics of the dataset. To identify the inherent grouping in the dataset, the k-means clustering was applied. The optimal number of clusters was determined using the silhouette scores. The computational was done with RStudio statistical software.

2.2 Data set

The data set we used in this study is the Kompas 100 stocks index. The data were obtained from [1]. The data set can be found in this link. It consists of 100 stocks but for further analysis the only 88 stocks are used due to incomplete data for 12 companies in 2018-2022 and excluded two shariah companies to make sure that all data sets have the same business principles.

2.3 Data Analysis

2.3.1 Descriptive Analysis

Descriptive analysis is an important first step in exploring and understanding the characteristics of a dataset. In this context, we explore data regarding Price Earnings Ratio (PER), Price Book Value (PBV), and Price Sales (PSR) from a sample. According to Kemp, Ng and Hollowood [16] descriptive analysis provides a framework for understanding several statistics that describe the distribution and patterns of observed data.

Standard deviation and variance provide information about the distribution of data. A very high standard deviation of PER indicates great variability in the price-to-earnings ratio, while PSR has a lower standard deviation, indicating greater consistency in the price-to-sales ratio [17]. Skewness and kurtosis describe the shape of the data distribution. Large positive skewness in all variables indicates the presence of a long tail on the right side of the distribution, indicating the presence of outliers with very high values [18]. The range shows the difference between the maximum and minimum values in the dataset.

2.3.2 Clustering Analysis

Suppose you have data observation where the characteristics (X) of each object have been measured. If we want to know whether there are certain groups based on the characteristics of these objects, then we can perform cluster analysis. Cluster analysis is used to group similar objects in one particular cluster and dissimilar objects in another cluster, [19]-[20] There are several similarity measurement, one of which is distance. In the similarity measure based on distance, two objects cluster differently or are not similar if the distance between them is greater [21].

There are two types of clustering methods, namely hierarchical and non-hierarchical. The K-means, K-medoid and FCM methods are included in the non-hierarchical type (partition-based method). Cluster analysis is an inter-dependency analysis. Therefore, in cluster analysis the dependent and independent variables are not known and the final result is not a model or functional relationship.

The steps in cluster analysis in general are formulate the problem (through defining the variables used in grouping) and choose the right distance measure. The function of the distance measure is to see the degree of similarity/dissimilarity of the research objects to be grouped [21]. In this research, the Euclidean distance similarity measure is used based on the scale of the available data are ratio (metrics), where there is no categorical and or nominal scales.

In this research, the k-means clustering method is used. The method is choosen based on some consideration: it is computationally efficient, easy to implement and best suited for continuous numerical variables. The steps in the k-means method with the Euclidean distance similarity measure are:

1. Determine the initial cluster, k, and obtain the initial centroids $m_1^1, m_2^1, \dots, m_k^1$
2. For each object (data) calculate the distance to each centroid using the distance similarity measure

$$S_i^t = \left\{x_p: \left\|x_p - m_i^t\right\|^2 \leq \left\|x_p - m_j^t\right\|^2 \forall j, \quad 1 \leq j \leq k\right\} \tag{1}$$

Allocate the data to the cluster with the closest distance to the centroid.

3. Calculate the centroid of each cluster, as a new centroid with the formula

$$m_i^{t+1} = \frac{1}{S_i^t} \sum_{x_j \in S_i^t} x_j \tag{2}$$

4. Calculate the total deviation (D) by calculating the value of the new total distance – the old total distance. If D < 0, then swap objects with cluster data to form a new set of k objects as centroids.
5. Repeat steps 1 to 4 until there are no changes in the centroid, such that you get the clusters and their respective cluster members.

To see whether the clustering we have done is good or not, we can assess it through two things as follows: within cluster the members have a high level of similarity (high homogeneity) and between clusters it is very heterogeneous (high heterogeneity), [21].

The research method in this study are statistical descriptive and clustering. The data set used in this research is the Kompas 100 stocks index which consists of 100 shares, but only 66 companies are used. First there is no data for 12 companies in the 2018-2022 and second, there are some companies have negative PER. Therefore, we eliminate further of these 88 to become 66. The elimination is based on Beidleman [22], which stated that if a company experiences a loss it will also cause a negative PER which cannot be compared and represent the ambiguous result. Subsequently, the researcher will create several clusters for each year to identify the optimal cluster that can be formed into a portfolio. Following this, the portfolio consisting of the best-performing clusters will be compared to the performance of several top portfolios in the real world. The best clusters are selected based on having the lowest average valuation ratios, namely PER, PBV, and PSR. Lower valuation ratios indicate that the stocks within the cluster are, on average, undervalued, making them fundamentally more attractive for portfolio formation compared to other clusters.

2.3.3 Portfolio Analysis

A portfolio comprises various investment assets, such as stocks and bonds, owned by an investor or institution to achieve capital appreciation or income through dividends, [23]. Diversification is a crucial strategy within a portfolio, involving investments in stocks across different companies, sectors, or industries to minimize overall risk. The aim of a stock portfolio is to enhance investment returns in alignment with the investor's risk tolerance, while taking into account factors like market volatility, company growth prospects, and global economic stability.

The portfolio derived from the best cluster will be equally weighted across all stocks within that cluster. For comparison in this study, the following portfolios will be used:

1. Jakarta Composite Index (JCI): The primary stock market index of the Indonesia Stock Exchange (IDX), including all listed companies on the IDX, and serving as a measure of overall performance for Indonesian stocks.
2. LQ45: Comprising the 45 most liquid and well-capitalized stocks on the Indonesia Stock Exchange, representing the leading companies in terms of market capitalization and trading volume.
3. KOMPAS 100: A selection of 100 stocks listed on the Indonesia Stock Exchange, chosen based on liquidity and market capitalization, offering a more comprehensive view of the Indonesian stock market compared to LQ45.
4. Sam Indonesian Equity Fund: An actively managed mutual fund investing primarily in Indonesian stocks, targeting long-term capital growth by selecting high-potential equities. As of 2023, it is recognized as the top mutual fund in Indonesia.
5. Vanguard Total Stock Market Index Fund: An index fund that aims to track the CRSP US Total Market Index, providing exposure to the entire U.S. stock market, including large-, mid-, small-, and micro-cap stocks across various sectors.
6. Fidelity 500 Index Fund: An index fund designed to mirror the performance of the S&P 500 Index, which includes 500 of the largest publicly traded U.S. companies, reflecting a broad spectrum of the American economy.

A portfolio performance is evaluated by calculating the annual returns of the portfolio formed from the best cluster [24]. This portfolio's performance will then be compared with the top 6 real-world portfolios to provide a thorough analysis of how the cluster-based portfolio measures up against established market portfolios. This evaluation seeks to demonstrate the effectiveness and competitiveness of the best cluster portfolio in achieving superior returns. In this paper, the performance valuation of each cluster will be compared based on the PER, PBV and PSR aspects. The best cluster for each year will be compared to the Indonesia and global portfolios, in order to see whether clustering is the best strategy for the buyer before doing investment to some certain companies.

2.4 Dunn Index

Dunn index is one of the internal validation indexes in determining the optimal number of clusters. It calculates the ratio of the minimum value of inter-cluster dissimilarity and the maximum value of intra-cluster diameter. The higher the index the more optimal cluster has been formed. The Dunn index (DI) formula [25] is

$$DI = \frac{\min d_{(c_p,c_q)}}{\max diam(c_r)} \tag{3}$$

with

$$\min d_{(c_p,c_q)} = \min_{x_i \in c_p, y_i \in c_q} d_{(x_i,y_i)} \tag{4}$$

$$\max diam(c_r) = \max_{z_i,z_j \in c_r} d_{(z_i,z_j)} \tag{5}$$

where

$\min d_{(c_p,c_q)}$	: distance between cluster p and cluster q
$diam(c_r)$	: diameter of the cluster r
x_i	: i-th object in cluster p
y_i	: i-th object in cluster p
z_i, z_j	: i-th and j-th objects in cluster r

2.5 Some Concepts

2.5.1 Valuation

Valuation is the process of determining the fair or intrinsic value of a stock. Valuation can help investors decide whether a stock is overvalued, fairly valued, or undervalued compared to its market price, [26]. Stock valuation is a key aspect of fundamental analysis, and various methods can be used to assess the fair value of a stock. According to Damodaran [7], several commonly used valuation methods include:

1. Real Option Valuation, developed by Black and Scholes in 1973, is a method for assessing the value of strategic flexibility in investment decisions using financial option principles. However, this method is not yet applicable in Indonesia due to the absence of options (Call and Put) in the stock trading market in Indonesia, and it is one of the most complex valuation methods in terms of calculation, [27].
2. Discounted Cash Flow (DCF) is a valuation method that calculates the present value of future cash flows expected to be generated by an asset or company. This method is widely used by professionals in the capital markets, but its calculations can be complex for laypersons, [28].
3. Dividend Discount Model (DDM) is a valuation method that calculates the intrinsic value of a stock based on the present value of all future dividends expected to be paid to shareholders. This method has a significant drawback as many stocks in Indonesia do not pay dividends regularly, making this calculation difficult, [29].
4. Relative Valuation [6], is a valuation method that assesses an asset or company by comparing it to similar assets or companies, typically using financial ratios such as Price-to-Earnings Ratio (PER), Price-to-Book Value (PBV), and Price-to-Sales Ratio (PSR). This method is most used by investors due to the ease of obtaining data and relatively simple calculations, [30].

2.5.2 Price to Earnings Ratio (PER)

The Price-Earnings Ratio (PER) is defined as a ratio used to measure the value of a share based on earnings per share, [31] and [32]. PER is calculated by dividing the share market price by earnings per share (EPS). A high PER shows that investors are willing to pay more for every rupiah of profit the company generates, [33] and [34]. A high PER can also indicate that the stock is overvalued . (because the price is high relative to earnings, reflecting investor expectations that may not align with the company's actual performance). Therefore, it is important to consider PER along with other factors, such as profit growth, dividends and company risk. The following is the formula used to calculate the price to earnings ratio:

$$Price\ to\ Earnings\ (PER) = \frac{End\ of\ year\ closing\ stock\ price\ n}{Net\ Profit\ n} \tag{6}$$

2.5.3 Price to Book Value Ratio (PBV)

According to Damodaran [7] Price-Book Value Ratio (PBV) is a ratio used to measure the value of a share based on its book value. PBV is calculated by dividing the stock market price by the book value per share (BVPS). A low PBV indicates that investors buy shares at a price that is cheaper than the asset value, [35]. This could reflect potential undervaluation. However, a low PBV can also indicate that the company has financial problems or low growth prospects. The following is the formula that will be used to calculate Price to Book Value:

$$Price\ to\ Book\ Value\ (PBV) = \frac{End\ of\ year\ closing\ stock\ price\ n}{Total\ Equity\ n} \tag{7}$$

2.5.4 Price to Sales Ratio (PSR)

Reilly & Brown [36] argue that the Price-Sales Ratio (PSR) is a ratio used to measure the value of a share based on its earnings per share. PSR is calculated by dividing the stock market price by earnings per share (Sales per Share). Tanius and Widjojo [37] explained that a high PSR shows that investors are willing to pay more for every rupiah of revenue generated by the company. A high PSR can also indicate that the stock is overvalued. Therefore, it is important to consider PSR along with other factors, such as revenue growth, profitability and Company risk. The following is the formula that will be used to calculate Price to Sales:

$$Price\ to\ Sales = \frac{End\ of\ year\ closing\ stock\ price\ n}{Total\ Sales\ n} \tag{8}$$

3. RESULTS AND DISCUSSION

Based on data from the Indonesian Central Securities Depository (KSEI), in 2023 the number of capital market investors in Indonesia will reach 11.4 million, up 10.75% from the previous year, see Table 1.

Table 1. Indonesia Investor Growth [1]

Products	Indonesian Investor Growth					
	2021	**2022**	**May 2023**	**June 2023**	**July 2023**	**2023 - YTD**
Capital Market	92.99%	37.68%	7.28%	1.50%	1.71%	10.75%
Mutual Fund	15.41%	40.41%	7.71%	1.56%	1.78%	11.35%
Shares and Other Securities	103.60%	28.64%	7.12%	1.19%	1.58%	10.11%
Government Securities (SBN)	32.75%	36.05%	8.57%	0.91%	2.00%	11.75%

Table 2 shows the numerical description of Kompas 100 stock index data, where the average (mean) of PER, PBV, and PSR is as follows: PER 30.15, PBV 6.28, and PSR 4.22. This average value provides a rough idea of the price level an investor might accept in relation to a company's earnings, book value, and sales. The ranges of PER, PBV, and PSR are 2409.9959, 670.6523, and 232.6774, respectively, reflecting significant variations in these three variables.

In addition, the standard deviation values for PER, PBV, and PSR are 137.79, 40.16, and 15.58, respectively, indicating high variability, especially for PER, which reflects greater fluctuations and inconsistent investor expectations. Skewness values of 15.97 (PER), 14.39 (PBV), and 11.17 (PSR) indicate a long right tail, showing the presence of a few companies with extremely high valuations. The kurtosis values of 273.30, 231.01, and 148.56 further confirm the existence of extreme outliers and heavy-tailed distributions.

Table 2. Kompas 100 Stocks Index Description

	PER	PBV	PSR
Mean	30.15	6.28	4.22
Standard Error	7.59	2.21	0.86
Median	13.69	1.21	1.62
Standard Deviation	137.79	40.16	15.58
Sample Variance	18,986.97	1,612.50	242.84
Kurtosis	273.30	231.01	148.56
Skewness	15.97	14.39	11.17
Range	2,410.00	670.65	232.68
Minimum	0.38	0.00	0.03
Maximum	2,410.37	670.65	232.71

3.1 Clustering Results

In the analysis, researchers have chosen to form 4 clusters each year. This is based on cluster validation using the *clValid* package in RStudio. The recommendations from the package, based on the Connectivity, Dunn and Silhouette indexes are 2, 3, 4 or 6 clusters. Then, the number of clusters chosen by the researchers was four (4), because we wanted to see the variability in the characteristics of the clusters in the Kompas 100 stocks index data and the Dunn index suggest in 4 clusters. If only two clusters were selected, the researchers would only characterize the shares of good and bad companies, where the distribution of cluster members is very extreme contains only one member. In the formation of these 4 clusters, the members of each cluster in 2018-2022 can be seen in Table 3.

Table 3. Members for Each Cluster

Year	Member of Cluster			
	1	2	3	4
2018	44	20	1	1
2019	4	60	1	1
2020	57	2	1	6
2021	59	5	1	1
2022	63	1	1	1

Each cluster in Tables 4 - 8 show the average value of each valuation variable, namely Price to Earnings Ratio (PER), Price to Book Value Ratio (PBV) and Price to Sales ratio (PSR), unless when the member is only one. A high PER indicates that a company's share price tends to be high compared to the profits it makes. A high PBV means that a company's share price tends to be high when compared to the book value or equity of the company. A high PSR means that a company's share price tends to be higher than the income earned by the company. Therefore, if PER, PBV and PSR have a high value, it means the share price is overvalued and not worth buying.

In this research we found something interesting, namely the Adi Sarana Armada (ASSA) company, from 2018-2022 is a company that is always self-clustered, see Figs. 1 - 5. If we examine further, ASSA's valuation variables (PER, PBV and PSR) are higher than average, especially for PBV which is already worth hundreds, see Tables 4 - 8. ASSA company is in cluster 3 at Tables 4-7 and in cluster 2 in Table 8. The tables show that the ASSA company shares are very overvalued, in comparison to others.

Next is a discussion of cluster profiling related to valuation variables for each year. To assess the portfolio performance, clusters that only consist of 1 or 2 shares cannot be used for comparison, a minimum of 3 shares are required.

3.1.1 Cluster Profiling Based on Valuation Variables and Year

Table 4. Average Stock Index per Cluster, 2018

Cluster	Member	PER	PBV	PSR
1	44	7.692	1.320	1.407
2	20	28.163	3.409	3.123
3	1	11.2	108.633	66.384
4	1	112.818	2.390	2.621

In 2018, Table 4 shows that the first cluster group was the cluster containing the best companies because all the valuation variables (PER, PBV and PSR) in it had the lowest value. This shows that cluster 1 consists of undervalued companies. The second cluster is still classified as a good company but PER, PBV and PSR get the second highest value but only PER has a value quite far from the lowest PER value which means the companies in this cluster are still classified as in value. Clusters 3 and 4 only have 1 share in the cluster so further cluster analysis cannot be carried out.

Table 5. Average Stock Index per Cluster, 2019

Cluster	Member	PER	PBV	PSR
1	4	133.371	1.704	1.753
2	60	15.428	2.861	2.375
3	1	28.850	187.882	107.708
4	1	562.602	1.628	2.011

In 2019, Table 5 shows that the first cluster had a high PER value with a value of 133 which made the shares of companies in this group appear quite overvalued even though the PBV and PSR were quite low. Cluster 2 can be said to be a cluster containing undervalued companies based on all valuation variables, even though the PBV and PSR values are not the lowest but are still in the low category. Clusters 3 and 4 only have 1 share in the cluster so further cluster analysis cannot be carried out. Cluster 4 has a very high PER, which means that if you look at the profit obtained by the issuer, the price of the shares being traded is 562 higher than the 1-year profit of the shares.

Table 6. Average Stock Index per Cluster, 2020

Cluster	Member	PER	PBV	PSR
1	57	17.059	2.893	2.897
2	2	210.209	1.729	1.825
3	1	24.756	149.891	71.029
4	6	80.935	1.882	3.695

In 2020, it can be seen from Table 6 that the first cluster can be categorized as the cluster with the best stocks because even though PBV and PSR are not the lowest, they are still in the stable category and have the lowest PER value. Clusters 2 and 3 only have 2 and 1 shares in the cluster so further analysis cannot be carried out. Cluster 4 can be said to be a cluster with in value stocks because all PER PBV and PSR are not at the highest or lowest positions. 2020 is where the pandemic has just hit the world and caused the economy to slow down or even decline, so if you pay attention to the valuation variables in 2020 it tends to be higher than other years.

Table 7. Average Stock Index per Cluster, 2021

Cluster	Member	PER	PBV	PSR
1	59	13.965	2.405	2.375
2	5	51.080	2.648	2.998
3	1	80.563	670.653	232.709
4	1	170.196	7.921	16.203

In the 2021, Table 7 shows that cluster 1 has the best PER, PBV and PSR values, namely having the lowest values compared to other clusters. Cluster 2 has quite good PER, PBV and PSR values and can be categorized as in value. If you look closely, the PER of cluster 2 is quite high although it is still within reasonable limits when compared to other clusters, PBV and PSR have the second lowest values so that the company shares can be categorized as quite cheap. Clusters 3 and 4 cannot be analysed further because there is only 1 share in this cluster so it cannot be categorized as a cluster.

Table 8. Average Stock Index per Cluster, 2022

Cluster	Member	PER	PBV	PSR
1	63	17.890	2.928	2.643
2	1	26.733	111.862	47.085
3	1	2410.371	1.212	1.523
4	1	74.915	4.774	59.051

In 2022, Table 8, cluster 1 has the best undervalued valuation because it has the lowest PER, PBV and PSR values. This means that cluster 1 contains shares of cheap companies that are worth buying. Clusters 2, 3 and 4 only have 1 member so in 2022 there will only be 1 cluster so it cannot be analysed further clusters like the previous cluster which had 2 clusters which can analyse comparisons between clusters in the same year. The clustering also reveals that concerning the stability of companies which their stocks are worth buying over the years, there are 52 companies where their valuation are high. This result come based on the cluster member where their clusters are the best valuation in 2018-2022.

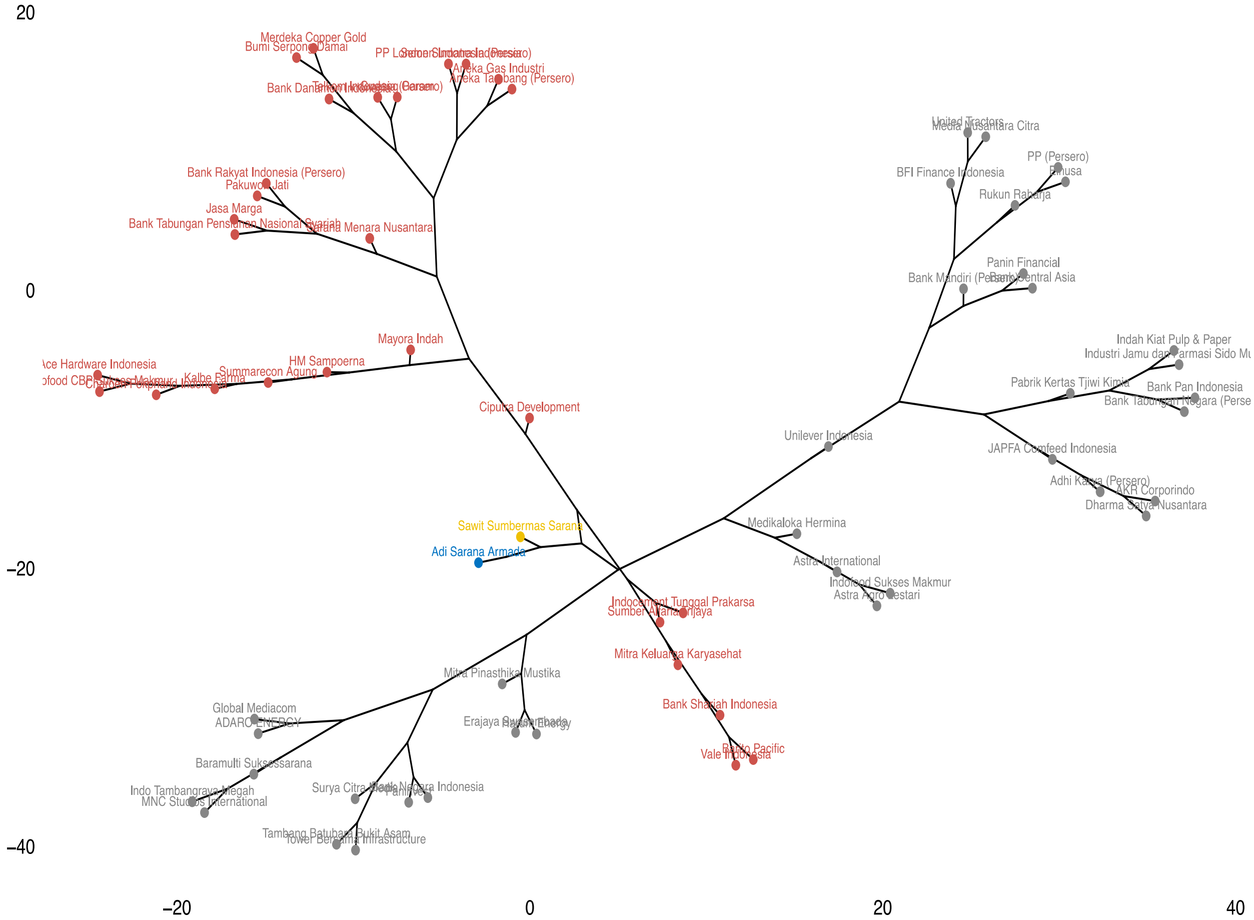

Figure 1. Dendrogram Cluster of Kompas 100 Stocks Index in 2018
(Source: RStudio Output)

The visualization of cluster for each year from 2018 till 2022, can be seen in Figs, 1 - 5. Clusters 1- 4 are represented by red, grey, blue and yellow colors respectively.

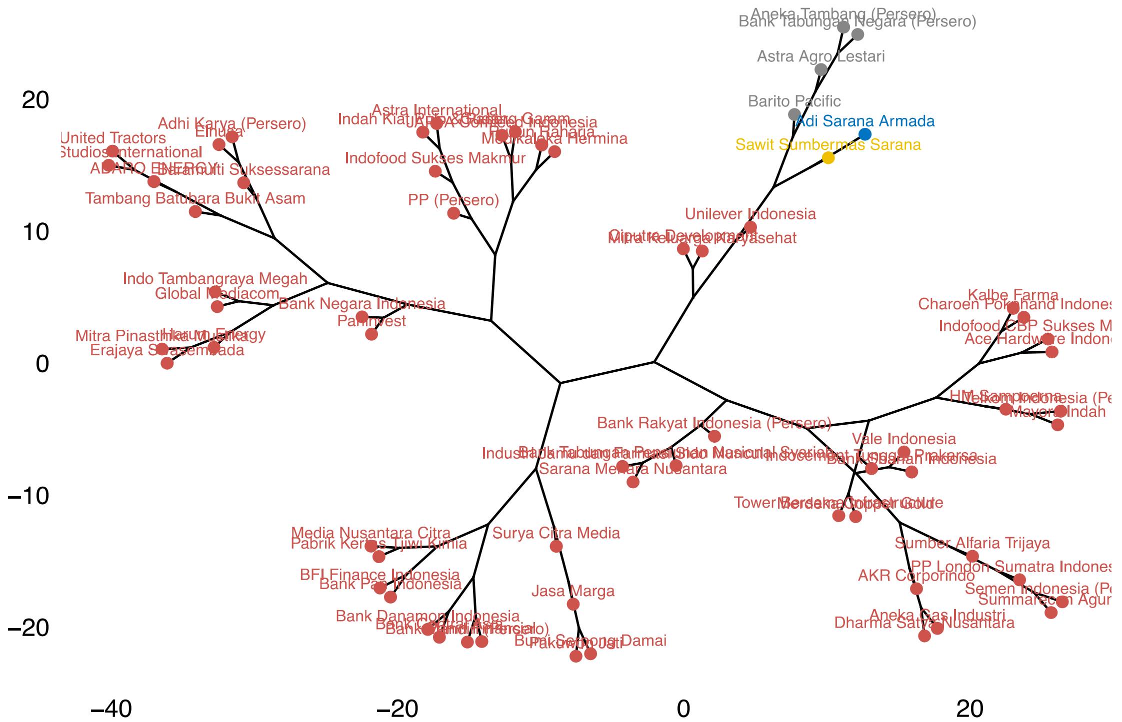

Figure 2. Dendrogram Cluster of Kompas 100 Stocks Index in 2019
(Source: RStudio Output)

Before the COVID-19, in 2028-2019, the Adi Sarana Armada and Sawit Sumbermas Sarana independently are in different clusters and a solely member (Figs. 1 and 2). This means that both of them are behaving differently in comparison to other companies and being consistent.

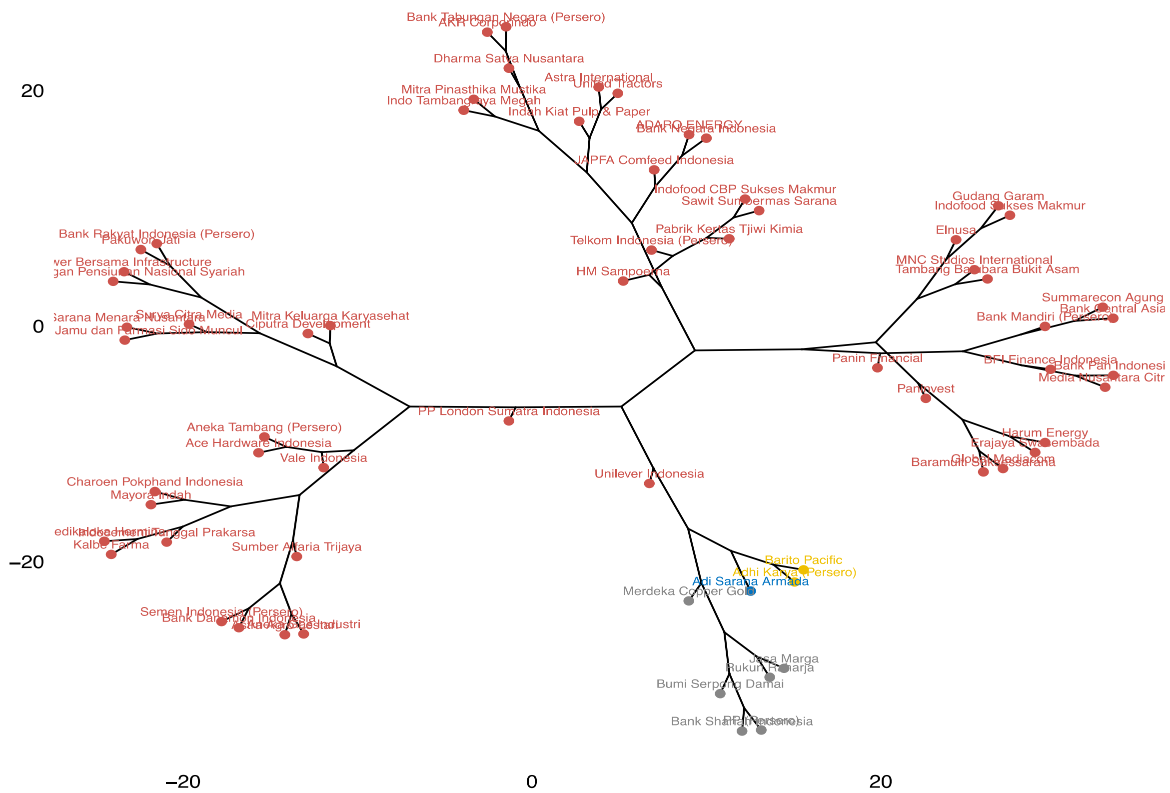

Figure 3. Dendrogram Cluster of Kompas 100 Stocks Index in 2020
(Source: RStudio Output)

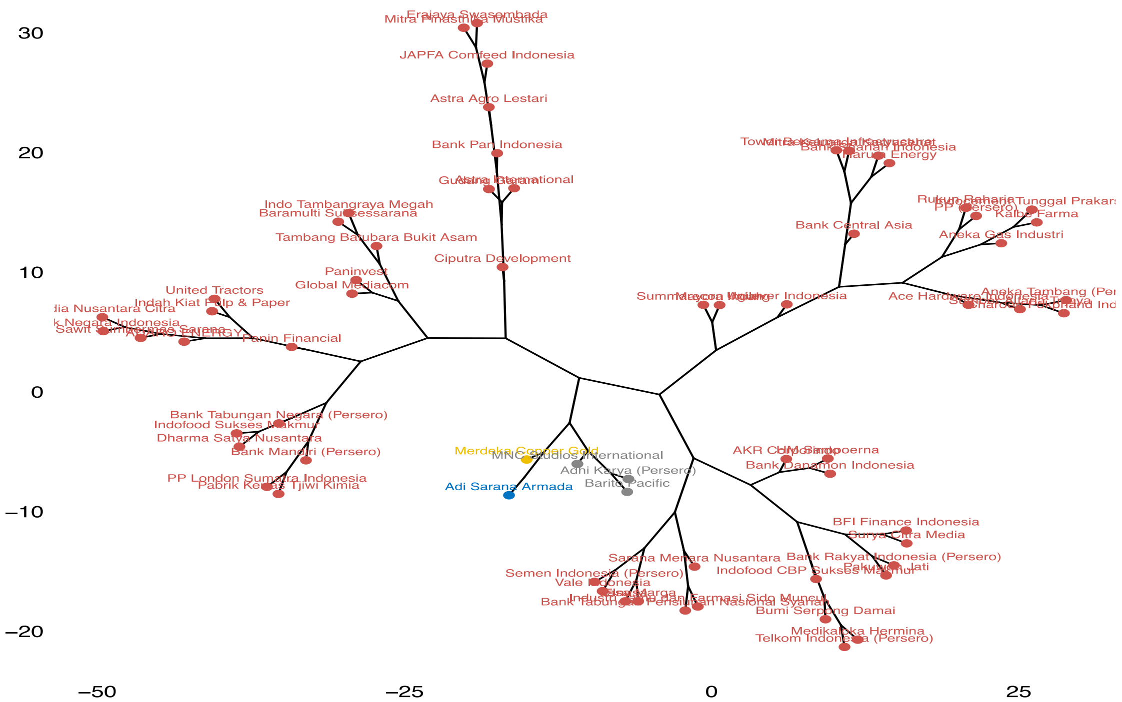

Figure 4. Dendrogram Cluster of Kompas 100 Stocks Index in 2021
(Source: RStudio Output)

During the COVID-19, in 2020-2021, it was only Adi Sarana Armada and two others (Barito Pacific and Adhi Karya in 2020 and Merdeka Copper Gold in 2021), in different clusters, which have different performances. Sawit Sumbermas Sarana moved to another cluster both in 2020 and 2021. See Figure 3 and Figure 4.

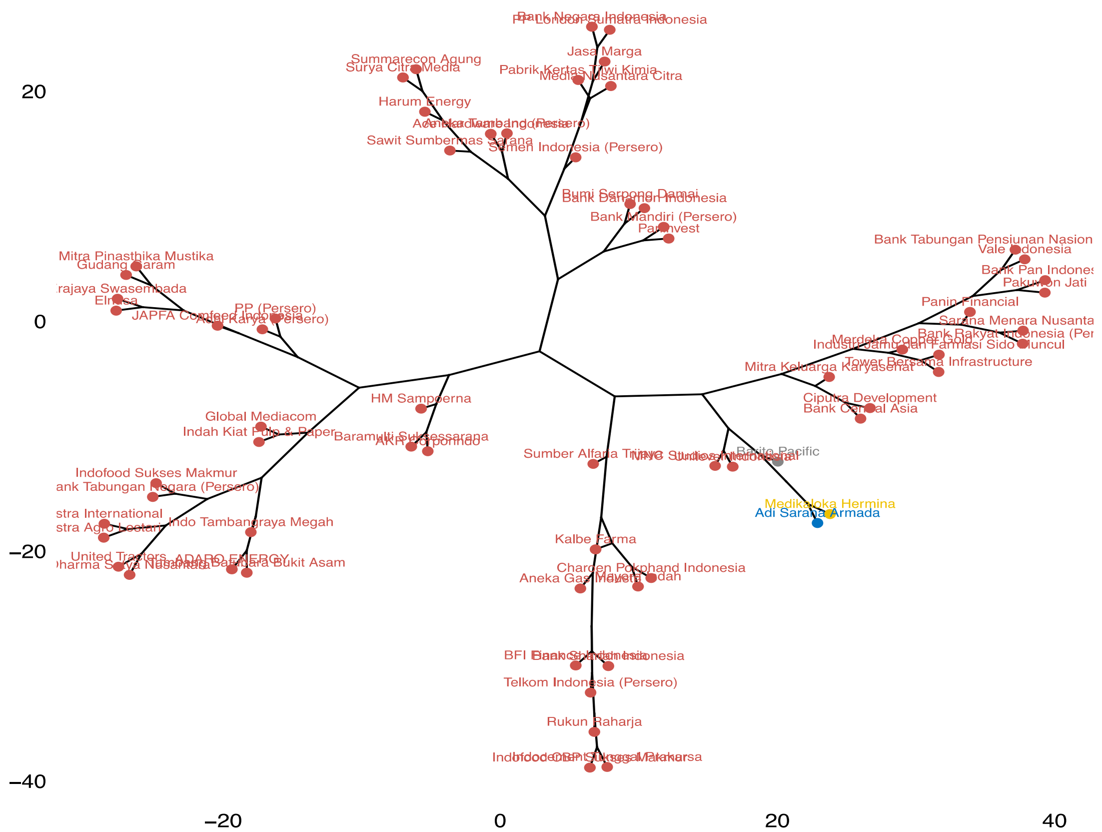

Figure 5. Dendrogram Cluster of Kompas 100 Stocks Index in 2022
(Source: RStudio Output)

After the COVID-19, in 2022, the Sawit Sumbermas Sarana still in a majority cluster together with Adhi Karya and Merdeka Copper Gold, but Adi Sarana Armada behaves the same as the previous years. In 2022, Medikaloka Hermina and Barito Pacific have similar performance as Adi Sarana Armada.

The dendrogram clusters in Figures 2-6 revealed which companies' stock are possibly become together in the same cluster. These results can be used by a buyer as a consideration which company stock should be buy in future. Furthermore, in further studies on how to form an optimal portfolio among these companies.

3.1. 2 Portfolio Analysis

After analyzing clusters and dividing them into 4 different clusters, the researcher will obtain the best cluster that will form the basis of which stocks to include in the portfolio. This portfolio is created under the assumption of equal weights among the stocks, as shown in Table 9. Next, the researcher will compare the performance of the portfolio formed from the best cluster with the best real-world portfolios in Indonesia and the world. In Indonesia, it will be compared with the Indonesia Composite Index (IDX), which includes all stocks in Indonesia, Liquid 45 (LQ45), which consists of the 45 most liquid stocks on the Indonesia Stock Exchange, Kompas 100, which includes the 100 best stocks based on liquidity and market capitalization, and SAM Equity, which is the best mutual fund in Indonesia in 2023.

Globally, the comparison will be made with the Vanguard Total Stock Market Index Fund (VSMPX), which is one of the best mutual funds in the world, and the Fidelity 500 Index Fund (FXAIX), which is one of the best index mutual funds in the world. These two mutual funds will serve as global benchmarks in this research.

Table 9. Portfolio Analysis

	Cluster	IDX	Kompas 100	LQ45	SAM Equity	VSMPX	FXAIX
2018	10.94%	-2.54%	-6.26%	-8.96%	-9.85%	-6.93%	-6.80%
2019	9.99%	1.70%	1.51%	3.23%	3.37%	28.32%	28.61%
2020	16.90%	-5.09%	-5.56%	-7.84%	-10.70%	18.89%	16.20%
2021	34.29%	10.08%	-3.40%	-0.37%	-4.19%	24.09%	27.00%
2022	21.83%	4.09%	-0.86%	0.62%	4.07%	-20.80%	-19.48%

In 2018, the first cluster, consisting of 44 stocks, was formed into a portfolio that generated a return of 10.94%. This portfolio's performance far exceeded the benchmark portfolios in Indonesia, which returned only -2.54% (IDX), -6.26% (Kompas 100), -8.96% (LQ45), and -9.85% (SAM Equity). Globally, the portfolio from this cluster also outperformed, surpassing VSMPX with a return of -6.93% and FXAIX with a return of -6.80%. In 2019, the second cluster, consisting of 60 stocks, was formed into a portfolio that generated a return of 9.99%. This portfolio's performance was better than the portfolios in Indonesia, which returned only 1.70% (IDX), 1.51% (Kompas 100), 3.23% (LQ45), and 3.37% (SAM Equity). However, the portfolio's performance this year was still lower than the global portfolios, with VSMPX returning 28.32% and FXAIX returning 28.51%.

In 2020, the first cluster, consisting of 57 stocks, was formed into a portfolio that generated a return of 16.90%. This portfolio's performance was better than the portfolios in Indonesia, which returned only -5.09% (IDX), -5.56% (Kompas 100), -7.84% (LQ45), and -10.70% (SAM Equity). However, the portfolio's performance appeared modest compared to the global portfolios, with VSMPX performing higher at 18.89% and FXAIX slightly lower at 16.20%. In 2021, the first cluster, consisting of 59 stocks, was formed into a portfolio that generated a return of 34.29%. This portfolio's performance far exceeded the benchmark portfolios in Indonesia, which returned only 10.08% (IDX), -3.40% (Kompas 100), -0.37% (LQ45), and -4.19% (SAM Equity). The performance also surpassed the global portfolios, with VSMPX returning 24.09% and FXAIX returning 27%.

In 2022, the first cluster, consisting of 63 stocks, was formed into a portfolio that generated a return of 21.83%. This portfolio's performance far exceeded the benchmark portfolios in Indonesia, which returned only 4.09% (IDX), -0.86% (Kompas 100), 0.62% (LQ45), and 4.07% (SAM Equity). The performance also significantly outpaced the global portfolios, with VSMPX returning -20.80% and FXAIX returning -19.48%.

The strong outperformance of the clustering-based portfolios in this study aligns closely with recent empirical findings in quantitative finance. For instance, Bolos et al. [37] demonstrated through k-means clustering on profitability, liquidity, and solvency metrics that equally-weighted cluster portfolios not only achieved higher returns but also offered superior risk-adjusted performance compared to traditional mean–

variance optimized portfolios Their approach echoes our use of valuation ratios (PER, PBV, PSR) in forming fundamentally driven clusters that consistently outperform conventional benchmarks. Additionally, Mattera et al. [38] investigated time-series clustering in high-dimensional asset selection and found that clustering offers significant out-of-sample gains over standard allocation methods, particularly in reducing estimation errors and enhancing Sharpe ratios. These studies reinforce our results; our clustering method not only captures undervalued stocks (as shown by lower valuation ratios) but also delivers tangible performance benefits across both return and risk dimensions in an emerging market context.

4. CONCLUSION

The results of clustering the shares that are members of the Kompas 100 in 2018-2022 show quite satisfactory results, researchers can divide shares into 4 clusters that have different characteristics and can find clusters that have similar shares. In 2018, cluster 1 was the best cluster with 44 shares included in it. In 2019, cluster 2 was the best cluster with 60 shares in the cluster. In 2020-2022, cluster 1 is the best cluster with 57 shares in 2020, 59 in 2021 and 63 in 2022. This cluster can be categorized as containing shares of the best companies in the year concerned, taken from the Kompas 100 stocks index.

In general, the formation of portfolios using the best cluster demonstrates very promising performance. Over the five-year testing period, in three years (2018, 2021, and 2022), the portfolio outperformed all other portfolios in Indonesia and even globally. Although the portfolio's performance in 2019 and 2020 was slightly lower than the global portfolios, it still managed to outperform the Indonesian portfolios.

The researcher believes that forming portfolios using clustering techniques is highly recommended, as it has proven to deliver excellent performance and is more cost-effective compared to traditional portfolio formation methods. Table 9 shows that the cluster-based portfolio consistently outperformed returns of IDX, Kompas 100, LQ45, and global benchmarks such as SAM Equity, VSMPX, and FXAIX in 2018, 2021, and 2022. It also achieved the highest return in 2020 despite market downturns in IDX and Kompas 100 but underperformed in 2019 compared to global indices such as VSMPX and FXAIX. Future research could include assessing the risk of each portfolio formation. This would allow for a comprehensive evaluation of not only the return performance but also the risk, providing a more complete basis for decision-making.

Author Contributions

Rohmatul Fajriyah: Conceptualization, Methodology, Writing-Original Draft, Validation. Yoel Christopher Tjen: Supervision, Methodology, Data Curation, Computational and Formal Analysis, Validation, Writing-Review and Editing. All authors discussed the results and contributed to the final manuscript.

Funding Statement

This research received financial support for publication from Universitas Islam Indonesia, Yogyakarta, Indonesia

Acknowledgment

The authors would like to express their deepest gratitude to Universitas Islam Indonesia, for providing financial support for this research. The author would also like to thank the support of the Master Program in Statistics of the Universitas Islam Indonesia, where this research was conducted

Declarations

The authors declare no competing interest.

REFERENCES

[1] Kustodian Sentral Efek Indonesia (KSEI), "STATISTIK PASAR MODAL INDONESIA", 2023. https://www.ksei.co.id/files/Statistik_Publik_Desember_2023_v3.pdf. Last accessed March 20, 2025.

[2] Z. Bodie, A. Kane, and A. J. Marcus. *INVESTMENTS* (11th ed.). New York: McGraw-Hill Education, 2020.

[3] MRA Ardana, WP Setiyono, and S Sriyono. "THE IMPACT OF FUNDAMENTALS AND TECHNICAL ANALYSIS ON STOCK RETURNS IN BANKING COMPANIES LISTED FOR INDONESIA STOCK EXCHANGE 2013-2022". *EKOMBIS REVIEW: Jurnal Ilmiah Ekonomi dan Bisnis*, vol. *12*(2), 2103-2118, 2024. doi: https://doi.org/10.37676/ekombis.v12i2.5505

[4] N. A. Lestari, A. Antony, and A. Alhidayatullah. "EVALUATION OF STOCK THROUGH FUNDAMENTAL ANALYSIS WITH THE DIVIDEND DISCOUNT MODEL (DDM) APPROACH". *International Journal of Indonesian Business Review*, vol. 2(1), 85-95, 2023. doi: https://doi.org/10.54099/ijibr.v2i1.499

[5] A. Masruroh, K. S. Rachmadanti, T. E. Tania, and M. Y. R. Pandin, "FUNDAMENTAL STOCK ASSESSMENT USING THE DIVIDEND DISCOUNT MODEL AND PRICE EARNING RATIO METHOD FOR INVESTMENT DECISIONS IN THE BANKING SECTOR LISTED ON THE INDONESIAN STOCK EXCHANGE". *Jurnal Manajemen Kreatif dan Inovasi*, vol. 2(1), 42-52, 2023. doi: https://doi.org/10.59581/jmki-widyakarya.v2i1.2065

[6] M. Skočir and I. Lončarski. "ON THE IMPORTANCE OF ASSET PRICING FACTORS IN THE RELATIVE VALUATION", *Research in International Business and Finance*, vol. 70, 2024. doi: https://doi.org/10.1016/j.ribaf.2024.102366

[7] A. Damodaran, *THE LITTLE BOOK OF VALUATION: HOW TO VALUE A COMPANY, PICK A STOCK, AND PROFIT*. New Jersey: John Wiley & Sons, 2024.

[8] F. Ferdinand, E. Hulu, G. Ugut, and R. Sembel. "VALUATIN OF STOCK VALUE WITH RELATIVE VALUATION AND ABSOLUTE VALUATION APPROACH IN BANKING SECTOR AT LQ 45 PERIOD 2015-2020". Ultima InfoSys : *Jurnal Ilmu Sistem Informasi*, vol. 15, no. 1, pp. 1-6, 2024. doi: https://doi.org/10.31937/si.v15i1.3410

[9] F. H. Christiansen and M. H. Søderholm. *RELATIVE VALUATION: MAN VS. MACHINE CAN AN ALGORITHM BEAT YOUR BROKER?*, Thesis, master of Science in Finance and Accounting, Denmark: Copenhagen Business School, 2020.

[10] B. K. Lukanima, "AN OVERVIEW OF RELATIVE VALUATION". In *Corporate Valuation A Practical Approach with Case Studies*, pp. 613-633, 2023. doi: https://doi.org/10.1007/978-3-031-28267-6_19

[11] T. Deng, Z. Ning, and K. Wan, "CORPORATE VALUE ASSESSMENT BASED ON THE EVA MODEL AND RELATIVE VALUATION", *Highlights in Business, Economics and Management*, vol. 40, pp. 1177-1183, Sep. 2024. doi: https://doi.org/10.54097/2ttasr58

[12] A. Saputra, R. Yustiana, and R. Hidayat, "ANALISIS PENGARUH PRICE TO SALES RATIO (PSR) DAN PRICE TO BOOK RATIO (PBR) TERHADAP RETURN SAHAM PADA PERUSAHAAN INDUSTRI BARANG KONSUMEN YANG TERDAFTAR DI BURSA EFEK INDONESIA PERIODE 2016-2020", *Jurnal Riset Ekonomi dan Manajemen*, vol. 8, no. 1, pp. 1-14, 2022. doi: https://doi.org/10.55606/jurnalrisetilmuakuntansi.v1i4.103

[13] T. Issakainen. "CLUSTERING-BASED VALUE INVESTING STRATEGY IN THE HELSINKI STOCK EXCHANGE: K-MEANS ALGORITHM". *Journal of Innovation in Business and Economics*, vol. 7, no. 2, pp. 115–128, 2023. doi: https://doi.org/10.22219/jibe.v7i02.26424

[14] B. Aslam, RA Bhuiyan, and C. Zhang. "PORTFOLIO CONSTRUCTION WITH K-MEANS CLUSTERING ALGORITHM BASED ON THREE FACTORS". *MATEC Web of Conferences: CGCHDRC 2022*, 2023. doi: https://doi.org/10.1051/matecconf/202337702006

[15] H.W. Zhang. "LIMITATIONS AND CRITIQUE OF MODERN PORTFOLIO THEORY: A COMPREHENSIVE LITERATURE REVIEW". *Advances in Economics, Management and Political Sciences*, vol. 60, pp. 24–29, 2024. doi: https://doi.org/10.54254/2754-1169/60/20231148

[16] S.E. Kemp, M. Ng, T. Hollowood, and J. Hort, *DESCRIPTIVE ANALYSIS IN SENSORY EVALUATION*, UK: John Wiley & Sons, 2018. doi: https://doi.org/10.1002/9781118991657

[17] N.T. Wulansari, E. Elwisam, and K. Digdowiseiso. "ANALYSIS OF THE INFLUENCE OF PRICE EARNING RATIO, PRICE TO BOOK VALUE, EARNING PER SHARE, DEBT TO EQUITY RATIO AND RETURN ON EQUITY ON SHARE PRICES IN LQ45 COMPANIES ON THE INDONESIA STOCK EXCHANGE 2016-2020". *Jurnal Syntax Admiration*, vol. 4, no. 2, pp. 332-348, 2023. doi: https://doi.org/10.46799/jsa.v4i2.883

[18] X. Chen, B. Li, and A.C. Worthington. "HIGHER MOMENTS AND US INDUSTRY RETURNS: REALIZED SKEWNESS AND KURTOSIS". *Review of Accounting and Finance,* vol. 20, no. 1, pp. 1–22, 2020. doi: https://doi.org/10.1108/RAF-06-2020-0171

[19] J. F. Hair, B. J. Babin, R. E. Anderson and W. C. Black, *MULTIVARIATE DATA ANALYSIS*, 8 ed. Cengage Learning, 2022.

[20] N. H. Spencer, *ESSENTIALS OF MULTIVARIATE DATA ANALYSIS* (1ST ED.), Florida: CRC Press, 2014. doi: https://doi.org/10.1201/b16344

[21] R. Fajriyah and A. Mahmudiati. *KESEHATAN MASYARAKAT INDONESIA 2013: SBRC SERIES ANALISIS DATA KESEHATAN 1.01*, Central Java: Eureka Media Aksara, 2023.

[22] C.R. Beidleman, "LIMITATIONS OF PRICE-EARNINGS RATIOS", *Financial Analysts Journal*, vol. 27, no. 5, pp. 86-91, 1971. doi: https://doi.org/10.2469/faj.v27.n5.86

[23] A. Gunawan, R. Fajriyah, M. Bimakasa, and S. N. Untari, "ANALYSIS OF PORTFOLIO FORMATION ON THE LQ45 STOCKS INDEX, USING THE MARKOWITZ AND SINGLE INDEX MODELS", *BAREKENG: J. Math. & App.*, vol. 18, no. 4, pp. 2363-2374, 2024. doi: https://doi.org/10.30598/barekengvol18iss4pp2363-2374

[24] L. Theron and G. Van Vuuren, "THE MAXIMUM DIVERSIFICATION INVESTMENT STRATEGY: A PORTFOLIO PERFORMANCE COMPARISON", *Cogent Economics & Finance*, vol. 6, no. 1, 2018. doi: https://doi.org/10.1080/23322039.2018.1427533

[25] R. Sary, N. Satyahadewi, and W. Andani, "APPLICATION OF K-MEANS++ WITH DUNN INDEX VALIDATION OF GROUPING WEST KALIMANTAN REGION BASED ON CRIME VULNERABILITY", *BAREKENG: J. Math. & App.*, vol. 18, no. 4, pp. 2283-2292, Oct. 2024.

[26] J. Fu, Q. Zhou, Y. Liu, and X. Wu, "PREDICTING STOCK MARKET CRISES USING DAILY STOCK MARKET VALUATION AND INVESTOR SENTIMENT INDICATORS". *The North American Journal of Economics and Finance*, vol. 51, 2020. doi: https://doi.org/10.1016/j.najef.2019.01.002

[27] D.S. Bates, "EMPIRICAL OPTION PRICING MODELS", Annual review of financial economics, vol. 14, no. 1, pp. 369-389, 2022. doi: https://doi.org/10.1146/annurev-financial-111720-091255

[28] M. Lof and H. Nyberg, "DISCOUNT RATES AND CASH FLOWS: A LOCAL PROJECTION APPROACH", *Journal of Banking and Finance*, vol. 162, 2024. doi: https://doi.org/10.1016/j.jbankfin.2024.107127

[29] B. Cornell and R. Gerger, "LET'S GET REAL ABOUT THE DIVIDEND DISCOUNT". 2022. Last access March 20, 2025. Available at SSRN. doi: https://doi.org/10.2139/ssrn.4015531

[30] R. Yi, Y. W. Chang, W. Xing, and J. Chen, "COMPARING RELATIVE VALUATION EFFICIENCY BETWEEN TWO STOCK MARKETS", *The Quarterly Review of Economics and Finance*, vol. 72, pp. 159-167, 2019. doi: https://doi.org/10.1016/j.qref.2018.11.008

[31] E. F. Brigham and M. C. Ehrhardt. FINANCIAL MANAGEMENT: THEORY AND PRACTICE (15TH ED.). New York: Cengage Learning, 2018.

[32] M. Neldi, H. Hady, Elfiswandi and Lusiana. "THE DETERMINANTS OF PRICE EARNING RATIO: EVIDENCE FROM INDONESIA", *Journal of Law and Sustainable Development*, vol. 11, no. 4, 2023. doi: https://doi.org/10.55908/sdgs.v11i4.1003

[33] F. Saputra, "ANALYSIS EFFECT RETURN ON ASSETS (ROA), RETURN ON EQUITY (ROE) AND PRICE EARNING RATIO (PER) ON STOCK PRICES OF COAL COMPANIES IN THE INDONESIA STOCK EXCHANGE (IDX) PERIOD 2018-2021". *Dinasti International Journal of Economics, Finance & Accounting*, vol. 3, no. 1, pp. 82-94, 2022. doi: https://doi.org/10.38035/dijefa.v3i1.1238

[34] MNP Waworuntu and A. Lelengboto "RELEVANSI PRICE EARNING RATIO DALAM ANALISA KERJA PERUSAHAAN", *Klabat Accounting Review*, vol 1, no. 1, pp. 71-80, 2020. doi: https://doi.org/10.31154/kar.v1i1.458.71-80

[35] M. Ikhsan, S. Jumono, A. Munandar and A. Abdurrahman, "THE EFFECT OF NON PERFORMING LOAN (NPL), INDEPENDENT COMMISSIONER (KMI), AND CAPITAL ADEQUACY RATIO (CAR) ON FIRM VALUE (PBV) MEDIATED BY RETURN ON ASSET (ROA)", *Quantitative Economics and Management Studies*, vol. 3, no. 5, pp. 810-824, 2022. doi: https://doi.org/10.35877/454RI.qems1063

[36] Q. Xu, "THE APPLICATION OF PRICE-SALES RATIO METHOD IN S COMPANY'S MERGER AND ACQUISITION OF F COMPANY", *BCP Business & Management*, vol. 26, pp. 1222-1227, 2022. doi: https://doi.org/10.54691/bcpbm.v26i.2123

[37] G.I. Tanius and R. Widjojo, "THE INFLUENCE OF PRICE/EARNING RATIO (PER), PRICE TO BOOK RATIO (PBR) AND PRICE TO SALES RATIO (PSR) ON NET ASSET VALUE (NAV) IN SHARIA EQUITY FUND", *Jurnal Manajemen dan Bisnis Indonesia*, vol. 6, no. 1, pp. 91-102, 2018. doi: https://doi.org/10.31843/jmbi.v6i1.185

[38] M. Bolos, S. Rusu, M. Leordeanu, C.D. Sabău-Popa, D.C. Perțicas and M. Crisan. "K-MEANS CLUSTERING FOR PORTFOLIO OPTIMIZATION: SYMMETRY IN RISK–RETURN TRADEOFF, LIQUIDITY, PROFITABILITY, AND SOLVENCY", *Symmetry*, vol. 17, no. 6, p. 847, 2025. doi: https://doi.org/10.3390/sym17060847

[39] R. Mattera, G. Scepi, and P. Kaur, P. "TIME SERIES CLUSTERING FOR HIGH-DIMENSIONAL PORTFOLIO SELECTION: A COMPARATIVE STUDY", *Soft Comput* vol. 29, pp. 4219–4231, 2025. doi: https://doi.org/10.1007/s00500-025-10656-2