

Klustering Data Mahasiswa Menggunakan Metode *K-Means* Sebagai Acuan dalam Penentuan Uang Kuliah Tunggal (UKT) Mahasiswa

Dwi Novia Prasetyanti^{1*}, Riyadi Purwanto², Cahya Vikasari³, Rostika Listyaningrum⁴

^{1, 2, 3, 4}Program Studi Teknik Informatika, Politeknik Negeri Cilacap

^{1, 2, 3, 4}Jln. Dr. Soetomo No.1 Karangcengis Sidakaya, Kabupaten Cilacap, 53212, Indonesia

E-mail: dnprasetyanti@pnc.ac.id¹, adidokbayu85@gmail.com², cahyavikasari@pnc.ac.id³, li_sa007@pnc.ac.id⁴

Abstrak

Penentuan kelompok UKT untuk mahasiswa baru menjadi hal penting dalam proses Penerimaan Mahasiswa Baru/PMB setelah proses seleksi mahasiswa baru. Penentuan kelompok UKT oleh panitia PMB di Politeknik Negeri Cilacap dilakukan satu persatu dengan melihat data perekonomian mahasiswa baru. Kondisi ini menjadi masalah tersendiri karena kenaikan kuota PMB di PNC, sehingga membutuhkan alternatif solusi yang dapat dijadikan salah satu acuan dalam penentuan kelompok UKT mahasiswa baru di PNC. Peneliti menggunakan clustering dengan fitur yang mewakili kondisi ekonomi mahasiswa baru dengan metode K-means untuk menyediakan solusi alternatif. Hasil dari penggunaan metode K-Means dalam clustering, menghasilkan nilai performa untuk jumlah *cluster* 8 sebesar 1669,283, dengan jumlah anggota *cluster* terbanyak pada cluster 4 sebesar 72 dari 275 data. Hasil pengujian Elbow method untuk menentukan jumlah cluster terbaik menghasilkan 4 cluster dengan nilai performa 2462,003.

Abstract

Determining Uang Kuliah Tunggal/UKT for new students is important in Penerimaan Mahasiswa Baru/PMB process after PMB selection process. The determination of UKT groups by The PMB committee at Politeknik Negeri Cilacap is carried out one by one by looking at the economic data of new students. This condition has become a special problem due to the increase in PMB quotas in the PNC, so it requires alternative solutions that can be used as one of the benchmarks in the determination of a new student UKT group in PNC. The researchers used clustering with features that represent the economic conditions of new students with the K-means method to provide alternative solutions. The result of using the K-Means method in clustering, yielding a performance value for the number of clusters 8 of 1669,283, with the highest number of cluster members in cluster members in cluster 4 being 72 out of 275 data. The Elbow method test results to determine the best number of clusters resulting in 4 cluster with a performance value of 2462,003.

Info Naskah:

Naskah masuk: 19 Juni 2024

Direvisi: 4 Juli 2024

Diterima: 23 Juli 2024

Keywords:

clustering;

K-Means;

uang kuliah tunggal.

*Penulis korespondensi:

Dwi Novia Prasetyanti

E-mail: dnprasetyanti@pnc.ac.id

1. Pendahuluan

Pelaksanaan Pendidikan, khususnya Pendidikan Tinggi, tidak terlepas dari pembiayaan yang harus dibayarkan oleh masing-masing mahasiswa yang digunakan untuk proses pembelajaran. Ketentuan ini berlaku baik bagi perguruan tinggi negeri/PTN maupun perguruan tinggi/PTS swasta. Pada PTN, sebagian biaya yang harus dibayarkan mahasiswa ditanggung oleh pemerintah. Besaran UKT dihitung dari Biaya Kuliah Tunggal/BKT PTN untuk setiap Program Studi. BKT merupakan keseluruhan biaya operasional penyelenggaraan Pendidikan untuk setiap tahun. Besaran UKT mahasiswa terbagi menjadi beberapa kelompok, dimana untuk penetapan kelompok UKT mahasiswa didasarkan pada pertimbangan kemampuan ekonomi mahasiswa baru.

Pemberlakuan UKT dilakukan pada semua PTN, tidak terkecuali di Politeknik Negeri Cilacap/PNC, yang memiliki beberapa program studi. Setiap program studi memiliki kebutuhan biaya yang berbeda-beda dalam penyelenggaraan Pendidikan. Sudah jelas bahwa persyaratan untuk program studi D3 Teknik Mesin dan Teknik Informatika berbeda. Persyaratan ini dapat dipengaruhi, antara lain, oleh jumlah bahan praktek yang dibutuhkan oleh masing-masing program studi. Perhitungan penentuan UKT pada setiap kelompok didasarkan dari BKT yang telah ditetapkan berdasarkan peraturan Menteri untuk setiap tahun[1].

Kuota Penerimaan Mahasiswa Baru/PMB di PNC setiap tahun mengalami kenaikan, dimana dalam hal ini, panitia harus melakukan penentuan UKT untuk setiap mahasiswa baru. Proses penentuan UKT dilakukan melalui rapat, dengan melakukan analisis data dan pertimbangan kondisi perekonomian untuk setiap mahasiswa. Kondisi tersebut menjadi permasalahan bagi panitia PMB, karena panitia harus menentukan UKT mahasiswa baru satu persatu, sementara setiap tahun terjadi penambahan jumlah mahasiswa baru yang diterima di PNC. Hasil akhir dari proses penentuan UKT adalah mahasiswa baru akan terbagi menjadi beberapa kelompok UKT, sesuai dengan kelompok UKT yang sudah ditetapkan untuk setiap program studi. Selain itu, penentuan UKT yang tidak tersistem, memiliki potensi terjadi perbedaan UKT untuk data yang sama ataupun mendekati.

Clustering adalah teknik statistik multivariat modern yang menggabungkan sekumpulan objek ke dalam kelas atau cluster sehingga objek dalam setiap cluster memiliki kemiripan satu sama lain untuk mempelajari kesamaan sample yang berbeda[2][3]. Berdasarkan masalah yang telah dijelaskan, pengelompokan/kelompokan dengan teknik tertentu dapat digunakan sebagai dasar untuk menentukan UKT mahasiswa baru. Untuk melakukan clustering, ada banyak metode yang dapat digunakan. Salah satunya adalah metode *K-Means*, yang merupakan salah satu algoritma *clustering unsupervised learning* yang digunakan untuk mengelompokkan data ke dalam berbagai kelompok atau partisipasi dengan tujuan untuk menggali informasi dari masing-masing kelompok atau partisipasi[4][5]. Algoritma *K-Means* banyak digunakan dalam proses clustering karena efisiensi dan sederhananya[6].

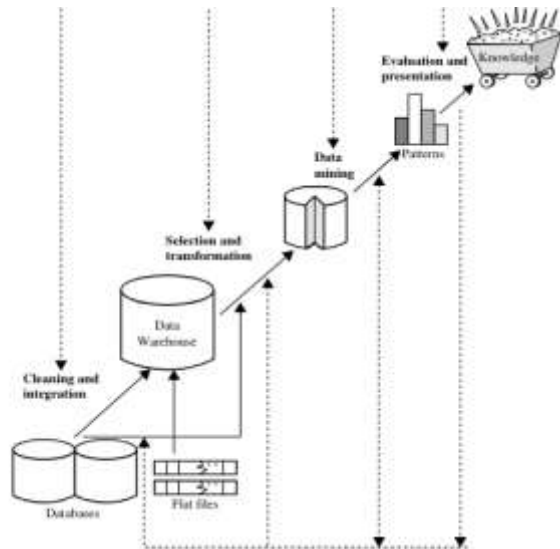
Untuk mengelompokkan data dengan metode *K-Means*, jarak antar data dengan masing-masing pusat kluster atau centroid dihitung. Pusat kluster yang paling dekat dengan data dikelompokkan dengan menggunakan jarak geometris.

Studi sebelumnya, seperti yang dilakukan oleh Fatkhudin dkk. (2023), membahas penerapan algoritma Clustering *K-Means* dalam pengelompokan mahasiswa studi kasus (Prodi Manajemen Informatika), menjadi salah satu acuan. Pada penelitian ini, kluster dibagi menjadi 4, 5 dan 6. Algoritma *K-Means* digunakan untuk menghasilkan lima kluster terbaik dengan nilai terkecil pada indeks Davies Bouldin. Hasil pengelompokan menunjukkan bahwa mahasiswa di kluster 2 dan kluster 3 harus memberikan perhatian khusus pada tugas akhir mereka agar mereka dapat menyelesaikan tugas akhir. Studi tambahan tentang penentuan UKT siswa dilakukan oleh Kurniawan et al. (2020) mengenai data mining menggunakan metode *clustering K-Means* untuk menentukan besaran uang kuliah tunggal. data dalam penelitian tersebut dikelompokkan menjadi lima kluster berdasarkan nilai UKT: kelompok pertama menerima 500.000 rupiah, kelompok kedua menerima 1.000.000 rupiah, kelompok tiga menerima 2.000.000 rupiah, kelompok empat menerima 3.000.000 rupiah, dan kelompok terakhir menerima 4.000.000 rupiah. Pengelompokan ini dilakukan berdasarkan pola yang mirip dengan kondisi ekonomi orang tua mahasiswa [7].

Penelitian saat ini mengumpulkan data menjadi 8 kluster, sesuai dengan kelompok UKT yang tercantum dalam Permendikbud Ristek [1]. Metode Clustering *K-Means* digunakan dalam penelitian ini. Pada penelitian ini, data penghasilan, pembayaran listrik, PDAM, dan PBB adalah data nominal pasti, bukan jangkauan, berbeda dengan penelitian yang dilakukan oleh Kurniawan et al. Selain itu, terdapat perbedaan dalam penggunaan beberapa variable. Penelitian ini memiliki manfaat berdasarkan masalah yang diuraikan di latar belakang masalah yaitu adalah untuk menawarkan referensi dan bantuan kepada pejabat terkait dalam menentukan UKT, dan untuk membuat penentuan UKT lebih seragam atau mengurangi perbedaan.

2. Metode Penelitian

Metode yang digunakan pada klustering ini adalah metode *K-Means* dengan proses mengikuti tahapan pada data Mining, dimana data mining merupakan proses penggalian informasi serta pola dari data yang berukuran besar[8]. Pengelompokan/*clustering* data merupakan salah satu bagian dari Data Mining. Sebagian besar orang menyebut data mining dengan istilah *knowledge discovery* atau biasa disebut dengan KDD[9]. Proses menemukan pola melalui pemrosesan data atau KDD ini melibatkan beberapa tahapan, seperti data cleaning (pembersihan) dan *integration* (integrasi data), *integration*, *selection and transformation* (pemilihan dan transformasi), data mining (pemrosesan data), *evaluation and presentation* (penilaian dan presentasi).



Gambar 1. Tahapan Knowledge discovery dari data mining

Penjelasan tentang masing-masing tahapan ini adalah sebagai berikut [9]:

- Data cleaning*, pada tahapan ini dilakukan pembersihan data mahasiswa baru dengan cara menghilangkan *noise* antara lain data-data yang tidak konsisten, data yang tidak valid atau tidak lengkap. Tahapan ini merupakan tahapan yang penting untuk menghasilkan data yang bermutu.
- Data integration*, dilakukan dengan menggabungkan beberapa sumber data, sehingga menjadi satu-kesatuan sumber data. Pada dunia sistem informasi, untuk proses *data cleaning* dan *integration* biasanya digabungkan menjadi satu pada tahap *preprocessing*. Data hasil dari *preprocessing* akan disimpan dalam data *warehouse*.
- Data Selection*, pada tahap ini, dilakukan pemilihan data yang akan digunakan dalam penelitian. Tidak semua atribut yang terdapat pada data mahasiswa baru digunakan pada penelitian ini. Atribut dari data mahasiswa baru yang digunakan dalam penelitian meliputi data pekerjaan ayah, penghasilan ayah, pekerjaan ibu, penghasilan ibu, pembayaran listrik, penmabayaran PBB, pembayaran PDAM, jumlah tanggungan, daerah asal, dan KIP/BSM. Jumlah data yang digunakan pada penelitian ini sebanyak 273 record.
- Data transformation*, data-data yang sudah dipilih pada tahap *selection*, akan dicek, apakah terdapat data yang memiliki tipe tidak sesuai dengan kebutuhan untuk melakukan klustering, yaitu semua data harus bertipe *numerik*. Pada data yang telah terpilih, terdapat data yang masih bertipe *polynomial*, yaitu data pekerjaan ayah, pekerjaan ibu, daerah asal dan KIP/BSM. Agar klustering dapat dilakukan, maka data tersebut harus diubah terlebih dahulu sehingga bertipe numerik.
- Data mining*, tahap ini merupakan tahapan untuk melakukan analisis terhadap data-data yang sudah disiapkan pada tahap *transformation*. Proses analisis yang dilakukan akan menghasilkan pengetahuan penting yang tersembunyi. Metode *K-Means* melakukan proses klusterisasi dengan menghitung jarak antara data dengan menggunakan persamaan Euclidean Distance.

$$d(x_i, \mu_j) = \sqrt{\sum (x_i - \mu_j)^2} \quad (1)$$

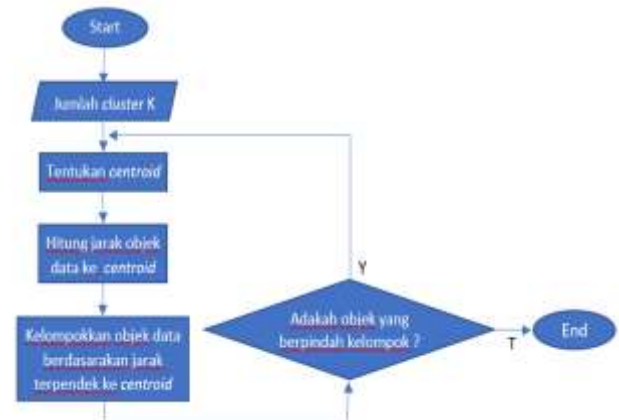
Keterangan :

x_i = data yang ke- i ,

μ_j = centroid yang ke- j .

Tahapan-tahapan proses pengelompokan data dengan menggunakan metode *K-Means* ini dapat dilihat pada Gambar 2.

+++



Gambar 2. Tahapan metode K-Means

- Pattern evaluation*, pada tahap ini, proses yang dilakukan adalah mengidentifikasi hasil dari proses data mining yang menghasilkan pola atau pattern. Selanjutnya pola yang didapatkan akan disimpan dalam *knowledge based* untuk selanjutnya dilakukan analisis.
- Knowledge presentation*, pada tahap ini digunakan teknik visualisasi dan representasi atau penyajian data yang terdapat pada *knowledge based* kepada pengguna yang melakukan *knowledge discover*.

3. Hasil dan Pembahasan

3.1 Data Cleaning

Penelitian ini menggunakan data primer, yaitu data PMB PNC tahun 2022. Karena masih berupa data mentah, data harus dibersihkan dari unsur-unsur yang tidak valid, atribut kosong, atau tidak memiliki nilai. Peneliti memastikan atau melakukan pengecekan terhadap masing-masing data mahasiswa baru. Terdapat beberapa data mahasiswa baru yang kosong pada salah satu variable atau beberapa variable, sehingga data-data tersebut tidak bisa digunakan dan harus dihapus dari data *sample*. Selain data mahasiswa baru yang kosong, terdapat data yang tidak valid, sehingga peneliti harus melakukan perbaikan data, antara lain menyamakan istilah yang digunakan dalam pengisian masing-masing variable yang digunakan.

3.2 Data integration

Peneliti melakukan penggabungan data dari beberapa sumber data yang didapatkan. Sumber data tersebut antara lain berasal dari data pendaftaran, data validasi yang dilakukan oleh tim Penerimaan Mahasiswa Baru, data seleksi, serta data pengumuman PMB.

Tabel 1. Data *sample*

NO	ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	0045633713	Swasta	2.500.000	IRT	500.000	0	0	0	3	Cilacap	Tidak
2	0041397032	Wiraswasta	1.000.000	Alm.		0	0	0	0	1	Cilacap
3	0041397588	BUMN	4.000.000	IRT	500.000	0	0	0	3	Bogor	Tidak
4	0041397845	Sopir	2.500.000	IRT		0	0	0	0	1	Bogor
5	0046339382	Buruh	8.532.600	IRT		0	85.426	0	0	3	Cilacap
6	0033453350	Buruh	1.000.000	IRT		0	0	0	0	1	Cilacap
7	0046156291	Buruh	1.500.000	IRT		0	18.608	0	3.920	2	Cilacap
8	0048921870	Swasta	1.800.000	IRT		0	55.248	0	23.220	1	Cilacap
9	0041195620	Pedagang	900.000	IRT	500.000	20.000	0	16.668	3	Cilacap	Tidak
...
...
...
272	0041397722	Buruh	1.500.000	Buruh	1.000.000	68.197	0	34.944	2	Cilacap	Tidak
273	0044574971	PNS	6.736.588	Guru	3.501.200	100.000	0	44.624	2	Cilacap	Tidak
274	0041397512	PNS Guru	5.120.500	IRT		0	20.000	79.500	30.014	3	Brebes
275	0042685308	Wiraswasta	900.000	IRT		0	108.326	0	73.760	2	Cilacap

Data yang didapatkan dibandingkan, untuk memperoleh data yang valid dan utuh. Data yang sudah valid yang digunakan dalam penelitian disimpan dalam folder khusus, sehingga mudah dalam mengakses data yang dibutuhkan

3.3 Data selection

Terdapat 103 variabel yang ada dalam data mahasiswa baru di PNC. Data pendaftaran PMB, nilai-nilai, kondisi perekonomian, dan biodata mahasiswa adalah beberapa variabel tersebut. Pada penelitian ini, data kondisi ekonomi mahasiswa baru termasuk variabel NISN/ID, pekerjaan ayah/V1, penghasilan ayah/V2, penghasilan ibu/V4, listrik/V5, PDAM/V6, PBB/V7, jumlah tanggungan/V8, kota asal/V9, dan KIP/BSM/V10. Peneliti menggunakan data kondisi ekonomi karena sesuai dengan tujuan peneliti: mengumpulkan data mahasiswa sebagai acuan untuk menentukan UKT mahasiswa baru; data penghasilan orang tua (ayah dan ibu) dipisahkan untuk memberikan detail lebih lanjut, memperlihatkan mahasiswa yang tidak memiliki salah satu orang tua atau keduanya. Selain itu, data asal kota, yang menunjukkan bagaimana perbedaan asal kota berdampak pada kebutuhan mahasiswa baru. Tabel 1 menunjukkan jumlah data yang diperoleh, yaitu berjumlah 275 record.

3.4 Data transformation

Peneliti melakukan *transformasi*, terhadap data dengan tipe *polynomial*, menjadi *numerik*, agar dapat dilakukan perhitungan jarak untuk setiap data. Dengan pengecualian variabel tanggungan yang sudah berbentuk numerik, semua variabel dalam data mahasiswa yang baru dikumpulkan memiliki tipe *polynomial*. Oleh karena itu, harus dilakukan *transformasi* sehingga semua variabel yang terdapat pada data mahasiswa baru yang digunakan berbentuk *numerik*. Hasil *transformasi* seperti pada Tabel 2.

Tabel 2. Data numerik

NO	ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
1	0045633713	0	0	0	0	0	0	0	3	0	0
2	0041397032	1	1	1	1	0	0	0	1	0	0
3	0041397588	2	2	0	0	0	0	0	3	1	0
4	0041397845	3	0	0	1	0	0	0	1	1	0
5	0046339382	4	3	0	1	1	0	0	3	0	0
...
...
...
274	0041397512	55	112	0	1	4	63	196	3	27	0
275	0042685308	1	6	0	1	220	0	197	2	0	0

3.5 Data mining

Pada proses *data mining* atau sering disebut sebagai KDD yang merupakan langkah penting dalam menemukan pengetahuan [10]. Proses clustering yang dilakukan dengan algoritma metode K-Means adalah sebagai berikut [11]:

- Tentukan jumlah cluster
- Tentukan pusat *cluster* atau *centroid*, penentuan pusat cluster biasanya dilakukan secara acak..
- Hitung jarak antara masing-masing data dengan pusat *cluster*, dengan menggunakan jarak eukledian (1).

Perhitungan jarak antara setiap data dengan pusat cluster (dalam penelitian ini, ada delapan pusat cluster), berdasarkan level UKT PNC saat ini. Data 1 s/d 8 diidentifikasi sebagai centroid 1 s/d 8. Perhitungan jarak data dengan pusat cluster pada perulangan/iterasi 1 adalah sebagai berikut:

$$d_{1,1} = \sqrt{((0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (0 - 0)^2)}$$

$$\begin{aligned}
 & ((0-0)^2 + (3-3)^2 + (0-0)^2 + (0-0)^2) = \sqrt{0} = 0 \\
 d_{1,2} &= \sqrt{((0-1)^2 + (0-1)^2 + (0-1)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (3-1)^2 + (0-0)^2 + (0-0)^2)} = \sqrt{8} = 2,82843 \\
 d_{1,3} &= \sqrt{((0-2)^2 + (0-2)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (3-3)^2 + (0-1)^2 + (0-0)^2)} = \sqrt{9} = 3,00000 \\
 d_{1,4} &= \sqrt{((0-3)^2 + (0-0)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (3-1)^2 + (0-1)^2 + (0-0)^2)} = \sqrt{15} = 3,87298 \\
 d_{1,5} &= \sqrt{((0-4)^2 + (0-3)^2 + (0-0)^2 + (0-1)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (3-3)^2 + (0-0)^2 + (0-0)^2)} = \sqrt{27} = 5,19615 \\
 d_{1,6} &= \sqrt{((0-4)^2 + (0-1)^2 + (0-0)^2 + (0-1)^2 + (0-0)^2 + (0-0)^2 + (0-0)^2 + (3-1)^2 + (0-0)^2 + (0-0)^2)} = \sqrt{22} = 4,690416 \\
 d_{1,7} &= \sqrt{((0-5)^2 + (0-4)^2 + (0-0)^2 + (0-1)^2 + (0-2)^2 + (0-0)^2 + (0-1)^2 + (3-2)^2 + (0-0)^2 + (0-0)^2)} = \sqrt{48} = 6,928203 \\
 d_{1,8} &= \sqrt{((0-0)^2 + (0-5)^2 + (0-0)^2 + (0-1)^2 + (0-3)^2 + (0-0)^2 + (0-2)^2 + (3-1)^2 + (0-0)^2 + (0-1)^2)} = \sqrt{44} = 6,63325
 \end{aligned}$$

Berdasarkan hasil perhitungan jarak data-1 dengan centroid 1 s/d 8, diketahui bahwa jarak terkecil berada pada centroid 1. Perhitungan jarak data dengan masing-masing centroid dilakukan untuk semua data, untuk setiap iterasi. Jika hasil iterasi sebelumnya sama, proses akan dihentikan[12]. Adapun statistic untuk masing-masing atribut yang digunakan terdapat pada Tabel 3.

Table 3. Statistik setiap atribut

Atribut	Min	Max	Average	Deviation
V1	0	61	12.385	14.507
V2	0	112	28.607	32.640
V3	0	30	3.273	6.468
V4	0	34	4.240	6.948
V5	0	220	93.989	68.696
V6	0	63	7.524	16.068
V7	0	197	72.491	64.979
V8	1	8	3.309	0.975
V9	0	27	1.735	4.979
V10	0	1	0.062	0.241

Table 4. Pengelompokan data

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	Cluster
0	0	0	0	0	0	0	3	0	0	Cluster 4
1	1	1	1	0	0	0	1	0	0	Cluster 5
2	2	0	0	0	0	0	3	1	0	Cluster 6
3	0	0	1	0	0	0	1	1	0	Cluster 7
4	3	0	1	1	0	0	3	0	0	Cluster 8
...
...
...
55	112	0	1	4	63	196	3	27	0	Cluster 7
1	6	0	1	220	0	197	2	0	0	Cluster 3

Mengelompokkan data berdasarkan jarak terpendek masing-masing data dengan setiap centroid. Pengelompokan data didasarkan kemiripan masing-masing data, sehingga dalam satu cluster memiliki kemiripan yang tinggi [13][14][15]. Hasil pengelompokan/clustering sebagaimana terdapat dalam Tabel 4. Cluster 0-7 terdiri dari delapan cluster.

3.6 Pattern evaluation

Berdasarkan hasil pada tahapan data mining, maka terbentuk pola/pattern yang membagi semua data pada setiap kelompok/cluster. Pola tersebut disimpan dalam knowledge base, sehingga jika terdapat data baru, maka dapat ditentukan data tersebut akan masuk ke dalam cluster berapa. Gambar 3 berikut menunjukkan jumlah data pada masing-masing cluster.

Cluster Model

```

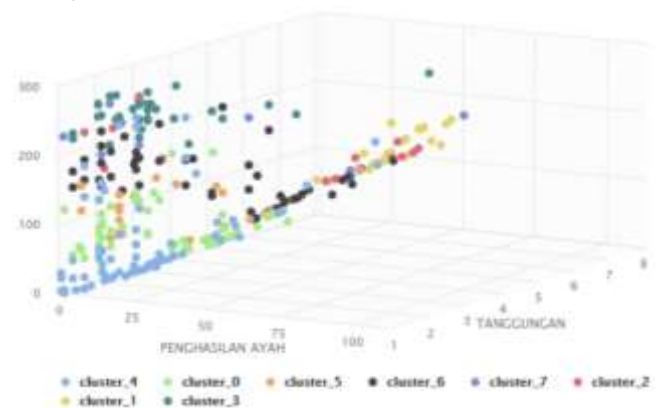
Cluster 0: 50 items
Cluster 1: 23 items
Cluster 2: 15 items
Cluster 3: 34 items
Cluster 4: 72 items
Cluster 5: 22 items
Cluster 6: 49 items
Cluster 7: 10 items
Total number of items: 275
    
```

Gambar 3. Cluster model

Jumlah data yang digunakan dalam penelitian ini adalah 275 data, dan masing-masing dikelompokkan berdasarkan jarak terdekat dengan pusat cluster.

3.7 Knowledge representation

Pada langkah ini, penyajian knowledge based dilakukan melalui visualisasi. Ini dilakukan dengan menggunakan grafik 3D, seperti yang ditunjukkan pada Gambar 4. Pada grafik tersebut, digunakan atribut penghasilan ayah sebagai x-axis column dan atribut tanggungan sebagai value column.



Gambar 4. Visualisasi knowledge based

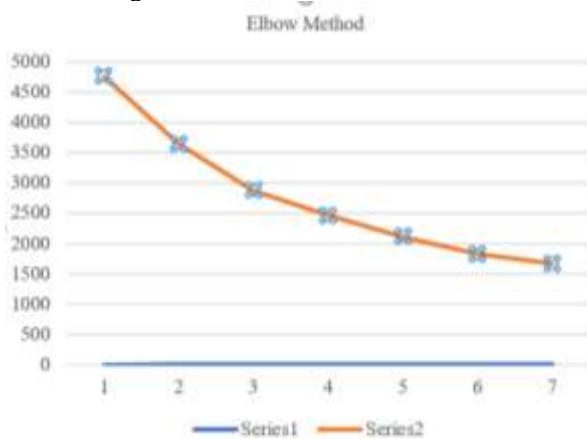
3.8 Pengujian

Untuk mengetahui seberapa baik model yang digunakan, pengujian dilakukan. Pada fase pengujian ini, berbagai cluster berbeda digunakan untuk menguji kinerja dan performa. Tabel 5 menunjukkan hasil pengujian performa dengan menggunakan jarak rata-rata antara centroid untuk setiap *cluster*.

Tabel 5. Pengujian *performa*

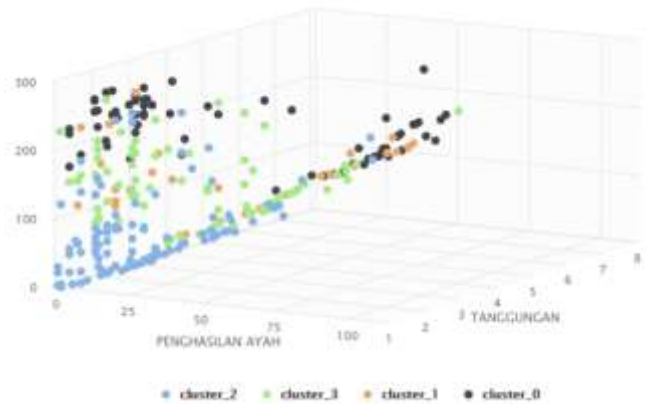
k	Avg. Centroid distance
2	4.760.096
3	3.648.675
4	2.873.882
5	2.462.003
6	2.114.789
7	1.828.224
8	1.669.283

Untuk menemukan nilai k (jumlah cluster) terbaik, metode elbow biasanya digambarkan sebagai grafik biaya untuk setiap nilai k yang membantu siku [16]. Metode *Elbow* digunakan untuk menentukan jumlah cluster yang ideal [17]. Gambar 5 menunjukkan implementasi metode Elbow dalam grafik.



Gambar 5. Grafik *elbow method*

Berdasarkan grafik tersebut, bagian yang membentuk sudut paling siku, berada pada $k = 4$. Kondisi tersebut menunjukkan bahwa jumlah *cluster* paling optimal untuk pengelompokan data mahasiswa baru adalah 4 *cluster*, dengan sebaran data sebagaimana terdapat pada Gambar 5. Nilai *performa* untuk jumlah *cluster* 4 sebesar 2462,003.



Gambar 5. Pengelompokan dengan $k=4$

4. Kesimpulan

Berdasarkan hasil penelitian, jumlah cluster atau pengelompokan data disesuaikan dengan tingkat UKT mahasiswa, yang berjumlah delapan cluster. Jumlah data terbanyak terdapat pada *cluster* 4, yaitu 72 data. Hasil pengujian yang dilakukan, *performa* yang didapatkan untuk jumlah *cluster* 8 sebesar 1669,283. *Elbow method* cocok digunakan untuk melakukan pengujian terhadap jumlah *cluster* optimum. Berdasarkan hasil pengujian dengan *elbow method*, menghasilkan jumlah *cluster* terbaik yaitu 4 *cluster*. Guna mengetahui konsistensi dari jumlah *cluster* terbaik, maka pada penelitian berikutnya perlu dilakukan pengujian dengan menggunakan metode yang berbeda.

Ucapan Terimakasih

Terimakasih kepada Politeknik Negeri Cilacap, khususnya Bagian Akademik dan Kemahasiswaan, yang telah menyediakan data yang diperlukan dalam penelitian. Saya Ucapan terima kasih kepada Politeknik Negeri Cilacap yang telah menyediakan fasilitas penelitian yang diperlukan.

Daftar Pustaka

- [1] Kemendikbud RI, "Peraturan Menteri Pendidikan Dan Kebudayaan Republik Indonesia Nomor 25 Tahun 2020," *Kementeri. Pendidik. dan Kebud. RI*, pp. 1–76, 2020.
- [2] and B. W. Y. Dai, G. Yuan, Z. Yang, "K-Modes Clustering Algorithm Based on Weighted Overlap Distance and Its Application in Intrusion Detection," *Sci Progr.*, vol. 2021, 2021, doi: 10.1155/2021/9972589.
- [3] A. Wirawan and D. Prasetyawan, "Analisis cluster data latar belakang ekonomi mahasiswa untuk rekomendasi penentuan uang kuliah tunggal dengan model K-Modes," *INFOTECH J. Inform. Teknol.*, vol. 4, no. 2, pp. 234–246, 2023, doi: 10.37373/infotech.v4i2.898.
- [4] Z. Nabila, A. R. Isnain, Permata, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [5] N. Dwitri, J. A. Tampubolon, S. Prayoga, F. I. R.H Zer, and D. Hartama, "Penerapan Algoritma K-Means Dalam Menentukan Tingkat Penyebaran Pandemi Covid-19 Di Indonesia," *J. Teknol. Inf.*, vol. 4, no. 1, pp. 128–132, 2020, doi: 10.36294/jurti.v4i1.1266.

- [6] A. Fatkhudin, A. Khambali, F. A. Artanto, and N. A. P. Zade, "Implementasi Algoritma Clustering K-Means Dalam Pengelompokan Mahasiswa Studi Kasus (Prodi Manajemen Informatika)," *J. Minfo Polgan*, vol. 12, no. 2, pp. 777–783, 2023, [Online]. Available: <https://jurnal.polgan.ac.id/index.php/jmp/article/view/12494/1682>
- [7] 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.
- [8] R. S. Wahono, *Data Mining Data mining*, vol. 2, no. January 2013. 2023. [Online]. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part
- [9] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [10] D. N. Anjar Wanto, Muhammad Noor Hasan Siregar, Agus Perdana Windarto, Dedy Hartama, Ni Luh Wiwik Sri Rahayu Ginantra, M. R. L. S. V. D. Edi Surya Negara, and C. Prianto, *Data Mining Algoritma dan Implementasi*, 1st ed. 2020.
- [11] and M. D. D. Tambun, S. Fauziah, "Pengelompokan Penerimaan Mahasiswa Baru Dengan Algoritma K-Means Untuk Meningkatkan Potensi Pemasaran," *Bull. Inf. Technol.*, vol. 4, pp. 294–298, 2023, doi: 10.47065/bit.v3i1.
- [12] M. R. Kusnaldi, T. Gulo, and S. Aripin, "Penerapan Normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 330–338, 2022, doi: 10.47065/josyc.v3i4.2112.
- [13] and Y. Y. S. Dewi, S. Defit, "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.
- [14] Narwati, "Pengelompokan Mahasiswa Menggunakan Algoritma K-Means," *J. Din. Informormatik*, vol. 2, pp. 1–7, 2010.
- [15] D. Gianyar, "PENERAPAN ALGORITMA K-MEANS CLUSTERING," vol. 8, no. 2, pp. 1543–1547, 2024.
- [16] N. P. M. N. D. and I. B. G. Dwidsamara, "Implementation of K-Modes Algorithm for Clustering of Stress Causes in University Students," *J. Elektron. Ilmu Komput. Udayana*, vol. 9 No. 3, pp. 419–427, 2021.
- [17] A. P. Riani, A. Voutama, and T. Ridwan, "Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD," vol. 6, pp. 164–172, 2023.