Portfolio Optimization Based on Clustering of Indonesia Stock Exchange: A Case Study of Index LQ45

Bakti Siregar[1]* & F. Anthon Pangruruk[2]
Department of Statistics, Matana University
Email: siregar.bakti@matanauniversity.ac.id[1]*

ABSTRACT: In general portfolio optimization is a technique for selecting the proportion of assets to make a better portfolio by maximizing the expectation return while also minimizing the risk. In this research, k-means clustering method is used to classify stocks are listed on the LQ45 Index and select stocks whose has the price tend to be increase. Then the Markowitz approach is used to analyze the performance of optimization portfolio models that have a minimum variance in expected return and risk. After understanding the performance this portfolio optimization, future works will be able to apply this model in cloud computing or artificial intelligence. In addition, investors will develop a better view of the latest performance of the stocks are listed in LQ45 index and support them decide which stocks that should be include to their portfolios, thus prevent wrong decisions.

INTRODUCTION

Investors must be able to do a fundamental analysis on the stocks they will choose to produce the expected profit. On the other hand, emphasized that the expected return and the risk as to the main parameters of the investment manner (Kulali, 2016). Therefore, investors always try to minimize risk while maximizing their returns as well. In the other words, individual investment strategies depends on how heavily the risk they can take to achieve the expected results. Classically, the risk can be measured by the standard deviation of expected returns. For that reason, when investors take more risks, the profits or losses will also increase. Therefore, to avoid this risk investors choose some stocks with the smallest variant in their portfolio investment. This technique is called portfolio optimization which has been widely used for years in the industry of capital market and financial analysis (Mayanja, Mataramyura, & Charles, 2013).

Markowitz in 1952 was officially the first introduced theory of portfolio optimization, then gradually become the most efficient technique in portfolio preparation and continuously developed until today. This model support investors how to reduce risk and analyzing the various possible portfolios by choosing some stock market are not "moving" together (Markowitz, 1952). In Principal, the Markowitz Mean-Variance (MV) is a quadratic model, where the variance of each stock is adjusted to measure the risk (Xu, Zhou, Jiang, Yu, & Niu, 2016). This is related to the habit of investors, which is more often think of rationality and is reluctant to take risks (avoid risk). As a result, we can conclude that investors tend to choose assets with higher returns given the lower level of risk.

There is a lot of research about the optimization portfolio of the Indonesia Stock Exchange but just a few of them have tried to apply the concept of machine learning to accelerate the formation of models. Nowadays, the era 4.0 industry needed everything in an automated manner including the formation of modeling (algorithms) especially research that involving a big data analysis. Thereupon, this research will focus on forming an algorithm or model to produce an optimal portfolio by applying k-means clustering to classify the Indonesian stock market in the LQ45 index, then training the allocation proportions of each possibility portfolio and in the end choose the best portfolio. The results of this research can be used as a basis for consideration in making decisions for investors, brokers, or individuals to arrange portfolios in allocating stocks.

THEORETICAL REVIEW

About Stocks

Stocks are instruments of market share in the field of financials which are indicated as a sign of ownership of the company. It can be defined as a sign of capital participation of a person, business entity in a company, and limited liability company. Many investors choose to buy stocks in the market share because it is possible to provide them an attractive level of profit (Trimulya, Syaifurrahman, & Setyaningsih, 2015). However, they are expected to be able to make an optimal portfolio for their investment.

K-means Clustering

K-means clustering is a non-hierarchical clustering method in order to classify data in one form or more clusters. In this field, data that have the same characteristics are classified in one cluster, and the data that has different characteristics are classified with another cluster. In consequence, if the data are listed in one cluster means that it has a small degree of variance. K-means clustering declares an initial value of cluster, where the center of the cluster is the first randomly selected, then calculates the distance of each cluster to the center of data using the Euclidean formula in (1) (Nanda, Mahanty, & Tiwari, 2010).

$$D_{ij} = \sqrt{\left(X_{1i} - X_{1j}\right)^2 + \left(X_{2i} - X_{2j}\right)^2 + \cdots \left(X_{ki} - X_{kj}\right)^2} \tag{1}$$

where,

D_{ij} : distance of the data from i to cluster center j.

X_{ki} : data i on the attribute data k.

X_{kj} : center point to j in attribute to k.

Recalculate the center of the cluster with the data in each cluster, where the center of the cluster is the average of all data. Next, perform the calculations of distance with data centers that have been updated, if there are no changes means the technique is complete.

Stock and Portfolio Return

Stock return is the result obtained from investment, it can be divided into two conditions; first is the form of realized returns that have occurred, second is the expected return that has not occurred but is expected to occur in the future. Equation (2) is the formula to calculate the return for a certain period for stocks (Kurniyati, 2007).

$$R_{ij} = \frac{(P_{ij} - P_{ij-1})}{P_{ij-1}} \tag{2}$$

where,

R_{ij} : return of the i stock price when j.

P_{ij} : the price of i stock at j.

The formula of expected return is

$$E(R_i) \approx \frac{\sum_{i=1}^{t} R_{ij}}{t} \tag{3}$$

where,

$E(R_i)$: expected return of investment in the i stock.

R_{ij} : actual return of investment in the i stock.

t : the number of periods (daily, weekly, and monthly).

The portfolio return is the average weighted realized return for every single security in a portfolio. (Capinski, & Zastawniak, 2003) The formula for calculating portfolio returns is

$$P_r = \sum_{i=1}^{t} W_i R_i \tag{4}$$

where,

W_i : weight of i stock portfolio.

R_i : return of investment in i stock.

Refer to the portfolio return (4), the formula for expected portfolio return is

$$E(P_r) = \sum_{i=1}^{t} W_i \, E(R_i) \tag{5}$$

therefore $W_1 + W_2 + \cdots + W_t = 1$.

Stock and Portfolio Risk

(Capinski, & Zastawniak, 2003) Stock risk is a condition faced by someone in the future, it contains a number of possible levels of benefits that deviate from the expected benefits, so it can be formulated as

$$\sigma_v{}^2 \approx \sum_{i=1}^{t} \frac{\left(R_{ij}-E(R_i)\right)^2}{t} \tag{6}$$

Portfolio risk is the magnitude of the deviation between the expected rate of return and the real rate of return achieved (Bodie, Kane, & Marcus, 2014). In this case, the measure of the risk is used to determine how far the possible value is obtained, from the expected value. The portfolio risk is influenced by the average for each asset risk and the covariance between assets, upon of the (7).

$$\sigma_v{}^2 = \sum_{A=1}^{t} W_A^2 \, \alpha_A^2 + \sum_{A=1}^{t} \sum_{B=1}^{t} W_A W_B \, \gamma_{AB} \tag{7}$$

where,

W_A : weight of asset A.
W_B : weight of asset B.
αA : standard deviation of asset A.
αB : standard deviation of asset B.
γ_{AB} : covariance of asset A against asset B.

The formulation of the standard deviation is

$$\alpha_v = \sqrt{\frac{1}{t-1}\sum_{i=1}^{t}[R_i - E(R_i)]^2} \tag{8}$$

The covariance calculation is done with

$$Cov(R_A, R_B) = \gamma_{AB} = \sum_{i=1}^{t} \frac{\left[\left(R_{A,i}-E(R_A)\right)\left(R_{B,i}-E(R_B)\right)\right]}{t-1} \tag{9}$$

where,

$R_{A,i}$: Actual return from investment in stock $R_{A,i}$
$R_{B,i}$: Actual return from investment in stock $R_{B,i}$

Markowitz Model

Harry Markowitz (1952) has developed a basic model of modern portfolio theory, based on a problem related to rational investor behavior. Markowitz uses profit fluctuation as an investment risk. Markowitz measures risk by using analysis of variance, then portfolio theory is further developed with a zero-correlation approach. By using the mean-variance method from Markowitz, stocks with correlations of less than +1 will reduce portfolio risk. The more recurrence included in a portfolio, the smaller the portfolio risk. The portfolio with the smallest variance in the set that can be achieved has weight as (Bodie, Kane, & Marcus, 2014).

$$W = \frac{\mu C^{-1}}{\mu C^{-1} \mu^{T}} \tag{10}$$

where,

μ : one-on-one matrix.

C^{-1} : the inverse covariance matrix of stocks.

μ^T : transpose one-on-one matrix.

The covariance matrix of asset as follows:

$$C = \begin{pmatrix} Cov(R_A, R_A) & Cov(R_A, R_B) \\ Cov(R_A, R_B) & Cov(R_B, R_B) \end{pmatrix} \tag{11}$$

where, $Cov(R_A, R_A)$ is the covariance of the average return of stock A to stock A, $Cov(R_A, R_B)$ is the covariance of the average return of stock A to stock B, and the one-on-one matrix is:

$$\mu = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \tag{12}$$

Accordingly, the formula of the expected return and the standard deviation in a portfolio is:

$$\mu v = m W^t \tag{13}$$

where, the expected return matrix as:

$$m = (E(R_A) \quad E(R_B) \quad \cdots \quad E(R_n)) \tag{14}$$

Portfolios with the smallest variance (Minimum Variance Line) among portfolios that can be achieved with the expected return μv have the weight as:

$$W = \frac{\begin{vmatrix} 1 & \mu C^{-1} m^T \\ \mu v & mC^{-1} m^T \end{vmatrix} \mu C^{-1} + \begin{vmatrix} \mu C^{-1} u^T & 1 \\ mC^{-1} u^T & \mu v \end{vmatrix} mC^{-1}}{\begin{vmatrix} \mu C^{-1} u^T & \mu C^{-1} m^T \\ mC^{-1} u^T & mC^{-1} m^T \end{vmatrix}} \tag{15}$$

Sharpe Ratio (SR)

SR was developed by Sharpe in 1966 while conducting a capital market analysis. This analysis is carried out by showing the expected return per total risk (Rana & Akhter, 2015). Therefore, SR explains the compensation that will be borne by investors by assuming additional risk to reflect investment security performance through a measure of return to volatility. (Bodie, Kane, & Marcus, 2014) Mathematically can be written as:

$$\text{Sharpe Ratio} = \frac{\text{Risk Premium}}{\sigma(\text{excess return})} = \frac{E(R_p) - r_f}{\sigma_p} \tag{16}$$

where,

$E(R_p)$: expected portfolio return

r_f : risk free interest rate

σ_p : minimum portfolio risk.

Risk Parity (RP)

RP is used in portfolio management to focus on risk allocation (volatility), not capital allocation. The RP approach advocates that when asset allocation is adjusted to the same level of risk, the RP portfolio can reach a higher SR and become more resilient to market declines than traditional portfolios (Lee, 2014). This approach to building an RP portfolio is similar to that which creates a minimum variance portfolio which refers to the constraint that each asset contributes equally to overall volatility. RP means that each asset (single stock, asset class, equity sector) has the same contribution to total portfolio risk.

Expected Shortfall and Conditional Value at Risk

Expected Shortfall (ES) and Conditional Value at Risk (CVaR) are measures of risk control. ES is a risk measure used to evaluate the market risk or credit risk of a portfolio. Whereas, CVaR helps to estimate the magnitude of the estimated losses on very bad days (Zhou, Jiang, Yu, & Niu, 2016). On the other hand, ES is also often expressed as a possible portfolio risk to be borne when VaR exceeds a reasonable limit (Bodie, Kane, & Marcus, 2014).

METHODOLOGY

In order to build an optimal portfolio of stocks, researchers applied the clustering method to the historical daily prices of the LQ45 index for the past 6 years (from January 2015 to September 2021), which is obtained from the https://finance.yahoo.com/. Calculations and modeling in this research are performed using the newest R language (R 4.1.1, 2021).

Flowchart

More detail about the methodology of this research is showing in Fig. 1. Flowchart of Portfolio Optimization Model.

Systematics for Portfolio Optimization Proses

The process of forming a portfolio optimization that has been carried out in this study, briefly explained in the following section:
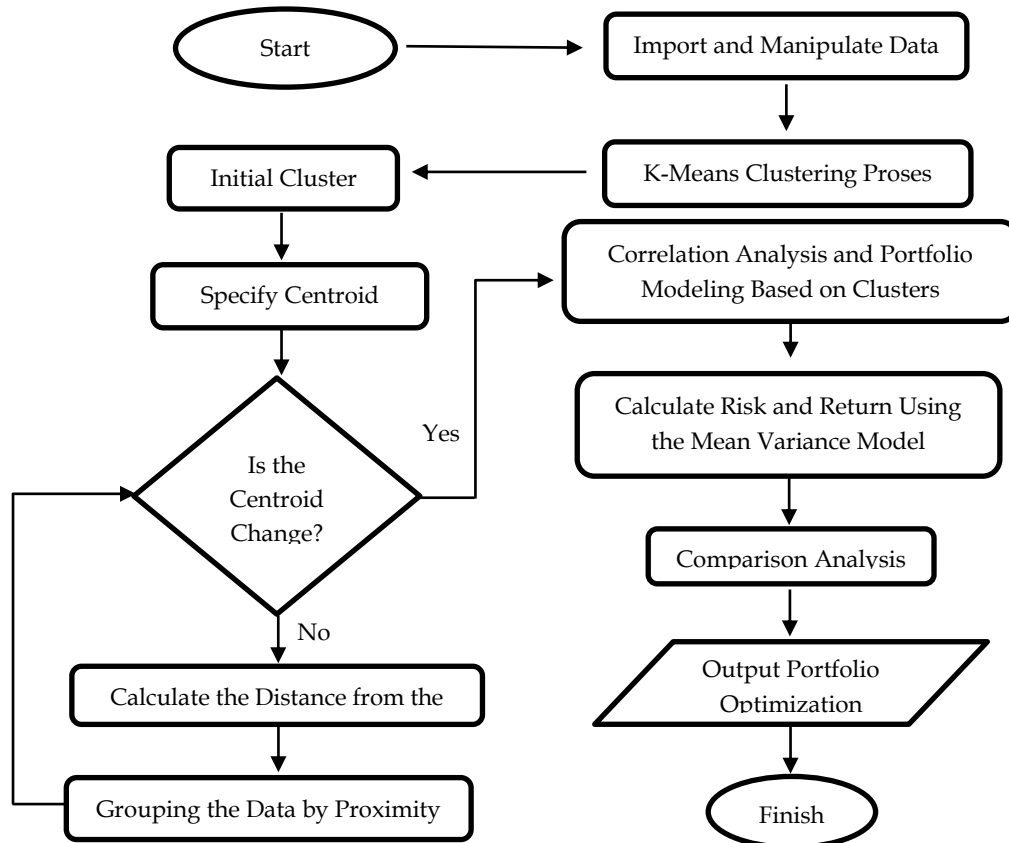
Fig. 1. Flowchart of Portfolio

i. Import and manipulating data, performed as a data frame, and managing missing values to align the time series objects.
ii. Applied K-Means clustering to classifies each stock according to the historical price in the LQ45 index, pay attention to the charts of stock price tend to be increase time by the time.
iii. Perform correlation analysis for each cluster and define specifications of the MV Markowitz model by the default settings. In this section, the researcher also considers the risk-free rate when we calculate the tangency portfolio and the Sharpe ratio when we calculate the whole efficient frontier (Zhou, Jiang, Yu, & Niu, 2016).
iv. In comparison analysis, researchers also combine each cluster in one portfolio by choosing the potential stocks according to the optimal portfolio result in step 3.
v. Output of portfolio optimization

RESULTS

Investors who want to invest stocks in a portfolio format need to find which stocks can provide the least risk with a certain rate of return. The formal model for creating a portfolio that provides minimal risk with a certain return, was developed by Harry Markowitz, but how to select stocks to the model is not diversified in detail. Therefore, this research applied clustering techniques to facilitate researchers and investors in analyzing time series data specifically in the stock market. Classically, investors tend to choose stocks that have prices continuously increase time after time for their portfolios. But the problem it is very difficult to analyze each stock price of the LQ45 index in a simultaneous way as showed in Fig. 2. Accordingly, we selected an appropriate number of clusters for the classification stocks using the K-Means clustering should be in 3 clusters refer to Elbow method as we can see Fig. 3
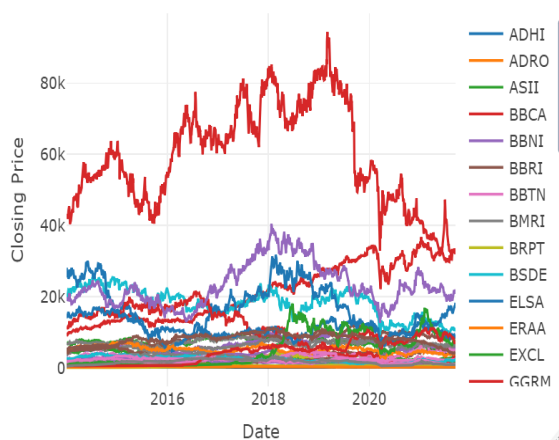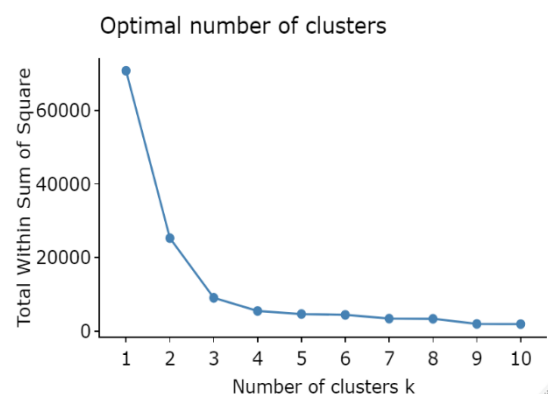

Fig. 2. The historical

Fig. 3. Elbow method

To convince the hypothesis about clusters that must be in 3 clusters, then we do cluster analysis using the principal component. Fig. 4. clearly performed that 93% accuracy of the cluster has been established. Fig. 5 is the result of the process of applying clustering techniques. In this section, it has been very helpful to do a preliminary analysis to select stocks. For instance, if you are a new investor with limited money, of course, you will easily choose the right stock to put in your portfolio pocket.
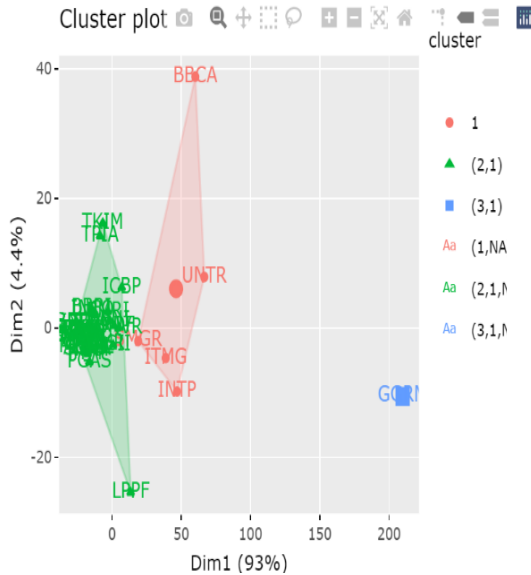

Fig. 4. Principal Component
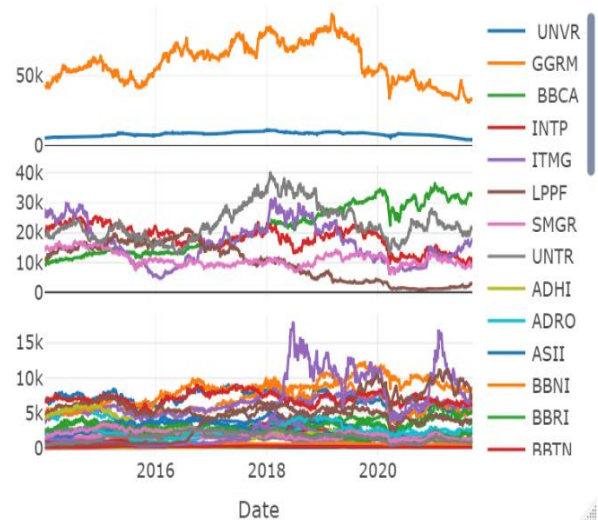
Fig. 5. Clustering by the Prices

Next, to ensure that each cluster does not have a strong correlation, we do correlation analysis for cluster 1, UNVR and GGRM do not show in a figure but we know they have a correlation is about 76.61%. This means they probably have no correlation, and it might be important to consider in detail. Cluster 2 in Fig.6, performs SMGR and ITMG 77% which means highly have a strong correlation. Therefore, we have an initial assumption it is not suitable to form in one portfolio.
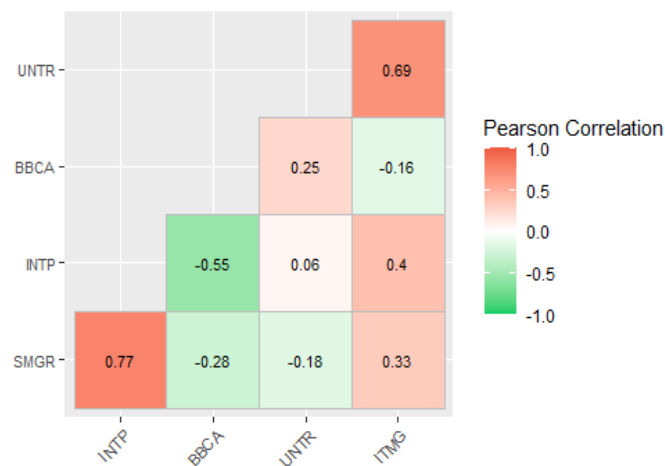
Fig.6. Correlation Cluster 2

Cluster 3 in Fig.7, there is a lot of information that must be further investigated. In this section I want to highlight stocks categorized as banking, from Fig. 7. we argue that they have a high correlation with each other. Meaning, if you want to form a portfolio then you should not unite them in your portfolio.
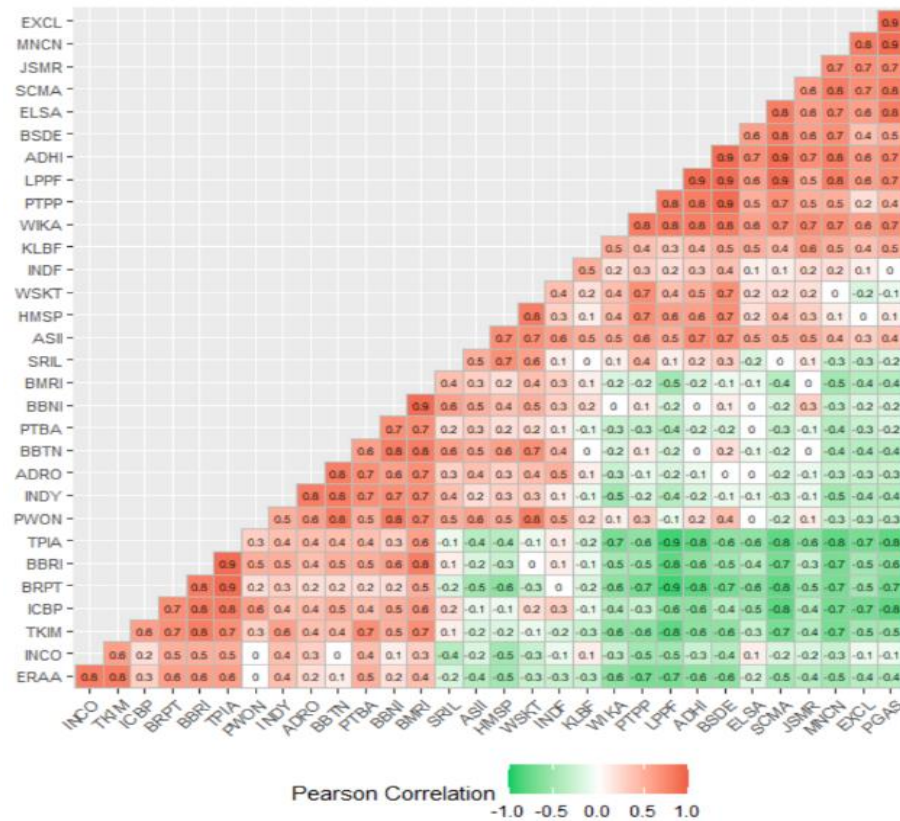
Fig.7. Correlation Cluster 3

In this section, we begin to consider the monthly return of cluster 2 in Appendix B. Then we can calculate the target return and risk at Fig. 8. ("the efficient frontier should be above of the red dot"). Next, define specifications of the MV Markowitz model by the default settings as can bee seen in Fig. 9.
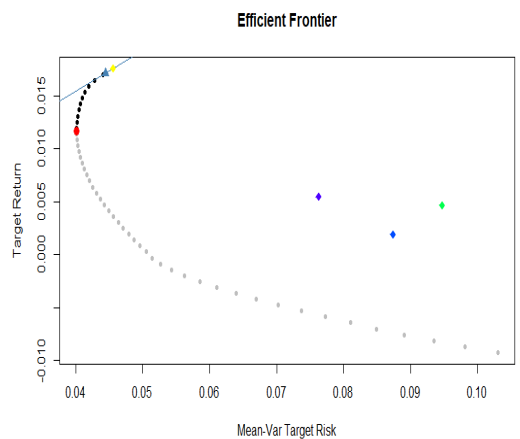
Fig. 8. Target Return and Risk
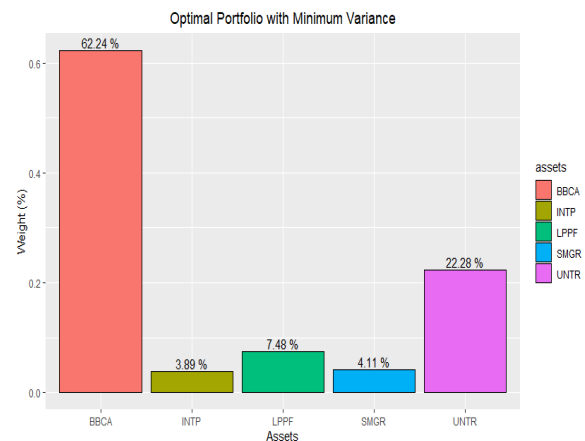
Fig. 9. Optimal Portfolio

Note: Here, the focus is on the "long-term" investment.

Now, let consider the risk-free rate of the portfolio by calculating the tangency portfolio and the Sharpe ratio using the whole efficient frontier. The result is showing in Fig. 10.
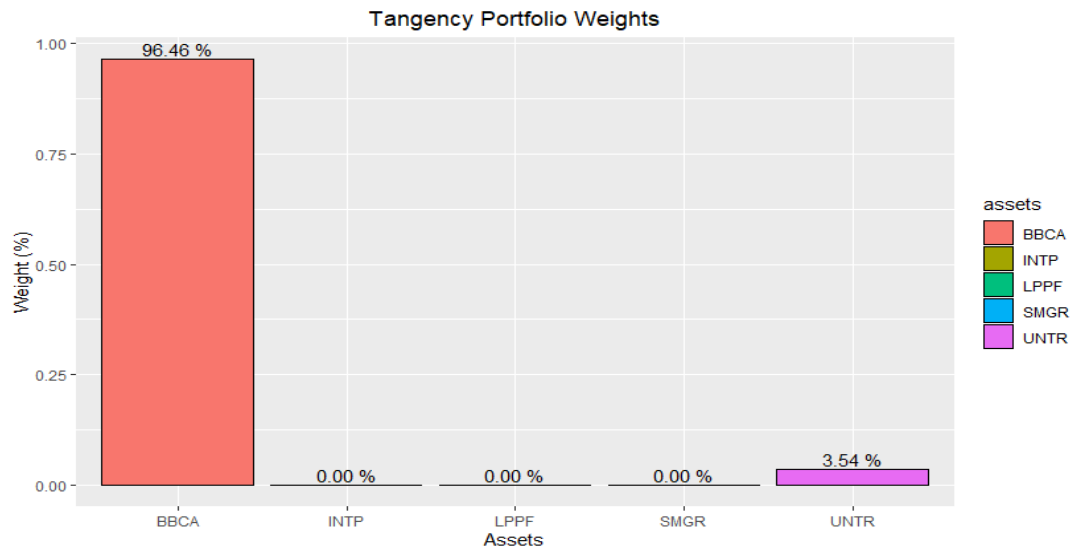

Fig. 10. Tangency Optimal Portfolio Cluster 2

According to the result of cluster 2, we do the same thing to cluster 3. For this cluster 3, we choose stocks with no correlation in our portfolio. Let consider the result in Fig. 11-12.
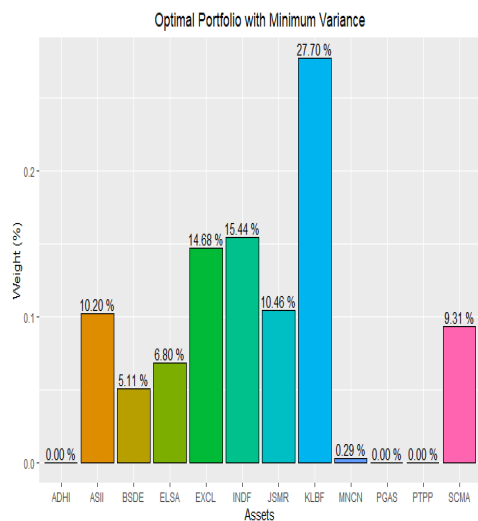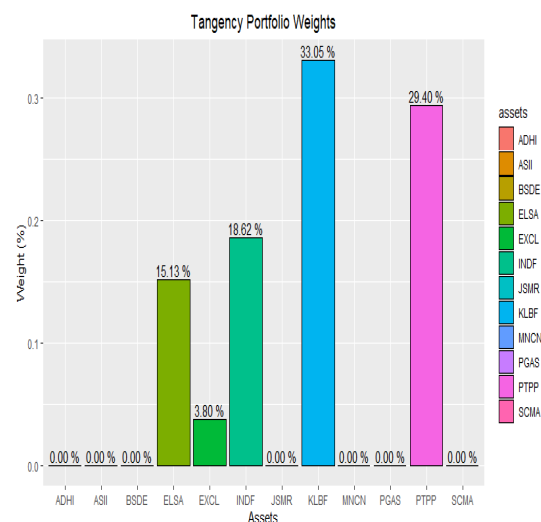

Fig. 11. Optimal Portfolio   Fig. 12. Tangency Optimal Portfolio

To summarize this discussion, we do a comparative analysis to combine the best of cluster 2 (BBCA, UNTR) and cluster 3 (ELSA, EXCL, INDF, KLBF, PTPP) in our portfolio. Then, we follow all the processes of forming a portfolio optimization from steps 1-5 (the result in Fig. 13-14).
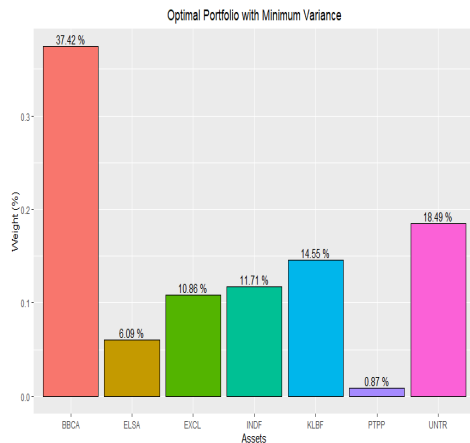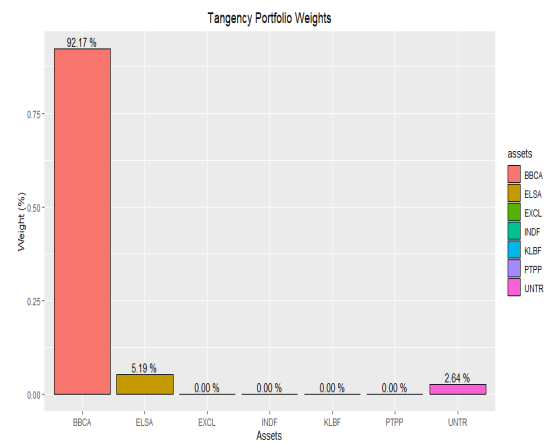
Fig. 13. Optimal Portfolio

Fig. 14. Tangency Optimal Portfolio

CONCLUSIONS

This K-Means clustering is highly recommended to select the stocks that you want to arrange in the portfolio. Especially, if you are a new investor with limited money in your pocket, you could easily choose the right stock to put in your portfolio. Then, to ensure that each stock of your portfolio does not have a strong correlation you can do correlation analysis. Moreover, you can approach Markowitz to analyze the performance of optimization portfolio models that have a minimum variance in expected return and risk. In addition, do not forget to consider the risk-free rate of the portfolio by calculating the tangency portfolio and the Sharpe ratio using the whole efficient frontier. Please note that as an investor you are also expected to do more analysis for each stock that you will obtain in the portfolio.

RECOMMENDATIONS AND FURTHER STUDY

In this research, K-Mean Clustering is used to classify the stock price fluctuations of LQ45 Index (January 2015-September 2021). It is possible to do with other methods or different data set (Index). For further research, it can be developed with more complex data set (other market stock indexes and other external factors). you might also do further research by comparing several clustering methods at the stage of determining the appropriate cluster for the LQ45 stock.

ACKNOWLEDGMENT

REFERENCES

I. Kulali. (2016). Portfolio Optimization Analysis with Markowitz Quadratic Mean-Variance Model. *European Journal of Business and Management*. https://iiste.org/Journals/index.php/EJBM/article/view/29452/30242

F. Mayanja, S. Mataramvura and W. Charles. (2013). A Mathematical Approach to a Stocks Portfolio Selection: The Case of Uganda Securities Exchange (USE). *Journal of Mathematical Finance*. https://www.scirp.org/journal/paperinformation.aspx?paperid=40096

H. M. Markowitz. (1952). Portfolio Selection. *The Journal of Finance*. https://doi.org/10.2307/2975974

Q.Xu, Y. Zhou, C. Jiang, K. Yu, & X. Niu. (2016). A Large CVaR-Based Portfolio Selection Model with Weight Constraints. *Economic Modelling*. https://ideas.repec.org/a/eee/ecmode/v59y2016icp436-447.html

Ayu Trimulya, Syaifurrahman, & F. A. Setyaningsih. (2015). Implemtasi Jaringan Syaraf Tiruan Metode Backpropagation Untuk Memprediksi Harga Saham. *Jurnal Coding*. http://dx.doi.org/10.26418/coding.v3i2.10784.

S. R. Nanda, B. Mahanty, & M. K. Tiwari. (2010). Clustering Indian stock market data for portfolio management. *Expert Systems with Applications*. https://doi.org/10.1016/j.eswa.2010.06.026

Yuli Kurniyati. (2007). Analisa Portofolio Optimal Di Bursa Efek Jakarta Dengan Menggunakan Indeks Beta. *Thesis of Universitas Negeri Semarang*. http://lib.unnes.ac.id/id/eprint/1786

Marek Capinski, & Tomasz Zastawniak. (2003). *Mathematics for Finance: An Introduction to Financial Engineering.* Springer-Verlag London Limited.

Z. Bodie, A. Kane, & A. J. Marcus. (2014). *Investments (10th ed, global edition)*, Berkshire, Mc Graw Hill Education.

M. E. Rana, & W. Akhter. (2015). Performance of Islamic and Conventional Stock Indices: Empirical Evidence from an Emerging Economy. *Financial Innovation*. https://doi.org/10.1186/s40854-015-0016-3

W. Lee. (2014). Constraints and Innovations for Pension Investment: The Cases of Risk Parity and Risk Premia Investing. *The Journal of Portfolio Management*. https://jpm.pm-research.com/content/40/3/12

BAKTI SIREGAR: Matana University.
E-mails: siregar.bakti@matanauniversity.ac.id

F. ANTHON PANGRURUK: Matana University.
E-mails: antpangruruk@matanauniversity.ac.id