

Klasterisasi Calon Mahasiswa Baru Universitas Almuslim Menggunakan K-Means Clustering

Noratul Iqramah¹, Imam Muslem², Munar³

^{1,2,3}Universitas Almuslim Bireuen

ARTICLE INFO

Article history:

Received : 08 Oktober 2025

Revised : 13 November 2025

Accepted : 15 November 2025

Keywords:

Clustering, K-Means,
Universitas Almuslim, Google
Colab, Academic Interests



This work is licensed under a
[Creative Commons
Attribution-ShareAlike 4.0
International License](https://creativecommons.org/licenses/by-sa/4.0/)

ABSTRACT

[Clustering of Prospective Students at Universitas Almuslim Using K-Means Clustering] The diverse academic backgrounds of prospective new students at Almuslim University often present challenges in determining the appropriate field of study. Determining the field of study is crucial because choosing the wrong study program can impact the learning process and the development of students' potential during their studies. Selecting the right field of study allows students to learn optimally and prepare themselves for the world of work according to their interests and abilities. It can also assist the university in recommending study programs with appropriate fields of study to develop a more targeted admission strategy. This study applies the clustering method with the K-Means algorithm to help group prospective students into two fields of study: science and social science. This field grouping is based on 1000 prospective new students' data with attributes of diploma grades (Mathematics, Science, Indonesian, and Social Studies), test scores, and field interests. The analysis process carried out using the K-Means clustering method on Google Colab resulted in a calculation of 17 iterations, C1 (science) with a total of 518 people who have higher interests and values in the field of science, and C2 (social) with a total of 482 people who have higher interests and values in the field of social. This division confirms that the K-Means algorithm is able to group data based on the characteristics in the dataset. With these results, K-Means Clustering is proven effective in grouping prospective students of Almuslim University based on their academic background and interests.

ABSTRAK

Keberagaman latar belakang akademik calon mahasiswa baru di Universitas Almuslim sering kali menjadi tantangan dalam menentukan bidang ilmu yang sesuai. Penentuan bidang ilmu sangat penting karena pemilihan program studi yang tidak tepat dapat memengaruhi proses pembelajaran dan pengembangan potensi mahasiswa selama masa studi. Pemilihan bidang ilmu yang sesuai memungkinkan mahasiswa untuk belajar secara optimal dan mempersiapkan diri menghadapi dunia kerja sesuai dengan minat dan kemampuannya serta dapat membantu pihak universitas dalam merekomendasikan program studi dengan bidang ilmu yang sesuai untuk menyusun strategi penerimaan yang lebih terarah. Penelitian ini menerapkan metode clustering dengan algoritma K-Means untuk

Kata kunci:

Clustering, K-Means,
Universitas Almuslim, Google
Colab, Minat Akademik

Corresponding Author:

[Noratul Iqramah](#)

Universitas Almuslim

membantu mengelompokkan calon mahasiswa ke dalam dua bidang ilmu yaitu sains dan sosial. Pengelompokan bidang ilmu ini didasarkan pada 1000 data calon mahasiswa baru dengan atribut nilai ijazah (Matematika, IPA, Bahasa Indonesia, dan IPS), nilai tes, serta minat bidang. Proses analisis yang dilakukan dengan menggunakan metode K-Means clustering pada Google Colab menghasilkan Perhitungan sebanyak 17 iterasi, C1 (sains) dengan jumlah 518 orang yang memiliki minat dan nilai lebih tinggi dalam bidang sains, dan C2 (sosial) dengan jumlah 482 orang yang memiliki minat dan nilai lebih tinggi dalam bidang sosial. Pembagian ini menegaskan bahwa algoritma K-Means mampu mengelompokkan data berdasarkan karakteristik yang ada dalam dataset. Dengan hasil tersebut, K-Means Clustering terbukti efektif dalam mengelompokkan calon mahasiswa Universitas Almuslim berdasarkan latar belakang dan minat akademik mereka.

PENDAHULUAN

Universitas di seluruh dunia termasuk Universitas Almuslim di hadapkan pada upaya untuk mengoptimalkan strategi penerimaan mahasiswa baru guna menjaga standar akademik yang tinggi dan memenuhi ekspektasi calon mahasiswa baru [1]. Calon mahasiswa baru merupakan komponen terpenting bagi suatu Universitas [2], [3]. Setiap mereka memiliki latar belakang, minat, dan potensi yang beragam [4]. Oleh karena itu, universitas harus memiliki pemahaman yang mendalam terhadap karakteristik calon mahasiswa, yang menjadi kunci utama dalam merancang sistem penerimaan yang efektif dan relevan [5]. Hal ini mengakibatkan pemilihan bidang ilmu yang kadang-kadang tidak sesuai dengan kebutuhan mahasiswa di masa depan [6], [7].

Tujuan dari pemilihan bidang ilmu yang tepat adalah agar mahasiswa dapat belajar secara optimal, dan mampu menyesuaikan diri serta mengasah potensi berdasarkan bakat, minat, dan kemampuannya dalam dunia kerja [8], [9]. Dengan menerapkan teknik clustering melalui data mining (DM), informasi yang strategis dapat diidentifikasi, dan kemudian dapat dimanfaatkan untuk menemukan peluang dan merancang strategi baru dalam proses penempatan mahasiswa ke bidang ilmu yang sesuai [10], [11]. Untuk mengatasi hal tersebut maka solusinya melakukan clustering [12]. clustering adalah salah satu teknik data mining yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik antar data yang ada [13], [14]. Clustering di lakukan menggunakan metode K-Means.

K-means adalah metode pengelompokan non-hierarki yang membagi data menjadi satu atau lebih cluster [15], [16]. Dalam metode ini, data dikelompokkan berdasarkan kesamaan karakteristik. Algoritma K-Means Clustering mampu mengelompokkan data yang memiliki kesamaan ke dalam satu cluster, serta memisahkan data yang berbeda ke dalam cluster yang lain [17], [18]. Cara kerja K-Means adalah dengan memilih secara acak titik-titik awal sebagai pusat cluster (centroid), kemudian mengelompokkan titik-titik data lainnya ke dalam cluster berdasarkan jarak Euclidean dari centroid terdekat [19], [20]. Setelah itu, Setiap cluster memiliki centroid baru yang diperoleh dari rata-rata posisi titik-titik data di dalamnya [21], [22]. Proses ini dilakukan secara berulang tanpa terjadi pergeseran yang signifikan dalam posisi centroid atau titik-titik data bergeser antar klaster.

Penelitian ini fokus pada penentuan bidang ilmu yang didasarkan dari nilai Ijazah (Matematika, IPA, IPS, Bahasa Indonesia) minat bidang dan nilai tes. Melalui penelitian ini diharapkan mampu mengarahkan mahasiswa ke bidang ilmu yang sesuai dengan potensi dan kemampuannya. Ada dua bidang yaitu : Sosial dan Sains. Data ijazah dikumpulkan melalui proses manual saat pendaftaran mahasiswa baru di Universitas Almuslim.

METODE

Langkah-langkah penelitian K-Means.

- Menentukan Jumlah Cluster (K), menentukan jumlah kluster (K) yang diperlukan, bisa berdasarkan kebutuhan atau menggunakan metode seperti elbow method untuk menemukan jumlah optimal kluster.
- Inisialisasi pusat cluster, memilih pusat awal setiap kluster k titik secara acak.
- Pengelompokkan data, beri label pada setiap titik data berdasarkan pusat kluster terdekat.
- Memperbarui pusat cluster, menghitung ulang pusat tiap cluster berdasarkan rata-rata dari semua titik data dalam cluster.
- Iterasi, ulangi langkah iterasi hingga tidak ada perubahan pada pengelompokkan atau mencapai jumlah iterasi yang telah ditentukan.

Adapun tahapan dalam mengimplementasikan algoritma K-Means terhadap data calon mahasiswa baru adalah sebagai berikut:

- Inisialisasi

Pada tahap inisialisasi, menentukan jumlah kluster yaitu dua kluster (sains dan sosial), dengan nilai centroid awal seperti yang ditampilkan pada Tabel 1 Nilai centroid awal dipilih secara acak untuk memulai proses clustering

Tabel 1 Jumlah cluster dan centeroid awal

Cluster	MTK	IPA	Tes Sains	Minat Sains	Bahasa Indonesia	IPS	Tes Sosial	Minaat Sosial
C1	90	85	85	1	60	60	60	0
C2	60	60	60	0	80	90	85	1

- Assignment

Setiap titik data didistribusikan ke kluster yang memiliki pusat terdekat dengan menggunakan suatu metrik jarak, biasanya jarak Euclidean dengan menggunakan persamaan (1), hasil perhitungan sebagai berikut. Perhitungan dari data atas nama camaba_1 dan camaba_2 ke tiap cluster yaitu:

$$d(1,1)$$

$$= \sqrt{(76 - 90)^2 - (71 - 85)^2 - (97 - 85)^2 - (0 - 1)^2 - (92 - 60)^2 - (70 - 60)^2 - (79 - 60)^2 - (1 - 0)^2}$$

$$d(1,2)$$

$$= \sqrt{(76 - 60)^2 - (71 - 60)^2 - (97 - 60)^2 - (0 - 0)^2 - (92 - 80)^2 - (70 - 90)^2 - (79 - 85)^2 - (1 - 1)^2}$$

$$d(2,1)$$

$$= \sqrt{(89 - 90)^2 - (75 - 85)^2 - (70 - 85)^2 - (0 - 1)^2 - (94 - 60)^2 - (86 - 60)^2 - (81 - 60)^2 - (0 - 0)^2}$$

$$d(2,2)$$

$$= \sqrt{(89 - 60)^2 - (75 - 60)^2 - (70 - 60)^2 - (0 - 0)^2 - (94 - 80)^2 - (86 - 90)^2 - (81 - 85)^2 - (0 - 1)^2}$$

Hasil perhitungan kluster dapat dilihat pada tabel 2.

Tabel 2 Iterasi Awal

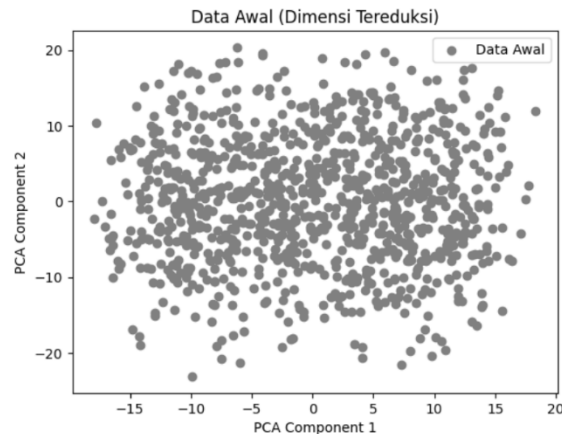
Nama	C1	C2	Cluster
Camaba_1	44,978	48,229	C1
Camaba_2	50,990	37,350	C2
Camaba_3	50,764	55,785	C1

Kemudian Menentukan ulang pusat masing-masing kelompok dengan menghitung rata-rata dari seluruh titik data dalam kelompok tersebut. Adapun hasil centroid-nya adalah sebagai berikut:

Tabel 3 Centeroid Kedua

Cluster	MTK	IPA	Tes Sains	Minat Sains	Bahasa Indonesia	IPS	Tes Sosial	Minaat Sosial
C1	80	80,5	96,5	0,5	94,5	77,5	79,5	1
C2	89	75	70,0	0,0	94,0	86,0	81,0	0,0

Perhitungan kembali data camaba_1 dan camaba_2 dengan menggunakan centeroid kedua:



$$d(1,1)$$

$$= \sqrt{(76 - 80)^2 - (71 - 80,5)^2 - (97 - 96,5)^2 - (0 - 0,5)^2 - (92 - 94,5)^2 - (70 - 77,5)^2 - (79 - 79,5)^2 - (1 - 1)^2}$$

$$d(1,2)$$

$$= \sqrt{(76 - 89)^2 - (71 - 75)^2 - (97 - 70)^2 - (0 - 0)^2 - (92 - 94)^2 - (70 - 86)^2 - (79 - 81)^2 - (1 - 0)^2}$$

$$d(2,1)$$

$$= \sqrt{(89 - 80)^2 - (75 - 80,5)^2 - (70 - 96,5)^2 - (0 - 0,5)^2 - (94 - 94,5)^2 - (86 - 77,5)^2 - (81 - 76,5)^2 - (0 - 1)^2}$$

$$d(2,2)$$

$$= \sqrt{(89 - 89)^2 - (75 - 75)^2 - (70 - 70)^2 - (0 - 0)^2 - (94 - 94)^2 - (86 - 86)^2 - (81 - 81)^2 - (0 - 0)^2}$$

Hasil perhitungan klaster pada centroid kedua dapat dilihat pada tabel 4.

Tabel 4 Iterasi Kedua

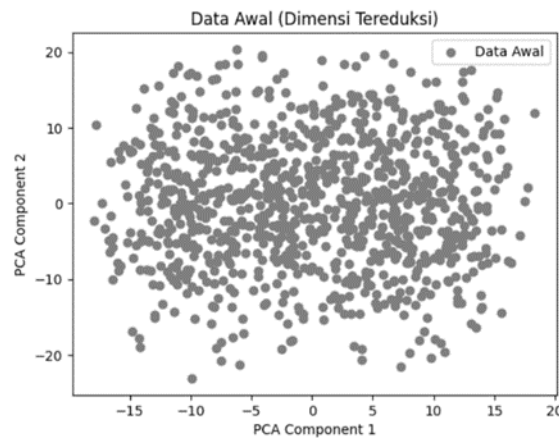
Nama	C1	C2	Cluster
Camaba_1	13,01922	34,33657	C1
Camaba_2	29,82449	0	C2
Camaba_3	13,01922	30,64311	C1

Karena pada iterasi kedua tidak terjadi perubahan pada hasil cluster dibandingkan dengan iterasi pertama, maka proses dihentikan.

HASIL DAN PEMBAHASAN

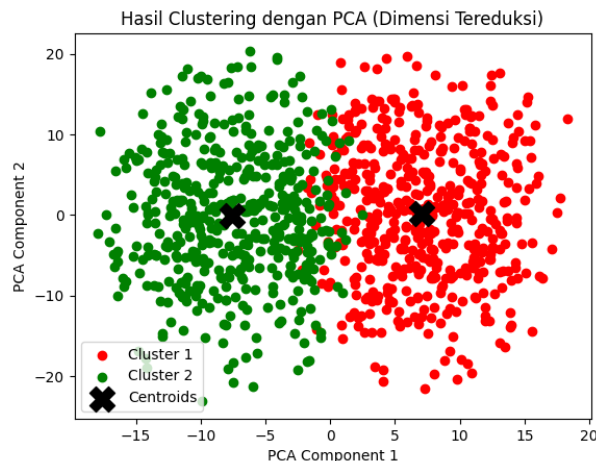
a. Hasil Clustering

Data calon mahasiswa Universitas Almuslim berjumlah 1000 data yang diolah menggunakan metode k-means clustering dengan bantuan google colab. Sebelum dilakukan proses clustering data tersebut telah dilakukan praproses data untuk memastikan kualitas dan konsistensi data yang akan digunakan dalam proses. Dengan demikian, data awal yang telah direduksi menggunakan PCA dapat divisualisasikan seperti yang ditampilkan pada gambar ini.



Gambar 1. Data awal

Hasil analisis klaster menggunakan metode K-Means Clustering terhadap calon mahasiswa baru Universitas Almuslim menunjukkan dua kelompok utama. Klaster pertama terdiri dari 518 data calon mahasiswa yang memiliki nilai tinggi pada variabel Nilai Matematika, Nilai IPA, Tes Sains, dan Minat Sains, yang mencerminkan kelompok dengan potensi besar di bidang sains. Klaster kedua terdiri dari 482 data calon mahasiswa dengan nilai tinggi pada variabel Nilai Bahasa Indonesia, Nilai IPS, Tes Sosial, dan Minat Sosial, yang lebih cenderung ke arah bidang sosial. Hasil ini memberikan gambaran yang jelas tentang profil calon mahasiswa berdasarkan potensi dan minat mereka, sehingga dapat menjadi dasar dalam proses seleksi dan penempatan bidang ilmu yang sesuai. Hasil analisis lengkapnya dapat dilihat pada Gambar 2, yang menampilkan hasil reduksi data hasil proses clustering menggunakan PCA.



Gambar 2. Hasil Clustering.

b. Hasil Akhir

Pada hasil clustering menggunakan K-Means didapatkan hasil yaitu cluster C1 dengan jumlah 518 orang dan cluster C2 berjumlah 482 orang pada dataset nilai calon mahasiswa. Proses ini dimulai dari data dari atribut calon mahasiswa teridentifikasi C1 sebagai kelompok dengan karakteristik yang lebih tinggi dari IPA. Sedangkan pada C2, yang berjumlah 482 menunjukkan minat dan nilai yang lebih tinggi dalam bidang IPS. Pembagian ini menjelaskan bahwa algoritma K-Means dikelompokkan berdasarkan dataset yang tercatat. Pembagian cluster dengan menggunakan K-Means clustering ini menunjukkan adanya persamaan karakteristik data calon mahasiswa yang berarti K-Means mampu mengelompokkan data, ini menjelaskan bahwa algoritma K-Means berhasil dilakukan dengan membentuk kelompok yang sesuai dengan pola-

pola datasetnya [23]. Penelitian ini menunjukkan bahwa teknik clustering K-Means efektif untuk mengelompokkan calon mahasiswa berdasarkan data berdasarkan atribut yang ada, yang digunakan untuk mendalami karakteristik lebih lanjut dari masing-masing kelompok tersebut. Clustering ini memberikan hasil yang relevan dengan tujuan penelitian, yaitu mengelompokkan calon mahasiswa berdasarkan minat yaitu minat IPA dan IPS. Hasilnya dapat dimanfaatkan oleh universitas untuk menyusun strategi penerimaan yang lebih terarah, terstruktur dan memahami preferensi calon mahasiswa.

KESIMPULAN

Dari hasil penelitian yang telah dilakukan, penulis menarik beberapa kesimpulan sebagai berikut:

1. Algoritma K-Means berhasil diterapkan dalam penelitian ini untuk mengelompokkan calon mahasiswa baru Universitas Almuslim ke dalam dua klaster utama, yaitu klaster ilmu sosial dan klaster ilmu sains.
2. Algoritma K-Means menunjukkan efektivitasnya dalam proses pengelompokan data calon mahasiswa baru dengan menggunakan nilai-nilai akademik dan minat calon mahasiswa menghasilkan klaster yang relevan, membuktikan bahwa algoritma ini sangat cocok diterapkan untuk pengelompokan berdasarkan karakteristik yang ada.

REFERENCES

- [1] M. M. Yafi and Taqwanur, "B. ANALISIS VALIDITAS DAN RELIABILITAS INSTRUMEN KEPUASAN PELAYANAN AKADEMIK," *Jurnal Teknik Industri dan Kimia*, vol. 5, no. 1, 2022, doi: 10.54980/jtik.v5i1.186.
- [2] Yovi, Ringgo Dwika, and Eka, "Penerapan Metode Monte Carlo pada Simulasi Prediksi Jumlah Calon Mahasiswa Baru Universitas Muhammadiyah Bengkulu," *Jurnal PROCESSOR*, vol. 17, no. 2, 2022, doi: 10.33998/processor.2022.17.2.1224.
- [3] G. W. N. Wibowo and M. A. Manan, "Penerapan Algoritma Naive Bayes Untuk Prediksi Heregistrasi Calon Mahasiswa Baru," *JTINFO (Jurnal Teknik Informatika)*, vol. 1, no. 1, 2022.
- [4] T. Bartin, "PENDIDIKAN ORANG DEWASA SEBAGAI BASIS PENDIDIKAN NON FORMAL," *Jurnal Teknodik*, 2018, doi: 10.32550/teknodik.v10i19.398.
- [5] R. Hasan, "MENUMBUHKAN SIKAP NASIONALISME DAN BELA NEGARA MAHASISWA MELALUI PENDIDIKAN KEWARGANEGARAAN DI PERGURUAN TINGGI," *Jurnal Tunas Pendidikan*, vol. 5, no. 1, 2022, doi: 10.52060/pgsd.v5i1.890.
- [6] W. Wasilah, I. Azzaroh, N. Rahmawati, and A. Q. Z. Fitriana, "Pengaruh Salah Pilih Jurusan Terhadap Nilai Akademik Mahasiswa Program Studi BKI UIN Khas Jember," *Jurnal Ilmu Sosial, Humaniora dan Seni*, vol. 1, no. 6, 2023, doi: 10.62379/jishs.v1i6.872.
- [7] W. Primayasa, I. Arifin, and M. Y. Baharsyah, "Pengaruh Salah Pilih Jurusan Terhadap Rasa Putus Asa Mahasiswa Teknik Informatika," *Nathiqiyyah*, vol. 3, no. 1, 2020, doi: 10.46781/nathiqiyyah.v3i1.76.
- [8] A. T. Santo and I. N. Alfian, "Hubungan Dukungan Sosial dan Kecemasan dalam Menghadapi Dunia Kerja pada Mahasiswa Akhir," *Buletin Riset Psikologi dan Kesehatan Mental (BRPKM)*, vol. 1, no. 1, 2021, doi: 10.20473/brpkm.v1i1.24895.
- [9] H. Haryati and N. Hasanah, "Kecemasan Mahasiswa Fakultas Dakwah Menghadapi Dunia Kerja," *INNOVATIO: Journal for Religious Innovation Studies*, vol. 19, no. 2, 2020, doi: 10.30631/innovatio.v19i2.88.
- [10] F. Handayani, "Aplikasi Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," *Jurnal Teknologi dan Informasi*, vol. 12, no. 1, 2022, doi: 10.34010/jati.v12i1.6733.
- [11] A. Agneresa, A. L. Hananto, S. S. Hilabi, A. Hananto, and T. Tukino, "Strategi Promosi Penerapan Data Mining Mahasiswa Baru Dengan Metode K-Means Clustering," *Dirgamaya: Jurnal Manajemen dan Sistem Informasi*, vol. 2, no. 2, 2022, doi: 10.35969/dirgamaya.v2i2.275.

- [12] D. Armiady, “Analisis Metode DBSCAN (Density-Based Spatial Clustering of Application with Noise) dalam Mendeteksi Data Outlier,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 6, 2022, doi: 10.30865/jurikom.v9i6.5080.
- [13] D. Armiady and I. Muslem R., “Penetapan Klaster Siswa Unggul Dengan Menggunakan Algoritma Roc-Smarter,” *Jurnal TIKA*, vol. 7, no. 2, 2022, doi: 10.51179/tika.v7i2.1229.
- [14] I. R. Muslem and T. M. Johan, “KLIK: Kajian Ilmiah Informatika dan Komputer Klasifikasi Citra Ikan Menggunakan Algoritma Convolutional Neural Network dengan Arsitektur VGG-16,” *Media Online*, vol. 4, no. 2, pp. 978–985, 2023, doi: 10.30865/klik.v4i2.1209.
- [15] D. Triyansyah and D. Fitriana, “Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing,” *Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 3, 2018, doi: 10.22441/incomtech.v8i3.4174.
- [16] S. Aulia, “KLAUSTERISASI POLA PENJUALAN PESTISIDA MENGGUNAKAN METODE K-MEANS CLUSTERING (STUDI KASUS DI TOKO JUANDA TANI KECAMATAN HUTABAYU RAJA),” *Djtechno: Jurnal Teknologi Informasi*, vol. 1, no. 1, 2021, doi: 10.46576/djtechno.v1i1.964.
- [17] R. NOVIANTO, “Penerapan Data Mining menggunakan Algoritma K-Means Clustering untuk Menganalisa Bisnis Perusahaan Asuransi,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 6, no. 1, 2019, doi: 10.35957/jatisi.v6i1.150.
- [18] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, 2020, doi: 10.25077/teknosi.v6i2.2020.82-89.
- [19] F. Grandoni, R. Ostrovsky, Y. Rabani, L. J. Schulman, and R. Venkat, “A refined approximation for Euclidean k-means,” *Inf Process Lett*, vol. 176, 2022, doi: 10.1016/j.ipl.2022.106251.
- [20] E. Rohaeti, I. M. Sumertajaya, A. H. Wigena, and K. Sadik, “MTSCLust with Handling Missing Data Using VAR-Moving Average Imputation,” *Mathematics and Statistics*, vol. 11, no. 2, 2023, doi: 10.13189/ms.2023.110201.
- [21] M. Andryan, M. Faisal, and R. Kusumawati, “K-Means Binary Search Centroid With Dynamic Cluster for Java Island Health Clustering,” *Jurnal Riset Informatika*, vol. 5, no. 3, 2023, doi: 10.34288/jri.v5i3.511.
- [22] K. Arai and A. Ridho Barakbah, “Hierarchical K-means: an algorithm for centroids initialization for K-means,” 2007.
- [23] L. Luthfi, R. Imam Muslem, D. Armiady, S. Sriwinar, R. Fajri, and I. Iqbal, “Analysis of CNN Method for Image Classification of Coconut Ripeness Levels,” in *2023 Eighth International Conference on Informatics and Computing (ICIC)*, IEEE, Dec. 2023, pp. 1–6. doi: 10.1109/ICIC60109.2023.10381964.