

Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering

Aditia Yudhistira^{1*}, Rio Andika²

^{1,2}Ilmu Komputer, Institut Pertanian Bogor, Indonesia

^{1*}aditia28aditia@apps.ipb.ac.id, ²rioandika@apps.ipb.ac.id

Abstrak: Data *mining* merupakan sebuah metode dalam bidang ilmu komputer yang digunakan dalam mencari pengetahuan dari data sehingga menjadi sebuah informasi yang bermanfaat. Tahapan dalam proses data *mining* berguna untuk mencari sebuah pola tertentu dari data penilaian yang sangat banyak. Tujuan penelitian ini yaitu mengetahui dan membentuk *cluster* data siswa berdasarkan nilai akademik, nilai sikap, dan nilai disiplin sehingga menjadi sebuah *cluster* sehingga hasil *cluster* siswa dapat menjadi acuan dalam meningkatkan nilai siswa dalam proses pembelajaran selanjutnya. Hasil evaluasi dan penilaian terhadap siswa dilakukan oleh tenaga pengajar atau guru dalam melakukan penilaian selama proses pembelajaran. Dalam proses pembelajaran terdapat 3 kategori penilaian yaitu nilai siswa, disiplin, serta sikap. Hasil pengelompokan data nilai siswa menggunakan metode *K-Means clustering* menunjukkan bahwa berdasarkan hasil cluster data siswa menggunakan dataset siswa dalam satu semester, maka didapatkan *cluster* 0 berjumlah 59 siswa, *cluster* 1 berjumlah 94 siswa, dan *cluster* 2 berjumlah 1 siswa. Hasil pengujian menggunakan *elbow method* maka jumlah cluster yang baik yang digunakan adalah 3 *cluster*, sehingga dalam penelitian ini menggunakan 3 *cluster* yaitu *cluster* 0, *cluster* 1, dan *cluster* 2. Hasil pengujian menggunakan *silhouette coefficient* maka jumlah *cluster* yang baik yang digunakan adalah 3 *cluster* dengan nilai *silhouette coefficient* yaitu 0.489, dan lebih baik dari nilai *silhouette coefficient cluster* lainnya.

Kata Kunci: *Clustering; Data Mining; Elbow Method; K-Means; Silhouette Coefficient;*

Abstract: Data mining is a method in the field of computer science that is used to seek knowledge from data so that it becomes useful information. The stages in the data mining process are useful for looking for a certain pattern from a large number of assessment data. The purpose of this study is to identify and form student data clusters based on academic values, attitude values, and disciplinary values so that they become a cluster so that the results of student clusters can be a reference in increasing student scores in the next learning process. The results of the evaluation and assessment of students are carried out by teaching staff or teachers in conducting assessments during the learning process. In the learning process

there are 3 categories of assessment, namely student scores, discipline, and attitudes. The results of grouping student value data using the K-Means clustering method show that based on the results of student data clusters using a student dataset in one semester, it is obtained that cluster 0 has a total of 59 students, cluster 1 has 94 students, and cluster 2 has 1 student. The results of the test using the elbow method, the good number of clusters used is 3 clusters, so that in this study 3 clusters are used, namely cluster 0, cluster 1, and cluster 2. The results of the test using the silhouette coefficient, the good number of clusters used is 3 clusters with the silhouette coefficient value is 0.489, and is better than the other silhouette coefficient cluster values.

Keywords: Clustering; Data Mining; Elbow Method; K-Means; Silhouette Coefficient;

1. PENDAHULUAN

Data merupakan salah satu komponen yang penting di era digital, seluruh aktivitas data akan terekam dan disimpan dalam media penyimpanan data atau *database*, sehingga data ini dapat diproses untuk mendapatkan informasi yang berguna[1]. Dalam era digital pemanfaatan teknologi dalam pengolahan data sudah diterapkan di berbagai sektor seperti aktivitas transaksi, pendidikan, pelayanan masyarakat, perbankan, transportasi, dan masih banyak lagi. Pengolahan data dengan menggunakan komputer sering dikenal dengan nama *Electronic Data Processing* (EDP). Siklus pengolahan data pada ini mempunyai tiga tahapan yaitu *Input*, *Processing*, dan *Output*. Tahapan tersebut dikembangkan dengan menambahkan tahap *Storage* dan menjadi model siklus pengolahan data (*Data Processing Cycle*). Permasalahan yang muncul dalam pengolahan data ialah bagaimana dapat mengetahui informasi yang akan dihasilkan dari basis data yang sangat besar[2]. *Knowledge Discovery in Database* (KDD) merupakan sebuah proses untuk menemukan pengetahuan dalam sebuah basis data. KDD dapat dijelaskan sebagai sebuah proses ekstraksi atau identifikasi pola, pengetahuan dan informasi dari sekumpulan data yang besar. Pengetahuan dan informasi yang dihasilkan dari KDD bersifat baru, mudah dimengerti, dan bermanfaat bagi para pengguna informasi.

Salah satu faktor yang menentukan kualitas pendidikan dari tinggi dan rendahnya tingkat keberhasilan siswa dalam proses penilaian pembelajaran. Dunia pendidikan pada era digital ini dituntut untuk memiliki kemampuan dalam bersaing dengan memanfaatkan sumber daya yang dimiliki, baik sumber daya sarana, sumber daya prasarana, sampai sumber daya manusia. Dengan memiliki sumber daya yang baik akan menunjang seluruh kegiatan operasional pembelajaran serta akan menunjang semua kegiatan dalam proses pengambilan keputusan.

Hasil evaluasi dan penilaian terhadap siswa dilakukan oleh tenaga pengajar atau guru dalam melakukan penilaian selama proses pembelajaran. Dalam proses pembelajaran terdapat 3 kategori penilaian yaitu nilai siswa, disiplin, serta sikap. Kategori nilai siswa didapat dari nilai mata pelajaran teori maupun praktek serta kehadiran siswa. Kategori disiplin dan sikap didapat dari penilaian individu perilaku siswa selama proses pembelajaran disekolah. Proses penggalian informasi berdasarkan data siswa terus bertambah setiap tahunnya diiringi dengan bertambahnya jumlah siswa, sehingga terjadi penumpukan data yang sangat besar dari data yang belum diolah secara optimal oleh pihak sekolah. Beberapa metode dalam mengolah data menjadi sebuah informasi dan pengetahuan menjadi bahan pertimbangan pihak sekolah dalam menentukan sebuah kebijakan terhadap pengelompokan prestasi tinggi dan rendah dari siswa, sehingga dengan prestasi yang tinggi dapat meningkatkan potensi kelulusan siswa.

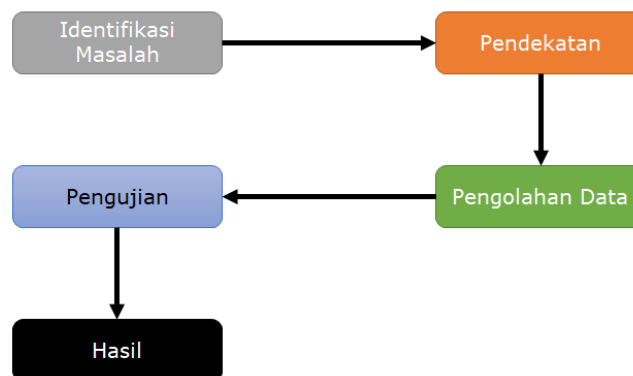
Data *mining* merupakan sebuah metode dalam bidang ilmu komputer yang digunakan dalam mencari pengetahuan dari data sehingga menjadi sebuah informasi yang bermanfaat[3]–[5]. Tahapan dalam proses data *mining* berguna untuk mencari sebuah pola tertentu dari data penilaian yang sangat banyak[6]. Bidang pendidikan pemanfaatan data mining dapat sebagai bahan evaluasi dari proses pembelajaran siswa, hasil dari data mining dapat menjadi bahan evaluasi instansi sekolah dalam proses apa yang akan diajarkan serta bagaimana cara mengajar yang baik yang dilakukan oleh guru.

K-Means clustering merupakan sebuah algoritma *unsupervised learning* yang digunakan dalam pengelompokan data dalam *dataset* yang tidak memiliki label kedalam sebuah *cluster-cluster* yang berbeda[7]–[9]. *K-Means clustering* memungkinkan pengguna melakukan pengelompokan data kedalam cluster berdasarkan *variabel-variabel* yang ada tanpa harus melalui proses *training* data terlebih dahulu. Tujuan utama dari algoritma *K-Means clustering* ini untuk meminimalisir jarak antara titik data dengan cluster yang sesuai.

Tujuan penelitian ini yaitu mengetahui dan membentuk cluster data siswa berdasarkan nilai akademik, nilai sikap, dan nilai disiplin sehingga menjadi sebuah cluster sehingga hasil cluster siswa dapat menjadi acuan dalam meningkatkan nilai siswa dalam proses pembelajaran selanjutnya.

2. METODE PENELITIAN

Tahapan penelitian merupakan sebuah metode yang digunakan untuk mendeskripsikan langkah-langkah yang dilakukan mulai dari identifikasi masalah sampai mendapatkan sebuah solusi dari permasalahan yang ada[10], [11]. Tahapan penelitian yang dilakukan dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Penelitian

Penjelasan tahapan penelitian yang dilakukan adalah

1. Identifikasi Masalah
Tahap mengidentifikasi masalah yang terjadi yaitu kurangnya informasi secara detail dan tepat mengenai pengelompokan data nilai siswa berdasarkan hasil penilaian proses pembelajaran yang dilakukan siswa setiap semester.
2. Pendekatan
Tahapan pendekatan menggunakan metode *clustering* dengan menggunakan algoritma *K-Means* data pengelompokan data nilai siswa berdasarkan hasil penilaian proses pembelajaran untuk mengetahui kelompok atau *atribut/variabel* apa saja yang berpengaruh di data penilaian siswa.
3. Pengolahan Data
Tahapan pengolahan data yaitu menyiapkan data nilai siswa yang akan digunakan, selanjutnya menentukan jumlah *cluster* yang akan digunakan. Jumlah cluster yang

digunakan yaitu dua *cluster*. Terakhir menghitung masing-masing kriteria untuk pengelompokan *cluster*.

4. Pengujian

Tahapan pengujian yang dilakukan dalam penelitian ini data yang di uji telah divalidasi atau ditentukan dan nilai data tidak berubah. Pengujian data dilakukan menggunakan metode *Silhoutte* serta *Elbow Method* untuk melihat seberapa baiknya *clustering* yang sudah dilakukan dengan menghitung kuantitas dan fitur turunan dari set data[12]-[15].

5. Hasil

Hasil penelitian ini berupa pengelompokan data nilai siswa berdasarkan hasil penilaian proses pembelajaran menggunakan algoritma *K-Means*.

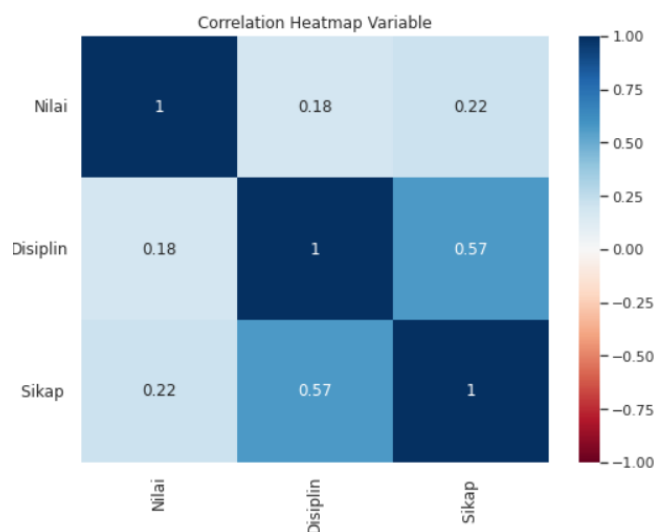
3. HASIL DAN PEMBAHASAN

Data yang diolah dalam penelitian ini merupakan sampel yang diambil dari data nilai semester siswa SMA XYZ kelas X tahun ajaran 2022. Dataset siswa terdiri dari atribut NIS, Nama, Nilai, Disiplin dan Sikap. Data sampel yang akan diuji cobakan terdiri dari 155 data peserta didik. Contoh data siswa yang digunakan dapat dilihat pada tabel 1.

Tabel 1. Dataset Siswa

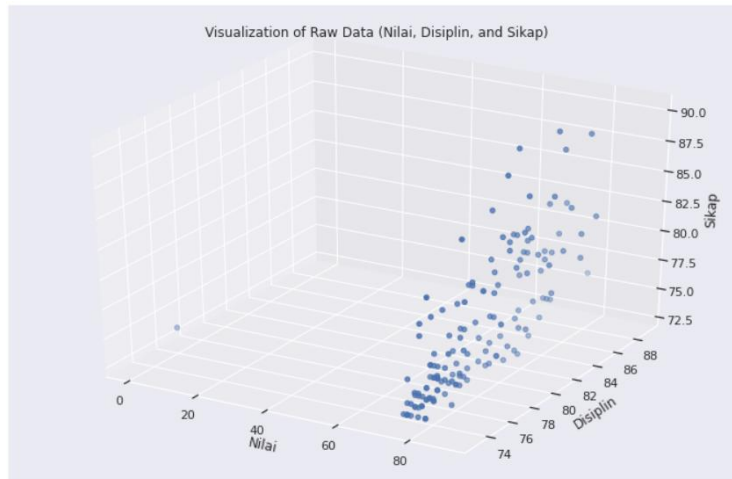
Nama	Nilai	Disiplin	Sikap
Abdullah Ibnu Aziz Hsb	87	82	81
Adinda Cynthia Salsabila	77	85	80
Adrian Bintang Mahendra	81	81	74
Afanda Dicki Apriyando	76	77	76
Afreza Peneduh	76	74	73
Ahmad Riyadi	78	83	82
Ainul Khusna Ramadhani	78	75	76
Aisyah Kamelia	76	76	80
Al Raneb	81	73	83
Aldila Wati	77	82	83

Berdasarkan dataset yang ada selanjutnya menggunakan *tools python dengan google colab* memasukan dataset siswa dan melihat korelasi antar 3 variabel yang digunakan, hasil korelasi variabel yang ada dapat dilihat pada Gambar 2.



Gambar 2. Hasil Korelasi Variabel

Selanjutnya melakukan visualisasi dataset siswa menggunakan *python* untuk melihat sebaran data siswa. Hasil visualisasi dataset siswa dapat dilihat pada Gambar 3.



Gambar 3. Hasil Visualisasi Dataset Siswa

Selanjutnya membentuk cluster yang diinginkan dari hasil visualisasi data siswa berdasarkan dataset siswa. Hasil pembentukan cluster menggunakan *python* dapat dilihat pada Gambar 4.

	Nilai	Disiplin	Sikap
Segment K-means			
0	80.101695	82.508475	81.355932
1	78.159574	75.723404	75.085106
2	0.000000	76.000000	74.000000

Gambar 4. Jumlah Cluster Yang Dibuat

Setelah jumlah cluster ditentukan, tahapan selanjutnya membuat kelompok dan cluster berdasarkan dataset siswa sehingga dataset siswa yang ada membentuk 3 cluster yang telah ditentukan. Hasil visualisasi cluster data siswa dapat dilihat pada Gambar 5.

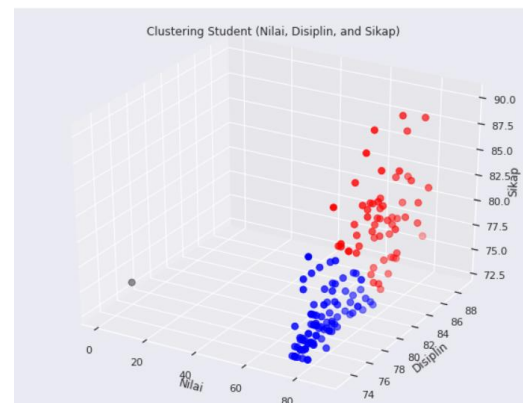
```
fig = plt.figure(figsize =(10,8))
ax = fig.add_subplot(projection='3d')

xrank = df_segm_kmeans['Nilai']
yrank = df_segm_kmeans['Disiplin']
zrank = df_segm_kmeans['Sikap ' ]

colors = {0:'red', 1:'blue', 2:'black'}
centers = np.array(km4.cluster_centers_)

ax.scatter(xrank, yrank,zrank, c=Cluster.map(colors), s=50)

ax.set_xlabel('Nilai')
ax.set_ylabel('Disiplin')
ax.set_zlabel('Sikap')
plt.title('Clustering Student (Nilai, Disiplin, and Sikap)')
plt.show()
```



Gambar 5. Hasil Cluster Data Siswa

Berdasarkan hasil cluster data siswa maka didapatkan cluster 0 berjumlah 59 siswa, cluster 1 berjumlah 94 siswa, dan cluster 2 berjumlah 1 siswa. Hasil cluster menggunakan python dapat dilihat pada Gambar 6.

	Nilai	Disiplin	Sikap	Cluster	N Obs
Segment K-means					
0	80.101695	82.508475	81.355932	0.0	59
1	78.159574	75.723404	75.085106	1.0	94
2	0.000000	76.000000	74.000000	2.0	1

Gambar 6. Hasil Cluster Jumlah Siswa

Hasil cluster 0 yang dilabel dengan nama siswa rajin berjumlah 59 siswa, berikut hasil cluster siswa dengan label siswa rajin seperti gambar 7.

NAMA	Disiplin	Sikap	Cluster K-means	Cluster
ABDULLAH IBNU AZIZ HSB	82	81		0 Rajin
Adinda Cynthia Salsabila	85	80		0 Rajin
Ahmad riyadi	83	82		0 Rajin
ALDILA WATI	82	83		0 Rajin
Aryaka Eko Syaputra	81	83		0 Rajin
ASHILAH AYU PRASOJO	83	86		0 Rajin
Attar Fadli Mawla	80	80		0 Rajin
Aulia Puspita Sari	82	90		0 Rajin
AZRIL TRIFAM BUDI	85	82		0 Rajin
Bimo Maulana	83	81		0 Rajin
Calista Prisilla Debby Irawan	84	75		0 Rajin
DENI FAHRI AKBAR	82	83		0 Rajin
DWI AMANDA LESTARI	82	76		0 Rajin
DWY ZAHRA RAMADHANI	83	83		0 Rajin
El Fazanisa Kinasih	80	84		0 Rajin
FAITH MIRZA SYARIF	89	76		0 Rajin
Faqih Achmad ridha	87	83		0 Rajin
Farah Aqeela Ramadhani	81	81		0 Rajin
HIZBUL WATHAN METEORA SAFANA	87	89		0 Rajin
INTAN NUR AINI	79	80		0 Rajin
Jenifer Alexandria	76	86		0 Rajin
Kanesya Putri Dermawan	81	84		0 Rajin
Kharida Nur Maiza	79	80		0 Rajin
M. Firdaus Efendi	82	83		0 Rajin
M. Rizky Al Miski	83	76		0 Rajin

Gambar 7. Hasil Cluster 0 Dengan Label Siswa Rajin

Hasil cluster 1 yang dilabel dengan nama siswa sedang berjumlah 94 siswa, berikut hasil cluster siswa dengan label siswa sedang seperti gambar 8.

NAMA	Disiplin	Sikap	Cluster K-means	Cluster
Adrian Bintang Mahendra	81	74		1 Sedang
Afanda Dicki Apriyando	77	76		1 Sedang
AFREZA PENEDUH	74	73		1 Sedang
Ainul Khusna Ramadhani	75	76		1 Sedang
AISYAH KAMELIA	76	80		1 Sedang
Al Raneb	73	83		1 Sedang
Aldo Hartadiansyah	77	74		1 Sedang
Amelinda Fadilah Aina Nugroho	76	73		1 Sedang
ANISA MUTIARA	75	75		1 Sedang
Arva Wicasono	73	73		1 Sedang
ASMA WATI	75	73		1 Sedang
Aura Zasqya Shallyha	75	75		1 Sedang
Avisesa Aquino	74	74		1 Sedang
Bernadeta Zora Valentina	76	75		1 Sedang
BIMA FADLU RAMADHAN	79	74		1 Sedang
CHAILA AMELIA	75	74		1 Sedang
CHALISTA KAMILA SIDABALOK	76	74		1 Sedang
CHELSY AULIA	75	75		1 Sedang
DENIS IRAWAN	75	73		1 Sedang
Desita Setyowati	80	75		1 Sedang
EZA ADE ANSYAH	76	78		1 Sedang
Fahmi	76	78		1 Sedang
FARREL RAMA ANDREAN	75	73		1 Sedang
Fiqih Alzier Dianis	75	73		1 Sedang

Gambar 8. Hasil Cluster 1 Dengan Label Siswa Sedang

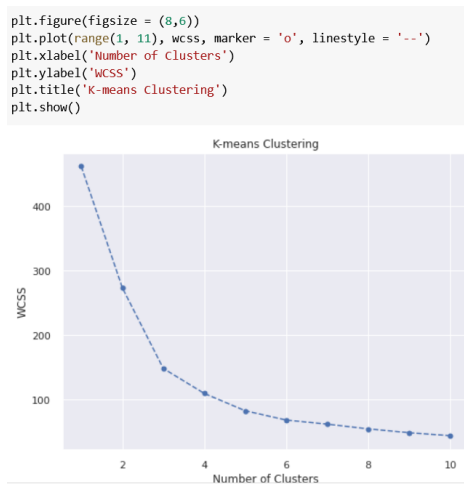
Hasil cluster 2 yang dilabel dengan nama siswa tidak rajin berjumlah 1 siswa, berikut hasil cluster siswa dengan label siswa tidak rajin seperti gambar 9.

NAMA	Disiplin	Sikap	Cluster K-means	Cluster
NOVIANA BERNADETH MANIHURUK	76	74		2 Tidak Rajin

Gambar 9. Hasil Cluster 2 Dengan Label Siswa Tidak Rajin

Hasil Pengujian

Tahapan pengujian ini penentuan *cluster* terbaik pada *K-Means* dilakukan dengan menggunakan *Elbow Methode* dengan cara menghitung nilai *Sum Square Error* (SSE) kemudian memvisualisasikannya dalam bentuk grafik dimana hasil perhitungan SSE digambarkan dalam bentuk siku lalu nilai yang memiliki penurunan secara drastis merupakan jumlah *K* yang optimal.



Gambar 10. Hasil Pengujian *Elbow Method*

Berdasarkan gambar hasil pengujian menggunakan *elbow method* maka jumlah cluster yang baik yang digunakan adalah 3 cluster, sehingga dalam penelitian ini menggunakan 3 cluster yaitu cluster 0, cluster 1, dan cluster 2.

Hasil pengujian menggunakan *silhouette coefficient* dari jumlah *cluster* yang terbaik ditunjukkan dengan nilai *Silhouette* yang semakin mendekati 1. Hasil yang didapat dari nilai cluster menggunakan *silhouette coefficient* dapat dilihat pada gambar 11 dibawah ini.

```
from sklearn.metrics import silhouette_score
from sklearn.cluster import KMeans

for n_cluster in range(3, 11):
    kmeans = KMeans(n_clusters=n_cluster).fit(X)
    label = kmeans.labels_
    sil_coeff = silhouette_score(X, label, metric='euclidean')
    print("For n_clusters={}, The Silhouette Coefficient is {}".format(n_cluster, sil_coeff))

For n_clusters=3, The Silhouette Coefficient is 0.4893575334428582
For n_clusters=4, The Silhouette Coefficient is 0.4832492381375905
For n_clusters=5, The Silhouette Coefficient is 0.3927416368504594
For n_clusters=6, The Silhouette Coefficient is 0.43783564449589696
For n_clusters=7, The Silhouette Coefficient is 0.44447880801160755
For n_clusters=8, The Silhouette Coefficient is 0.39321471294943483
For n_clusters=9, The Silhouette Coefficient is 0.38348523648658134
For n_clusters=10, The Silhouette Coefficient is 0.3834447274075178
```

Gambar 10. Hasil Pengujian *Silhouette Coefficient*

Berdasarkan gambar hasil pengujian menggunakan *silhouette coefficient* maka jumlah cluster yang baik yang digunakan adalah 3 cluster dengan nilai *silhouette coefficient* yaitu 0.489, dan lebih baik dari nilai *silhouette coefficient cluster* lainnya.

4. KESIMPULAN

Hasil pembahasan pengelompokan data nilai siswa menggunakan metode K-Means clustering menunjukkan bahwa berdasarkan hasil cluster data siswa menggunakan dataset siswa dalam satu semester, maka didapatkan cluster 0 berjumlah 59 siswa, cluster 1 berjumlah 94 siswa, dan cluster 2 berjumlah 1 siswa. Hasil pengujian menggunakan *elbow method* maka jumlah cluster yang baik yang digunakan adalah 3 cluster, sehingga dalam penelitian ini menggunakan 3 cluster yaitu cluster 0, cluster 1, dan cluster 2. Hasil pengujian menggunakan *silhouette coefficient* maka jumlah cluster yang baik yang digunakan adalah 3 cluster dengan nilai *silhouette coefficient* yaitu 0.489, dan lebih baik dari nilai *silhouette coefficient cluster* lainnya.

5. REFERENCES

- [1] H. Sulistiani, S. Setiawansyah, and D. Darwis, "Penerapan Metode Agile untuk Pengembangan Online Analytical Processing (OLAP) pada Data Penjualan (Studi Kasus: CV Adilia Lestari)," *J. CoreIT J. Has. Penelit. Ilmu Komput. dan Teknol. Inf.*, vol. 6, no. 1, pp. 50–56, 2020.
- [2] A. Aldino, A. Saputra, A. Nurkholis, and S. Setiawansyah, "Application of Support Vector Machine (SVM) Algorithm in Classification of Low-Cape Communities in Lampung Timur," *Build. Informatics, Technol. Sci.*, vol. 3, no. 3 SE-Articles, Dec. 2021, doi: 10.47065/bits.v3i3.1041.
- [3] N. Nurdin and D. Astika, "Penerapan Data Mining Untuk Menganalisis Penjualan Barang Dengan Menggunakan Metode Apriori Pada Supermarket Sejahtera Lhokseumawe," *TECHSI-Jurnal Tek. Inform.*, vol. 7, no. 1, pp. 132–155, 2019.
- [4] M. J. Zaki and W. J. Meira, *Data Mining and Machine Learning Fundamental Concepts and Algorithms*, vol. 53, no. 9. 2020.
- [5] C. Rygielski, J.-C. Wang, and D. C. Yen, "Data mining techniques for customer relationship management," *Technol. Soc.*, vol. 24, no. 4, pp. 483–502, 2002.
- [6] M. S. Sungkar and M. T. Qurohman, "Penerapan Algoritma C5.0 Untuk Prediksi Kelulusan Pembelajaran Mahasiswa Pada Matakuliah Arsitektur Sistem Komputer," *J. Media Inform. Budidarma*, vol. 5, no. 3, p. 1166, 2021, doi: 10.30865/mib.v5i3.3116.
- [7] E. Muningsih, H. M. Nur, F. F. Dwi Imaniawan, Saifudin, V. R. Handayani, and F. Endiarto, "Comparative Analysis on Dimension Reduction Algorithm of Principal Component Analysis and Singular Value Decomposition for Clustering," *J. Phys. Conf. Ser.*, vol. 1641, no. 1, 2020, doi: 10.1088/1742-6596/1641/1/012101.
- [8] N. Noviyanto, "Penerapan Data Mining dalam Mengelompokkan Jumlah Kematian Penderita COVID-19 Berdasarkan Negara di Benua Asia," *Paradig. - J. Komput. dan Inform.*, vol. 22, no. 2, pp. 183–188, 2020, doi: 10.31294/p.v22i2.8808.
- [9] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [10] M. N. D. Satria, "Application of SAW in the Class Leader Selection Decision Support System," *Chain J. Comput. Technol. Comput. Eng. Informatics*, vol. 1, no. 1, pp. 27–31, 2023.
- [11] S. Maryana and D. Suhartini, "Implementasi Certainty Factor Untuk Diagnosa Penyakit Sapi," *Chain J. Comput. Technol. Comput. Eng. Informatics*, vol. 1, no. 1, pp. 14–20, 2023.
- [12] M. Cui, "Introduction to the k-means clustering algorithm based on the elbow

- method," *Accounting, Audit. Financ.*, vol. 1, no. 1, pp. 5–8, 2020.
- [13] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration k-means clustering method and elbow method for identification of the best customer profile cluster," in *IOP conference series: materials science and engineering*, 2018, vol. 336, no. 1, p. 12017.
- [14] 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)*, 2020, pp. 747–748.
- [15] F. Batool and C. Hennig, "Clustering with the average silhouette width," *Comput. Stat. Data Anal.*, vol. 158, p. 107190, 2021.