
Klasterisasi Data Kompetisi Mahasiswa Menggunakan Algoritma K-Modes

Raditya Pradhyatama Nugroho¹, Angelina Prima Kurniati², Yanuar Firdaus Arie Wibowo³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹radityanugroho@student.telkomuniversity.ac.id, ²angelina@telkomuniversity.ac.id,

³yanuar@telkomuniversity.ac.id

Abstrak

Kompetisi adalah salah satu aspek penting dalam dunia pendidikan. Salah satu fungsi utama partisipasi dalam kompetisi adalah dapat mengukur kemampuan diri. Selain itu, kompetisi juga memiliki berbagai manfaat, yakni dapat menciptakan inovasi baru dalam pemecahan suatu masalah, memberikan pengalaman baru, dan mengenal/memahami persaingan antar peserta. Peserta kompetisi di bidang pendidikan, tidak lepas dari dukungan universitas yang mewadahi segala fasilitas dan penunjang dalam aspek lainnya. Untuk lebih mengenal perkembangan kompetisi peserta dalam suatu universitas, diperlukan analisis data yang dapat memberikan gagasan baru. Teknik analisis data yang efektif adalah dengan pendekatan penambangan data, lebih rincinya klasterisasi. Klasterisasi memiliki banyak variasi algoritma dengan salah satu contohnya K-Modes yang mampu menangani data kategorikal seperti data kompetisi. Sebagai perguruan tinggi dengan peraih prestasi yang cukup banyak dan menjadi salah satu universitas swasta terbaik di Indonesia, Universitas Telkom menjadi subjek yang menarik untuk implementasi algoritma K-Modes. Penelitian ini bertujuan untuk melakukan klasterisasi K-Modes pada data kompetisi Mahasiswa Universitas Telkom dan ditambah dengan metode Silhouette Score dan Davies-Bouldin Index untuk membantu menentukan jumlah klaster yang optimal. Hasil penelitian menunjukkan jumlah klaster optimal adalah 13 klaster dengan Silhouette Score bernilai 0,23 (jangkauan nilai -1 hingga 1) dan Davies-Bouldin Index bernilai 1,95 (jangkauan nilai mulai dari 0 ke bilangan positif). Selanjutnya penelitian ini memberikan visualisasi klaster-klaster K-Modes dari perhitungan metode Principal Component Analysis (PCA) dan hasil identifikasi berupa informasi analisis setiap klaster.

Kata kunci: Davies-Bouldin Index, K-Modes, Klasterisasi, PCA, Silhouette Score, Universitas Telkom

Abstract

Competition is an important aspect of education. One of the main functions of participating in competitions is to measure personal abilities. In addition, competitions have various benefits, namely creating new innovations in solving a problem, providing new experiences, and recognizing/understanding competition between participants. Competition participants in the educational field cannot be separated from the support of universities that offer all facilities in many aspects. To better recognize the development of competition participants in a university, data analysis is needed that can provide new ideas. An effective data analysis technique is the data mining approach, specifically clustering. Clustering has many variations of algorithms with one example being K-Modes that can handle categorical data such as competition data. As a university with numerous achievements and one of the best private universities in Indonesia, Telkom University is an interesting subject for the implementation of the K-Modes algorithm. This study aims to conduct K-Modes clustering on Telkom University Student competition data and couple with the Silhouette Score and Davies-Bouldin Index methods to help determine the optimal number of clusters. The results showed that the optimal number of clusters was 13 clusters with a Silhouette Score of 0.23 (value range -1 to 1) and a Davies-Bouldin Index of 1.95 (value range from 0 to positive number). Furthermore, this research provides visualization of K-Modes clusters from the Principal Component Analysis (PCA) method calculation and identification results in the form of analysis information for each cluster.

Keywords: Clustering, Davies-Bouldin Index, K-Modes, PCA, Silhouette Score, Telkom University

1. Pendahuluan

Kompetisi menjadi komponen penting dalam bidang pendidikan yang memiliki tujuan untuk mengembangkan penelitian, mendorong kolaborasi, dan memotivasi setiap peserta. Tentunya, partisipasi kompetisi memberikan manfaat seperti memberikan pengalaman pemecahan masalah/isu sebagai pengembangan inovasi baru [1], menghasilkan produk kerja berkualitas tinggi [2], dan memungkinkan berbagai pihak untuk berpartisipasi pada penyelesaian isu ilmiah tertentu [3].

Perkembangan partisipasi kompetisi juga menjadi hal yang penting bagi setiap institusi pendidikan khususnya universitas, mengingat kompetisi merupakan salah satu kegiatan yang menjadi dasar pengukuran kemampuan mahasiswanya [4]. Perkembangan ini dapat dilihat dengan melakukan identifikasi pola pada kompetisi mahasiswa melalui analisis pengelompokan dengan metode klusterisasi. Beberapa variabel yang dapat menjadi acuan klusterisasi, yaitu identitas mahasiswa (angkatan, program studi, fakultas) dan identitas kompetisi (tahun pelaksanaan, tingkat kompetisi, prestasi/juara yang diraih). Namun semua variabel tersebut, tergolong dalam variabel kategorikal sehingga perlu metode klusterisasi yang khusus menangani data kategorikal.

K-Modes merupakan salah satu algoritma klusterisasi yang khusus untuk dapat menangani data kategorikal. Beberapa implementasi algoritma K-Modes, yakni prediksi penyakit yang disebabkan oleh pola diet anak-anak [5], mengelompokkan gaya belajar mahasiswa [6], dan mengelompokkan berbagai faktor stress mahasiswa berdasarkan tingkat-tingkat tertentu [7].

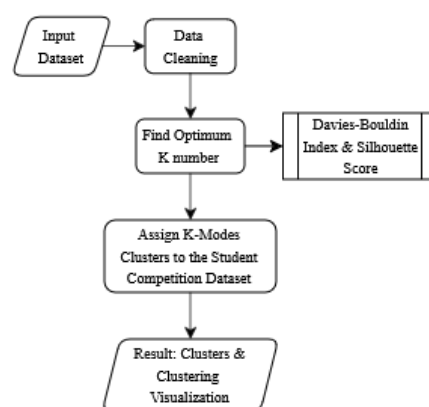
Untuk menunjang kinerja algoritma K-Modes agar semakin baik, diperlukan beberapa metode tambahan seperti metode untuk pemilihan jumlah kluster. Beberapa diantaranya adalah Davies-Bouldin Index (DBI) dan Silhouette Score yang memungkinkan peneliti menghasilkan luaran kluster yang baik. Selain dapat menentukan jumlah kluster terbaik, DBI dan Silhouette Score juga dapat menampilkan nilai masing-masing untuk memberikan informasi apakah kluster yang dibuat sudah baik dan tepat. Hal ini membuat hasil penentuan kluster optimal dan terbaik lebih akurat dan meminimalisasi kesalahan klusterisasi.

Universitas Telkom merupakan salah satu universitas swasta terbaik di Indonesia [8]. Mahasiswa Universitas Telkom telah mengukir banyak prestasi yang membanggakan. Tercatat dalam data kompetisi mahasiswa Universitas Telkom, terdapat lebih dari 950 prestasi sepanjang 2020 hingga 2024. Oleh sebab itu, dengan jumlah prestasi dan jumlah partisipasi kompetisi yang menjanjikan, menjadi subjek menarik untuk analisis berbasis penambangan data melalui klusterisasi K-Modes dengan metode Silhouette Score dan Davies-Bouldin Index.

2. Studi Terkait

Terdapat beberapa studi yang mengadopsi teknik klusterisasi pada analisis data untuk penyelesaian isu tertentu khususnya pada bidang pendidikan dan dapat menjadi acuan bagi penelitian ini. Pertama pada studi [6], membuat aplikasi untuk analisis gaya belajar mahasiswa menggunakan algoritma K-Modes. Pada aplikasi tersebut juga diberikan fitur input kuesioner bagi pengguna yang kemudian dikumpulkan menjadi suatu dataset sehingga dapat dilihat oleh admin dan menjadi bahan untuk proses klusterisasi. Studi [6] ini mengombinasikan bidang rekayasa perangkat lunak dengan penambangan data sebagai penyelesaian isu di bidang pendidikan. Selanjutnya pada studi [7], melakukan implementasi algoritma K-Modes untuk klusterisasi penyebab stress mahasiswa. Elbow Method juga diimplementasikan dalam studi [7] untuk optimalisasi proses klusterisasi K-Modes sehingga kluster yang dibuat menjadi lebih tepat dan akurat. Berikutnya pada studi [9], menganalisis data akreditasi salah satu universitas menggunakan metode klusterisasi algoritma K-Medoids. Untuk memperkuat hasil klusterisasi K-Medoids, diberikan evaluasi Davies-Bouldin Index dan Silhouette Coefficient. Hasil klusterisasi diperoleh nilai evaluasi Davies-Bouldin Index terbaik 0,4 dan Silhouette Coefficient terbaik 0,6.

3. Sistem yang Dibangun

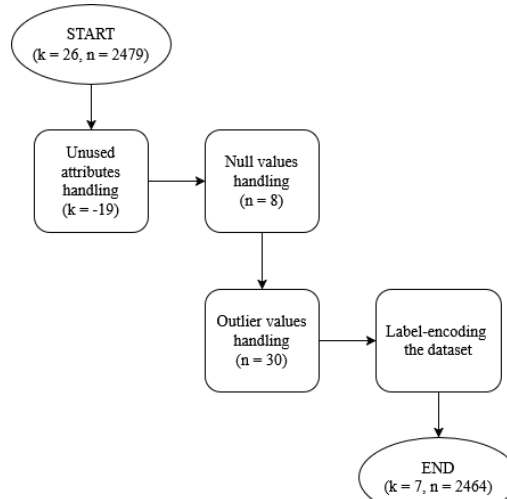


Gambar 3.1. Tahap penelitian klusterisasi kompetisi Mahasiswa Universitas Telkom

Proses penelitian secara visual digambarkan pada Gambar 3.1., menguraikan tahapan penelitian yang berurutan. Dimulai dari masukan dataset kompetisi Mahasiswa Universitas Telkom, selanjutnya melakukan serangkaian proses pembersihan data agar tidak terjadi kesalahan atau error pada saat pengujian klusterisasi algoritma K-Modes, lalu menentukan jumlah kluster paling optimal dengan metode evaluasi Silhouette Score dan Davies-Bouldin Index, kemudian menetapkan kluster berdasarkan jumlah yang paling optimal ke dalam dataset, terakhir menampilkan luaran hasil klusterisasi K-Modes berupa kluster-kluster, visualisasi kluster dengan metode Principal Component Analysis (PCA), dan beberapa luaran lainnya.

Dataset

Data kompetisi mahasiswa Universitas Telkom diperoleh melalui Lembaga Administrasi Akademik Kemahasiswaan, Universitas Telkom. Dalam keadaan “mentah”, dataset ini berisi 2479 baris rekaman data dan 26 atribut. Dataset ini memiliki isi rentang waktu 2020 hingga 2024. Oleh sebab adanya kekurangan pada dataset ini, diperlukan serangkaian tahap pembersihan data agar pengujian klasterisasi berjalan dengan lancar dan mengurangi kendala error. Prosedur yang dilakukan pada tahap pembersihan data (pada Gambar 3.2.):



Gambar 3.2. Prosedur tahap pembersihan data

Gambar 3.2. memberikan keterangan tambahan mengenai kekurangan yang terdapat dalam data kompetisi Mahasiswa Universitas Telkom. Diantaranya adalah atribut yang kosong dan tidak relevan dengan tujuan penelitian, nilai kosong pada beberapa baris data, dan adanya pencilan (*outlier*) pada beberapa atribut. Sehingga, pembersihan menangani segala kekurangan tersebut dengan cara menghapus, menambahkan, atau mengganti data. Hasilnya, terdapat 7 atribut dan 2464 baris data yang digunakan, semula berjumlah 26 atribut dan 2479 baris data. Untuk informasi atribut yang dipilih dan digunakan dalam penelitian ini adalah (lihat Tabel 3.1):

Tabel 3.1. Atribut dataset kompetisi Mahasiswa Universitas Telkom terpilih

Atribut	Tipe Data	Keterangan
Fakultas	String	Fakultas dari mahasiswa
Tahun	Numerik	Tahun kompetisi diselenggarakan
Kuartal	String	Kuartal tahun kompetisi
Tingkat	String	Tingkat kompetisi
Juara	String	Perolehan prestasi
Angkatan	Numerik	Tahun angkatan mahasiswa
Tingkat Kelas	Numerik	Tingkat kelas mahasiswa saat mengikuti kompetisi

4. Evaluasi

Berdasarkan pengujian untuk implementasi algoritma K-Modes pada data kompetisi Mahasiswa Universitas Telkom memperoleh hasil yang dipaparkan dalam 2 bagian: hasil pengujian dan analisis. Bagian hasil pengujian membahas mekanisme dan sistem kerja penelitian sesuai dengan tahap penelitian (pada Gambar 3.1.). Sedangkan bagian 2 menganalisis tentang pengujian dan hasil pengujian yang diperoleh. Lebih lengkapnya:

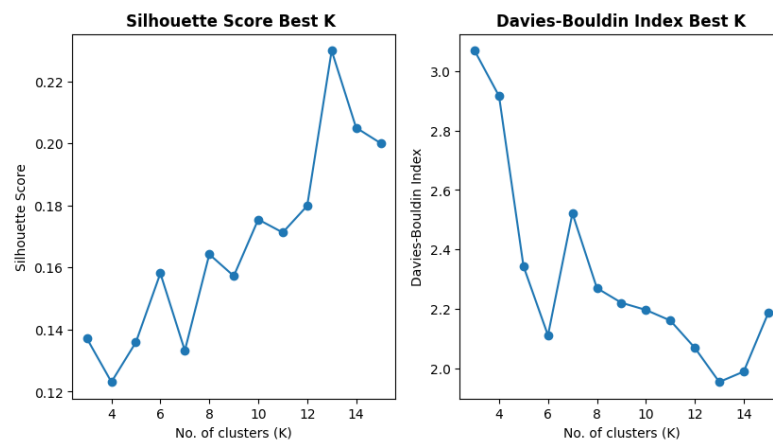
4.1. Hasil Penelitian

Setelah proses pembersihan data selesai dan sudah dipastikan tidak ada lagi kekurangan pada data, proses eksekusi algoritma K-Modes dapat dilakukan. Proses ini dimulai dengan mencari jumlah kluster optimal agar kluster yang dipetakan ke setiap titik data lebih tepat dan hasil klasterisasi keseluruhan yang lebih baik. Tahap pencarian kluster optimal dilakukan dengan menentukan kandidat nilai jumlah kluster yang akan diuji coba. Pada penelitian ini, beberapa kandidat nilai jumlah kluster yang ditentukan adalah 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15. Seluruh kandidat ini berikutnya diterapkan dalam program perulangan untuk mengkalkulasi Silhouette Score dan DBI. Hasil dari program perulangan kalkulasi Silhouette Score dan DBI dibuat menjadi keluaran menjadi *list* yang kemudian menjadi tabel seperti pada Tabel 4.1. serta menjadi grafik seperti pada Gambar 4.1.

Tabel 4.1. Silhouette Score & DBI setiap jumlah percobaan kluster

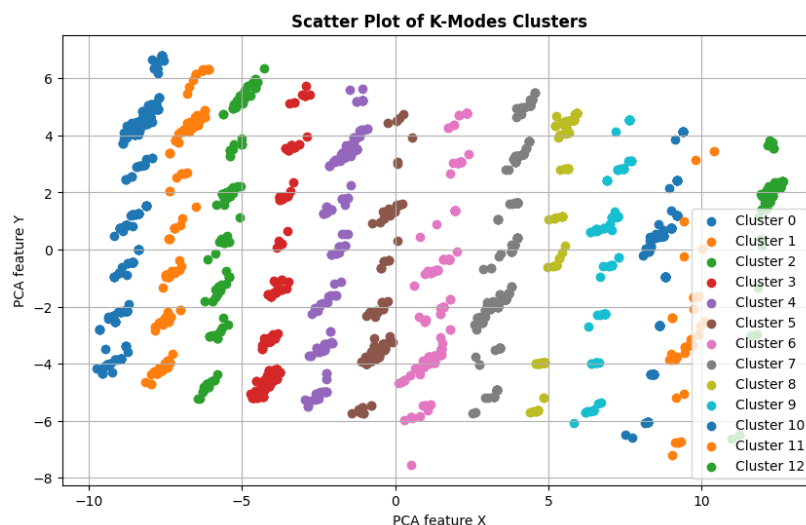
Jumlah Kluster (K)	Silhouette Score	Davies-Bouldin Index
3	0,14	3,07
4	0,12	2,92
5	0,14	2,34
6	0,16	2,11
7	0,13	2,52
8	0,16	2,27
9	0,16	2,22
10	0,18	2,20
11	0,17	2,16
12	0,18	2,07
13	0,23	1,95
14	0,21	1,99
15	0,20	2,19

Hasil keluaran dari kalkulasi Silhouette Score dan DBI untuk seluruh kandidat nilai yang ditampilkan pada Tabel 4.1. dan Gambar 4.1. menunjukkan bahwa jumlah kluster dengan Silhouette Score dan Davies-Bouldin Index tertinggi adalah K = 13 dengan nilai evaluasi masing-masing sebesar 0,23 dan 1,95.



Gambar 4.1. Grafik jumlah kluster terhadap Silhouette Score & DBI

Setelah mendapat (jumlah K = 13) sebagai jumlah kluster optimal untuk klusterisasi K-Modes ini, tahap selanjutnya adalah menetapkan kluster ke dalam dataset. Kemudian tahap terakhir adalah menampilkan visualisasi untuk hasil keseluruhan kluster yang dihasilkan dengan menggunakan metode PCA (lihat Gambar 4.2.). Visualisasi kluster-kluster K-Modes dilakukan dengan inisialisasi 2 komponen PCA yang merepresentasikan jumlah dimensi grafik (2 dimensi) dan ditampilkan dalam *scatter plot*.



Gambar 4.2. Hasil klusterisasi K-Modes kompetisi Mahasiswa Universitas Telkom

Setelah visualisasi kluster dengan metode PCA berhasil ditampilkan, selanjutnya melakukan identifikasi informasi setiap kluster melalui pengamatan dari hasil luaran klusterisasi K-Modes menjadi file (.xlsx) yang terinspirasi dari studi [7]. Tabel 4.2. berikut adalah informasi lengkap untuk masing-masing kluster:

Tabel 4.2. Informasi setiap kluster-kluster K-Modes

Kluster	Penjelasan
1	- Jumlah titik data terkandung: 319 - Mencakup Fakultas Rekayasa Industri dan Fakultas Informatika - Tahun kompetisi yang terselenggarakan pada 2023 - Mayoritas pada tingkat kompetisi nasional dengan tanpa perolehan prestasi - Angkatan mahasiswa terbanyak adalah 21 pada tingkat kelas 3
2	- Jumlah titik data terkandung: 194 - Mencakup Fakultas Rekayasa Industri dan Fakultas Ekonomi & Bisnis - Tahun kompetisi yang terselenggarakan pada 2021 - Mayoritas pada tingkat kompetisi nasional dengan perolehan prestasi terbanyak juara 2 - Angkatan mahasiswa terbanyak adalah 18 pada tingkat kelas 3
3	- Jumlah titik data terkandung: 157 - Mencakup Fakultas Teknik Elektro dan Fakultas Komunikasi & Bisnis - Tahun kompetisi yang terselenggarakan pada 2023 - Mayoritas pada tingkat kompetisi nasional dengan perolehan prestasi terbanyak juara 3 - Angkatan mahasiswa terbanyak adalah 21 pada tingkat kelas 2
4	- Jumlah titik data yang terkandung: 263 - Mencakup Fakultas Ekonomi & Bisnis dan Fakultas Ilmu Terapan - Tahun kompetisi yang terselenggarakan pada 2021 - Mayoritas pada tingkat kompetisi nasional dengan tanpa perolehan prestasi - Angkatan mahasiswa terbanyak adalah 19 pada tingkat kelas 2
5	- Jumlah titik data terkandung: 192 - Mencakup Fakultas Rekayasa Industri dan Fakultas Komunikasi & Bisnis - Tahun kompetisi yang terselenggarakan pada 2022 - Mayoritas pada tingkat kompetisi nasional dengan perolehan prestasi terbanyak juara 2 - Angkatan mahasiswa terbanyak adalah 21 pada tingkat kelas 2
6	- Jumlah titik data terkandung: 196 - Mencakup Fakultas Ilmu terapan dan Fakultas Industri Kreatif - Tahun kompetisi yang terselenggarakan pada 2021 - Mayoritas pada tingkat kompetisi nasional dengan tanpa perolehan prestasi - Angkatan mahasiswa terbanyak adalah 20 pada tingkat kelas 1
7	- Jumlah titik data terkandung: 217 - Mencakup Fakultas Ilmu Terapan dan Fakultas Informatika - Tahun kompetisi yang terselenggarakan pada 2023 - Mayoritas pada tingkat kompetisi nasional dengan tanpa perolehan prestasi - Angkatan mahasiswa terbanyak adalah 22 pada tingkat kelas 2
8	- Jumlah titik data terkandung: 363 - Mencakup Fakultas Industri Kreatif dan Fakultas Rekayasa Industri - Tahun kompetisi yang terselenggarakan pada 2022 - Mayoritas pada tingkat kompetisi internasional dengan tanpa perolehan prestasi - Angkatan mahasiswa terbanyak adalah 20 pada tingkat kelas 3
9	- Jumlah titik data yang terkandung: 127 - Mencakup Fakultas Teknik Elektro dan Fakultas Ilmu Terapan - Tahun kompetisi yang terselenggarakan pada 2022 - Mayoritas pada tingkat kompetisi nasional dengan tanpa perolehan prestasi - Angkatan mahasiswa terbanyak adalah 20 pada tingkat kelas 3
10	Jumlah titik data yang terkandung: 122

	• Mencakup Fakultas Komunikasi & Bisnis dan Fakultas Rekayasa Industri • Tahun kompetisi yang terselenggarakan pada 2023 • Mayoritas pada tingkat kompetisi nasional dengan tanpa perolehan prestasi • Angkatan mahasiswa terbanyak adalah 21 pada tingkat kelas 4
11	• Jumlah titik data yang terkandung: 156 • Mencakup Fakultas Industri Kreatif dan Fakultas Informatika • Tahun kompetisi yang terselenggarakan pada 2022 • Mayoritas pada tingkat kompetisi internasional dengan tanpa perolehan prestasi • Angkatan mahasiswa terbanyak adalah 21 pada tingkat kelas 2
12	• Jumlah titik data yang terkandung: 47 • Mencakup Fakultas Industri Kreatif dan Fakultas Informatika • Tahun kompetisi yang terselenggarakan pada 2023 • Mayoritas pada tingkat kompetisi wilayah dengan perolehan prestasi juara 1 • Angkatan mahasiswa terbanyak adalah 21 pada tingkat kelas 3
13	• Jumlah titik data yang terkandung: 115 • Mencakup Fakultas Rekayasa Industri dan Fakultas Informatika • Tahun kompetisi yang terselenggarakan pada 2022 • Mayoritas pada tingkat kompetisi nasional dengan tanpa perolehan prestasi • Angkatan mahasiswa terbanyak adalah 18 pada tingkat kelas 4

4.2. Analisis

Dari seluruh kandidat yang diuji coba (3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, & 15) bahwa jumlah kluster paling optimal dan terbaik untuk data kompetisi Universitas Telkom adalah jumlah $K = 13$ dengan masing-masing nilai evaluasi Silhouette Score dan Davies-Bouldin Index (DBI) sebesar 0,23 dan 1,95 (lihat Tabel 4.1.). Lalu pada Gambar 4.1. menunjukkan bahwa perkembangan nilai evaluasi masing-masing metode terbilang cukup stabil, yakni semakin besar jumlah kluster, besaran nilai evaluasi mengalami kenaikan dan penurunan yang juga stabil serta puncak jumlah kluster terbaik yang diperoleh sama (pada jumlah $K = 13$).

Nilai evaluasi Silhouette Score pada jumlah $K = 13$ sebesar 0,23 memberikan pemisahan antar-kluster yang jelas, tetapi kesederhanaan titik-titik data intra-kluster yang kurang baik. Terlihat pada Gambar 4.2., masih terdapat kumpulan titik data intra-kluster yang memiliki jarak koordinat yang cukup jauh. Hal ini dapat terjadi oleh sebab beberapa kemungkinan: proses pembersihan data yang belum sempurna, pemilihan atribut yang kurang tepat maupun kurang relevan, atau pemilihan parameter eksekusi algoritma dan metode evaluasi yang tidak sesuai. Jika titik-titik data intra-kluster memiliki jarak yang pendek, maka nilai evaluasi Silhouette Score dapat semakin meningkat mendekati angka 1 yang berarti semakin baik [10]. Berikutnya, nilai evaluasi DBI pada jumlah $K = 13$ sebesar 1,95 menunjukkan masing-masing kluster terpisah dengan baik dengan tingkat *overlap* yang rendah. Menurut [11], DBI dengan nilai 1 hingga 2 dinyatakan valid untuk klusterisasi efektif. Sedangkan DBI dengan nilai di atas 3 menunjukkan *overlap* yang signifikan antar-kluster yang dapat menyebabkan kesalahan interpretasi struktur data [12].

Berdasarkan identifikasi kluster-kluster K-Modes yang dilakukan, diperoleh informasi-informasi penting yang dapat membantu pihak terkait untuk menganalisis perkembangan kompetisi di Universitas Telkom. Salah satu contohnya (lihat Tabel 4.2.) seperti pada kluster 8 yang merupakan kluster dengan jumlah titik data terbanyak sebesar 363 dengan cakupan Fakultas Rekayasa Industri dan Fakultas Komunikasi & Bisnis di tahun penyelenggaraan 2022. Kluster 8 ini merupakan satu-satunya kluster dengan partisipan tingkat kompetisi internasional terbanyak namun tanpa adanya perolehan juara/prestasi. Dari salah satu contoh kluster tersebut, diharapkan penelitian ini memberikan wawasan dan pengetahuan bagi pihak terkait (dalam hal ini, Universitas Telkom) yang selanjutnya dapat menciptakan ide agar dapat menghadapi segala situasi dan kondisi secara tidak terkecuali untuk seluruh kluster K-Modes lainnya.

5. Kesimpulan

Berdasarkan seluruh rangkaian proses implementasi algoritma K-Modes untuk data kompetisi mahasiswa Universitas Telkom dapat disimpulkan bahwa beberapa hasil yang diperoleh: nilai evaluasi metode Silhouette Score dan Davies-Bouldin Index untuk setiap kandidat jumlah kluster untuk menentukan jumlah kluster paling optimal, visualisasi kluster-kluster K-Modes melalui metode PCA dengan *scatter plot*, dan informasi setiap kluster yang diperoleh melalui identifikasi atau pengamatan luaran file (.xlsx) klusterisasi K-Modes. Selanjutnya, informasi setiap kluster menunjukkan jumlah titik data yang terkandung, cakupan fakultas, tahun penyelenggaraan dan tingkat kompetisi, serta perolehan prestasi dari Angkatan terbanyak menyesuaikan tingkat kelas mahasiswanya.

Oleh sebab nilai evaluasi Silhouette score yang belum optimal, diharapkan penelitian ini dapat dimaksimalkan kembali dengan saran: penyempurnaan tahap pembersihan data, pemilihan atribut yang lebih sesuai dan relevan, serta percobaan eksekusi algoritma dan metode-metode lainnya dengan jenis dan parameter yang bervariasi sehingga klusterisasi K-Modes mencapai hasil yang lebih baik.

Daftar Pustaka

- [1] Y. Wei, N. Zhao, and L. Ge, "The research on the cultivation of college students' innovation ability based on academic competition," *SHS Web of Conferences*, vol. 166, p. 01024, 2023, doi: 10.1051/shsconf/202316601024.
- [2] E. A. Ivanchenko, T. V. Vorotilina, S. S. Teligisova, I. S. Shulzhenko, and K. A. Selivanova, "Phenomenon of competition in the educational environment," *Revista on line de Política e Gestão Educacional*, p. e022065, Mar. 2022, doi: 10.22633/rpge.v26iesp.2.16563.
- [3] H. J. Escalante and A. Kruchinina, "Academic competitions," Nov. 2023, [Online]. Available: <http://arxiv.org/abs/2312.00268>
- [4] L. Dan, C. Wei, and D. Binyan, "Investigation and research on the motivation of college students' participation in academic competitions: Taking the English competition of Panzhuhua University as an example," *International Journal of Advanced Academic Studies*, vol. 5, no. 12, pp. 44–47, Dec. 2023, doi: 10.33545/27068919.2023.v5.i12a.1092.
- [5] C. Bhanuprakash and A. Setty, "Prediction of diseases caused in Child Dieting System by using K-Modes Algorithm," in *2024 International Conference on Smart Systems for applications in Electrical Sciences (ICSSES)*, IEEE, May 2024, pp. 1–6. doi: 10.1109/ICSSES62373.2024.10561366.
- [6] M. F. Kurniawan, D. Sugianti, A. S. Darmawan, A. P. Wibowo, and W. Widiyono, "IMPLEMENTASI CLUSTERISASI UNTUK PENGELOMPOKAN GAYA BELAJAR MAHASISWA DENGAN METODE K MODES," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 8, no. 1, pp. 20–25, Apr. 2024, doi: 10.46880/jmika.Vol8No1.pp20-25.
- [7] N. P. M. N. Dewi and I. B. G. Dwidasmara, "Implementation of K-Modes Algorithm for Clustering of Stress Causes in University Students," *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, vol. 9, no. 3, p. 419, Feb. 2021, doi: 10.24843/JLK.2021.v09.i03.p17.
- [8] F. Fahlevi, "Perguruan Tinggi Swasta Indonesia Masuk Peringkat QS World Rankings 2025," *Tribun News, Pendidikan*. Accessed: Jan. 09, 2025. [Online]. Available: <https://www.tribunnews.com/pendidikan/2024/11/12/perguruan-tinggi-swasta-indonesia-masuk-peringkat-qs-world-rankings-2025>
- [9] G. Ghufro, B. Surarso, and R. Gernowo, "The Implementations of K-medoids Clustering for Higher Education Accreditation by Evaluation of Davies Bouldin Index Clustering," *Jurnal Ilmiah Kursor*, vol. 10, no. 3, pp. 119–128, Nov. 2020, doi: 10.21107/kursor.v10i3.232.
- [10] K. R. Shahapure and C. Nicholas, "Cluster Quality Analysis Using Silhouette Score," in *2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, Oct. 2020, pp. 747–748. doi: 10.1109/DSAA49011.2020.00096.
- [11] Nanda Shalsadilla, Shantika Martha, and Hendra Perdana, "Penentuan Jumlah Cluster Optimum Menggunakan Davies Bouldin Index dalam Pengelompokan Wilayah Kemiskinan di Indonesia," *STATISTIKA Journal of Theoretical Statistics and Its Applications*, vol. 23, no. 1, pp. 63–72, Jun. 2023, doi: 10.29313/statistika.v23i1.1743.
- [12] Y. Hasan, "Pengukuran Silhouette Score dan Davies-Bouldin Index pada Hasil Cluster K-Means dan DBSCAN," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3S1.5001.