



Penerapan algoritma k-means untuk pengelompokan data mahasiswa baru Program Studi Teknik Informatika di Universitas Pahlawan Tuanku Tambusai

Kasini¹✉, Hidayati Rusnedy¹, Lailatul Syifa Tanjung², Novi Yona Sidratul Munti³

Teknik Informatika, Fakultas Teknik, Universitas Pahlawan Tuanku Tambusai ⁽¹⁾

Teknik Industri, Fakultas Teknik, Universitas Pahlawan Tuanku Tambusai ⁽²⁾

Teknik Informatika, Universitas Islam Sumatera Barat ⁽³⁾

DOI: 10.31004/jutin.v8i1.41449

✉ Corresponding author:

[kasiniaqm@gmail.com]

Article Info

Abstrak

Kata kunci:

Data Mining;

Klustering;

K-Means;

Teknik Informatika;

Universitas Pahlawan

Tuanku Tambusai

Universitas Pahlawan Tuanku Tambusai (UP) di Provinsi Riau memiliki Program Studi Teknik Informatika yang menerima mahasiswa baru setiap tahunnya dari berbagai daerah di sekitar Bangkinang. Data mahasiswa yang masuk diolah untuk membantu pengambilan keputusan, terutama dalam bidang promosi. Penelitian ini bertujuan untuk menerapkan algoritma K-Means pada data mahasiswa Program Studi Teknik Informatika, dengan atribut nama mahasiswa dan kecamatan daerah asal, untuk mengelompokkan daerah berdasarkan potensi promosi. Metode K-Means digunakan untuk mengelompokkan data ke dalam tiga cluster: Prioritas Tinggi, Prioritas Sedang, dan Prioritas Rendah. Hasil analisis menunjukkan bahwa terdapat 22 daerah yang termasuk dalam Cluster Prioritas Rendah, 23 daerah dalam Cluster Prioritas Sedang, dan 43 daerah dalam Cluster Prioritas Tinggi. Daerah-daerah dalam Cluster Prioritas Tinggi menjadi prioritas utama untuk strategi promosi sehingga tidak memerlukan cara pendekatan promosi yang lebih terfokus, sementara daerah dalam Cluster Prioritas Sedang dan Prioritas Rendah memerlukan pendekatan promosi yang lebih terfokus. Penelitian ini memberikan kontribusi penting bagi strategi promosi Program Studi Teknik Informatika di UP dengan menggunakan pendekatan data mining untuk meningkatkan visibilitas program studi di masyarakat.

Abstract

Keywords:

Data Mining;

Clustering;

K-Means;

Informatics Engineering;

Universitas Pahlawan

Tuanku Tambusai

Pahlawan Tuanku Tambusai University (UP) in Riau Province has an Informatics Engineering Study Program that accepts new students every year from various regions around Bangkinang. Incoming student data is processed to assist decision making, especially in the field of promotion. This study aims to apply the K-Means algorithm to Informatics Engineering Study Program student data, with attributes of student name and district of origin, to group regions based on promotion potential. The K-Means method is used to group data into three clusters: High Priority, Medium Priority, and Low Priority. The results of the analysis show that there are 22 regions included in the High Priority Cluster, 23 regions in the Medium Priority Cluster, and 43 regions in the Low Priority Cluster. Regions in the High Priority Cluster are the main priority for promotion strategies,

while regions in the Medium Priority and Low Priority Clusters require a more focused promotion approach. This study provides an important contribution to the promotion strategy of the Informatics Engineering Study Program at UP by using a data mining approach to increase the visibility of the study program in the community.

1. INTRODUCTION

Universitas Pahlawan Tuanku Tambusai (UPTT) adalah satu-satunya Universitas yang ada di Kabupaten Kampar yang terletak di wilayah Bangkinang, Provinsi Riau (Fadhilaturrahmi et al., 2020). Saat ini, UPTT memiliki 8 Fakultas yang terdiri dari 26 Program Studi, salah satu nya Prodi S1 Teknik Informatika yang menjadi salah satu program studi paling diminati di Fakultas Teknik. Berdasarkan data dari sekretariat Program Studi S1 Teknik Informatika, sejak pertama kali dibuka, jumlah pendaftar baru terus mengalami peningkatan yang signifikan setiap tahunnya, menunjukkan tingginya minat calon mahasiswa terhadap program studi ini.

Meskipun jumlah pendaftar baru terus mengalami peningkatan, potensi penurunan minat di masa depan perlu diantisipasi. Hal ini menjadi penting karena UPTT harus bersaing dengan berbagai perguruan tinggi lainnya, seperti Politeknik Kampar dan sejumlah universitas terkemuka di Kota Pekanbaru. Oleh karena itu, diperlukan strategi promosi yang berkelanjutan dan terfokus untuk memastikan bahwa Program Studi Teknik Informatika tetap menarik perhatian calon mahasiswa di tengah persaingan yang semakin intensif. Untuk mendukung strategi promosi yang lebih efektif, diperlukan analisis berbasis data guna mengidentifikasi daerah-daerah dengan potensi tinggi sebagai target promosi. Data mahasiswa yang berasal dari berbagai kecamatan dapat diolah dan dianalisis dengan proses yang tepat, sehingga informasi tersembunyi dari data mahasiswa dapat diketahui dan bermanfaat bagi pihak perguruan tinggi untuk memudahkan dalam menentukan strategi promosi baru berdasarkan pengelompokan daerah asal mahasiswa (Rahmalinda & Jananto, 2022).

Pendekatan yang dapat diterapkan untuk pengelompokan data adalah menggunakan metode Data Mining, khususnya Clustering (Rusnedy et al., 2021). Clustering merupakan teknik penting dalam data mining yang berfungsi untuk mengolah dan mengelompokkan data secara unsupervised. Pendekatan ini mampu memberikan analisis yang efisien dalam menangani masalah kompleks dengan mengidentifikasi pola-pola menarik pada data, sehingga mendukung proses penemuan pengetahuan (Okfalisa et al., 2021). Algoritma K-Means adalah salah satu metode clustering yang banyak digunakan dalam berbagai bidang, seperti pemasaran dan penjualan (Hidayati & Rahmah, 2022), kesehatan (Rahmayani & Hidayati, 2022), pendidikan (Hidayati & Kasini, 2023) dan masih banyak lainnya.

Tiga variabel utama yang digunakan dalam penelitian ini yaitu nama, kelurahan/desa, dan kecamatan. Variabel nama berfungsi untuk mengidentifikasi individu dalam dataset, sedangkan variabel kelurahan/desa dan kecamatan merepresentasikan asal geografis mahasiswa. Data ini akan diolah menggunakan metode clustering K-Means, dengan pembagian cluster menjadi tiga kategori: Prioritas Tinggi, Prioritas Sedang, dan Prioritas Rendah. Cluster Prioritas Tinggi mencakup kecamatan dengan jumlah mahasiswa terbanyak dan menjadi target utama promosi, karena memiliki potensi besar untuk menjaring calon mahasiswa baru sehingga tidak diperlukan strategi promosi untuk menarik minat mahasiswa baru. Cluster Prioritas Sedang terdiri dari kecamatan dengan jumlah mahasiswa yang cukup signifikan, sehingga memerlukan pendekatan promosi yang terarah. Sementara itu, Cluster Prioritas Rendah mencakup kecamatan dengan jumlah mahasiswa relatif sedikit, yang tetap membutuhkan strategi promosi khusus untuk meningkatkan partisipasi. Pembagian cluster ini bertujuan untuk membantu institusi menyusun strategi promosi yang lebih efektif dan efisien, sehingga sumber daya dapat dialokasikan secara optimal ke daerah-daerah dengan potensi yang berbeda.

Penelitian ini berangkat dari kebutuhan untuk meningkatkan efektivitas promosi Program Studi Teknik Informatika UPTT melalui pendekatan berbasis data. Dengan menganalisis pola distribusi asal mahasiswa, diharapkan institusi dapat merancang strategi promosi yang lebih terarah dan efisien, sekaligus meningkatkan visibilitas program studi di masyarakat.

2. METHODS

Metode yang diterapkan dalam penelitian ini bertujuan untuk menganalisis dan mengolah data mahasiswa Program Studi Teknik Informatika di Universitas Pahlawan Tuanku Tambusai (UP). Penelitian ini memanfaatkan teknik data mining untuk mengidentifikasi pola dan informasi yang terkandung dalam data mahasiswa, dengan penekanan pada potensi promosi berdasarkan asal daerah. Selanjutnya, algoritma K-Means digunakan untuk mengelompokkan data mahasiswa ke dalam beberapa klaster yang didasarkan pada daerah asal mereka, yang kemudian dianalisis untuk mengukur potensi promosi. Proses implementasi metode ini dilakukan menggunakan perangkat lunak Rapid Miner, yang mendukung pelaksanaan data mining dengan cara yang lebih efektif dan efisien.

2.1. Data Mining

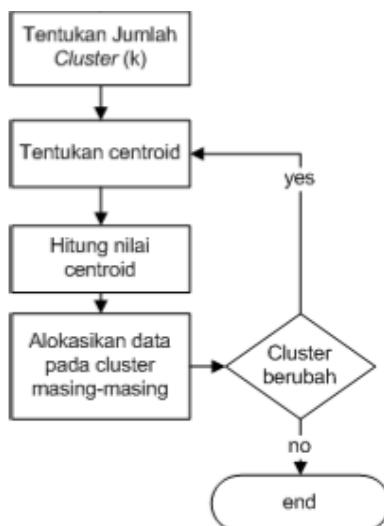
Teknologi data mining merupakan salah satu alat untuk mengekstrak data dari database besar dengan spesifikasi yang sangat kompleks. Data mining sendiri memiliki metode dan teknik untuk memenuhi kebutuhan,

termasuk kebutuhan akan informasi yang luas, dan informasi yang dihasilkan dapat digunakan sebagai alat pengambilan keputusan atau digunakan untuk meningkatkan kualitas proses pengambilan keputusan. Data atau informasi memiliki potensi yang besar untuk dijadikan sebagai kesimpulan pengambilan keputusan dengan menganalisis data yang dikandungnya (Pattipeilohy & Pakereng, 2023).

2.2. K-Means

K-Means adalah salah satu algoritma yang digunakan dalam clustering. Clustering sendiri merupakan sebuah teknik, di mana algoritma clustering berfungsi untuk mengelompokkan sejumlah data ke dalam kelompok-kelompok tertentu (cluster) (Noviyanto, 2020). Algoritma K-Means dikenal luas karena keunggulannya dalam kemudahan penggunaan serta kemampuannya mengklasifikasikan data dalam jumlah besar dan menangani outlier dengan cepat (Aswan et al., 2021). Metode K-Means sudah banyak digunakan dalam berbagai kasus selain penjualan, misalnya pengelompokan buku huruf, tingkat kesembuhan COVID-19, dan pengelompokan data buku di perpustakaan (Hidayati & Rahmah, 2022). Dengan algoritma K-Means diharapkan dapat menemukan clusterisasi yang baik, sehingga menghasilkan informasi berupa nilai minat baca (Fitriani et al., 2020).

Metode dalam data mining yang melakukan proses data clustering non-hirarki yang dimana data dikelompokkan dalam satu bahkan lebih cluster dan merupakan salah satu metode yang dilakukan dengan sistem partisi adalah K-Means. Langkah-langkah melakukan clustering dengan metode K-Means adalah (Pattipeilohy & Pakereng, 2023):



Gambar 1. Tahapan K-Means

1. Tentukan jumlah nilai k sebagai jumlah cluster.
2. Alokasikan data kedalam kelompok secara random.
3. Hitung pusat cluster (centroid) menggunakan mean untuk masing-masing cluster dengan persamaan Euclidean yaitu sebagai berikut :

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - X_{kj})^2}$$

Dimana:

$D_{(i,j)}$ = jarak data ke i ke pusat *cluster* j

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

X_{kj} = Titik pusat ke j pada atribut ke k

4. Alokasikan data berdasarkan jarak terdekat antara data dengan centroidnya.
5. Kembali kelangkah sebelumnya, jika ternyata masih ada data yang berpindah cluster atau jika nilai centroid diatas nilai ambang, atau jika nilai pada fungsi objektif yang digunakan masih diatas ambang.

2.3. Implementasi Rapid Miner

Tahapan berikutnya yang akan dilakukan dalam penelitian ini adalah melakukan implementasi. Pada tahap implementasi ini penulis akan menghitung data karakter siswa yang telah diperoleh menggunakan aplikasi RapidMiner dengan perhitungan algoritma K-Means. Rapidminer adalah sebuah tools atau software untuk

mempermudah pengolahan data, rapid miner tersedia dalam bentuk software free to use atau gratis jika pemakaiannya untuk kepentingan edukasi atau pembelajaran (Nahjan et al., 2023).

2.4. Hasil

Pada tahapan hasil penulis akan mengetahui hasil perhitungan algoritma *K-Means* menggunakan excel dan aplikasi *RapidMiner*. Dari hasil ini didapatkan kesimpulan dan saran dalam pengelompokan karakter siswa sesuai dengan atribut yang diteliti dan hasil ini akan diberikan kepada pihak sekolah untuk kebijakan selanjutnya

3. RESULT AND DISCUSSION

Data yang digunakan dalam proses pengelompokan mahasiswa didasarkan pada kelurahan atau desa dan kecamatan asal mahasiswa yang telah tersedia di bagian sekretariat Program Studi Teknik Informatika. Data mahasiswa ini mencakup daerah di Kabupaten Kampar dan kemudian dikategorikan berdasarkan kelurahan dan kecamatan. Data tersebut selanjutnya akan diproses untuk menghasilkan wawasan yang berguna dalam merumuskan strategi promosi Program Studi Teknik Informatika. Dalam penerapan data mining, pengelompokan mahasiswa akan dibagi menjadi tiga klaster, yaitu Prioritas Tinggi, Prioritas Sedang, dan Prioritas Rendah. Untuk itu, perlu diterapkan variabel-variabel yang digunakan sebagai acuan dalam pengujian algoritma *K-Means*, yang terdiri dari variabel-variabel berikut:

Table 1. Data Variabel yang Digunakan

No	Nama Variabel	Keterangan
1.	Nama Mahasiswa	Nama mahasiswa berdasarkan angkatan
2.	Kelurahan/desa	Daerah desa asal
3.	Kecamatan	Kecamatan daerah asal

Penelitian ini menggunakan data primer mahasiswa Program Studi Teknik Informatika Universitas Pahlawan Tuanku Tambusai (UPTT) sebagai bahan utama untuk analisis dan pengelompokan. Data ini mencakup informasi penting seperti nomor induk mahasiswa (NIM), nama mahasiswa, dan alamat tempat tinggal yang meliputi kelurahan, kecamatan, hingga kabupaten/kota asal mahasiswa. Pengolahan data ini bertujuan untuk menemukan pola-pola strategis yang relevan dalam menentukan daerah prioritas promosi bagi program studi. Pada Tabel 2 ditampilkan sebagian data primer mahasiswa yang menjadi objek penelitian ini.

Table 2. Data Primer Mahasiswa

No	NIM	Nama	Alamat
1.	2155201022	Novira Wahyuni	Jln. Melati RT.02 RW.05, Desa Bangkinang, Kec. Bangkinang, Kab. Kampar
2.	2155201028	Roqi Nazif	Jln. Prof. M. Yamin SH RT.02 RW.07, Desa Langgini, Kec. Bangkinang Kota, Kab. Kampar
3.	2155201027	Rofiq Nazif	Jln. Prof. M. Yamin SH RT.02 RW.07, Desa Langgini, Kec. Bangkinang Kota, Kab. Kampar
4.	2155201005	Dandy Aditya Rafael	Jln. KH. Agussalim RT.01 RW.06, Desa Bangkinang Kota, Kec. Bangkinang Kota, Kab. Kampar
5.	2155201014	M. Yusuf Ariyandi. S	Desa Kumantan RT.01 RW.04, Desa Kumantan, Kec. Bangkinang Kota, Kab. Kampar
6.	2155201013	Lilis Mahdalena	Ridan Permai RT.01 RW.02, Desa Ridan Permai, Kec. Bangkinang, Kab. Kampar
7.	2155201008	Djoel Jalalianifallah	Petapanaha Jaya RT.14 RW.07, Desa Petapanaha, Kec. Tapung, Kab. Kampar
8.	2155201024	Parhan Prayoga Firmansyah	Dusun Lan RT.03 RW.02, Desa Merangin, Kec. Kuok, Kab. Kampar
9.	2155201001	Adam Elfiansyah	Dusun Rantau Berangin RT.03 RW.02, Desa Merangin, Kec. Kuok, Kab. Kampar
10.	2155201004	Alman Azizi	Desa Sipungguk RT.03 RW.02, Desa Sipungguk, Kec. Bangkinang, Kab. Kampar
...
88	2255201049	Micky Ari Finanda	Dusun Lan RT.04 RW.02, Desa Merangin, Kec. Kuok, Kab. Kampar
89	2255201051	Okta Bernaldi	Dusun 1 Balai Jering RT.02 RW.01, Desa Sungai Jalau, Kec. Kampar Utara, Kab. Kampar
90	2255201050	Muhammad Dhifa Ferdian	Dusun Tanjung RT.02 RW.01, Desa Sawah, Kec. Kampar Utara, Kab. Kampar
91	2255201052	Aditia Ananda Putra	Jln. A. Rahman Saleh Tunas Mandiri RT.02 RW.08, Desa Bangkinang, Kec. Kampar, Kab. Kampar

No	NIM	Nama	Alamat
92	2255201053	Ahmad Rayhan Wahidy	Dusun 1 Naumbai RT.01 RW.01, Desa Naumbai, Kec. Kampar, Kab. Kampar
93	2255201054	Jefri Efendi	Dusun 1 Tanjung Alai RT.03 RW.01, Desa Tanjung Alai, Kec. XIII Koto Kampar, Kab. Kampar
94	2255201055	Marsa Aulia Hasanah	Jln. Stasiun RT.00 RW.00, Desa Kampung Lalang, Kec. Sunggal, Kab. Deli Serdang
95	2255201056	Arif Rahman	Dusun Sungai Deras RT.01 RW.01, Desa Lereng, Kec. Kuok, Kab. Kampar
96	2255201057	Dina Meysi Chandra	Limau Manis RT.04 RW.02, Desa Limau Manis, Kec. Kampar, Kab. Kampar
97	2255201058	Sri Wiji Astuti	Dusun Ii Pasar Selatan RT.02 RW.04, Desa Kampar, Kec. Kampa, Kab. Kampar
98	2255201059	Amar Rozi	Bungur Besar GG IX RT.14 RW.01, Desa Bungur, Kec. Senen, Kota Jakarta Pusat

Berdasarkan data yang ditampilkan pada Tabel 2 di atas informasi yang tersedia masih perlu disederhanakan agar sesuai dengan variabel pengolahan data yang relevan. Penyederhanaan data ini bertujuan untuk mempermudah analisis serta pengelompokan wilayah berdasarkan potensi promosi. Tabel 3 berikut menampilkan data mahasiswa yang telah disederhanakan berdasarkan variabel penelitian yang telah ditentukan, sehingga dapat digunakan untuk analisis lebih lanjut.

Table 3. Data Sesuai Variabel

No	Nama	Kelurahan/Desa	Kecamatan
1.	Novira Wahyuni	Bangkinang	Bangkinang
2.	Roqi Nazif	Langgini	Bangkinang Kota
3.	Rofiq Nazif	Langgini	Bangkinang Kota
4.	Dandy Aditya Rafael	bangkinang Kota	Bangkinang Kota
5.	M. Yusuf Ariyandi. S	Kumantan	Bangkinang Kota
6.	Lilis Mahdalena	Ridan Permai	Bangkinang Kota
7.	Djoel Jalalianifallah	Petapahan Jaya	Tapung
8.	Parhan Prayoga Firmansyah	Merangin	Kuok
9.	Adam Elfiansyah	Merangin	Kuok
10.	Alman Azizi	Sipungguk	Salo
...
79	Zilla Tulhusna	Koto Tuo	XIII Koto Kampar
80	Micky Ari Finanda	Merangin	Kuok
81	Okta Bernaldi	Sungai Jalau	Kampar Utara
82	Muhammad Dhifa Ferdian	Sawah	Kampar Utara
83	Aditia Ananda Putra	Bangkinang	Bangkinang Kota
84	Ahmad Rayhan Wahidy	Naumbai	Kampar
85	Jefri Efendi	Tanjung Alai	XIII Koto Kampar
86	Arif Rahman	Lereng	Kuok
87	Dina Meysi Chandra	Limau Manis	Kampar
88	Sri Wiji Astuti	Kampar	Kampar Timur

Data yang tercantum pada Tabel 3 belum bisa diolah langsung menggunakan metode K-Means. Hal ini disebabkan karena format data yang masih berupa teks, sehingga tidak memenuhi syarat untuk dilakukan penghitungan berbasis angka. Oleh karena itu, diperlukan proses transformasi data agar variabel yang digunakan dapat direpresentasikan dalam bentuk numerik.

Proses transformasi ini dilakukan dengan memberikan inisial numerik untuk setiap kecamatan yang terdapat dalam data mahasiswa. Inisial ini bertujuan untuk menggantikan nama kecamatan menjadi nilai angka, yang lebih kompatibel untuk perhitungan dalam algoritma K-Means. Berikut ini data transformasi yang akan diolah menggunakan metode K-Means.

Table 4. Data Transformasi Variabel Kecamatan

No	Kecamatan	Inisial
1.	Bangkinang Kota	1
2.	Bangkinang	2
3.	Tapung	3
4.	Kuok	4
5.	Kampar Timur	5
6.	Kampar	6
7.	Kampar Utara	7
8.	Tambang	8

No	Kecamatan	Inisial
9.	Salo	9
10.	XIII Koto Kampar	10
11.	Koto Kampar Hulu	11

Setelah data kecamatan ditransformasikan, langkah serupa juga dilakukan untuk variabel kelurahan/desa. Data nama kelurahan/desa diubah menjadi format numerik melalui pemberian inisial angka, adapun data transformasi untuk variable Kelurahan/Desa tedapat pada tabel 5 berikut.

Table 5. Data Transformasi Variabel Kelurahan/Desa

No	Kelurahan/Desa	Inisial	No	Kelurahan/Desa	Inisial
1.	Bangkinang	1	23.	Salo	23
2.	Langgini	2	24.	Batu Belah	24
3.	Bangkinang Kota	3	25.	Siabu	25
4.	Kumantan	4	26.	Batu Bersurat	26
5.	Ridan Permai	5	27.	Tanjung Rambutan	27
6.	Petapahan Jaya	6	28.	Kel. Pulau	28
7.	Merangin	7	29.	Padang Mutung	29
8.	Sipungguk	8	30.	Koto Tuo	30
9.	Koto Tibun	9	31.	Simpang Kubu	31
10.	Bandur Picak	10	32.	Kenantan	32
11.	Nagah Beralih	11	33.	Rumbio	33
12.	Pulau	12	34.	Ranah Baru	34
13.	Air Tiris	13	35.	Laboy Jaya	35
14.	Kuok	14	36.	Petapahan	36
15.	Kualu NENAS	15	37.	Bukit Sembilan	37
16.	Tanjung berulak	16	38.	Pasir Sialang	38
17.	Muara Mahat Baru	17	39.	Sawah	39
18.	Ganting Damai	18	40.	Naumbai	40
19.	Sungai Jalau	19	41.	Tanjung Alai	41
20.	Ganting	20	42.	Lereng	42
21.	Pulau Permai	21	43.	Limau Manis	43
22.	Salo Timur	22	44.	Kampar	44

Setelah data melalui proses transformasi, langkah berikutnya adalah mengolahnya menggunakan algoritma K-Means. Tahapan awal pada metode ini dimulai dengan menentukan jumlah cluster atau kelompok yang akan digunakan untuk mengelompokkan data. Setelah jumlah cluster ditentukan, tahapan selanjutnya adalah menentukan nilai pusat atau centroid awal. Penentuan centroid awal dilakukan secara acak dari dataset yang telah tersedia untuk memulai iterasi pertama. Dalam penelitian ini, tiga cluster yang dipilih sebagai centroid awal adalah data ke-20 untuk cluster Prioritas Rendah (C1), data ke-39 untuk cluster Prioritas Sedang (C2), dan data ke-85 untuk cluster Prioritas Tinggi (C3).

Table 6. Nilai Pusat Centroid Awal

Data ke-	Centroid	Kelurahan/Desa	Kecamatan
20	C1	14	4
39	C2	23	9
85	C3	41	10

Setelah menentukan nilai pusat centroid awal, tahapan selanjutnya adalah menghitung jarak antara setiap data dan pusat cluster (centroid) yang telah ditentukan pada langkah sebelumnya. Perhitungan jarak ini biasanya menggunakan rumus jarak Euclidean, yang merupakan metode standar untuk menentukan jarak terdekat antara dua titik dalam ruang multidimensi. Proses ini bertujuan untuk mengukur sejauh mana setiap data terkait dengan setiap centroid, sehingga dapat menentukan keanggotaan data terhadap cluster tertentu, seperti C1 (Cluster Prioritas Rendah), C2 (Cluster Prioritas Sedang) atau C3 (Cluster Prioritas Tinggi).

Pada iterasi pertama, perhitungan dimulai dengan menentukan jarak semua data ke pusat C1. Setiap data akan dibandingkan dengan centroid C1, dan hasilnya digunakan untuk mengevaluasi apakah data tersebut lebih dekat ke cluster C1 atau mungkin ke cluster lainnya seperti C2 atau C3. Proses ini dilakukan untuk seluruh data dalam dataset, sehingga pola awal pembagian cluster mulai terbentuk berdasarkan kedekatan data dengan masing-masing centroid. Perhitungan ini merupakan langkah krusial yang memengaruhi hasil akhir pengelompokan pada iterasi berikutnya. Hasil cluster pada iterasi 1 dapat dilihat pada tabel 7 berikut.

Table 7. Hasil Perhitungan Iterasi-1

No	Nama	Kelurahan/ Desa	Kecamatan	C1	C2	C3	Jarak Terdekat	Cluster
1.	Novira Wahyuni	1	2	41	41	13	13	3
2.	Roqi Nazif	2	1	40	40	12	12	3
3.	Rofiq Nazif	2	1	40	40	12	12	3
4.	Dandy Aditya Rafael	3	1	39	39	11	11	3
5.	M. Yusuf Ariyandi. S	4	1	38	38	10	10	3
6.	Lilis Mahdalena	5	1	37	37	9	9	3
7.	Djoel Jalalianifallah	6	3	36	36	8	8	3
8.	Parhan Prayoga Firmansyah	7	4	35	35	7	7	3
9.	Adam Elfiansyah	7	4	35	35	7	7	3
10.	Alman Azizi	8	9	33	33	8	8	3
...
79	Zilla Tulhusna	30	10	11	7	17	7	2
80	Micky Ari Finanda	7	4	35	17	7	7	3
81	Okta Bernaldi	19	7	22	4	6	4	2
82	Muhammad Dhifa Ferdian	39	7	4	16	25	4	1
83	Aditia Ananda Putra	1	1	41	23	13	13	3
84	Ahmad Rayhan Wahidy	40	6	4	17	26	4	1
85	Jefri Efendi	41	10	0	18	28	0	1
86	Arif Rahman	42	4	6	20	28	6	1
87	Dina Meysi Chandra	43	6	4	20	29	4	1
88	Sri Wiji Astuti	44	5	6	21	30	6	1

Berdasarkan tabel di atas, pengelompokan data pada iterasi 1 menghasilkan Cluster 1 (C1) dengan 16 data, Cluster 2 (C2) dengan 23 data, dan Cluster 3 (C3) dengan 49 data. Langkah selanjutnya adalah menentukan nilai centroid baru, yang dihitung berdasarkan data yang telah masuk ke dalam masing-masing cluster pada iterasi sebelumnya.

Setelah nilai centroid baru ditemukan, proses perhitungan jarak antara data dan centroid diulang, seperti pada langkah sebelumnya, hingga semua data dikelompokkan kembali ke dalam cluster yang sesuai. Proses ini terus dilakukan berulang-ulang dengan cara yang sama hingga data dalam setiap cluster tidak mengalami perubahan posisi antara iterasi sebelumnya dan iterasi berikutnya. Dengan kata lain, ketika data dalam setiap cluster tetap stabil dan tidak berpindah, maka perhitungan nilai centroid dapat dihentikan.

Dalam penelitian ini, posisi data dalam cluster mencapai kestabilan pada iterasi ke-6. Artinya, data dalam setiap cluster pada iterasi ke-6 tetap sama dengan iterasi sebelumnya. Oleh karena itu, proses iterasi dihentikan, dan hasil akhir pengelompokan data pada iterasi ke-6 dapat dilihat pada Tabel 8.

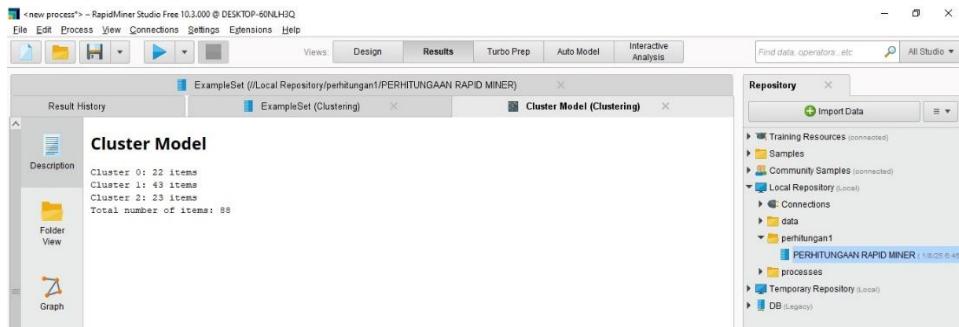
Table 8. Hasil Perhitungan Iterasi-6

No	Nama	Kelurahan/ Desa	Kecamatan	C1	C2	C3	Jarak Terdekat	Cluster
1.	Novira Wahyuni	1	2	34	21	3	3	3
2.	Roqi Nazif	2	1	34	20	3	3	3
3.	Rofiq Nazif	2	1	33	20	3	3	3
4.	Dandy Aditya Rafael	3	1	32	19	2	2	3
5.	M. Yusuf Ariyandi. S	4	1	31	18	2	2	3
6.	Lilis Mahdalena	5	1	30	17	2	2	3
7.	Djoel Jalalianifallah	6	3	29	16	2	2	3
8.	Parhan Prayoga Firmansyah	7	4	28	14	3	3	3
9.	Adam Elfiansyah	7	4	28	14	3	3	3
10.	Alman Azizi	8	9	27	13	7	7	3
...
79	Zilla Tulhusna	30	10	7	9	27	7	1
80	Micky Ari Finanda	7	4	28	14	3	3	3
81	Okta Bernaldi	19	7	16	2	16	2	2
82	Muhammad Dhifa Ferdian	39	7	4	18	35	4	1
83	Aditia Ananda Putra	1	1	34	21	4	4	3
84	Ahmad Rayhan Wahidy	40	6	5	19	36	5	1
85	Jefri Efendi	41	10	8	20	38	8	1
86	Arif Rahman	42	4	7	21	38	7	1
87	Dina Meysi Chandra	43	6	8	22	39	8	1
88	Sri Wiji Astuti	44	5	9	23	40	9	1

Hasil akhir perhitungan menggunakan algoritma K-Means Clustering pada data mahasiswa menunjukkan wilayah yang memerlukan peningkatan strategi promosi agar Program Studi Teknik Informatika

dapat dikenal lebih luas oleh masyarakat maupun sekolah-sekolah di sekitar Bangkinang. Dari proses pengelompokan, ditemukan bahwa Cluster 1 (Tinggi) mencakup 22 daerah, Cluster 2 (Sedang) mencakup 23 daerah, dan Cluster 3 (Rendah) mencakup 43 daerah. Wilayah-wilayah tersebut akan menjadi fokus utama dalam strategi promosi Program Studi Teknik Informatika.

Sementara itu, hasil implementasi menggunakan RapidMiner Studio 10.3 berdasarkan data mahasiswa dengan variabel wilayah menghasilkan hasil pengelompokan sebagai berikut:



Gambar 2. Hasil Perhitungan Rapid Miner

4. CONCLUSION

Penelitian ini bertujuan untuk meningkatkan efektivitas strategi promosi Program Studi Teknik Informatika Universitas Pahlawan Tuanku Tambusai (UPTT) melalui penerapan metode K-Means untuk mengelompokkan data mahasiswa berdasarkan asal geografis. Sebagai satu-satunya universitas di Kabupaten Kampar, UPTT menghadapi persaingan ketat dengan perguruan tinggi lain, seperti Politeknik Kampar dan universitas di Pekanbaru, sehingga diperlukan strategi promosi yang berkelanjutan dan berbasis data untuk tetap menarik minat calon mahasiswa. Dengan menggunakan variabel nama, kelurahan/desa, dan kecamatan, hasil analisis menunjukkan bahwa data mahasiswa terbagi menjadi 22 daerah dalam Cluster Prioritas Rendah, 23 daerah dalam Cluster Prioritas Sedang, dan 43 daerah dalam Cluster Prioritas Tinggi.. Daerah-daerah dengan peningkatan strategi promosi pada daerah Cluster Prioritas Tinggi merupakan daerah-daerah yang paling diprioritaskan untuk dilakukan promosi yang lebih lagi dari daerah lainnya, Cluster Prioritas Sedang merupakan daerah yang tingkat promosinya menengah dan Cluster Prioritas Rendah merupakan daerah dengan tingkat promosi yang tidak begitu diprioritaskan karena sudah banyak mahasiswa yang mengetahui Prodi Teknik Informatika. Penelitian ini memberikan kontribusi penting bagi strategi promosi UPTT dengan menggunakan pendekatan data mining untuk meningkatkan visibilitas program studi di masyarakat, menyusun strategi promosi yang efisien, dan mengoptimalkan alokasi sumber daya berdasarkan potensi daerah masing-masing.

5. REFERENCES

- Aswan, Y., Defit, S., & Nurcahyo, G. W. (2021). Algoritma K-Means Clustering dalam Mengklasifikasi Data Daerah Rawan Tindak Kriminalitas (Polres Kepulauan Mentawai). *Jurnal Sistem Informasi Dan Teknologi*, 3, 245–250. <https://doi.org/10.37034/jsisfotek.v3i4.73>
- Fadhilaturrahmi, F., Erlinawati, E., & Ananda, R. (2020). Workshop Sinta 2 dan Google Scholar di Universitas Pahlawan Tuanku Tambusai. *Jurnal Abdidas*, 1(4), 203–209. <https://doi.org/10.31004/abdidias.v1i4.45>
- Hidayati, N., & Kasini, K. (2023). Classification of books at SMP YPK Pematang Siantar using the k-means clustering method. *Jurnal Mantik*, 7(2). <http://iocscience.org/ejournal/index.php/mantik/article/view/3846%0Ahttp://iocscience.org/ejournal/index.php/mantik/article/download/3846/2825>
- Hidayati, N., & Rahmah, S. A. (2022). Clasterization Of Zeeida Product Sales Using K - Means Method In Medan Distributors. *Jurnal Mantik*, 6(36), 1685–1692. <https://ejournal.iocscience.org/index.php/mantik/article/view/2545%0Ahttps://ejournal.iocscience.org/index.php/mantik/article/download/2545/2108>
- Nahjan, M. R., Nono Heryana, & Apriade Voutama. (2023). Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 101–104. <https://doi.org/10.36040/jati.v7i1.6094>
- Noviyanto, N. (2020). Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia. *Paradigma - Jurnal Komputer Dan Informatika*, 22(2), 183–188. <https://doi.org/10.31294/p.v22i2.8808>
- Okfalisa, Angraini, Novi, S., Rusnedy, H., Handayani, L., & Mustakim. (2021). Identification of the Distribution Village Maturation : Village Classification Using Density-Based Spatial Clustering of Application with Noise. *Jurnal Teknologi Dan Sistem Komputer*, Apr. <https://doi.org/10.14710/jtsiskom.2021.13998>

- Pattipeilohy, R. L., & Pakereng, M. A. I. (2023). Penerapan K-Means Clustering Pada Data Mahasiswa Fakultas Interdisiplin Program Studi D4 Destinasi Pariwisata Untuk Menentukan Strategi Promosi. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 7(1), 320–331.
- Rahmalinda, N. A., & Jananto, A. (2022). Penerapan Metode K-Means Clustering Dalam Menentukan Strategi Promosi Berdasarkan Data Penerimaan Mahasiswa Baru. *Jurnal Tekno Kompak*, 16(2), 163. <https://doi.org/10.33365/jtk.v16i2.1971>
- Rahmayani, M. T. I., & Hidayati, N. (2022). Implemention K-Means Algorithm Determine the Recovery Rate Of Covid-19 Patients In Indonesia. *Jurnal Mantik*, 6(36), 127–135. <http://iocscience.org/ejournal/index.php/mantik/article/view/2059%0Ahttps://iocscience.org/ejournal/index.php/mantik/article/download/2059/1754>
- Rusnedy, H., Nurcahyo, G. W., & Sumijan, S. (2021). Identifikasi Tingkat Pemakaian Obat Menggunakan Metode Fuzzy C-Means. *Jurnal Informasi Dan Teknologi*, 3, 196–201. <https://doi.org/10.37034/jidt.v3i4.152>