



KLASTERISASI PENDIDIKAN SD UNTUK MENGETAHUI DAERAH DENGAN PENDIDIKAN TERENDAH MENGGUNAKAN ALGORITMA K-MEANS

Kevin Riyas Robbani^{a*}, Zaehol Fatah^b

^a Sains dan Teknologi / Sistem Informasi, kevinriyasrobbani003@gmail.com, Universitas Ibrahimy Jawa Timur

^b Sains dan Teknologi / Sistem Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy Jawa Timur

* Korespondensi

ABSTRACT

Elementary education serves as the foundational stage in efforts to improve the overall quality of education in Indonesia. Identifying regions with the lowest levels of elementary education is essential for effectively targeting initiatives to enhance education quality. The K-Means clustering algorithm is employed to group regions based on specific indicators, such as the number of students, dropout rates, classrooms, teaching staff, school principals, and others. The objective of this method is to identify regions with the lowest levels of elementary education by pinpointing clusters of areas that require the most support and development. K-Means clustering operates by dividing data into several clusters based on the similarity of feature patterns. This process facilitates the identification of regional groups with varying priorities for support and development. The clustering analysis results reveal that from 39 datasets related to elementary education across various regions in Indonesia, three clusters were formed. Cluster 0 consists of 34 data points, Cluster 1 contains only 1 data point, and Cluster 2 comprises 4 data points.

Keywords: *K-Means clustering algorithm, elementary school education, data clustering*

Abstrak

Pendidikan di tingkat sekolah dasar (SD) berperan sebagai fondasi awal yang mendukung upaya peningkatan kualitas pendidikan di Indonesia. Mengetahui daerah dengan pendidikan SD terendah diperlukan untuk lebih tepat sasaran dalam meningkatkan kualitas pendidikan SD. Algoritma K-Means clustering digunakan untuk mengelompokkan daerah pendidikan SD berdasarkan indikator tertentu, seperti jumlah siswa, siswa putus sekolah, ruang kelas, jumlah tenaga pengajar, kepala sekolah dan sebagainya. Tujuan dari metode ini untuk mengetahui daerah dengan pendidikan SD terendah dengan mengidentifikasi kelompok daerah yang paling membutuhkan dukungan dan pengembangan. K-Means clustering bekerja dengan membagi data menjadi beberapa cluster tergantung pada kesamaan pola fitur yang digunakan. Proses ini memungkinkan untuk mengidentifikasi kelompok daerah dengan prioritas dukungan dan pengembangan yang berbeda. Hasil analisis clustering menunjukkan bahwa dari 39 dataset terkait jumlah pendidikan SD di berbagai wilayah Indonesia, terbentuk tiga cluster. Cluster 0 mencakup 34 data, cluster 1 hanya memiliki 1 data, sedangkan cluster 2 terdiri dari 4 data.

Kata Kunci: algoritma K-Means clustering, pendidikan sd, pengelompokan data

1. PENDAHULUAN

Pendidikan merupakan salah satu aspek penting dalam kehidupan manusia yang memiliki peran besar dalam membentuk sikap dan perilaku sehari-hari. Melalui pendidikan, seseorang dapat meraih tujuan hidupnya. Hal ini disebabkan oleh kondisi manusia yang sejak lahir berada dalam keadaan lemah dan bergantung pada orang lain untuk memenuhi kebutuhannya serta mencapai aspirasinya.[1]

Pendidikan di tingkat sekolah dasar (SD) berperan sebagai fondasi awal dalam upaya meningkatkan mutu pendidikan di Indonesia. SD merupakan jenjang pendidikan yang menyelenggarakan proses pembelajaran

dasar dan menjadi landasan bagi pendidikan pada tingkat berikutnya. Tujuannya adalah memberikan siswa kemampuan dasar, seperti membaca, menulis, berhitung, serta pengetahuan dan keterampilan dasar yang relevan dengan tahap perkembangan mereka.[2].

Rendahnya tingkat pendidikan tidak hanya berdampak pada individu, tetapi juga memiliki konsekuensi yang luas bagi masyarakat dan negara[3]. Langkah untuk meningkatkan mutu pendidikan menjadi hal yang krusial dalam mengatasi berbagai permasalahan ini.

Salah satu metode yang bisa diterapkan untuk mengidentifikasi dan mengelompokkan daerah berdasarkan tingkat pendidikan adalah dengan menggunakan algoritma K-Means. K-Means merupakan algoritma yang tergolong dalam kategori *unsupervised learning* dan digunakan untuk mengelompokkan data ke dalam beberapa cluster tanpa memerlukan label kategori [4]. Algoritma ini mengelompokkan data berdasarkan kemiripan yang terdapat dalam dataset.

Dengan menerapkan algoritma K-Means, pemerintah dan pemangku kepentingan di bidang pendidikan dapat memperoleh gambaran yang lebih jelas mengenai distribusi kualitas pendidikan di berbagai daerah. Hal ini dapat membantu untuk merancang kebijakan yang lebih tepat sasaran dalam meningkatkan kualitas pendidikan, terutama di daerah-daerah yang teridentifikasi memiliki tingkat pendidikan yang rendah. Penelitian ini bertujuan untuk menggunakan algoritma K-Means dalam proses pengelompokan data pendidikan sekolah dasar guna mengetahui daerah dengan pendidikan terendah, sehingga dapat menjadi dasar bagi pengambilan keputusan sebagai langkah untuk meningkatkan kualitas pendidikan di Indonesia.

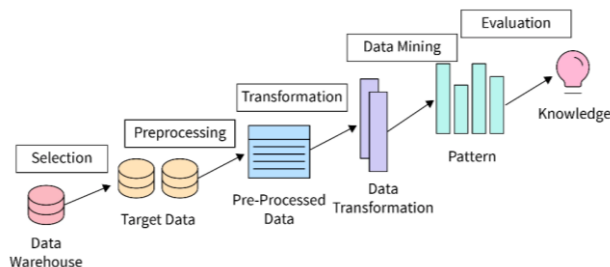
2. METODOLOGI PENELITIAN

2.1. Data Mining

Data mining adalah alat yang memungkinkan pengguna untuk dengan cepat mengakses data dalam jumlah besar. Proses ini melibatkan ekstraksi atau penggalian informasi penting dan belum diketahui sebelumnya dari database yang besar. Informasi tersebut kemudian dapat dipahami, dimanfaatkan, dan digunakan untuk mendukung pengambilan keputusan bisnis yang signifikan[5].

2.2. Knowledge Discovery In Database

Dalam bidang penelitian computing, Knowledge Discovery in Database (KDD) merupakan salah satu metode yang dapat diterapkan[6]. Proses KDD terdiri dari lima tahapan, yaitu pemilihan data dari sumber ke data target, tahap pra-pemrosesan, transformasi, penambangan data (data mining), dan tahap evaluasi[7]. Tahap seleksi berfokus pada penentuan data yang akan digunakan dalam penelitian. Pada tahap pra-pemrosesan, dilakukan pengintegrasian atau penggabungan data serta pembersihan data, termasuk menghilangkan noise, data redundan, inkonsistensi, dan data yang tidak relevan. Tahap transformasi bertujuan untuk menggabungkan data dan menyesuaikan formatnya agar siap diproses pada tahap penambangan data (data mining). Pada tahap data mining, algoritma yang sesuai dengan permasalahan data diterapkan. Terakhir, tahap evaluasi digunakan untuk menguji hasil pada data yang telah diproses.



Gambar 1. Tahapan KDD

2.3. Clustering

Clustering adalah metode yang digunakan untuk menemukan dan mengelompokkan data berdasarkan kemiripan karakteristik antar data. Tujuan utamanya adalah mengelompokkan sejumlah data atau objek ke dalam cluster sehingga setiap cluster berisi data yang memiliki kemiripan tinggi. Metode ini berupaya menempatkan objek yang serupa dalam satu cluster, sambil menjaga jarak antar cluster sejauh mungkin. Dengan demikian, objek dalam satu cluster memiliki kemiripan yang tinggi satu sama lain, sementara objek di cluster lain memiliki perbedaan yang signifikan [8]. Clustering juga sangat berguna untuk mengidentifikasi cluster yang sebelumnya tidak dikenali dalam data.

2.4. Algoritma K-Means

K-Means adalah algoritma yang diawali dengan menentukan titik pusat awal sebagai pusat sementara dari centroid atau cluster. Selanjutnya, algoritma ini menghitung jarak setiap data ke pusat menggunakan rumus tertentu. Data yang memiliki jarak lebih dekat ke pusat akan dikelompokkan bersama, sedangkan data yang lebih jauh akan dimasukkan ke kelompok lain[9]. Algoritma K-Means mempunyai kelebihan mampu mengelompokkan objek besar serta meningkatkan kecepatan proses pengelompokan[10].

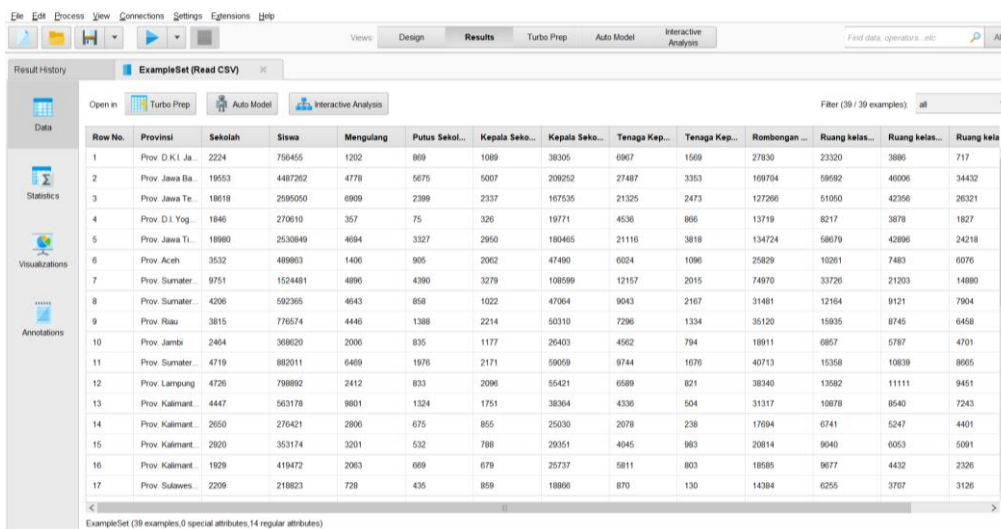
3. HASIL DAN PEMBAHASAN

Penelitian ini menerapkan algoritma K-Means clustering menggunakan alat RapidMiner Studio. RapidMiner mempermudah penggunaannya dalam menjalankan berbagai operator. Operator berfungsi untuk memodifikasi data, yang kemudian dihubungkan dengan node-node operator lainnya.

Hasil dari proses penelitian menunjukkan bahwa terdapat 39 dataset yang berkaitan dengan jumlah sekolah dasar (SD) di setiap wilayah di Indonesia. Cluster 0 berisi 34 data. Cluster ini merupakan kelompok terbesar dan mungkin mencerminkan karakteristik umum dari siswa dengan pendidikan SD terendah. Data dalam cluster ini bisa jadi menunjukkan kesamaan dalam faktor-faktor seperti latar belakang ekonomi, akses pendidikan, atau hasil belajar yang rendah. Cluster 1: Berisi 1 data. Cluster ini sangat kecil dan mungkin mewakili kasus unik atau outlier. Data ini bisa jadi menunjukkan kondisi yang sangat berbeda dibandingkan dengan cluster lainnya, seperti siswa yang memiliki latar belakang yang sangat berbeda atau yang mengalami situasi khusus yang mempengaruhi pendidikan. Cluster 2: Berisi 4 data. Cluster ini juga kecil, tetapi lebih besar dari cluster 1. Data dalam cluster ini mungkin menunjukkan siswa yang memiliki beberapa kesamaan, tetapi tidak cukup untuk membentuk kelompok yang lebih besar. Ini bisa mencakup siswa dengan akses pendidikan yang sedikit lebih baik dibandingkan dengan cluster 0, tetapi masih tergolong dalam kategori pendidikan SD terendah.

3.1. Dataset

Dataset yang digunakan menggunakan sumber data terbuka, atau open source, yang dapat diakses melalui situs web Kaggle.com. Kaggle adalah platform terkemuka yang memungkinkan akses ke berbagai jenis dataset untuk penelitian, pembelajaran mesin, dan proyek data science lainnya. Dalam penelitian ini menggunakan dataset kelayakan pendidikan Indonesia, yang mencakup informasi tentang masalah pelayanan pendidikan SD di berbagai macam daerah di Indonesia.



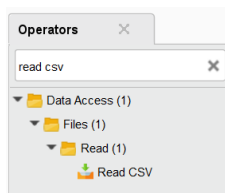
Row No.	Provinsi	Sekolah	Siswa	Mengulang	Pulus Sekol...	Kepala Seko...	Kepala Seko...	Tenaga Kep...	Tenaga Kep...	Rombongan ...	Ruang kelas...	Ruang kelas...	Ruang kela
1	Prov. D.K.I. Ja.	2224	750455	1202	869	1089	30305	1569	27830	23320	3690	717	
2	Prov. Jawa Ba.	19553	4487262	4778	6675	5007	209252	27487	3353	109704	99592	46006	34432
3	Prov. Jawa Te.	18018	2595050	6909	2399	2337	167535	21325	2473	127266	51050	42356	26321
4	Prov. D.I. Yog.	1846	270610	357	75	326	19771	4536	866	13719	8217	3878	1827
5	Prov. Jawa Ti.	18980	2530849	4094	3327	2950	180465	21110	3818	134724	58679	42896	24218
6	Prov. Aceh	3532	489803	1406	905	2062	47490	6024	1096	25829	10261	7483	6076
7	Prov. Sumate	9751	1524481	4896	4390	3279	108599	12157	2015	74970	33726	21203	14880
8	Prov. Sumate...	4206	592365	4643	858	1022	47064	9043	2167	31481	12164	9121	7904
9	Prov. Riau	3815	776574	4446	1388	2214	50310	7296	1334	35120	15935	8745	6458
10	Prov. Jamba	2464	308620	2006	835	1177	26403	4562	794	18911	6857	5787	4701
11	Prov. Sumate	4719	882011	6469	1976	2171	59059	9744	1676	40713	15358	10639	8605
12	Prov. Lampung	4726	798892	2412	833	2096	55421	6589	821	38340	13582	11111	9451
13	Prov. Kalimantan	4447	563178	9801	1324	1751	38364	4336	504	31317	10878	8540	7243
14	Prov. Kalimantan	2650	276421	2806	675	855	25030	2079	238	17694	6741	5247	4401
15	Prov. Kalimantan	2920	353174	3201	532	788	29351	4045	963	20814	9040	6053	5091
16	Prov. Kalimantan	1929	419472	2063	689	679	25737	5811	803	18585	9677	4432	2326
17	Prov. Sulawesi	2209	218823	728	435	859	18866	870	130	14384	6255	3707	3126

Gambar 1. Dataset

3.2. Pengolahan Data dengan software Rapidminer

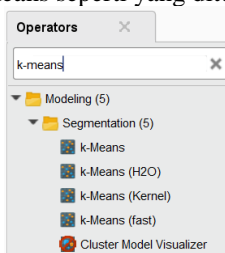
Proses pengelompokan data dalam penelitian ini dilakukan menggunakan perangkat lunak RapidMiner versi 10.3 dengan langkah-langkah sebagai berikut:

- Jalankan perangkat lunak RapidMiner Studio versi 10.3.
- Selanjutnya, buka menu file dan pilih opsi "new process."
- Cari operator "read csv" di panel operator, lalu klik dua kali atau seret operator tersebut ke panel proses. Hal ini memungkinkan operator "read csv" muncul di panel proses dan digunakan untuk mengimpor data CSV yang telah melewati tahap pembersihan (data cleaning) dan transformasi.



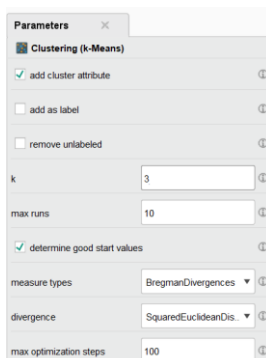
Gambar 2. Read CSV

- d. Lakukan proses clustering data mining menggunakan algoritma K-Means dengan cara membuka menu operator, kemudian memilih opsi K-Means seperti yang ditunjukkan pada gambar 4 berikut:



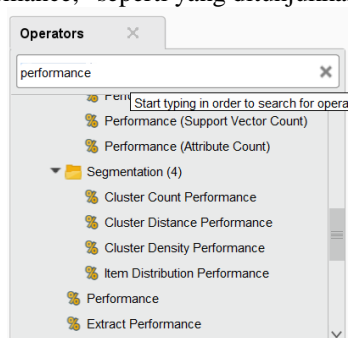
Gambar 3. Memilih Algoritma K-Means

- e. Tentukan jumlah cluster (k) yang akan digunakan, misalnya k=3, seperti yang ditampilkan pada gambar 5 berikut:



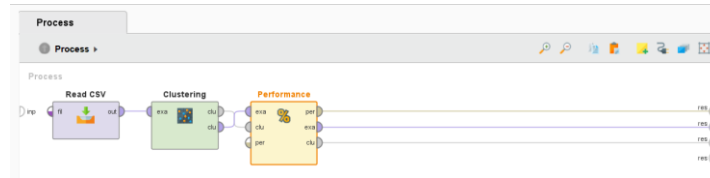
Gambar 4. Menentukan Nilai K

- f. Masukkan model untuk mengevaluasi kualitas hasil clustering dengan membuka menu operator, lalu pilih opsi "cluster distance performance," seperti yang ditunjukkan pada gambar 6 berikut:



Gambar 5. Cluster Distance Performance

- g. Sambungkan konektor dari setiap proses ke main process seperti yang ditampilkan pada gambar 7, lalu klik tombol "run" untuk menjalankan proses tersebut.



Gambar 6. Main Proses Data Mining

3.3. Hasil Klasterisasi dengan Algoritma K-Means

Setelah melalui langkah-langkah proses clustering K-Means dengan aplikasi Rapidminer Studio 10.3, proses selanjutnya adalah melihat hasilnya. Hasil clustering yang diperoleh akan ditampilkan seperti yang terlihat pada Gambar 8 berikut:

Gambar 7. Tampilan Hasil Cluster Pada Data View

Cluster Model

Cluster 0: 34 items
 Cluster 1: 1 items
 Cluster 2: 4 items
 Total number of items: 39

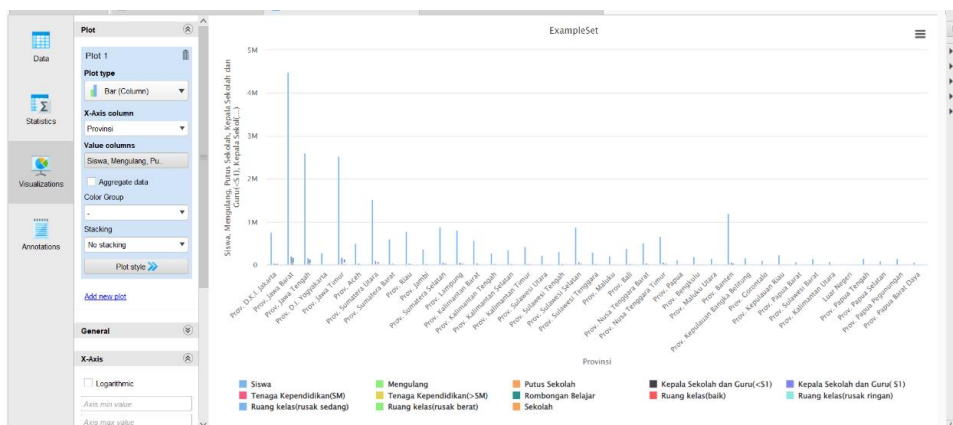
Gambar 8. Cluster Model

Hasil yang ditunjukkan pada Gambar 9 merupakan hasil pengujian dataset sebanyak 39 dataset dengan software Rapidminer 10.3, yang menghasilkan 3 cluster. Cluster 0 (cluster pertama) terdiri atas 34 data, cluster 1 (cluster kedua) berisi 1 data, dan cluster 2 (cluster ketiga) mencakup 4 data.

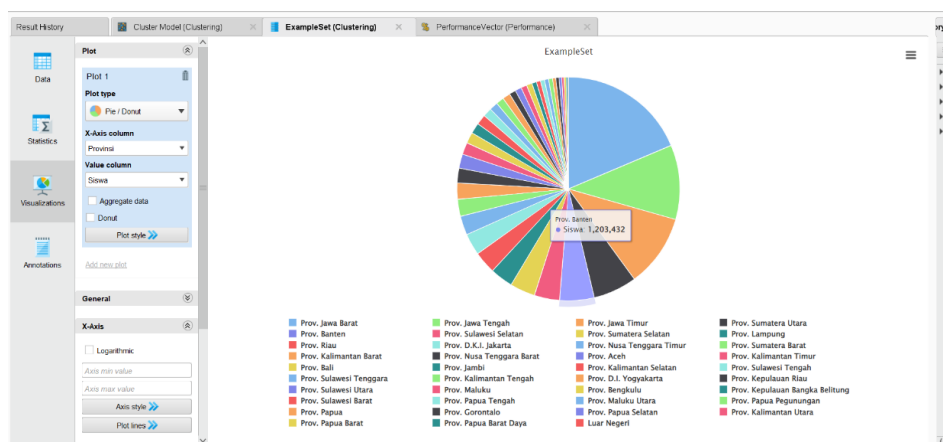
Pada gambar 10 di bawah ini merupakan gambar tampilan folder jumlah pendidikan SD di setiap daerah Indonesia:

Gambar 9. Folder View

Chart adalah representasi grafis dari hasil pengelompokan data pendidikan SD di berbagai daerah di Indonesia, menggunakan 13 variabel penelitian dan menghasilkan 3 cluster. Berikut ini adalah tampilan kolom pada chart serta diagram pie yang ditunjukkan pada Gambar 11 dan Gambar 12 di bawah ini.



Gambar 10. Chart



Gambar 11. Pie

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian ini, dapat disimpulkan bahwa penerapan metode algoritma K-Means Clustering berhasil digunakan untuk mengidentifikasi daerah dengan tingkat pendidikan terendah. Pengelompokan ini memberikan wawasan penting tentang kondisi pendidikan di tingkat SD. Dengan memahami karakteristik masing-masing cluster, dapat merancang intervensi yang lebih tepat sasaran untuk meningkatkan pendidikan di cluster 0 dengan jumlah 34 data, yang merupakan kelompok terbesar. Selain itu, cluster 1, yang hanya memiliki 1 data, memerlukan perhatian khusus. Cluster 2, yang mencakup 4 data, harus dipastikan tidak diabaikan dan menerima dukungan yang diperlukan. Melalui analisis ini, kebutuhan spesifik dari setiap kelompok dapat diidentifikasi, sehingga memungkinkan perumusan strategi yang lebih efektif untuk meningkatkan kualitas pendidikan di tingkat dasar.

Ucapan Terima Kasih

Ucapan terima kasih ditujukan kepada semua pihak yang telah memberikan dukungan hingga penelitian ini dapat terlaksana. Segala puji dan syukur dihaturkan kepada Allah SWT atas limpahan taufik dan hidayah-Nya, sehingga jurnal ini berhasil diselesaikan. Penghargaan sebesar-besarnya juga diberikan kepada dosen pembimbing atas bimbingan dan kesabarannya selama proses penyusunan. Dan juga terima kasih untuk segenap dukungan, khususnya kepada teman – teman yang selalu mengingatkan akan terselesaikannya jurnal ini.

DAFTAR PUSTAKA

- [1] W. R. H. Moh. Fahmi Nugraha, Budi Hendrawan, Anggia Suci Pratiwi, Rahmat Permana, Yopa Taufik Saleh, Meiliana Nurfitriani, Milah Nurkamilah, Asti Trilestari, *Pengantar Pendidikan dan Pembelajaran di Sekolah Dasar*. Edu Publisher, 2020.
- [2] U. M. Rokhmaniyah, Kartika Chrysti Suryandari, Siti Fatimah, *Anak Putus Sekolah, Dampak dan strategi mengatasinya*. CV. Pajang Putra Wijaya, 2022.
- [3] F. Cheong, “Using a Problem-Based Learning Approach to Teach an Intelligent Systems Course,” *J. Inf. Technol. Educ. Res.*, vol. 7, pp. 047–060, 2008, doi: 10.28945/178.
- [4] M. F. Amalia and D. B. Arianto, “Implementasi Algoritma K-Means Clustering Dalam Klasterisasi Kabupaten/Kota Provinsi Jawa Barat Berdasarkan Faktor Pemicu Stunting Pada Balita,” *Simkom*, vol. 9, no. 1, pp. 36–46, 2024, doi: 10.51717/simkom.v9i1.356.
- [5] I. E. Mustika, Yunita Ardilla, Abraham Manuhutu, Nazaruddin Ahmad, Imanuddin Hasbi, Guntoro, Melda Agnes Manuhutu, Mohamad Ridwan, Hozairi, Anindya Khrisna Wardhani, Syariful Alim, Ikhsan Romli, Yoga Religia, D Tri Octafian, Unggul Utan Sufandi, *DATA MINING DAN APLIKASINYA*. Penerbit Widina, 2021.
- [6] R. S. Wahono, “Research Methodology,” *brainmatics*. <http://romisatriawahono.net/rm>
- [7] Yuli Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4.5,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [8] M. Anjelita, A. P. Windarto, A. Wanto, and I. Sudahri, “Pengembangan Datamining Klastering Pada Kasus Pencemaran Lingkungan Hidup,” pp. 309–313, 2020.
- [9] W. P. Prima Dina Atika, *MODUL DATA MINING*. Fakultas Ilmu Komputer Universitas Bhayangkara, 2020.
- [10] H. Rosika *et al.*, “IMPLEMENTASI RAPIDMINER UNTUK CLESTERING DATA PENJUALAN PAKAIAN MENGGUNAKAN METODE K-MEANS,” vol. 5, pp. 221–231, 2024.