



KEPUTUSAN DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR

NOMOR : K/UBL/FTI/000/002/09/25

TENTANG:

**PENUGASAN KEGIATAN TRI DHARMA & PENUNJANG BAGI DOSEN
FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR
SEMESTER GASAL TAHUN AKADEMIK 2025/2026**

DEKAN FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR

- Menimbang : 1) Bahwa Dosen adalah pendidik profesional dan ilmu dengan tugas utama mentrans-formasikan, mengembangkan, dan menyebarkan ilmu pengetahuan, teknologi, dan seni melalui pendidikan/pengajaran penelitian & karya ilmiah, dan Pengabdian pada masyarakat yang dikenal dengan istilah Tri Dharma Perguruan Tinggi;
- 2) Bahwa untuk meningkatkan profesionalitas dan kompetensi sebagai pendidik profesional maka dipandang perlu untuk memberikan tugas-tugas tambahan/penunjang dalam lingkup kegiatan penunjang Tri Dharma;
- Mengingat : 1) Undang – undang Republik Indonesia Nomor 12 Tahun 2012 tentang Pendidikan Tinggi;
- 2) Undang – undang Republik Indonesia Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional;
- 3) Peraturan Pemerintah Republik Indonesia Nomor 17 Tahun 2010 tentang Pengelolaan dan Penyelenggaraan Pendidikan;
- 4) Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Republik Indonesia Nomor 53 Tahun 2023 tentang Penjaminan Mutu Perguruan Tinggi;
- 5) Keputusan Ketua Yayasan Pendidikan Budi Luhur Cakti Nomor: K/YBLC/KEP/000/216/06/2023 tentang Statuta Universitas Budi Luhur;
- 6) SK YPBLC No: K/YBLC/KEP/000/020/01/24 tanggal 05 Januari 2024 tentang Pengangkatan Para Pejabat Struktural Universitas Budi Luhur Periode 2024-2028

MEMUTUSKAN

- Menetapkan :
PERTAMA : Menugaskan dosen-dosen Fakultas Teknologi Informasi Universitas Budi Luhur untuk melaksanakan kegiatan **Tri Dharma Perguruan Tinggi dan penunjangnya** pada Semester Gasal Tahun Akademik 2025/2026 yang meliputi:
- a. **Kegiatan partisipasi aktif** dalam Pertemuan Ilmiah sebagai Ketua/Anggota/Peserta/Pembicara/Penulis/Narasumber pada kegiatan Seminar, Workshop, Konferensi, Pelatihan, Simposium, Lokakarya, Forum Diskusi, Sarasehan dan sejenisnya;
- b. **Publikasi Ilmiah** pada Prosiding, Jurnal/majalah/surat kabar dan sejenisnya;
- c. **Partisipasi dalam organisasi** profesi, organisasi keilmuan dan/atau organisasi lain yang menunjang kegiatan Tri Dharma Pendidikan Tinggi;
- d. **Pengabdian Kepada Masyarakat (PPM)**, dalam kegiatan terprogram, terjadwal atau insidental;
- KEDUA : Dosen-dosen yang melaksanakan penugasan wajib membuat Laporan Kegiatan, dengan mengikuti pedoman dari Fakultas/Program Studi, sebagai pertanggungjawaban atas kegiatan yang diikuti;
- KETIGA : Kegiatan Tri Dharma yang tidak termasuk dalam surat keputusan ini akan memiliki penugasan tersendiri;
- KEEMPAT : Keputusan ini berlaku sejak tanggal ditetapkan dan akan diubah sebagaimana mestinya apabila di kemudian hari terdapat kekeliruan.

Ditetapkan di : Jakarta

Tanggal : 02 September 2025

Dekan Fakultas Teknologi Informasi



Dr. Ir. Achmad Solichin, S.Kom., M.T.I



**LAMPIRAN KEPUTUSAN DEKAN FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR**

NOMOR : K/UBL/FTI/000/002/09/25

**TENTANG:
PENUGASAN KEGIATAN TRI DHARMA & PENUNJANG BAGI DOSEN
FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR
SEMESTER GASAL TAHUN AKADEMIK 2025/2026**

No	NUPTK	Nama	Program Studi
1	6356750651130093	ABDUL MUIS SOBRI	Teknik Informatika (S1)
2	5934758659137112	ACHMAD ADITYA ASHADUL USHUD	Teknik Informatika (S1)
3	4437767668130323	ACHMAD ARDIANSYAH	Teknik Informatika (S1)
4	7937760661130282	ACHMAD SOLICHIN	Ilmu Komputer (S3)
5	5454763664230162	AGNES ARYASANTI	Sistem Informasi (S1)
6	1947743644130112	AGUNG PRIHARTONO	Sistem Informasi (S1)
7	4652761662130272	AGUNG SAPUTRA	Teknik Informatika (S1)
8	8141761662130183	AGUS UMAR HAMDANI	Sistem Informasi (S1)
9	2636769670130302	AHMAD PUDOLI	Teknik Informatika (S1)
10	3955753654130082	AKHMAD UNGGUL PRIANTORO	Ilmu Komputer (S2)
11	1653757658130122	ANDY RIO HANDOKO	Teknik Informatika (S1)
12	1646766667130292	ANGGA KUSUMA NUGRAHA	Teknik Informatika (S1)
13	8947761662230262	ANITA DIANA	Sistem Informasi (S1)
14	0544751652130173	ANTON SATRIA PRABUWONO	Ilmu Komputer (S2)
15	4535772673130233	ANWAR RIFA'I	Teknik Informatika (S1)
16	5060770671130293	AQMAL MAULANA	Teknik Informatika (S1)
17	6647764665131142	ARI SAPUTRO	Manajemen Informatika (D3)
18	5239757658130173	ARIEF WIBOWO	Ilmu Komputer (S3)
19	0543756657130133	ARIF BRAMANTORO	Ilmu Komputer (S2)
20	4162753654131073	ARMAN YUSUF	Teknik Informatika (S1)
21	2533753654130132	ARSANTO NARENDRO	Teknik Informatika (S1)
22	5251757658130183	ASEP ABDUL ROHMAN	Sistem Informasi (S1)
23	7752762663237012	ATIK ARIESTA	Manajemen Informatika (D3)
24	3733759660130242	BASUKI HARI PRASETYO	Teknik Informatika (S1)
25	9846770671130352	BAYU SATRIA PRATAMA	Sistem Informasi (S1)
26	9551750651130082	BRURI TRYA SARTANA	Sistem Informasi (S1)
27	2555742643130063	BULLION DRAGON ANDAH	Sistem Informasi (S1)
28	3251756657130123	DARMAWAN BAGINDA NAPITTUPULU	Ilmu Komputer (S2)
29	5560751652130083	DENI MAHDIANA	Sistem Informasi (S1)



No	NUPTK	Nama	Program Studi
30	8556757658137103	DENNI KURNIAWAN	Ilmu Komputer (S2)
31	3535770671130233	DEVIT SETIONO	Sistem Informasi (S1)
32	1542762663230293	DEWI KUSUMANINGSIH	Sistem Informasi (S1)
33	4454761662130162	DIAN ANUBHAKTI	Sistem Informasi (S1)
34	7637741642130122	DJATI KUSDIARTO	Sistem Informasi (S1)
35	2243767668130313	DOLLY VIRGIAN SHAKA YUDHA SAKTI	Teknik Informatika (S1)
36	4556758659231082	DWI PEBRIANTI	Ilmu Komputer (S2)
37	9560763664230232	DWI PUSPITA ANGGRAENI	Teknik Informatika (S1)
38	2155762663131103	FERDIANSYAH	Komputerisasi Akuntansi (D3)
39	3453751652130073	FX BIMA CAHYA PUTRA	Sistem Informasi (S1)
40	2538753654130102	GANDUNG TRIYONO	Sistem Informasi (S1)
41	9043744645130083	GATOT PURWANTO	Sistem Komputer (S1)
42	4751753654230082	GRACE GATA	Komputerisasi Akuntansi (D3)
43	0537746647130122	GUNAWAN PRIA UTAMA	Teknik Informatika (S1)
44	0740763664130282	HADIDTYO WISNU WARDANI	Teknik Informatika (S1)
45	5846747648130102	HARI SOETANTO	Ilmu Komputer (S3)
46	9838763664130292	HARIS MUNANDAR	Teknik Informatika (S1)
47	8857759660131082	HENDRI IRAWAN	Sistem Informasi (S1)
48	0652765666130282	HILLMAN AKHYAR DAMANIK	Teknik Informatika (S1)
49	4735758659130162	HUMISAR HASUGIAN	Sistem Informasi (S1)
50	0434764665230262	IKA SUSANTI	Teknik Informatika (S1)
51	8949771672130282	IKHSAN RAHDIANA	Teknik Informatika (S1)
52	3941771672130302	IMAN PERMANA	Sistem Komputer (S1)
53	7437754655230112	IMELDA	Teknik Informatika (S1)
54	7746771672230342	INDAH PUSPASARI HANDAYANI	Sistem Informasi (S1)
55	2654764665130222	INDRA	Teknik Informatika (S1)
56	7454765666130203	INDRA HERTANTO	Teknik Informatika (S1)
57	9950765666130302	INDRA NUGRAHA ABDULLAH	Ilmu Komputer (S2)
58	0537752653130122	IRAWAN	Sistem Komputer (S1)
59	6435760661230183	ITA NOVITA	Sistem Informasi (S1)
60	7734743644130092	JAN EVERHARD RIWUROHI	Ilmu Komputer (S3)
61	1944770671130422	JEREMY JONATHAN	Sistem Informasi (S1)
62	9456761662130143	JOKO CHRISTIAN	Manajemen Informatika (D3)
63	2935754655130132	JOKO SUTRISNO	Sistem Informasi (S1)
64	2851769670130282	KUKUH HARSANTO	Sistem Informasi (S1)
65	9849754655130112	LAUW LI HIN	Sistem Informasi (S1)



No	NUPTK	Nama	Program Studi
66	5460755656230082	LESTARI MARGATAMA	Teknik Informatika (S1)
67	6849759660131132	LIS SURYADI	Komputerisasi Akuntansi (D3)
68	3457756657130123	LUHUR BAYUAJI	Ilmu Komputer (S2)
69	1654747648130072	MARDI HARDJIANTO	Ilmu Komputer (S2)
70	8639765666237002	MARINI	Sistem Informasi (S1)
71	5540767668230303	MEPA KURNIASIH	Teknik Informatika (S1)
72	4562753654230103	MERRY ANGGRAENI	Teknik Informatika (S1)
73	2453748649130073	MOHAMMAD ANIF	Teknik Informatika (S1)
74	9248752653130093	MOHAMMAD SYAFRULLAH	Ilmu Komputer (S2)
75	0643760661230242	MOTIKA DIAN ANGGRAENI	Sistem Informasi (S1)
76	2961757659200032	MUFTI	Teknik Informatika (S1)
77	0333764665130313	MUHAMAD SALMAN ALFARISI	Manajemen Informatika (D3)
78	1961760661130172	MUHAMMAD AINUR RONY	Teknik Informatika (S1)
79	7050757658237093	NAWINDAH	Sistem Informasi (S1)
80	6050754655230123	NIDYA KUSUMAWARDHANY	Sistem Informasi (S1)
81	3547763664230252	NOFIYANI	Sistem Informasi (S1)
82	5037758659230233	NONI JULIASARI	Sistem Informasi (S1)
83	4847756657231432	NURWATI	Sistem Informasi (S1)
84	1834757658230202	PAINEM	Sistem Informasi (S1)
85	2543764665230232	PEPI PERMATASARI	Sistem Informasi (S1)
86	4554760661230252	PIPIN FARIDA ARIYANI	Teknik Informatika (S1)
87	4151756657130113	PURWANTO	Teknik Informatika (S1)
88	8540769670230272	PUTRI HAYATI	Teknik Informatika (S1)
89	2362766667131233	RAHMAT OKTAVIAN	Teknik Informatika (S1)
90	5947771672230352	RATNA KUSUMAWARDANI	Sistem Informasi (S1)
91	3537759660230223	RATNA UJIAN DARI	Sistem Informasi (S1)
92	4656758659230152	RETNO WULANDARI	Sistem Informasi (S1)
93	0949761662230182	REVA RAGAM SANTIKA	Teknik Informatika (S1)
94	0443759660230253	RIRI IRAWATI	Sistem Komputer (S1)
95	1660744645230082	RIRIT ROESWIDIAH	Teknik Informatika (S1)
96	1745767668230302	RISKIANA WULAN	Teknik Informatika (S1)
97	2959764665237002	RIZKA TIAHARYADINI	Teknik Informatika (S1)
98	4456766667130233	RIZKY PRADANA	Sistem Informasi (S1)
99	4943758659130162	RIZKY TAHARA SHITA	Teknik Informatika (S1)
100	6235757658230143	RUSDAH	Ilmu Komputer (S2)
101	6249760661230213	SAFITRI JUANITA	Sistem Informasi (S1)



No	NUPTK	Nama	Program Studi
102	4554753654230092	SAFRINA AMINI	Teknik Informatika (S1)
103	3444749650130102	SAMIDI	Ilmu Komputer (S2)
104	4261760661230183	SAMSINAR	Sistem Informasi (S1)
105	9937760661130262	SEJATI WALUYO	Teknik Informatika (S1)
106	0157741642130083	SETYAWAN WIDYARTO	Ilmu Komputer (S2)
107	0241752653237043	SRI MULYATI	Sistem Informasi (S1)
108	3542749650230153	SRI WAHYUNINGSIH	Sistem Informasi (S1)
109	0246748649131143	SUBANDI	Teknik Informatika (S1)
110	7944752653130152	SUBANDI	Teknik Informatika (S1)
111	5937767668130372	SYAMSUDIN ZUBAIR	Teknik Informatika (S1)
112	4549736637130032	TATANG WIRAWAN WISNUADJI	Sistem Komputer (S1)
113	5539750651131093	TEJA ENDRA ENG TJU	Sistem Informasi (S1)
114	7552757658230133	TITIN FATIMAH	Sistem Informasi (S1)
115	7449765666230222	TRI IKA JAYA KUSUMAWATI	Sistem Informasi (S1)
116	6447751652130113	UTOMO BUDIYANTO	Teknik Informatika (S1)
117	4639763664130282	WAHYU PRAMUSINTO	Manajemen Informatika (D3)
118	9252739640130053	WENDI USINO	Ilmu Komputer (S3)
119	4749764665137022	WINDARTO	Teknik Informatika (S1)
120	7854758659230162	WINDHY WIDHYANTY	Teknik Informatika (S1)
121	9758748649230072	WIWIN WINDIHASTUTY	Sistem Informasi (S1)
122	2257766667230243	WULANDARI	Sistem Informasi (S1)
123	7863755656130092	YANI PRABOWO	Sistem Komputer (S1)
124	3948765666230332	YESI PUSPITA DEWI	Sistem Informasi (S1)
125	0448750651130092	YUDI SANTOSO	Sistem Informasi (S1)
126	6945763664130252	YUDI WIHARTO	Teknik Informatika (S1)
127	4057766667230303	YULIANAWATI	Sistem Informasi (S1)
128	7061753654230083	YULIAZMI	Sistem Informasi (S1)
129	6952768669130332	ZAQI KURNIAWAN	Teknik Informatika (S1)

Ditetapkan di : Jakarta

Tanggal : 02 September 2025

Dekan Fakultas Teknologi Informasi



Dr. Ir. Achmad Solichin, S.Kom., M.T.I

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.

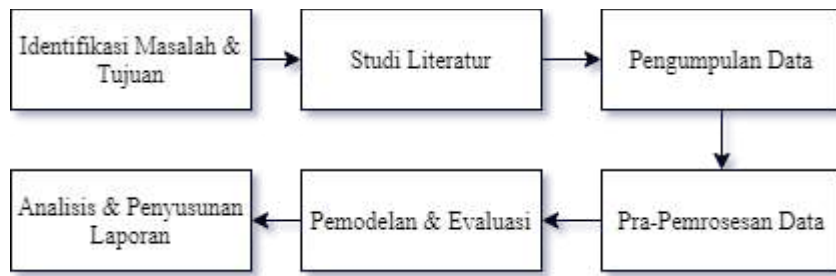
Pendekatan yang diusulkan dalam penelitian ini adalah penerapan teknik *clustering* berbasis data untuk mengelompokkan saham secara objektif. Metode yang digunakan adalah *Gaussian Mixture Model (GMM)* dan *Bayesian Gaussian Mixture (BGM)*, dua algoritma probabilistik yang efektif dalam mengidentifikasi 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

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

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:

$$s_{ij} = \frac{b_{ij} - a_{ij}}{\max(a_{ij}, b_{ij})} \quad (7)$$

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.:

$$DB = \frac{1}{K} \sum_{i=1}^K \min_{j \neq i} \frac{\sigma_i + \sigma_j}{d(C_i, C_j)} \quad (8)$$

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:

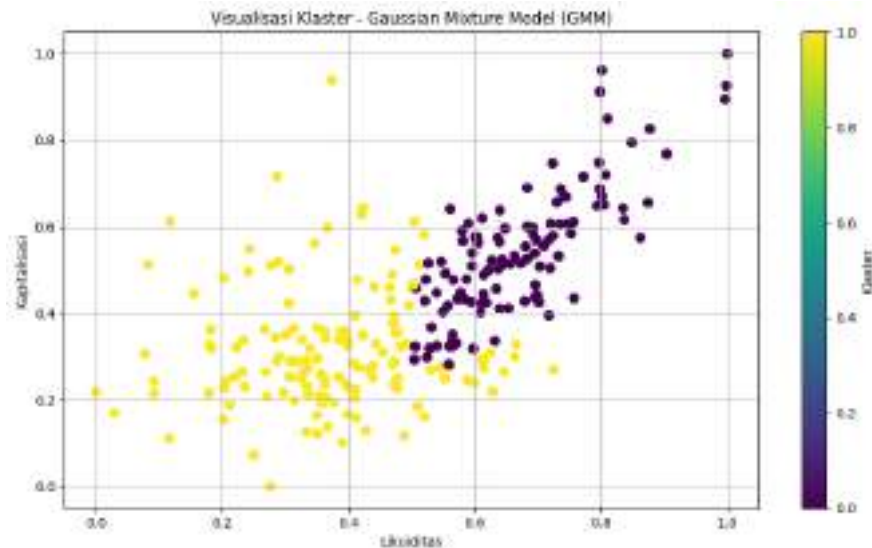
Tabel 6. Hasil Komparasi Model (*Silhouette Score*)

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.