

PENERAPAN ALGORITMA *K-MEANS CLUSTERING* UNTUK MENGELOMPOKKAN DATA MAHASISWA BARU

Erni Ilmiyah¹, Agus Bahtiar²

¹Manajemen Informatika, STMIK IKMI Cirebon

²Sistem Informasi Stmik Ikmi Cirebon

Jalan Perjuangan No 10 B Majasem, Kec. Kesambi, Kota Cirebon

erniilmiyah22@gmail.com

ABSTRAK

Penelitian ini mengeksplorasi metode pengelompokan calon mahasiswa baru berdasarkan preferensi program studi menggunakan *algoritma K-Means Clustering*. Aplikasi *RapidMiner* digunakan sebagai platform untuk menjalankan proses pengelompokan, memanfaatkan *Davies Bouldin Index* sebagai parameter evaluasi kualitas klaster. Dari hasil penelitian, didapatkan pemahaman mendalam mengenai pola pengelompokan calon mahasiswa baru. Metode *K-Means Clustering* menghasilkan dua klaster optimal, ditandai dengan nilai *Davies Bouldin Index* terendah sebesar -0.278 . Penemuan ini memberikan pandangan strategis dalam perancangan kebijakan penerimaan mahasiswa baru dan membantu meningkatkan efisiensi dalam proses seleksi. Dengan adanya informasi ini, perguruan tinggi dapat mengoptimalkan strategi promosi untuk menarik calon mahasiswa sesuai dengan preferensi program studi, meningkatkan kualitas penerimaan, dan mendukung keputusan berbasis data untuk perbaikan kebijakan penerimaan di masa depan.

Kata kunci : *Pengelompokan, Pemilihan Jumlah Cluster, Data Mahasiswa.*

1. PENDAHULUAN

Pendidikan tinggi mengalami perkembangan dinamis seiring perubahan sosial dan teknologi. Pemilihan calon mahasiswa menjadi aspek krusial dalam memastikan kecocokan dengan visi dan misi program studi. Penelitian ini menggunakan metode *K-Means Clustering* untuk mengelompokkan calon mahasiswa baru berdasarkan kategori Lulus pada Prodi dan Lulus Pilihan program studi. Tujuannya adalah memberikan pemahaman lebih mendalam terhadap karakteristik penerimaan mahasiswa, sehingga institusi pendidikan dapat merancang strategi penerimaan yang lebih efektif.

Mengelompokkan data berdasarkan karakteristik tertentu, efektivitas metode ini dalam mengelompokkan calon mahasiswa berdasarkan kategori pendaftaran dan preferensi program studi. Penelitian bertujuan mengeksplorasi kontribusi positif metode *K-Means Clustering* terhadap perancangan strategi penerimaan mahasiswa yang efektif.

Penelitian terdahulu yang relevan dengan topik ini berjudul "*KLASTERISASI CALON MAHASISWA BARU MENGGUNAKAN ALGORITMA K-MEANS*". Teknologi Informasi berkembang dengan pesat dan penggunaannya juga semakin luas. Pengolahan data hingga pengambilan keputusan dapat dilakukan dengan bantuan perangkat Teknologi Informasi. Demikian juga dengan pengambilan keputusan untuk penerimaan mahasiswa baru. Sistem seleksi Proses penerimaan mahasiswa baru di STMM melibatkan dua jenis ujian, yakni CBT (Computer Based Test) dan wawancara, yang dilaksanakan secara bersamaan pada hari yang sama. Penilaian dilakukan dengan menggabungkan kedua nilai tersebut dengan proporsi tertentu, sehingga dapat diperoleh peringkat. Calon mahasiswa baru. Berdasar pada permasalahan

tersebut penulis pada kesempatan Ini akan menghasilkan pengelompokan data calon mahasiswa berdasarkan nilai *Computer Based Test (CBT)*. dan wawancara menggunakan *algoritma K-Means*. Tujuan dari penelitian ini adalah untuk mengeksplorasi pola pengelompokan nilai CBT dan melakukan analisis *klasterisasi* terhadap data calon mahasiswa. wawancara calon mahasiswa baru STMM Yogyakarta. Hasilnya Ada empat klaster [1].

Setelah dilakukan pengelompokan data mahasiswa berdasarkan program studi, asal daerah, dan jenis sekolah menggunakan *K-Means clustering* terbentuk 3 (tiga) *cluster*. *Cluster 1* lebih banyak memilih program studi SIA berasal dari sekolah SMA di daerah. Denpasar, Badung, Dari Gianyar, Tabanan, dan Jembrana, mayoritas calon mahasiswa pada *Cluster 2* cenderung memilih program studi Informatika. dari sekolah SMK di daerah Denpasar, Badung, Gianyar, Karangasem, dan Tabanan. Mayoritas calon mahasiswa dalam *Cluster 3* cenderung berasal dari daerah tersebut. memilih program studi SI berasal dari sekolah SMA di daerah Denpasar, Badung, Gianyar, Tabanan, dan Jembrana. Hasil clustering ini selanjutnya dapat digunakan untuk menentukan strategi promosi pada persebaran daerah berdasarkan jenis sekolah dan program studi yang diminati[1].

2. TINJAUAN PUSTAKA

2.1. Penerimaan Calon Mahasiswa baru

Apabila melibatkan proses penerimaan mahasiswa baru di sebuah perguruan tinggi, data baru yang terbentuk berupa profil mahasiswa baru. Selanjutnya, mahasiswa akan menjalani kegiatan pembelajaran setiap semester, yang menghasilkan data indeks prestasi mahasiswa pada akhir semester. Proses

ini berulang di setiap semester perguruan tinggi. Jika dilakukan pengolahan data dari kedua sumber tersebut, dapat diperoleh berbagai informasi yang berguna untuk menentukan strategi promosi penerimaan mahasiswa baru pada tahun-tahun berikutnya. Pengolahan data tersebut dapat menggunakan metode Data Mining dengan menggunakan *algoritma K-Means Clustering*[2].

Penerimaan mahasiswa baru, yang rutin diselenggarakan sebelum awal semester Gasal setiap tahunnya, melibatkan pengumpulan informasi dari calon mahasiswa baru yang berasal dari berbagai daerah. Dari data yang terkumpul tersebut, dapat ditarik kesimpulan penting dengan melakukan pengolahan data agar menghasilkan informasi baru. Informasi yang dihasilkan melalui pengolahan data ini dapat digunakan sebagai dasar dalam pengambilan keputusan bagi Universitas. Namun, pengolahan data dalam skala besar yang mengandung banyak *record* tidak dapat dilakukan dengan mudah dan cepat untuk mendapatkan informasi yang diinginkan[3].

2.2. Data Mining

Data Mining merupakan konsep yang digunakan untuk menggali pengetahuan yang tersembunyi dalam suatu basis data. Proses *Data Mining* melibatkan pemanfaatan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengidentifikasi informasi berharga serta pengetahuan yang terdapat dalam berbagai basis data besar[4].

Data Mining mencakup metode *Clustering* yang menggunakan *algoritma K-Means*. Kegiatan *Clustering* bertujuan untuk mengelompokkan data berdasarkan tingkat kemiripan antar data, sehingga data dalam satu kelompok memiliki kemiripan tinggi dan data antar kelompok memiliki kemiripan yang rendah[5].

Data Mining mencakup metode *Clustering* yang menggunakan *algoritma K-Means*. Kegiatan *Clustering* bertujuan untuk mengelompokkan data berdasarkan tingkat kemiripan antar data, sehingga data dalam satu kelompok memiliki kemiripan tinggi dan data antar kelompok memiliki kemiripan yang rendah(Narwati, 2020)

2.3. K-Means Clustering

Teknik *clustering* adalah salah satu pendekatan dalam pengelompokan data yang tetap menjadi salah satu metode yang populer hingga saat ini. Penggunaan analisis pengelompokan data ini telah meluas di berbagai bidang, termasuk dalam analisis data medis, analisis pasar, analisis gambar, video, dan bidang lainnya. *Clustering* merupakan proses pengelompokan data yang bersifat *unsupervised*, di mana analisis dilakukan pada data yang tidak memiliki label, sehingga memungkinkan untuk mengidentifikasi pola-pola tersembunyi dalam data tersebut.[7].

Langkah-langkah melakukan *Clustering* dengan metode *K-Means* adalah sebagai berikut:

- a) Pilih jumlah *Cluster K*.
- b) Inisialisasi *K* pusat *Cluster* ini bisa di lakukan dengan berbagai cara. Namun yang palig sering di lakukan adalah dengan cara random. Pusat-pusat *Cluster* di beri nilai awal dengan langkah-langkah random.
- c) Alokasikan semua data atau objek ke *Cluster* terdekat. Kedekatan dua objek di tentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *Cluster* tertentu di tentukan jarak antara data dengan pusat *Cluster*. Jarak paling antara satu data dengan satu *Cluster* tertentu akan menentukan suatu data masuk dalam *Cluster* mana.

Untuk menghitung jarak semua data ke setiap titik pusat *Cluster* dapat menggunakan teori jarak *Euclidean* yang di rumuskan sebagai berikut:

$$D(i,j) = (X_{xi} - X_{1j}) + (X_{2i} - X_{2j}) + \dots + (X_{ki} + X_{kj})$$

Dimana:

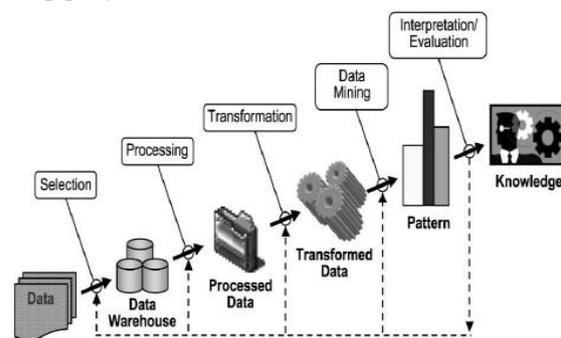
$D(i,j)$ = jarak data ke *i* ke pusat cluster *j*

X_{ki} = Data ke *i* atribut data ke *k*

X_{kj} = Titik pusat ke *j* pada atribut k [8].

3. METODE PENELITIAN

Metode yang diterapkan dalam analisis adalah *Knowledge Discovery in Databases (KDD)*. Metode ini merupakan suatu proses yang tidak sederhana untuk menemukan pola yang baru, memiliki manfaat, dan dapat dipahami dengan mudah.[9]. Tahapan perancangan Tahapan perancangan adalah pedoman dan langkah-langkah yang diterapkan dalam suatu penelitian guna mencapai tujuan penelitian tersebut. Tahapan perancangan dalam Penelitian ini disusun dengan menggunakan struktur atau format tertentu berdasarkan proses *Knowledge Discovery In Database*, dengan menerapkan langkah demi langkah yang dilaksanakan secara berurutan untuk mencapai tahap penyelesaian.

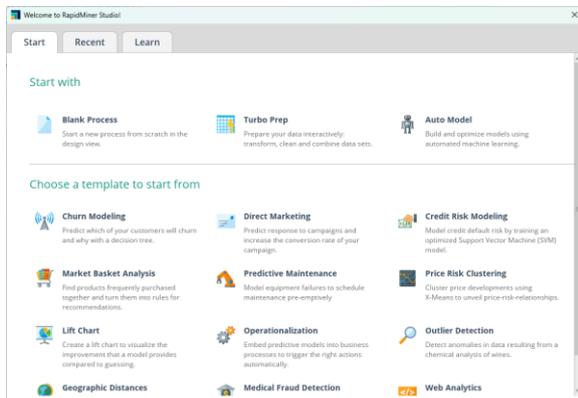


Gambar 1. Tahap proses KDD

Pada gambar 1 Metodologi yang diterapkan pada penelitian kali ini Metodologi yang digunakan dalam penelitian ini adalah *Knowledge Discovery in Databases (KDD)* yang telah terstruktur dalam lima tahap, meliputi seleksi data, *preprocessing* atau pembersihan data, transformasi data, data mining untuk mencari pola atau informasi, dan interpretasi atau evaluasi data.

4. HASIL DAN PEMBAHASAN

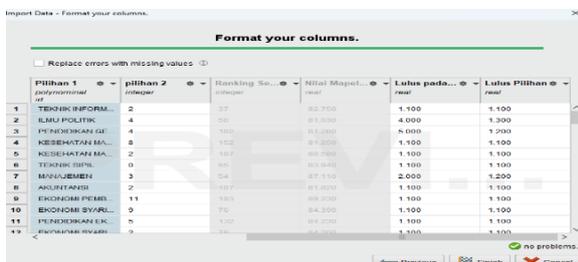
Pada gambar 2 Merupakan tampilan awal dari aplikasi RapidMiner Studio v9.10 ketika pertama kali dibuka setelah proses loading selesai.



Gambar 2. Tampilan Awal RapidMiner

Pada tahap awal ini dilakukan data selection dimulai dengan pemilihan data yang akan digunakan dengan memilih atribut-atribut yang benar-benar berpengaruh dan berguna untuk menjawab pertanyaan atau tujuan analisis. Data yang didapat diberikan dalam bentuk tabel dengan jumlah 999 Data Mahasiswa dengan 10 atribut. Selanjutnya tahap preprocessing dilakukan pengecekan missing value. Pada dataset hasil data selection tidak terdapat fields dengan data kosong atau strip (-).

Kemudian, data akan mengalami tahap transformasi di mana perubahan dilakukan pada data yang kompleks untuk menjadikannya lebih mudah diolah, seperti melakukan normalisasi data. Proses ini melibatkan penyesuaian skala atau rentang data agar data tersebut dapat diolah dengan lebih efektif. i bertujuan untuk memastikan bahwa rentang nilai dari setiap variabel menjadi seragam, sehingga perbedaan skala tidak menjadi masalah dalam analisis. memengaruhi output dari analisis. Transformasi data membantu mengubah informasi menjadi bentuk yang lebih berguna atau lebih efektif. sesuai untuk analisis atau presentasi.

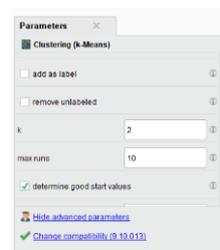


Gambar 3. Data Transformasi

Setelah proses tersebut, data dapat digunakan untuk analisis atau penyajian. Analisis data yang telah tersedia menunjukkan bahwa tidak ada data yang memerlukan transformasi tambahan. Data yang ada sudah dapat diolah dan dianalisis tanpa perlunya penyesuaian tambahan. diberikan telah diverifikasi,

terbukti akurat, dan sesuai dengan kebutuhan analisis. Oleh karena itu, proses dapat melanjutkan ke tahap berikutnya dalam evaluasi tanpa perlu melakukan perubahan lebih lanjut pada data yang telah diberikan

Setelah dilakukan tahap transformasi data seperti pada gambar di atas, tahapan selanjutnya yaitu proses data mining digunakan untuk mengeksplorasi pola-pola yang ada dalam data. Teknik ini digunakan pada tahap ini yaitu pendekatan pengelompokan yaitu teknik yang mengimplementasikan pengelompokan optimasi penentuan jumlah cluster dengan metode "Elbow." Algoritma K-Means bertujuan untuk mengelompokkan data ke dalam beberapa cluster, di mana data dalam satu cluster memiliki karakteristik serupa satu sama lain dan berbeda dengan data dalam cluster lainnya. Metode "Elbow" digunakan untuk menentukan jumlah cluster optimal dengan melihat penurunan varians intra-cluster seiring dengan penambahan jumlah cluster Davies Bouldin Index (DBI) digunakan untuk menentukan banyaknya cluster optimum berdasarkan kedekatan objek terhadap centroidnya dalam satu cluster dan jarak antar centroid dan cluster.



Gambar 4. Penentuan jumlah Cluster

Pada gambar 4 dilakukan pengaturan nilai k. Nilai k ini ditentukan dengan mengambil nilai Davies Bouldin Index (DBI) yang paling kecil. Dalam dataset ini nilai DBI yang paling kecil berada di k=2, dengan max run = 10. Dibawah ini adalah tabel perbandingan nilai DBI.

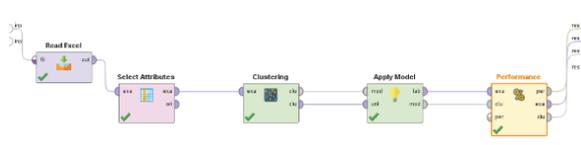
Tabel 1. Perbandingan Nilai DBI

Tabel Perbandingan DBI			
Nilai K	Max Run	Max Optimization Steps	Davies Bouldin Index
2	10	100	-0.274
3	10	100	-0.463
4	10	100	-0.458
5	10	100	-0.415
6	10	100	-0.390
7	10	100	-0.380
8	10	100	-0.407
9	10	100	-0.384
10	10	100	-0.390

Dapat dilihat dari tabel 1 di atas menunjukkan perbandingan DBI dari k = 2 hingga k= 10 dengan max run = 10, tabel di atas merupakan perbandingan hasil dari pengolahan di RapidMiner. Disini menunjukkan bahwa dikelompokkan Lulus pada pilihan dan Lulus pada prodi menjadi dua (2) cluster. Pemilihan jumlah

cluster berdasarkan nilai yang paling kecil pada nilai *Davies Bouldin Index* (DBI).

Dengan mempertimbangkan nilai DBI untuk $k = 2$ hingga $k = 10$, analisis ini memastikan bahwa dua (2) *cluster* memberikan representasi yang optimal terhadap struktur data Calon Masiswa Baru yang dianalisis. Pemilihan jumlah *cluster* yang sesuai merupakan langkah kritis dalam analisis *cluster*, dan nilai DBI memberikan panduan yang kuat dalam menentukan model *cluster* yang paling baik untuk data yang diberikan.



Gambar 5. Pemodelan Algoritma K-Means Clustering

Pada Gambar 5 Perancangan pengolahan data menggunakan algoritma *K-Means Clustering* pada *RapidMiner* menggunakan lima (5) operator, menggunakan *Read Excel* untuk membaca dataset Pilihan 1 berdasarkan Program Studi, menggunakan *Select Attributes* untuk memilih atribut mana yang akan digunakan, menggunakan operator *K-Means Clustering* untuk mengelompokkan dataset dengan nilai $K=2$ dan $\max \text{runs} = 10$. Operator *Apply Model* digunakan untuk membaca kumpulan data dalam *cluster* berdasarkan data yang diselidiki. Operator *Cluster Performance Distance* digunakan untuk mengukur kinerja algoritma *K-Means clustering* pada dataset yang telah dibaca.

Dalam tahap selanjutnya, yaitu proses evaluasi, dilakukan penilaian dan pengukuran kualitas, akurasi, dan keberlanjutan data yang digunakan dalam suatu analisis atau model. Evaluasi ini merupakan langkah penting untuk memastikan bahwa hasil yang diperoleh dari analisis atau model tersebut dapat diandalkan dan relevan. Dengan melakukan evaluasi secara cermat, kita dapat mengidentifikasi potensi kekurangan atau ketidakpastian dalam data yang digunakan, sehingga memungkinkan pengguna untuk membuat keputusan yang lebih informasional dan tepat.

Proses evaluasi juga bertujuan untuk menentukan sejauh mana model atau analisis dapat mencapai tingkat akurasi yang diharapkan. Dengan mengukur performa model secara sistematis, kita dapat memahami sejauh mana model tersebut dapat menggeneralisasi data dengan baik dan memberikan hasil yang konsisten. Selain itu, evaluasi juga memungkinkan penyesuaian atau peningkatan model jika ditemukan kelemahan atau ketidaksesuaian dengan kebutuhan pengguna. Selanjutnya, proses *Evaluation* juga melibatkan pengevaluasian keberlanjutan data yang digunakan. Ini mencakup pemantauan apakah data yang digunakan masih relevan dan representatif terhadap kondisi aktual. Dengan memastikan keberlanjutan data, kita dapat meningkatkan keandalan model dan memastikan

bahwa analisis yang dihasilkan tetap relevan seiring waktu.

Hasil dari evaluasi ini nantinya akan menjadi dasar bagi pengambilan keputusan lebih lanjut terkait dengan penerapan atau pengembangan model. Oleh karena itu, langkah evaluasi ini bukan hanya sebagai penentu kualitas, tetapi juga sebagai panduan untuk perbaikan dan pengoptimalan yang berkelanjutan. Dengan demikian, evaluasi menjadi elemen integral dalam siklus analisis data yang komprehensif. Hasil pada gambar 4.4 ini memunculkan banyaknya anggota dari beberapa kelompok (*cluster*) total semua anggota yang ada pada dataset ini adalah 999 items.

5. KESIMPULAN DAN SARAN

Berdasarkan data yang telah dianalisis, penelitian ini menyimpulkan beberapa hal : Dapat diketahui cara pengelompokan Data pendaftaran mahasiswa baru pada data diatas menerapkan pendekatan *K-Means Clustering* yang dijalankan dengan aplikasi *RapidMiner*. Dengan melihat nilai *Davies Bouldin Index* kita dapat mengetahui kelompok atau *cluster* dari masing-masing data. Dapat diketahui hasil pengelompokan jumlah *cluster* (K) yang dihasilkan dari nilai K yang optimal adalah pengelompokan ini menghasilkan 2 *cluster* yang memiliki nilai $k -0.278$ yang optimal dilihat dari nilai *Davies Bouldin Index* yang paling rendah.

Untuk peneliti selanjutnya dalam mengelompokan data mahasiswa berdasarkan program studi, sebaiknya Pada langkah-langkah pengolahan data, data calon mahasiswa baru yang melibatkan metode *K-Means Clustering* menggunakan operator-operator pada *RapidMiner* dihasilkan beberapa kesimpulan. selain metode algoritma *K-Means Clustering*. Penelitian ini memberikan wawasan yang berharga dalam Kesimpulan dari pengelompokan data mahasiswa dengan metode *K-Means* dapat digunakan sebagai landasan untuk membuat keputusan strategis dalam pemilihan program studi. Harapannya, kesimpulan ini akan memberikan dampak positif terhadap peningkatan kualitas dan efisiensi di masa mendatang.

DAFTAR PUSTAKA

- [1] A. A. I. I. P. Nengah Widya Utami, "Penerapan Data Mining Untuk Mengetahui Pola Pemilihan Program Studi Di Stmik Primakara Menggunakan Algoritma K-Means ...," *J. Teknol. Inf. dan ...*, vol. 3, pp. 456–463, 2021, [Online]. Available: <http://jurnal.undhirabali.ac.id/index.php/jutik/article/view/1540>
- [2] M. R. Alhapizi, M. Nasir, and I. Effendy, "Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang," *J. Softw. Eng. Ampera*, vol. 1, no. 1, pp. 1–14, 2020, doi: 10.51519/journalsea.v1i1.10.

- [3] A. L. Maukar, F. Marisa, and A. A. Widodo, "Analisis Data Penerimaan Mahasiswa Baru Berbasis K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 6, no. 2, p. 142, 2022, doi: 10.26798/jiko.v6i2.558.
- [4] T. Syahputra, "Penerapan Data Mining Dalam Pengelompokan Kelas Mahasiswa Berdasarkan Hasil Ujian Saringan Masuk Dengan Algoritma K-Means," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 5, no. 2, pp. 161–166, 2019, doi: 10.33330/jurteks.v5i2.350.
- [5] S. Dewi, S. Defit, and Y. Yuhandri, "Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 28–33, 2021, doi: 10.37034/jsisfotek.v3i1.40.
- [6] Narwati, "Pengelompokan Mahasiswa Menggunakan Algoritma K-Means," *J. Din. Inform.*, vol. 2, no. 2, pp. 1–7, 2010.
- [7] D. Tambun, S. Fauziah, and M. Danny, "Pengelompokan Penerimaan Mahasiswa Baru Dengan Algoritma K-Means Untuk Meningkatkan Potensi Pemasaran," *Bull. Inf. Technol.*, vol. 4, no. 2, pp. 294–298, 2023, doi: 10.47065/bit.v3i1.
- [8] Haris Kurniawan, Sarjon Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.
- [9] N. A. Rahmalinda and A. Jananto, "Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru," *J. Tekno Kompak*, vol. 16, no. 2, pp. 163–175, 2022.