

CLUSTERING DAFTAR SAHAM BERDASARKAN LIKUIDITAS DAN KAPITALISASI PASAR MENGGUNAKAN ALGORITMA GMM DAN BGM

Angel Patrecia^{1*}, Dian Anubhakti², Kukuh Harsanto³

^{1,2,3} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta Selatan, Indonesia

Email: ^{1*}2112501412@student.budiluhur.ac.id, ²dian.anubhakti@budiluhur.ac.id, ³kukuh.harsanto@budiluhur.ac.id

Abstrak- Indeks saham merupakan alat penting dalam membantu investor menyaring saham unggulan secara efisien. Namun, indeks konvensional seperti LQ45 belum tentu mencerminkan kebutuhan analisis spesifik bagi setiap investor, karena seleksi saham dilakukan berdasarkan kriteria yang ditentukan oleh bursa dan tidak selalu berbasis pada kondisi pasar terkini. Penelitian ini bertujuan untuk membentuk indeks saham alternatif melalui pendekatan klusterisasi terhadap saham papan utama Bursa Efek Indonesia (BEI), berdasarkan dua fitur utama yaitu kapitalisasi pasar (*marketcap*) dan likuiditas (*turnover*). Dua algoritma yang digunakan dalam proses klusterisasi adalah *Gaussian Mixture Model* (GMM) dan *Bayesian Gaussian Mixture* (BGM), yang sama-sama merupakan pendekatan berbasis probabilitas dan mampu mengakomodasi distribusi data yang kompleks. Data saham diperoleh dari platform *Yahoo Finance* dengan cakupan saham-saham papan utama. Data yang diperoleh kemudian diproses melalui tahapan transformasi distribusi data untuk mengatasi skewness, dilanjutkan dengan normalisasi menggunakan *MinMax Scaler*. Selanjutnya, klusterisasi dilakukan untuk mengelompokkan saham berdasarkan pola kemiripan data. Evaluasi kualitas model dilakukan menggunakan *Silhouette Score* untuk memilih jumlah kluster terbaik, serta *Davies-Bouldin Index* (DBI) untuk menilai efektivitas pemisahan antar kluster. Berdasarkan hasil evaluasi, algoritma *Gaussian Mixture Model* (GMM) dengan dua kluster memberikan performa terbaik dalam membedakan karakteristik saham. Hasil penelitian menunjukkan bahwa algoritma *Gaussian Mixture Model* (GMM) dengan dua kluster memberikan performa terbaik dengan *Silhouette Score* sebesar 0,406770 dan DBI sebesar 0,8504, menandakan pemisahan kluster yang cukup jelas. Kluster pertama didominasi oleh saham dengan kapitalisasi pasar dan likuiditas tinggi, sementara kluster kedua berisi saham dengan nilai lebih rendah pada kedua fitur tersebut. Temuan ini menunjukkan bahwa pendekatan GMM dan BGM efektif untuk segmentasi saham secara objektif, serta berpotensi menjadi dasar pembentukan indeks saham alternatif yang lebih fleksibel dan relevan dengan kebutuhan investor modern.

Kata Kunci: Saham, *Gaussian Mixture Model*, *Bayesian Gaussian Mixture*, *Davies-Bouldin Index*.

CLUSTERING OF STOCK LISTINGS BASED ON LIQUIDITY AND MARKET CAPITALIZATION USING THE GMM AND BGM ALGORITHMS

Abstract- Stock indices are essential tools that help investors efficiently filter top-performing stocks. However, conventional indices such as LQ45 may not always reflect the specific analytical needs of every investor, as stock selection is based on criteria set by the exchange and may not necessarily align with current market conditions. This study aims to develop an alternative stock index using a clustering approach on stocks listed on the Indonesia Stock Exchange (IDX) main board, based on two key features: market capitalization (*market cap*) and liquidity (*turnover*). Two algorithms were employed in the clustering process: the *Gaussian Mixture Model* (GMM) and the *Bayesian Gaussian Mixture* (BGM), both of which are probabilistic approaches capable of accommodating complex data distributions. Stock data were obtained from the *Yahoo Finance* platform, focusing on main-board stocks. The collected data underwent a distribution transformation to address skewness, followed by normalization using the *MinMax Scaler*. Subsequently, clustering was performed to group stocks based on data similarity patterns. The quality of the models was evaluated using the *Silhouette Score* to determine the optimal number of clusters and the *Davies-Bouldin Index* (DBI) to assess the effectiveness of cluster separation. Based on the evaluation, the *Gaussian Mixture Model* (GMM) with two clusters delivered the best performance in distinguishing stock characteristics. The results show that the GMM algorithm with two clusters achieved a *Silhouette Score* of 0.406770 and a DBI of 0.8504, indicating relatively clear cluster separation. The first cluster is dominated by stocks with high market capitalization and liquidity, while the second cluster consists of stocks with lower values in both features. These findings indicate that the GMM and BGM approaches are effective for objective stock segmentation and have the potential to serve as the foundation for developing a more flexible and investor-relevant alternative stock index.

Keywords: Stock, *Gaussian Mixture Model*, *Bayesian Gaussian Mixture*, *Davies-Bouldin Index*.

1. PENDAHULUAN

Di Indonesia, indeks saham seperti LQ45 dan IDX30 telah disusun oleh Bursa Efek Indonesia (BEI) berdasarkan kriteria tertentu, termasuk likuiditas dan kapitalisasi pasar, yang diperbarui secara berkala untuk mencerminkan kinerja saham-saham unggulan. Meskipun demikian, indeks konvensional ini memiliki keterbatasan, di antaranya adalah kurangnya fleksibilitas dan kecepatan dalam merespons perubahan pasar yang dinamis, karena evaluasinya hanya dilakukan secara periodik [1]. Selain itu, jumlah kluster optimal serta karakteristik masing-masing kelompok saham belum teridentifikasi secara jelas, sehingga diperlukan penelitian lebih lanjut untuk menentukan struktur pengelompokan yang tepat [2]. Keterbatasan ini mendorong perlunya pendekatan alternatif untuk menyusun indeks saham yang lebih adaptif dan sesuai dengan kebutuhan investor modern.

Penelitian terbaru oleh Jumairi et al pada tahun 2025 [3] mengimplementasikan *K-Means Clustering* untuk mengelompokkan saham di Bursa Efek Indonesia berdasarkan rasio keuangan seperti PER, EPS, BVPS, dan PBV. Hasil penelitian tersebut membuktikan bahwa pendekatan berbasis *clustering* dapat membantu investor memahami karakteristik saham dan profil risikonya. Berbeda dengan penelitian tersebut, studi ini mengembangkan analisis serupa dengan menggunakan pendekatan *Gaussian Mixture Model (GMM)* dan *Bayesian Gaussian Mixture (BGM)*. Pendekatan berbasis probabilistik ini mampu menangkap distribusi data yang lebih kompleks, sehingga menghasilkan segmentasi yang lebih akurat dan fleksibel dalam pengelompokan saham.

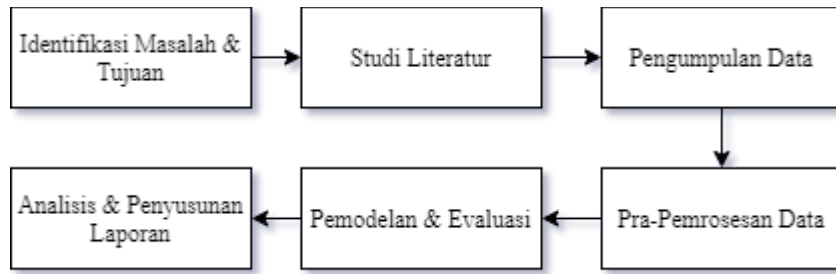
Pendekatan yang diusulkan dalam penelitian ini adalah penerapan teknik *clustering* berbasis data untuk mengelompokkan saham secara objektif. Metode yang digunakan adalah *Gaussian Mixture Model (GMM)* dan *Bayesian Gaussian Mixture (BGM)*, dua algoritma probabilistik yang efektif dalam mengidentifikasi pola kompleks pada data keuangan [4]. Analisis difokuskan pada dua parameter utama, yakni likuiditas (*turnover*) dan kapitalisasi pasar (*market capitalization*), yang juga menjadi dasar seleksi indeks oleh BEI. Penelitian ini bertujuan membentuk indeks saham alternatif dengan menentukan jumlah kluster terbaik pada daftar saham papan utama, sekaligus membandingkan kinerja kedua algoritma tersebut untuk menghasilkan pengelompokan yang optimal.

Berbagai studi sebelumnya telah mengkaji pengelompokan saham menggunakan karakteristik perdagangan seperti volatilitas dan likuiditas [5], serta metrik *expected return* dan *value at risk* [6]. Terdapat pula penelitian yang menerapkan *Model-Based Clustering (GMM)* pada saham LQ45 menggunakan rasio keuangan dan membuktikan keunggulannya dibandingkan *K-Means* [7]. Selain itu, penelitian oleh Tohendry dan Jollyta [8] menunjukkan efektivitas *K-Means Clustering* dalam mengelompokkan saham berdasarkan rasio keuangan seperti PER dan PBV, yang memberikan wawasan tambahan bagi investor. Penelitian ini berbeda karena secara langsung membandingkan GMM dan BGM dalam mengelompokkan saham berdasarkan likuiditas dan kapitalisasi pasar di seluruh segmen papan utama BEI.

Tidak seperti penelitian terdahulu yang menggunakan cakupan data terbatas, penelitian ini memanfaatkan keseluruhan saham papan utama yang aktif diperdagangkan, sehingga hasilnya merepresentasikan kondisi pasar modal secara lebih menyeluruh dan terkini. Selain itu, penelitian ini memanfaatkan kombinasi GMM dan BGM yang jarang digunakan secara bersamaan dalam studi pasar modal untuk menentukan struktur kluster terbaik. Relevansi topik ini semakin kuat mengingat tren pengambilan keputusan investasi berbasis data (*data-driven decision making*) dan penerapan kecerdasan buatan yang terus berkembang. Dengan demikian, penelitian ini diharapkan memberikan kontribusi nyata, baik secara akademis maupun praktis, sebagai landasan bagi pengembangan metode seleksi saham yang lebih responsif terhadap perubahan pasar.

2. METODE PENELITIAN

Seluruh proses penelitian ini dilaksanakan menggunakan *Jupyter Notebook* dengan bahasa pemrograman *Python*. Tahapan yang dilakukan meliputi identifikasi masalah, studi literatur, pengumpulan data saham papan utama Bursa Efek Indonesia, pra-pemrosesan data, pemodelan menggunakan algoritma *Gaussian Mixture Model (GMM)* dan *Bayesian Gaussian Mixture (BGM)*, evaluasi hasil klusterisasi, serta penyusunan laporan akhir. Diagram alir metode penelitian ditunjukkan pada Gambar 1:



Gambar 1. Diagram Alir Metode Penelitian

2.1 Pengumpulan Data

Pengumpulan data dilakukan dari situs resmi Bursa Efek Indonesia (BEI) dan *Yahoo Finance*. Bursa Efek Indonesia menyediakan daftar saham papan utama yang menjadi dasar pemilihan sampel penelitian, sedangkan *Yahoo Finance* menyediakan data sekunder seperti *volume* perdagangan, harga saham, dan kapitalisasi pasar.

Kapitalisasi pasar diambil berdasarkan nilai terbaru yang merujuk pada laporan keuangan perusahaan tahun 2025, sehingga data bersifat statis sesuai kondisi terkini saat pengambilan. Data dari kedua sumber ini kemudian diolah untuk menghasilkan dua parameter utama yang digunakan dalam penelitian, yaitu likuiditas (*turnover*) dan kapitalisasi pasar (*market capitalization*).

Proses pengumpulan data dilakukan secara manual dengan mengunduh daftar saham dari situs BEI dan mengekstrak data pasar dari *Yahoo Finance*, kemudian seluruh data digabungkan ke dalam satu *dataset* terstruktur setelah melalui proses penyamaan format, pembersihan nilai yang hilang (*missing values*), dan pemadanan kode saham untuk menjamin konsistensi serta akurasi data.

2.2 Pra-pemrosesan Data

Pra-pemrosesan data dilakukan untuk memastikan data saham yang digunakan memiliki kualitas baik, konsisten, dan sesuai untuk dianalisis. Proses ini mencakup beberapa tahap yang akan dijelaskan lebih lanjut pada subbab berikut:

2.2.1 Pembersihan Data

Pembersihan data (*data cleaning*) bertujuan untuk menghilangkan elemen data yang tidak konsisten, mengandung kesalahan, atau tidak relevan. Pada penelitian ini, pembersihan dilakukan dengan memeriksa keberadaan data kosong (*missing values*) dan nilai yang tidak wajar, kemudian menghapus (*drop*) baris yang terdeteksi bermasalah agar *dataset* siap digunakan pada tahap analisis.

2.2.2 Transformasi Distribusi Data

Transformasi logaritmik atau transformasi distribusi data dilakukan untuk menyesuaikan distribusi data agar lebih cocok dengan sistem atau metode yang digunakan dalam proses data mining, terutama jika algoritma yang dipakai peka terhadap bentuk distribusi data [9]. Persamaan transformasi distribusi data yang digunakan adalah:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

Keterangan:

1. x : Nilai asli.

2.2.3 Penyamaan Skala Fitur

Penyamaan skala fitur atau normalisasi data dilakukan untuk menyamakan skala nilai antar variabel sehingga tidak ada fitur yang mendominasi proses analisis. Pada penelitian ini digunakan metode *Min-Max Scaling*, yang mengubah seluruh nilai fitur ke dalam rentang [0, 1] agar algoritma *clustering* dapat bekerja secara optimal. Persamaan normalisasi yang digunakan adalah::

$$X_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (2)$$

Keterangan:

1. X : Nilai asli dari data.

2. x_{min} : Nilai minimum dari variabel.
3. x_{max} : Nilai maksimum dari variabel.
4. $X_{normalized}$: Nilai hasil normalisasi.

2.3 Pemodelan

Pemodelan dilakukan untuk membentuk kluster saham papan utama berdasarkan dua fitur utama, yaitu likuiditas (*turnover*) dan kapitalisasi pasar (*market capitalization*). Pada penelitian ini digunakan pendekatan *Model-Based Clustering* dengan membandingkan dua algoritma, yaitu *Gaussian Mixture Model* (GMM) dan *Bayesian Gaussian Mixture* (BGM). Jumlah kluster yang diuji bervariasi, dan model terbaik dipilih berdasarkan nilai *Silhouette Score* tertinggi. Penjelasan dari masing-masing algoritma yang digunakan disajikan pada subbab berikut:

2.3.1 Gaussian Mixture Model

Gaussian Mixture Model (GMM) adalah algoritma *Model-Based Clustering* yang memandang data sebagai gabungan dari beberapa distribusi *Gaussian*. Setiap kluster direpresentasikan oleh parameter rata-rata, kovariansi, dan bobot tertentu, serta memberikan probabilitas keanggotaan bagi setiap data dalam tiap kluster (*soft clustering*). Persamaan GMM untuk data berdimensi lebih dari satu adalah:

$$p(x) = \sum_{k=1}^K \pi_k \cdot N(x | \mu_k, \Sigma_k) \quad (3)$$

Keterangan:

1. K : Jumlah komponen (kluster) *Gaussian*
2. π_k : Bobot campuran dari komponen ke- k
3. μ_k : rata-rata (*mean*) dari kluster ke- k
4. Σ_k : matriks kovarian dari kluster ke- k

2.3.2 Bayesian Gaussian Mixture

Bayesian Gaussian Mixture Model (BGM) adalah perluasan dari GMM yang mengadopsi pendekatan Bayesian. *Bayesian Gaussian Mixture Model* (BGM) adalah metode yang digunakan untuk menganalisis data yang bersifat beragam, dengan asumsi bahwa data tersebut terbentuk dari gabungan beberapa distribusi Gaussian [10]. Formulasi BGM:

1. *Dirichlet prior* untuk bobot kluster

$$p(\phi) = \text{Dir}(\alpha) \quad (4)$$

Keterangan:

- a. ϕ = vektor bobot
- b. α_0 = parameter konsentrasi dari distribusi Dirichlet

2. *Gaussian prior* untuk *mean* kluster

$$p(\mu_k | \Lambda_k) = N(\mu_k | \mu_0, \beta_0 \Lambda_k^{-1}) \quad (5)$$

Keterangan:

- a. μ_k = *mean* dari kluster ke- k
- b. μ_0 = nilai rata-rata awal (*prior*)
- c. β_0 = bobot kekuatan *prior* terhadap *mean*
- d. Λ_k = *precision matrix*

3. *Wishart prior* untuk *precision matrix*

$$p(\Lambda_k) = W \cdot W \cdot v \quad (6)$$

2.3.3 Silhouette

Silhouette Score digunakan untuk mengukur sejauh mana suatu objek mirip dengan kluster asalnya dibandingkan dengan kluster lain, serta berfungsi sebagai alat evaluasi kualitas hasil pengelompokan pada berbagai jumlah kluster yang diuji [11]. Nilai yang lebih tinggi (mendekati 1) menunjukkan bahwa objek tersebut sangat

cocok dengan klasternya dan terpisah dengan baik dari klaster lain. *Silhouette Score* $s(i)$ untuk satu sampel i didefinisikan sebagai *Silhouette* adalah sebagai berikut:

$$s \cdot i \cdot = \frac{b \cdot i \cdot - a \cdot i \cdot}{\dots \cdot a \cdot i \cdot \cdot b \cdot j \cdot} \quad (7)$$

Keterangan:

1. $a(i)$: Rata-rata jarak antara data ke- i dengan semua data lain dalam klasternya sendiri.
2. $b(i)$: Rata-rata jarak minimum antara data ke- i dengan semua data kluster terdekat lainnya.

2.4 Pengujian

Pengujian dilakukan untuk memvalidasi hasil kluster yang dihasilkan, dengan fokus hanya pada algoritme terbaik menggunakan jumlah kluster antara 2 hingga 10. Validasi dilakukan menggunakan *Davies-Bouldin Index* (DBI), yaitu indeks yang mengukur kepadatan suatu kluster berdasarkan rata-rata jarak antara titik-titik dalam kluster dengan pusat kluster. Rumus DBI ditunjukkan sebagai berikut.:

$$DB = \frac{\dots \sum_{i=1}^K \dots \cdot \frac{\sigma_i + \sigma_j}{d \cdot C_i \cdot C_j} \cdot}{K} \quad (8)$$

Keterangan:

1. K : Jumlah kluster.
2. C_i : Pusat dari kluster ke- i .
3. σ_i : Rata-rata jarak antara setiap data dalam kluster ke- i terhadap pusat kluster (C_i).
4. d : Jarak antara pusat kluster C_i dan C_j .

3. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari proses pengumpulan data serta klusterisasi saham papan utama yang telah dilakukan, disertai pembahasan mengenai pola-pola yang terbentuk dari masing-masing metode yang digunakan.

3.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh secara otomatis melalui pemrograman Python dengan bantuan *library pandas* dan *yfinance*. Pengambilan data dilakukan dengan membaca daftar saham papan utama dari file Excel, kemudian melengkapi informasi pasar setiap saham melalui *API Yahoo Finance* dengan penyesuaian *ticker symbol* (.JK). Dari proses ini diperoleh data fundamental terkini berupa kapitalisasi pasar (*market capitalization*), harga rata-rata 50 hari terakhir, dan rata-rata *volume* transaksi.

Seluruh data disimpan dalam satu *dataset* terstruktur yang kemudian diolah lebih lanjut menjadi dua fitur utama penelitian, yaitu likuiditas (*turnover*) dan kapitalisasi pasar (*market capitalization*). *Dataset* ini terdiri dari total 254 baris data setelah proses pembersihan, dari jumlah awal 255 baris data, di mana setiap baris merepresentasikan satu saham papan utama BEI. Deskripsi fitur yang digunakan pada penelitian ini dapat dilihat pada **Tabel 1** berikut:

Tabel 1. Deskripsi Fitur untuk Pembentukan Kluster

No	Fitur	Keterangan
1.	Kode	Kode saham perusahaan yang terdaftar di Bursa Efek Indonesia.
2.	Nama Perusahaan	Nama lengkap perusahaan emiten.
3.	Tanggal Pencatatan	Tanggal pertama kali saham perusahaan dicatatkan di Bursa Efek Indonesia.
4.	Saham	Jumlah total saham yang beredar.
5.	Papan Pencatatan	Kategori papan pencatatan saham (contoh: Utama).
6.	Price	Harga rata-rata saham selama 50 hari terakhir (<i>fifty day average</i>).
7.	Marketcap	Nilai kapitalisasi pasar perusahaan, dihitung dari harga saham x jumlah saham.
8.	Volumeavg	Rata-rata <i>volume</i> perdagangan saham dalam periode tertentu.
9.	Turnoveravg	Rata-rata nilai transaksi perdagangan saham (<i>turnover</i>).

3.2 Pra Pemrosesan Data

Tahapan pra-pemrosesan data mencakup pembersihan data, transformasi distribusi data, dan penyamaan skala fitur. Uraian lebih lanjut dijelaskan pada subbab berikut:

3.2.1 Pembersihan Data

Pembersihan data dilakukan untuk memastikan data saham yang digunakan dalam analisis bersih dan *valid*. Proses ini mencakup penanganan *error*, nilai kosong, serta data duplikat. Hasil pembersihan menunjukkan bahwa tidak terdapat perubahan jumlah data pada tahap penanganan *error* dan duplikat, sedangkan pada tahap penanganan nilai kosong terjadi pengurangan satu data (-0,39%). Ringkasan perubahan data setelah pembersihan disajikan pada **Tabel 2** berikut:

Tabel 2. Deskripsi Hasil Pembersihan Data

No	Nama Kegiatan	Jumlah Data Awal	Jumlah Data Akhir	Persentase Perubahan
1	Penanganan <i>Error</i>	255	255	0,00%
2	Penanganan Nilai Kosong	255	254	-0,39%
3	Penanganan Data Duplikat	254	254	0,0%

3.2.2 Transformasi Distribusi Data

Transformasi distribusi data dilakukan karena fitur *marketcap* dan *turnoveravg* memiliki sebaran yang sangat condong ke kanan (*right-skewed*). Transformasi distribusi data diterapkan pada kedua fitur tersebut untuk mengurangi efek outlier dan menyeimbangkan distribusi. Ringkasan hasil transformasi pada setiap fitur ditampilkan pada **Tabel 3** berikut:

Tabel 3. Jenis Distribusi Data Sebelum dan Sesudah Transformasi

No	Fitur	Jenis Distribusi Data	Proses yang dilakukan
1.	Kapitalisasi Pasar	<i>Right-skewed</i> (positif)	Transformasi distribusi data
2.	Likuiditas	<i>Right-skewed</i> (positif)	Transformasi distribusi data

3.2.3 Penyamaan Skala Fitur

Meskipun distribusi data telah diperbaiki melalui transformasi logaritmik, setiap fitur masih memiliki skala yang berbeda. Perbedaan ini berpotensi membuat fitur dengan nilai lebih besar menjadi dominan dalam pembentukan kluster. Untuk itu digunakan *MinMax Scaler* agar seluruh fitur berada pada rentang 0–1. Skala fitur sebelum normalisasi ditunjukkan pada **Tabel 4** berikut:

Tabel 4. Skala Setiap Fitur Sebelum Penyamaan Skala

No	Fitur	Min	Max	Mean
1.	<i>Turnover</i>	22,358,845	2,781,073,886	2,066,782,695
2.	<i>Marketcap</i>	31,041,251	3,466,734,783	2,776,669,366

Skala data setelah dilakukan penyamaan skala menggunakan metode *MinMax Scaler* ditunjukkan pada **Tabel 5** berikut:

Tabel 5. Skala Setiap Fitur Sesudah Penyamaan Skala

No	Fitur	Min	Max	Mean
1.	<i>Turnover</i>	0	1	0.50
2.	<i>Marketcap</i>	0	1	0.50

3.3 Pemodelan

Pemodelan dilakukan menggunakan data yang telah melalui proses pembersihan dan transformasi. Dua algoritme yang digunakan adalah *Gaussian Mixture Model* dan *Bayesian Gaussian Mixture*.

3.3.1 Komparasi Model

Pemodelan dilakukan menggunakan jumlah kluster 2 hingga 10, kemudian setiap model dievaluasi menggunakan *Silhouette Score*. Hasil komparasi *Silhouette Score* dari algoritme *Gaussian Mixture Model* (GMM) dan *Bayesian Gaussian Mixture* (BGM) ditunjukkan pada **Tabel 6** berikut:

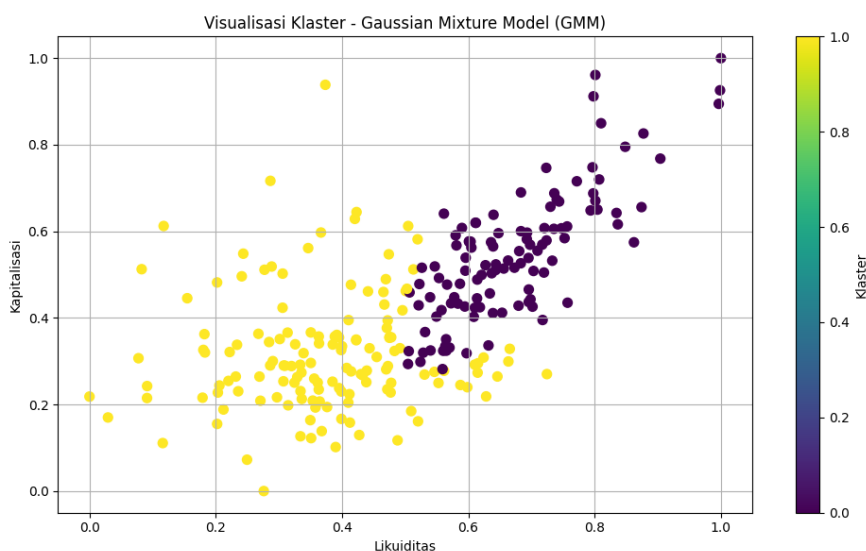
Tabel 6. Hasil Komparasi Model (*Silhouette Score*)

Kluster	Algoritma	Skor <i>Silhouette</i>
0	<i>Gaussian Mixture Model (GMM)</i>	0,406770
1	<i>Bayesian Gaussian Mixture (BGM)</i>	0,362131

Berdasarkan **Tabel 9**, GMM menghasilkan skor tertinggi sebesar 0,406770, sedikit lebih baik dibandingkan BGM dengan skor 0,362131. Oleh karena itu, GMM dipilih sebagai algoritme terbaik dalam penelitian ini.

3.3.2 Penyajian Model Terbaik

Berdasarkan hasil evaluasi, ditetapkan bahwa model GMM dengan dua kluster merupakan model terbaik. Visualisasi kluster yang terbentuk berdasarkan model ini dapat dilihat pada **Gambar 3** berikut:



Gambar 2. Kluster Model Terbaik

Analisis kluster dilakukan untuk mengetahui karakteristik tiap kelompok saham. Ringkasan statistik saham dalam Kluster 0, yang didominasi oleh likuiditas dan kapitalisasi pasar rendah, ditampilkan pada **Tabel 7** berikut:

Tabel 7. Karakteristik Kluster berdasarkan Rata-rata

Statistik	<i>Turnoveravg</i>	<i>Marketcap</i>
Minimum	Rp 3.787.873.120,44	Rp 2.435.292.856.320,00
Maksimum	Rp 1.196.879.485.560,00	Rp 1.137.202.650.152.960,00
Rata-rata	Rp 76.195.085.958,71	Rp 65.320.159.676.708,57

Selain itu, Kluster 1 berisi saham-saham dengan likuiditas dan kapitalisasi pasar yang sangat tinggi, mewakili kelompok big cap yang aktif diperdagangkan. Ringkasan statistik kluster ini dapat dilihat pada **Tabel 8** berikut:

Tabel 8. Hasil Pengujian Menggunakan *Davies-Bouldin Index*

Statistik	<i>Turnoveravg</i>	<i>Marketcap</i>
Minimum	Rp 10.812.071,88	Rp 218.467.205.120,00
Maksimum	Rp 48.740.686.604,40	Rp 671.665.977.556.992,00
Rata-rata	Rp 2.670.182.728,79	Rp 11.288.257.870.025,92

3.4 Pengujian

Untuk mengevaluasi kualitas kluster yang terbentuk, dilakukan pengujian pada algoritme terbaik yaitu *Gaussian Mixture Model* (GMM) dengan jumlah kluster 2 menggunakan metrik *Davies-Bouldin Index* (DBI). DBI digunakan untuk mengukur kepadatan internal suatu kluster sekaligus keterpisahan antar kluster, di mana nilai yang semakin rendah menunjukkan kualitas kluster yang semakin baik.

Berdasarkan hasil perhitungan, diperoleh nilai DBI sebesar 0,8504. Nilai ini menunjukkan bahwa dua kluster yang terbentuk memiliki tingkat keterpisahan yang cukup jelas dengan penyebaran internal yang relatif kecil. Meskipun nilai tersebut belum ideal karena masih cukup jauh dari nol, hasil ini sudah dianggap memadai untuk data saham yang kompleks. Dengan demikian, model GMM dengan dua kluster dinilai layak digunakan dalam analisis lebih lanjut terhadap kelompok saham yang terbentuk.

4. KESIMPULAN

Penelitian ini berhasil mengelompokkan saham papan utama Bursa Efek Indonesia (BEI) menggunakan dua algoritma berbasis probabilistik, yaitu *Gaussian Mixture Model* (GMM) dan *Bayesian Gaussian Mixture* (BGM). Hasil evaluasi menunjukkan bahwa GMM memberikan performa terbaik, dengan *Silhouette Score* sebesar 0,406770 dan *Davies-Bouldin Index* (DBI) sebesar 0,8504, yang menandakan pemisahan kluster yang lebih jelas dibandingkan dengan BGM. Kluster pertama terdiri dari saham dengan kapitalisasi pasar dan tingkat likuiditas tinggi, sedangkan kluster kedua mencakup saham dengan kapitalisasi pasar dan tingkat likuiditas lebih rendah.

Kontribusi utama penelitian ini adalah memberikan pendekatan alternatif dalam pembentukan indeks saham berbasis data yang lebih fleksibel dan sesuai dengan kondisi pasar terkini. Temuan ini dapat menjadi dasar pengembangan indeks alternatif untuk mendukung pengambilan keputusan investasi. Untuk penelitian ke depan, model ini dapat diperluas dengan menambahkan fitur lain seperti volatilitas harga, *volume* perdagangan, atau indikator keuangan lain agar hasil klusterisasi semakin akurat.

DAFTAR PUSTAKA

- [1] L. Winny and Y. Yulfiswandi, "Macroeconomics and the LQ45 Index: Is the COVID-19 pandemic making a difference?," *J. Manaj. Strateg. dan Apl. Bisnis*, vol. 5, no. 2, pp. 217–230, 2022, doi: 10.36407/jmsab.v5i2.612.
- [2] S. Mukherjee, "Bayesian Analysis of Stochastic Volatility Model using Finite Gaussian Mixtures with Unknown Number of Components," pp. 1–15, 2021, [Online]. Available: <http://arxiv.org/abs/2110.12824>
- [3] N. Jumairi, J. Jasmir, and B. Purnama, "Penerapan Data Mining Dalam Pengelompokan Saham Melalui Rasio Keuangan Pada Saham Papan Utama Bursa Efek Indonesia Menggunakan Metode K-Means Clustering," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 3, pp. 4863–4869, 2025, doi: 10.36040/jati.v9i3.13670.
- [4] Y. Zhang *et al.*, "Gaussian Mixture Model clustering with incomplete data," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 17, no. 1s, 2021, doi: 10.1145/3408318.
- [5] B. Siregar and Y. Yosia, "Implementation of K-means Clustering Algorithm for the Indonesian Stock Exchange," *J. Sisfotek Glob.*, vol. 14, no. 1, p. 49, 2024, doi: 10.38101/sisfotek.v14i1.10860.
- [6] A. Fawaid Ridwan and S. Supian, "IDX30 Stocks Clustering with K-Means Algorithm based on Expected Return and Value at Risk," *Int. J. Quant. Res. Model.*, vol. 2, no. 4, pp. 201–208, 2021.
- [7] I. S. D. Hasnida and R. Kusumawati, "Penerapan Model-Based Clustering pada Pengelompokan Saham Berdasarkan Rasio Keuangan," *J. Apl. Stat. Komputasi Stat.*, vol. 15, no. 1, pp. 37–50, 2023, doi: 10.34123/jurnalasks.v15i1.510.
- [8] D. Tohendry and D. Jollyta, "Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Saham Berdasarkan Price Earning Ratio Dan Price To Book Value," *J. Mhs. Apl. Teknol. Komput. dan Inf.*, vol. 5, no. 1, pp. 3–9, 2023.
- [9] R. Raja, K. K. Nagwanshi, S. Kumar, and K. R. Laxmi, *Data Mining and Machine Learning Applications*. Beverly, MA: Scrivener Publishing, 2022.
- [10] B. Grün, "Model-Based Clustering," in *Handbook of Mixture Analysis*, S. Frühwirth-Schnatter, G. Celeux, and C. P. Robert, Eds., Boca Raton, Florida : CRC Press, [2019]: Chapman and Hall/CRC, 2019, ch. 8, pp. 157–188. doi: 10.1201/9780429055911.
- [11] P. Rahayu *et al.*, *Buku Ajar Data Mining*. Bandung: PT. Sonpedia Publishing Indonesia, 2024.