

PENGELOMPOKKAN DATA MAHASISWA MENGGUNAKAN CLUSTERING UNTUK OPTIMALISASI PENERIMAAN MAHASISWA BARU

Diana Yusuf¹⁾, Elliya Sestri²⁾, Fahrul Razi¹

¹ Program Studi Sistem Informasi Fakultas Teknik dan Desain

^{2,3} Program Studi Teknologi Informasi Fakultas Teknik dan Desain

Institut Teknologi dan Bisnis Ahmad Dahlan, Jl. Ir. H. Juanda No. 77, Ciputat, Kota Tangerang Selatan

Co Responden Email: dianayusuf01@gmail.com

Abstract

Article history

Received 30 Sep 2024

Revised 12 Oct 2024

Accepted 22 Oct 2024

Available online 31 Oct 2024

Keywords

Clustering

New student

K-Means Clustering

Student data clustering

Student admissions

One of the important stages in managing a university is the new student admission process, where this process will affect the quality and quantity of students accepted at the university. Optimizing this process requires an effective approach to analyzing student potential data. Where student data will be grouped using the K-Means clustering algorithm to find patterns and characteristics that can optimize new student admissions. The application of the K-Means algorithm variables such as study programs, GPA, sub-districts, cities, provinces, and types of schools. The results of the grouping are expected to provide deeper insight into the segmentation of prospective students, so that universities can develop more targeted admission strategies. The research conducted has the benefit of helping to increase the efficiency and effectiveness of institutions in determining new student admission targets, minimizing costs in the recruitment process, and maximizing the potential of students to be accepted. In addition, it is also expected to be a basis for decision makers to be more data-based to improve the quality of student admissions in the future.

Abstrak

Riwayat

Diterima 30 Sep 2024

Revisi 12 Okt 2024

Disetujui 22 Okt 2024

Terbit online 31 Okt 2024

Kata Kunci

Clustering

K-Means

Mahasiswa baru

Pengelompokan data mahasiswa

Penerimaan mahasiswa

Salah satu tahapan penting dalam pengelolaan perguruan tinggi yakni proses penerimaan mahasiswa baru, dimana proses ini akan mempengaruhi kualitas dan kuantitas mahasiswa yang diterima di perguruan tinggi. Mengoptimalkan proses ini memerlukan pendekatan yang efektif untuk menganalisis data potensi mahasiswa. Dimana akan dilakukan pengelompokan data mahasiswa menggunakan algoritma clustering K-Means untuk menemukan pola dan karakteristik yang dapat mengoptimalkan penerimaan mahasiswa baru. Penerapan algoritma K-Means variabel-variabel seperti program studi, IPK, kelurahan, kota, provinsi, dan jenis sekolah. Hasil pengelompokan diharapkan dapat memberikan wawasan lebih dalam mengenai segmentasi calon mahasiswa, sehingga perguruan tinggi dapat menyusun strategi penerimaan yang lebih tepat sasaran. Penelitian yang dilakukan ini memiliki manfaat membantu meningkatkan efisiensi dan efektivitas institusi dalam menentukan target penerimaan mahasiswa baru, meminimalisir biaya pada proses rekrutmen, dan dapat memaksimalkan potensi mahasiswa yang akan diterima. Selain itu diharapkan juga bisa menjadi dasar bagi pengambil keputusan agar lebih berbasis data untuk meningkatkan kualitas penerimaan mahasiswa pada masa mendatang.

PENDAHULUAN

Proses penerimaan mahasiswa baru yang saat ini terjadi di ITB Ahmad Dahlan terdapat beberapa jalur pendaftaran yang bisa diikuti oleh calon mahasiswa seperti jalur prestasi, jalur beasiswa, jalur reguler, dan jalur lainnya yang tertuang dalam buku pedoman penerimaan mahasiswa baru di lingkungan ITB Ahmad Dahlan. Ketatnya persaingan antara perguruan tinggi yang ada di sekitaran ITB Ahmad Dahlan

membuat tim penerimaan mahasiswa harus mampu berinovasi lebih luas dalam memperkenalkan dan mempromosikan ITB Ahmad Dahlan agar dikenal oleh masyarakat luas sehingga dapat meningkatkan jumlah mahasiswa baru.

Berbagai kegiatan promosi pada dasarnya telah dilakukan seperti mengikuti pameran, kunjungan sekolah, website, iklan sosial media, iklan billboard, dan lain sebagainya. Akan tetapi

belum khusus membedah data historis yang ada untuk melihat sebaran mahasiswa ITB Ahmad Dahlan. Oleh karena itu, ITB Ahmad Dahlan membutuhkan suatu model clustering yang akan memberikan rekomendasi-rekomendasi strategi promosi kampus ITB Ahmad Dahlan berdasarkan pada data historis dari mahasiswa aktif. Promosi bisa lebih mengerucut pada daerah tertentu atau untuk segmen dengan status pekerjaan.

Algoritma K-Means pada teknik clustering adalah satu dari beberapa algoritma yang masuk dalam kelompok **unsupervised learning** dan cukup populer digunakan dalam proses pengelompokan data (Sari et al., 2020). Algoritma ini bekerja dengan cara membagi data menjadi **k** kelompok (cluster) berdasarkan kemiripan antar data (Hendrastuty, 2024).

K-Means clustering juga didefinisikan sebagai metode yang berguna untuk mengelompokkan set objek data dengan karakteristik yang mirip. Sehingga alasan memilih algoritma k-means clustering untuk menganalisis dan mengelompokkan data mahasiswa untuk menentukan strategi promosi sangat tepat (Afiasari et al., 2023).

Terdapat beberapa penelitian yang membahas mengenai teknik clustering seperti penelitian dengan judul implementasi k-means clustering untuk pengelompokan daerah rawan bencana kebakaran menggunakan model CRISP-DM, penelitian ini berhasil melakukan pengelompokan terhadap daerah yang berpotensi tinggi mengalami kebakaran sehingga masyarakat sekitar mengenali bahaya sedari awal yang mungkin terjadi pada daerah tersebut (Dhewayani et al., 2022).

Penelitian lain yang membahas teknik clustering dengan algoritma k-means yakni dengan judul implementasi algoritma k-means clustering dalam penentuan siswa kelas unggulan, hasil penelitian ini menyatakan bahwa implementasi algoritma k-means ke dalam sistem informasi memberikan hasil yang efektif dan waktu prosesnya juga lebih ringkas dalam menentukan klasterisasi kelas unggulan (Sulistiyawati & Supriyanto, 2021).

Penelitian yang berkaitan dengan teknik clustering yang pernah dilakukan oleh Afiasari N dengan judul Implementasi data mining transaksi penjualan menggunakan algoritma clustering

Sebuah penelitian yang juga menerapkan data mining dengan Algoritma K-Means Clustering bertujuan untuk membuat kebijakan-

kebijakan terkait promosi kepada mahasiswa baru di lingkungan Universitas Bina Darma Palembang, hasil penelitian menyatakan pengelompokan data mahasiswa jika dilihat dari sebaran daerah asal, sekolah asal, pendapatan orang tua dan informasi-informasi lainnya berkaitan dengan institusi, diperoleh sebuah pengetahuan bahwa cluster terbanyak adalah strategi promosi yang menggunakan brosur dan promosi dari teman atau kerabat dekat (Alhapizi et al., 2020).

Penelitian yang akan dilakukan ini nantinya untuk mengidentifikasi pola-pola yang relevan pada data historis mahasiswa ITB Ahmad Dahlan, dimana dapat merekomendasikan strategi promosi yang lebih tepat sasaran.

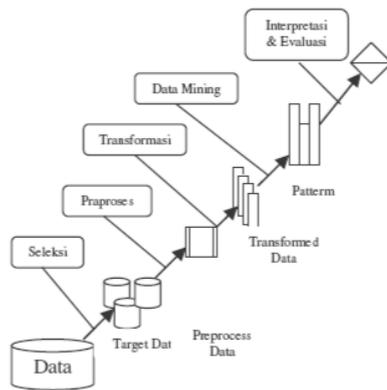
Mengidentifikasi pola-pola yang relevan dalam data mahasiswa sehingga dapat digunakan untuk pembuatan strategi-strategi penerimaan mahasiswa baru yang lebih tepat sasaran. Merekomendasikan strategi promosi yang optimal berdasarkan hasil pengelompokan data mahasiswa menggunakan Algoritma K-Means agar jumlah mahasiswa baru mengalami peningkatan

Kebanyakan kasus data mining digunakan untuk menghasilkan wawasan/pengetahuan dari data yang ada. Tidak hanya itu hasil data mining bahkan lebih banyak digunakan dalam kasus praktis seperti proses pengambilan keputusan suatu organisasi. Misalnya, suatu organisasi bisnis bisa mendapatkan informasi detail terkait segmentasi pelanggan yang banyak menggunakan produk mereka dengan memanfaatkan teknik Data Mining. Selain itu hasil dari data mining tersebut nantinya akan dimanfaatkan perusahaan untuk menyusun strategi kedepannya (Yusuf et al., 2023).

Knowledge Discovery in Databases (KDD)

Beberapa ahli berpendapat bahwa data mining adalah suatu proses untuk menganalisis wawasan/pengetahuan yang ada dalam kumpulan data yang besar atau bisa juga dikatakan data mining merupakan salah satu tahapan dalam KDD (Muzakir & Wulandari, 2016). Namun seringkali orang salah pemahaman antara *Knowledge Discovery in Database* (KDD) dengan Data Mining yang beranggapan bahwa keduanya adalah proses yang sama. Namun pada dasarnya antara KDD dan data mining berbeda secara konsep karena jika dilihat berdasarkan tahapan yang ada dalam KDD diketahui bahwa data mining merupakan salah satu tahapan dalam proses KDD, seperti

pada gambar dibawah ini:(Yusuf & Sestri, 2020)



Gambar 1. Tahapan Knowledge Discovery in Databases (KDD)

Tahapan KDD ini antara lain: (Haris Kurniawan et al., 2020)

1. Data Selection

Tahapan ini melakukan pengumpulan data dari sekumpulan data besar dengan cara memilih atribut yang akan digunakan untuk teknik data mining.

2. Data Preprocessing

Tahapan ini biasa disebut juga dengan data cleaning/pembersihan data, dimana pada tahapan ini dilakukan proses perbaikan terhadap kesalahan data, menghapus data, dan memeriksa data-data yang tidak konsisten.

3. Data Transformation

Tahapan ini adalah tahap pengkodean atau perubahan data. Tujuannya adalah untuk mempermudah proses pengolahan data mining nantinya.

4. Data Mining

Tahapan ini merupakan tahap penemuan pola atau pengetahuan yang bermanfaat dalam sekumpulan data yang telah dipilih.

5. Interpretation

Tahapan menampilkan pola atau pengetahuan dari hasil proses data mining ke dalam bentuk visualisasi yang mudah dimengerti oleh pemangku kepentingan.

CRISP-DM

CRISP-DM merupakan suatu metodologi yang digunakan secara luas sebagai panduan pengembangan data mining (Dhewayani et al., 2022). Terdapat beberapa fase dalam CRISP-DM yang akan digunakan dalam pengembangan proyek data mining, yang mana fase-fase tersebut bersifat iteratif dalam pengembangan

proyek data mining (Sulistiyawati & Supriyanto, 2021).

1. Business Understanding

Fase business understanding merupakan fase memahami tujuan bisnis dari proyek data mining dan bagaimana teknik data mining dapat mencapai tujuan tersebut. Pada fase ini dilakukan penetapan mengenai tujuan dan target data mining serta mengidentifikasi kebutuhan.

2. Data Understanding

Data understanding merupakan fase pengumpulan data dari berbagai sumber kemudian melakukan eksplorasi terhadap data tersebut dan memeriksa apakah terdapat anomali pada data. Sumber data pada penelitian ini ialah data historis mahasiswa.

3. Data Preparation

Fase ini dilakukan persiapan terhadap data yang sudah dikumpulkan untuk digunakan pada tahapan modeling. Pada fase ini melibatkan proses transformasi data menjadi format yang cocok digunakan pada model data mining. Penelitian ini melakukan normalisasi atribut terhadap data yang digunakan.

4. Modeling

Fase modelling merupakan fase penerapan salah satu teknik data mining untuk membangun model data mining sehingga menghasilkan knowledge dari data tersebut. Pada penelitian ini digunakan teknik clustering untuk memperoleh model data mining.

5. Evaluation

Setelah fase modelling berhasil dilakukan maka pada fase evaluation akan dilakukan evaluasi untuk memastikan model tersebut memenuhi tujuan dan target data mining yang telah ditetapkan pada tahap awal.

6. Deployment

Fase ini diterapkan model data mining yang dihasilkan dan telah dievaluasi ke dalam lingkungan bisnis nyata.

Data Mining

Data mining merupakan salah satu metode dalam ilmu komputer yang dapat digunakan untuk memperoleh pengetahuan dari hasil penambangan terhadap sekumpulan data dengan volume besar. Teknik yang ada dalam

data mining dapat dimanfaatkan sesuai pengetahuan apa yang ingin diperoleh dari sumber data yang tersedia (Yudhistira & Andika, 2023).

Data mining adalah proses menemukan pola hingga pengetahuan dalam kumpulan data yang jumlahnya besar dan dapat dijadikan acuan untuk mengambil kebijakan ataupun keputusan (Sugianto et al., 2020). Sumber data yang digunakan dalam data mining mencakup basisdata, data warehouse, web, repository, atau data lainnya yang dikumpulkan dalam suatu sistem yang dinamis (Zai, 2022). Data mining juga didefinisikan sebagai proses mengekstraksi informasi dari data yang jumlahnya besar dan kompleks, yang bertujuan untuk informasi-informasi yang bermanfaat bagi pengguna (Rafi Nahjan et al., 2023).

Clustering

Clustering adalah salah satu teknik data mining dengan metode unsupervised learning, yang mana dalam permasalahan clustering nilai yang ada belum memiliki target atau kelasnya belum memiliki label. Sehingga perlu dilakukan prediksi ke dalam kelas mana nantinya objek tersebut akan dimasukkan. Terdapat beberapa algoritma clustering yang dapat digunakan, salah satunya yang paling umum dan sederhana digunakan ialah k-means (Veronica & Effendi, 2023).

METODE PENELITIAN

Tujuan dari penelitian ini ialah untuk mengelompokkan data mahasiswa ITB Ahmad Dahlan Jakarta dengan memanfaatkan algoritma K-Means Clustering agar mampu mengidentifikasi pola dan karakteristik tertentu dari kelompok mahasiswa. Penelitian ini diharapkan dapat mengembangkan strategi-strategi yang optimal dalam proses penerimaan mahasiswa baru, serta juga mampu memberikan rekomendasi keputusan atau kebijakan berbasis data kepada manajemen ITB Ahmad Dahlan Jakarta untuk meningkatkan jumlah mahasiswa baru. Agar tujuan tersebut tercapai maka peneliti akan menguji data mahasiswa aktif ITB Ahmad Dahlan Jakarta Tahun Akademik 2019-2022 dengan memanfaatkan teknik data mining. Adapun beberapa metode dan tahapan yang dilakukan antara lain sebagai berikut:

a. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan melalui sumber data historis mahasiswa ITB Ahmad Dahlan Jakarta. Data historis yang dikumpulkan mencakup antara

lain prodi, jenis kelas, kota/kab, provinsi, IPK, dan sekolah asal mahasiswa aktif ITB Ahmad Dahlan Jakarta. Sampel yang digunakan adalah pada Fakultas Teknik dan Desain ITB Ahmad Dahlan Jakarta yakni tahun akademik 2019-2022. Tabel dibawah ini merupakan data yang akan digunakan dalam proses penelitian ini:

Tabel 1. Data historis mahasiswa

No	Nama	Prodi	Kelas	Kota	Prov	IPK	Asal Sekolah
1	Mohamad Farhan Hafidzudin	DKV	Pagi	Aceh Besar	Aceh	3.0169	SMK
2	Faiz Afza	DKV	Pagi	Balikpapan	Kaltim	3.9355	SMA
3	Riska Sulistian	TI	Malam	Bandar Lampung	Lampung	3.6842	SMK
4	Ardini Annisah Azzahra	ARS	Pagi	Bandar Lampung	Lampung	2.8889	SMA
5	M Haqiqi Arnun Pangestu	DKV	Pagi	Bandar Lampung	Lampung	3.4746	MAN
6	Ferry Alamsyah	TI	Pagi	Bandung	Jawa Barat	2.7483	SMA
.
386	Heri Wahyudin	DKV	Pagi	Wonosobo	Jawa Tengah	3.4098	MAN

b. Transformasi Data

Transformasi data adalah tahapan dilakukannya perubahan data agar algoritma K-Means dapat mengolah data tersebut untuk memperoleh model atau pola. Transformasi yang dilakukan yakni dengan merubah data yang dalam bentuk kategori ke dalam bentuk numerik dengan melakukan proses inialisasi.

Variabel yang digunakan dalam penelitian ini ialah Prodi, Kelas, Kota/Kab, Provinsi, IPK, dan Asal Sekolah.

Tabel 2. Inialisasi variabel prodi

Prodi	Inialisasi
Sistem Informasi (SI)	57
Teknologi Informasi (TI)	59
Arsitektur (ARS)	23
Desain Komunikasi Visual (DKV)	90

Tabel 3. Inialisasi variabel kelas

Kelas	Inialisasi
Pagi	10
Malam	11

Tabel 4. Inialisasi variabel kota/kab

Kota/Kab	Inialisasi
Aceh Besar	1
Balikpapan	2
...	...
Wonosobo	65

Tabel 5. Inialisasi variabel provinsi

Provinsi	Inialisasi
Aceh	11

Kaltim	99
...	...

Tabel 6. Inisialisasi asal sekolah

Asal Sekolah	Inisialisasi
SMA	31
SMK	41
MAN/Ponpes	51

Hasil transformasi dari data yang akan digunakan pada tahapan ini:

Tabel 7. Data hasil transformasi

Prodi	Kelas	Kota/Kab	Prov	IPK	Asal Sekolah
90	10	1	11	3.0169	41
90	10	2	99	3.9355	31
59	11	3	12	3.6842	41
23	10	3	12	2.8889	31
90	10	3	12	3.4746	51
59	10	4	55	2.7483	31
...
90	10	65	66	3.4098	51

c. Pengolahan Data dengan Clustering

Tahapan berikutnya setelah seluruh data selesai ditransformasikan ialah mengolah data tersebut menggunakan algoritma K-Means Clustering. Berikut adalah proses algoritma K-Means Clustering dalam penelitian ini:

1. Proses pertama ialah penentuan jumlah cluster yang akan dibentuk. Penelitian ini penulis menentukan jumlah cluster ialah 3.
2. Berikutnya menentukan titik centroid awal dari setiap cluster yang dibentuk. Penentuan titik centroid bisa dipilih secara random dari data yang digunakan.

Tabel 8. Centroid awal cluster

Centroid	Nama	Prodi	Kelas	Kota/Kab	Prov	IPK	Sekolah
Cluster 1	Muhamad Annafi Anwar	57	10	12	55	2,96	31
Cluster 2	Achmad Arifin	90	10	21	33	3,94	41
Cluster 3	Rifky Addrian	90	10	60	22	3,59	51

3. Proses selanjutnya ialah menghitung jarak dari setiap data ke masing-masing centroid dengan mengimplementasikan rumus Euclidean Distance seperti dibawah ini:

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

Menggunakan rumus ini akan dihitung setiap data yang jaraknya paling dekat dengan titik pusat. Sebagai contoh dibawah ini bentuk perhitungan jarak dari setiap data ke titik pusat cluster pada data mahasiswa pertama ke pusat cluster pertama:

$$D(1,1) = \sqrt{(90 - 57)^2 + (10 - 10)^2 + (1 - 12)^2 + (11 - 55)^2 + (3,02 - 2,96)^2 + (41 - 31)^2} = 56,97$$

Lakukan perhitungan sampai dengan data terakhir terhadap titik pusat cluster pertama. Lanjutkan proses perhitungan jarak dari data mahasiswa pertama sampai data

mahasiswa terakhir ke titik pusat cluster kedua. Dan data mahasiswa pertama ke titik pusat cluster ketiga.

$$D(1,2) = \sqrt{(90 - 90)^2 + (10 - 10)^2 + (1 - 21)^2 + (11 - 33)^2 + (3,02 - 3,94)^2 + (41 - 41)^2} = 29,75$$

$$D(1,3) = \sqrt{(90 - 90)^2 + (10 - 10)^2 + (1 - 60)^2 + (11 - 22)^2 + (3,02 - 3,59)^2 + (41 - 51)^2} = 60,85$$

Tabel 9. Hasil perhitungan euclidean distance setiap data ke setiap cluster iterasi 1

No	Nama	Prodi	Kelas	Kota/Kab	Provinsi	IPK	Sekolah	Jarak Ke-			Jarak Terdekat
								C1	C2	C3	
1	MOHAMAD FARHAN HAFIDZUDIN	90	10	1	11	3,02	41	56,97	29,75	60,85	C2
2	FAIZ AFZA	90	10	2	99	3,94	31	55,91	69,40	98,45	C1
3	RISKA SULISTIANI	59	11	3	12	3,68	41	45,12	41,56	66,42	C2
4	ARDINI ANNISYAH AZZAHRA	23	10	3	12	2,89	31	55,55	73,18	90,77	C1
5	M HAQIQI ARNUN PANGESTU	90	10	3	12	3,47	51	58,47	29,41	57,87	C2
6	FERRY ALAMSYAH	59	10	4	55	2,75	31	8,25	42,84	74,74	C1
7	YENIKA SUGIARTO YUSUF	90	10	5	88	3,97	31	47,20	58,15	88,21	C1
8	YOLANDA PRATIWI	90	10	5	88	2,95	41	48,24	57,29	86,50	C1
9	NISMARA MAYA MANIKA	90	10	6	66	3,86	41	36,70	36,25	70,37	C2
10	FEBRIAN EKO HADI	23	11	7	86	3,77	41	47,37	86,57	107,21	C1
11	ALFIAN NUR KHASAN	90	11	8	66	3,54	41	36,43	35,48	68,85	C2
12	PRADANNI KRESNA MUKTI	59	11	9	77	3,62	41	24,46	55,16	81,78	C1
13	MUHAMMAD AL RAFLI	59	10	10	55	2,75	41	10,39	39,59	68,20	C1
14	MOCH AZZAM FALIH	90	10	10	55	3,69	51	38,65	26,55	59,91	C2
15	ANISA NUR FITRIANA	90	10	10	55	3,95	51	38,65	26,55	59,91	C2
16
386

4. Setelah jarak terdekat setiap data dihitung, selanjutnya ialah hitung kembali titik pusat cluster (centroid) yang baru berdasarkan masing-masing anggota setiap cluster.

$$C = \frac{\sum m}{n} \quad (2)$$

5. Hitung kembali jarak tiap-tiap data dengan centroid baru yang telah dihitung.

Tabel 10. Hasil perhitungan setiap data ke setiap cluster iterasi 2

No	Nama	Prodi	Kelas	Kota/Kab	Provinsi	IPK	Sekolah	Jarak Ke-			Jarak Terdekat
								C1	C2	C3	
1	MOHAMAD FARHAN HAFIDZUDIN	90	10	1	11	3,02	41	62,55	35,86	61,97	C2
2	FAIZ AFZA	90	10	2	99	3,94	31	69,41	61,57	96,80	C2
3	RISKA SULISTIANI	59	11	3	12	3,68	41	46,69	42,85	58,65	C2
4	ARDINI ANNISYAH AZZAHRA	23	10	3	12	2,89	31	53,40	71,35	75,77	C1
5	M HAQIQI ARNUN PANGESTU	90	10	3	12	3,47	51	62,21	36,07	61,03	C2
6	FERRY ALAMSYAH	59	10	4	55	2,75	31	27,93	34,96	65,35	C1
7	YENIKA SUGIARTO YUSUF	90	10	5	88	3,97	31	60,90	50,32	86,56	C2
8	YOLANDA PRATIWI	90	10	5	88	2,95	41	60,41	49,79	86,18	C2
9	NISMARA MAYA MANIKA	90	10	6	66	3,86	41	49,09	29,32	70,41	C2
10	FEBRIAN EKO HADI	23	11	7	86	3,77	41	49,73	78,24	95,13	C1
11	ALFIAN NUR KHASAN	90	11	8	66	3,54	41	48,21	28,41	68,92	C2
12	PRADANNI KRESNA MUKTI	59	11	9	77	3,62	41	34,79	46,25	74,56	C1
13	MUHAMMAD AL RAFLI	59	10	10	55	2,75	41	21,51	31,80	59,85	C1
14	MOCH AZZAM FALIH	90	10	10	55	3,69	51	46,40	21,83	62,19	C2
15	ANISA NUR FITRIANA	90	10	10	55	3,95	51	46,40	21,84	62,19	C2
16
386

6. Ketika hasil perhitungan antara centroid yang baru dengan yang lama sesuai atau tidak terdapat perubahan lagi maka iterasi dihentikan, sebaliknya ketika masih terjadi perubahan anggota cluster maka iterasi masih terus dilanjutkan.
7. Pada penelitian ini membutuhkan perhitungan sampai dengan iterasi ke 5 agar kondisi cluster tidak lagi berubah. Ketika kelompok data pada cluster tidak lagi berubah maka proses perhitungan dihentikan. Berikut hasil akhir dari perhitungan iterasi ke 5.

Tabel 11. Hasil perhitungan setiap data ke setiap cluster iterasi 5

No	Nama	Prodi	Kelas	Kota/Kab	Provinsi	IPK Sekolah	Jarak Ke-			Jarak Terdekat	
							C1	C2	C3		
1	MOHAMAD FARHAN HAFIZUDIN	90	10	1	11	3,02	41	61,31	44,40	61,57	C2
2	FAIZ AFZA	90	10	2	99	3,94	31	68,44	52,04	97,61	C2
3	RISKA SULISTIANI	59	11	3	12	3,68	41	45,49	52,74	56,87	C1
4	ARDINI ANNISYAH AZZAHRA	23	10	3	12	2,89	31	52,93	79,83	73,04	C1
5	M HAQIQI ARNUN PANGESTU	90	10	3	12	3,47	51	60,86	44,22	60,78	C2
6	FERRY ALAMSYAH	59	10	4	55	2,75	31	26,30	36,19	64,50	C1
7	YENIKA SUGIARTO YUSUF	90	10	5	88	3,97	31	59,89	40,89	87,35	C2
8	YOLANDA PRATIWI	90	10	5	88	2,95	41	59,22	40,10	87,05	C2
9	NISMARA MAYA MANIKA	90	10	6	66	3,86	41	47,68	20,80	71,10	C2
10	FEBRIAN EKO HADI	23	11	7	86	3,77	41	49,13	76,79	94,07	C1
11	ALFIAN NUR KHASAN	90	11	8	66	3,54	41	46,86	19,50	69,66	C2
12	PRADANNI KRESNA MUKTI	59	11	9	77	3,62	41	33,46	42,06	74,36	C1
13	MUHAMMAD AL RAFU	59	10	10	55	2,75	41	19,43	32,96	59,15	C1
14	MOCH AZZAM FALIH	90	10	10	55	3,69	51	44,86	16,13	62,92	C2
15	ANISA NUR FITRIANA	90	10	10	55	3,95	51	44,86	16,14	62,92	C2
16
386

HASIL DAN PEMBAHASAN

Hasil perhitungan berhenti sampai pada iterasi ke 5. Setelah melakukan iterasi sebanyak 5 kali diketahui bahwa titik pusat cluster (centroid) sudah tidak mengalami perubahan lagi dan tidak ada perpindahan data dari satu cluster ke cluster yang lain.

Berdasarkan hasil perhitungan diketahui pada cluster 1, terlihat bahwa karakteristik mahasiswa pada cluster ini berasal dari prodi Sistem Informasi dan Teknologi Informasi. Sedangkan untuk kelasnya didominasi dari kelas pagi dan daerah asalnya didominasi dari wilayah Jakarta Selatan dan Bogor. Cluster 1 untuk asal sekolah didominasi dari SMK dan SMA, dengan rata-rata IPK pada cluster 1 : 3,25.

Tabel 12. Hasil analisis clustering

Hasil Cluster 1	Hasil Cluster 2	Hasil Cluster 3
Cluster 1 terdiri dari 99 orang mahasiswa yang berasal dari program studi: SI: 41 TI: 36 DKV: 0 ARS: 22 Kab/Kota asal mahasiswa: Jaksel: 23 Bogor: 13 Provinsi asal mahasiswa: Aceh: 1 Lampung: 2 Banten: 7 DKI Jakarta: 33 Jambi: 1 Jawa Barat: 26 Jawa Tengah: 13 Jawa Timur: 4 NTT: 3 Yogyakarta: 1 Sulawesi Selatan: 1 Sulawesi Barat: 1 NTB: 6 Asal sekolah mahasiswa: SMK: 37 SMA: 23 MAN/Pesantren: 17 Rata-rata IPK pada Cluster 1 yakni 3,25	Cluster 2 terdiri dari 62 orang mahasiswa yang seluruhnya berasal dari program DKV Kab/Kota asal mahasiswa: Bogor: 9 Depok: 11 Jaksel: 10 Jakarta: 5 Lamongan: 2 Jakbar: 3 Bekasi: 2 Bandung: 2 Provinsi asal mahasiswa: Aceh: 2 DKI Jakarta: 18 Jawa Barat: 26 Jawa Tengah: 6 Jawa Timur: 5 Kalsel: 2 Kalim: 2 Asal sekolah mahasiswa: SMA: 19 SMK: 33 MAN/Pesantren: 10 Rata-rata IPK mahasiswa pada cluster 2 yakni 3,48	Cluster 3 terdiri dari 225 orang mahasiswa yang berasal dari program studi: SI: 55 TI: 59 ARS: 17 DKV: 97 Kab/Kota asal mahasiswa Lampung Barat: 2 Lebak: 5 Ogan Komering: 3 Serang: 2 Tangerang: 115 Tangsel: 88 Tasikmalaya: 3 Provinsi asal mahasiswa Lampung: 4 Bone: 2 Sumsel: 4 Banten: 209 Jawa Barat: 3 Sekolah asal mahasiswa: SMA: 77 SMK: 128 MAN/Pesantren: 20 Rata-rata IPK mahasiswa pada cluster 3 yakni 3,57

Hasil cluster 2 diatas dapat terlihat bahwa mahasiswa berasal dari program studi Desain Komunikasi Visual (DKV) dan kelasnya didominasi dari kelas pagi. Sedangkan daerah asal provinsinya berasal dari Jawa Tengah dan DKI Jakarta. Pada asal sekolah didominasi dari SMK, dengan rata-rata IPK untuk cluster 2 yakni 3,48.

Pada cluster 3 bisa kita lihat bahwa masiswa berasal didominasi dari program studi DKV, TI, SI dan kelas asalnya dari kelas pagi. Sedangkan daerah asal provinsinya didominasi dari Banten.

Untuk asal sekolah dominan dari SMK, dengan rata-rata IPK cluster 3 yakni 3,57.

Penulis juga melakukan analisis dengan menggunakan bahasa pemrograman python dengan script seperti berikut ini:

```

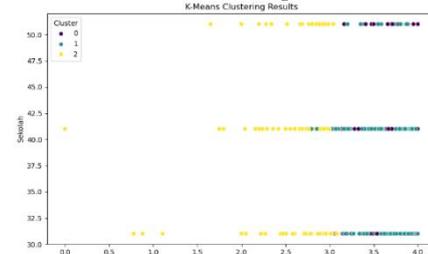
24 data_preprocessed = preprocessor.fit_transform(features)
25
26 # K-Means Clustering
27 n_clusters = 3 # Tentukan jumlah cluster yang diinginkan
28 kmeans = KMeans(n_clusters=n_clusters, random_state=12)
29 kmeans.fit(data_preprocessed)
30 clusters = kmeans.predict(data_preprocessed)
31
32 # Tambahkan hasil clustering ke data asli
33 data['cluster'] = clusters
34
35 # Visualisasi hasil clustering
36 plt.figure(figsize=(10, 6))
37 sns.scatterplot(data=data, x='IPK', y='Sekolah', hue='cluster', palette='viridis')
38 plt.title('K-Means Clustering Results')
39 plt.xlabel('IPK')
40 plt.ylabel('Sekolah')
41 plt.legend(title='cluster')
42 plt.show()
43
44 # Evaluasi hasil clustering
45 score = silhouette_score(data_preprocessed, clusters)
46 print(f'Silhouette Score: {score}')
47 print(f'Inertia: {kmeans.inertia_}')
    
```

Gambar 2. Script Python untuk Menentukan Cluster dan Visualisasi

Sehingga hasil dari algoritma clustering dengan nilai silhouette skor adalah 0.16487130938367947; nilai inertia ialah 886.0532411451791. Hasil visualisasi dari analisis ini menggunakan scatterplot seperti pada tampilan berikut ini:



Gambar 2. Hasil Nilai Silhouette dan Inertia Clustering



Gambar 3. Hasil Visualisasi Clustering

KESIMPULAN

Penelitian ini berhasil mengelompokkan data mahasiswa ITB Ahmad Dahlan dengan memakai algoritma K-Means Clustering. Tujuan utamanya yakni untuk mengidentifikasi karakterik dan pola mahasiswa yang ada di ITB Ahmad Dahlan mencakup prodi, IPK, asal sekolah, kota dan provinsi. Penelitian ini mengelompokkan mahasiswa ke dalam tiga cluster yang dapat memberikan tambahan wawasan dalam menetapkan strategi penerimaan mahasiswa baru. Hasil pengelompokkan yang diperoleh diketahui bahwa karakteristik mahasiswa cukup beragama atau berbeda-beda pada masing-masing cluster, misalnya cluster 1 didominasi oleh prodi A yang

berasal dari kota X. Adanya temuan-temuan tersebut dapat dipergunakan oleh ITB Ahmad Dahlan untuk menyusun strategi promosi yang paling cocok dan tepat sasaran, sehingga dapat meningkatkan jumlah mahasiswa baru.

REFERENSI

- Afiasari, N., Suarna, N., & Rahaningsi, N. (2023). Implementasi Data Mining Transaksi Penjualan Menggunakan Algoritma Clustering dengan Metode K-Means. *Jurnal SAINTEKOM*, 13(1), 100–110.
- Alhapizi, M. R., Nasir, M., & Effendy, I. (2020). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Promosi Mahasiswa Baru Universitas Bina Darma Palembang. *Journal of Software Engineering Ampera*, 1(1), 1–14.
- Dhewayani, F. N., Amelia, D., Alifah, D. N., Sari, B. N., & Jajuli, M. (2022). Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM. *Jurnal Teknologi Dan Informasi*, 12(1), 64–77.
- Haris Kurniawan, Sarjon Defit, & Sumijan. (2020). Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal. *Journal of Applied Computer Science and Technology*, 1(2), 80–89.
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56.
- Muzakir, A., & Wulandari, R. A. (2016). Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree. *Scientific Journal of Informatics*, 3(1), 19–26.
- Rafi Nahjan, M., 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.
- Sari, Y. P., Primajaya, A., & Irawan, A. S. Y. (2020). Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang. *INOVTEK Polbeng - Seri Informatika*, 5(2), 229.
- Sugianto, C. A., Rahayu, A. H., & Gusman, A. (2020). Algoritma K-Means untuk Pengelompokan Penyakit Pasien pada Puskesmas Cigugur Tengah. *Journal of Information Technology*, 2(2), 39–44.
- Sulistiyawati, A., & Supriyanto, E. (2021). Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan. *Jurnal Tekno Kompak*, 15(2), 25.
- Veronica, M., & Effendi, H. (2023). Clustering Tingkat Kedisiplinan Pegawai Pada Pengadilan Tinggi Palembang Menggunakan Algoritma K-Means. *Prosiding CORISINDO 2023*.
- Yudhistira, A., & Andika, R. (2023). Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *Journal of Artificial Intelligence and Technology Information (JAITI)*, 1(1), 20–28.
- Yusuf, D., & Sestri, E. (2020). Metode Decision Tree Dalam Klasifikasi Kredit Pada Nasabah PT Bank Perkreditan Rakyat (Studi Kasus : PT BPR Lubuk Raya Mandiri). *Jurnal Sistem Informasi (JUSIN)*, 1(1), 21–28.
- Yusuf, D., Sestri, E., & Razi, F. (2023). Implementasi Teknik Clustering Untuk Pengelompokan Mobil Bekas Berdasarkan Grade Pada Mobi Auto. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 6(2), 307.
- Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data. *Jurnal Portal Data*, 2(3).