Jurnal Matematika Integratif.
Vol. 22, No. 1 (2026), pp. 59–72.

p-ISSN:1412-6184, e-ISSN:2549-903
doi:10.24198/jmi.v22.n1.69287.59-72

Clustering of Banking Sector Stocks using Integration of Fourier Transform, Spectral Clustering, and Fuzzy C-Means as a Basis for Mean-Variance Portfolio Optimization

Dimitri Salsabila Fakhriyah Ricardo, Nurul Gusriani, and Herlina Napitupulu

Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran
Jl. Raya Bandung Sumedang KM.21, Hegarmanah, Kec. Jatinangor, Kabupaten Sumedang, Jawa Barat 45363, Indonesia
Email: *dimitri22001@mail.unpad.ac.id

Abstract

The Indonesian capital market has experienced significant growth accompanied by high volatility, particularly in the banking sector which holds a substantial contribution to market capitalization. Extreme volatility during the 2019-2024 period triggered by the impact of the COVID-19 pandemic, economic recovery phases, and global macroeconomic challenges has created complexities in investment decision-making and portfolio optimization, which often produces unstable solutions under uncertain market conditions. Various studies have applied frequency domain analysis to uncover hidden patterns in stock price movements or clustering to group stocks based on their characteristics to support optimal investment decision-making, however the integration of these two approaches remains limited in its application to generate robust portfolio optimization solutions in the Indonesian capital market. This study aims to generate clustering of banking sector stocks through the integration of Fourier Transform, spectral clustering, and Fuzzy C-Means and to construct an optimal portfolio using the Mean-Variance method based on the clustering results. This study uses closing price data of 41 banking sector stocks on the Indonesia Stock Exchange for the 2019-2024 period through an integrated approach of Fourier Transform to extract frequency patterns, spectral clustering as a basis for grouping, Fuzzy C-Means to generate cluster membership degrees, and Mean-Variance for portfolio optimization. The results show that the integration of these methods produces four optimal stock clusters consisting of nine stocks with a medium risk-low return profile, six stocks with a high risk-high return profile, fifteen stocks with a low risk-low return profile, and eleven stocks with a medium risk-high return profile. Based on the clustering results, four representative stocks were selected from each cluster for portfolio optimization, resulting in an optimal portfolio at a risk aversion value of $\rho = 6.83$ *with a portfolio ratio of* 3.4128877*. This optimal portfolio is constructed from four representative stocks with weight allocations of* 11.48% *BMAS,* 11.76% *ARTO,* 72.08% *BNGA, and* 4.68% *BBHI, with an expected return value of 0.0263613 and a portfolio variance of* 0.0077241*.*

2020 Mathematics Subject Classification: 91B28, 62H30
Received: 15-01-2026, accepted: 02-02-2026.

Keywords: *Banking Stocks, Clustering, Fourier Transform, Spectral Clustering, Fuzzy C-Means, Mean-Variance.*

1. Introduction

The capital market plays a central role as a pillar of the financial system and a reflection of a country's economic dynamics [1]. In Indonesia, the capital market has shown significant growth, where the number of listed companies and market capitalization on the Indonesia Stock Exchange (IDX) increased sharply during the 2019-2024 period [2, 3]. Nevertheless, this growth has been accompanied by a fundamental challenge in the form of extreme stock price fluctuations, reflecting non-stationary market conditions. This high volatility has become more pronounced due to significant external factors, such as the impact of the COVID-19 pandemic, the economic recovery phase, and global macroeconomic challenges. These uncertain market conditions create great complexity and high risk in investment decision-making and effective portfolio optimization, especially in key sectors that dominate the market.

Among the various listed sectors, the banking sector holds a central position and leads the IDX market capitalization, accounting for more than 33% of the total market capitalization [4]. The dominance of the banking sector makes it highly sensitive to strict regulations, digital transformation, and changes in macroeconomic variables. The complex dynamics that have occurred, particularly in the 2019–2024 period, have caused extreme volatility in banking stocks. Therefore, a more comprehensive analytical approach is needed to uncover hidden price movement patterns in order to support more optimal investment decision-making.

Various approaches have been developed to improve portfolio performance. Lyle and Yohn [5] integrated fundamental analysis into the Mean-Variance framework. However, this approach produced unstable solutions in non-stationary market environments. In response to this limitation, Scalzo et al. [6] shifted optimization to the frequency domain to address the non-stationarity issue, which proved to be superior. The utilization of this frequency domain analysis was then integrated with K-Means clustering by[7], but the resulting instability indicated the need for a more suitable clustering method. Gubu et al. [8] subsequently demonstrated that Fuzzy C-Means exhibits superior performance compared to K-Means in generating stable portfolios due to its ability to provide partial membership degrees, although its integration with frequency domain analysis remains limited.

Based on this literature review, frequency domain analysis is effective in uncovering hidden patterns and addressing non-stationarity, but its integration with clustering as a basis for portfolio optimization still yields less stable solutions. Meanwhile, membership-based clustering such as Fuzzy C-Means excels in producing stable portfolios, yet its integration with frequency domain analysis is still limited. This indicates that both approaches have advantages, but their integration remains limited in its application to generate robust portfolio optimization solutions in the Indonesian capital market. This research develops a hybrid model that integrates the Fourier Transform to extract frequency patterns, spectral clustering as a basis for grouping, Fuzzy C-Means to generate cluster membership degrees, and Mean-Variance for portfolio optimization. This model is applied to banking sector stocks on the IDX for the 2019–2024 period to form a more stable and diversified portfolio based on stock grouping that considers risk-return characteristics and hidden patterns in the Indonesian capital market.

2. Research Methods

2.1. **Stock.** A stock is a capital market instrument representing ownership in a company that grants its holder claims to dividends, assets, and voting rights in the General Meeting of Shareholders (GMS), while also carrying the potential for gains and losses [9, 10]. The performance of this stock ownership is reflected in stock return, which represents the rate of return on investment from price differences (capital gain/loss) and dividend payments. Mathematically,

stock return can be calculated using equation (1),

$$R_{i,t} = \frac{P_{i,t} - P_{i,(t-1)}}{P_{i,(t-1)}} \tag{1}$$

where $R_{i,t}$ is the return of the i-th stock in period t, $P_{i,t}$ is the closing price of the i-th stock in period t, $i = 1, 2, \ldots, n$ is the stock index with n being the number of stocks, and $t = 1, 2, \ldots, T$ is the period index with T being the number of data periods.

2.2. **Fast Fourier Transform.** Fast Fourier Transform (FFT) is an efficient algorithm for computing the *Discrete Fourier Transform* (DFT), which functions to transform data from the time domain to the frequency domain. DFT is the discrete form of the *Fourier Transform*, a mathematical transformation method that analyzes data using sine and cosine functions to represent it in the form of complex numbers [11].

The FFT algorithm overcomes the computational limitations of the conventional DFT, which has a complexity of $O(T^2)$, by reducing the complexity to $O(T \log T)$, where T is the number of data periods. The FFT calculation uses the DFT formula as its basis. For a function $x(t^*)$ in the time domain with T observation periods, the DFT is defined as equation (2),

$$X(k) = \sum_{t^*=0}^{T-1} x(t^*) \cdot e^{-I\frac{2\pi t^* k}{T}} \tag{2}$$

where $X(k)$ is the DFT coefficient at the k-th frequency, $x(t^*)$ is the function in the time domain at the t^*-th time index, $t^* = 0, 1, \ldots, T-1$ is the time index, $k = 0, 1, \ldots, T-1$ is the frequency index, and I is the imaginary unit.

2.2.1. *Fast Fourier Transform Algorithm.* The Fast Fourier Transform algorithm is applied to the stock return data $R_{i,t}$ to extract hidden patterns in the frequency domain. The process begins with demeaning the return data to remove the constant mean component using equation (3),

$$\tilde{R}_{i,t} = R_{i,t} - \bar{R}_i \tag{3}$$

The demeaned data is then transformed into the frequency domain using the Fast Fourier Transform (FFT). The FFT coefficient $X_i(k)$ for the i-th stock at the k-th frequency is calculated using equation (4),

$$X_i(k) = \sum_{t=1}^{T} \tilde{R}_{i,t} \cdot e^{-I\frac{2\pi k(t-1)}{T}} \tag{4}$$

To obtain a scalar value representing the strength of each frequency component, the spectrum magnitude $M_i(k)$ is calculated from the complex coefficient $X_i(k)$ using equation (5),

$$M_i(k) = \sqrt{[\mathrm{Re}\,(X_i(k))]^2 + [\mathrm{Im}\,(X_i(k))]^2} \tag{5}$$

Since the return data are real numbers, the resulting frequency spectrum is symmetric. Therefore, only the first half of the spectrum contains unique (non-redundant) information. The number of non-redundant frequencies is expressed as $m = \lfloor \frac{T}{2} \rfloor + 1$. The magnitudes at these non-redundant frequencies are subsequently denoted as $M_i(k^*)$, where $k^* = 0, 1, \ldots, m-1$ is the non-redundant frequency index. Furthermore, dominant frequencies are identified based on the highest average magnitude across all stocks. In this study, three dominant frequencies are selected and denoted as k_l, where $l = 1, 2, 3$ is the dominant frequency index based on the order of the largest average magnitude. The final result of the FFT process is a dominant frequency magnitude matrix $\mathbf{M}$ of size $n \times 3$, which contains the magnitude of each stock at these three dominant frequencies.

2.3. **Data Normalization.** Data normalization is a process in data preprocessing that aims to standardize the scale of the data. This uniformity of data affects matrix calculations and the measurement of similarity between objects, particularly in cluster analysis [12]. One commonly used normalization method is the Min-Max Scaler, which scales the data into a specific range. Normalization is applied to the magnitudes resulting from the FFT using equation (6),

$$M_i'(k_l) = \frac{M_i(k_l) - \min \mathbf{M}(k_l)}{\max \mathbf{M}(k_l) - \min \mathbf{M}(k_l)} \tag{6}$$

where $\mathbf{M}(k_l)$ is the magnitude vector at the l-th dominant frequency, $M_i(k_l)$ is the magnitude for the i-th stock at the l-th dominant frequency, and l is the dominant frequency index.

2.4. **Spectral Clustering.** Spectral clustering is a clustering method based on the similarity between each data object. This method uses information obtained from eigenvalues and eigenvectors of a similarity matrix to partition a graph [13]. In this graph representation, each vertex or node represents a data object, while edges represent relationships between data points calculated based on the distance between two related data points.

The spectral clustering algorithm utilizes the spectrum (eigenvalues) of the data Laplacian matrix to reduce dimensions to a lower space before performing clustering. This approach is rooted in spectral graph partitioning theory, where each data point is transformed into a node in a graph and the relationships between nodes are measured by edges representing the similarity between data. This graph-based representation enables the algorithm to cluster data that cannot be separated linearly, making spectral clustering proven to be superior to many traditional algorithms in various cases [14]. Other advantages include the ability to handle large and complex data, flexibility with various kernels, and no requirement for specific cluster shape assumptions.

2.4.1. *Spectral Clustering Algorithm.* The spectral clustering algorithm is implemented on the normalized dominant frequency magnitude matrix $\mathbf{M'}$, which contains the normalized magnitudes $M_i'(k_l)$. According to Yusuf et al. in Wulandari [15], the clustering process begins by constructing a similarity matrix $\mathbf{W} \in \mathbb{R}^{n \times n}$ that represents the spectral similarity between stocks. The similarity weight w_{ij} between stocks i and j is calculated using a Gaussian kernel function based on the Euclidean distance between them, which can be expressed as equation (7),

$$w_{i,j} = \begin{cases} \exp\left(-\frac{\sum_{l=1}^{3}\left(M_i'(k_l) - M_j'(k_l)\right)^2}{2\sigma^2}\right), & \text{untuk } i \neq j \\ 0, & \text{untuk } i = j \end{cases} \tag{7}$$

The degree matrix $\mathbf{D}$ is formed as a diagonal matrix with its diagonal elements calculated using equation (8),

$$\begin{aligned} d_{i,i} &= \sum_{j=1}^{n} w_{ij} \\ d_{i,j} &= 0, \quad \text{untuk } i \neq j \end{aligned} \tag{8}$$

The normalized Laplacian matrix $\mathbf{L}_{sym}$ is then constructed from the similarity matrix $\mathbf{W}$ and the degree matrix $\mathbf{D}$ using equation (9),

$$\mathbf{L}_{sym} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2} \tag{9}$$

Eigen decomposition is performed on $\mathbf{L}_{sym}$ to obtain eigenvalues λ and eigenvectors $\mathbf{v}$ that fulfill $\mathbf{L}_{sym}\mathbf{v} = \lambda\mathbf{v}$. The obtained eigenvectors are then sorted based on their corresponding eigenvalues from smallest to largest. For comparative analysis purposes, the clustering process is carried out for three variations of the number of clusters, namely c, $c+1$, and $c+2$. Define $c^* := c$ as the initial number of clusters, with iterations continuing until $c^* = c + 2$. The

spectral embedding matrix $\mathbf{U}^{(c^*)}$ is formed from the c^* smallest eigenvectors (excluding the first eigenvector for $\lambda_1 = 0$) as in equation (10),

$$\mathbf{U}^{(c^*)} = [\mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_{c^*+1}] = \begin{bmatrix} u_{1,1}^{(c^*)} & u_{1,2}^{(c^*)} & \cdots & u_{1,c^*}^{(c^*)} \\ u_{2,1}^{(c^*)} & u_{2,2}^{(c^*)} & \cdots & u_{2,c^*}^{(c^*)} \\ \vdots & \vdots & \ddots & \vdots \\ u_{n,1}^{(c^*)} & u_{n,2}^{(c^*)} & \cdots & u_{n,c^*}^{(c^*)} \end{bmatrix} \in \mathbb{R}^{n \times c^*} \tag{10}$$

Each row of the matrix $\mathbf{U}^{(c^*)}$ is then normalized to produce the normalized spectral embedding matrix $\mathbf{T}^{(c^*)}$ using equation (11),

$$t_{i,s}^{(c^*)} = \frac{u_{i,s}^{(c^*)}}{\left\| \mathbf{u}_i^{(c^*)} \right\|^2} = \frac{u_{i,s}^{(c^*)}}{\sqrt{\sum_{s=1}^{c^*} \left(u_{i,s}^{(c^*)} \right)^2}} \tag{11}$$

The final result of the spectral clustering algorithm is the normalized spectral embedding matrix $\mathbf{T}^{(c^*)}$.

2.5. ***Fuzzy C-Means.*** Fuzzy C-Means (FCM) is a clustering method that groups data based on characteristic similarity by applying the concept of fuzzy sets [16]. Unlike conventional clustering methods that are crisp (hard clustering), FCM applies a soft partition where each data point can belong to several clusters simultaneously with membership degree values between 0 and 1.

The FCM method was first introduced by Dunn (1973) and later developed by Jim Bezdek (1981) for pattern recognition [17]. The FCM algorithm works iteratively by minimizing an objective function that measures the total weighted distance between each data point and cluster centers until optimal convergence is achieved.

2.5.1. *Fuzzy C-Means Algorithm.* The Fuzzy C-Means (FCM) algorithm is applied to the feature matrix $\mathbf{T}^{(c^*)}$ obtained from spectral clustering. According to Bezdek [18], the clustering process with Fuzzy C-Means begins by preparing the data in the form of a feature matrix. In this study, a feature matrix of size $n \times c^*$ is used, where n is the number of stocks and c^* is the number of clusters being evaluated. This algorithm requires several input parameters: the number of clusters being evaluated c^*, the fuzzifier parameter α, the error threshold ε, the initial iteration τ, the maximum iteration $\tau_{\max}$, and the initial value of the objective function $P^{(c^*,0)}$.

The process begins with the random initialization of the initial fuzzy partition matrix $\mathbf{F}^{(c^*,0)} \in \mathbb{R}^{n \times c^*}$ by generating a matrix $\mathbf{A}^{(c^*)}$ with elements $a_{i,s}^{(c^*)} \sim \text{Uniform}(0,1)$, which is then normalized using equation (12),

$$f_{i,s}^{(c^*,0)} = \frac{a_{i,s}^{(c^*)}}{\sum_{s=1}^{c^*} a_{i,s}^{(c^*)}} \tag{12}$$

After initialization, the algorithm enters an iterative process where at each iteration τ, the cluster centers $\mathbf{G}^{(c^*,\tau)}$ are calculated using equation (13),

$$g_{s,d}^{(c^*,\tau)} = \frac{\sum_{i=1}^{n} \left(\left(f_{i,s}^{(c^*,\tau-1)} \right)^{\alpha} t_{i,d}^{(c^*)} \right)}{\sum_{i=1}^{n} \left(f_{i,s}^{(c^*,\tau-1)} \right)^{\alpha}} \tag{13}$$

The objective function $P^{(c^*,\tau)}$ is then calculated using equation (14),

$$P^{(c^*,\tau)} = \sum_{i=1}^{n} \sum_{s=1}^{c^*} \left(\left(f_{i,s}^{(c^*,\tau-1)} \right)^{\alpha} \left[\sum_{d=1}^{c^*} \left(t_{i,d}^{(c^*)} - g_{s,d}^{(c^*,\tau)} \right)^2 \right] \right) \tag{14}$$

The fuzzy partition matrix is updated using equation (15),

$$f_{i,s}^{(c^*,\tau)} = \frac{\left(\sum_{d=1}^{c^*}\left(t_{i,d}^{(c^*)} - g_{s,d}^{(c^*,\tau)}\right)^2\right)^{\frac{-1}{(\alpha-1)}}}{\sum_{q=1}^{c^*}\left(\sum_{d=1}^{c^*}\left(t_{i,d}^{(c^*)} - g_{q,d}^{(c^*,\tau)}\right)^2\right)^{\frac{-1}{(\alpha-1)}}} \tag{15}$$

The algorithm stops if $\left|P^{(c^*,\tau)} - P^{(c^*,\tau-1)}\right| < \varepsilon$ or $\tau > \tau_{\max}$. If these conditions are not met, the value of τ is increased to $\tau = \tau + 1$ and the iteration process continues again from the cluster center calculation stage. The final results of the algorithm are the fuzzy partition matrix $\mathbf{F}^{(c^*,\tau)}$, the cluster center matrix $\mathbf{G}^{(c^*,\tau)}$, and the objective function value $P^{(c^*,\tau)}$. This process is repeated for each variation of the number of clusters $c^* \in \{c, c+1, c+2\}$.

2.6. **Xie-Beni Index.** Xie-Beni Index (XB) is a cluster validation index specifically used for fuzzy clustering. This index was first introduced by Xie and Beni in 1991. Mathematically, the Xie-Beni Index measures the ratio between the total variation within clusters and the separation between clusters [19]. In the context of Fuzzy C-Means, minimizing the Xie-Beni Index aims to produce more compact clusters, where a lower Xie-Beni Index value indicates better cluster partition quality. The formula for the Xie-Beni Index can be expressed as equation (16),

$$\mathrm{XB}^{(c^*)} = \frac{\sum_{i=1}^{n}\sum_{s=1}^{c^*}\left(\left(f_{i,s}^{(c^*,\tau)}\right)^{\alpha}\left[\sum_{d=1}^{c^*}\left(t_{i,d}^{(c^*)} - g_{s,d}^{(c^*,\tau)}\right)^2\right]\right)}{n \cdot \left(\min\limits_{s\neq q}\left[\sum_{d=1}^{c^*}\left(g_{s,d}^{(c^*,\tau)} - g_{q,d}^{(c^*,\tau)}\right)^2\right]\right)} \tag{16}$$

where $\mathrm{XB}^{(c^*)}$ is the Xie-Beni Index value for the number of clusters c^*.

2.7. **Mean-Variance Portfolio Optimization.** Mean-Variance portfolio optimization is a fundamental framework in Modern Portfolio Theory (MPT) developed by Markowitz in 1952. This model aims to determine the optimal asset allocation by considering the trade-off between portfolio expected return (mean) and risk (variance) [20]. The basic concept of this model is to form an efficient frontier that provides the highest expected return for a given level of risk, or the lowest risk for a given level of expected return. Portfolio expected return and risk can be expressed in vector form as in equations (17) and (18),

$$\mu_p = E(R_p) = \boldsymbol{\mu}^T\mathbf{b} = \mathbf{b}^T\boldsymbol{\mu} \tag{17}$$

$$\sigma_p^2 = \mathrm{Var}(R_p) = \mathbf{b}^T\boldsymbol{\Sigma}\mathbf{b} \tag{18}$$

To obtain an efficient portfolio, optimization is performed by maximizing the objective function in equation (19),

$$\mu_p - \frac{\rho}{2}\sigma_p^2, \quad \rho \geq 0 \tag{19}$$

where ρ is the risk aversion parameter. For an investor with risk aversion ρ ($\rho \geq 0$), portfolio optimization can be formulated as in equation (20),

$$\begin{aligned} &\max\left\{\mu_p - \frac{\rho}{2}\sigma_p^2\right\}, \\ &\text{s.t.} \quad \sum_{i=1}^{n} b_i = 1, \\ &\qquad b_i \geq 0, \\ &\qquad \text{atau} \\ &\max\left\{\mathbf{b}^T\boldsymbol{\mu} - \frac{\rho}{2}\mathbf{b}^T\boldsymbol{\Sigma}\mathbf{b}\right\}, \\ &\text{s.t.} \quad \mathbf{b}^T\mathbf{e} = 1, \\ &\qquad b_i \geq 0, \end{aligned} \tag{20}$$

Thus, the solution of the Mean-Variance portfolio model is obtained as in equation (21),

$$\mathbf{b} = \frac{1}{\rho}\mathbf{\Sigma}^{-1}(\boldsymbol{\mu} + \lambda\mathbf{e}), \quad \text{dengan } \lambda = \frac{\rho - \boldsymbol{\mu}^T\mathbf{\Sigma}^{-1}\mathbf{e}}{\mathbf{e}^T\mathbf{\Sigma}^{-1}\mathbf{e}} \tag{21}$$

3. Results and Discussion

3.1. **Transformation of Return Data to the Frequency Domain.** This study uses monthly closing price data of 41 banking sector stocks for the period December 2019–December 2024. The stock price data were then converted into monthly returns using equation (1), resulting in 60 return periods for each stock. Furthermore, a demeaning process was performed to remove the constant mean component from the return data of each stock using equation (3), yielding data with zero mean. Then, the demeaned data were transformed to the frequency domain using the Fast Fourier Transform as expressed in equation (4). This process calculates the complex coefficients $X_i(k)$ for each stock i at each frequency k, which are then converted into magnitude values $M_i(k)$ using equation (5). The transformation results produce magnitudes for 60 frequencies, namely $k = 0, 1, \ldots, 59$ from all 41 stocks. Since the return data are real numbers, the resulting frequency spectrum is symmetric. Therefore, the analysis focuses on the first half of the spectrum which is non-redundant, totaling $m = 31$ frequencies. These non-redundant frequencies are denoted as $k^* = 0, 1, \ldots, 30$. The average magnitude for each non-redundant frequency was then calculated from all stocks. The results of the average magnitude calculation are presented in graphical form in Figure 1.

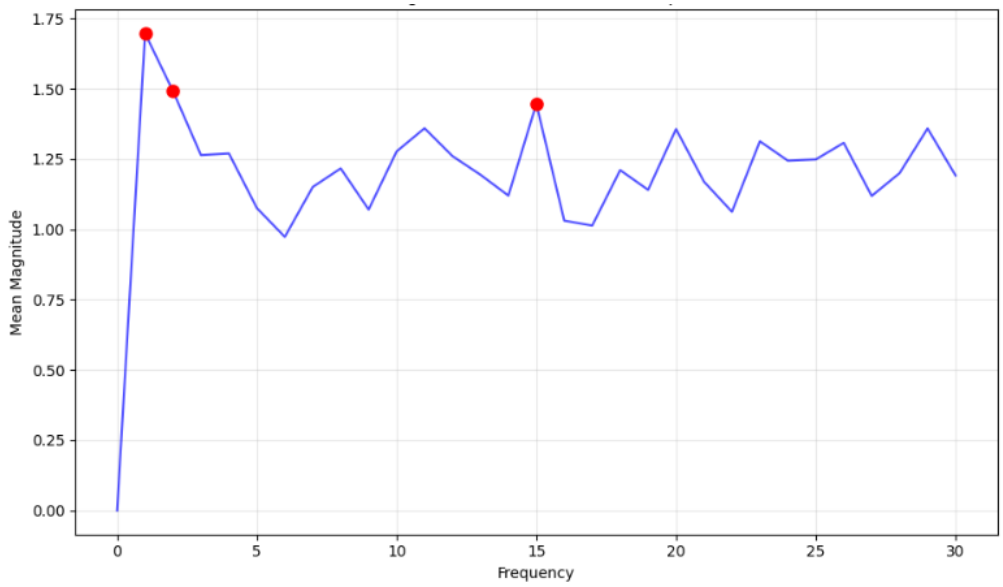

Figure 1. Graph of Average Magnitude for Each Frequency

Based on Figure 1, three dominant frequencies with the highest average magnitudes were identified, namely $k_1 = 1$, $k_2 = 2$, and $k_3 = 15$. These three dominant frequencies represent cycles with specific time periods. Conversion to time periods is done by dividing the number of observation periods by the frequency index, resulting in a long cycle of 60 months for frequency 1, a medium cycle of 30 months for frequency 2, and a short cycle of 4 months for frequency 15.

The final result of the Fast Fourier Transform is a dominant frequency magnitude matrix $\mathbf{M}$ of size 41×3 containing the magnitude of each stock at the three dominant frequencies, which can be expressed as follows:

$$\mathbf{M} = \begin{bmatrix} 4,45831 & 3,87188 & 2,90839 \\ 1,78604 & 1,59996 & 0,85113 \\ \vdots & \vdots & \vdots \\ 0,05000 & 0,37788 & 0,99397 \end{bmatrix}$$

3.2. **Stock Clustering.** Stock clustering begins by applying the spectral clustering algorithm to the normalized dominant frequency magnitude matrix $\mathbf{M}'$. The spectral clustering process includes constructing the similarity matrix $\mathbf{W}$ using equation (7), forming the degree matrix $\mathbf{D}$ using equation (8), constructing the normalized Laplacian matrix $\mathbf{L}_{\text{sym}}$ using equation (9), and performing eigen decomposition on the matrix $\mathbf{L}_{\text{sym}}$. Based on the results of the eigen decomposition of the matrix $\mathbf{L}_{\text{sym}}$, the spectral embedding matrix $\mathbf{U}^{(c^*)}$ and the normalized spectral embedding matrix $\mathbf{T}^{(c^*)}$ were obtained for various cluster number variations being evaluated.

For $c^* = 3$, the spectral embedding matrix $\mathbf{U}^{(3)}$ and the normalized spectral embedding matrix $\mathbf{T}^{(3)}$ are as follows:

$$\mathbf{U}^{(3)} = \begin{bmatrix} -0,41252 & 0,12300 & -0,10949 \\ 0,00599 & -0,08097 & -0,14253 \\ \vdots & \vdots & \vdots \\ 0,08310 & 0,09035 & 0,09516 \end{bmatrix}, \quad \mathbf{T}^{(3)} = \begin{bmatrix} -0,92874 & 0,27692 & -0,24650 \\ 0,03654 & -0,49361 & -0,86891 \\ \vdots & \vdots & \vdots \\ 0,53502 & 0,58169 & 0,61269 \end{bmatrix}$$

For $c^* = 4$, the spectral embedding matrix $\mathbf{U}^{(4)}$ and the normalized spectral embedding matrix $\mathbf{T}^{(4)}$ are as follows:

$$\mathbf{U}^{(4)} = \begin{bmatrix} -0,41252 & 0,12300 & -0,10949 & 0,15157 \\ 0,00599 & -0,08097 & -0,14253 & -0,24391 \\ \vdots & \vdots & \vdots & \vdots \\ 0,08310 & 0,09035 & 0,09516 & 0,11291 \end{bmatrix}$$

$$\mathbf{T}^{(4)} = \begin{bmatrix} -0,87897 & 0,26208 & -0,23329 & 0,32296 \\ 0,02039 & -0,27546 & -0,48490 & -0,82981 \\ \vdots & \vdots & \vdots & \vdots \\ 0,43276 & 0,47051 & 0,49558 & 0,58800 \end{bmatrix}$$

For $c^* = 5$, the spectral embedding matrix $\mathbf{U}^{(5)}$ and the normalized spectral embedding matrix $\mathbf{T}^{(5)}$ are as follows:

$$\mathbf{U}^{(5)} = \begin{bmatrix} -0,41252 & 0,12300 & -0,10949 & 0,15157 & 0,12149 \\ 0,00599 & -0,08097 & -0,14253 & -0,24391 & -0,19941 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,08310 & 0,09035 & 0,09516 & 0,11291 & 0,29901 \end{bmatrix}$$

$$\mathbf{T}^{(5)} = \begin{bmatrix} -0,85092 & 0,25372 & -0,22585 & 0,31266 & 0,25060 \\ 0,01688 & -0,22796 & -0,40127 & -0,68670 & -0,56141 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0,23384 & 0,25424 & 0,26779 & 0,31772 & 0,84144 \end{bmatrix}$$

The clustering process was continued with the Fuzzy C-Means (FCM) algorithm on each matrix $\mathbf{T}^{(c^*)}$. FCM parameter initialization was performed with fuzzifier exponent $\alpha = 2$, error tolerance threshold $\varepsilon = 10^{-5}$, initial iteration $\tau = 1$, maximum iteration $\tau_{\max} = 100$, and initial objective function value $P^{(c^*,0)} = 0$. For each cluster number variation evaluated, an iterative process was carried out starting with the initialization of the initial fuzzy partition matrix $\mathbf{F}^{(c^*,0)}$ using equation (12), followed by the calculation of cluster centers $\mathbf{G}^{(c^*,\tau)}$ using equation (13), calculation of the objective function $P^{(c^*,\tau)}$ using equation (14), and updating

the fuzzy partition matrix $\mathbf{F}^{(c^*,\tau)}$ using equation (15). The iterative process continued until the convergence condition $|P^{(c^*,\tau)} - P^{(c^*,\tau-1)}| < \varepsilon$ or $\tau > \tau_{\max}$ was satisfied.

The iteration results show that for $c^* = 3$, convergence was achieved at the 18th iteration with $|P^{(3,18)} - P^{(3,17)}| = 0.0000060 < \varepsilon$. For $c^* = 4$, convergence was achieved at the 27th iteration with $|P^{(4,27)} - P^{(4,26)}| = 0.0000093 < \varepsilon$. Meanwhile, for $c^* = 5$, convergence was achieved at the 32nd iteration with $|P^{(5,32)} - P^{(5,31)}| = 0.0000095 < \varepsilon$.

Furthermore, cluster quality evaluation was performed using the Xie-Beni Index. The calculation results of the Xie-Beni Index for the three cluster number variations evaluated are presented in Table 1.

Table 1. Cluster Quality Evaluation Results Using Xie-Beni Index

c^*	Xie-Beni Index Value
3	0,11949
4	0,10098
5	0,12229

Based on Table 1, the smallest Xie-Beni Index value was obtained for $c^* = 4$. Therefore, the optimal number of clusters was determined as $c_{\text{opt}} = 4$. Thus, the clustering results for $c_{\text{opt}} = 4$ based on the fuzzy partition matrix $\mathbf{F}^{(4,27)}$ containing the membership degree values of each stock in the four clusters can be expressed as follows:

$$\mathbf{F}^{(4,27)} = \begin{bmatrix} 0,06543 & 0,78004 & 0,08056 & 0,07397 \\ 0,03551 & 0,03095 & 0,02216 & 0,91138 \\ 0,02668 & 0,91068 & 0,03456 & 0,02808 \\ 0,07454 & 0,11046 & 0,06285 & 0,75214 \\ 0,82728 & 0,05143 & 0,03549 & 0,08580 \\ \vdots & \vdots & \vdots & \vdots \\ 0,17780 & 0,12988 & 0,59810 & 0,09422 \end{bmatrix}$$

The main cluster for each stock was determined by selecting the highest membership degree value for each stock. The results of the main cluster determination are presented in Table 2.

Table 2. Main Clusters of Banking Stocks

Main Cluster	Number of Stocks	Stock Code List
1	9	BACA, BBTN, BEKS, BKSW, BMAS, BSIM, DNAR, INPC, MAYA
2	6	AGRO, ARTO, BBYB, BGTG, BINA, BNBA
3	15	BBCA, BBMD, BBNI, BBRI, BDMN, BJBR, BJTM, BMRI, BNGA, BTPN, BTPS, MCOR, MEGA, NISP, SDRA
4	11	AGRS, BABP, BBHI, BBKP, BNII, BNLI, BRIS, BVIC, NOBU, PNBN, PNBS

The characteristics of the four clusters were further analyzed by comparing the monthly return statistics calculated from the stocks in each cluster. Descriptive statistics for the four clusters are presented in Table 3.

TABLE 3. Descriptive Statistics of Return Based on Main Clusters

Statistic	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Mean Return	0, 00962	0, 04599	0, 00114	0, 02229
Standard Deviation of Return	0, 20176	0, 32987	0, 08558	0, 19284
Minimum Return	−0, 03127	0, 03170	−0, 01750	−0, 00042
Maximum Return	0, 06669	0, 06391	0, 01286	0, 10749
Range	0, 09796	0, 03221	0, 03036	0, 10790

Based on Table 3, it can be identified that the four clusters have different risk and return characteristics. Cluster 1 shows a medium risk-low return profile with an average return of 0.96% and volatility of 20.18%. Cluster 2 shows a high risk-high return profile with the highest average return of 4.60% and the highest volatility of 32.99%. Cluster 3 represents a low risk-low return profile with the lowest average return of 0.11% and the highest stability with a standard deviation of 8.56%. Cluster 4 has a medium risk-high return profile with an average return of 2.23% and the highest appreciation potential indicated by the maximum return value of 10.75%.

3.3. **Mean-Variance Portfolio Optimization Based on Clustering Results.** Portfolio optimization was performed using the Mean-Variance method based on the stock clustering results. From the four formed clusters, representative stocks were selected from each cluster based on the highest Sharpe Ratio value to be used in forming the optimal portfolio. The selected representative stocks were BMAS from cluster 1, ARTO from cluster 2, BNGA from cluster 3, and BBHI from cluster 4.

The optimization calculation began with an initial risk aversion value of $\rho = 0$. In this calculation, it was found that there were still negative portfolio weights. Therefore, the ρ value was gradually increased with $\Delta\rho = 0.01$ until all portfolio weights were non-negative and the total portfolio weight summed to one. For each ρ value, the portfolio expected return (μ_p), portfolio variance (σ_p^2), and portfolio ratio μ_p/σ_p^2 were calculated. The results of the portfolio optimization calculations for various ρ values are presented in Table 4.

TABLE 4. Portfolio Optimization Results

ρ	BMAS	ARTO	BNGA	BBHI	$\mathbf{b}^T\mathbf{e}$	μ_p	σ_p^2	μ_p/σ_p^2
1,57	0,40925	0,26610	-0,00375	0,32840	1	0,0691961	0,0412789	1,6860370
1,58	0,40683	0,26488	0,00220	0,32608	1	0,0688440	0,0408319	1,6860370
1,59	0,40444	0,26368	0,00808	0,32380	1	0,0684964	0,0403933	1,6957396
1,60	0,40208	0,26249	0,01389	0,32154	1	0,0681532	0,0399628	1,7054141
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
6,81	0,11505	0,11776	0,72017	0,04703	1	0,0263989	0,0077351	3,4128836
6,82	0,11492	0,11769	0,72049	0,04690	1	0,0263801	0,0077295	3,4128874
6,83	0,11479	0,11763	0,72080	0,04678	1	0,0263613	0,0077241	3,4128877
6,84	0,11466	0,11756	0,72112	0,04666	1	0,0263426	0,0077186	3,4128844
6,85	0,11453	0,11750	0,72144	0,04653	1	0,0263240	0,0077131	3,4128776
6,86	0,11441	0,11743	0,72175	0,04641	1	0,0263054	0,0077077	3,4128674
6,87	0,11428	0,11737	0,72206	0,04629	1	0,0262869	0,0077023	3,4128536
6,88	0,11415	0,11731	0,72238	0,04617	1	0,0262684	0,0076969	3,4128364
6,89	0,11402	0,11724	0,72269	0,04605	1	0,0262500	0,0076916	3,4128157
6,90	0,11390	0,11718	0,72300	0,04593	1	0,0262316	0,0076863	3,4127917

Based on Table 4, a set of efficient portfolios was obtained in the interval $\rho = [1.58, \infty)$. The portfolio weight compositions with $\rho < 1.58$ produced negative weights and were therefore

not included in the set of efficient portfolios, although the total portfolio weights summed to one. This set of efficient portfolios was then visualized in graphical form. The efficient frontier graph can be seen in Figure 2.

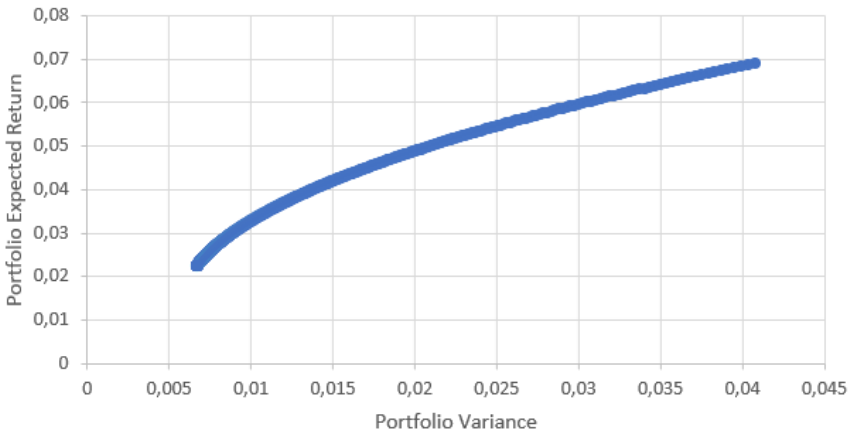

FIGURE 2. Efficient Frontier Graph

Based on Figure 2, a positive relationship between portfolio expected return and portfolio variance can be observed. The points on this curve represent efficient portfolios that provide optimal combinations of expected return and risk level. Furthermore, a graph comparing the portfolio ratio $\frac{\mu_p}{\sigma_p^2}$ against the risk aversion value ρ can be seen in Figure 3.

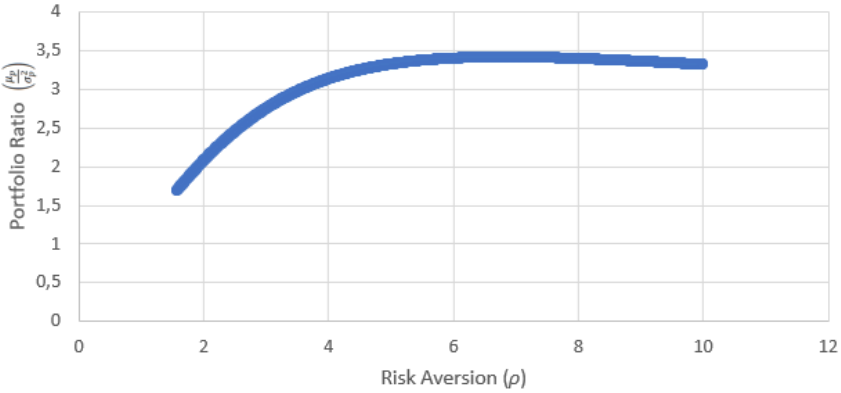

FIGURE 3. Portfolio Ratio Graph

Based on Figure 3, the portfolio ratio $\frac{\mu_p}{\sigma_p^2}$ increases as ρ increases until reaching an optimal point at $\rho = 6.83$, after which it decreases. The increase in the ratio over the range $1.58 \leq \rho \leq 6.83$ indicates that within this range, an increase in risk aversion still provides a commensurate return compensation. The decrease in the ratio after $\rho > 6.83$ shows that further increases in risk aversion no longer provide proportional additional returns. Thus, the optimal portfolio was obtained at $\rho = 6.83$ with a ratio of $\frac{\mu_p}{\sigma_p^2} = 3.4128877$.

Based on Table 4, the stock weight composition for $\rho = 6.83$ is 0.11479 (11.48%) for BMAS stock, 0.11763 (11.76%) for ARTO stock, 0.72080 (72.08%) for BNGA stock, and 0.04678 (4.68%) for BBHI stock. A visualization of this weight composition can be seen in Figure 4.

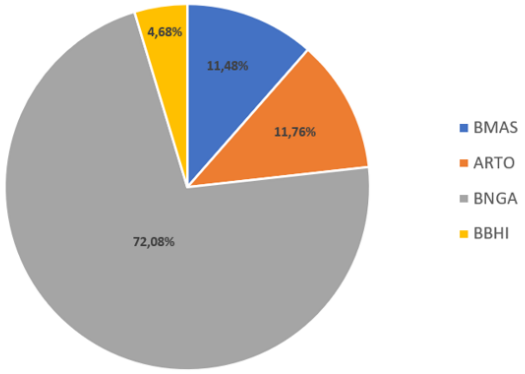

FIGURE 4. Optimal Portfolio Stock Weight Composition

This portfolio composition has an expected return of 0.0263613 and a portfolio variance of 0.0077241.

4. CONCLUSION

Based on the results of the analysis and discussion that have been conducted, it is concluded that the integration of the Fourier Transform, spectral clustering, and Fuzzy C-Means methods successfully grouped 41 banking stocks into four optimal clusters with different risk and return characteristics. Cluster 1 consists of nine stocks with a medium risk-low return profile, cluster 2 consists of six stocks with a high risk-high return profile, cluster 3 consists of fifteen stocks with a low risk-low return profile, and cluster 4 consists of eleven stocks with a medium risk-high return profile. Furthermore, from these clustering results, the optimal portfolio formed using the Mean-Variance method was obtained at a risk aversion value of $\rho = 6.83$ with a portfolio ratio of 3.4128877. This portfolio has a stock weight composition of 11.48% BMAS, 11.76% ARTO, 72.08% BNGA, and 4.68% BBHI, with an expected return value of 0.0263613 and a portfolio variance of 0.0077241.

REFERENCES

[1] Kusuma, A.A., Farhan, M., Maedianda, M., and Rahmah, F., 2024, Kebijakan moneter terhadap stabilitas pasar modal di Indonesia, *EQUILIBRIUM: Jurnal Ilmiah Ekonomi dan Pembelajarannya*, 12(1), pp. 55–64. doi: 10.25273/equilibrium.v12i1.18701.

[2] Bursa Efek Indonesia, 2019, Laporan Tahunan 2019, *Bursa Efek Indonesia*, https://www.idx.co.id/id/tentang-bei/laporan-tahunan/.

[3] Bursa Efek Indonesia, 2024, Laporan Tahunan 2024, *Bursa Efek Indonesia*, https://www.idx.co.id/id/tentang-bei/laporan-tahunan/.

[4] Bursa Efek Indonesia, 2023, Tambah Alternatif Acuan Investasi Sub-Sektor Bank, BEI dan PEFINDO Luncurkan Indeks IDX-PEFINDO Prime Bank, *Bursa Efek Indonesia*, https://www.idx.co.id/id/berita/siaran-pers/2028.

[5] Lyle, M.R., and Yohn, T.L., 2020, Fundamental analysis and Mean-Variance optimal portfolios, *The Accounting Review*, 96(6), pp. 303–327. doi: 10.2308/TAR-2019-0622.

[6] Scalzo, B., Arroyo, A., Stanković, L., and Mandic, D.P., 2021, Nonstationary portfolios: Diversification in the spectral domain, *ICASSP 2021 – IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 5155–5159. doi: 10.1109/ICASSP39728.2021.9413769.

[7] Alam, K.E., Saputra, K.V.I., and Margaretha, H., 2021, Stocks clustering with Fourier transformation towards Indonesian stocks for pairs trading and investment diversification, *AIP Conference Proceedings*, 020033. doi: 10.1063/5.0075346.

[8] Gubu, L., Cahyono, E., Arman, A., Budiman, H., and Djafar, M.K., 2023, Optimasi portofolio Mean-Variance dengan analisis klaster Fuzzy C-Means, *Jurnal Gaussian*, 12(4), pp. 593–604. doi: 10.14710/j.gauss.12.4.593-604.

[9] Sunariyah, S., 2006, *Pengantar pengetahuan pasar modal*, 5th ed., Yogyakarta: UPP STIM YKPN.

[10] Anoraga, P. and Pakarti, P., 2008, *Pengantar pasar modal*, Jakarta: Rineka Cipta.
[11] Elawati, E., Hayati, R., and Hanafi, H., 2020, Analisis perbandingan metode Descrete Fourier Transform dan metode Descrete Cosine Transform pada teknik menyembunyikan sinyal suara, *Jurnal Tektro*, 4(2), pp. 82–90. doi: 10.30811/tektro.v6i1.3219.
[12] Yusuf, A., and Tjandrasa, H., 2014, Prediksi nilai dengan metode spectral clustering dan clusterwise regression, *Jurnal Simantec*, 4(1), pp. 1–8.
[13] Sharma, A., Kumar, R., and Mansotra, V., 2016, Proposed stemming algorithm for Hindi information retrieval, *International Journal of Innovative Research in Computer and Communication Engineering*, 3297(6), pp. 11449–11455. doi: 10.15680/IJIRCCE.2016.
[14] von Luxburg, U., 2007, A tutorial on spectral clustering, *Statistics and Computing*, 17(4), pp. 395–416. doi: 10.1007/s11222-007-9033-z.
[15] Wulandari, S., 2020, Clustering microarray adenoma menggunakan spectral clustering dengan algoritma Partitioning Around Medoid (PAM), *Prosiding Seminar Nasional Sains 2020*, pp. 345–351.
[16] Aliansa, W., and Saputra, R.A., 2023, Perbandingan harga saham perbankan dengan metode Fuzzy C-Means clustering, *Prosiding Seminar Nasional Pemanfaatan Sains dan Teknologi Informasi*, pp. 369–376.
[17] Rahmatika, L., Suparti, S., and Safitri, D., 2015, Analisis kelompok dengan algoritma Fuzzy C-Means dan Gustafson Kessel clustering pada Indeks LQ45, *Jurnal Gaussian*, 4(3), pp. 543–551.
[18] Bezdek, J.C., 1981, *Pattern recognition with fuzzy objective function algorithms*, New York: Plenum Press.
[19] Ningrat, D.R., Maruddani, D.A.I., and Wuryandari, T., 2016, Analisis cluster dengan algoritma K-Means dan Fuzzy C-Means clustering untuk pengelompokan data obligasi korporasi, *Jurnal Gaussian*, 5(4), pp. 641–650. doi: 10.14710/j.gauss.5.4.641-650.
[20] Guo, Q., 2022, Review of research on Markowitz model in portfolios, *2022 7th International Conference on Social Sciences and Economic Development (ICSSED 2022)*, pp. 786–790. doi: 10.2991/aebmr.k.220405.131.