

K-Means dan Data Mining Tools: Strategi Efektif untuk Menganalisis Siswa Putus Sekolah

Ade Christian¹, Hariyanto², Ahmad Yani³, Sumanto^{4*}

¹ Teknologi Informasi, Universitas Bina Sarana Informatika, Indonesia

² Sistem Informasi, Universitas Bina Sarana Informatika, Indonesia

³ Sistem Informasi Akuntansi, Universitas Bina Sarana Informatika, Indonesia

⁴ Informatika, Universitas Bina Sarana Informatika, Indonesia

¹ade.adc@bsi.ac.id, ²hariyanto.hro@bsi.ac.id, ³ahmad.amy@bsi.ac.id ,

^{4*}sumanto@bsi.ac.id

Abstrak: Pendidikan memiliki peran penting dalam membangun sumber daya manusia yang berkualitas, namun permasalahan putus sekolah masih menjadi tantangan serius, terutama di tingkat Sekolah Menengah Kejuruan (SMK). Untuk mengatasi masalah ini, penelitian ini membandingkan tiga aplikasi data mining, yaitu RapidMiner, Orange, dan Weka, dalam mengelompokkan siswa putus sekolah menggunakan algoritma K-Means. Data yang digunakan diperoleh dari berbagai sumber dan diproses melalui tahapan pengujian, penerapan algoritma K-Means, serta perbandingan hasil klasterisasi. Hasil penelitian menunjukkan bahwa RapidMiner memiliki akurasi tertinggi sebesar 86%, diikuti oleh Orange dengan 80%, dan Weka dengan 73%. Perbedaan akurasi ini menunjukkan bahwa setiap aplikasi memiliki keunggulan dan keterbatasan masing-masing dalam pemrosesan data dan pengelompokan siswa berdasarkan pola tertentu. Dari hasil perbandingan ini, RapidMiner terbukti lebih optimal dalam menghasilkan klaster yang lebih akurat dan stabil dibandingkan dengan dua aplikasi lainnya. Meskipun penelitian ini menunjukkan hasil yang signifikan, masih terdapat beberapa keterbatasan, seperti jumlah dataset yang terbatas dan penggunaan satu algoritma saja (K-Means). Oleh karena itu, penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan beragam, serta mengeksplorasi algoritma lain, seperti DBSCAN atau Hierarchical Clustering, untuk meningkatkan kualitas analisis. Selain itu, integrasi teknik machine learning yang lebih kompleks juga direkomendasikan guna meningkatkan akurasi prediksi. Hasil penelitian ini diharapkan dapat membantu lembaga pendidikan dalam mengidentifikasi pola siswa berisiko putus sekolah, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan dan strategi intervensi pendidikan yang lebih efektif.

Kata Kunci: Data Mining; K-Means, RapidMiner; clustering; Siswa Putus Sekolah.

Abstract: Education plays a crucial role in developing high-quality human resources, yet the issue of student dropout remains a significant challenge, particularly at the vocational high school (SMK) level. To address this problem, this study compares three data mining applications—RapidMiner, Orange, and Weka—in clustering dropout students using the K-Means algorithm. The dataset was collected from various sources and processed through testing, K-Means clustering application, and comparative analysis of clustering results. The findings indicate that RapidMiner achieved the highest accuracy at 86%, followed by Orange at 80% and Weka at 73%. This accuracy difference highlights that each application has its own strengths and limitations in data processing and student classification based on specific patterns. Among the three tools, RapidMiner proved to be the most optimal, producing more accurate and stable clusters than the other applications. Although this study provides significant insights, it has several limitations, such as a limited dataset size and the use of only one algorithm (K-Means). Future research should consider using larger and more diverse datasets while also exploring alternative clustering algorithms, such as DBSCAN or Hierarchical Clustering, to enhance analysis quality. Additionally, integrating more advanced machine learning techniques is recommended to improve predictive accuracy. This study is expected to help educational institutions identify patterns of students at risk of dropping out, providing a valuable foundation for decision-making and more effective educational intervention strategies.

Keywords: Data Mining; K-Means; RapidMiner; clustering; Dropout Students.

1. PENDAHULUAN

Pendidikan saat ini merupakan kebutuhan dasar yang memiliki peran krusial dalam kehidupan. Di tengah upaya banyak individu untuk meraih pendidikan setinggi mungkin, masih terdapat sebagian masyarakat yang tidak mendapatkan akses pendidikan yang layak, baik pada tingkat dasar maupun jenjang yang lebih tinggi. Bahkan, beberapa anak yang telah mengenyam pendidikan dasar tetap mengalami putus sekolah akibat berbagai faktor [1]. Penyebab utama putus sekolah meliputi keterbatasan dana pendidikan akibat kondisi ekonomi yang sulit, kurangnya fasilitas pendidikan, serta faktor lingkungan seperti pergaulan yang kurang mendukung [2]. Hak atas pendidikan dapat diperoleh secara formal melalui sekolah maupun secara informal dalam keluarga, tetapi dalam kenyataannya, tidak semua anak memperoleh kesempatan pendidikan formal akibat keterbatasan yang dialami oleh orang tua dalam memenuhi kebutuhan pendidikan mereka. Oleh karena itu, permasalahan putus sekolah menjadi isu yang perlu mendapatkan perhatian lebih dalam upaya meningkatkan kualitas pendidikan dan pemerataan akses bagi seluruh lapisan masyarakat [3].

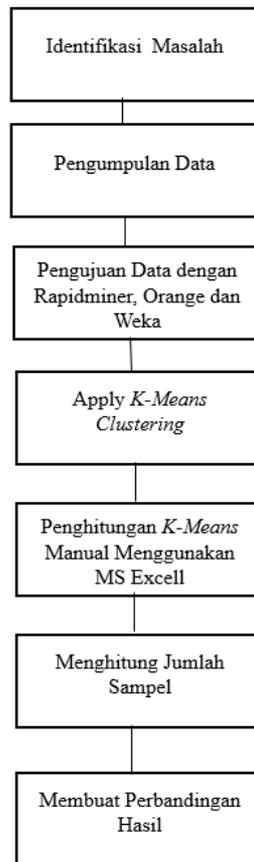
Machine learning telah digunakan secara efektif di berbagai bidang penelitian, termasuk dalam memprediksi kemungkinan siswa putus sekolah. Prediksi ini biasanya dimodelkan sebagai masalah klasifikasi biner, di mana kinerjanya dapat dipengaruhi oleh tingkat ketidakseimbangan data dropout. Masalah utama dalam prediksi ini adalah tingkat putus sekolah yang sangat bervariasi tergantung pada institusi pendidikan [4]. Sebagai contoh, pada tahun 2018, tingkat putus sekolah di universitas 4 tahun di Korea Selatan sekitar 5%, sedangkan pada universitas siber mencapai 18,9%. Di sisi lain, pada sistem

pembelajaran daring terbuka seperti MOOCs, tingkat putus sekolah bisa sangat tinggi, yaitu antara 80–95% [5]. Ketidakseimbangan data akibat perbedaan tingkat putus sekolah ini dapat berdampak signifikan pada akurasi model prediksi. Oleh karena itu, untuk mencapai akurasi yang tinggi, penting untuk menyesuaikan ketidakseimbangan data dengan baik dalam proses pembelajaran. Selain itu, diperlukan indikator evaluasi yang mempertimbangkan ketidakseimbangan data agar hasil prediksi lebih akurat dan dapat diandalkan [6]. K-Means adalah metode machine learning yang efektif untuk klusterisasi dan sangat cocok dalam mengidentifikasi penyebab siswa putus sekolah. Dengan membagi data ke dalam kelompok berdasarkan kemiripan pola, metode ini mampu mengelompokkan faktor-faktor seperti kondisi ekonomi, lingkungan sosial, dan prestasi akademik secara lebih terstruktur [7][8]. Keunggulannya terletak pada kesederhanaan, efisiensi, dan kemampuannya menemukan pola tersembunyi, sehingga membantu dalam analisis dan pengambilan keputusan yang lebih tepat untuk mencegah putus sekolah [9].

RapidMiner, Orange, dan Weka merupakan aplikasi data mining yang banyak digunakan untuk berbagai analisis data, termasuk klusterisasi dan prediksi [10]. Meskipun ketiganya memiliki fungsi dan fitur unggulan masing-masing, masih diperlukan evaluasi lebih lanjut untuk menentukan aplikasi mana yang paling efektif dalam memberikan hasil yang akurat, cepat, dan mudah digunakan [11][12]. Perbandingan ini penting untuk mengetahui keunggulan dan kelemahan masing-masing tools, terutama dalam aspek akurasi, efisiensi pemrosesan, serta kemudahan dalam pengoperasian. Dengan memahami performa dari setiap aplikasi, pengguna dapat memilih tools yang paling sesuai dengan kebutuhan analisis data, khususnya dalam proses pengelompokan dan prediksi [12], [13], [14]. Meskipun metode yang digunakan sudah baik, hasil yang diperoleh mungkin kurang optimal jika aplikasi yang digunakan tidak cocok dengan karakteristik data atau kebutuhan analisis. Sebaliknya, meskipun sebuah aplikasi memiliki fitur yang canggih, jika metode yang diterapkan tidak sesuai, maka hasil analisis tetap tidak akurat atau kurang efektif. Oleh karena itu, kesesuaian antara metode dan aplikasi sangat penting untuk mendapatkan hasil yang optimal dalam proses analisis data [15]. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa berbagai data mining tools, yaitu RapidMiner, Orange, dan Weka, dalam pengelompokan siswa putus sekolah menggunakan algoritma K-Means. Melalui penelitian ini, diharapkan dapat diketahui tools yang paling efektif berdasarkan akurasi, kecepatan pemrosesan, dan kemudahan penggunaan. Selain itu, penelitian ini juga bertujuan untuk mengidentifikasi pola faktor-faktor penyebab siswa putus sekolah guna memberikan wawasan yang lebih mendalam bagi lembaga pendidikan dalam merancang strategi intervensi yang lebih tepat.

2. METODE PENELITIAN

Penelitian ini dilakukan melalui serangkaian tahapan sistematis untuk menganalisis dan membandingkan performa berbagai tools data mining dalam pengelompokan siswa putus sekolah menggunakan algoritma K-Means. Gambar 1 menunjukkan alur penelitian yang dimulai dengan identifikasi masalah, diikuti oleh pengumpulan data yang digunakan dalam analisis. Selanjutnya, data diuji menggunakan RapidMiner, Orange, dan Weka untuk menerapkan K-Means Clustering. Untuk validasi, dilakukan perhitungan manual K-Means menggunakan Microsoft Excel, kemudian dihitung jumlah sampel untuk analisis yang lebih akurat. Tahap akhir adalah membandingkan hasil dari setiap metode dan tools yang digunakan untuk menentukan tools yang paling efektif dalam pengelompokan siswa putus sekolah.



Gambar 1. Research Framework

1. Identifikasi Masalah

Tahap awal penelitian dilakukan dengan mengidentifikasi permasalahan terkait pengelompokan siswa putus sekolah dan pentingnya penggunaan data mining tools untuk analisis lebih lanjut.

2. Pengumpulan Data

Data mengenai siswa putus sekolah dikumpulkan dari berbagai sumber terpercaya, termasuk dataset pendidikan, laporan akademik, atau data dari instansi terkait. Data ini mencakup variabel yang relevan, seperti kondisi ekonomi, lingkungan sosial, dan prestasi akademik siswa.

3. Pengujian Data dengan RapidMiner, Orange, dan Weka

Dataset yang telah dikumpulkan diuji menggunakan tiga tools data mining, yaitu RapidMiner, Orange, dan Weka. Setiap tools akan digunakan untuk menerapkan algoritma K-Means guna melihat efektivitasnya dalam pengelompokan siswa putus sekolah.

4. Apply K-Means Clustering

Pada tahap ini, algoritma K-Means diterapkan untuk mengelompokkan data siswa berdasarkan kemiripan pola tertentu. Proses ini dilakukan secara otomatis menggunakan tools data mining, dengan tujuan mengidentifikasi kelompok siswa yang berisiko putus sekolah.

5. Penghitungan K-Means Manual Menggunakan MS Excel

Untuk validasi hasil yang diperoleh dari tools data mining, dilakukan perhitungan manual K-Means menggunakan Microsoft Excel. Hal ini bertujuan untuk memastikan akurasi dan konsistensi hasil pengelompokan yang dilakukan oleh masing-masing aplikasi.

6. Menghitung Jumlah Sampel

Setelah hasil klasterisasi diperoleh, dilakukan penghitungan jumlah sampel pada setiap klaster. Tahap ini membantu dalam analisis distribusi data, sehingga dapat dilihat bagaimana pola siswa yang termasuk dalam kelompok tertentu.

7. Membuat Perbandingan Hasil

Tahap akhir penelitian adalah membandingkan hasil klasterisasi dari setiap tools data mining yang digunakan. Analisis dilakukan berdasarkan akurasi, kecepatan pemrosesan, dan kemudahan penggunaan dari masing-masing tools. Hasil perbandingan ini akan digunakan untuk menentukan tools yang paling efektif dalam mengelompokkan siswa putus sekolah dan memberikan rekomendasi terbaik bagi lembaga pendidikan/sekolah.

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini, data siswa putus sekolah yang diambil dari situs web <https://data.kemdikbud.go.id/> dianalisis menggunakan MS Excell dan tiga alat data mining: RapidMiner, Weka, dan Orange. Masing-masing alat digunakan untuk mengolah dan menganalisis data melalui beberapa langkah, antara lain memuat data, melakukan preprocessing data, menetapkan parameter K-Means, menerapkan clustering K-Means, mengevaluasi cluster dengan Silhouette Plot, dan memvisualisasikan cluster dengan Scatter Plot. Berikut ini adalah hasil penelitian dari masing-masing tools:

Menentukan titik pusat centroid secara acak (untuk x, y, dan z)

Tabel 1. Titik pusat centroid menggunakan MS Excel

Provinsi		x	y	z
Prov. D.I. Yogyakarta	C1	9	6	2
Prov. Sulawesi Tengah	C0	81	43	54
Prov. Nusa Tenggara Timur	C2	682	425	385

Pada Tabel 1, pemilihan titik pusat centroid dilakukan secara acak sebagai langkah awal dalam algoritma K-Means. Setelah itu, sistem akan menghitung jarak setiap data terhadap titik centroid yang telah dipilih dan menentukan kelompok (cluster) yang paling sesuai. Dalam proses ini, data akan dikelompokkan ke dalam tiga cluster, yaitu Cluster 1, Cluster 2, dan Cluster 3, berdasarkan kemiripan pola yang ditemukan. Proses ini dilakukan secara iteratif hingga setiap data berada dalam cluster yang paling sesuai dengan karakteristiknya.

Tabel 2. Pusat Cluster

Distance C1	Distance C0	Distance C2
91.29074433	48.03123983	806.8215416
680.4263369	633.8430405	169.7380335
259.5168588	501.4987537	388.8585861
9	96.21330469	880.4425024
1065.466564	962.7673655	185.4642823
167.1107417	76.072334	723.2572157
1336.047155	1260.78309	668.6972409
290.5098966	154.6641523	683.4171493
486.2314264	21.11871208	804.9186294
118.8738827	24.12467616	761.8556294
675.5989935	98.12746812	693.6634631
82.95782061	33.07567082	807.2558454
468.9413609	105.4134716	689.6912353
202.4969136	52.1919534	832.0829286
18.86796226	46.5295605	818.6055216
70.49822693	45.607017	824.9690903
156.8088008	65.90902821	756.5989691
67.50555533	0	785.2935757
390.6174599	391.1150726	406.995086
144.034718	67.41661516	721.3383395
32.57299495	54.27706698	837.7123611
25.07987241	75.69676347	860.7914962
298.0972996	268.6801072	557.7104984
826.7109531	785.2935757	0
327.687351	70.29224708	848.1043568
51.72040216	60.07495318	752.2692337
74.40430095	14.49137675	779.6011801
302.3855155	222.7868937	569.9701747

Pada Tabel 2. didapatkan hasil cluster 0 sampai 3 dengan menggunakan persamaan 1.

$$d(b_i, a_t) = \sqrt{\sum (b_{ij} - a_{tj})^2} \quad (1)$$

Keterangan:

d = jarak antara nilai data dan nilai pusat cluster

b_i = nilai data, $i = 1, 2, \dots, n$, n = jumlah data

a_t = nilai pusat cluster, $t = 1, 2, \dots, K$, K = jumlah cluster

l = jumlah atribut atau dimensi

Dalam proses **K-Means Clustering**, langkah berikutnya adalah **menghitung jarak minimum** antara setiap data dengan titik **centroid** yang telah ditentukan. Jarak ini digunakan untuk menentukan **keanggotaan klaster**, yaitu mengelompokkan setiap data ke dalam **cluster yang memiliki jarak terdekat** dengan centroid. Setelah data diklasifikasikan ke dalam cluster yang sesuai, dilakukan perhitungan **jarak minimum kuadrat (jarak minimum²)** untuk mengukur seberapa dekat data dengan titik centroid yang telah ditetapkan. Proses ini akan terus diulang hingga perubahan dalam pembagian cluster menjadi minimal atau mencapai kondisi konvergen.

Sumanto : *Penulis Korespondensi



Copyright © 2025, Ade Christian, Hariyanto, Ahmad Yani, Sumanto.

Tabel 3. Jarak minimum, keanggotaan klaster dan jarak minimum²

Minimum Distance	C	(Minimum Distance) ²
48.03123983	c0	2307
169.7380335	c2	28811
259.5168588	c1	67349
9	c1	81
185.4642823	c2	34397
76.072334	c0	5787
668.6972409	c2	447156
154.6641523	c0	23921
21.11871208	c0	446
24.12467616	c0	582
98.12746812	c0	9629
33.07567082	c0	1094
105.4134716	c0	11112
52.1919534	c0	2724
18.86796226	c1	356

Pada **Tabel 3**, nilai **jarak minimum** diperoleh dengan mencari **nilai terkecil** dari **Cluster 0 hingga Cluster 2** yang telah dihitung sebelumnya. Setelah menemukan **jarak minimum**, langkah berikutnya adalah menghitung **(Jarak Minimum)²**, yang dilakukan dengan **mengkuadratkan nilai jarak minimum** menggunakan rumus: **(Jarak Minimum)² = (Jarak Minimum) × (Jarak Minimum)**. Perhitungan ini berguna untuk mengevaluasi tingkat kedekatan setiap data terhadap **centroid** dalam proses klusterisasi.

Menghitung dan menentukan titik pusat baru (untuk x, y, dan z)

Dalam proses K-Means Clustering, setelah setiap data dikelompokkan ke dalam klaster berdasarkan jarak minimum ke centroid awal, langkah selanjutnya adalah menghitung dan menentukan titik pusat (centroid) baru. Titik pusat baru ini dihitung dengan mengambil rata-rata koordinat (x, y, dan z) dari seluruh data dalam setiap klaster. Hasil centroid baru terdapat pada Tabel 4.

Tabel 4. Centroid baru

Provinsi		x	y	z
Prov. D.I. Yogyakarta	C1	107.6666667	40.5	56.333
Prov. Sulawesi Tengah	C0	100.5454545	53.68181818	56.182
Prov. NTT	C2	669.5	472.75	563.25

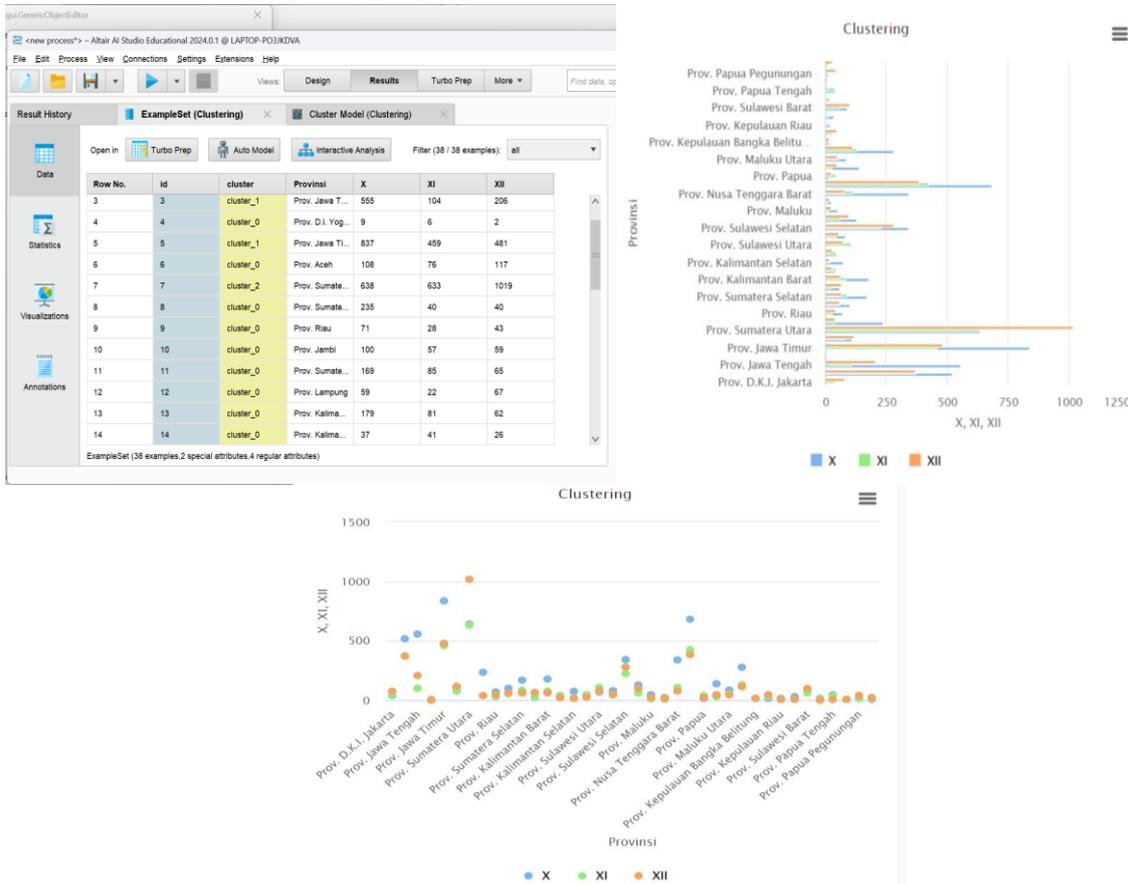
Setelah menghitung centroid baru, langkah berikutnya adalah menghitung kembali jarak setiap data ke centroid baru dan memperbarui keanggotaan klaster berdasarkan jarak terdekat. Jika terjadi perubahan dalam klaster, proses ini diulang hingga mencapai konvergensi, yaitu ketika tidak ada lagi perpindahan data antar-klaster. Setelah klusterisasi stabil, hasilnya dievaluasi menggunakan metrik seperti SSE, Silhouette Score, atau Davies-Bouldin Index untuk menilai kualitas pengelompokan. Hasil akhir ini membantu mengidentifikasi pola dan faktor utama penyebab siswa putus sekolah.

Tabel 5. Hasil akhir tanpa perubahan

Distance C1	Distance C0	Distance C2	Minimum Distance	C	(Minimum Distance)^2
326.0540446	43.59133536	902.5490984	43.59133536	C0	1900
378.5092337	657.8671895	264.4359563	264.4359563	C2	69926
215.23299	525.5941771	526.0364769	215.23299	C1	46325
386.864886	71.54066387	984.4375932	71.54066387	C0	5118
681.7970666	987.3780663	187.1105956	187.1105956	C2	35010
247.7386526	97.77065547	819.6535091	97.77065547	C0	9559
1054.152095	1282.725908	484.1284695	484.1284695	C2	234380
174.9023728	174.5587473	806.1348367	174.5587473	C1	30471
310.9897748	13.57854676	909.2111278	13.57854676	C0	184
271.6745111	47.95361661	866.8600666	47.95361661	C0	2300
200.5712841	121.2412714	805.6695197	121.2412714	C0	14699
317.0126811	29.52435058	906.7248067	29.52435058	C0	872
193.6477214	128.3111696	803.3118168	128.3111696	C0	16464
343.8770711	28.38706057	935.4687996	28.38706057	C0	806

Hasil perhitungan K-Means dengan RapidMiner, Orange dan Weka

Berikut ini merupakan hasil perhitungan K-Means Clustering menggunakan RapidMiner, Orange dan Weka yang menampilkan distribusi data berdasarkan klasterisasi provinsi dan kategori tingkat pendidikan.



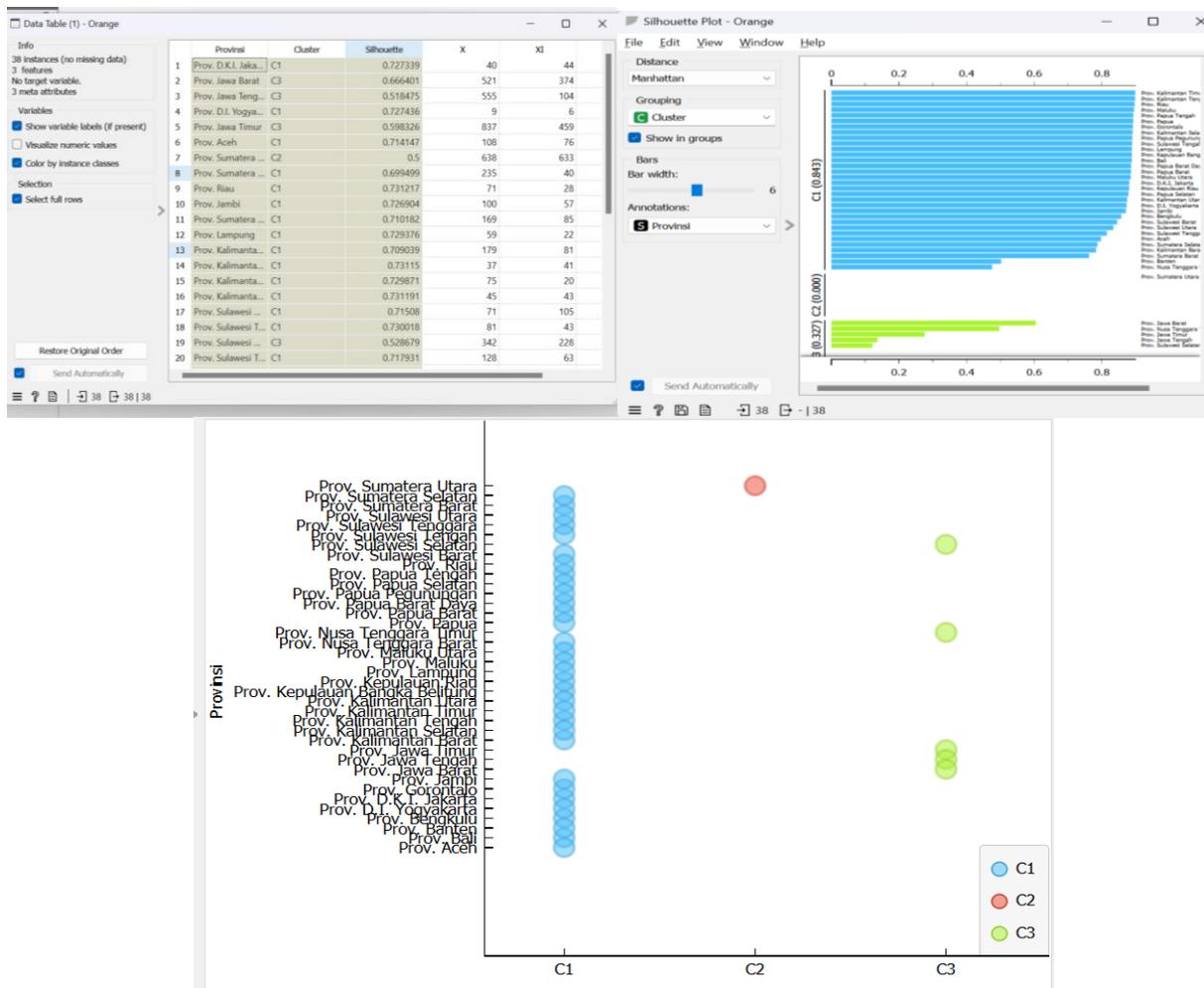
Gambar 2. Hasil perhitungan K-Means dengan Rapid Miner

Sumanto : *Penulis Korespondensi



Copyright © 2025, Ade Christian, Hariyanto, Ahmad Yani, Sumanto.

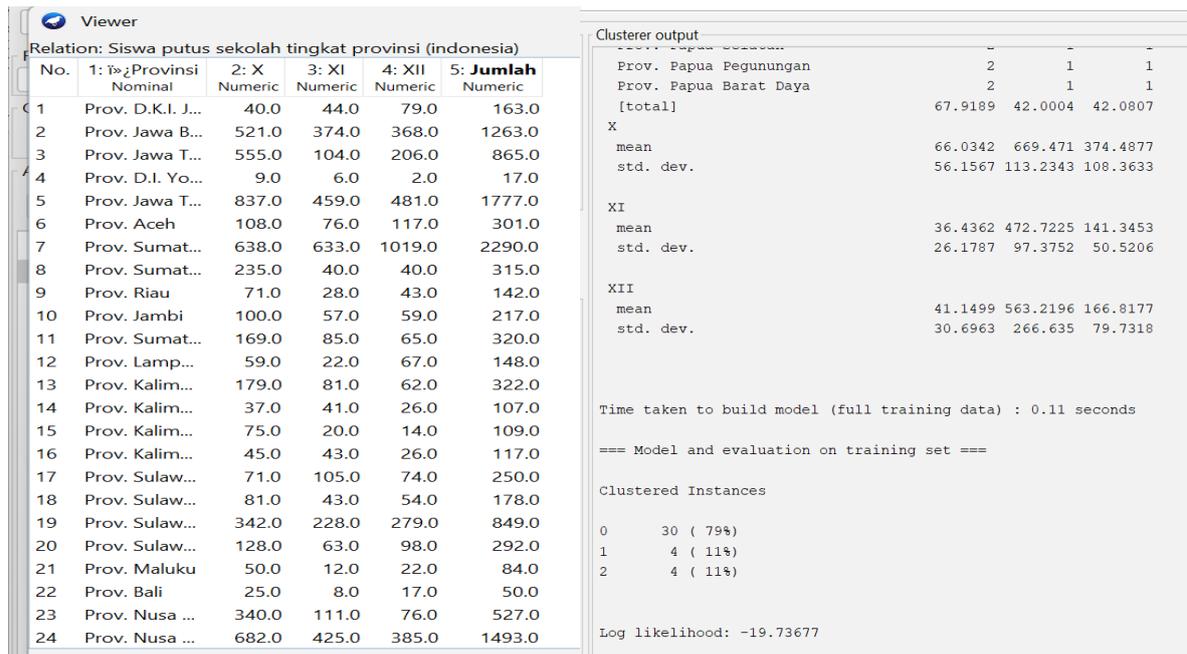
Gambar 2 menunjukkan hasil perhitungan K-Means Clustering menggunakan RapidMiner, yang mengelompokkan data berdasarkan pola tertentu. Tampilan pertama memperlihatkan hasil klasterisasi dalam bentuk tabel, di mana setiap provinsi dikelompokkan ke dalam klaster tertentu berdasarkan variabel yang digunakan. Grafik batang di bagian atas kanan menunjukkan distribusi jumlah data dalam masing-masing klaster untuk setiap provinsi, sedangkan scatter plot di bagian bawah menggambarkan penyebaran data antarprovinsi berdasarkan kategori X, XI, dan XII. Dari hasil ini, dapat diamati bahwa beberapa provinsi memiliki jumlah siswa yang lebih dominan dalam kelompok tertentu, yang mengindikasikan adanya pola atau faktor tertentu yang berkontribusi terhadap perbedaan antar klaster. Hasil klasterisasi ini dapat digunakan untuk mengidentifikasi hubungan antara faktor-faktor pendidikan dengan tingkat putus sekolah di berbagai wilayah, sehingga dapat menjadi dasar dalam pengambilan keputusan untuk strategi pendidikan yang lebih efektif.



Gambar 3. Hasil perhitungan K-Means dengan Orange data mining

Gambar 3 menunjukkan hasil perhitungan K-Means Clustering menggunakan Orange Data Mining, yang mengelompokkan data ke dalam tiga klaster berbeda (C1, C2, dan C3). Tampilan pertama menunjukkan tabel data hasil klasterisasi, di mana setiap provinsi dikelompokkan berdasarkan nilai silhouette score, yang mengukur sejauh mana data berada dalam klaster yang tepat. Diagram batang di bagian kanan menampilkan distribusi

silhouette score untuk setiap provinsi, menunjukkan kualitas pengelompokan data. Visualisasi scatter plot di bagian bawah memperlihatkan penyebaran data berdasarkan klaster dengan warna berbeda untuk masing-masing klaster (biru untuk C1, merah untuk C2, dan hijau untuk C3). Dari hasil ini, dapat diamati bahwa sebagian besar provinsi masuk dalam klaster C1, sedangkan hanya beberapa provinsi yang masuk dalam klaster C2 dan C3. Hasil klasterisasi ini dapat digunakan untuk menganalisis pola penyebab siswa putus sekolah di berbagai provinsi, serta membantu dalam pengambilan keputusan terkait strategi intervensi pendidikan.



Gambar 4. Hasil perhitungan K-Means dengan Weka

Gambar 4 menunjukkan hasil perhitungan K-Means Clustering menggunakan Weka, yang mengelompokkan data siswa putus sekolah berdasarkan provinsi di Indonesia. Tampilan pertama memperlihatkan tabel data input, yang mencakup jumlah siswa putus sekolah untuk tingkat X, XI, dan XII di setiap provinsi. Pada bagian kanan, terdapat hasil output klaster, yang menunjukkan distribusi provinsi ke dalam kelompok tertentu berdasarkan nilai rata-rata dan standar deviasi dari masing-masing variabel. Model ini berhasil membentuk tiga klaster, dengan persentase distribusi sebesar 74% untuk klaster 0, 14% untuk klaster 1, dan 11% untuk klaster 2. Selain itu, perhitungan mean dan standar deviasi untuk setiap tingkat pendidikan digunakan untuk mengidentifikasi pola dalam dataset. Hasil ini memberikan gambaran tentang pola putus sekolah di berbagai provinsi, yang dapat digunakan untuk analisis lebih lanjut dalam menentukan faktor dominan penyebab siswa tidak melanjutkan pendidikan.

Perbandingan Hasil Akhir

Untuk mengevaluasi kinerja masing-masing alat dalam pengelompokan siswa putus sekolah, dilakukan perbandingan akurasi hasil klasterisasi. Hasil akurasi dari setiap alat dihitung menggunakan Rumus 2 dan ditampilkan pada Tabel 6, yang menunjukkan tingkat keakuratan masing-masing metode dalam mengolah data.

$$\text{Accuracy} = \frac{\text{Total Correct Predictions}}{\text{TotData Samples}} \times 100 \quad (2)$$

Tabel 6. Hasil perbandingan aplikasi

Provinsi	MS.Excel	Rapidminer	Orange	Weka
Prov. D.K.I. Jakarta	C0	C0	C0	C0
Prov. Jawa Barat	C2	C1	C2	C1
Prov. Jawa Tengah	C1	C1	C2	C2
Prov. D.I. Yogyakarta	C0	C0	C0	C0
Prov. Jawa Timur	C2	C1	C2	C1
Prov. Aceh	C0	C0	C0	C0
Prov. Sumatera Utara	C2	C2	C1	C1
Total Benar	30	26	24	22
Akurasi		86%	80%	73%

Tabel 6 menunjukkan hasil perbandingan aplikasi data mining, yaitu RapidMiner, Orange, dan Weka, dalam mengelompokkan siswa putus sekolah berdasarkan centroid yang diperoleh dari MS Excel sebagai acuan perhitungan manual. Setiap provinsi dikelompokkan ke dalam salah satu klaster (C0, C1, atau C2), dan hasil klasterisasi dari masing-masing aplikasi dibandingkan dengan hasil dari MS Excel untuk menilai akurasi. Berdasarkan jumlah data yang dikelompokkan dengan benar, RapidMiner menunjukkan performa terbaik dengan total 26 data yang sesuai, menghasilkan akurasi sebesar 86%. Sementara itu, Orange memiliki 24 data yang benar, dengan akurasi 80%, dan Weka memiliki akurasi terendah, yaitu 73%, dengan hanya 22 data yang benar. Perbedaan akurasi ini menunjukkan bahwa RapidMiner lebih unggul dalam mendekati hasil perhitungan manual, diikuti oleh Orange, sedangkan Weka memiliki tingkat ketepatan yang lebih rendah dalam proses klasterisasi. Hasil ini dapat disebabkan oleh perbedaan dalam cara setiap aplikasi menangani algoritma K-Means, termasuk optimasi centroid, pemrosesan data, dan teknik pembagian klaster yang diterapkan oleh masing-masing software. Dengan demikian, RapidMiner dapat dianggap sebagai tools yang paling efektif untuk analisis klasterisasi siswa putus sekolah, meskipun Orange juga menunjukkan hasil yang cukup baik, sementara Weka cenderung memiliki tingkat akurasi yang lebih rendah dibandingkan kedua aplikasi lainnya.

4. KESIMPULAN

Penelitian ini membahas perbandingan tiga aplikasi data mining, yaitu RapidMiner, Orange, dan Weka, dalam pengelompokan siswa putus sekolah menggunakan algoritma K-Means. Hasil penelitian menunjukkan bahwa RapidMiner memiliki akurasi tertinggi sebesar 86%, diikuti oleh Orange dengan 80%, dan Weka dengan 73%. Perbedaan akurasi ini mengindikasikan bahwa setiap aplikasi memiliki kemampuan yang berbeda dalam mengimplementasikan algoritma K-Means, terutama dalam aspek optimasi centroid, pemrosesan data, dan metode evaluasi klaster. Dari hasil perbandingan ini, dapat disimpulkan bahwa RapidMiner adalah tools yang paling optimal untuk pengelompokan siswa putus sekolah, sedangkan Orange juga menunjukkan hasil yang cukup baik, sementara Weka memiliki tingkat akurasi yang lebih rendah. Penelitian ini memberikan manfaat bagi lembaga pendidikan dalam memilih tools yang tepat untuk mengidentifikasi pola siswa yang berisiko putus sekolah, sehingga dapat digunakan untuk mengambil keputusan yang lebih efektif dalam strategi intervensi pendidikan. Meskipun penelitian ini telah menghasilkan wawasan yang berguna, terdapat beberapa batasan yang perlu diperhatikan untuk penelitian selanjutnya. Salah satu batasan utama adalah jumlah

dataset yang digunakan masih terbatas, sehingga penelitian lanjutan dapat menggunakan dataset yang lebih besar dan lebih beragam untuk meningkatkan keakuratan dan generalisasi model. Selain itu, penelitian ini hanya menggunakan algoritma K-Means, sehingga dalam penelitian mendatang, dapat dilakukan eksperimen dengan algoritma lain, seperti DBSCAN, Hierarchical Clustering, atau kombinasi dengan metode supervised learning untuk melihat apakah ada pendekatan yang lebih efektif. Rekomendasi lainnya adalah mengintegrasikan teknik machine learning yang lebih kompleks, seperti Deep Learning atau ensemble methods, guna meningkatkan akurasi prediksi dan mendapatkan hasil klasterisasi yang lebih akurat. Dengan pengembangan lebih lanjut, penelitian ini dapat memberikan kontribusi yang lebih besar dalam pemanfaatan data mining untuk analisis pendidikan, terutama dalam mengurangi angka putus sekolah melalui sistem deteksi dini berbasis data.

5. REFERENCES

- [1] M. Madani and R. Risfaisal, "Perilaku Sosial Anak Putus Sekolah," *Equilib. J. Pendidik.*, vol. 4, no. 2, pp. 184–193, 2017, doi: 10.26618/equilibrium.v4i2.500.
- [2] M. L. Y. Yoridi and M. A. I. Pakereng, "Klasifikasi Anak Berpotensi Putus Sekolah dengan Metode Naïve Bayes Di Kabupaten Manokwari," *J. Sains Komput. Inform.*, vol. 7, no. 2, pp. 968–976, 2023, [Online]. Available: <https://tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/700/675>
- [3] E. Alyahyan and D. Düstegör, "Predicting academic success in higher education: literature review and best practices," *Int. J. Educ. Technol. High. Educ.*, vol. 17, no. 1, 2020, doi: 10.1186/s41239-020-0177-7.
- [4] C. H. Cho, Y. W. Yu, and H. G. Kim, "A Study on Dropout Prediction for University Students Using Machine Learning," *Appl. Sci.*, vol. 13, no. 21, 2023, doi: 10.3390/app132112004.
- [5] N. Mduma, K. Kalegele, and D. Machuve, "A survey of machine learning approaches and techniques for student dropout prediction," *Data Sci. J.*, vol. 18, no. 1, pp. 1–10, 2019, doi: 10.5334/dsj-2019-014.
- [6] D. Kim and S. Kim, "Sustainable education: Analyzing the determinants of university student dropout by nonlinear panel data models," *Sustain.*, vol. 10, no. 4, pp. 1–18, 2018, doi: 10.3390/su10040954.
- [7] Laila Khoirun Nisa, Tari Fitri Ningsih, Burhanuddin Izzul Salam, F. Fauzi, and Eny Winaryati, "Clustering Model K-Means Pada Kasus Angka Putus Sekolah Tingkatan Sekolah Dasar Di Provinsi Jawa Tengah," *LogicLink*, vol. 1, no. 1, pp. 13–20, 2024, doi: 10.28918/logiclink.v1i1.7793.
- [8] F. Simanjourang, R. Winanjaya, and F. Rizki, "Penerapan Algoritma K-Means Dalam Pengelompokan Rasio Angka Partisipasi Kasar di Tingkat Pendidikan Perguruan Tinggi Menurut Provinsi," *J. Terap. Inform. Nusant.*, vol. 2, no. 7, pp. 454–459, 2021.
- [9] M. Norshahlan, H. Jaya, and R. Kustini, "Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru," *J. Sist. Inf. Triguna Dharma (JURSI TGD)*, vol. 2, no. 6, p. 1042, 2023, doi: 10.53513/jursi.v2i6.9148.
- [10] E. Y. T. P. Dewi and I. Kamila, "Pengelompokan Wilayah Berdasarkan Faktor Pendukung Pendidikan Dengan Jumlah Sekolah Dan Jumlah Guru Menggunakan Algoritma K-Means," *Interval J. Ilm. Mat.*, vol. 2, no. 1, pp. 1–12, 2022, doi: 10.33751/interval.v2i1.5161.
- [11] S. Oktarian, S. Defit, and Sumijan, "Clustering Students' Interest Determination in School Selection Using the K-Means Clustering Algorithm Method," *J. Inf. dan Teknol.*, vol. 2, pp. 68–75, 2020, doi: 10.37034/jidt.v2i3.65.

- [12] Ainurrohma, "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," *Prism. Pros. Semin. Nas. Mat.*, vol. 4, pp. 493–499, 2021, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/>
- [13] S. Informasi *et al.*, "Perbandingan Penggunaan Aplikasi Rapidminer Dengan Weka Untuk Penerapan Metode K-Means Clustering Pada Pengelompokan Penderita Demensia," *J. Ilm. Komputasi*, vol. 18, no. 2, 2019, doi: 10.32409/jikstik.18.2.2584.
- [14] A. Surip, M. A. Pratama, I. Ali, A. R. Dikananda, and A. I. Purnamasari, "Penerapan Machine Learning menggunakan algoritma C4.5 berbasis PSO dalam Menganalisa Data Siswa Putus Sekolah," *INFORMATICS Educ. Prof. J. Informatics*, vol. 5, no. 2, p. 147, 2021, doi: 10.51211/itbi.v5i2.1530.
- [15] R. Ridwan, I. Irawaty, and A. H. Momo, "Faktor Penyebab Anak Putus Sekolah (Studi di Desa Mapila Kecamatan Kabaena Utara Kabupaten Bombana)," *Selami Ips*, vol. 12, no. 1, p. 62, 2020, doi: 10.36709/selami.v12i1.10838.