# PENERAPAN *MODEL-BASED CLUSTERING* PADA PENGELOMPOKAN SAHAM BERDASARKAN RASIO KEUANGAN

**Irena Sekar Dwi Hasnida[1], Rosita Kusumawati[2]**
[1,2]Statistika, Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Yogyakarta
e-mail: [1]irenasekardwihasnida@gmail.com

**Abstrak**

Untuk meminimalkan kerugian dengan tingkat keuntungan tertentu, investor perlu memilih saham potensial agar keuntungan yang diperoleh optimal. Penelitian ini bertujuan untuk mengetahui penerapan *Model-Based Clustering* (MBC) dalam mengelompokkan perusahaan berdasarkan kinerja keuangan saham. Indikator keuangan yang digunakan yaitu data rasio likuiditas, profitabilitas, dan solvabilitas tahun 2020 untuk perusahaan yang terdaftar pada indeks LQ45. Pemilihan model dan jumlah komponen akan dilakukan berdasarkan nilai *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL). Dari proses *clustering*, terbentuk 6 *cluster* dengan nilai BIC dan ICL berturut-turut sebesar 709,3757 dan -709,376, dan model optimal terpilih VEV (*Variable volume, Equal shape, Variable orientation*). Berdasarkan nilai rata-rata setiap rasio, *cluster* 6 merupakan *cluster* terbaik karena memiliki mayoritas rasio likuiditas dan profabilitas terbaik serta rasio solvabilitas terendah. *Cluster* 6 memiliki kemampuan yang tinggi dibandingkan perusahaan *cluster* lain untuk membiayai kegiatan operasional perusahaan dan memenuhi kewajiban keuangannya jangka pendek.

**Kata kunci:** *Model-Based Clustering*, pengelompokan saham, rasio keuangan.

***Abstract***

*To minimize losses at a certain profit level, investors need to choose potential stocks so that the profits are optimal. This study aims to determine the application of model-based clustering (MBC) in classifying companies based on stock performance. The financial indicators used are data on liquidity ratios, profitability, and solvency for 2020 for companies listed on the LQ45 index. The selection of the model and the number of components will be based on the Bayesian Information Criterion (BIC) and Completed Likelihood Criterion (ICL) values. From the clustering process, six clusters were formed with BIC and ICL values of 709.3757 and -709.376, respectively, and the optimal model was selected as VEV (variable volume, equal shape, variable orientation). Based on the average value of each ratio, cluster 6 is the best cluster because it has the majority of the best liquidity and profitability ratios and the lowest solvency ratio. Cluster 6 has a high ability, compared to other cluster companies, to finance the company's operational activities and fulfill its short-term financial obligations.*

***Keywords**: Model-Based Clustering, stock clustering, financial ratios.*

## PENDAHULUAN

Pengambilan keputusan investasi oleh investor perlu dibuat secara rasional untuk memaksimalkan keuntungan. Dalam hal ini analisis fundamental perusahaan dapat dimanfaatkan oleh investor sebagai pertimbangan dalam keputusan investasinya. Analisis fundamental bertujuan untuk memperkirakan harga saham di masa mendatang dengan memperkirakan nilai faktor fundamental yang akan mempengaruhi harga saham di masa mendatang, seperti perkiraan laba per saham, perputaran pendapatan perusahaan, risiko pendapatan masa depan perusahaan, penggunaan hutang oleh manajemen dan kebijakan deviden (Weston dan Brigham, 1993). Analisis fundamental melihat kondisi internal perusahaan salah satunya melalui analisis laporan keuangan menggunakan rasio-rasio keuangan yang menginterpertasikan kondisi keuangan dan hasil operasi suatu perusahaan (Suad, 2004).

Untuk menilai keuntungan perusahaan dapat dianalisis melalui data yang tercermin dalam laporan keuangan perusahaan, dengan menganalisis laporan keuangan perusahaan sendiri (Andy dan Megawati, 2019). Analisis rasio keuangan adalah analisis yang menggabungkan evaluasi laporan keuangan dan laporan laba rugi terhadap satu dengan lainnya, yang memberikan gambaran tentang sejarah perusahaan dan penilaian kesehatan suatu perusahaan tertentu. Analisis rasio dapat membimbing investor membuat keputusan atau pertimbangan tentang apa yang akan dicapai oleh perusahaan dan atau bagaimana prospek yang akan dihadapi dimasa yang akan datang (Andayani, 2016). Masing-masing rasio keuangan mencerminkan aspek tertentu. Saat menganalisis rasio keuangan, pemberi pinjaman mencari informasi tentang kemampuan perusahaan untuk membayar kembali pinjamannya tepat waktu, sementara investor lebih tertarik pada kemampuan perusahaan untuk menghasilkan keuntungan.

Beberapa penelitian telah dilakukan pada portofolio dan strategi pemilihan aset/saham. Salah satunya dengan melakukan pengelompokkan terhadap saham menggunakan teknik *clustering* seperti yang dilakukan oleh Nanda et.al (2010) yang memanfaatkan beberapa teknik clustering pada saham yaitu *K-means*, SOM (*Self-Organization Map*), dan *Fuzzy C-means* yang menggunakan data pasar Bombay Stock Exchange. Subekti dkk (2017) juga melakukan *clustering* saham menggunakan *K-Means* dan *Average Linkage* untuk menyusun portofolio yang optimal dalam rangka strategi diversifikasi yang makin beragam. Pada penelitian lainnya, Subekti et.al (2018) menggunakan *Ant colony algorithm* untuk *clustering* dapat memberikan portofolio dengan performa yang lebih baik berdasarkan nilai *Sharpe index*.

Dalam perspektif *data mining*, masalah pemilihan saham bertujuan untuk mengidentifikasi saham potensial atau saham yang memberikan keuntungan tinggi. *Clustering* saham bisa digunakan untuk mengelompokkan saham berdasarkan beberapa indikator kinerja perusahaan. Seorang investor bisa memilih saham hasil analisis karakteristik saham dari *clustering* yang terbentuk. Terdapat beberapa jenis clustering yang digunakan, diantara nya adalah *Centroid-based Clustering, Density-based Clustering, Distribution-based Clustering, dan Hierarchical Clustering*. Pada artikel ini penulis akan menggunakan *Model/Distribution-based Clustering* sebagai metode *clustering* saham. Jenis *clustering* ini mengelompokkan data berdasarkan fungsi peluang campuran atau distribusi campuran (*mixture distribution*), dimana setiap kelompok dimodelkan oleh fungsi peluangnya sendiri. *Expectation-maximization* merupakan algoritma yang digunakan untuk mengimplementasikan metode *Model-Based Clustering*.

*Model-Based Clustering* (MBC), menganggap sebaran data mengikuti distribusi campuran (*mixture distribution*). Mixture distribution adalah gabungan dari beberapa komponen distribusi nilai keanggotaan data padasuatu cluster ditentutan oleh nilai peluang terbesar dari komponen distribusi. Tidak seperti metode *k-means*, MBC menggunakan *soft assignment*, di mana setiap titik data memiliki probabilitas untuk menjadi milik setiap kelompok. Qona'ah et. al. (2020) melakukan pengelompokan pada 100 laboraturium di Institut Teknologi Sepuluh Nopember (ITS) berdasarkan produktivitas

laboratorium dalam menjalankan fungsinya. Penelitian ini membandingkan metode *K-Means*, *K-Medoids*, and *Model-Based Clustering* (MBC) dalam proses pengelompokan. Hasil penelitian menunjukkan bahwa MBC memberikan performa yang lebih baik dari *K-Means* dan *K-Medoids*.

Penelitian ini bertujuan untuk mengetahui penerapan algoritma MBC untuk mengelompokkan saham berdasarkan rasio keuangan sebagai indikator kinerja perusahaan dengan MBC berdasarkan distribusi multivariate mixture gaussian. Eksplorasi *cluster* akan dilakukan serta rata-rata r*eturn* dari setiap *cluster* akan digunakan penulis sebagai ukuran untuk menilai *cluster* yang terbentuk.

## METODOLOGI

Penelitian ini menerapkan *Model-Based Clsutering* (MBC) untuk membentuk *cluster* perusahaan terlebih dahulu berdasarkan nilai rasio keuangan. Selanjutnya dilakukan eksplorasi data terhadap masing-masing kelompok yang terbentuk dari proses *clustering.* Rata-rata r*eturn* dari setiap *cluster* akan digunakan penulis sebagai ukuran untuk menilai *cluster* yang terbentuk. Data yang digunakan untuk proses *clustering* merupakan data rasio keuangan saham pada indeks LQ45 yang tersedia pada *website* www.yahoofinance.com periode tahun 2020. Jumlah saham yang digunakan hanya sebanyak 36 saham, saham ini dipilih dengan pertimbangan kelengkapan data yang tersedia.

### Rasio Keuangan

Rasio keuangan dihitung dengan menggunakan nilai numerik yang diambil dari laporan keuangan untuk mendapatkan informasi yang berarti tentang perusahaan. Ini memungkinkan untuk mengikuti kinerja perusahaan dari waktu ke waktu dan menemukan tanda-tanda masalah. Sehingga dalam penelitian ini rasio keuangan akan digunakan sebagai variabel penelitian.

Menurut Munawir (2002) ada 4 kelompok rasio keuangan yaitu rasio likuiditas, rasio aktivitas, rasio profitabilitas dan rasio solvabilitas. Artikel ini mengunakan 3 aspek rasio keuangan, diantaranya adalah Rasio Likuiditas, Rasio Probabilitas, dan Rasio Solvabilitas. Secara lebih terperinci, variabel dari masing-masing aspek rasio keuangan yang digunakan akan dijelaskan pada poin-poin berikut:

a. Rasio likuiditas adalah rasio yang menentukan kemampuan perusahaan untuk membiayai operasi dan memenuhi kewajiban keuangannya jangka pendek. Rasio ini biasanya digunakan terutama untuk menilai kemampuan perusahaan dalam memenuhi kewajibannya. Adapun untuk jenis rasio likuiditas yang digunakan pada penelitian adalah sebagai berikut (Rist dan Pizzica, 2015):

   1) *Cash Ratio*

   Persamaan untuk menghitung nilai *cash ratio* adalah sebagai berikut:

$$Cash\ Ratio = \frac{Cash\ or\ Cash\ Equipment}{Current\ Liabilities}$$

   2) *Current Ratio*

   Persamaan untuk menghitung nilai *current ratio* adalah sebagai berikut:

$$Current\ Ratio = \frac{Current\ Assets}{Current\ Liabilities}$$

   3) *Quick Ratio*

   Persamaan untuk menghitung nilai *quick ratio* adalah sebagai berikut:

$$Quick\ Ratio = \frac{Current\ Assets - Liabilities}{Current\ Liabilities}$$

b. Rasio profitabilitas merupakan rasio yang menentukan kemampuan perusahaan untuk menghasilkan laba dari berbagai kebijakan dan keputusan yang telah diambil. Rasio profitabilitas menilai kemampuan perusahaan untuk menghasilkan pendapatan, aset, dan keuntungan modal, dan mengukur pengembalian yang diperoleh dari modal perusahaan. Adapun untuk jenis rasio profitabilitas yang digunakan pada penelitian adalah sebagai berikut (Rist dan Pizzica, 2015):

   1) *Return on Assets* (ROA)

   Persamaan untuk menghitung nilai return on assets adalah sebagai berikut:

$$ROA = \frac{Net\ Income}{Total\ Assets}$$

   2) *Return on Equity* (ROE)

Persamaan untuk menghitung nilai return on equity adalah sebagai berikut:

$$ROE = \frac{Net\ Income}{Equity}$$

3) *Gross Profit Margin*

Persamaan untuk menghitung nilai gross profit margin adalah sebagai berikut:

$$Gross\ Profit\ Margin = \frac{Net\ Sales - Cost\ of\ Goods\ Sold}{Sales}$$

4) *Net Profit Margin*

Persamaan untuk menghitung nilai *net* profit margin adalah sebagai berikut:

$$Net\ Profit\ Margin = \frac{Net\ Income}{Sales}$$

c. Rasio solvabilitas adalah rasio untuk mengukur seberapa jauh aset perusahaan dibiayai oleh hutang. Rasio ini mengukur perbandingan dana yang disediakan oleh pemiliknya dengan dana yang dipinjam dari kreditur perusahaan tersebut. Semakin tinggi rasio ini, semakin sedikit ekuitas dibandingkan dengan hutang. Bagi perusahaan, sebaiknya besarnya hutang tidak boleh melebihi modal sendiri agar beban tetapnya tidak terlalu tinggi. Adapun untuk jenis rasio solvabilitas yang digunakan pada penelitian adalah sebagai berikut (Rist dan Pizzica, 2015):

1) *Debt to Equity Ratio*

Persamaan untuk menghitung nilai *debt to equity* ratio adalah sebagai berikut:

$$Debt\ to\ Equity\ Ratio = \frac{Total\ Debt}{Equity}$$

2) *Long-term Debt Ratio*

Persamaan untuk menghitung nilai *long-*term *debt ratio* adalah sebagai berikut:

$$Long - term\ Debt\ Ratio = \frac{Long - term\ Debt}{Total\ Assets}$$

***Model-Based Clustering* (MBC)**

MBC merupakan algoritma *clustering* yang dikembangkan berdasarkan model probabilistik *finite mixture* untuk *probability densities*. Kata "model" dalam metode tersebut biasa digunakan untuk menunjukkan jenis kendala dan sifat geometris dari matriks kovarians (Guojun Gan et. al., 2007). Dalam pendekatan MBC, data dipandang berasal dari campuran distribusi probabilitas, data berdistribusi *finite mixture* atau berdistribusi campuran. Distribusi mixture merupakan distribusi yang tersusun dari beberapa komponen distribusi.

Algoritma *clustering* tradisional seperti *K-means* dan pengelompokan hierarki lainnya adalah algoritma berbasis heuristik yang menurunkan klaster secara langsung berdasarkan data daripada mempertimbangkan ukuran probabilitas dari masing-masing data (Kassambra, 2017). Tidak seperti metode tersebut, MBC mencoba menggunakan *soft assignmen*t, di mana setiap titik data memiliki probabilitas untuk menjadi milik setiap kelompok. MBC menggabungkan korelasi antar variabel ke dalam matriks jaraknya, sehingga lebih fleksibel terhadap bentuk dan ukuran *cluster* (Kessler, 2019).

Menurut Fraley dan Raftery (2002), untuk menemukan jumlah *cluster* dan model *clustering* terbaik, dibutuhkan penyelesaian dari masalah pemilihan model statistik, dimana model dengan jumlah dan jenis distribusi komponen yang berbeda dibandingkan. MBC mengimplementasikan teori ini dengan menggunakan *Expectation Maximization* (EM) yang diinisialisasi dan menggunaan kriteria *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL) untuk membandingkan model. Oleh karena itu, MBC memberikan dasar statistika yang lebih kuat untuk melakukan inferensi dan interpretasi karena berbasis distribusi.

MBC adalah jenis pemodelan *finite mixture*, dengan asumsi bahwa data berasal dari campuran subpopulasi yang berbeda mengikuti distribusi yang diberikan, biasanya menggunaan multivariat normal (Gergely dan Vargha, 2021). *Mixture distribution* adalah distribusi yang terbentuk dari beberap komponen penyusun distribusi. Persamaan *mixture* yang diadaptasi dari Bouveyron et. al. (2019) adalah sebagai berikut:

$$p(\boldsymbol{x}_i) = \sum_{k=1}^{K} \tau_k f_k(\boldsymbol{x}_i|\boldsymbol{\theta_k}) \qquad (1)$$

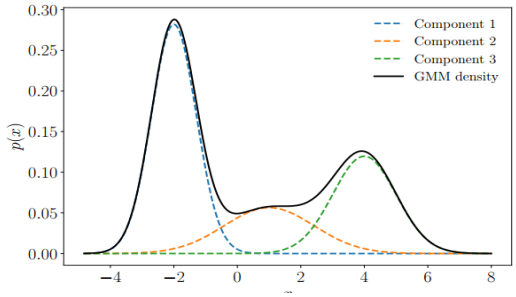
Dalam persamaan (1), $\pi_k$ merupakan bobot probabilitas bahwa suatu pengamatan dihasilkan oleh komponen ke-$k$, dengan batasan $\pi_k \geq 0, k = 1, \dots, K$ dan $\sum_{k=1}^{K} \pi_k = 1$. Sedangkan $f_k(\boldsymbol{x_n}|\boldsymbol{\theta}_k)$ adalah kepadatan komponen ke-$k$ dengan parameter $\boldsymbol{\theta}_k$.

a. *Gaussian Mixture Model* (GMM)

Distribusi probabilitas data yang digunakan pada artikel ini adalah distribusi gaussian, sehingga dari model *finite mixture* sebelumnya akan diarahkan pada *Gaussian Mixture Models* (GMM). GMM adalah model densitas yang terdiri dari sejumlah distribusi $K$ Gaussian $N(\boldsymbol{x}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})$, sehingga dari persamaan (1) GMM bisa dituliskan (Bishop, 2006):

$$p(\boldsymbol{x_i}|\boldsymbol{\theta}_k) = \sum_{k=1}^{K} \tau_k N(x|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}) \quad (2)$$

GMM memberikan pendekatan klasik dan kuat untuk analisis *clustering* (Banfield dan Raftery,1993), GMM juga berguna untuk memahami dan menyarankan kriteria pengelompokan yang kuat (Gan, Ma, dan Wu 2007). Ilustrasi GMM bisa dilihat pada Gambar 1.




Gambar 1. Ilustrasi *Gaussian Mixture Model*.

(Deisenroth, Faisal, & Ong, 2020)

gaussian dan lebih ekspresif dari komponen individu. Garis putus-putus mewakili komponen gaussian tertimbang. Fungsi densitas peluang dari distribusi multivariate GMM adalah sebagai berikut:

$$N(\mathbf{x}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}) = \frac{1}{(2\pi)^{\frac{d}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \quad (3)$$

b. Algoritma *Expectation-Maximization* (EM) untuk GMM

Seperti metode *k-means* yang melakukan iterasi parameter untuk mendapatkan jarak yang konvergen dari pusat *cluster*, sehingga didapatkan hasil *cluster* berdasarkan jarak terdekat. *Model-Based Clustering* (MBC) juga memerlukan iterasi untuk update parameter. Proses iterasi untuk mendapatkan estimasi parameter *clustering* dilakukan dengan menggunakan algoritma EM untuk menyederhanakan perolehan estimasi parameter. Algoritma EM adalah algoritma optimasi iterative untuk memaksimumkan fungsi likelihood dari model probabilistik dengan data yang hilang (*missing data*). Pada artikel ini algoritma EM digunakan untuk memaksimumkan fungsi ekspektasi likelihood dengan data yang tidak lengkap, yaitu data berasal dari komponen distribusi tertentu.

Metode ini merupakan pendekatan paling umum digunakan untuk mencari kemungkinan maksimum data yang terdiri dari $n$ pengamatan multivariate$(x_n, z_n)$, di mana $x_n$ merupakan variabel teramati dan $z_n$ tidak teramati atau variabel laten (label kelompok belum diketahui). Jika label ini diketahui, maka akan didapatkan estimasi parameter di setiap distribusi komponen dengan membagi observasi ke dalam kelompok masing-masing. Tujuan utama dari proses ini adalah untuk mengetahui parameter dari distribusi yaitu $\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}, \pi_k$. Berikut ini adalah algoritma EM untuk mencari parameter distribusi *mixture* pada kasus distribusi gaussian:

1. Inisialisasi parameter awal $\tau^{(0)}, \boldsymbol{\mu^{(0)}}, \boldsymbol{\Sigma^{(0)}}$
2. Langkah Ekspektasi (*E-step*)

   Evaluasi probabilitas posterior $v_{i,k}^{(s+1)}$ untuk setiap data point menggunakan parameter saat ini. Dimana probabilitas posterior diberikan oleh $v_{i,k}^{(s+1)} = \frac{\hat{\tau}_k^{(s)} N_k\left(\boldsymbol{x_i}|\boldsymbol{\theta}_k^{(s)}\right)}{\sum_{j=1}^{K} \tau_j N_j(\boldsymbol{x_i}|\boldsymbol{\theta}_j)}$. Oleh karena penelitian ini menggunakan pendeketan distrubusi *multivariate gaussian*, $v_{i,k}^{(s+1)}$ bisa dituliskan sebagai berikut:

$$v_{i,k}^{(s)} = \frac{\tau_k^{(s)}\left|\boldsymbol{\Sigma}_k^{(s)}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x_i}-\boldsymbol{\mu}^{(s)})^T\left[\boldsymbol{\Sigma}_k^{(s)}\right]^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu}^{(s)})\right)}{\sum \tau^{(s)}\left|\boldsymbol{\Sigma}_l^{(s)}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x_i}-\boldsymbol{\mu}^{(s)})^T\left[\boldsymbol{\Sigma}_l^{(s)}\right]^{-1}(\boldsymbol{x_i}-\boldsymbol{\mu}^{(s)})\right)} \quad (4)$$

3. Langkah Maksimisasi (*M-step*)

   Langka Maksimiasi merupakan estimasi ulang parameter $\tau^{(s)}, \boldsymbol{\mu^{(s)}}, \boldsymbol{\Sigma^{(s)}}$ menggunakan $v_{i,k}^{(s+1)}$ yang diperoleh dari *E-step*. Langkah ini dilakukan dengan persamaan yang diperoleh dari turunan ekspektasi log likelihood dengan parameter yang akan dicari. Menurut Deisenroth, Faisal, & Ong (2020),

turunan untuk mendapatkan nilai estimasi parameter dapat dilakukan dengan langkah-langkah berikut:

a.) Mendapatkan bobot ($\tau_k$)

Untuk menemukan turunan parsial log likehood dari bobor parameter $\tau_k, k = 1, \dots, K$, dengan kendala $\sum_{k=1}^{K} \tau_k = 1$ dapat digunakan *Lagrange multipliers* sebagai berikut:

$$\mathcal{L} = \sum_{i=1}^{n} \log \sum_{k=1}^{K} \tau_k N(\boldsymbol{x_i}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}) + \lambda\left(\sum_{k=1}^{K} \tau_k - 1\right) \quad (5)$$

Maka untuk menentukan turunan partial terhadap $\pi_k$ adalah sebagai berikut:

$$\frac{\partial \mathcal{L}}{\partial \pi_k} = \sum_{i=1}^{n} \frac{N(x_i|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})}{\sum_{j=1}^{K} \pi_j N(x|\boldsymbol{\mu_j}, \boldsymbol{\Sigma_j})} + \lambda$$

$$= \frac{1}{\tau_k} \sum_{n=1}^{N} v_{i,k}^{(s+1)} + \lambda$$

Dan turunan parsial terhadap $\lambda$ adalah sebagai berikut:

$$\frac{\partial \mathcal{L}}{\lambda} = \sum_{k=1}^{K} \tau_k - 1$$

Dengan menyamakan kedua persamaan diatas menjadi 0, diperoleh

$$\tau_k = -\left(\sum_{i=1}^{n} v_{i,k}^{(s+1)}\right)/\lambda \quad (6)$$

$$\sum_{k=1}^{K} \tau_k = 1 \quad (7)$$

Dari persamaan diatas $\pi_k^{(s+1)}$ diperoleh:

$$\sum_{k=1}^{K} \tau_k = 1$$

$$-\sum_{k=1}^{K}\left(\sum_{i=1}^{n} v_{i,k}^{(s+1)}\right)/\lambda = 1$$

$$-\sum_{k=1}^{K}\left(\sum_{i=1}^{n} v_{i,k}^{(s+1)}\right) = \lambda$$

Substitusikan $\lambda$ ke persamaan (6), sehingga didapatkan:

$$\tau_k^{(s+1)} = \frac{\sum_{i=1}^{n} v_{i,k}^{(s+1)}}{\sum_{k=1}^{K}\sum_{i=1}^{n} v_{i,k}^{(s+1)}} \quad (8)$$

b.) Mendapatkan $\boldsymbol{\mu}$

$$\frac{\partial \log p(\boldsymbol{x}_i|\boldsymbol{\theta})}{\partial \boldsymbol{\mu_k}} = \sum_{i=1}^{n} \frac{1}{p(\boldsymbol{x_i}|\boldsymbol{\theta})} \frac{\partial p(\boldsymbol{x_i}|\boldsymbol{\theta})}{\partial \mu_k}$$

$$= \sum_{i=1}^{n} \frac{1}{p(\boldsymbol{x_i}|\boldsymbol{\theta})}\left(\sum_{j=1}^{K} \tau_j \frac{\partial N(\boldsymbol{x_i}|\boldsymbol{\mu_j}, \boldsymbol{\Sigma_j})}{\partial \mu_k}\right)$$

$$= \sum_{i=1}^{n} \frac{1}{p(\boldsymbol{x_i}|\boldsymbol{\theta})}\left(\tau_k \frac{\partial N(\boldsymbol{x_i}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})}{\partial \mu_k}\right)$$

$$= \sum_{i=1}^{n} \frac{1}{p(\boldsymbol{x_i}|\boldsymbol{\theta})}(\tau_k(\boldsymbol{x_i} - \boldsymbol{\mu_k})\boldsymbol{\Sigma}^{-1} N(\boldsymbol{x_i}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}))$$

$$= \sum_{i=1}^{n} \frac{\tau_k N(\boldsymbol{x_i}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})}{\sum_{j=1}^{K} \tau_j N(x|\boldsymbol{\mu_j}, \boldsymbol{\Sigma_j})}(\boldsymbol{x_i} - \boldsymbol{\mu_k})\boldsymbol{\Sigma}^{-1}$$

$$= \sum_{i=1}^{n} v_{i,k}^{(s+1)}(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma}^{-1}$$

Sehingga dengan menyamakan persamaan di atas dengan 0, diperoleh:

$$\sum_{i=1}^{n} v_{i,k}^{(s+1)} \boldsymbol{x_i} = \sum_{i=1}^{n} v_{i,k}^{(s+1)} \boldsymbol{\mu_k}^{(s+1)}$$

$$\boldsymbol{\mu_k}^{(s+1)} = \frac{\sum_{i=1}^{n} v_{i,k}^{(s+1)} x_n}{\sum_{i=1}^{n} v_{i,k}^{(s+1)}} \quad (9)$$

c.) Mendapatkan $\boldsymbol{\Sigma}$

$$\frac{\partial \log p(x_i|\theta)}{\partial \boldsymbol{\Sigma_k}} = \sum_{i=1}^{n} \frac{1}{p(x_i|\theta)} \frac{\partial p(x_i|\theta)}{\partial \boldsymbol{\Sigma_k}}$$

$$= \sum_{i=1}^{n} \frac{1}{p(x_i|\theta)}\left(\frac{\partial}{\partial \boldsymbol{\Sigma_k}} \tau_k (2\pi)^{-\frac{d}{2}} |\boldsymbol{\Sigma_k}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu_k})\right)\right)$$

$$= \sum_{i=1}^{n} \frac{1}{p(x_i|\theta)}\left(\tau_k (2\pi)^{-\frac{d}{2}}\left[\frac{\partial}{\partial \boldsymbol{\Sigma_k}} |\boldsymbol{\Sigma_k}|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu_k})\right) + |\boldsymbol{\Sigma_k}|^{-\frac{1}{2}} \frac{\partial}{\partial \boldsymbol{\Sigma_k}} \exp\left(-\frac{1}{2}(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu_k})\right)\right]\right)$$

Dengan menggunakan identitas untuk menghitung gradien berdasarkan Deisenroth, Faisal, dan Ong (2020), $\frac{\partial}{\partial \boldsymbol{\Sigma_k}} |\boldsymbol{\Sigma_k}|^{-\frac{1}{2}}$ dan $\frac{\partial}{\partial \boldsymbol{\Sigma_k}} (x_i - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1}(x_i - \boldsymbol{\mu_k})$ dapat diselesaikan sebagai berikut:

$$\frac{\partial}{\partial \boldsymbol{\Sigma_k}} |\boldsymbol{\Sigma_k}|^{-\frac{1}{2}} = -\frac{1}{2} |\boldsymbol{\Sigma_k}|^{-\frac{1}{2}} \boldsymbol{\Sigma_k}^{-1}$$

$$\frac{\partial}{\partial \boldsymbol{\Sigma_k}} (x_i - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1}(x_i - \boldsymbol{\mu_k}) = -\boldsymbol{\Sigma_k}^{-1}(x_i - \boldsymbol{\mu_k})(x_i - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1}$$

Sehingga,

$$\frac{\partial \log p(\boldsymbol{x_i}|\boldsymbol{\theta})}{\partial \boldsymbol{\Sigma_k}} = \sum_{i=1}^{n} \frac{1}{p(x_i|\theta)}\left(\tau_k N(\boldsymbol{x_i}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})\left[-\frac{1}{2}\left(\boldsymbol{\Sigma_k}^{-1} - \boldsymbol{\Sigma_k}^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu_k})(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1}\right)\right]\right) \quad (8)$$

$$= \sum_{i=1}^{n} \frac{\tau_k N(x_i|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k})}{\sum_{j=1}^{K} \tau_j N(x|\boldsymbol{\mu_j}, \boldsymbol{\Sigma_j})}\left[-\frac{1}{2}(\boldsymbol{\Sigma_k}^{-1} - \boldsymbol{\Sigma_k}^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu_k})(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1})\right]$$

$$= -\frac{1}{2}\sum_{i=1}^{n} v_{i,k}^{(s+1)}(\boldsymbol{\Sigma_k}^{-1} - \boldsymbol{\Sigma_k}^{-1}(\boldsymbol{x_i} - \boldsymbol{\mu_k})(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \boldsymbol{\Sigma_k}^{-1})$$

$$= -\frac{1}{2}\boldsymbol{\Sigma_k}^{-1}\sum_{i=1}^{n} v_{i,k}^{(s+1)} + \frac{1}{2}\boldsymbol{\Sigma_k}^{-1}\left(\sum_{i=1}^{n} v_{i,k}^{(s+1)}(\boldsymbol{x_i} - \boldsymbol{\mu_k})(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T\right)\boldsymbol{\Sigma_k}^{-1}$$

Oleh karena $\sum_{i=1}^{n} v_{i,k}^{(s+1)} = N_k$ dan dengan menyamakan persamaan di atas dengan 0, maka:

$$N_k \boldsymbol{\Sigma}_{\mathbf{k}}^{-1} = \boldsymbol{\Sigma}_{\mathbf{k}}^{-1} \left( \sum_{i=1}^{n} v_{i,k}^{(s+1)} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \right) \boldsymbol{\Sigma}_{\mathbf{k}}^{-1}$$

$$N_k \boldsymbol{I} = \boldsymbol{\Sigma}_{\mathbf{k}}^{-1} \left( \sum_{i=1}^{n} v_{i,k}^{(s+1)} (\boldsymbol{x_n} - \boldsymbol{\mu_k})(\boldsymbol{x_n} - \boldsymbol{\mu_k})^T \right)$$

$$\frac{\boldsymbol{I}}{\boldsymbol{\Sigma}_{\mathbf{k}}^{-1}} = \frac{1}{N_k} \left( \sum_{i=1}^{n} v_{i,k}^{(s+1)} (\boldsymbol{x_i} - \boldsymbol{\mu_k})(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \right)$$

Sehingga $\boldsymbol{\Sigma}_{\boldsymbol{k}}^{(s+1)}$ dapat dituliskan:

$$\boldsymbol{\Sigma}_{\boldsymbol{k}}^{(s+1)} = \frac{1}{\sum_{i=1}^{n} v_{i,k}^{(s+1)}} \left( \sum_{i=1}^{n} v_{i,k}^{(s+1)} (\boldsymbol{x_i} - \boldsymbol{\mu_k})(\boldsymbol{x_i} - \boldsymbol{\mu_k})^T \right) \quad (10)$$

4. Ulangi terus-menerus langkah tersebut sampai mendapatkan hasil iterasi yang konvergen jika iterasi belum konvergen kembali ke langkah b).
5. Batas Geometris *Multivariate Gaussian Mixture Model*

Parameter yang perlu diduga untuk $\boldsymbol{\Sigma}_{\boldsymbol{k}}$ dipersamaan (7) terlalu banyak sehingga dilakukan pendekatan baru untuk menduga $\boldsymbol{\Sigma}_{\boldsymbol{k}}$**.** Banfield dan Raftery (1993) mengusulkan kerangka umum untuk batasan *cross-cluster* geometris dalam *multivariate mixture model* dengan parameterisasi matriks kovarians melalui dekomposisi nilai eigen, hal tersebut bisa dilakukan dengan persamaan berikut (Guojun Gan et.al, 2007):

$$\boldsymbol{\Sigma}_{\boldsymbol{k}} = \boldsymbol{\lambda}_{\boldsymbol{k}} \boldsymbol{D}_{\boldsymbol{k}} \boldsymbol{A}_{\boldsymbol{k}} \boldsymbol{D}_{\boldsymbol{k}}^{\boldsymbol{T}} \quad (11)$$

Menurut Bouveyron et. al. (2019) terdapat 14 kemungkinan model dari matriks kovarians $\boldsymbol{\Sigma}_{\boldsymbol{k}}$ dalam MBC (lihat Tabel 2.1 pada buku Bouveyron et. al., 2019), 14 model tersebut berbeda untuk data, bentuk (bola atau ellipsoidal) dan volume. Dalam kasus model ellipsoidal, penyelarasan sumbu dan perbedaan bentuk ellipsoidal yang dipasang ditentukan. Ini dikenal sebagai dekomposisi *Volume-Shape-Orientation* (VSO). Untuk model tertentu, volume, bentuk, dan orientasi dapat dibatasi ke varian yang sama, dilambangkan dengan 'E'. Jika varians bebas berubah, model dilambangkan 'V'. Selain itu, orientasi klaster relatif terhadap satu sama lain dapat dibatasi ke *Equal* (sama) atau *Varying* (bervariasi), atau model dapat memiliki keselarasan terbatas pada sumbu koordinat, dan diberi label 'I'. Sebagai contoh (Kassambara, 2017):

1. EVI menunjukkan model di mana volume semua *cluster* adalah sama (E), bentuk cluster dapat bervariasi (V), dan orientasinya adalah identitas (I) atau "sumbu koordinat".
2. EEE berarti *cluster* memiliki volume, bentuk, dan orientasi yang sama dalam ruang dimensi-d.
3. VEI berarti *cluster* memiliki volume variabel, bentuk dan orientasi yang sama dengan sumbu koordinat.

MBC menerapkan 14 model dan mengidentifikasikan salah satu yang model yang paling mendefinisikan data. Untuk memilih model yang optimal penelitian ini akan menggunakan *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL), kemudian hasil clustering terpilih akan dibandingkan dengan metode *K-means* menggunakan indeks *Davies-Bouldin* (DB).

## HASIL DAN PEMBAHASAN

Penelitian ini mendeskripsikan algoritma *Mode-Based Clsutering* (MBC) untuk membentuk *cluster* perusahaan terlebih dahulu berdasarkan nilai rasio keuangan. Selanjutnya Eksplorasi *cluster* akan dilakukan serta rata-rata r*eturn* dari setiap *cluster* akan digunakan penulis sebagai ukuran untuk menilai *cluster* yang terbentuk

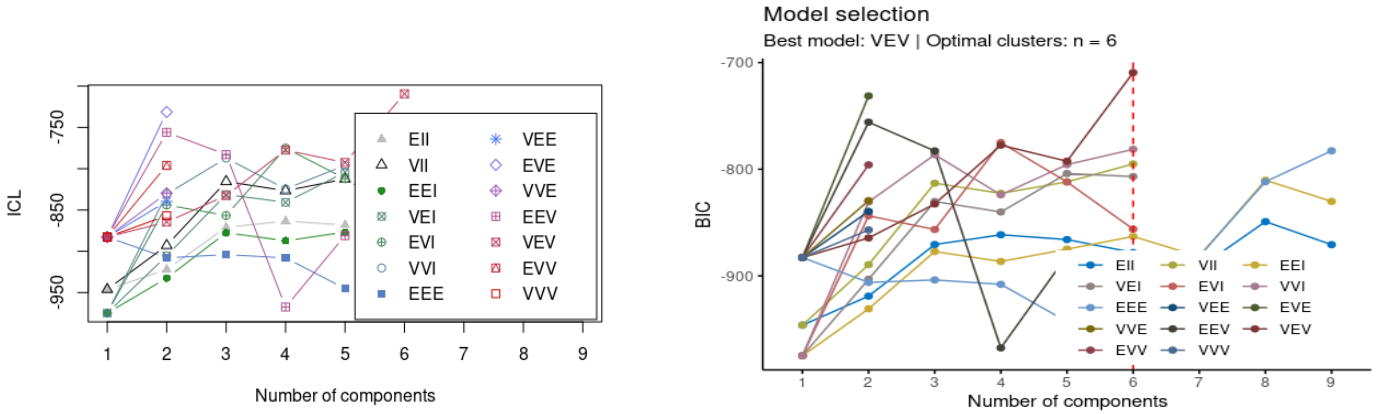
Proses pengelompokan saham berdasarkan nilai rasio keuangan dilakukan dengan menggunakan fungsi *Mclust* pada *package mclust* (Fraley, Raftery, dan Scrucca, 2016) pada program R. Jumlah model yang mampu diidentifikasi menggunakan *package* tersebut ada sebanyak 14 model dengan jumlah kelompok maksimal 9 kelompok apabila tidak ada spesifikasi jumlah kelompok tertentu dari peneliti. Estimasi parameter MBC menggunakan *gaussian mixture model* dan dilakukan dengan algoritma *Expectation–Maximization* (EM). Pemilihan model terbaik didasarkan pada kriteria *Bayesian information criterion* (BIC) dan *Completed Likelihood Criterion* (ICL). Model dengan nilai BIC dan ICL tertinggi diantara semua iterasi yang dilakukan akan dipilih menjadi model terbaik.

Jumlah komponen yang diperoleh ekuivalen dengan jumlah *cluster* optimal. Proses perhitungan MBC menggunakan program R, mengidentifikasi model optimal dan jumlah cluster untuk data. Hasil

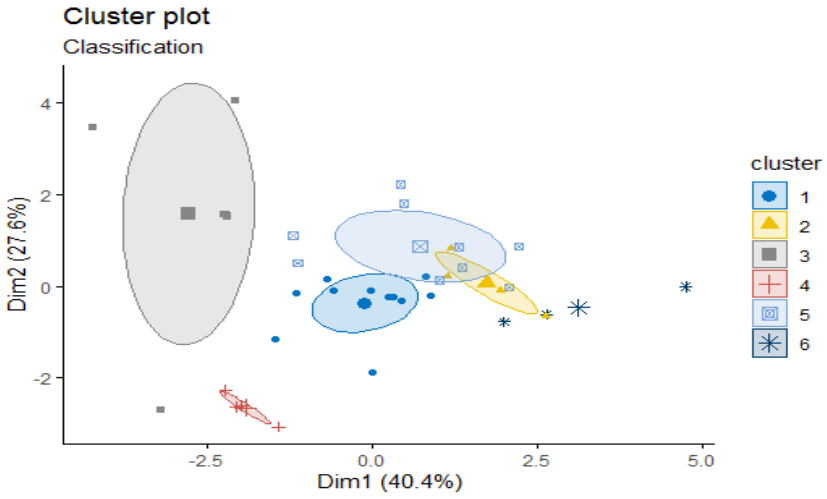
perhitungan masing-masing kriteria dengan nilai BIC dan ICL terbaik pada masing-masing model dapat dilihat pada Tabel 1. Dari Tabel 1

Tabel 1. BIC dan ICL Masing-masing Model

| Model | *Bayesian Information Criterion* (BIC) | Jumlah *Cluster* | *Completed Likelihood Criterion* (ICL) | Jumlah *Cluster* |
|---|---|---|---|---|
| EII | -849,104 | 8 | -851,725 | 8 |
| VII | -795,065 | 6 | -795,53 | 6 |
| EEI | -810,312 | 8 | -810,342 | 8 |
| VEI | -803,994 | 5 | -804,218 | 5 |
| EVI | -775,037 | 4 | -775,043 | 4 |
| VVI | -781,276 | 6 | -781,443 | 6 |
| EEE | -782,71 | 9 | -782,724 | 9 |
| VEE | -839,531 | 2 | -839,667 | 2 |
| EVE | -731,269 | 2 | -731,296 | 2 |
| VVE | -829,456 | 2 | -829,693 | 2 |
| EEV | -755,865 | 2 | -755,94 | 2 |
| VEV | -709,3757 | 6 | -709,376 | 6 |
| EVV | -795,893 | 2 | -795,894 | 2 |
| VVV | -856,887 | 2 | -856,891 | 2 |




Gambar 2. Pemilihan Model Terbaik Berdasarkan Nilai BIC dan ICL




Gambar 3. Visualisasi Hasil *Clustering Model-Based Clustering*

diketahui bahwa dari masing-masing kriteria model, nilai BIC dan ICL tertinggi sama-sama pada model VEV dengan nilai BIC dan ICL berturut-turut sebesar -709,3757 dan -709,376, dan jumlah komponen sebanyak 6 komponen. Visualiasi pemilihan model terbaik dapat dilihat pada Gambar 2. Dalam model VEV (*Variable volume, Equal shape, Variable orientation*), V berarti bahwa *cluster* memiliki volume yang bervariasi yang ditunjukan dengan jumlah anggota *cluster* yang berbeda, E berarti setiap dimensi memiliki variansi yang kurang lebih sama, dan V berarti memiliki orientasi yang bervariasi ini artinya *cluster* tidak mengikuti garis sumbu (bisa dilihat pada Gambar 2, tidak seperti model dengan huruf ketiga I, model dengan huruf ketiga V memiliki arah yang berbeda). Visualisasi setiap kelompok bisa dilihat pada gambar 3. Model dan jumlah *cluster* yang terpilih memberikan hasil *clustering* yang lebih baik dibandingkan menggunakan metode *K-means* dengan jumlah *cluster* optimal sebanyak 3 *cluster* berdasarkan indeks *Davies-Bouldin* (DB). Hasil perbandingan tersebut dapat dilihat pada Tabel 4 berikut :

Tabel 4. Perbandingan Indeks *Davies-Bouldin* (DB) MBC dengan *K-means*

| Metode | Indeks DB |
|---|---|
| *Model-Based Clustering* | 1,253362 |
| *K-Means* | 1,374315 |

Seperti yang dapat dilihat pada Tabel 4, dengan menggunakan indeks DB, metode MBC memiliki nilai yang lebih kecil dibandingkan dengan metode *K-means*. Sehingga berdasarkan kriteria penilaian indeks DB, metode MBC memberikan hasil yang lebih baik (Charrad et al. 2010).

Estimasi moodel *finite mixture* seperti pada persamaan (2) yang diperoleh dari output fungsi *mclust* pada program R adalah sebagai berikut:

$$\begin{aligned} p(\boldsymbol{x_n}) = {} & 0{,}30555556 N_1(\boldsymbol{x_n}|\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) \\ & + 0{,}11111111 N_2(\boldsymbol{x_n}|\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2) \\ & + 0{,}13888889\, N_3(\boldsymbol{x_n}|\boldsymbol{\mu}_3, \boldsymbol{\Sigma}_3) \\ & + 0{,}11111111 N_4\, (\boldsymbol{x_n}|\boldsymbol{\mu}_4, \boldsymbol{\Sigma}_4) \\ & + 0{,}25000000\, N_5(\boldsymbol{x_n}|\boldsymbol{\mu}_5, \boldsymbol{\Sigma}_5) \\ & + 0{,}08333333\, N_6(\boldsymbol{x_n}|\boldsymbol{\mu}_6, \boldsymbol{\Sigma}_6) \end{aligned}$$

Dengan,

$$N(\mathbf{x}|\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma_k}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x_n - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(x_n - \boldsymbol{\mu}_k)\right)$$

untuk $k = 1,2,3,4,5,6$

Hasil estimasi untuk masing-masing komponen adalah sebagai berikut:

a. *Means* ($\boldsymbol{\mu}_k$)

$$\boldsymbol{\mu}_1 = \begin{bmatrix} 0{,}2016216 \\ 0{,}2869895 \\ -0{,}4182307 \\ 0{,}2075350 \\ -0{,}2597670 \\ -0{,}2265134 \\ -0{,}3204897 \\ -0{,}3771211 \\ -0{,}3684507 \end{bmatrix} \quad \cdots \quad \boldsymbol{\mu}_6 = \begin{bmatrix} 1{,}6730409 \\ 0{,}6308741 \\ 0{,}1272041 \\ 0{,}4878447 \\ -0{,}8976315 \\ -0{,}7618192 \\ 1{,}6083048 \\ 1{,}9652379 \\ 0{,}8360117 \end{bmatrix}$$

b. Matriks kovarians ($\boldsymbol{\Sigma}_\mathrm{k}$)

$$\boldsymbol{\Sigma_1} = \begin{bmatrix} 44844685 & 27248523 & \cdots & 26557783{,}9 \\ 27248523 & 17454993 & \cdots & 8509959{,}8 \\ \vdots & \vdots & \ddots & \vdots \\ 26557784 & 8509960 & \cdots & 101467242{,}0 \end{bmatrix}$$

$$\vdots$$

$$\boldsymbol{\Sigma_6} = \begin{bmatrix} 0{,}026483 & 0{,}006483 & \ldots & 0{,}035579 \\ 0{,}006483 & 0{,}002005 & \cdots & 0{,}007546 \\ \vdots & \vdots & \ddots & \vdots \\ 0{,}035578 & 0{,}00755 & \cdots & 0{,}648961 \end{bmatrix}$$

Anggota bagi masing-masing *cluster* dipilih berdasarkan nilai peluang menggunakan persamaan (4). Sampel dengan peluang tertinggi dari keenam *cluster* akan masuk dalam *cluster* tersebut. Besaran peluang dapat dilihat pada bagian tabel *z* dari output fungsi *Mclust.* Peluang sampel pada masing-masing *cluster* dapat dilihat pada Tabel 2.

Sehingga dari Tabel 2 diperoleh rincian hasil pengelompokan rasio keuangan pada masing-masing saham sebagai berikut:

a. *Cluster* 1 : ADHI, AKRA, ANTM, ASII, ELSA, EXCL, INDF, PTPP, SMGR, TLKM, dan WIKA
b. *Cluster* 2 : ADRO, PTBA, TPIA, dan UNTR
c. *Cluster* 3 : ASRI, BBTN, BKSL, LPKR, dan WSKT
d. *Cluster* 4 : BBCA, BBNI, BBRI, dan BMRI

e. *Cluster* 5 : BSDE, ICBP, INDY, INKP, JSMR, MEDC, PGAS, SRIL, dan SSMS
f. *Cluster* 6 : CPIN, KLBF, dan MNCN

Setelah *cluster* terbentuk, bisa diketahui karakteristik data pada masing-masing kelompok berdasarkan nilai rata-rata masing-masing rasio, yang bisa dilihat pada Tabel 3. Dari Tabel 3 bisa diketahui karakteristik dari masing-masing kelompok yang terbentuk berdasarkan nilai rasio, dapat dijelaskan sebagai berikut:

a. *Cluster* 1

*Cluster* 1 bisa dibilang memiliki rata-rata nilai rasio yang stabil dibandingkan *cluster* lainnya. Tidak ada nilai yang ekstrim, seperti nilai rata-rata tertinggi atau yang terendah. Akan tetapi, apabila dibandingkan dengan rata-rata rasio secara keseluruhan, *cluster* 1 memiliki rasio profitabilitas yang lebih tinggi. Ini artinya kemampuan *cluster* 1 dalam menghasilkan profit masih di atas rata-rata.

Untuk ukuran risiko perusahaan, cluster ini memiliki risiko yang relative rendah, mengingat nilai rasio solvabilitas cluster ini lebih kecil dibandingkan rata-rata rasio solvabilitas secara keseluruhan. Akan tetapi, kemampuan dalam membiayai kegiatan operasional perusahaan lebih rendah dari rata-rata rasio likuiditas secara keseluruhan.

b. *Cluster* 2

*Cluster* 2 memiliki *cash rasio* yang paling tinggi dibandingkan dibandingkan *cluster* yang lain, sehingga *cluster* ini memiliki kemampuan yang lebih tinggi dibandingkan *cluster* lain dalam membiayai kegiatan operasional perusahaan. Akan tetapi, hal ini tidak sejalan dengan kemampuan *cluster* ini dalam menghasilkan profit, salah rasio profitabilitas *cluster* ini terendah dibandingkan *cluster* yang lain. Untuk ukuran risiko perusahaan, cluster ini memiliki risiko yang relative rendah dibandingkan *cluster* lain, mengingat nilai rasio solvabilitas *cluster* ini dekat dengan nilai rasio solvabilitas terendah.

c. *Cluster* 3

*Cluster* 3 memiiliki nilai rasio likuiditas yang lebih rendah dibandingkan rata-rata rasio likuiditas secara keseluruhan, selain itu rasio likuiditas *cluster* 3 juga relatif dekat dengan rasio likuiditas terendah, sehingga *cluster* ini memiliki kemampuan dalam membiayai kegiatan operasional perusahaan lebih rendah dari rata-rata. *Cluster* 3 juga memiiliki nilai rata-rata RoA, RoE, dan *net profit margin* terendah diandingkan dengan *cluster* lainnya, artinya secara garis besar kemampuan perusahaan *cluster* 3 dalam menghasilkan profit untuk perusahaan lebih rendah dibandingkan dengan cluster lainnya. Risiko perusahaan pada *cluster* 3 juga cukup besar, hal ini dikarenakan salah satu rasio solvabilitas pada *cluster* ini yaitu *debt to equity ratio* paling rendah diantara *cluster* lainnya.

d. *Cluster* 4

*Cluster* 4 memiiliki nilai rata-rata *cash ratio* dan *quick ratio* terendah diandingkan dengan *cluster* lainnya, sehingga *cluster* ini memiliki kemampuan yang lebih rendah dalam membiayai kegiatan operasional perusahaan. Meskipun demikian, kemampuan perusahaan perusahaan pada *cluster* ini dalam menghasilkan profit sangat baik, hal ini ditunjukan dengan rasio profitabilitas *cluster* 4 yang lebih tinggi dari *cluster* lainnya yaitu *gross profit margin* dan *net profit margin.* Untuk ukuran risiko perusahaan, *cluster* ini memiliki risiko yang relatif rendah dibandingkan *cluster* lain, mengingat salah satu nilai rasio solvabilitas *cluster* ini adalah yang terendah.

e. *Cluster* 5

*Cluster* 5 memiliki kemampuan dalam membiayai kegiatan operasional perusahaan yang cukup baik, hal ini ditunjukan dengan rasio likuiditas *cluster* ini dekat dengan nilai rasio likuidtas tertinggi. Akan tetapi, hal ini tidak sejalan dengan kemampuan *cluster* ini dalam menghasilkan profit, rasio profitabilitas *cluster* ini lebih rendah dari rata-rata rasio profitabilitas secara keseluruhan dan relatif dekat dengan yang terendah. Risiko perusahaan *cluster* ini juga cukup tinggi, mengingat salah satu nilai rasio solvabilitas *cluster* ini adalah yang tertinggi.

f. *Cluster* 6

*Cluster* 6 memiliki nilai rata-rata *current ratio* dan *quick ratio* tertinggi diandingkan dengan *cluster* lainnya, sehingga cluster ini memiliki kemampuan yang lebih tinggi dibandingkan *cluster* lain dalam membiayai kegiatan operasional perusahaan. Dalam menghasilkan profit, *cluster* 6 juga memiliki

Tabel 2. Peluang pada Maing-Masing *Cluster*

| Kode Saham | *Cluster* 1 | *Cluster* 2 | *Cluster* 3 | *Cluster* 4 | *Cluster* 5 | *Cluster* 6 |
|---|---|---|---|---|---|---|
| ADHI | 1 | 0 | 0 | 0 | 6.96E-69 | 0 |
| ADRO | 2.57E-72 | 1 | 0 | 0 | 5.60E-130 | 0 |
| AKRA | 1 | 0 | 0 | 0 | 4.70E-146 | 0 |
| ANTM | 1 | 0 | 0 | 0 | 5.68E-188 | 0 |
| ASII | 1 | 0 | 0 | 0 | 2.77E-98 | 0 |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| UNTR | 7.40E-130 | 1 | 0 | 0 | 7.67E-200 | 0 |
| WIKA | 1 | 0 | 0 | 0 | 6.49E-66 | 0 |
| WSKT | 0 | 0 | 1 | 0 | 0 | 0 |

Tabel 3. Rata-rata Masing-Masing Rasio di Setiap *Cluster*

| Rasio | *Cluster* 1 | *Cluster* 2 | *Cluster* 3 | *Cluster* 4 | *Cluster* 5 | *Cluster* 6 | Rata-rata Rasio |
|---|---|---|---|---|---|---|---|
| *Cash Ratio* | 0,3538 | 1,0472 | 0,1261 | 0,0912 | 0,7380 | 0,8582 | 0,5081 |
| *Curent Ratio* | 1,1919 | 1,8804 | 1,2208 | 0,2600 | 2,0005 | 3,4979 | 1,5632 |
| *Quick Ratio* | 0,9194 | 1,6314 | 0,4419 | 0,2586 | 1,5953 | 2,2276 | 1,1367 |
| *RoA* | 0,03072 | 0,0492 | -0,0637 | 0,0132 | 0,0109 | 0,1123 | 0,0195 |
| *RoE* | 0,0751 | 0,0775 | -0,3072 | 0,0895 | 0,0111 | 0,1484 | 0,0139 |
| *Gross Profit Margin* | 0,2606 | 0,2024 | 0,3566 | 0,9798 | 0,3629 | 0,4250 | 0,3867 |
| *Net Profit Margin* | 0,05161 | 0,0809 | -0,6217 | 0,2529 | 0,0228 | 0,1427 | -0,0159 |
| *Long-Term Debt Ratio* | 0,1528 | 0,1020 | 0,2232 | 0,0350 | 0,3674 | 0,0616 | 0,1899 |
| *Debt to Equity Ratio* | 2,1719 | 0,6684 | 6,5092 | 6,2590 | 2,4089 | 0,2987 | 2,9645 |

kemampuan yang sangat baik, hal ini ditunjukkan dengan RoA dan RoA tertingi diandingkan dengan *cluster* lainnya. Untuk ukuran risiko perusahaan, *cluster* ini memiliki risiko yang rendah dibandingkan *cluster* lain, mengingat *debt to equity ratio cluster* ini adalah yang terendah dan nilai *long term debt ratio cluster* ini juga mendekati nilai terendah.

Berdasarkan karakteristik masing-masing *cluster*, diketahui bahwa mayoritas nilai rasio pada *cluster* 6 memiliki nilai yang paling baik. Perusahaan pada *cluster* ini memiliki kemampuan yang tinggi dibandingkan perusahaan dari *cluster* lain untuk membiayai kegiatan operasional perusahaan dan memenuhi kewajiban jangka pendek. Sedangkan untuk rasio solvabilitas, *cluster* ini memiliki nilai *debt to equity ratio cluster* yang terendah dan nilai *long term debt ratio cluster* yang juga mendekati nilai terendah Hal ini berarti, perusahaan pada *cluster* ini memiliki risiko relatif lebih rendah dibandingkan perusahaan pada *cluster* yang lain. *Cluster* 6 merupakan *cluster* terbaik karena memiliki mayoritas rasio likuiditas dan profabilitas terbaik serta rasio solvabilitas terendah.

## KESIMPULAN DAN SARAN

### Kesimpulan

Dari proses *clustering*, terbentuk 6 *cluster* dengan nilai nilai *Bayesian Information Criterion* (BIC) dan *Completed Likelihood Criterion* (ICL) berturut-turut sebesar -709,3757 dan -709,376, dengan model optimal yang dipilih adalah model VEV.

Berdasarkan nilai rata-rata setiap rasio, *cluster* 6 merupakan *cluster* terbaik karena memiliki mayoritas rasio likuiditas dan profabilitas terbaik serta rasio solvabilitas terrendah. *Cluster* 6 memiliki kemampuan yang tinggi dibandingkan perusahaan *cluster* lain untuk membiayai kegiatan operasional perusahaan dan memenuhi kewajiban keuangannya jangka pendek.

**Saran**

Penelitian ini menggunakan distribusi *Gaussian* dalam *Model-Based Clustering* (MBC), penelitian selanjutnya bisa digunakan distribusi lain seperti distribusi *t multivariate* untuk mengetahui pengelompokan yang lebih baik terutama untuk data dengan pencilan.

**DAFTAR PUSTAKA**


Andy, and Melly Megawati. 2019. "*Analysis of Liquidity , Profitability and Solvency Ratios to Assess the Financial Performance of Companies in Cigarette Industries Listed on the Indonesia Stock Exchange*." *ECo-Fin* 1 (1): 22–34. https://doi.org/https://doi.org/10.32877/ef.v1i1.54.

Andayani, Mery. 2016. "Analisis Rasio Likuiditas Dan Rasio Profitabilitas Terhadap Perubahan Laba." *Jurnal Ilmu Dan Riset Akuntansi* 5 (7): 1–19.

Banfield, Jeffrey D., and Adrian E. Raftery. 1993. "*Model-Based Clustering and Non-Gaussian Clustering*." *Biometrics* 49 (3): 803–21.

Bennett, Mark J., and Dirk L. Hugen. 2016. *Financial Analytics with R: Building a Laptop Laboratory for Data Science*. United Kingdom: Cambridge University Press.

Bishop, Christoper M. 2006. *Pattern Recognition and Machine Learning*. New York: Springer

Bouveyron, Charles, Gilless Celeux, T. Brenden Murphy, and Adrian E. Raftery. 2019. *Model-Based Clustering and Classification for Data Science: With Application in R*. United Kingdom: Cambridge University Press.

Charrad, Malika, Yves Lechevallier, Mohamed Ben Ahmed, and Gilbert Saporta. 2010. "On the Number of Clusters in Block Clustering Algorithms." In *23rd International FLAIRS Conference*, 392–97. Florida, United States.

Deisenroth, Marc Peter, A Aldo Faisal, and Cheng Soon Ong. 2020. *Mathematics For Machine Learning*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108679930

Fraley, Chris, and Adrian E Raftery. 2002. "*Model-Based Clustering , Discriminant Analysis , and Density Estimation*." *Journal of the American Statistical Association* 97 (458): 611–31.

Gan, Guojun, Chaoqun Ma, and Jianhong Wu. 2007. *Data Clustering: Theory, Algorithms, and Applications (ASA-SIAM Series on Statistics and Applied Probability)*. Philadelphia, Pennsylvania: Society for Industrial and Applied Mathematics American.

Gergely, Bence, and András Vargha. 2021. "How to Use Model-Based Cluster Analysis Efficiently in Person-Oriented Research." *Journal for Person-Oriented Research* 7 (1): 22–35. https://doi.org/10.17505/jpor.2021.23449.

Husnan, Suad, Enny Pudjiastuti, 2004. Dasar-Dasar Manajemen Keuangan. Edisi. Keempat, Yogyakarta, UPP AMP YKPN.

Kassambara, Alboukadel. 2017. *Practical Guide To Cluster Analysis in R: Unsupervised Machine Learning (Multivariate Analysis Book 1)*. (http://www.sthda.com): STHDA.

Kessler, Dave. 2019. "Introducing the MBC Procedure for Model-Based Clustering." In *SAS Proceeding*, 1–21.

Mcnicholas, Paul D. 2016. "*Model-Based Clustering*." *Journal of Classificatio* 373 (November): 331–73. https://doi.org/10.1007/s0035.

Munawir, S, 2002. *Akuntansi Keuangan dan Manajemen*, Edisi Pertama, Penerbit BPFE, Yogyakarta.

Qona'ah, Niswatul, Alvita Rachma Devi, and I Made Gde Meranggi Dana. 2020. "*Laboratory Clustering Using K- Means, K-Medoids, and Model- Based*

*Clustering."* Indonesian Journal of Applied Statistics 3 (1): 64–77.

Rist, Michael, and Albert J. Pizzica. 2015. *Financial Ratios For Executives*. New York: Apress Berkeley, CA.

Scrucca L., Fop M., Murphy T. B. and Raftery A. E. (2016) *mclust 5: clustering, classification and density estimation using Gaussian finite mixture models The R Journal* 8/1, pp. 289-317

Shalev-shwartz, Shai, and Shai Ben-david. 2014. *Understanding Machine Learning Machine: From Theory to Algorithms*. New York: Cambridge University Press.

S.R. Nanda, B. Mahanty, M.K. Tiwari (2010) *Clustering Indian Stock Market Data for Portfolio Management. Expert Systems with Applications* 37 (2010) 8793-8798

Subekti, R, E R Sari, and R Kusumawati. 2018. "*Ant Colony Algorithm for Clustering in Portfolio Optimization.*" In *IOP Conf. Series: Journal of Physics: Conf. Series 983*.

Subekti, Retno, Rosita Kusumawati, and Eminugroho Ratna Sari. 2017. "*K-Means Clustering* dan *Average Linkage* dalam Pembentukan Portfolio Saham." In *Seminar Matematika Dan Pendidikan Matematika UNY T-31*, 219–24.

Weston, J. Fred. dan Eugene F. Brigham.1993. Manajemen Keuangan, Edisi Kesembilan. Jilid 1dan 2. Alih Bahas: Alfonsus Sirait. Erlangga, Jakarta.