

Analisis Pengelompokan Prestasi Mahasiswa Universitas Malang Menggunakan Metode *K-Means Clustering*

Moh. Aqil Mukhtar Alfarera ^{a,1*}, Zaehol Fatah ^{b,2}

^a Fakultas Sains dan Teknologi, Universitas Ibrahimy, Jalan KHR. Syamsul Arifin No.1-2, Situbondo 68374, Indonesia
¹ Aqilfarel86@gmail.com *; ² zaeholfatah@gmail.com
* Korespondensi Penulis

Submission:31/10/2024, Revision: 11/11/2024, Accepted : 30/11/2024

Abstract

The right choice to use data mining in analyzing the application of the K-Means Clustering technique in grouping Malang University (UM) student achievement data. With the increasing number of students and variations in achievement, managing achievement data in higher education has become more complex, so manual methods are not sufficient. K-Means Clustering was chosen because of its ability to group data based on certain attributes, which makes it easier to identify patterns and trends. This research aims to prove the effectiveness of K-Means in analyzing achievement data, as well as adding to the literature regarding the application of data mining in education. The dataset used includes student achievement indexes from various study programs at the University of Malang in the period 2018 to 2022. The data is processed to group student achievements efficiently. The clustering model was built using one of the algorithms in the clustering method, namely K-Means. This research produced the best cluster with a total of 3 clusters. The process to determine the best grouping was carried out by testing 6 different cluster models. Selections of the best cluster was carried out using the Davies Bouldin index test. Based on research with the results, these 3 groups can be categorized as, cluster 0 with a low value category with a value of 100, cluster 1 with a high category with a value value of 4.100, and cluster 2 with a middle category with a value 1.900.

Keywords: Data Analysis, K-Means Clustering, Student Achievement, University Of Malang.

Abstrak

Pilihan tepat menggunakan data mining dalam menganalisis penerapan teknik K-Means Clustering dalam pengelompokan data prestasi mahasiswa Universitas Malang (UM). Dengan meningkatnya jumlah mahasiswa dan variasi prestasi, pengelolaan data prestasi di perguruan tinggi menjadi lebih kompleks, sehingga metode manual tidak cukup memadai. K-Means Clustering dipilih karena kemampuannya untuk mengelompokan data berdasarkan atribut tertentu, yang memudahkan identifikasi pola dan tren. Penelitian ini bertujuan membuktikan efektivitas K-Means dalam menganalisis data prestasi, serta menambah literatur mengenai penerapan data mining di pendidikan. Dataset yang digunakan mencakup indeks prestasi mahasiswa dari berbagai program studi di Universitas Malang pada periode 2018 hingga 2022. Data diolah untuk mengelompokan prestasi mahasiswa secara efisien. Model klastering dibangun menggunakan salah satu algoritma dalam metode clustering yaitu K-Means. Penelitian ini menghasilkan klaster terbaik dengan jumlah 3 klaster, proses untuk menentukan pengelompokan terbaik dilakukan dengan menguji model 6 klaster yang berbeda. Pemilihan klaster terbaik dilakukan menggunakan pengujian indeks Davies Bouldin. Berdasarkan penelitian dengan hasil 3 kelompok tersebut dapat dikategorikan sebagai, cluster 0 dengan kategori rendah dengan nilai 100, cluster 1 dengan kategori tinggi dengan nilai 4.100, dan cluster 2 dengan kategori menengah dengan nilai 1.900.

Kata Kunci: Analisis data, K-Means Clustering, Prestasi Mahasiswa, Universitas Malang.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



1. Pendahuluan

Dalam era informasi saat ini, pengelolaan dan analisis data menjadi sangat penting terutama dalam konteks pendidikan tinggi. Universitas Malang (UM) sebagai salah satu institusi pendidikan terkemuka di Indonesia memiliki tanggung jawab untuk mengelola pendidikan dan proses pembelajaran di perguruan tinggi fokus pada mahasiswa, menggunakan metode pembelajaran yang efisien serta memaksimalkan penggunaan teknologi. Salah satu aspek penting dalam pencapaian ini adalah prestasi mahasiswanya, yang mencakup hasil akademik dan motivasi prestasi mahasiswa dipengaruhi oleh berbagai faktor internal dan eksternal[1]. Dalam upaya untuk mendukung mahasiswa mencapai prestasi yang lebih baik, universitas sering kali mengumpulkan data terkait capaian mahasiswa dalam berbagai aspek. Namun seiring dengan meningkatnya jumlah data, proses pengelolaan dan analisis data menjadi semakin kompleks. Di era digital saat ini, data mining menjadi solusi yang efektif untuk mengolah dan menganalisis data.

Data mining adalah suatu proses mengungkap pengetahuan atau informasi berharga dari kumpulan data yang besar dan kompleks. Tujuan utamanya ialah menemukan pola, hubungan, atau informasi tersembunyi dalam data, yang mungkin tidak tampak secara langsung. Proses ini menggunakan berbagai teknik statistik, matematika, dan kecerdasan buatan untuk menganalisis data secara otomatis dan sistematis[2]. Data mining umumnya memproses data dari database berukuran besar. Dari data tersebut, dilakukan pencarian pola atau tren yang sesuai dengan tujuan penerapan data mining. Hasil dari proses ini kemudian dapat digunakan untuk pengambilan keputusan maupun analisis yang diperlukan[3].

Prestasi mahasiswa merupakan hasil dari proses pembelajaran yang telah diselesaikan berdasarkan potensi atau kemampuan individu, serta sesuai dengan kompetensi yang ditetapkan oleh lembaga pendidikan terkait. Keberhasilan mahasiswa dalam menyelesaikan pembelajaran tercermin dari capaian indeks prestasi kumulatif, yang diklasifikasikan ke dalam lima kategori berdasarkan potensi nilai tertinggi pada setiap indikator yang telah ditentukan[4]. Di Universitas Malang (UM) data prestasi mahasiswa terus bertambah setiap tahunnya, data tersebut mencakup prestasi akademik dan non-akademik, serta pencapaian individu dan kelompok. Pengolahan data prestasi ini menjadi tantangan karena jumlahnya yang besar dan variatif, sehingga sulit dilakukan secara manual. Oleh karena itu, diperlukan metode yang tepat untuk mengelompokkan dan menganalisis data prestasi mahasiswa secara efisien.

Salah satu teknik yang dapat digunakan yaitu K-Means clustering, metode K-Means clustering adalah salah satu metode dalam data mining yang populer dikenal karena kecepatannya untuk mengelompokkan data berdasarkan kesamaan atribut. Metode ini mampu mengidentifikasi pola dan tren dalam data yang dapat digunakan untuk pengambilan keputusan dan bertujuan mempartisipasi objek ke dalam satu atau lebih kelompok berdasarkan karakteristik yang serupa akan dikelompokkan dalam satu cluster sementara objek dengan karakteristik berbeda akan berada di cluster lain. Selain itu, K-Means clustering memiliki kelemahan dalam menangani data dengan banyak dimensi, terutama ketika berhadapan dengan data yang tidak dapat dipisahkan secara non-linear[5]. Dalam konteks prestasi mahasiswa K-Means clustering, dapat membantu mengelompokkan data dan pengambilan keputusan yang lebih baik. Dengan menggunakan pendekatan ini, pihak universitas lebih mudah memahami profil prestasi mahasiswa dan strategi yang lebih tepat untuk pengembangan akademik.

Namun pengelolaan data prestasi mahasiswa Universitas Malang saat ini belum memanfaatkan teknologi analisis data secara optimal. Data yang ada masih tersebar dan kurang terstruktur, sehingga sulit untuk menemukan pola atau tren prestasi mahasiswa. Hal ini berdampak pada lambatnya proses evaluasi dan pengambilan keputusan terkait kebijakan pendidikan. Untuk mengatasi hal ini, penerapan K-Means clustering sebagai bagian pendekatan data mining dapat membantu menyusun rekapitulasi data prestasi mahasiswa secara terstruktur. Algoritma K-Means adalah metode klustering yang bertujuan untuk membagi data ke dalam satu atau lebih kelompok cluster. Metode ini mengelompokkan data ke dalam cluster yang sama, sementara data dengan karakteristik yang berbeda dimasukkan ke cluster lain[6].

Melalui analisis klaster, diharapkan dapat menghasilkan rekapitulasi yang lebih akurat dan informatif, mengenai prestasi mahasiswa di Universitas Malang menggunakan algoritma K-Means clustering. Analisis klaster merupakan metode fundamental dalam analisis data yang diterapkan di berbagai bidang untuk mengidentifikasi pola dan hubungan tersembunyi dalam dataset. Salah satu metode yang umum digunakan untuk analisis klaster adalah algoritma K-Means, membagi dataset menjadi kelompok yang berbeda berdasarkan kemiripan di antara data[7]. Dengan cara ini, penelitian akan memberikan wawasan tentang distribusi prestasi mahasiswa, juga akan memberikan kontribusi dalam pengambilan keputusan strategis, seperti penentuan program, peningkatan kualitas akademik, dan pengembangan kebijakan yang lebih efektif.

Penelitian ini bertujuan untuk menyediakan informasi yang relevan bagi pengambil keputusan di tingkat universitas dalam hal kebijakan akademik dan pengembangan program studi berdasarkan hasil analisis data prestasi mahasiswa. Penelitian ini dibatasi hanya pada data prestasi mahasiswa dan diolah menggunakan model

clustering, algoritma yang diterapkan dalam proses pengembangan model adalah algoritma K-Means dan evaluasi model dilakukan dengan pengujian menggunakan metode davies bouldin[8].

2. Metode Penelitian

a. Sumber Data

Penelitian ini menggunakan data sekunder yang diambil dari platform online yang menyediakan informasi tentang prestasi mahasiswa Universitas Malang. Sumber data diperoleh melalui portal resmi universitas dan sumber-sumber publik yang menyajikan data prestasi akademik dan non-akademik mahasiswa dalam bentuk terstruktur. Data yang dikumpulkan mencakup prestasi mahasiswa Universitas Malang dalam berbagai bidang prestasi di masing-masing program studi, yang telah dipublikasikan oleh universitas terkait. Data tersebut tersedia secara online melalui aplikasi *kaggle*[9].

b. Pengumpulan Data

Pengumpulan data ini dilakukan untuk memperoleh dan mengetahui informasi mengenai prestasi akademik dan non-akademik mahasiswa Universitas Malang (UM). Data prestasi mahasiswa diakses melalui arsip prestasi universitas yang memuat catatan resmi prestasi akademik dan non-akademik yang telah dicapai oleh mahasiswa. Dokumen yang digunakan meliputi laporan tahunan prestasi mahasiswa di setiap masing-masing program studi.

c. Pengolahan Data

Dalam penelitian ini, bertujuan untuk mengelompokkan data prestasi mahasiswa Universitas Malang berdasarkan program studi mulai periode 2018 hingga 2022 menggunakan metode K-Means clustering. Hasil dari pengelompokan tersebut kemudian menentukan kluster terbaik dengan membandingkan nilai yang terkecil dari model cluster yang dibuat, dengan melakukan proses pengujian menggunakan metode Davies Bouldin.

d. Data Mining

Data mining merupakan sebuah proses yang secara otomatis mencari informasi berguna dalam penyimpanan data berukuran besar, teknik data mining diterapkan untuk menganalisis basis data besar sebagai metode untuk menemukan pola baru yang bermanfaat. Namun tidak semua aktivitas pencarian informasi dapat dikategorikan sebagai data mining[10].

e. Clustering

Clustering adalah proses membagi sekumpulan objek data ke dalam kelompok yang disebut cluster. Objek dalam satu cluster memiliki karakteristik yang mirip satu sama lain, tetapi berbeda dari objek dalam cluster lain. Pembagian ini dilakukan secara otomatis menggunakan algoritma clustering, karena itu clustering berguna untuk mengidentifikasi kelompok atau pola yang tidak terlihat sebelumnya dalam data. Clustering juga dikenal sebagai segmentasi data karena memisahkan data ke dalam beberapa kelompok berdasarkan kesamaan tertentu[11].

f. K-Means

K-Means adalah salah satu algoritma pembelajaran tanpa pengawasan yang berfungsi untuk mengelompokkan data ke dalam cluster, algoritma ini dapat bekerja dengan data tanpa label kategori. K-Means clustering ialah dimana data dikelompokkan berdasarkan karakteristik yang mirip dalam satu cluster, namun berbeda dengan data dalam cluster lain[12].

g. Davies Bouldin

Davies bouldin merupakan metode untuk mengevaluasi nilai hasil dari algoritma clustering atau pengelompokan data, berdasarkan kuantitas dan karakteristik yang ada dalam kumpulan data. Namun pendekatan ini, memiliki kelemahan karena nilai baik yang dihasilkan tidak selalu mencerminkan kualitas informasi yang diperoleh[13].

h. Rapid Miner

Rapid miner adalah perangkat lunak open source yang menawarkan solusi untuk analisis data mining, text mining, dan analisis prediktif. Dengan memanfaatkan berbagai teknik deskriptif dan prediktif, rapidminer memberikan wawasan yang membantu pengguna dalam membuat keputusan yang lebih baik[14].

i. Visualisasi Data

Visualisasi data merupakan tampilan berupa grafis dari informasi dan data, dengan kata lain visualisasi data mengubah kumpulan data mengubah format yang lebih sederhana untuk ditampilkan. Melalui elemen visual ini, pembaca dapat lebih mudah mengidentifikasi tren dan pola dalam data[15].

3. Hasil dan Pembahasan

3.1. Datasheet

Datasheet ini disusun berdasarkan data yang diperoleh dari portal perguruan tinggi dan mencakup prestasi mahasiswa Universitas Malang. Data prestasi mahasiswa yang diolah mencakup periode tahun 2018 hingga tahun 2022, dari setiap program studi di Universitas Malang. Dataset ini terdiri 71 entri dan 5 atribut. Hasil pengolahan dataset ditampilkan pada gambar 1.

1	Fakultas	Prodi	2018	2019	2020	2021	2022
2	FEB	S1 Pendidikan Ekono...	0	0	396	36	120
3	FEB	S1 Akuntansi	1	1	1562	473	4152
4	FEB	S1 Pendidikan Akunt...	0	0	722	285	1903
5	FEB	S1 Manajemen	0	0	707	61	339
6	FEB	S1 Pendidikan Admini...	0	0	586	86	549
7	FEB	S1 Pendidikan Tata N...	0	0	266	18	78
8	FEB	S1 Ekonomi Peman...	0	0	421	24	251
9	FEB	D3 Akuntansi	0	0	69	18	0
10	FEB	D3 Manajemen Pem...	0	0	35	7	7
11	FIK	S1 Pendidikan Kepel...	1	0	291	159	37
12	FIK	S1 Ilmu Keolahragaa...	0	0	64	73	29
13	FIK	S1 Ilmu Kesehatan M...	1	2	441	748	233
14	FIK	S1 Pendidikan Jasm...	0	2	220	178	88
15	FIP	S1 Pendidikan Guru ...	0	0	184	127	106
16	FIP	S1 Pendidikan Luar S...	1	0	41	94	84
17	FIP	S1 Administrasi Pend...	0	0	107	49	74
18	FIP	S1 Pendidikan Luar P...	0	0	90	22	54

Gambar 1. Datasheet Prestasi Mahasiswa Universitas Malang

3.2. Data Cleaning and Integration

Proses pembersihan data dilakukan dengan memeriksa dataset yang digunakan. Langkah ini mencakup pengecekan data kosong, dan data yang berada di luar batas normal (outlier).

a. Pengecekan data kosong

Data kosong diperiksa dengan melihat kolom missing value pada datasheet. Gambar 2 menunjukkan hasil pengecekan missing value. Berdasarkan hasil pengecekan, dataset tidak mengandung data kosong.

Name	Type	Missing
Id Fakultas	Polynomial	0
Label Prodi	Polynomial	0
Cluster cluster	Nominal	0
2018	Integer	0
2019	Integer	0
2020	Integer	0
2021	Integer	0
2022	Integer	0

Gambar 2. Pengecekan Data Kosong

b. Pengecekan data yang menyimpang jauh (outlier)

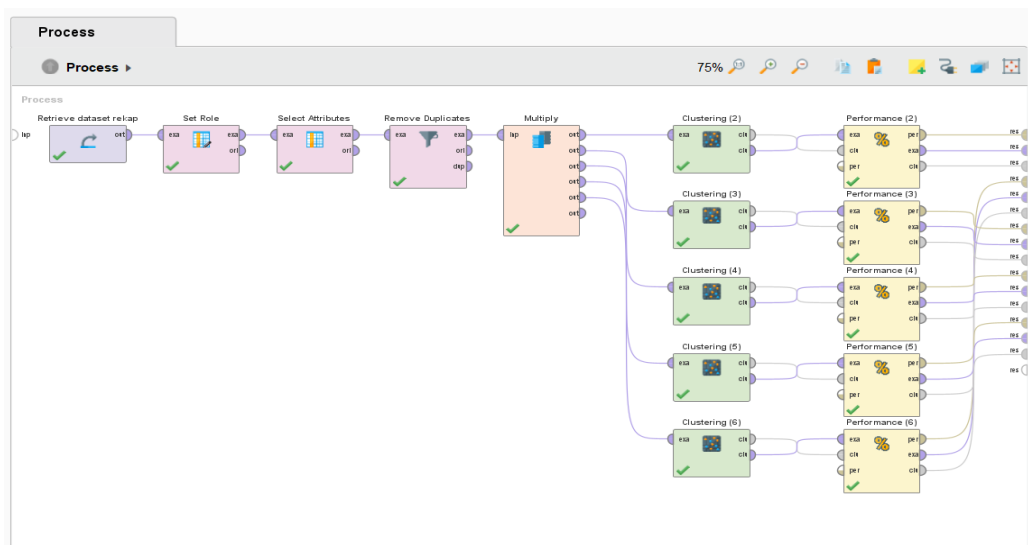
Pengecekan data pencilan dilakukan dengan memeriksa nilai minimum dan maksimum dari setiap atribut. Gambar 3 menunjukkan informasi nilai minimum dan maksimum pada dataset. Dari gambar tersebut, nilai minimum dan maksimum masih sesuai dengan data yang dibutuhkan, misalnya data prestasi mahasiswa pada periode 2018 masih berada dalam rentang 0-3.

Name	Type	Missing	Statistics	Filter
id	Polynomial	0	Least FPPSi (1)	Most FT (19)
Fakultas	Polynomial	0	Least S1 Tekno [...] dikan (1)	Most D3 Akuntansi (1)
Label	Polynomial	0	Least S1 Tekno [...] dikan (1)	Most D3 Akuntansi (1)
Prodi	Polynomial	0	Least S1 Tekno [...] dikan (1)	Most D3 Akuntansi (1)
Cluster	Nominal	0	Least cluster_0 (2)	Most cluster_1 (68)
cluster	Nominal	0	Least cluster_0 (2)	Most cluster_1 (68)
2018	Integer	0	Min 0	Max 3
2019	Integer	0	Min 0	Max 9
2020	Integer	0	Min 0	Max 1562
2021	Integer	0	Min 0	Max 827
2022	Integer	0	Min 0	Max 4152

Gambar 3. Pengecekan data yang menyimpang jauh

3.3. Data Mining

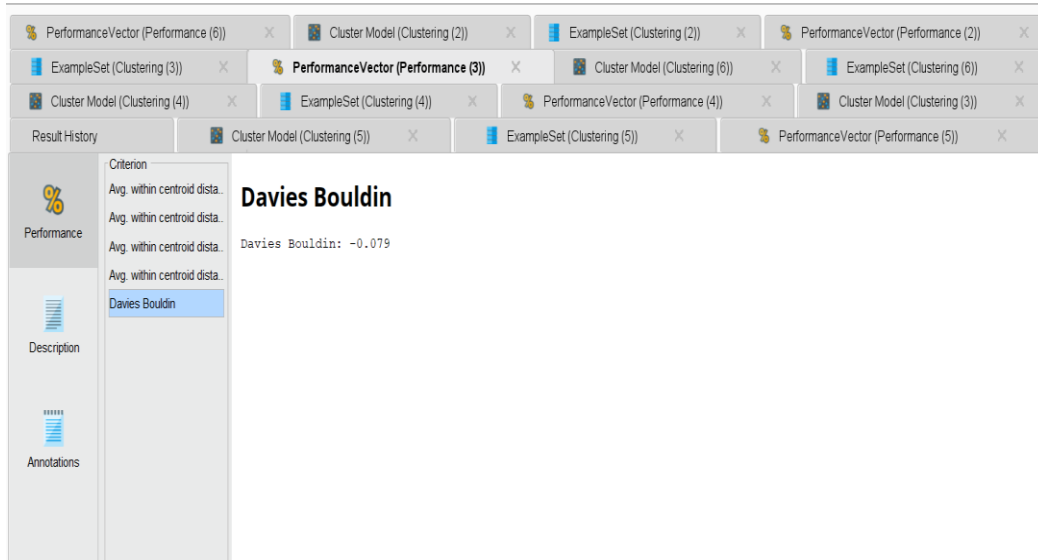
Pembuatan model clustering dilakukan dengan algoritma K-Means. Model ini dibangun di rapidminer menggunakan operator K-Means, dan untuk menguji kinerja model, digunakan evaluasi performa dengan cluster distance performance. Uji coba dilakukan untuk mendapatkan cluster optimal, mulai dari cluster 2 hingga sampai cluster 6. Setiap hasil clustering kemudian dievaluasi performanya. Gambar 4 menunjukkan proses pembuatan model clustering dengan algoritma K-Means.



Gambar 4. Pembuatan Model Algoritma K-Means Clustering

3.4. Evaluasi Model

Model dievaluasi untuk menemukan pengelompokan data yang optimal. Pemilihan cluster terbaik dilakukan dengan memperhatikan nilai evaluasi menggunakan Davies Bouldin. Semakin kecil nilai Davies Bouldin, semakin baik kualitas clusternya. Gambar 5 menunjukkan hasil model K-Means dengan nilai dari K=2 hingga K=6 diukur dan dievaluasi performanya.



Gambar 5. Hasil dari proses model yang dibuat

Gambar tersebut menunjukkan hasil evaluasi performa model clustering menggunakan Davies Bouldin di rapidminer. Pada bagian kiri, terlihat beberapa kriteria performa yang dapat dipilih, termasuk Davies Bouldin yang saat ini dipilih, Davies Bouldin yang ditampilkan adalah -0,079.

Nilai Davies Bouldin terkecil diperoleh pada cluster dengan jumlah 0,079, sehingga model clustering yang dipilih adalah dengan K=3. Tabel 1 menunjukkan hasil nilai Davies Bouldin untuk setiap nilai dari K=2 hingga K=6.

Tabel 1 : Perbandingan Nilai Davies Bouldin

Banyak Cluster	Nilai Davies Bouldin
2	0.448
3	0.079
4	0.511
5	0.419
6	0.498

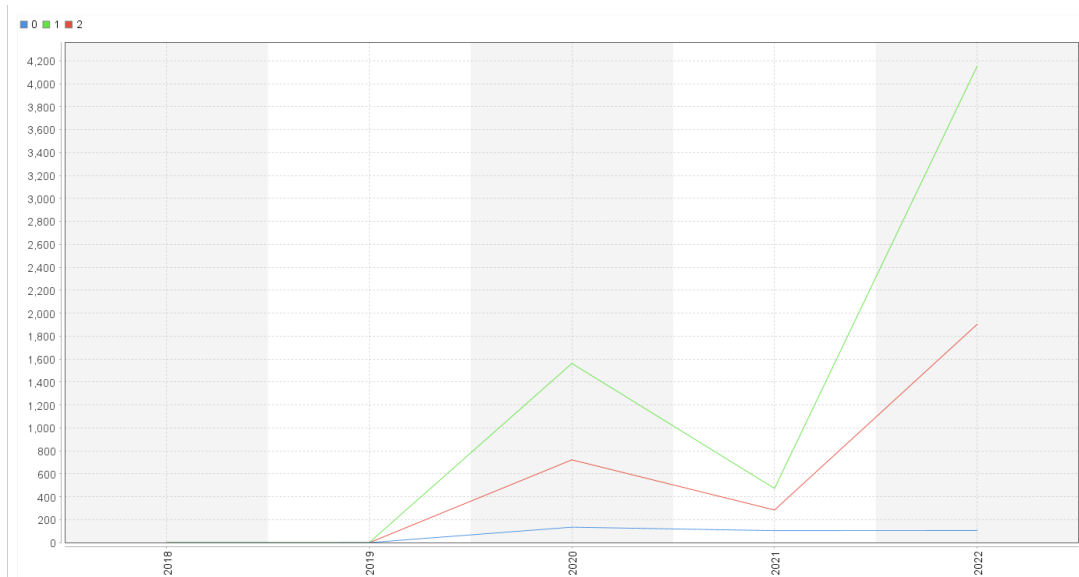
Davies Bouldin indeks mengukur seberapa baik cluster yang dihasilkan, dimana nilai yang lebih rendah menunjukkan cluster yang lebih baik. Nilai negatif ini tidak umum, karena Davies Bouldin biasanya bernilai positif disebabkan oleh kondisi atau normalisasi khusus dalam dataset. Hasil ini menandakan bahwa model K-Means telah dievaluasi dan pengguna perlu memeriksa nilai Davies Bouldin ini untuk menentukan kualitas clustering yang dihasilkan.

Pengelompokan dataset dengan K=3 menghasilkan tiga cluster, yaitu cluster 0 yang berisi 68 data, cluster 1 yang berisi 1 data, dan cluster 2 yang berisi 1 data. Tabel 2 menampilkan jumlah data dalam setiap cluster untuk setiap nilai K.

Tabel 2 : Banyak Cluster untuk masing-masing nilai K

Banyak Cluster	Pengelompokan Data
2	Kelompok 0 : 2 data Kelompok 1 : 68 data
3	Kelompok 0 : 68 data Kelompok 1 : 1 data
4	Kelompok 0 : 63 data Kelompok 1 : 1 data Kelompok 2 : 1 data Kelompok 3 : 5 data
5	Kelompok 0 : 59 data Kelompok 1 : 1 data Kelompok 2 : 6 data Kelompok 3 : 1 data Kelompok 4 : 3 data
6	Kelompok 0 : 36 data Kelompok 1 : 1 data Kelompok 2 : 3 data Kelompok 3 : 1 data Kelompok 4 : 26 data Kelompok 5 : 3 data

Pengelompokan dataset menjadi 3 kategori memungkinkan untuk penamaan masing-masing cluster. Pemberian nama kategori dapat dilihat pada plot yang ditampilkan dalam gambar 6. Gambar tersebut menunjukkan grafik hasil clustering data menggunakan model K-Means. Pada grafik ini terdapat tiga cluster yang ditampilkan dengan warna berbeda : biru (0), hijau (1), dan merah (2). Sumbu horizontal mewakili tahun, sedangkan sumbu vertikal menunjukkan nilai yang terkait dengan data yang dikelompokkan.



Gambar 6. Plot hasil dari cluster 0, cluster 1, dan cluster 2

Berdasarkan grafik diatas, distribusi prestasi mahasiswa Universitas Malang dari berbagai program studi, yang terbagi ke dalam tiga cluster yaitu cluster 0, cluster 1, dan cluster 2, memperlihatkan perubahan nilai untuk masing-masing cluster dari periode tahun 2018 sampai tahun 2022. Cluster berwarna biru (0) memiliki jumlah prestasi yang paling rendah sepanjang tahun. Cluster berwarna hijau (1) memiliki jumlah

prestasi tertinggi dan prestasi terbaik di dibandingkan cluster lainnya. Sementara itu, Cluster berwarna merah (2) memiliki jumlah prestasi dengan peningkatan yang konsisten, program studi ini memiliki performa yang cukup baik namun masih berada di bawah cluster 1.

Hasil pengelompokan data berdasarkan beberapa atribut, seperti fakultas dan prodi, dengan hasil clustering yang dibagi menjadi beberapa cluster yaitu cluster 0, cluster 1, dan cluster 2. Tabel ini mencakup data dari tahun 2018 hingga tahun 2022 untuk setiap baris atau contoh data serta memberikan gambaran bagaimana data dari berbagai program studi dan fakultas dikelompokkan ke dalam cluster tertentu. Cluster 0, hasil program studi dengan jumlah prestasi yang paling rendah, dibandingkan dengan cluster yang lain. Cluster 1, hasil program studi yang memiliki jumlah prestasi tertinggi, dan paling terbaik di dibandingkan cluster lainnya. Sedangkan Cluster 2, hasil program studi dengan peningkatan yang konsisten, setelah mengalami penurunan pada tahun 2021.

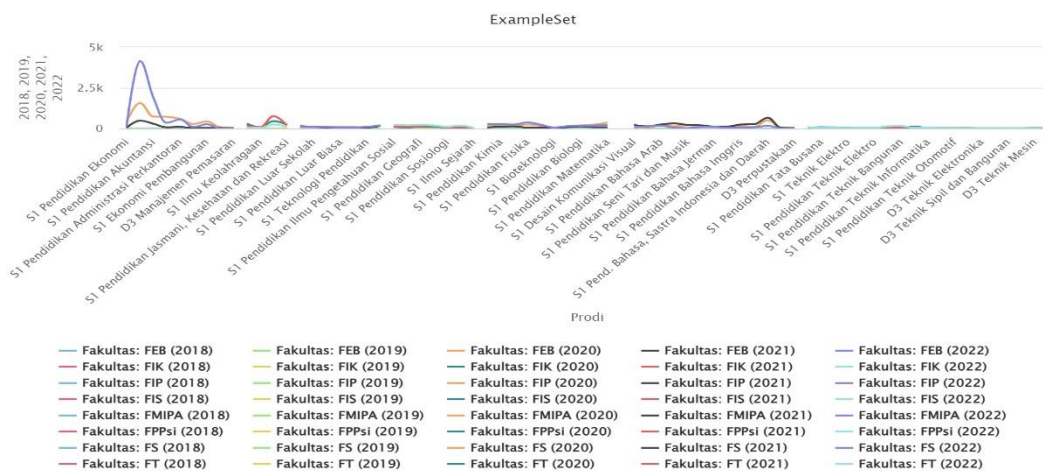
Row No.	Fakultas	Prodi	cluster	2018	2019	2020	2021	2022
1	FEB	S1 Pendidikan ...	cluster_0	0	0	396	36	120
2	FEB	S1 Akuntansi	cluster_1	1	1	1562	473	4152
3	FEB	S1 Pendidikan ...	cluster_2	0	0	722	285	1903
4	FEB	S1 Manajemen	cluster_0	0	0	707	61	339
5	FEB	S1 Pendidikan ...	cluster_0	0	0	586	86	549
6	FEB	S1 Pendidikan ...	cluster_0	0	0	266	18	78
7	FEB	S1 Ekonomi P...	cluster_0	0	0	421	24	251
8	FEB	D3 Akuntansi	cluster_0	0	0	69	18	0
9	FEB	D3 Manajeme...	cluster_0	0	0	35	7	7
10	FIK	S1 Pendidikan ...	cluster_0	1	0	291	159	37
11	FIK	S1 Ilmu Keolah...	cluster_0	0	0	64	73	29
12	FIK	S1 Ilmu Keseh...	cluster_0	1	2	441	748	233
13	FIK	S1 Pendidikan ...	cluster_0	0	2	220	178	88
14	FIP	S1 Pendidikan ...	cluster_0	0	0	184	127	106
15	FIP	S1 Pendidikan ...	cluster_0	1	0	41	94	84
16	FIP	S1 Administra...	cluster_0	0	0	107	49	74

ExampleSet (70 examples,3 special attributes,5 regular attributes)

Gambar 7. Hasil cluster untuk masing-masing data

3.5. Visualisasi Hasil

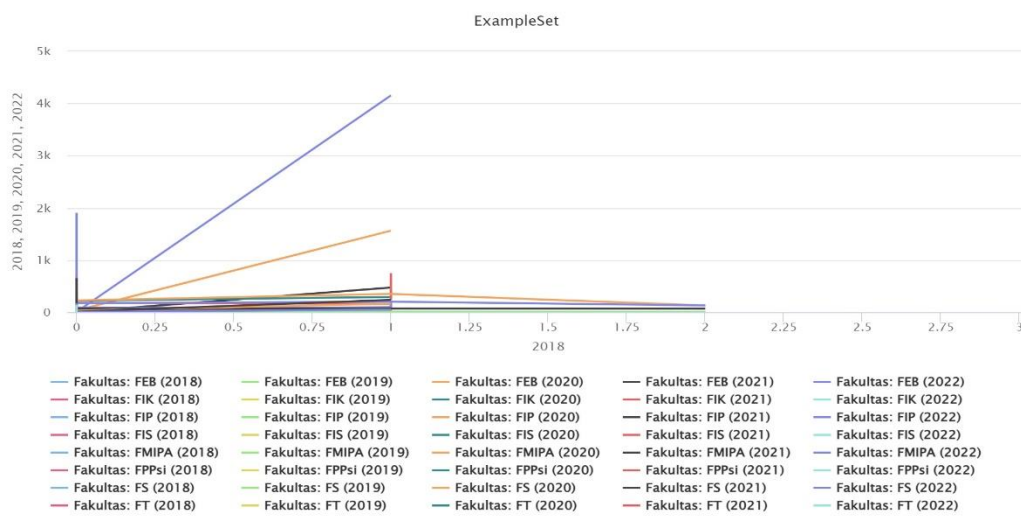
Visualisasi data pada K-Means clustering membantu dalam memahami cara algoritma mengelompokkan data berdasarkan kesamaan. Visualisasi ini memudahkan untuk melihat bagaimana data terdistribusi di dalam cluster dan seberapa baik pemisahannya. Dengan visualisasi yang tepat kita dapat melihat apakah K-Means clustering menghasilkan cluster yang terpisah dengan baik.



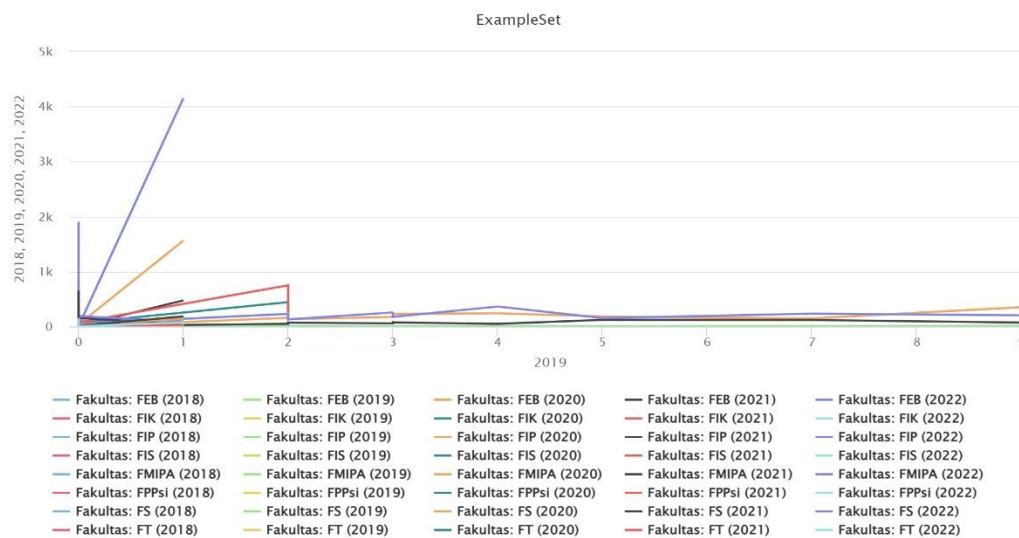
Gambar 8 : Visualisasi dari masing-masing program studi

Gambar diatas menunjukkan grafik tren jumlah data dengan berbagai program studi dari beberapa periode mulai tahun 2018 hingga tahun 2022. Grafik tersebut menyajikan jumlah atau identitas data per tahun untuk setiap program studi dan menyoroti variasi data antar fakultas yang diwakili oleh warna garis yang berbeda. Gambar ini memungkinkan kita untuk membandingkan perubahan jumlah atau nilai data antar program studi pada tahun-tahun yang berbeda serta memeriksa perbedaan antar fakultas. Puncak dan penurunan dalam grafik ini mencerminkan perubahan yang terjadi dalam rentang waktu tersebut, dengan beberapa program studi yang memiliki nilai tinggi pada tahun tertentu, sementara lainnya mungkin stabil atau mengalami sedikit perubahan.

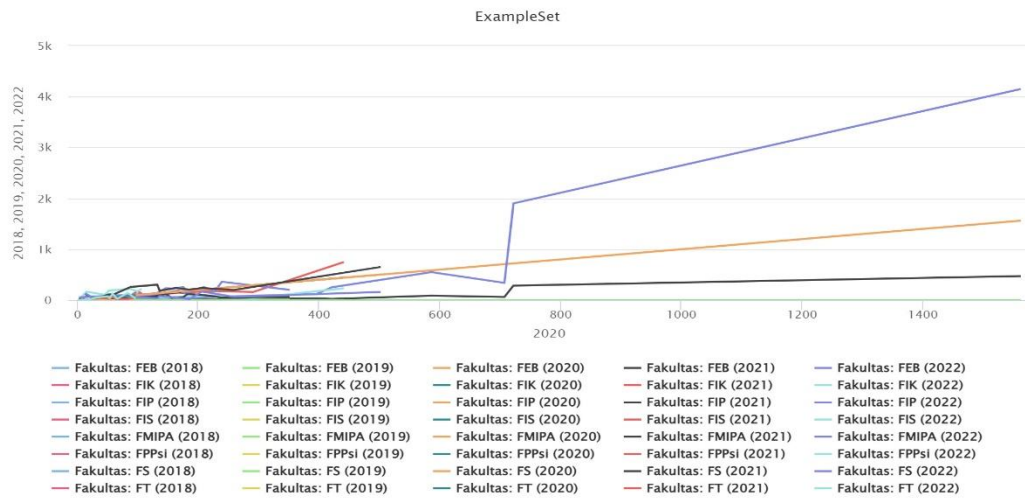
Berikut adalah gambar sebuah grafik yang menunjukkan data dari berbagai fakultas untuk periode tahun 2018 hingga tahun 2022. Setiap garis mewakili data dari sebuah fakultas yang berbeda dan terdapat beberapa fakultas yang berbeda dan terdapat beberapa fakultas yang dicatat untuk setiap tahun seperti FEB, FIK, FIP, FMIPA, FPPsi, FS, dan FT. Secara keseluruhan grafik ini tampaknya menunjukkan tren data, berdasarkan fakultas dan tahun. Untuk interpretasi yang lebih mendalam, akan bermanfaat untuk mengetahui lebih lanjut tentang apa yang diukur oleh grafik ini (misalnya, prestasi mahasiswa dari setiap program studi dan kelulusan) agar dapat memberikan konteks yang lebih spesifik.



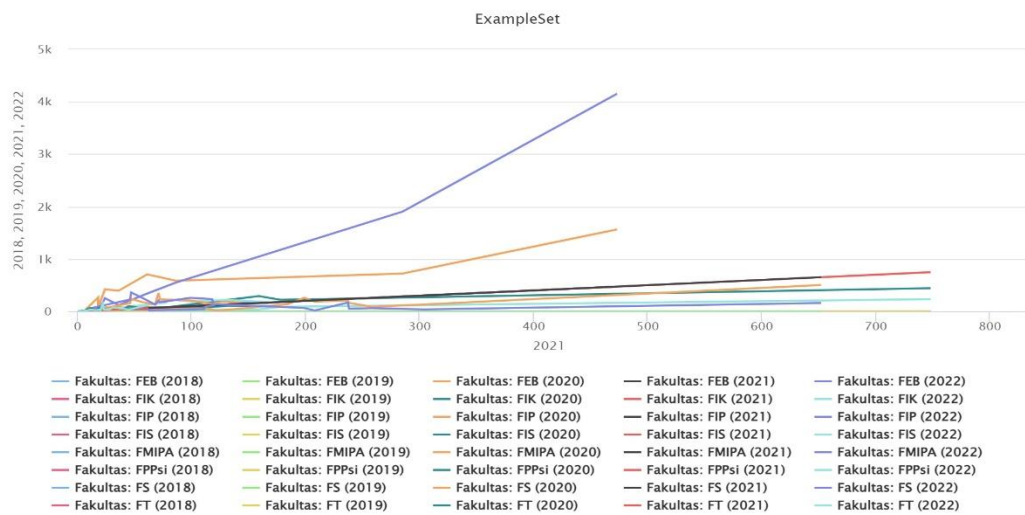
Gambar 9. Visualisasi prodi periode tahun 2018



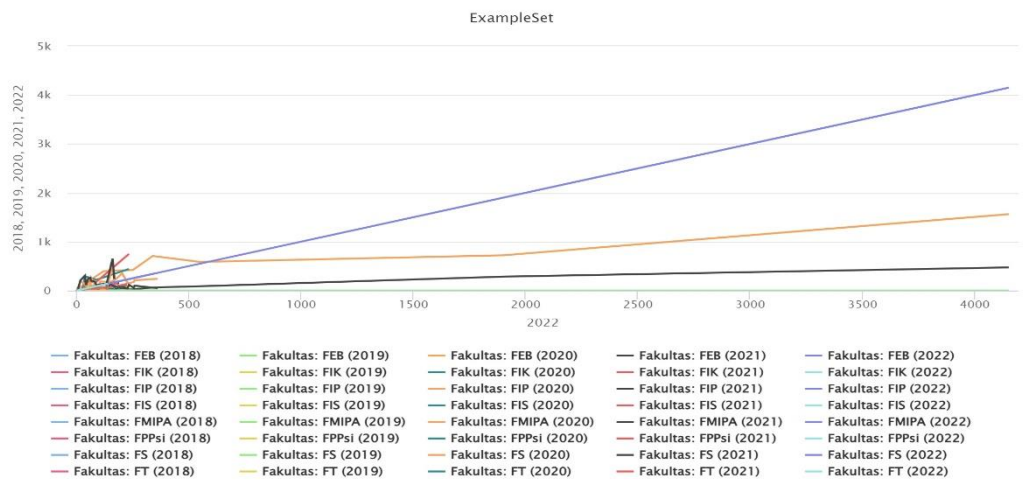
Gambar 10. Visualisasi prodi periode tahun 2019



Gambar 11. Visualisasi prodi periode tahun 2020



Gambar 12. Visualisasi prodi periode tahun 2021



Gambar 13. Visualisasi prodi periode tahun 2022

4. Kesimpulan

Model clustering dapat digunakan untuk mengelompokkan data dalam sebuah dataset. Pengelompokan ini membantu dalam pengambilan keputusan, terutama untuk menentukan kelompok yang memerlukan kebijakan atau strategi tertentu dengan meningkatkan performa, sehingga data awal berada dalam kelompok kurang baik dapat ditingkatkan sehingga masuk ke kelompok yang lebih baik. Dengan menggunakan dataset prestasi mahasiswa dari periode tahun 2018 sampai tahun 2022, menghasilkan 3 kelompok. Cluster 0, memiliki jumlah prestasi yang paling rendah sepanjang tahun, dengan sedikit peningkatan tetapi relatif stabil dengan nilai 100. Cluster 1, memiliki jumlah prestasi tertinggi terutama pada tahun 2022 yang menunjukkan lonjakan signifikan, hal ini menunjukkan bahwa program studi memiliki performa prestasi yang lebih baik dibandingkan cluster lainnya yaitu dengan nilai 4.100. Sedangkan cluster 2, menunjukkan peningkatan yang konsisten, dengan lonjakan pada tahun 2022 setelah mengalami penurunan pada tahun 2021, program studi dalam cluster ini memiliki performa yang cukup baik namun masih berada di bawah cluster 1 dengan nilai 1.900.

5. Ucapan Terima Kasih

Penulis ingin mengucapkan terima kasih yang mendalam kepada dosen pembimbing yang terhormat, atas segala bimbingan, dukungan serta arahan yang telah diberikan. Berkat perhatian dan kesabaran beliau, saya dapat berkembang secara pribadi dan akademis. Setiap masukan dan saran yang beliau berikan sangat berarti bagi kemajuan saya serta menjadikan perjalanan ini sebagai pengalaman yang sangat berharga.

6. Daftar Pustaka

- [1] D. R. Retnowati, A. Fatchan, and K. Astina, "Prestasi Akademik dan Motivasi Berprestasi Mahasiswa Universitas Negeri Malang," *J. Pendidik.*, vol. 1, pp. 521–525, 2016, [Online]. Available: <https://media.neliti.com/media/publications/212159-prestasi-akademik-dan-motivasi-berpresta.pdf>
- [2] A. Wasik *et al.*, "Implementasi data mining untuk memprediksi penjualan accessoris handphone dan handphone terlaris menggunakan metode k-nearest neighbor (k-nn) 1," vol. 1, no. 2, pp. 469–479, 2024.
- [3] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm," *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
- [4] Z. Zaifullah and T. Yulianto, "Analisis Cluster Untuk Pengelompokan Prestasi Mahasiswa Angkatan 2013 Fakultas MIPA Universitas Islam Madura," *Zeta - Math J.*, vol. 7, no. 1, pp. 1–10, 2022, doi: 10.31102/zeta.2022.7.1.1-10.
- [5] I. Vhalla, S. Sumijan, and J. Santony, "Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering K-Means," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 2, no. 2, pp. 572–577, 2018, doi: 10.29207/resti.v2i2.308.
- [6] Narwati, "Pengelompokan Mahasiswa Menggunakan Algoritma K-Means," *J. Din. Inform.*, vol. 2, no. 2, pp. 1–7, 2010.
- [7] I. T. Umagapi and B. Umaternate, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa," *Sist. Inf. dan Teknol.*, vol. 7, pp. 303–308, 2023.
- [8] S. Suraya, M. Sholeh, and D. Andayati, "Penerapan Metode Clustering Dengan Algoritma K-Means Pada Pengelompokan Indeks Prestasi Akademik Mahasiswa," *Skanika*, vol. 6, no. 1, pp. 51–60, 2023, doi: 10.36080/skanika.v6i1.2982.
- [9] R. P. Nugroho, "Rekapitulasi Statistik Prestasi Mahasiswa Universitas Negeri Malang," no. 2022, [Online]. Available: <https://www.kaggle.com/datasets/vzrengamani/statistik-prestasi-mahasiswa-um>
- [10] K. A. Gede Aditra Pradnyana, "Konsep Dasar Data Mining," vol. 1, pp. 1–16, 2018, [Online]. Available: <https://pustaka.ut.ac.id>
- [11] E. Irwansyah, "Pengertian Clustering, Manfaat dan Konsep Dasar," 2017, [Online]. Available: <https://socs.binus.ac.id/2017/03/09/clustering/>

-
- [12] E. P. P. Bryan Oerleans, "Clustering Algoritma K-Means," 2022, [Online]. Available: <https://sis.binus.ac.id/2022/01/31/clustering-algoritma-k-means/>
- [13] Wikipedia, "Pengertian Indeks Davies Bouldin," 2023, [Online]. Available: https://en.wikipedia.org/wiki/Davies–Bouldin_index#Definition
- [14] A. Umar, "RapidMiner, Definisi dan Fitur-Fiturnya," 2021, [Online]. Available: <https://www.abdumar.com/2021/03/rapidminer-definisi-dan-fitur-fiturnya.html>
- [15] A. Perdana, "Visualisasi Data: Pengertian, Fungsi, dan Tipe-tipenya," 2022, [Online]. Available: <https://glints.com/id/lowongan/data-visualization-adalah/>