

Penerapan K-Means Clustering Dalam Pengelompokan Data Nilai Siswa Pada SMP PGRI Waingapu

(Application of K-Means Clustering in Grouping Student Value Data at SMP PGRI Waingapu)

Sonia Yina¹, Arini Aha Pekuwali²

^{1,2}Program Studi Teknik Informatika, Universitas Kristen Wira Wacana Sumba

E-mail: ¹soniayina022@gmail.com, ²arini.pekuwali@unkriswina.ac.id

KEYWORDS:

K-Means Clustering, Student Grade Data, RapidMiner.

KATA KUNCI:

K-Means Clustering, Data Nilai Siswa, RapidMiner.

ABSTRACT

Education is one of the important factors in shaping quality human resources, especially in the digital era that is full of challenges. SMP PGRI Waingapu in East Sumba Regency faces problems in managing and utilizing student grade data, which has an impact on the lack of personalized learning. This study aims to cluster seventh grade students based on knowledge scores and project scores to strengthen the Pancasila learner profile (P5) according to the independent curriculum using the k-means clustering method. Data was taken from 70 students, including grades in 10 subjects and 1 P5 project in semester 1 of the 2024/2025 school year. By using three variables, namely gender, knowledge score, and P5 project score, three clusters were produced, namely high, medium, and low achievement. The results of this clustering are expected to be a reference in applying more focused and personalized learning strategies, in order to help students reach their maximum potential.

ABSTRAK

Pendidikan ialah satu di antara faktor penting untuk membentuk SDM yang bermutu, terutama di era digital yang penuh dengan tantangan. SMP PGRI Waingapu di Kabupaten Sumba Timur menghadapi masalah dalam pengelolaan dan pemanfaatan data nilai siswa, yang berdampak pada kurangnya personalisasi pembelajaran. Penelitian ini bertujuan mengelompokkan siswa kelas VII berdasarkan nilai pengetahuan dan nilai proyek penguatan profil pelajar Pancasila (P5) sesuai kurikulum merdeka dengan metode k-means clustering. Data diambil dari 70 siswa, meliputi nilai 10 mata pelajaran dan 1 proyek P5 pada semester 1 tahun ajaran 2024/2025. Dengan menggunakan tiga variabel yaitu jenis kelamin, nilai pengetahuan, dan nilai proyek P5 dihasilkan tiga kluster yaitu prestasi tinggi, sedang, dan rendah. Hasil pengelompokan ini diharapkan menjadi acuan dalam penerapan strategi pembelajaran yang lebih terfokus dan personal, guna membantu siswa mencapai potensi maksimal.

PENDAHULUAN

Pendidikan di Indonesia terus mengalami perkembangan seiring meningkatnya kebutuhan untuk melanjutkan ke jenjang yang lebih tinggi. Peningkatan kualitas sumber daya manusia menjadi langkah awal yang krusial dalam mewujudkan berbagai tujuan. Dalam hal ini, Pendidikan memainkan peran penting sebagai pondasi utama dalam meraih cita-cita, sekaligus menjadi faktor penentu arah dan masa depan suatu bangsa [1]. Pendidikan yakni satu di antara faktor penting utama untuk membentuk SDM yang bermutu. Di era digital saat ini, tantangan yang dihadapi dunia Pendidikan semakin kompleks, khususnya dalam hal peningkatan kualitas pembelajaran dan pemanfaatan data Pendidikan. SMP PGRI Waingapu, yang berlokasi di Kelurahan

Temu, Kecamatan Kanatang, Kabupaten Sumba Timur, dengan jumlah siswa keseluruhan 165 dari kelas VII hingga IX, menghadapi beberapa masalah spesifik yang mempengaruhi kualitas pembelajaran.

Satu di antara persoalan utama di SMP PGRI Waingapu ialah belum maksimalnya pengelolaan serta pemanfaatan data nilai siswa untuk memahami keberagaman prestasi dan kebutuhan belajar siswa. Proses evaluasi yang dilakukan selama ini masih bersifat umum dan tidak memberikan wawasan yang cukup tentang kekuatan dan kelemahan individual siswa. Dampaknya adalah terbatasnya penyesuaian pembelajaran terhadap kebutuhan individu siswa, baik yang memiliki prestasi tinggi, sedang, maupun rendah [2].

Menyelesaikan persoalan ini, pemakaian metode data mining misal *K-Means Clustering* menjadi penting. Metode *data mining* digunakan dalam ilmu komputer untuk menggali pengetahuan dari data mentah dan mengubahnya menjadi informasi yang bermanfaat [3]. *Data mining* yakni proses pengumpulan, pengolahan, serta analisis data dalam jumlah besar guna menjumpai pola, korelasi, atau informasi yang relevan [4]. Salah satu metode yang dapat digunakan secara efektif untuk mengatasi permasalahan ini adalah *Clustering* dengan algoritma *K-Means* [5]. *K-Means Clustering* adalah suatu metode dalam *data mining* yang dipergunakan dalam mengelompokkan data kedalam sejumlah *cluster* sesuai kesamaan karakteristik. Metode ini bekerja melalui cara menentukan pusat (*centroid*) untuk setiap *cluster*, lalu mengelompokkan data berdasarkan jarak terdekat ke *centroid* tersebut. *Clustering* adalah suatu Teknik dalam pengelompokan beberapa data ke beberapa kelompok berdasarkan karakteristik yang identitas tidak diketahui [6]. Metode ini memberi peluang pada sekolah untuk mengelompokkan siswa sesuai dengan prestasi akademik yang tercermin dari nilai pengetahuan, serta nilai proyek penguatan profil pelajar pancasila (P5) sesuai dengan kurikulum merdeka [7]. Dengan demikian, dapat lebih mudah mengidentifikasi kebutuhan spesifik siswa dalam setiap cluster dan merancang strategi pembelajaran yang lebih efektif. Pada penerapan metode *k-means*, data yang tersedia di proses mempergunakan algoritma *k-means* untuk menghasilkan pengelompokan siswa sesuai tiga kategori prestasi : tinggi, sedang, dan cukup. Pengelompokan ini didasarkan pada nilai *indeks davies-bouldin* [8]. *Davies- Bouldin Index* merupakan salah satu metode yang sering digunakan dalam proses *clustering* untuk menilai seberapa baik hasil pengelompokan yang dilakukan serta mengevaluasi kualitas dari masing-masing *cluster* [9]. Perangkat lunak ini menyediakan berbagai operator yang sudah terintegrasi, yang menjadi komponen dasar dari alur kerja, meliputi seluruh tahap pada proses data mining misalnya pembersihan data, pemilihan fitur, serta pemodelan [10].

Studi ini bertujuan untuk mengaplikasikan *K-Means Clustering* pada data nilai siswa di SMP PGRI Waingapu guna membantu sekolah dalam mengelompokkan siswa. Besar harapannya, temuan studi ini bisa dijadikan acuan dalam menaikkan prestasi belajar siswa melalui pendekatan pembelajaran yang lebih terfokus serta personalisasi, dengan demikian tiap siswa bisa mendapatkan potensi maksimalnya.

METODE PENELITIAN

Tahapan penelitian ini meliputi sejumlah tahapan yakni metode *K-Means Clustering*. Tahapan-tahapan tersebut yaitu pada gambar 1.



Gambar 1. Metode Penelitian

- 1) Pengumpulan Data, di tahap ini dilaksanakan pengumpulan data yaitu data nilai siswa kelas VII SMP PGRI Waingapu semester 1 tahun ajaran 2024/2025. Data yang dikumpulkan

meliputi Nilai pengetahuan dari 10 mata pelajaran, nilai proyek penguatan profil pelajar Pancasila (P5).

- 2) Pra-Pemrosesan Data, tahapan ini merupakan tahapan awal yang melibatkan persiapan serta pembersihan data mentah sebelum dilaksanakan analisis lebih lanjut. Proses ini meliputi penghapusan data yang tidak sesuai atau duplikat.
- 3) Penerapan K-Means Clustering, di tahap ini, dilakukan clustering dengan mempergunakan metode clustering yang sudah dipilih yakni k-means.
- 4) Visualisasi, hasil clustering yang diperoleh divisualisasi dengan menganalisis ciri dan karakteristik tiap kelompok data yang tercipta. Dimana hasil ini disajikan melalui bentuk format tabel serta grafik.
- 5) Evaluasi, dalam tahapan ini dilakukan evaluasi guna menilai mutu tiap cluster saat mengelompokkan data ialah indeks Davies-Bouldin Index.

HASIL DAN PEMBAHASAN

1. Mengumpulkan Data

Sumber data diambil dari nilai pengetahuan dari 10 mata pelajaran dan nilai proyek penguatan profil pelajar Pancasila (P5) berdasarkan pada kurikulum merdeka. Data akademik yang diambil dalam penelitian ini adalah data nilai 1 semester, dengan 10 mata pelajaran yaitu Agama, Pancasila, Bahasa Indonesia, Matematika, Ipa, Ips, Bahasa Inggris, PJOK, Tikom, Seni Budaya, dan nilai proyek P5. Gambar berikut merupakan data akademik siswa, yang terdiri dari 70 siswa dan 11 atribut penilaian 1 semester.

Tabel 1. Dataset Mentah

N O	Nama Siswa/i	L/ P	NILAI PENGETAHUAN										
			1	2	3	4	5	6	7	8	9	10	P 5
			A G A M A	P A N C A S I L A	B I N D O	M A T E M A T I K A	I P A	I P S	B I N G	P J O K	T I K O M	S E N I B U D A	
1	Agung Punda Hudang	L	88	88	88	88	88	88	75	88	88	75	B
2	Aldiron Putra	L	88	88	88	88	88	88	78	88	88	78	B
3	Alfonsa Safira Taka	P	76	84	84	84	84	84	78	84	84	78	B
4	Anastasia Jevita Bani	P	85	85	85	85	85	85	77	85	85	77	B
5	Ardianus Rano Wala Ndara	L	90	85	85	85	85	85	80	85	85	80	B
6	Deponsius Umbu Hina	L	76	83	83	83	83	83	79	83	83	79	B
7	Desiyanti Kalita Mburu	P	90	90	90	90	90	90	75	90	90	75	B
8	Devino Domu Manukara	L	90	90	90	90	90	90	78	90	90	78	B

9	Falentinus Ion Bili	L	82	83	83	83	83	75	83	83	75	B
...		...	...	...	...	...	...	...	...	...	...	...
...		...	...	...	...	...	...	...	...	...	...	...
...		...	...	...	...	...	...	...	...	...	...	...
70	ZEPRI LAKA PEKUWALI	L	86	85	80	84	80	76	81	82	85	B

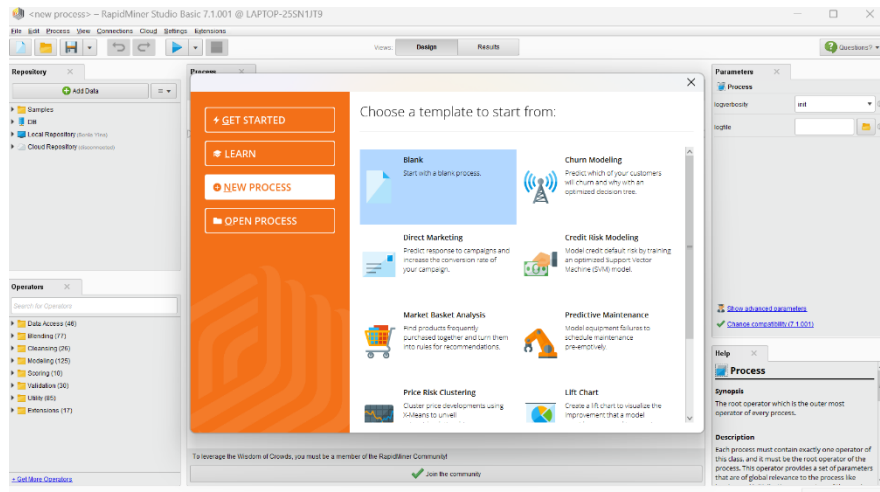
2. Pra-Pemrosesan Data

Selepas dilaksanakan pengumpulan data, maka akan dilaksanakan pengecekan kembali data-data yang sudah dikumpulkan, dan dilakukan inisialisasi data. Data diatas merupakan dataset mentah disiapkan untuk diolah, dengan data yang sudah didapatkan dengan demikian, data yang dipergunakan ialah data dengan atribut, nama siswa, jenis kelamin, nilai pengetahuan dari 10 mata pelajaran dan nilai proyek penguatan profil pelajar Pancasila (P5). Dengan jumlah data sebanyak 70 siswa kelas VII maka sudah dilakukan normalisasi data, yaitu pada tabel dibawah ini.

Tabel 2. Data Normalisasi

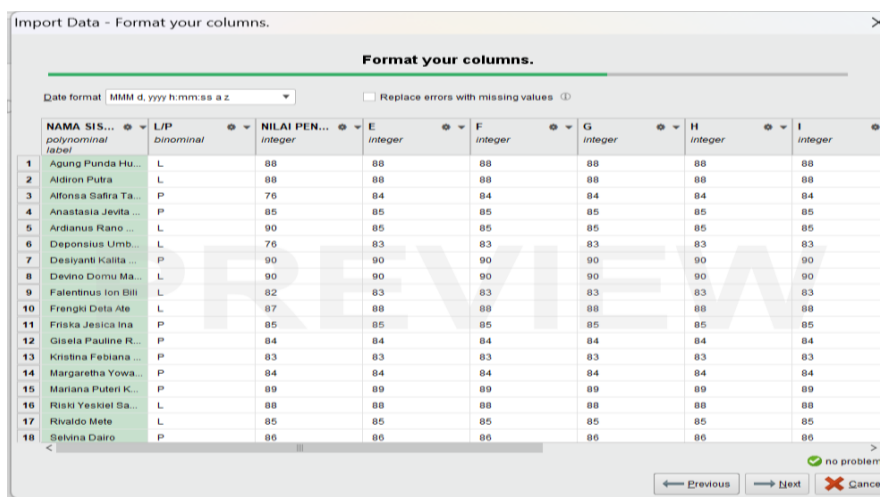
N O	NAMA SISWA / I	L/P	AG AM A	PAN CASI LA	BIN DO	MATE MATI KA	IP A	IPS	BIN G	PJ OK	TIK OM	SB D	P 5
1	Agung Punda Hudang	0	88	88	88	88	88	88	75	88	88	75	1
2	Aldiron Putra	0	88	88	88	88	88	88	78	88	88	78	1
3	Alfonsa Safira Taka	1	76	84	84	84	84	84	78	84	84	78	1
4	Anastasi a Jevita Bani	1	85	85	85	85	85	85	77	85	85	77	1
5	Ardianus Rano Wala Ndara	0	90	85	85	85	85	85	80	85	85	80	1
...			...									...	...
70	Zepri Laka Pekuwali	0	86	85	80	84	80	76	81	82	85	85	1

3. Penerapan K-Means Clustering dengan Aplikasi RapidMiner



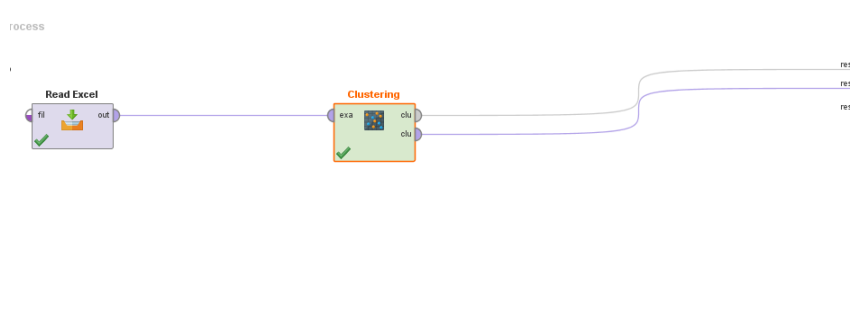
Gambar 2. Tampilan Menu Utama RapidMiner

Menu utama *RapidMiner* mempunyai tampilan awal yang meliputi beragam komponen fitur. Satu diantaranya yakni “new process”, yang ialah Langkah pertama dalam mempergunakan aplikasi *RapidMiner*.



Gambar 3. Tampilan Pemilihan Tipe Data

Gambar 3 merupakan tabel pemilihan tipe data serta menetapkan tipe data untuk setiap kolom, seperti nominal, numerik, dan binominal. Fungsi ini memastikan bahwa setiap kolom diinterpretasikan dengan benar.



Gambar 4. Tampilan Clustering RapidMiner

Gambar 4 menampilkan hasil *clustering* menggunakan metode *k-means* di *RapidMiner*. Data siswa telah dikelompokkan ke dalam beberapa klaster berdasarkan kesamaan nilai. Setiap baris data diberi label klaster (*cluster 1*, *cluster 2*, dan *cluster 3*) dan ditampilkan dalam bentuk tabel hasil output. Proses ini membantu mengidentifikasi pola kelompok siswa secara otomatis berdasarkan atribut yang dianalisis.

ExampleSet (70 examples, 2 special attributes, 12 regular attributes) Filter (70 / 70 examples): all

Row No.	id	cluster	NO	nama siswa	Agama	pancasila	bindo	matematika	ipa	ips	bing	piok
1	1	cluster_0	1	Agung Punda...	88	88	88	88	88	88	75	88
2	2	cluster_0	2	Aldron Putra	88	88	88	88	88	88	78	88
3	3	cluster_0	3	Alfonsa Safr...	76	84	84	84	84	84	78	84
4	4	cluster_0	4	Anastasia Je...	85	85	85	85	85	85	77	85
5	5	cluster_0	5	Arrianus Ran...	90	85	85	85	85	85	80	85
6	6	cluster_0	6	Deponsius U...	76	83	83	83	83	83	79	83
7	7	cluster_0	7	Desiyanti Kail...	90	90	90	90	90	90	75	90
8	8	cluster_0	8	Devino Domu...	90	90	90	90	90	90	78	90
9	9	cluster_0	9	Falentinus Io...	82	83	83	83	83	83	75	83
10	10	cluster_0	10	Frengki Deta ...	87	88	88	88	88	88	80	88
11	11	cluster_0	11	Friska Jessica...	85	85	85	85	85	85	75	85
12	12	cluster_0	12	Gisela Paulin...	84	84	84	84	84	84	75	84
13	13	cluster_0	13	Kristina Febl...	83	83	83	83	83	83	78	83
14	14	cluster_0	14	Margaretha Y...	84	84	84	84	84	84	80	84
15	15	cluster_0	15	Mariana Pute...	89	89	89	89	89	89	76	89
16	16	cluster_0	16	Riski Yeskiel ...	88	88	88	88	88	88	75	88
17	17	cluster_0	17	Rivaldo Mate	85	85	85	85	85	85	75	85
18	18	cluster_0	18	Selvina Dairo	86	86	86	86	86	86	75	86
19	19	cluster_0	19	Timotius Kal...	88	88	88	88	88	88	77	88
20	20	cluster_0	20	Valentina Jely...	88	88	88	88	88	88	76	88
21	21	cluster_0	21	Yohanis Bulu	87	87	87	87	87	87	75	87
22	22	cluster_0	22	Yosef Jekson...	83	83	83	83	83	83	75	83
23	23	cluster_0	23	Yuliana Ivank...	93	93	93	93	93	93	78	93

Gambar 5. Tampilan Hasil Clustering

Gambar 5 menunjukan hasil akhir dari pengelompokan data *cluster*. Sesuai hasil pengujian, bisa diambil suatu simpulan bahwasannya ada 23 siswa *cluster 1*, terdapat 22 siswa *cluster 2*, dan 25 siswa *cluster 3*.

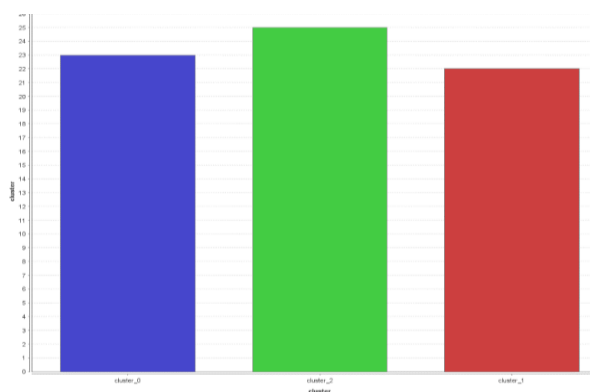
Cluster Model

```
Cluster 0: 23 items
Cluster 1: 22 items
Cluster 2: 25 items
Total number of items: 70
```

Gambar 6. Tampilan Cluster Model

Pada model *cluster*, ada tampilan yang menampilkan hasil *cluster*, tampilan teks yang memperlihatkan hasil pengelompokan sesuai *cluster* serta total anggota pada tiap *cluster* pada nilai $K=3$.

4. Visualisasi



Gambar 7. Hasil Visualisasi

Gambar 7 merupakan diagram batang yang menunjukkan jumlah siswa dalam masing-masing kluster hasil pengelompokan menggunakan metode *k-means clustering*. *Cluster 2* (warna hijau), mempresentasikan *cluster 1* (prestasi tinggi), dengan jumlah siswa terbanyak. *Cluster 0* (warna biru) menunjukkan *cluster 2* (prestasi sedang), dengan jumlah siswa sedikit lebih rendah. *Cluster 1* (warna merah) merupakan *cluster 3* (prestasi rendah), dengan jumlah siswa paling sedikit. Visualisasi ini menunjukkan distribusi siswa berdasarkan tingkat pencapaian akademik, dimana mayoritas siswa berada dalam kategori berprestasi tinggi.

5. Evaluasi

Tabel 3. Perhitungan DBI

<i>K-Means</i>		
K	<i>Avg. within centroid distance</i>	DBI
2	106.811	0.768
3	190.846	0.705
4	89.533	0.959

Berdasarkan tabel 3, dilakukan tiga kali percobaan dengan jumlah kluster berbeda ($K=2$, $K=3$, dan $K=4$) untuk menemukan hasil *clustering* yang paling optimal. Hasil menunjukkan bahwa nilai DBI terendah diperoleh saat jumlah kluster adalah 3, yaitu sebesar 0,705, sehingga dipilih sebagai konfigurasi kluster terbaik dalam penelitian ini.

KESIMPULAN [DAN SARAN (Optional)]

Studi ini berhasil mengaplikasikan metode *K-Means Clustering* untuk mengelompokkan siswa kelas VII di SMP PGRI Waingapu berdasarkan nilai pengetahuan dan proyek P5. Hasil evaluasi dengan DBI sebesar 0.705 menunjukkan kualitas *clustering* yang sangat baik. Tiga kluster terbentuk yaitu *cluster 1* (prestasi tinggi) dengan siswa sebanyak 23 items, *cluster 2* (prestasi sedang) dengan siswa sebanyak 22 items, dan *cluster 3* (prestasi rendah) dengan siswa sebanyak 25 items. Hasil ini membantu sekolah memahami capaian akademik siswa dan mendukung penerapan pembelajaran yang lebih terarah dan personal.

Hasil penelitian menyarankan agar pihak sekolah memanfaatkan hasil pengelompokan sebagai dasar dalam menyusun strategi pembelajaran yang lebih adaptif. Siswa berprestasi tinggi dapat diberikan program

pengayaan, sementara siswa dengan prestasi sedang dan rendah perlu mendapatkan bimbingan atau pendampingan khusus. Guru juga dapat menggunakan hasil clustering untuk mengelompokkan siswa di kelas dan menyesuaikan metode pembelajaran. Penelitian selanjutnya disarankan untuk menambahkan variabel non-akademik, seperti kehadiran atau keaktifan, guna mendapatkan hasil analisis yang semakin komprehensif.

DAFTAR PUSTAKA

Jurnal:

- [1] Syafina Haviyola, Susilawati, Mohamad Jajuli. 2023. Pengelompokan Prestasi Siswa Guna Kualifikasi Beasiswa Berdasarkan Data Nilai Menggunakan Algoritma K-Means. *JATI (Jurnal Mahasiswa Teknik Informatika)*. 7(4): 2786–2791.
- [2] Maspa Makkawaru. 2019. Pentingnya Pendidikan Bagi Kehidupan dan Pendidikan Karakter dalam Dunia Pendidikan. *Jurnal Konsepsi*. 8(3): 1–4.
- [3] Aditia Yudhistira, Rio Andika. 2023. Pengelompokan Data Nilai Siswa Menggunakan Metode K-Means Clustering. *JAITI (Journal of Artificial Intelligence and Technology Information)*. 1(1): 20–28.
- [4] Ryci Rahmatil Fiska. 2017. Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus SMKN 1 Sutera). *JSIR (Jaringan Sistem Informasi Robotik)*. 1(1): 42–51.
- [5] Mochammad Syukron Ramadani, Zaehol Fatah. 2021. Analisis Pengelompokan Data Nilai Siswa untuk Menentukan Siswa Berprestasi Menggunakan Metode Clustering K-Means. *JRSI (Jurnal Riset Sistem Informasi)*. 3(3): 424–439.
- [6] Ramadhana, Islamiyah, Amin Padmo Azam Masa. 2023. Penerapan Data Mining Menggunakan Metode K-Means Clustering Pada Data Ekspor Batubara. *ATASI (Adopsi Teknologi dan Sistem Informasi)*. 2(1): 35–42.
- [7] Muhamad, Agus Perdana Windarto, Suhada. 2019. Penerapan Algoritma C4 . 5 Pada Klasifikasi Potensi Siswa Drop Out. *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*. 3(1): 753–760.
- [8] Diwa Oktario Dacwanda, Yessica Nataliani. 2021. Implementasi k-Means Clustering untuk Analisis Nilai Akademik Siswa Berdasarkan Nilai Pengetahuan dan Keterampilan. *AITI (Jurnal Teknologi Informasi)*. 18(2): 125–138.
- [9] Falih Pramataning Dewi, Priskila Siwi Aryni, Yuyun Umaidah. 2022. Implementasi Algoritma K-Means Clustering Seleksi Siswa Berprestasi Berdasarkan Keaktifan dalam Proses Pembelajaran. *JISKA (Jurnal Informatika Sunan Kalijaga)*. 7(2): 111–121.
- [10] Muhammad Rafi Nahjan, Nono Heryana, Apriade Voutama. 2023. Implementasi Rapidminer Dengan Metode Clustering K-Means Untuk Analisa Penjualan Pada Toko Oj Cell. *JATI (Jurnal Mahasiswa Teknik Informatika)*. 7(1): 101–104.