Contents lists available at GrowingScience

Decision Science Letters

homepage: www.GrowingScience.com/dsl

Decision-making in formation of mean-VaR optimal portfolio by selecting stocks using K-means and average linkage clustering

Ahmad Fawaid Ridwan[a*], Herlina Napitupulu[b] and Sukono[b]

[a]*Master's Program of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, West Java, Indonesia*
[b]*Depertment of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, West Java, Indonesia*

C H R O N I C L E

Article history:
Received April 18, 2022
Received in revised format: June 18, 2022
Accepted July 3 2022
Available online July 3, 2022

Keywords:
Average Linkage
K-Means
Clustering
Investment Portfolio
Mean-VaR

A B S T R A C T

Stock is one of the investment assets that has its charm for investors. It is very liquid and has a high rate of return, but it has a high risk. The strategy commonly used to minimize investment risk is to diversify through portfolio formation. A good allocation of funds must be determined in forming an optimal portfolio. In addition, the method of stock selection needs to be considered so the stocks are well diversified and the portfolio developed has good performance. This study aims to compare stock selection between K-Means and Average Linkage clustering approaches in forming an investment portfolio. Clustering analysis is used to group IDX80 stocks based on their attributes. In forming a portfolio with the Mean-VaR model, the stock selection decision criteria used are by selecting stocks with the highest positive returns from each cluster. As a result, the two clustering techniques show the superiority of the Silhouette score for a certain number of clusters, but there are still more advantages in Average Linkage. The portfolio approached by Average Linkage resulted in a better performance than the portfolio approached by K-Means. Therefore, Average Linkage clustering can be used as a better recommendation in decision-making to select stocks so as to produce optimal portfolio performance.

1. Introduction

Investment is an activity of placing funds in one or more types of assets during a certain period in the hope of obtaining income or increasing the value of funds in the future (Hidayati, 2017). The decision to invest starts from the determination of the number of assets that must be owned by the company to the consideration of the risks that will be faced when making an investment and the return that will be received when making an investment (Safelia, 2012). One of the investment assets that is the main attraction for investors is stocks. The advantage of stock instruments compared to other instruments is that stocks are very liquid, meaning that stockholders can trade them easily on the stock exchange (Setyawati, 2011). In proportion to the potential for high returns, the risk of investing in stocks is also high. The strategy that investors usually use to minimize risk is to diversify through the formation of an investment portfolio. The purpose of this diversification is to spread investment in several assets so that the risk to be received is also diversified and can be minimized (Sulistiyowati & Santoso, 2017).

The problem in forming a portfolio is how to decide on the right composition for each asset so that it can minimize risk and maximize return. One model that is often used in portfolio optimization is the Mean-Variance (MV) model (Markowitz, 1952). The MV model in portfolio formation is increasingly diverse by modifying and adding or changing the statistical measures used. One of the developments of the MV model is the Mean-Value at Risk (Mean-VaR) model. The formation of the Mean-VaR portfolio model is carried out by utilizing the stock's Value at Risk (VaR) and expected return as a measure

* Corresponding author.
E-mail address: ahmad20034@mail.unpad.ac.id (A. F. Ridwan)

doi: 10.5267/dsl.2022.7.002

used in the component of portfolio risk measurement. VaR is a market risk calculation method to determine the maximum risk of loss that can occur in a portfolio, either single-instrument or multi-instruments, at a certain level of confidence, during a certain holding period, and under normal market conditions (Ismanto, 2016).

The problem in stock investment portfolios is not only focused on optimization modeling but also on how to decide the combination of the stock so that they can be well-diversified. In terms of selecting the combined assets, the more diverse the assets in the portfolio, the more diversified it can be so that it is considered capable of reducing the risks that occur in the portfolio (Subekti et al., 2017). Clustering analysis is a multivariate technique that has the main objective of grouping objects based on their characteristics. The clustering technique is one of the unsupervised data mining methods, meaning that this method is applied without training and teaching and does not require a target output (Utami et al., 2019). There are two types of clustering analysis, namely hierarchical and non-hierarchical. In non-hierarchical methods such as K-Means (KM) which is a grouping of objects by determining the number of groups that will be formed first (Jiang et al., 2014). The purpose of this grouping is to reduce diversity within a group and maximize diversity between groups (Cebeci & Yildiz, 2015). The advantages of the KM algorithm are that it can handle large data, and cluster members can be adjusted, but this technique has drawbacks which are sensitive to outliers, sensitive to data scale, and ineffective for various clusters (Widyadhana et al., 2021).

In contrast to hierarchical clustering, the number of groups is not predetermined. Grouping in hierarchical clustering has two approaches, it is agglomerative and divisive. Some methods that are often used to group objects based on the size of their resemblance in the agglomerative method include Single Linkage, Complete Linkage, Average Linkage, and Ward. The Average Linkage (AL) algorithm was chosen because it has advantages, such as grouping objects based on the average distance between all objects in a cluster with all objects in other clusters, producing a dendrogram that provides a graphic description, and can detect various shapes and sizes of clusters (Yusniyanti et al., 2021). AL clustering starts with a single object as a cluster and moves to converge to form a new cluster according to the similarity of the characteristics of the object so that a single cluster is formed (Bhattacharjee et al., 2019). However, Average Linkage has drawbacks, namely, it is difficult to determine which clusters are considered insignificant, adjustments cannot be made after the cluster is formed, and clusters depend on the distance metric used (Xu et al., 2021).

There are previous studies that discuss clustering techniques using KM and AL in the formation of investment portfolios. Arimarista (2017) examines the expected return and risk of LQ-45 stocks for investment decision-making and optimal portfolio formation. Based on the discussion in this study, LQ-45 stocks were analyzed using AL with expected return and risk indicators. Subekti et al. (2017) examine KM and AL in the formation of a Stock Portfolio. Based on the discussion of the study, the stocks that are included in the Jakarta Islamic Index are analyzed using KM and AL which no different validation result between those two clustering techniques. The results of the clustering are then used as the basis for selecting stocks in forming a portfolio using the MV model. Kumari et al. (2019) examine the formation of a portfolio based on Mean-VaR with K-Means clustering. The study explains that KM can reduce time efficiency in the selection of similar stocks that are grouped into a cluster and the best stocks from these groups can be selected for the formation of an optimal portfolio with the Mean-VaR model.

In the previous study that has been described, there are still several aspects that need to be developed, including the Arimarista (2017), although it has conducted a clustering analysis on the LQ-45 stock, in the formation of a portfolio using the Capital Asset Pricing Model it does not involve the results of the cluster analysis that has been done, so it does not explain the cluster analysis approach in the formation of the portfolio. In the study of Subekti et al. (2017), clustering using the KM and AL resulted in the optimal number of clusters being 2 clusters, but the allocated stocks were only selected from all members in one of the clusters. This shows that the selected stocks have not been well diversified because the stocks selected in one cluster have uniform characteristics. In addition, although the study of Arimarista (2017) and Subekti et al. (2017) have used cluster analysis on stocks used for portfolio formation, the portfolio model used does not use the Mean-VaR model. On the other hand, Kumari et al. (2019) despite using the Mean-VaR model and the KM clustering in the formation of an optimal portfolio, no comparison was made with the AL or other clustering techniques.

In this study, we intend to compare the stock selection strategy between KM and AL clustering to form a Mean-VaR model portfolio. The aim is to get the best clustering technique for deciding the stock combination used in the portfolio. The advantages of this study are the use of KM and AL clustering analysis with the expected return and VaR attributes of each stock, and the use of decision selection criteria by selecting one stock that has the largest positive expected return from each cluster formed. The approach with the clustering technique can accommodate stock data that have certain characteristics. Therefore, the K-Means and Average Linkage clustering technique approach are expected to be a consideration in making decisions to choose the stocks used in the formation of the Mean-VaR model portfolio to produce better portfolio performance.

2. Material and Methods

The object in this study is 80 stocks traded on the Indonesia Stock Exchange (IDX) and included in the IDX80 index in the period February 2022 to July 2022. The stocks on IDX80 were chosen because the price performance of these stocks has high liquidity and large market capitalization also is supported by good company fundamentals (IDX, 2021). The data used is the daily closing price of stocks for the period August 2, 2021, to January 31, 2022. This study was conducted by using the simulation method of the stock selection with KM and AL clustering approach for the formation of an investment portfolio with the Mean-VaR model. Furthermore, the simulation results are compared based on the level of clustering validation and portfolio performance which is built based on the Mean-VaR model. The simulation stages in this study can be seen in Fig. 1.

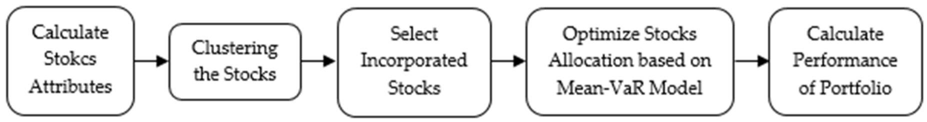

Fig. 1. Simulation Steps.

In carrying out the steps of simulating the formation of a portfolio based on the stock selection approach by the two clustering techniques, several theories and concepts used in this study are explained as follows.

2.1. Stock Investment

Investment in general is an activity of allocating a number of funds to certain assets, which aims to obtain increased funds in the future. Stock investment offers an advantage compared to other investment instruments in that stocks are liquid and offer a high rate of return. Stock return is the result obtained from investing that can be expressed as

$$R_i = \frac{S_j - S_{j-1}}{S_{j-1}}, \tag{1}$$

where R_i is stock return for period i, S_j is stock price for period j, and S_{j-1} is stock price for period $j-1$ (Chairunnisa et al., 2018). Based on equation (1) we can calculate the expected stock return with the equation of

$$E(R_i) = \frac{\sum_{j=1}^{n} R_{ij}}{n}, \tag{2}$$

where $E(R_i)$ is the expected return of the stock for period i, R_{ij} is the stock return from period i investment in period j, and n is the number of observations (Chairunnisa et al., 2018).

Stock investments with high returns have a high level of risk. In the calculation, stock risk is usually measured by the standard deviation of the return. Since the portfolio model used is Mean-VaR, then investment risk is measured using VaR. VaR is defined as the maximum loss rate at the confidence level and over a certain period of time which is obtained using the equation of

$$VaR = -(z_\alpha \sigma + \mu), \tag{3}$$

where z_α is quantile $1-\alpha$ of the return distribution with α being the significant level, σ is standard deviation of stock return, and μ is expected return of stock (Sukono et al., 2019).

2.2. K-Means Clustering

KM is one of the partitional clustering techniques because it is based on determining the initial number of groups by defining the initial centroid value (Madhulatha, 2012). The KM algorithm uses an iterative process to obtain a cluster database. KM will produce a center point that is continuously updated in each iteration (Ediyanto et al., 2013). After the KM iteration stops, each associated object in the dataset becomes a member of a cluster. KM clustering is conducted with the following algorithm.

1. Determine k as the number of clusters formed.
2. Determine the initial k centroids at random from the objects.
3. Calculate the distance of each object to each centroid of each cluster. The distance is being calculated using the Euclidian distance.

$$d(x.y) = \|x - y\| = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}\ ; \quad i = 1.2.3. \dots .n, \tag{4}$$

   where x_i is the i-th attribute of object x, y_i is the i-th attribute of object y, and n is the number of objects (Cebeci & Yildiz, 2015).
4. Allocate each object to the nearest centroid.
5. Determining the position of the new centroid using the following equation.

$$v = \frac{\sum_{i=1}^{n} x_i}{n}. \quad i = 1.2.3. \dots .n, \tag{5}$$

   where v is cluster centroid, x_i is i-th object, and n is the number of objects (Naeem & Wumaier, 2018).

6. Repeat steps 3 to 5 if the new centroid position is not the same.

2.3. Average Linkage Clustering

The AL method is a clustering method with the principle of the average distance between each possible pair of objects in one cluster and all objects in another cluster. AL calculates the distance between two clusters which is referred to as the average distance where the distance is calculated for each cluster. AL clustering is conducted with the following algorithm (Abdurrahman, 2019).

1. Defines the distance matrix of the objects used, it is defined as $\mathbf{D} = [d_{ik}]$ where d_{ik} is the Euclidean distance between object i and object k.
2. Merge objects that have the shortest distance to form a new cluster.
3. Calculate the distance between clusters that have been merged with other clusters with the equation of

$$d(UV)W = \frac{\sum_i \sum_k d_{ik}}{N_{UV} N_W}, \tag{6}$$

   where d_{ik} is the distance between the i-th object in the UV cluster and the k-th object in the W cluster, N_{UV} is the number of objects in the UV cluster, N_W is the number of objects in the W cluster (Eliguzel & Ozceylan, 2019).

4. Repeat steps 2 and 3 to form a single cluster containing all objects as members.

2.4. Silhouette Score

The Silhouette score is a measure of the uniformity of each cluster and how well the clusters are separated. This value is also a useful indicator for validating the suitability for a given clustering solution as it can be used to compare clustering solutions quantitatively. Silhouette score (s) is obtained based on the average distance between objects and has a value that varies in the range -1 to 1 where the higher the resulting value, the better the results of the clustering (Corporal-Lodangco et al., 2014). Silhouette Score is defined as

$$s(k) = \frac{b(k) - a(k)}{max\{a(k).b(k)\}}, \tag{7}$$

where $s(k)$ is the Silhouette score, $a(k)$ is the average distance between objects within the cluster, $b(k)$ is the average distance between objects over the clusters and k is number of cluster (Bienvenido-Huertas et al., 2021).

2.5. Mean-VaR Investment Portfolio Model

Investors in terms of forming a portfolio must determine the portfolio weight of each stock. If w_i are the proportion of funds to be allocated to the i-th stock, then all funds invested are 100%, the assumption can be symbolized as

$$\sum_{i=1}^{n} w_i = \mathbf{e}^{\mathbf{T}}\mathbf{w} = 1, \tag{8}$$

where $\mathbf{w}$ is the vector of the proportion of funds from stocks or can be written $\mathbf{w}^{\mathbf{T}} = [w_1 . w_2 . , , , . w_n]$, $\mathbf{e}^{\mathbf{T}}$ is a vector whose all entries are 1, or can be written $\mathbf{e}^{\mathbf{T}} = [1.1.1. \ldots .1]$ as many as n, where n is the number of stocks (Sukono et al., 2017).

From that assumptions, the return (R_p), expected return (μ_p), and variance (σ_p^2) of portfolio are determined as

$$R_p = \mathbf{w}^{\mathbf{T}}\mathbf{R}, \tag{9}$$

$$\mu_p = E(R_p) = \boldsymbol{\mu}^{\mathbf{T}}\mathbf{w}, \tag{10}$$

$$\sigma_p^2 = Var(R_p) = \mathbf{w}^{\mathbf{T}}\boldsymbol{\Sigma}\mathbf{w}, \tag{11}$$

where $\mathbf{R}$ is a vector of returns, $\boldsymbol{\mu}$ is a vector of expected returns and $\boldsymbol{\Sigma}$ is a covariance matrix of stock returns (Sukono et al., 2017). Then, VaR of portfolio can be written as

$$VaR_p = -\left(z_\alpha (\mathbf{w}^{\mathbf{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}} + \boldsymbol{\mu}^{\mathbf{T}}\mathbf{w}\right). \tag{12}$$

To obtain an efficient portfolio with $\tau \geq 0$, the weight of portfolio are determined by an optimization process (Sukono et al., 2017). The objective function is

$$\begin{aligned} &\text{maximize: } 2\tau\boldsymbol{\mu}^{\mathbf{T}}\mathbf{w} + z_\alpha (\mathbf{w}^{\mathbf{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}} + \boldsymbol{\mu}^{\mathbf{T}}\mathbf{w} \\ &\text{subject to: } \mathbf{e}^{\mathbf{T}}\mathbf{w} = 1. \end{aligned} \tag{13}$$

where τ indicating risk tolerance which is the level of a risk to be accepted by investors (Sukono et al., 2017). The problem is an optimization with constraints, so to find the optimal weight of Eq. (13) it can be defined by the Lagrange function as

$$L(\mathbf{w}.\lambda) = (2\tau\boldsymbol{\mu}^{\mathbf{T}}\mathbf{w} + z_{\alpha}(\mathbf{w}^{\mathbf{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}} + \boldsymbol{\mu}^{\mathbf{T}}\mathbf{w}) + \lambda(\mathbf{w}^{\mathrm{T}}\mathbf{e} - 1). \tag{14}$$

where λ is a Lagrange multiplier, so we define the conditions that need to be optimized with the first derivative of the Lagrange function with respect to $\mathbf{w}$ and λ as

$$\frac{\partial L}{\partial \mathbf{w}} = (2\tau + 1)\boldsymbol{\mu} + \frac{z_{\alpha}\boldsymbol{\Sigma}\mathbf{w}}{(\mathbf{w}^{\mathbf{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}}} + \lambda\mathbf{e} = 0 \tag{15}$$

and

$$\frac{\partial L}{\partial \lambda} = \mathbf{e}^{\mathrm{T}}\mathbf{w} - 1 = 0. \tag{16}$$

From Eq. (16) it can be obtained

$$\frac{z_{\alpha}\boldsymbol{\Sigma}\mathbf{w}}{(\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}}} = -\big((2\tau + 1)\boldsymbol{\mu} + \lambda\mathbf{e}\big). \tag{17}$$

If (17) is multiplied by $\boldsymbol{\Sigma}^{-1}$, then we get

$$\frac{z_{\alpha}\mathbf{w}}{(\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}}} = -\Big((2\tau + 1)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \lambda\boldsymbol{\Sigma}^{-1}\mathbf{e}\Big). \tag{18}$$

Furthermore, if equation (18) is multiplied by $\mathbf{e}^{\mathbf{T}}$, then we get

$$\frac{z_{\alpha}\mathbf{e}^{\mathrm{T}}\mathbf{w}}{(\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}}} = -\Big((2\tau + 1)\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \lambda\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}\Big). \tag{19}$$

Based on Eq. (19) it is known that $\mathbf{e}^{\mathbf{T}}\mathbf{w} = 1$, so that

$$(\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}} = \frac{z_{\alpha}}{-\Big((2\tau+1)\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}+\lambda\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}\Big)}. \tag{20}$$

Next, by substituting (20) into (18), we get

$$\mathbf{w} = \frac{(2\tau+1)\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}+\lambda\boldsymbol{\Sigma}^{-1}\mathbf{e}}{(2\tau+1)\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}+\lambda\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}}. \tag{21}$$

Based on the description above, Eq. (21) is the optimum weight vector ($\mathbf{w}$) for the investment portfolio with $\tau \geq 0$ (Sukono et al., 2019). To find the value of λ, Eq. (17) is multiplied by $\mathbf{w}^{\mathbf{T}}$, then we get

$$z_{\alpha}(\mathbf{w}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{w})^{\frac{1}{2}} = -\big((2\tau + 1)\mathbf{w}^{\mathbf{T}}\boldsymbol{\mu} + \lambda\big). \tag{22}$$

Substituting EQ. (20) and Eq. (21) into EQ. (22), then will be obtained

$$(\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e})\lambda^2 + \Big((2\tau + 1)\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}\Big)\lambda + (2\tau + 1)^2\boldsymbol{\mu}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - z_{\alpha}^2 = 0. \tag{23}$$

Eq. (23) is a quadratic equation in λ, so it can be calculated using the root formula of the quadratic equation as follows

$$\lambda_{1.2} = \frac{-b\pm\sqrt{b^2-4ac}}{2a}, \text{ where } \lambda \geq 0, \tag{24}$$

with $a = \mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}$, $b = (2\tau + 1)\mathbf{e}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} + \boldsymbol{\mu}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\mathbf{e}$, and $c = (2\tau + 1)^2\boldsymbol{\mu}^{\mathbf{T}}\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu} - z_{\alpha}^2$ (Sukono et al., 2019).

2.6. Sharpe Ratio

Sharpe ratio is a method used to measure portfolio performance. In general, the Sharpe ratio is a performance measurement through a portfolio return approach and measures the risk adjusted return where the higher the value generated, the better the portfolio performance formed (Azis & Purnamasari, 2017). Risk adjusted return is the calculation of return that is adjusted to the risk that must be borne. Sharpe ratio can be defined as

$$S = \frac{R_p - R_{RF}}{VaR_p}, \tag{25}$$

where S is Sharpe ratio, R_p is return of portfolio, R_{RF} is free risk return, and VaR_p is risk of portfolio that measured using VaR (Anggara & Yulianto, 2017).

3. Result

The formation of an investment portfolio with a stock selection strategy through a clustering approach is carried out according to the simulation steps in Fig. 1. The results of each step are analyzed using python programming language and presented as follows.

3.1. Stocks Attribute

The attributes used for clustering are expected return and value at risk that calculated using Eq. (2) and Eq. (3). These values are used to represent the characteristics of the stock. The values are shown in Table 1.

Table 1
Stocks Attributes

No	Code	Exp. Return	VaR	No	Code	Exp. Return	VaR
1	AALI	0.001839	0.029089	41	INCO	-0.000927	0.030411
2	ACES	-0.000196	0.036652	42	INDF	0.000290	0.022997
3	ADRO	0.004341	0.044632	43	INKP	0.001234	0.046706
4	AGII	-0.001241	0.055928	44	INTP	0.001611	0.035486
5	AKRA	0.000562	0.039689	45	ISAT	0.000058	0.044536
6	AMRT	-0.001195	0.048356	46	ITMG	0.002419	0.048927
7	ANTM	-0.002477	0.042552	47	JPFA	0.000504	0.033719
8	ASII	0.001218	0.032485	48	JSMR	-0.001403	0.033356
9	ASRI	-0.000283	0.036552	49	KAEF	-0.001296	0.041722
10	ASSA	0.001755	0.058065	50	KLBF	0.001942	0.024447
11	BBCA	0.002055	0.021155	51	LPKR	-0.000453	0.040976
12	BBNI	0.003526	0.029571	52	LPPF	0.006412	0.050116
13	BBRI	0.000828	0.028300	53	LSIP	0.001224	0.038732
14	BBTN	0.002409	0.033584	54	MAPI	0.001832	0.040374
15	BFIN	0.003134	0.051215	55	MDKA	0.002237	0.045487
16	BJBR	0.000750	0.021639	56	MEDC	0.001616	0.049068
17	BJTM	0.000620	0.018802	57	MIKA	0.000092	0.031876
18	BMRI	0.002324	0.024908	58	MNCN	0.000777	0.032059
19	BMTR	-0.000231	0.020915	59	MYOR	-0.001254	0.027002
20	BRPT	-0.000956	0.039967	60	PGAS	0.002910	0.035840
21	BSDE	-0.000055	0.033576	61	PTBA	0.002150	0.033811
22	BTPS	0.003890	0.046688	62	PTPP	0.000943	0.044076
23	BUKA	-0.008068	0.076167	63	PWON	0.000200	0.032923
24	CPIN	-0.000012	0.028131	64	SCMA	-0.002955	0.053778
25	CTRA	0.000201	0.039866	65	SIDO	0.001371	0.023840
26	DGNS	-0.002310	0.056078	66	SMGR	-0.000901	0.039824
27	DMAS	-0.000164	0.026561	67	SMRA	-0.000644	0.044187
28	DOID	-0.001411	0.047667	68	SRTG	0.003585	0.045472
29	DSNG	0.001454	0.045504	69	TAPG	0.000685	0.052540
30	ELSA	0.000697	0.034411	70	TBIG	-0.000594	0.035127
31	EMTK	-0.002677	0.063477	71	TINS	-0.001322	0.039299
32	ERAA	-0.001377	0.038225	72	TKIM	0.000391	0.041007
33	ESSA	0.002993	0.070263	73	TLKM	0.002013	0.025802
34	EXCL	0.001976	0.030048	74	TOWR	-0.002380	0.028000
35	GGRM	-0.000399	0.026136	75	TPIA	-0.000215	0.033244
36	HEAL	-0.000513	0.028913	76	UNTR	0.001728	0.038305
37	HMSP	-0.000734	0.024323	77	UNVR	-0.000438	0.038531
38	HOKI	-0.001329	0.034042	78	WIKA	0.001133	0.046622
39	HRUM	0.005671	0.064340	79	WMUU	-0.001078	0.047169
40	ICBP	0.000651	0.021017	80	WSKT	-0.003001	0.053129

3.2. Clustering Result

Clustering is done using KM and AL. The clustering technique is used to group stocks with similar characteristics based on predetermined attributes. The number of clusters is simulated from 2 to 10 with the results of the two clustering techniques as follows. Clusters are presented by **Ci**(a, b, c, etc) where i is cluster number and a, b, c, ect. is cluster member which is listed based on the stock number in Table 1.

3.2.1. K-Means

KM clustering groups objects based on the nearest distance to the centroids which is continuously updated in each iteration according to the number of clusters. KM clustering is conducted by following the steps in section 2.2. The results of KM clustering can be seen in Table 2.

Table 2
Result of KM Clustering

k	2	3	4	5	6	7	8	9	10
Member	**C1**(1, 2, 5, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 20, 21, 24, 25, 27, 30, 32, 34, 35, 36, 37, 38, 40, 41, 42, 44, 47, 48, 50, 51, 53, 54, 57, 58, 59, 60, 61, 63, 65, 66, 70, 71, 72, 73, 74, 75, 76, 77) **C2**(3, 4, 6, 7, 10, 15, 22, 23, 26, 28, 29, 31, 33, 39, 43, 45, 46, 49, 52, 55, 56, 62, 64, 67, 68, 69, 78, 79, 80)	**C1**(1, 8, 11, 12, 13, 14, 16, 17, 18, 19, 21, 24, 27, 30, 34, 35, 36, 37, 38, 40, 41, 42, 44, 47, 48, 50, 57, 58, 59, 60, 61, 63, 65, 70, 73, 74, 75) **C2**(2, 3, 5, 6, 7, 9, 15, 20, 22, 25, 28, 29, 32, 43, 45, 46, 49, 51, 52, 53, 54, 55, 56, 62, 66, 67, 68, 71, 72, 76, 77, 78, 79) **C3**(4, 10, 23, 26, 31, 33, 39, 64, 69, 80)	**C1**(1, 11, 12, 13, 16, 17, 18, 19, 24, 27, 34, 35, 36, 37, 40, 41, 42, 50, 59, 65, 73, 74) **C2**(2, 5, 7, 8, 9, 14, 20, 21, 25, 30, 32, 38, 44, 47, 48, 49, 51, 53, 54, 57, 58, 60, 61, 63, 66, 70, 71, 72, 75, 76, 77) **C3**(3, 4, 6, 10, 15, 22, 26, 28, 29, 43, 45, 46, 52, 55, 56, 62, 64, 67, 68, 69, 78, 79, 80) **C4**(23, 31, 33, 39)	**C1**(1, 2, 8, 9, 12, 14, 21, 30, 34, 38, 41, 44, 47, 48, 57, 58, 60, 61, 63, 70, 75) **C2**(3, 5, 7, 20, 22, 25, 29, 32, 43, 45, 49, 51, 53, 54, 55, 62, 66, 67, 68, 71, 72, 76, 77, 78, 79) **C3**(4, 6, 10, 15, 26, 28, 46, 52, 56, 64, 69, 80) **C4**(11, 13, 16, 17, 18, 19, 24, 27, 35, 36, 37, 40, 42, 50, 59, 65, 73, 74) **C5**(23, 31, 33, 39)	**C1**(1, 8, 12, 14, 21, 30, 34, 36, 38, 41, 44, 47, 48, 57, 58, 60, 61, 63, 70, 75) **C2**(2, 5, 7, 9, 20, 25, 32, 49, 51, 53, 54, 66, 71, 72, 76, 77) **C3**(3, 6, 15, 22, 28, 29, 43, 45, 46, 52, 55, 56, 62, 67, 68, 78, 79) **C4**(4, 10, 26, 64, 69, 80) **C5**(11, 13, 16, 17, 18, 19, 24, 27, 35, 37, 40, 42, 50, 59, 65, 73, 74) **C6**(23, 31, 33, 39)	**C1**(1, 8, 12, 14, 21, 30, 34, 36, 38, 41, 44, 47, 48, 57, 58, 60, 61, 63, 70, 75) **C2**(2, 5, 7, 9, 20, 25, 32, 49, 51, 53, 54, 66, 71, 72, 76, 77) **C3**(3, 6, 15, 22, 28, 29, 43, 45, 46, 52, 55, 56, 62, 67, 68, 78, 79) **C4**(4, 10, 26, 64, 69, 80) **C5**(11, 13, 16, 17, 18, 19, 24, 27, 35, 37, 40, 42, 50, 59, 65, 73, 74) **C6**(23) **C7**(31, 33, 39)	**C1**(1, 12, 13, 24, 27, 34, 35, 36, 41, 59, 73, 74) **C2**(2, 8, 9, 14, 21, 30, 38, 44, 47, 48, 57, 58, 60, 61, 63, 70, 75) **C3**(3, 6, 15, 22, 28, 29, 43, 45, 46, 52, 55, 56, 62, 67, 68, 78, 79) **C4**(4, 10, 26, 64, 69, 80) **C5**(5, 7, 20, 25, 32, 49, 51, 53, 54, 66, 71, 72, 76, 77) **C6**(11, 16, 17, 18, 19, 37, 40, 42, 50, 65) **C7**(23) **C8**(31, 33, 39)	**C1**(1, 12, 13, 24, 27, 34, 35, 36, 41, 59, 73, 74) **C2**(2, 8, 9, 14, 21, 30, 38, 44, 47, 48, 57, 58, 60, 61, 63, 70, 75) **C3**(3, 6, 22, 28, 29, 43, 45, 55, 62, 67, 68, 78, 79) **C4**(4, 10, 26, 31, 64, 80) **C5**(5, 7, 20, 25, 32, 49, 51, 53, 54, 66, 71, 72, 76, 77) **C6**(11, 16, 17, 18, 19, 37, 40, 42, 50, 65) **C7**(15, 46, 52, 56, 69) **C8**(23) **C9**(33, 39)	**C1**(1, 12, 13, 24, 27, 34, 35, 36, 41, 59, 73, 74) **C2**(2, 5, 7, 9, 20, 25, 32, 49, 51, 53, 54, 66, 71, 72, 76, 77) **C3**(3, 15, 22, 46, 52, 56, 68) **C4**(4, 26, 64, 69, 80) **C5**(6, 28, 29, 43, 45, 55, 62, 67, 78, 79) **C6**(8, 14, 21, 30, 38, 44, 47, 48, 57, 58, 60, 61, 63, 70, 75) **C7**(10, 31) **C8**(11, 16, 17, 18, 19, 37, 40, 42, 50, 65) **C9**(23) **C10**(33, 39)
Silhouette	0.5405	0.5127	0.5067	0.4416	0.4729	0.4726	0.4665	0.4432	0.4340

3.2.2. Average Linkage

AL clustering groups objects based on the nearest distance of objects to other objects as the initial cluster. The clustering process runs by combining clusters that have the nearest average distance in each iteration to form a single cluster in the end. AL clustering is conducted by following the steps in section 2.3. The results of AL clustering can be seen in Table 3.

Table 3
Result of AL Clustering

k	2	3	4	5	6	7	8	9	10
Member	**C1**(1, 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 27, 28, 29, 30, 32, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79) **C2**(4, 10, 23, 26, 31, 33, 39, 64, 80)	**C1**(1, 2, 3, 5, 6, 7, 8, 9, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 24, 25, 27, 28, 29, 30, 32, 34, 35, 36, 37, 38, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79) **C2**(174, 10, 26, 31, 33, 39, 64, 80) **C3**(23)	**C1**(1, 2, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 21, 24, 27, 30, 34, 35, 36, 37, 38, 40, 41, 42, 44, 47, 48, 50, 57, 58, 59, 60, 61, 63, 65, 70, 73, 74, 75) **C2**(3, 5, 6, 7, 15, 20, 22, 25, 28, 29, 32, 43, 45, 46, 49, 51, 52, 53, 54, 55, 56, 62, 66, 67, 68, 69, 71, 72, 76, 77, 78, 79) **C3**(4, 10, 26, 31, 33, 39, 64, 80) **C4**(23)	**C1**(1, 2, 8, 9, 11, 12, 13, 14, 16, 17, 18, 19, 21, 24, 27, 30, 34, 35, 36, 37, 38, 40, 41, 42, 44, 47, 48, 50, 57, 58, 59, 60, 61, 63, 65, 70, 73, 74, 75) **C2**(3, 5, 6, 7, 15, 20, 22, 25, 28, 29, 32, 43, 45, 46, 49, 51, 52, 53, 54, 55, 56, 62, 66, 67, 68, 69, 71, 72, 76, 77, 78, 79) **C3**(4, 10, 26, 31, 64, 80) **C4**(23) **C5**(33, 39)	**C1**(1, 2, 8, 9, 12, 13, 14, 21, 24, 27, 30, 34, 35, 36, 38, 41, 44, 47, 48, 57, 58, 59, 60, 61, 63, 70, 74, 75) **C2**(3, 5, 6, 7, 15, 20, 22, 25, 28, 29, 32, 43, 45, 46, 49, 51, 52, 53, 54, 55, 56, 62, 66, 67, 68, 69, 71, 72, 76, 77, 78, 79) **C3**(4, 10, 26, 31, 64, 80) **C4**(11, 16, 17, 18, 19, 37, 40, 42, 50, 65, 73) **C5**(23) **C6**(33, 39)	**C1**(1, 2, 8, 9, 12, 13, 14, 21, 24, 27, 30, 34, 35, 36, 38, 41, 44, 47, 48, 57, 58, 59, 60, 61, 63, 70, 74, 75) **C2**(3, 5, 6, 7, 15, 20, 22, 25, 28, 29, 32, 43, 45, 46, 49, 51, 52, 53, 54, 55, 56, 62, 66, 67, 68, 69, 71, 72, 76, 77, 78, 79) **C3**(4, 10, 26, 64, 80) **C4**(11, 16, 17, 18, 19, 37, 40, 42, 50, 65, 73) **C5**(23) **C6**(31) **C7**(33, 39)	**C1**(1, 2, 8, 9, 12, 13, 14, 21, 24, 27, 30, 34, 35, 36, 38, 41, 44, 47, 48, 57, 58, 59, 60, 61, 63, 70, 74, 75) **C2**(3, 6, 15, 22, 28, 29, 43, 45, 46, 52, 55, 56, 62, 67, 68, 69, 78, 79) **C3**(4, 10, 26, 64, 80) **C4**(5, 7, 20, 25, 32, 49, 51, 53, 54, 66, 71, 72, 76, 77) **C5**(11, 16, 17, 18, 19, 37, 40, 42, 50, 65, 73) **C6**(23) **C7**(31) **C8**(33, 39)	**C1**(1, 2, 8, 9, 12, 13, 14, 21, 24, 27, 30, 34, 35, 36, 38, 41, 44, 47, 48, 57, 58, 59, 60, 61, 63, 70, 74, 75) **C2**(3, 6, 15, 22, 28, 29, 43, 45, 46, 52, 55, 56, 62, 67, 68, 69, 78, 79) **C3**(4, 10, 26, 64, 80) **C4**(5, 7, 20, 25, 32, 49, 51, 53, 54, 66, 71, 72, 76, 77) **C5**(11, 16, 17, 18, 19, 37, 40, 42, 50, 65, 73) **C6**(23) **C7**(31) **C8**(33) **C9**(39)	**C1**(1, 12, 13, 24, 27, 34, 35, 36, 41, 59, 74) **C2**(2, 8, 9, 14, 21, 30, 38, 44, 47, 48, 57, 58, 60, 61, 63, 70, 75) **C3**(3, 6, 15, 22, 28, 29, 43, 45, 46, 52, 55, 56, 62, 67, 68, 69, 78, 79) **C4**(4, 10, 26, 64, 80) **C5**(5, 7, 20, 25, 32, 49, 51, 53, 54, 66, 71, 72, 76, 77) **C6**(11, 16, 17, 18, 19, 37, 40, 42, 50, 65, 73) **C7**(23) **C8**(31) **C9**(33) **C10**(39)
Silhouette	0.5369	0.4801	0.4983	0.4871	0.4360	0.4125	0.4199	0.4138	0.4444

3.3. Selected Stocks

The stocks that have been clustered are then selected 1 from each cluster with the criteria of stocks having the highest positive expected return in the cluster. Different clusters have different member characteristics. This selection aims to maximize the diversity of the combined stocks so that the portfolio formed is more diversified. Furthermore, there are a number of k that produce 1 selected stock and the same stock combination. The results presented are different stock combinations and more than 1 stock is selected from the approaches of the two clustering techniques.

3.3.1. Stock Combination Approached by K-Means

Different stock combinations were selected based on the results of KM clustering according to the number of clusters presented based on the stock code in Table 1 can be seen in Table 4.

Table 4
Stocks Combination Approached by KM

k	Number of Stock	Stock Combination	Portfolio
2	2	BBNI, LPPF	P1
3	3	BBNI, HRUM, LPPF	P3
4	4	BBNI, HRUM, LPPF, PGAS	P4
5	5	ADRO, BBNI, BMRI, HRUM, LPPF	P6
6, 7	6	ASSA, BBNI, BMRI, HRUM, LPPF, MAPI	P8
8	7	ASSA, BBNI, BMRI, HRUM, LPPF, MAPI, PGAS	P9
9	8	ADRO, ASSA, BBNI, BMRI, HRUM, LPPF, MAPI, PGAS	P11
10	8	ASSA, BBNI, BMRI, HRUM, LPPF, MAPI, MDKA, PGAS	P12

3.3.2. Stock Combination Approached by Average Linkage

Different stock combinations were selected based on the results of AL clustering according to the number of clusters presented based on the stock code in table 1 can be seen in Table 5.

Table 5
Stocks Combination Approached by AL

k	Number of Stock	Stock Combination	Portfolio
2, 3	2	HRUM, LPPF	P2
4	3	BBNI, HRUM, LPPF	P3
5	4	ASSA, BBNI, HRUM, LPPF	P5
6, 7	5	ASSA, BBNI, BMRI, HRUM, LPPF	P7
8	6	ASSA, BBNI, BMRI, HRUM, LPPF, MAPI	P8
9	7	ASSA, BBNI, BMRI, ESSA, HRUM, LPPF, MAPI	P10
10	8	ASSA, BBNI, BMRI, ESSA, HRUM, LPPF, MAPI,PGAS	P13

3.4 Mean-VaR Portfolio Performance Optimization

The formation of the investment portfolio is carried out by an optimization process based on the Mean-VaR model. Portfolio optimization based on the Mean-VaR model aims to obtain an efficient portfolio composition by maximizing returns and minimizing risk levels as measured by VaR. The Mean-VaR portfolio optimization problem is structured referring to the equation (13). For example, the optimization process for P6 with 5 stocks is presented.

The value of expected return stocks in P6 is arranged in the form of an vector $\boldsymbol{\mu}_{\mathbf{P6}}$, a unit vector of $\mathbf{e}_{\mathbf{P6}}$ is formed for 5 stocks, and the results of the calculation of the covariance between stocks S_1 to S_5 are formed into a covariance matrix $\boldsymbol{\Sigma}_{\mathbf{P6}}$. Using vectors $\boldsymbol{\mu}_{\mathbf{P6}}$, $\mathbf{e}_{\mathbf{P6}}$, and matrices $\boldsymbol{\Sigma}_{\mathbf{P6}}$, the weight vector $\mathbf{w}$ is calculated using equation (21). The risk tolerance τ with the condition 0 in this portfolio optimization is simulated by taking several values that meet the conditions $\lambda > 0$ and $\mathbf{e}^{\mathbf{T}}\mathbf{w} = 1$. Taking the risk tolerance value is discontinued for a risk tolerance value substituted into equation (21) produces a weight $\mathbf{w}_{\mathbf{i}}$ $(i = 1. \dots .5)$ which is not a positive real number that satisfies $\lambda > 0$ and $\mathbf{e}^{\mathbf{T}}\mathbf{w} = 1$. The portfolio formed with the allocation of stocks that have been determined is then calculated the expected return using Equation (10), VaR using Equation (12), and performance of portofolios by Sharpe ratio using Equation (25) with R_{RF} of 0.002759. The taking of risk tolerance values and the results of optimizing portfolios are given in Table 6.

Tabel 6
Simulation Result of Mean-VaR Optimization of P6

τ	λ	S_1	S_2	S_3	S_4	S_5	$\mathbf{e}^{\mathbf{T}}\mathbf{w}$	μ_p	VaR_p	Sharpe
0.00	0.0193	0.1009	0.1775	0.4885	0.0724	0.1607	1.00	0.00364	0.019305	0.172244
0.01	0.0193	0.1010	0.1778	0.4875	0.0726	0.1611	1.00	0.003643	0.019305	0.172392
0.02	0.0192	0.1011	0.1782	0.4865	0.0728	0.1614	1.00	0.003646	0.019305	0.172525
0.03	0.0192	0.1012	0.1786	0.4854	0.0730	0.1618	1.00	0.003649	0.019305	0.172679
0.04	0.0191	0.1014	0.1789	0.4844	0.0731	0.1621	1.00	0.003651	0.019303	0.172803
…	…	…	…	…	…	…	…	…	…	…
2.37	0.0005	0.1353	0.2828	0.1940	0.1245	0.2633	1.00	0.00443	0.021268	0.193470
2.38	0.0003	0.1356	0.2838	0.1915	0.1249	0.2642	1.00	0.004437	0.021301	0.193498
2.39	0.0002	0.1360	0.2847	0.1889	0.1254	0.2651	1.00	0.004444	0.021337	0.193526
2.40	0.00004	0.1363	0.2856	0.1862	0.1259	0.2660	1.00	0.004451	0.021369	0.193549

Based on Table 6, it can be seen that for each $0 \leq \tau \leq 2,40$ risk tolerance value, the portfolio expected return μ_{P6} and the risk level VaR_{P6} are different. For each increase in the value of risk tolerance, it causes an increase in the value of the portfolio's expected return μ_{P6} and is accompanied by an increase in the level of risk VaR_{P6}. Meanwhile, the risk tolerance value $\tau > 2,40$ is not feasible to calculate the expected return μ_{P6} and the risk level VaR_{P6}, because it is not satisfies $\lambda > 0$. The graph of the efficient frontier can be seen in Fig. 2.

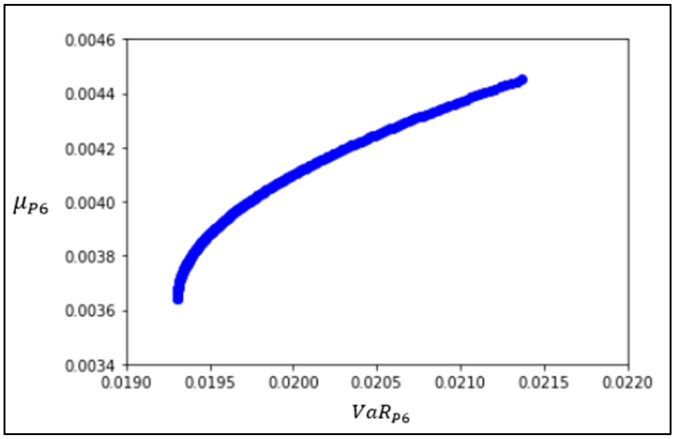

Fig. 2. Efficient Frontier of P6

Among the efficient frontier there is an optimal portfolio and his optimal portfolio that needs to be sought. Based on the optimization performed, the results show that efficient portfolios lie along the line with a risk tolerance of $0 \leq \tau \leq 2,40$. For the risk tolerance value $\tau = 2,40$ produces an expected return portfolio $\mu_{P6} = 0,004451$ with a risk level of $VaR_{P6} = 0,021369$. These values are the expected return and VaR with the optimum weight of $\mathbf{w}_{P6}^{\mathbf{T}} = (0,1363 \quad 0,2856 \quad 0,1862 \quad 0,1259 \quad 0,2660)$ and optimum performance of $S_{P6} = 0,193549$. The optimal weight of the fund allocation for the entire portfolio formed is presented in Table 7.

Table 7
Optimal Portfolio Weight and Performance

Portfolio	Weight	Exp. Return	VaR	Sharpe
P1	0.6122, 0.3878	0.004645	0.026981	0.160485
P2	0.3657, 0.6343	0.006141	0.037932	0.153592
P3	0.4799, 0.1884, 0.3317	0.004888	0.025001	0.182914
P4	0.3885, 0.1717, 0.2896, 0.1502	0.004638	0.023157	0.186684
P5	0.1051, 0.4205, 0.1791, 0.2952	0.004576	0.022666	0.187993
P6	0.1363, 0.2856, 0.1862, 0.1259, 0.2660	0.004451	0.021369	0.193549
P7	0.0834, 0.2842, 0.1997, 0.1591, 0.2736	0.004269	0.020756	0.190501
P8	0.0852, 0.2780, 0.2139, 0.1557, 0.2670, 0.0002	0.004222	0.020532	0.190290
P9	0.0823, 0.2419, 0.1901, 0.1448, 0.2374, 0.0002, 0.1034	0.003685	0.017437	0.193269
P10	0.0848, 0.2785, 0.1779, 0.0557, 0.1457, 0.2571, 0.0003	0.004184	0.020015	0.193307
P11	0.1082, 0.0713, 0.2541, 0.1294, 0.1245, 0.2326, 0.0002, 0.0798	0.004220	0.019733	0.197894
P12	0.0888, 0.1896, 0.2540, 0.1268, 0.1939, 0.0417, 0.0000, 0.1053	0.003756	0.018464	0.186365
P13	0.0823, 0.2467, 0.1604, 0.0502, 0.1371, 0.2320, 0.0002, 0.0910	0.004067	0.019274	0.194668

4. Discussion

The clustering process for each number of k produces different combinations of stocks with the number of stocks selected varies for both methods. Since each stock has different characteristics according to its expected return and VaR, the results depend on the distribution of the analyzed stocks. The distribution of analyzed stocks can be seen in Fig. 2.

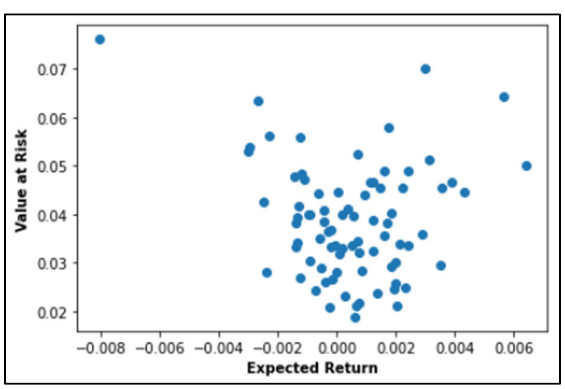

Fig. 2. Scatter Plot of IDX80 Stocks.

Based on Fig. 2, it can be seen that the stocks are spread based on the attributes they have. The distribution has an expected return range from -0.008068 to 0.006412 with a mean of 0.000472 and a VaR range from 0.018802 to 0.076167 with a mean of 0.038750. It can also be seen that the stocks spread randomly and some stocks are close to other stocks. This indicates that the stocks that are close to each other can form a cluster. On the other hand, some stocks spread far from the data center and far from other stocks. This also indicates that these stocks can form separate clusters depending on the clustering method used.

It can be seen in Table 2, KM which is a non-hierarchical partitional clustering groups objects based on their closest distance to the centroids which is continuously updated for each k cluster. As a result, for k of 6 to 10 forming 1 cluster which has 1 member stock. This stock, as seen in Fig. 2, is located quite far from the data set, so that the stock forms its own cluster. In contrast to the results of AL which can be seen in Table 3. AL which is an agglomerative hierarchical clustering grouping objects by allocating objects that have the closest distance. As a result, based on the analyzed stocks, starting from k of 3, a cluster consisting of 1 stock is formed. Since AL is hierarchical, this result continues for k to 10 and clusters with 1 stock of members increase when k is 6, 8 and at k of 10 form 4 clusters whose members are 1 stock. This happens because these stocks are located quite far from the data group and are not close to other stocks as shown in Fig. 2.

To see which clustering results produce the best cluster structure for each k, we can see the silhouette scores generated from the two clustering techniques. Silhouette score comparison for each k can be seen in Fig. 3.

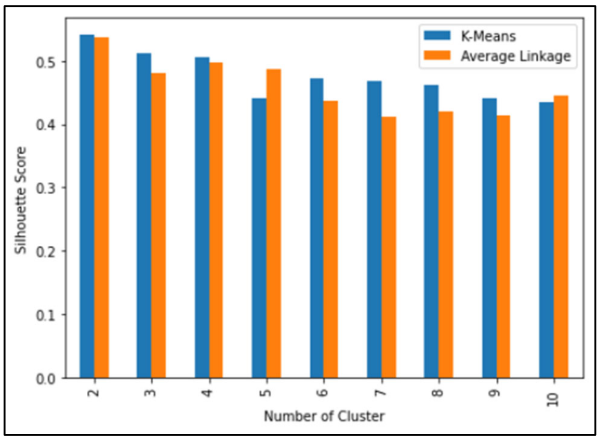

Fig. 3. Silhouette score of KM and AL

Based on Figure 3, KM has a higher silhouette score than AL, except for k of 6 and 10. This shows that for stocks analyzed with the distribution as shown in Fig. 2, KM performs better than AL for every k, except at k of 6 and 10. Both clustering techniques have optimal silhouettes respectively at k of 2 where at this optimal value the clustering process represents the best results. This means that for k of 2, each clustering techniques forms a cluster whose cluster members have the most uniform characteristics and each cluster has the most diverse characteristics. However at k of 2, the cluster generated by KM is better than AL.

The different results of the two clustering techniques produce different stock combinations as can be seen in Table 4 and 5. However, in some k, the same stock combination is formed for the KM and AL approaches. In the KM approach, for k of 6 and 7 obtained 6 stocks with the same combination. In addition, for k of 9 and 10 obtained the same number of stocks, namely 8 stocks, but with different combinations. On the other hand, the AL approach produces the same combination of k of 2 and 3 with a total of 2 stocks, and k of 6 and 7 with a total of 5 stocks. As for k of 3, both KM and AL produce the same stock combination. The same combination of stocks is also obtained for a total of 6 stocks when k is 6 and 7 in KM with k equal to 8 in AL.

Based on the stock combinations obtained, then a portfolio is formed by determining the optimal weight of each stock. Portfolios with the same stock combinations produce the same weights and performances as seen as Table 7. It is also seen that the largest expected return is owned by P3 with a value of 0.004888, but P3 also has the highest VaR value of 0.037932, so that P3 provides a performance of 0.182914. On the other hand, P9 has the smallest VaR with a value of 0.017437, but P9 also has the smallest expected return of 0.003685, so P9 gives a performance of 0.193269. If the investment decision is chosen based on the highest portfolio performance, then P11 becomes the optimal portfolio that can be used as an investment choice which produces a performance of 0.197894 by offering an expected return of 0.004220 and a VaR of 0.019733. As a comparison of portfolio performance based on the number of stocks incorporated and the clustering approach used, it can be seen in Fig. 4. It can be seen in Fig. 4 that the highest portfolio performance is owned by P11 which is a portfolio with the formation of a combination of stocks selected based on the results of KM clustering with a number of k of 9 and produces a combination of 8 stocks. Therefore, P11 can be used as a recommendation to be used as optimal portfolio for investing in stocks. Further looking at the number of clusters formed, in this case, the combination of selected stocks with the KM approach gives a better portfolio performance than AL because the number of stocks selected is greater with the same number of k used. If we look at the number of stocks incorporated, the KM approach also provides a higher portfolio performance than the AL approach. However, the AL approach gives a higher portfolio performance on the total of 4 stocks and 7 stocks.

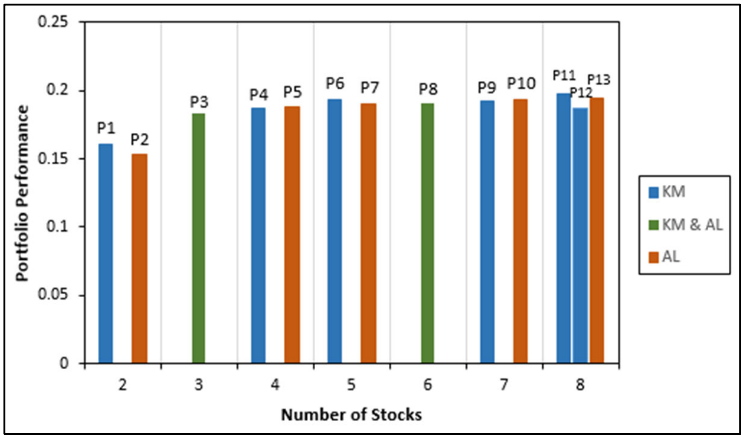

Fig. 4. Portfolio Performance Grouped by Number of Stock

The results of stock selection to form an optimal and well-diversified portfolio are not only seen from the number of clusters formed, but also from the number of selected stocks. From this cases and methods used, it can be seen that the number of k used does not necessarily produce the same number of selected stocks. Therefore, it is still necessary to evaluate the stock selection algorithm so that the use of clustering techniques can be assessed more objectively in producing a combination of stocks that have good diversification and provide optimal portfolio performance.

5. Conclusions

The use of KM and AL in determining the combination of stocks to form a portfolio produces 13 different portfolios with the number of stocks from 2 to 8 stocks. The P11 portfolio is obtained as the optimal portfolio with a combination of stocks, namely ADRO, ASSA, BBNI, BMRI, HRUM, LPPF, MAPI, and PGAS which have weights respectively of 0.1082, 0.0713, 0.2541, 0.1294, 0.1245, 0.2326, 0.0002, and 0.0798. P11 offers an expected return of 0.004222 with a VaR of 0.019733 and provides a portfolio performance of 0.197894. The combination of P11 stocks is obtained using the KM clustering approach with a total of k of 9. Therefore, the KM clustering can be used as a recommendation in making a decision on the selection of stock combinations to form an optimal portfolio. The results of this study still require further evaluation, especially in the stock selection algorithm that uses the clustering technique. The selection of stocks that have positive returns can be done earlier so that the clustering process is more effective and produces a stock combination that matches the number of clusters formed. The stock attributes used in clustering are still limited to stock returns and risks which can be further developed by measuring the correlation of the return movement of each stock in the clustering process so that the cluster formed can be diversified based on the movement of its stock return.

Acknowledgments

We would like to thank Universitas Padjadjaran, which has provided the Universitas Padjadjaran Internal Research Grant, Fiscal Year 2022, the "Academic Leadership Grant (ALG)" program under Prof. Dr. Sukono, MM., M.Si., with contract Number: 2203/UN6.3.1/PT.00/2022.

References

Abdurrahman, G. (2019). Clustering Data Kredit Bank Menggunakan Algoritma Agglomerative Hierarchical Clustering Average Linkage. *JUSTINDO (Jurnal Sistem Dan Teknologi Informasi Indonesia)*, *4*(1), 13. https://doi.org/10.32528/justindo.v4i1.2418

Anggara, A. W., & Yulianto, A. (2017). Analisis Kinerja Reksa Dana Saham dengan Metode Sharpe, Treynor dan Jensen. *Management Analysis Journal*, *6*(1), 13–22.

Arimarista, L. (2017). Expected Return dan Risiko Saham LQ-45 untuk Pengambilan Keputusan Investasi serta Pembentukan Portofolio Optimal (Capital Asset Pricing Model). *Journal of Accounting Science*, *1*(1), 62–70. https://doi.org/10.21070/jas.v1i1.776

Azis, M., & Purnamasari, I. (2017). Analisis Kinerja Portofolio Reksadana dengan Metode Rasio Sharpe dan Non Risk Premium. *Prosiding Seminar Nasional Manajemen Dan Ekonomi Bisnis*, *1*(1), 246–264.

Bhattacharjee, B., Shafi, M., & Acharjee, A. (2019). Network Mining Based Elucidation of The Dynamics of Cross-Market Clustering and Connectedness in Asian Region: An MST and Hierarchical Clustering Approach. *Journal of King Saud University - Computer and Information Sciences*, *31*(2), 218–228. https://doi.org/10.1016/j.jksuci.2017.11.002

Bienvenido-Huertas, D., Marín-García, D., Carretero-Ayuso, M. J., & Rodríguez-Jiménez, C. E. (2021). Climate Classification for New and Restored Buildings in Andalusia: Analysing the Current Regulation and A New Approach

based on K-Means. *Journal of Building Engineering*, *43*, 102829. https://doi.org/10.1016/j.jobe.2021.102829
Cebeci, Z., & Yildiz, F. (2015). Comparison of K-Means and Fuzzy C-Means Algorithms on Different Cluster Structures. *Journal of Agricultural Informatics*, *6*(3), 13–23. https://doi.org/10.17700/jai.2015.6.3.196
Chairunnisa, C. A., Yozza, H., & Devianto, D. (2018). Pengukuran Nilai Risiko Portofolio berdasarkan Mean-VaR. *Jurnal Matematika UNAND*, *7*(1), 24. https://doi.org/10.25077/jmu.7.1.24-32.2018
Corporal-Lodangco, I. L., Richman, M. B., Leslie, L. M., & Lamb, P. J. (2014). Cluster Analysis of North Atlantic Tropical Cyclones. *Procedia Computer Science*, *36*(C), 293–300. https://doi.org/10.1016/j.procs.2014.09.096
Ediyanto, Mara, N., & Satyahadewi, N. (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*, *02*(2), 133–136.
Eliguzel, I. M., & Ozceylan, E. (2019). Comparison of Different Clustering Methods for Cellular Manufacturing: A Case of Gym Centre. *Procedia Computer Science*, *158*, 1–8. https://doi.org/10.1016/j.procs.2019.09.021
Hidayati, A. N. (2017). Investasi : Analisis dan Relevansinya dengan Ekonomi Islam. *Jurnal Ekonomi Islam*, *8*(2), 227–242.
IDX. (2021). *IDX Stock Index Handbook V1.2*. Indonesia Stock Exchange. https://www.idx.co.id/media/9816/idx-stock-index-handbook-v12-_-januari-2021.pdf
Ismanto, H. (2016). Analisis Value at Risk dalam Pembentukan Portofolio Optimal (Studi Empiris pada Saham-saham). *University Research Colloquium 2016*, *3*(1), 243–255.
Jiang, K., Li, D., Gao, J., & Yu, J. X. (2014). Factor Model Based Clustering Approach for Cardinality Constrained Portfolio Selection. *IFAC Proceedings*, *19*(3), 10713–10718. https://doi.org/10.3182/20140824-6-za-1003.00663
Kumari, S. K., Kumar, P., Priya, J., Surya, S., & Bhurjee, A. K. (2019). Mean-Value at Risk Portfolio Selection Problem using Clustering Technique : A Case Study. *AIP Conference Proceedings*, *2112*(020178), 1–7.
Madhulatha, T. S. (2012). An overview of clustering methods. *IOSR Journal of Engineering*, *2*(4), 719–725. https://doi.org/10.3233/ida-2007-11602
Markowitz, H. (1952). Portofolio Selection. *Journal of Finance*, *7*(1), 77–91. https://doi.org/10.2307/2345307
Naeem, S., & Wumaier, A. (2018). Study and Implementing K-mean Clustering Algorithm on English Text and Techniques to Find the Optimal Value of K. *International Journal of Computer Applications*, *182*(31), 7–14. https://doi.org/10.5120/ijca2018918234
Safelia, N. (2012). Konsep Dasar Keputusan Investasi dan Portofolio. *Jurnal Manajemen Terapan Dan Keuangan*, *1*(3), 217–226.
Setyawati, I. (2011). Perdagangan Saham Kaitannya dengan Perolehan Expected Return pada Bursa Efek Indonesia. *Jurnal Ilmiah Manajemen STIMA KOSGORO*, *17*(1), 31–41.
Subekti, R., Kusumawati, R., & Sari, E. R. (2017). K-Means Clustering dan Average Linkage dalam Pembentukan Portfolio Saham. *Seminar Matematika Dan Pendidikan Matematika UNY 2017*, 219–224.
Sukono, Lesmana, E., Napitupulu, H., Hidayat, Y., Saputra, J., & Ghazali, P. L. B. (2019). Mean-var portfolio optimisations: An application of multiple index models with non-constant volatility and long memory effects. *International Journal of Innovation, Creativity and Change*, *9*(12), 364–381.
Sukono, Sidi, P., Bon, A. T. Bin, & Supian, S. (2017). Modeling of Mean-VaR Portfolio Optimization by Risk Tolerance when The Utility Function is Quadratic. *AIP Conference Proceedings*, *1827*(020035). https://doi.org/10.1063/1.4979451
Sulistiyowati, E., & Santoso, B. H. (2017). Analisis Portofolio Optimal berdasarkan Model Indeks Tunggal pada Perusahaan Property and Real Estate. *Jurnal Ilmu Dan Riset Manajemen*, *6*(5), 1–17.
Utami, M. P., Saepudin, D., & Rohmawati, A. A. (2019). Pengaruh Teknik Clustering Harga Saham dalam Manajemen Portofolio. *E-Proceeding of Engineering*, *6*(1), 2491–2509.
Widyadhana, D., Hastuti, R. B., Kharisudin, I., & Fauzi, F. (2021). Perbandingan Analisis Klaster K-Means dan Average Linkage untuk Pengklasteran Kemiskinan di Provinsi Jawa Tengah. *PRISMA, Prosiding Seminar Nasional Matematika 4*, *4*, 584–594.
Xu, N., Finkelman, R. B., Dai, S., Xu, C., & Peng, M. (2021). Average Linkage Hierarchical Clustering Algorithm for Determining the Relationships between Elements in Coal. *ACS Omega*, *6*(9), 6206–6217. https://doi.org/10.1021/acsomega.0c05758
Yusniyanti, A. L., Virgantari, F., & Faridhan, Y. E. (2021). Comparison of Average Linkage and K-Means Methods in Clustering Indonesia's Provinces Based on Welfare Indicators. *Journal of Physics: Conference Series*, *1863*(1), 1–10. https://doi.org/10.1088/1742-6596/1863/1/012071