

Cacah Klaster pada Klasterisasi dengan Algoritma K-Means Menggunakan Silhouette Coeficient dan Elbow Method

Mohammad Guntara¹, Nafisatul Lutfi²

^{1,2}Informatika, Universitas Teknologi Digital Indonesia
Jl Raya Janti Majapahit No.143, Yogyakarta,
Indonesia

¹guntara@utdi.ac.id

²nafisatullutfi@utdi.ac.id

Disubmit: 07-07-23; diterima: 31-07-23; dipublikasikan: 08-08-23

Cara mengutip:

M. Guntara, et.al., 2023, " Cacah Klaster pada Klasterisasi dengan Algoritma K-Means Menggunakan Silhouette Coeficient dan Elbow Method", JuTI "Jurnal Teknologi Informasi", Vol. 2, No. 1, pp.43 – 52, DOI: 10.26798/juti.v2i1.944

Ringkasan

Mahasiswa dropout (DO) atau undur diri merupakan masalah yang perlu ditangani sedini mungkin. Banyaknya mahasiswa dropout (DO) akan menjadikan kualitas kinerja suatu perguruan tinggi akan menurun, karena akan berdampak pada apresiasi masyarakat terhadap sistem pendidikan di perguruan tinggi. Untuk mengetahui terjadinya mahasiswa DO diperlukan klasterisasi, yang dalam hal ini akan digunakan Algoritma K-Means. Untuk menentukan cacah klaster yang paling baik (optimal) digunakan 2 metode yakni Elbow Method dan Silhouette Coefficient penggunaan 2 metode ini sekalgus untuk menguji sejauh mana kesamaan hasil penentuan cacah klasternya. Data training yang akan dibuat kelas (label) berjenis unsupervised learning dengan atribut: NIM, jumlah semester tidak aktif, jumlah SKS, IPK, angkatan dan jumlah record (mahasiswa) 955 mahasiswa. Berdasarkan data training tersebut kemudian menggunakan Elbow Method dan Silhouette dihasilkan penentuan cacah klaster optimal yang sama yakni $K=2$. Berdasarkan nilai k ini kemudian dilakukan proses klasterisasi menggunakan algoritma K-Means dan menghasilkan kelas Cluster_0 dan Cluster_1. Dari sisi jumlah klaster untuk Cluster_0 terdapat 85,1% (813 mahasiswa) dan Cluster_1 terdapat 14,9% (142 mahasiswa). Cluster_0 ini didominasi mahasiswa DO karena $IPK < 2,00$, jumlah sks yang diambil < 90 sks, dan tidak aktif kuliah (cuti) ≥ 4 semester.

Kata kunci: Silhouette Coefficient, Elbow Method, K-Means, dropout

Abstract

A student dropping out (DO) or resigning is a problem that needs to be addressed as early as possible. The large number of dropout students (DO) will make the quality of the performance of a tertiary institution decrease, besides that as much as possible to reduce the cases of dropout students because it will have an impact on people's appreciation of the education system in tertiary institutions. To find out the occurrence of DO students clustering is needed. In this case, the clustering process uses the K-Means Algorithm. A good cluster has a high degree of separation. To determine the best (optimal) cluster count, 2 methods were used, namely the Elbow Method and The Silhouette Coefficient uses these 2 methods at the same time to test the extent to which the results of determining the cluster count are similar. The training data that will be made into classes/labels is of the unsupervised learning type with the attributes: NIM, number of inactive semesters, number of credits, GPA, batch, and number of records (students) of 836 students. Based on the training data then using the Elbow Method and Silhouette resulted in determining the same optimal cluster count, namely $k = 2$. Based on this k value, the clustering process is then carried out using the K-Means

algorithm and produces Cluster_0 and Cluster_1 classes. In terms of the number of clusters for Cluster_0, there are 755 students and for Cluster_1 there are 85 students. Cluster_0 is dominated by DO students because of a GPA <2.00, the number of credits taken <90 credits, and > 4 semesters of inactivity.

Kata kunci: Silhouette Coefficient, Elbow Method, K-Means, dropout

1. Pendahuluan

Klasterisasi adalah teknik pengelompokan record dalam database berdasarkan kondisi tertentu. Konsep dasar *clustering* adalah mengelompokkan beberapa objek ke dalam klaster, dimana klaster yang baik adalah klaster yang memiliki tingkat kemiripan yang tinggi antara objek-objek dalam klaster tersebut dan tingkat ketidaksamaan yang tinggi dengan objek klaster lainnya[1].

K-Means *Clustering* merupakan teknik dalam klaster data yang sangat terkenal dalam kecepatannya dalam mengklasterkan data. Akan tetapi metode ini memiliki kelemahan yakni dalam memproses data berdimensi banyak., khususnya untuk data yang bersifat bersifat non-linierly separable [2]. Undur diri adalah keadaan dimana seorang mahasiswa dinyatakan sudah tidak terdaftar sebagai mahasiswa . Proses undur diri ini dilakukan setelah melalui berbagai evaluasi yang dilakukan secara komprehensif [3].

Mahasiswa undur diri di STMIK AKAKOM berkisar 30% dari mahasiswa yang terdaftar pada satu angkatan [4]. Keadaan ini perlu untuk dicermati mengingat tingginya mahasiswa yang undur diri. Untuk mengambil langkah strategis agar mahasiswa undur diri dapat dikurangi diperlukan analisis dengan cara membuat klasterisasi berdasarkan data dari mahasiswa yang di akumulasi beberapa periode. Klasterisasi ini diperlukan untuk mengidentifikasi kecenderungan mahasiswa undur diri ini apakah dikarenakan faktor indeks prestasi, faktor jumlah SKS yang sudah diambil (masa studi), atau faktor lainnya yang tidak memenuhi syarat. Proses klasterisasi mahasiswa undur diri dilakukan dengan menggunakan metode K-Means.

Untuk memperoleh cacah klaster yang optimal digunakan Silhouette Coefficient methode dan Elbow methode. Hasil klasterisasi ini selanjutnya dianalisis dan digunakan untuk pengambilan keputusan dalam upaya pencegahan mahasiswa undur diri.

2. Metode Penelitian

2.1. Persiapan data dan transfer data

Dataset yang digunakan pada riset merupakan data mentah (raw data) yang diambil dari lampiran dalam format spreadsheet untuk mahasiswa yang dinyatakan *dropout* selama kurun waktu 5 tahun. Untuk dapat diolah dengan metode k-Means diperlukan tahapan sebagai berikut.

2.1.1. Pengambilan data

Data mahasiswa *dropout* diambil dari daftar mahasiswa yang dinyatakan undur diri [4] dan melalui proses sebagai berikut.

1. Normalisasi Data.

Normalisasi data dilakukan dengan memilih atribut yang digunakan pada proses klasterisasi. Atribut yang digunakan adalah : NIM, IPK, Jumlah SKS, angkatan, jumlah semester tidak aktif

2. Validasi Data.

Data dikumpulkan dari berkas-berkas yang berbeda sehingga dimungkinkan terjadinya ketidakseragaman urutan atribut, selanjutnya urutan atribut dibuat standar.

3. Penggabungan Data.

Data yang sudah valid dari masing-masing berkas digabungkan menjadi 1 berkas yang siap untuk diolah.

2.1.2. Persiapan dan pelaksanaan proses klasterisasi

1. Transformasi data dalam format spreadsheet ke format CSV (*comma separator value*)
2. Mengeksekusi data training menggunakan program Python dengan IDE Jupyter Notebook
3. Menganalisis hasil klasterisasi

2.2. Alat dan bahan yang digunakan

2.2.1. Alat yang digunakan

1. Piranti Lunak pemrograman Python dengan IDE Jupyter Notebook.
2. Piranti lunak spreadsheet, digunakan untuk menganalisis hasil klasterisasi menggunakan statistik deskriptif

2.2.2. Bahan

Sebagai data training digunakan data mahasiswa *dropout* yang dikemas dalam berkas dengan format Excel.

2.3. Algoritma Proses Klasterisasi (menggunakan bahasa pemrograman Python)

Algoritma proses klasterisasi ini terdiri atas 2 bagian, algoritma secara umum dan algoritma secara khusus dari 3 metode yang digunakan (Elbow Method, Silhouette Coefficient, dan K-Means)

2.3.1. Algoritma untuk proses klasterisasi

Silhouette Coefficient digunakan untuk melihat kualitas dan kekuatan klaster, seberapa baik suatu objek ditempatkan dalam suatu klaster. Metode ini merupakan gabungan dari metode cohesion dan separation [?], dengan tahapan perhitungan Silhouette Coefficient adalah sebagai berikut.

1. Buka IDE untuk Python
2. Import library : pandas, numpy,matplotlib,seaborn,sklearn, yellowbrick, KelbowVisualizer
3. Baca berkas data DO dalam format CSV
4. Eliminasi kolom yang tidak digunakan dalam proses
5. Mengubah *dataframe* menjadi *array*
6. Penskalaan *dataset*
7. Membuat grafik Elbow Method
8. Membuat grafik Silhouetten Coefficient
9. Membuat grafik/*plotting* grafik scatter ipk vs sks
10. Menentukan cacah klater terbaik berdasarkan 2 grafik diatas
11. Proses klasterisasi menggunakan algoritma K-Means
12. Ekspor hasil klasterisasi ke format *spreadsheet*
13. Analisis hasil menggunakan statistik deskriptif.

2.3.2. Elbow Method

Metode Elbow Metode dan analisis ini digunakan untuk pemilihan jumlah klaster atau kelompok yang optimal. Berikut disajikan algoritma Elbow dalam menentukan jumlah kelompok yang terbentuk, yaitu berdasarkan pada *sum of square error-SSE* [5]

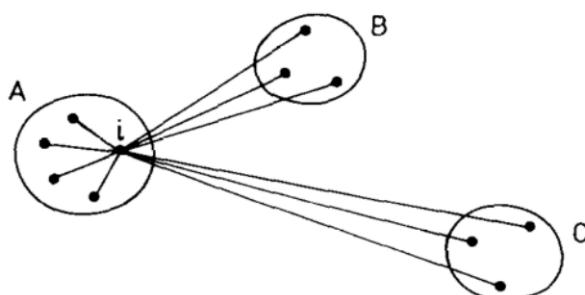
$$SSE = \sum_{k=1}^k \sum_{x_i \in S_k} ||X_i - C_k||^2 \quad (1)$$

Dimana k adalah banyaknya kelompok yang digunakan pada algoritma K-Means, X_i adalah jumlah data dan C_k adalah banyaknya klaster pada klaster ke K. Berikut ini tahapan algoritma metode Elbow dalam menentukan nilai K pada K-Means [6]:

1. Menginisialisasi awal nilai k ;
2. Menaikkan nilai k ;
3. Menghitung hasil *sum of square error* dari tiap nilai k ;
4. Analisis hasil *sum of square error* dari nilai k yang mengalami penurunan secara drastis ;
5. Cari dan tetapkan nilai k yang berbentuk siku.

2.3.3. Algoritma Silhouette Coefficient

Silhouette Coefficient digunakan untuk melihat kualitas dan kekuatan klaster, seberapa baik suatu objek ditempatkan dalam suatu klaster. Metode ini merupakan gabungan dari metode *cohesion* dan *separation* [7]. Ilustrasi Silhouette Coefficient terdapat pada Gambar 1 [8].



Gambar 1. Ilustrasi Proses Penentuan Silhouette Score

Misalkan didefinisikan $s(i)$ sebagai kasus *separation* (ketidaksamaan), dimana i dalam kumpulan data, dan dilambangkan dengan A cluster yang telah ditetapkan. Ketika klaster A berisi objek lain selain i, maka dapat dihitung $(i) =$ perbedaan rata-rata i untuk semua data lain dari A. Seperti terlihat pada Gambar 1 maka dapat diketahui panjang rata-rata semua data dalam klaster A. Untuk setiap klaster C yang berbeda dari A, dapat dihitung $d(i, C) =$ perbedaan rata-rata i untuk semua objek dari C.

Adapun persamaan untuk menghitung Silhouette Coefficient adalah :

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (2)$$

dengan $a(i)$ adalah rata-rata jarak antara entitas i ke entitas lain dalam klaster, dan $b(i)$ adalah rata-rata jarak minimum ke entitas di klaster lain.

Algoritma Silhouette Coefficient sebagai berikut [9].

1. Mencari jarak rerata di klaster yang sama, diasumsikan data ke-i berada di klaster A. Rumus dari $a(i)$ dituliskan dalam persamaan

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (3)$$

di mana, A = banyaknya data di klaster A

2. Menghitung nilai $b(i)$ yang merupakan nilai minimum dari jarak rata-rata data ke-i dengan semua data di klaster berbeda. Sekarang, mari asumsikan klaster berbeda selain A dengan klaster C. Maka, perhitungan jarak rata-rata data ke-i dengan semua data di klaster C ditulis sebagai berikut:

$$d(i, C) := \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (4)$$

di mana $C =$ banyaknya data di klaster C.

3. Memilih nilai jarak paling minimum sebagai nilai $b(i)$, dengan persamaan

$$b(i) := \min_{C \neq A} d(i, j) \quad (5)$$

Jika klaster B memiliki nilai jarak minimum, maka $d(i, B) = b(i)$ yang disebut sebagai tetangga dari data ke-i dan merupakan klaster terbaik kedua untuk data ke-i setelah klaster A.

4. Menghitung Silhouette coefficient sesuai persamaan (2), dengan nilai $s(i)$ berada antara -1 dan 1. Bila $s(i)$ sama atau mendekati 1, maka separasi data ke-i dikatakan baik, bila antara 0 dan 1 maka separasi data ke-i antara klaster A dan B, dan bila $s(i)$ sama atau mendekati -1 maka separasi kategori lemah

2.3.4. Algoritma untuk K-Means

Clustering termasuk dalam klasifikasi unsupervised classification. Pengertian *Clustering* adalah proses mengelompokkan atau penggolongan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimumkan kesamaan antar kelas/klaster. *Clustering* dalam textitdata mining berguna untuk menemukan pola distribusi di dalam sebuah dataset yang berguna untuk proses analisa data. Kesamaan objek biasanya diperoleh dari kedekatan nilai-nilai atribut yang menjelaskan objek-objek data, sedangkan objek-objek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multi dimensi [6].

Algoritma K-Means *clustering* merupakan salah satu algoritma pengelompokan data dengan sistem partisi, dengan aturan sebagai berikut [5] : jumlah klaster atau k harus diinisialisasikan terlebih dahulu, atribut bersifat numerik , keterbatasan atribut, kompleksitas algoritma linear (n). Algoritma K-Means termasuk dalam metode *non hierarchical* yang mempartisi data ke dalam bentuk satu atau lebih klaster, sehingga data yang mempunyai karakteristik yang sama dikelompokkan dalam satu klaster yang sama dan data yang memiliki karakteristik berbeda dikelompokkan ke dalam klaster lain. Algoritma ini merupakan algoritma yang paling umum digunakan karena mudah untuk diimplementasikan. Adapun kelemahannya adalah algoritma ini sangat sensitif terhadap inisialisasi klaster.

Berikut ini urutan Algoritma K-Means [10]

1. Menentukan kembali pusat klaster dengan anggota klaster yang baru.
2. Ulangi lagi setiap objek memakai pusat klaster yang baru. Jika pusat klaster sudah tidak berubah lagi, maka proses selesai.
3. Kembali ke langkah a sampai pusat klaster tidak mengalami perubahan lagi. Berikut adalah rumus untuk perhitungan jarak terdekat:

$$d_{(x,y)} = \|x - y\|^2 = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (6)$$

3. Hasil dan Pembahasan

3.1. Format Sumber data.

Sumber data berupa berkas Excel yang merupakan data mahasiswa DO dari tahun 2015 sampai tahun 2019 adalah sebagai berikut. Berdasarkan sumber data di atas yang telah difilter, diformat, dan divalidasi menjadi data yang siap diolah, maka akan tampak pada Tabel 1.

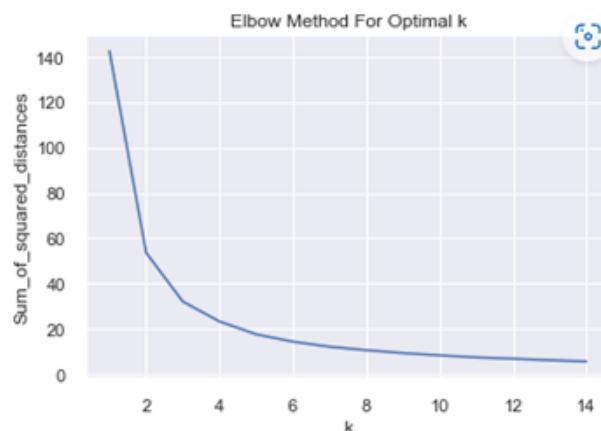
Tabel 1. Data Setelah Melalui Proses *Filtering, Formating, dan Validasi*

ID	NIM	JUMNA	SKS	IPK	ANG
1	113310030	3	25	0,48	11
2	113310033	4	47	0,72	11
3	123310004	4	70	1,4	12
4	123310017	3	13	1,54	12
5	133310018	4	21	0,33	13
....
950	155410126	4	22	2,64	15
951	155410132	5	72	3,39	15
952	155410137	4	22	3,82	15
953	155410183	4	42	1,38	15
954	165410093	1	15	0,00	16
955	165410187	1	22	0,00	16

3.2. Luaran Proses

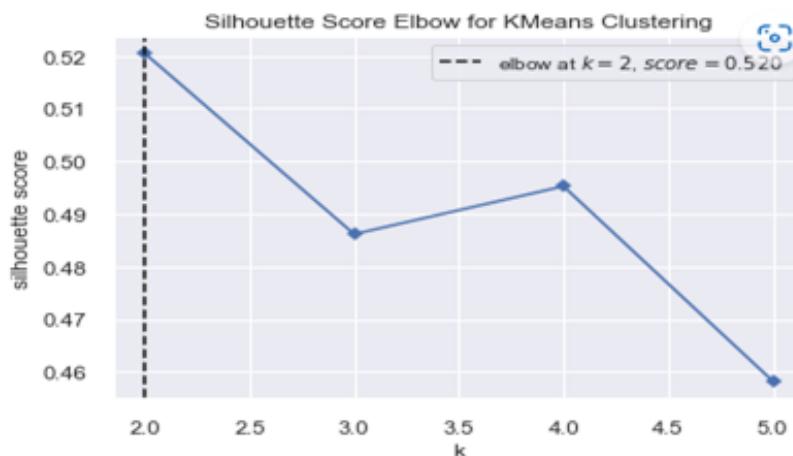
Berikut ini luaran hasil proses klasterisasi menggunakan metode K-Means dengan K=2 :

1. Grafik analisis kualitas klaster menggunakan Elbow Methode. Grafik Gambar 2 menunjukkan grafik yang dibuat berdasarkan Elbow Metode dengan nilai k=1 sampai k=15, untuk koordinat mendatar dan *sum of square distance* untuk koordinat vertikal

**Gambar 2.** Grafik jumlah square distance untuk nilai k=1 sampai k=14

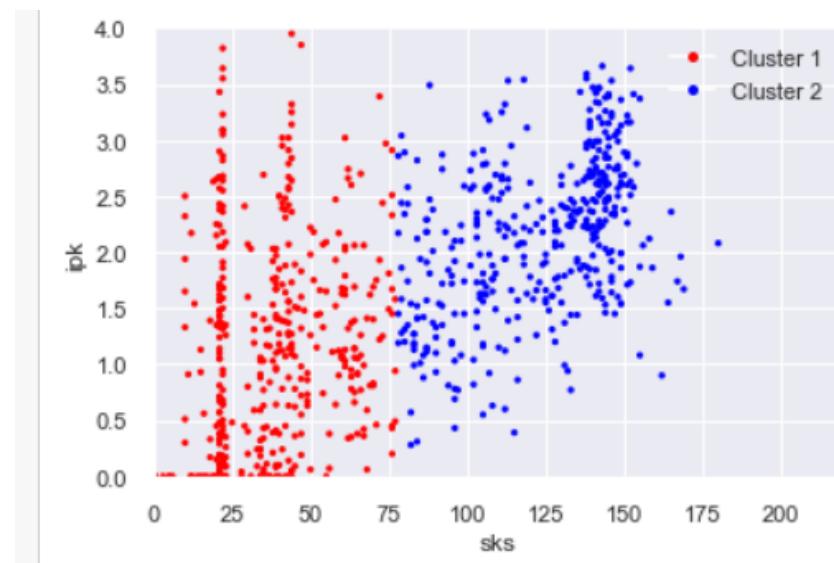
Berdasarkan grafik Gambar 2 tersebut maka didapat nilai optimal untuk K, yakni k=2 , karena sudut potongan antar garis memiliki sudut siku terbesar.

2. Grafik Silhouette Coefficient Berdasarkan data pada Tabel 2, bila dibuat grafik nilai K=2 sampai dengan K=5, menggunakan metode Silhouette Coefficient didapat grafik seperti Gambar 3.



Gambar 3. Grafik Silhouette Score (Silhouette Coefficient) untuk nilai k=2 sampai k=5

Berdasarkan grafik Gambar 3 terlihat bahwa nilai k paling optimal adalah k=2 dimana pada metode ini tingkat separasi paling besar terletak pada k=2 dengan silhouette score 0,52 (dari skala 0 sampai dengan 1)



Gambar 4. Scatter Plot SKS vs IPK Hasil Klasterisasi Menggunakan Algoritma K-Means

3. Hasil Klaster tiap Obyek berdasarkan *data training* Tabel 1, kemudian dicari berapa nilai k optimal yakni yang memiliki tingkat separasi terbaik menggunakan 2 metode: Elbow Method dan Silhouette Coefficient. Berdasarkan 2 metode diatas dihasilkan nilai k optimum yang sama yakni k=2, sehingga dilakukan proses klasterisasi menggunakan algoritma K-Means dengan hasil seperti pada Tabel 2.

Tabel 2. Tabel Hasil Klasterisasi Menggunakan Metode K-Means dengan K=2

ID	NIM	JUMNA	SKS	IPK	ANG	Klaster
1	113310030	3	25	,48	11	cluster_0
2	113310033	4	47	,72	11	cluster_0
3	123310004	4	70	1,40	12	cluster_0
4	123310017	3	13	1,54	12	cluster_0
5	133310018	4	21	,33	13	cluster_0
.....
451	165410217	2		,00	16	cluster_1
452	165410218	2	22	,00	16	cluster_1
453	165410222	1		,00	16	cluster_1
454	165410223	2		,00	16	cluster_1
455	165410224	2		,00	16	cluster_1
456	133310024	4	57	1,11	13	cluster_0
.....
951	155410132	5	72	3,39	15	cluster_0
952	155410137	4	22	3,82	15	cluster_0
953	155410183	4	42	1,38	15	cluster_0
954	165410093	1	15	,00	16	cluster_1
955	165410187	1	22	,00	16	cluster_1

- Scatter Plot SKS vs IPK Berdasarkan hasil klasterisasi pada Tabel 3 kemudian dibuat Scatter Plot seperti Gambar 4. Berdasarkan Gambar 4 terlihat bahwa tingkat separasi dengan nilai k=2 cukup baik. Klaster ke-1 (Cluster_0) cenderung dominan di sebelah kiri atau pada sks lebih kecil dari 75 sks dan IPK dibawah 2. Sedang klaster ke-2 (Cluster_1) cenderung memiliki IPK lebih tinggi dari 1,5 dengan jumlah SKS diatas 75 SKS.
- Tabel Statistik tiap Klaster Berdasarkan data hasil klasterisasi sesuai Tabel 2 maka dibuat statistik ditribusi sesuai dengan klasernya berdasarkan IPK, jumlah sks, dan ketidak-aktifan maka akan dihasilkan distribusi seperti Tabel3, Tabel 4, dan Tabel 5.

Tabel 3. Distribusi Klaster Berdasarkan IPK

IPK	IPK<2,00	2,00<=IPK<3,00	IPK>=3,00	Total
Cluster_0	423	315	76	814
Cluster_1	66	48	27	141

Tabel 4. Distribusi Klaster Berdasarkan Jumlah SKS

Jumlah SKS	SKS<90	90≤SKS≤120	120≤SKS≤144	SKS≥144	Total
Cluster_0	420	226	155	71	762
Cluster_1	176	8	7	2	193

Tabel 5. Distirbusi Klaster Berdasarkan Ketidakaktifan (semester)

Tidak Aktif	1 Sem	2 Sem	3 Sem	≥4 Sem	Total
Cluster_0	53	97	99	409	658
Cluster_1	42	89	44	122	297

3.3. Pembahasan

Berdasarkan Gambar 2, yang merupakan grafik dari Elbow Method terlihat bahwa siku atau sudut perpotongan paling kecil terjadi pada k=2 sehingga pada nilai k ini akan menghasilkan jumlah klaster terbaik

dimana terjadi separasi antar klaster paling optimal. Sementara pada Gambar3 yang merupakan grafik dari Silhouette Coefficient, puncak grafik atau disebut dengan Silhouette Coefficient/Score terjadi pada nilai k=2, dengan Silhouette Score = 0,52. Ini menunjukkan bahwa separasi optimal terjadi bila hanya terdapat 2 buah klaster.

Pembuktian secara visual separasi 2 klaster optimal terlihat pada grafik titik (scatter plot) yang ada pada Gambar 4. Berdasar grafik ini terlihat kumpulan titik terpusat di 2 bagian koordinat grafik, yakni di kiri bawah dan kanan atas. Hal ini menunjukkan bahwa adanya 2 klaster ini merupakan jumlah yang paling optimal.

Tabel 2 menunjukkan hasil klasterisasi yang diindikasikan dengan label klaster dan terdapat 2 nilai : Cluster_0 dan Claster_1. Dari segi jumlah obyek, Cluster_0 sejumlah 813 obyek (mahasiswa DO) dan Cluster_1 sejumlah 142 mahasiswa sehingga prosentasi Cluster_0 mendominasi jumlah obyeknya (85,1%). Berdasarkan Tabel 3 dapat diketahui bahwa mahasiswa DO 48%-nya ada di Cluster_0 dimana IPK<2,00, sedangkan pada Tabel 4 4 terlihat bahwa dominasi mahasiswa DO juga karena jumlah SKS yang diambil masih <90 sks. Jika dilihat dari Tabel 5 diketahui bahwa 44% mahasiswa DO juga karena tidak aktif 4 semester atau lebih.

4. Simpulan

Berdasarkan pembahasan terdahulu disimpulkan bahwa

1. Telah berhasil dilakukan proses klasterisasi untuk mahasiswa DO, grafik Elbow Method, Silhouette Coefficient, dan *scatter plot* menggunakan bahasa pemrograman Python dengan metode K-Means
2. Telah berhasil diketahui bahwa cacah klaster yang optimal untuk menghasilkan klasterisasi terbaik pada baik menggunakan Elbow method maupun Silhouette Coefficient pada metode K-Means adalah k=2.
3. Telah dihasilkan klasterisasi untuk tiap mahasiswa atau obyek dengan nilai klas Cluster_0 dan Cluster_1 dimana jumlah Cluster_0 adalah 813 mahasiswa, dan Cluster_1 sejumlah 142 mahasiswa.
4. Terdapat 48% mahasiswa DO dikarenakan IPK <2, 00 dan 44% mahasiswa DO karena ketidakaktifannya sebagai mahasiswa 4 semester atau lebih.

Pustaka

- [1] R. A. Farissa, R. Mayasari, and Y. Umaidah, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Data Obat dengan Silhouette Coefficient di Puskesmas Karangsambung," *Journal of Applied Informatics and Computing*, vol. 5, no. 2, pp. 109–116, 2021.
- [2] I. Vhalla, S. Sumijan, and J. Santony, *Pengelompokan Mahasiswa Potensial Drop Out Menggunakan Metode Clustering K-Means*. RESTI (Rekayasa Sist. dan Teknol. Informasi, 2018, vol. 2, no. 2.
- [3] Akakom, *Peraturan Akademik STMIK Akakom*. STMIK AKAKOM, 2019. [Online]. Available: <https://if.utdi.ac.id/wp/wp-content/uploads/2017/12/SK-Peraturan-Akademik-2017.pdf>
- [4] ——, "SK Ketua STMIK AKAKOM tentang Pengunduran Diri Mahasiswa," *STMIK AKAKOM*, pp. 2015–2019.
- [5] A. F. Febrianti, A. H. Cabral, and G. Anuraga, "K-Means Clustering Dengan Metode Elbow Untuk Pengelompokan Kabupaten Dan Kota Di Jawa Timur," *Seminar Nasional Hasil Riset dan Pengabdian -SNHRP*, vol. 1, pp. 863–870, 2018. [Online]. Available: <http://karyailmiah.unipasby.ac.id/wp-content/uploads/2019/04/K-Means-Artikel.pdf>
- [6] A. T. Rahman, Wiranto, and A. Rini, "Coal Trade Data Clustering Using K-Means (Case Study Pt. Global Bangkit Utama)," *ITSMART: Jurnal Teknologi dan Informasi*, vol. 6, no. 1, pp. 24–31, 2017.
- [7] A. Mario, S. Herry, and H. Nasution, "Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness," *Jurnal Sistem dan Teknologi Informasi*, vol. 1, no. 1, pp. 1–6, 2016.
- [8] M. Robani and A. Widodo, "Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia," *Jurnal Sistem Informasi Bisnis*, vol. 6, no. 2, p. 164, 2016.

- [9] R. Hidayati, A. Zubair, A. H. Pratama, and L. Indana, “Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering,” *Techno.Com*, vol. 20, no. 2, pp. 186–197, 2021.
- [10] N. T. Hartanti, “Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian Nasional,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 6, no. 2, pp. 82–89, 2020.