SKEWNESS
Jurnal Statistika, Aktuaria dan Sains Data
Volume 2, No. 1, April 2025

Clustering Jakarta Islamic Index Stocks Using Fuzzy C-Means Clustering and Determining Optimal Portfolio Using Capital Asset Pricing Model

Zidni Ilma Nafi'a [1], Agung Prabowo [2*], Agus Sugandha [3], Diah Paramita Amitarwati [4]

[1,3]*Mathematics Study Program, Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Purwokerto, Indonesia.*

[2]*Statistics Study Program, Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Jenderal Soedirman, Purwokerto, Indonesia.*

[4]*Sharia Economics Study Program, Faculty of Islamic Economics and Business, Universitas Islam Negeri Prof. K.H. Saifuddin Zuhri, Purwokerto, Indonesia.*

Corresponding e-mail: agung.prabowo@unsoed.ac.id

Abstract. Investment is an activity of investing fund or capital to earn profits in the future. One investment instrument that many people prefer is stocks, especially those included in stock market index, such as Jakarta Islamic Index (JII). To maximize returns and minimize risks, portfolio is formed. Stock clustering by their financial performance can be done using the so-called Fuzzy C-Means Clustering, a method to cluster data based on membership degrees. This research aims at clustering the stocks included in JII using the Fuzzy C-Means Clustering algorithm. Four financial ratios, namely Return on Assets, Current Ratio, Debt to Equity Ratio, and Total Asset Turnover, were employed to do this. The resulting clustering was then used to form an optimal portfolio using a method known as Capital Asset Pricing Model (CAPM). This model aimed to calculate the expected returns and risks of each portfolio. Two methods were used in this research, they were literature review and a case study in the stocks included in JII. Based on the Silhouette coefficient value, this research generated three optimal clusters of stocks included in JII. The portfolio made using CAPM method produced INKP and UNVR stocks in cluster 1, at an investment weight of 5% and 95% respectively. In cluster 2, the stocks of ACES, ADMR, EMTK, ICBP, ITMG, and MIKA formed a portfolio with investment weights of 1%, 1%, 1%, 1%, 1%, and 95% respectively. In cluster 3, the stocks of ADRO, AKRA, ANTM, ASII, BRIS, BRMS, CPIN, EXCL, INDF, INTP, PGAS, SMGR, and UNTR, formed a portfolio with investment weights of 6.66%, 2.01%, 4.95%, 20.86%, 7.65%, 10.68%, 6.69%, 5.64%, 21.08%, 3.14%, 1.91%, 3.98%, and 4.75% respectively. The portfolio performance was evaluated using Sharpe ratio and it led to negative results, i.e., -0.2501 for cluster 1, -0.0405 for cluster 2, and -0.0587 for cluster 3. It is, therefore, recommended that investors invest more in the cluster 2 portfolio or choose risk-free assets instead.

Keywords: *investment, jakarta islamic index, fuzzy c-means clustering, capital asset pricing model, optimal portfolio.*

1 Introduction

According to [1], investment is where a certain amount of money or other resources being committed today while expecting of gaining benefits in the future. The advancement of communication and information technology in the current globalization era makes

investment no longer something new to the society. Stock investment in capital market is one of ways to earn profits without having to exert too much energy and can be done anywhere as long as one is connected to the Internet. Stock index is one of important indicator for investors to place their investment [2]. Jakarta Islamic Index (JII) can be used for investors' consideration when deciding on investing in a capital market [Hartono]. JII is also an index of choice in terms of the liquidity and financial condition of its companies, which have promising performance development for periods to come [3]. Clustering is a method of grouping data into some clusters based on their characteristic commonality [4],[5]. One of soft clustering methods is Fuzzy C-Means Clustering [6]. This method allows stock clustering by certain levels or degrees of membership [7]. This method provides greater flexibility in analyzing the stock data and providing a more realistic clustering in determining the optimal portfolio [7]. The resulting stock clustering could be used to minimize risks and maximize returns when investing by forming a portfolio. One method to analyze a portfolio is Capital Asset Pricing Model (CAPM) [8]. This model aims to determine the *expected return* rate in minimizing investment risks [9],[10]. Having discussed above, this research aims to cluster JII stocks using the Fuzzy C Means Clustering algorithm and form a portfolio using the Capital Asset Pricing Model (CAPM) method.

2 Materials and Methods

The methods used in this research were literature review and case study. In literature review, the author learned, comprehended, and analyzed textbooks, journals, and other references regarding Fuzzy C-Means Clustering and Capital Asset Pricing Model (CAPM). Meanwhile, in a case study, the author used those companies belonging to JII index for 1 December 2023 to December 2024 period to cluster those companies based on their financial ratios using Fuzzy C-Means Clustering. Later on, the resulting clustering was used to form a portfolio using Capital Assets Pricing Model (CAPM) by employing the data on stock return rates that could be classified as an optimal cluster.

The data analysis was also assisted by some software, i.e., Microsoft Excel and RStudio. The steps taken in this research involved:

1. Collecting the data on current ratio, DER, TATO, and ROA in JII stock index based on 2023 financial statements in IDX's site at www.idx.co.id.

2. Clustering the data using the Fuzzy C-Means Clustering method.
   a. Testing the cluster analysis assumption
      1) Sample adequacy assumption test
      2) Non-multicollinearity assumption test
   b. Clustering process using Fuzzy C-Means [7], [11], [12],
      i. Inputting the data to matrix $\boldsymbol{X}$ of $n \times h$ size, where n was the number of data to be formed into clusters and h was the number of variables.
      ii. Determining the initial parameters, i.e., the number of cluster c, parameters of fuzzy m, maximum iteration, expected error ε, and determining the initial objective function $P_0 = 0$, and initial iteration $q = 1$.
      iii. Forming initial partition matrix $\boldsymbol{U}^0$, which was used to initiate the membership degree of each data to every cluster.
      iv. Calculating the central elements of cluster **V** for every cluster using the following equation [13] [14].

$$v_{dj} = \frac{\sum_{i=1}^{n} u_{ij}^{m} x_{id}}{\sum_{i=1}^{n} u_{ij}^{m}} \quad (1)$$

      which was used to determine the average position of each cluster.

      v. Calculating the objective function in the q-th iteration using equation [15].

$$P_q = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} dist^2(x_{id}, v_{dj}) \quad (2)$$

      which was used to measure the total weighted distance between the data and cluster center [14].

      vi. Calculating the change in membership degree of every data in each cluster in the the q-th iteration using equation [15].

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left(\frac{\sum_{d=1}^{h} dist(x_{id}, v_{jd})}{\sum_{d=1}^{h} dist(x_{id}, v_{kd})} \right)^{\frac{2}{m-1}}} \quad (3)$$

      to update the cluster center.

      vii. Investigating the stop condition

      If $\left|P_q - P_{q-1}\right| < \varepsilon$ or $q >$ maximum iteration, then it stopped.

      If $\left|P_q - P_{q-1}\right| > \varepsilon$ or $q <$ maximum iteration), then $(q = q + 1)$, recalculate the cluster center.

3. Validating the number of clusters using Silhouette Coefficient [16].
4. Compiling the optimal portfolio using Capital Asset Pricing Model (CAPM).
5. Evaluating the portfolio performance using Sharpe ratio.

3 Results and Discussion

The data used in this research consisted of four independent variables which constituted the financial ratios of stocks included in Jakarta Islamic Index. These financial ratios included Current Ratio, Debt to Equity Ratio, Total Assets Turnover, and Return on Assets. The values of these ratios were obtained based on the 2023 financial statements of companies included in the Jakarta Islamic Index (JII) stock index. The description of data characteristics of these variables can be seen in the descriptive statistic of research data in Table 1.

Table 1. Research Data Descriptive Statistics

Variable	*n*	*Minimum*	*Maximum*	*Mean*	*Std. Deviation*
CR	30	0.356	7.415	2.435	1.720
DER	30	0.115	6.880	1.177	1.376
TATO	30	0.042	2.317	0.700	0.482
ROA	30	−1.671	0.288	0.023	0.329

Based on Table 1, current ratio showed the companies' liquidity. The average 2.435 meant that companies had current assets 2.435 times greater than their current liabilities. DER reflected solvability, at an average 1.177, indicating that the companies' debts were less than their equity. TATO measured the efficiency of asset usage, at an average 0.700, showing that the assets were capable of producing income at 70%. ROA showed the companies' profitability, at an average 0.023 or 2.3% .

3.1 Cluster Analysis Assumption Test

1. Sample adequacy assumption test

Table 2. KMO Value Output

Kaiser-Meyer-Olkin factor adequacy			
CR	DER	TATO	ROA
0.50	0.51	0.51	0.50

Based on Table 2, it was known that the KMO values of each variable were more than or equal to 0.5, meaning that the used data were representative to analyze clustering [16].

3. Non-multicollinearity assumption test

Table 3. VIF Value Output

Variable	VIF
Current ratio	1.34039
DER	1.143236
TATO	1.101702
ROA	1.105078

Based on Table 3, it was found that the VIF values of each variable were less than 10, thus it could be concluded that no multicollinearity was found between variables [6].

3.2 Result of Clustering Using Fuzzy C-Means

The Fuzzy C-Means algorithm was applied to form 3 clusters as follows:

1. The data were inputted to matrix $\boldsymbol{X}$ of $n \times h$ size, where n was the number of data to be formed into clusters and h was the number of variables (criteria) [11][12][15]. The provided data consisted of 30 stock data with 4 variables, hence the matrix $\boldsymbol{X}$ of 30×4 size was as follows:

$$\begin{bmatrix} 7.4151 & 0.2551 & 0.9818 & 0.0985 \\ 4.2172 & 0.6733 & 0.6405 & 0.2601 \\ 2.0148 & 0.4524 & 0.6223 & 0.1567 \\ 1.4458 & 1.4394 & 1.3911 & 0.0919 \\ 2.3395 & 0.3813 & 0.9579 & 0.0718 \\ \vdots & \vdots & \vdots & \vdots \\ 1.1795 & 1.1094 & 0.2809 & 0.0374 \end{bmatrix}$$

2. To determine the initial parameter values, $m = 2$ was selected, since this value helped produce more accurate clustering while separating clusters clearly, allowing the data to be interpreted better. The initial parameters used were as follows:

c	: 3
m	: 2
MaxIter	: 1000
ε	: 10^{-6}
initial objective function	: $P_0 = 0$
initial iteration	: $q = 1$

3. To form the initial partition matrix U^0, as many as 3 clusters were to be formed in this discussion. The initial partition matrix U^0 used to initiate the initial membership degree

was of 30 × 3 size, the element of which had conditions $u_{ij} \epsilon [0,1]$ and $\sum_{j=1}^{c} u_{ij} = 1$. Thie initial partition matrix U^0 was formed using the software RStudio. The initial partition matrix U^0 was as follows:

$$\begin{bmatrix} 0.4759 & 0.0054 & 0.5187 \\ 0.1379 & 0.4540 & 0.4081 \\ 0.3210 & 0.1481 & 0.5309 \\ 0.2500 & 0.2891 & 0.4609 \\ 0.8230 & 0.0973 & 0.0797 \\ \vdots & \vdots & \vdots \\ 0.2651 & 0.2651 & 0.4698 \end{bmatrix}$$

4. The cluster center ***V*** was calculated for each center of the *j*-th cluster in the *d*-th variable using equation (1) [14], and the following matrix ***V*** was obtained:

$$\begin{bmatrix} 2.6510 & 2.2690 & 2.2854 \\ 1.1313 & 1.1274 & 1.1372 \\ 0.6034 & 0.8300 & 0.7601 \\ -0.1853 & 0.0883 & 0.0779 \end{bmatrix}$$

5. The objective function value in the 1st iteration was calculated using equation (2) as follows:

$$P_1 = \sum_{i=1}^{30} \sum_{j=1}^{3} u_{ij}^2 dist^2(x_{id}, v_{dj}) = 23.2995 + 19.0392 + 22.9887 = 65.3275$$

The change in membership degree of each data in every cluster was calculated based on the partition matrix change using equation (3) as follows:

Table 4. New Membership Degree Calculation Results

i	u_{i1}	u_{i2}	u_{i3}
1.	0.3648	0.3169	0.3183
2.	0.4132	0.2908	0.2961
3.	0.2240	0.3878	0.3882
4.	0.2025	0.4170	0.3805
5.	0.2591	0.3814	0.3595
⋮	⋮	⋮	⋮
30.	0.2410	0.3748	0.3842

6. Check the stop condition, after calculating the new membership value, then a calculation would be made for the change in P_q value with P_{q-1}. Based on the previous calculation, it could be seen that the objective function value was 65.68175244, thus if $|P_q - P_{q-1}|$, that is $|P_1 - P_0| = |65.3275 - 0| = 65.3275 > 1 \times 10^{-6}$, the result was greater than the error value. Thus, the iteration was continued once again.

Based on the data analysis using Rstudio software and the complete result of clustering process with Fuzzy C-Means algorithm, the iteration process was continuously performed for 57 times until the optimal objective function value was obtained at 30.49125 with the cluster center as follows:

$$\boldsymbol{V} = \begin{bmatrix} 2.2690 & 4.6807 & 1.4235 \\ 6.0092 & 0.3702 & 1.0923 \\ 0.6876 & 0.6736 & 0.6904 \\ 0.0739 & 0.0848 & 0.0306 \end{bmatrix}$$

Then, the new membership degree was:

$$\boldsymbol{U} = \begin{bmatrix} 0.0938 & 0.7510 & 0.1552 \\ 0.0097 & 0.9503 & 0.0400 \\ 0.0215 & 0.0965 & 0.8820 \\ 0.0250 & 0.0471 & 0.9279 \\ 0.0333 & 0.1963 & 0.7704 \\ \vdots & \vdots & \vdots \\ 0.0085 & 0.0171 & 0.9744 \end{bmatrix}$$

The result of stock clustering with members of each cluster is as follows:

- Cluster 1: INKP and UNVR stocks
- Cluster 2: ACES, ADMR, EMTK, ICBP, INCO, ITMG, KLBF, MIKA, and TPIA stocks
- Cluster 3: ADRO, AKRA, ANTM, ASII, BRIS, BRMS, CPIN, EXCL, GOTO, INDF, INTP, MAPI, MDKA, PGAS, PTBA, SMGR, TLKM, UNTR, and WIFI stocks

The silhouette coefficient value in the formation of 3 clusters was 0.7366, indicating that the formed cluster distribution had a clear and good separation between one cluster and another.

3.3 Portfolio Preparation

The result of clustering was used to prepare the optimal portfolio using CAPM:

- Cluster 1: INKP and UNVR stocks with investment weight of 5% and 95% respectively.
- Cluster 2: ACES, ADMR, EMTK, ICBP, ITMG, and MIKA stocks with investment weight of 1%, 1%, 1%, 1%, 1%, and 95% respectively.
- Cluster 3: This cluster had 13 stocks, namely ADRO, AKRA, ANTM, ASII, BRIS, BRMS, CPIN, EXCL, INDF, INTP, PGAS, SMGR, and UNTR with investment weight of 6.66%, 2.01%, 4.95%, 20.86%, 7.65%, 10.68%, 6.69%, 5.64%, 21.08%, 3.14%, 1.91%, 3.98%, and 4.75% respectively.

3.4 Portfolio Performance Evaluation

Sharpe Ratio showed negative results for all clusters:

Cluster 1: -0.2501

Cluster 2: -0.0405

Cluster 3: -0.0587

This result showed that the risks taken were greater than the resulting expected return. In this situation, investors are suggested to invest more in the cluster 2 portfolio or choose risk-free assets such as bonds or deposits that give higher return.

4 Conclusion

Based on the result and discussion, it could be concluded that, from forming the clusters of stocks registered in JII with Fuzzy C-Means, 3 optimal clusters were obtained based on the silhouette coefficient value. Members of each cluster were INKP and UNVR in cluster 1; ACES, ADMR, EMTK, ICBP, INCO, ITMG, KLBF, and MIKA in cluster 2; and ADRO, AKRA, ANTM, ASII, BRIS, BRMS, CPIN, EXCL, GOTO, INDF, INTP, MAPI, MDKA, PGAS, PTBA, SMGR, TLKM, UNTR, and WIFI in cluster 3.

Based on the formed clustering results, a portfolio was made using Capital Asset Pricing Model. Cluster 1 had INKP and UNVR stocks with investment weight of 5% and 95% respectively. Cluster 2 had ACES, ADMR, EMTK, ICBP, ITMG, and MIKA stocks with investment weight of 1%, 1%, 1%, 1%, 1%, and 95% respectively. Finally, cluster 3 had ADRO, AKRA, ANTM, ASII, BRIS, BRMS, CPIN, EXCL, INDF, INTP, PGAS, SMGR, and UNTR stocks with investment weight of 6.66%, 2.01%, 4.95%, 20.86%, 7.65%, 10.68%, 6.69%, 5.64%, 21.08%, 3.14%, 1.91%, 3.98%, and 4.75% respectively.

References

[1] Tandelilin, E. (2001). *Analisis investasi manajemen portofolio* (Cet. 1). Yogyakarta: BPFE.

[2] Bodie, Z., Kane, A., & Marcus, A. J. (2014). *Investments* (10th ed.). New York, NY: McGraw-Hill Education.

[3] Hartono, J. (2013). *Teori portofolio dan analisis investasi* (Edisi ke-8). Yogyakarta: BPFE.

[4] Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

[5] Faujia, R. A., Setianingsih, E. S., & Pratiwi, H. (2022). Analisis klaster K-means dan agglomerative nesting pada indikator stunting balita di Indonesia. *Seminar Nasional Official Statistics*, *2022*(1), 1249–1258.

[6] Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). *Multivariate data analysis* (8th ed.). England: Pearson Prentice.

[7] Aliansa, W., & Saputra, R. A. (2023). Perbandingan harga saham perbankan dengan metode fuzzy C-means clustering. *Prosiding Seminar Nasional Pemanfaatan Sains dan Teknologi Informasi*, *1*(1), 369–376.

[8] Sharpe, W. F., Alexander, G. J., & Bailey, J. V. (1999). *Investments* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

[9] Adnyana, I. M. (2020). *Manajemen investasi dan portofolio*. Jakarta: Lembaga Penerbitan Universitas Nasional.

[10] Aunillah, M. W., & Wahyudi, W. (2022). Analisis portofolio optimal CAPM dan single index model pada perusahaan IDX30. *Jurnal Ilmiah Ekonomi Islam*, *8*(2), 2231–2240.

[11] Höppner, F., Kruse, R., Klawonn, F., & Runkler, T. (1999). *Fuzzy cluster analysis: Methods for classification, data analysis and image recognition*. Chichester: John Wiley & Sons, Ltd.

[12] Lilis, A. (2022). *Pengelompokan saham dengan fuzzy C-means clustering untuk menentukan portofolio optimal indeks saham IDX30* (Skripsi, Universitas Diponegoro).

[13] Shihab, A. I. (2000). *Fuzzy clustering algorithms and their application to medical image analysis* (Disertasi, University of London).

[14] Utami, D. S. (2018). *Clustering pada data saham indeks Liquates 45 (LQ45) dengan algoritme 2-means dan 2-medoids* (Skripsi, Universitas Sebelas Maret).

[15] Kusumadewi, S., Hartati, S., Harjoko, A., & Wardoyo, R. (2006). *Fuzzy multi-attribute decision making*. Yogyakarta: Graha Ilmu.

[16] Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. New York, NY: Plenum Press.