



ANALISIS PENGELOMPOKAN DATA NILAI SISWA UNTUK MENENTUKAN SISWA BERPRESTASI MENGGUNAKAN METODE CLUSTERING K- MEANS

Mochammad Syukron Ramadani^{a*}, Zaehol Fatah^b

^a Sains & teknologi / Sistem Informasi, syukronrdani1@gmail.com Universitas Ibrahimy Situbondo Jawa Timur

^b Sains & teknologi / Sistem Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy Situbondo Jawa Timur

*Korespondensi

ABSTRACT

Identifying high-achieving students is a critical step in evaluating learning outcomes to enhance the quality of education. This study aims to analyze the clustering of student grade data using the K-Means Clustering method to identify groups of high-achieving students. The K-Means method is utilized due to its effectiveness in grouping data based on value similarity. The data used in this study consist of students' academic scores across various subjects. The research stages include data collection, preprocessing, applying the K-Means algorithm, and validating the clustering results. The results show that the K-Means method successfully grouped students into several categories, such as high-achieving, moderate-achieving, and low-achieving students. The clustering analysis indicates that high-achieving students exhibit consistent performance across all subjects, whereas low-achieving students tend to show significant variations in their scores. This method also provides data visualization that helps schools make informed decisions to improve student performance. Thus, the implementation of the K-Means method in clustering student grade data can serve as an effective and efficient approach to support evaluation processes and data-driven decision-making.

Keywords: *K-Means Clustering, Data Grouping, High-Achieving Students, Data Analysis, Performance Evaluation.*

Abstrak

Penentuan siswa berprestasi merupakan langkah penting dalam evaluasi hasil belajar guna mendorong peningkatan kualitas pendidikan. Penelitian ini bertujuan untuk menganalisis pengelompokan data nilai siswa menggunakan metode Clustering K-Means untuk mengidentifikasi kelompok siswa berprestasi. Metode K-Means digunakan karena kemampuannya yang efektif dalam mengelompokkan data berdasarkan kedekatan nilai (similarity). Data yang digunakan terdiri dari nilai akademik siswa dari berbagai mata pelajaran. Tahapan penelitian meliputi pengumpulan data, preprocessing, penerapan algoritma K-Means, dan validasi hasil clustering. Hasil penelitian menunjukkan bahwa metode K-Means mampu mengelompokkan siswa ke dalam beberapa kategori, seperti siswa berprestasi tinggi, berprestasi sedang, dan berprestasi rendah. Analisis hasil clustering mengindikasikan bahwa kelompok siswa berprestasi memiliki pola nilai yang konsisten di semua mata pelajaran, sedangkan siswa dengan prestasi rendah cenderung memiliki variasi nilai yang signifikan. Metode ini juga memberikan visualisasi data yang memudahkan pihak sekolah dalam mengambil kebijakan untuk peningkatan prestasi siswa. Dengan demikian, penerapan metode K-Means dalam pengelompokan data nilai siswa dapat menjadi pendekatan yang efektif dan efisien dalam mendukung proses evaluasi serta pengambilan keputusan berbasis data.

Kata Kunci: Clustering K-Means, Pengelompokan Data, Siswa Berprestasi, Analisis Data, Evaluasi Prestasi.

1. PENDAHULUAN

Pendidikan memiliki peran strategis dalam membentuk kualitas sumber daya manusia yang berdaya saing tinggi. Salah satu indikator keberhasilan pendidikan adalah prestasi belajar siswa yang diukur melalui nilai akademik dari berbagai mata pelajaran. Penilaian prestasi siswa tidak hanya berfungsi sebagai evaluasi, tetapi juga sebagai dasar dalam merancang program pembinaan yang lebih tepat sasaran. Namun, dalam praktiknya, identifikasi siswa berprestasi sering kali menjadi tantangan tersendiri bagi pihak sekolah, terutama ketika jumlah siswa yang besar dan data nilai yang kompleks harus dianalisis secara manual. [1]

Proses identifikasi siswa berprestasi secara konvensional cenderung kurang efektif, membutuhkan waktu lama, serta berisiko menimbulkan bias atau kesalahan manusia. Selain itu, adanya variasi nilai di antara siswa membuat proses pengelompokan prestasi menjadi sulit dilakukan tanpa dukungan metode analisis data yang sistematis. Hal ini mendorong perlunya penerapan metode berbasis teknologi untuk menganalisis data nilai siswa secara objektif dan efisien. [2]

Metode Clustering K-Means menjadi salah satu pendekatan yang relevan untuk mengatasi masalah ini. Clustering K-Means merupakan algoritma unsupervised learning yang bertujuan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan kemiripan karakteristik atau pola nilai. Metode ini efektif untuk menangani data numerik dalam jumlah besar dan menghasilkan pengelompokan yang dapat dianalisis lebih lanjut. Dengan menerapkan metode K-Means, siswa dapat dikelompokkan ke dalam kategori tertentu, seperti siswa berprestasi tinggi, siswa berprestasi sedang, dan siswa berprestasi rendah. [3]

Penggunaan metode ini diharapkan mampu memberikan solusi dalam identifikasi siswa berprestasi secara lebih akurat dan efisien. Hasil pengelompokan dapat membantu pihak sekolah dalam mengambil kebijakan yang lebih terarah, seperti memberikan penghargaan kepada siswa berprestasi atau menyusun program pembinaan khusus bagi siswa dengan prestasi rendah. Dengan demikian, pendekatan ini tidak hanya mendukung evaluasi pendidikan, tetapi juga berkontribusi pada peningkatan kualitas belajar siswa secara keseluruhan. [4]

Berdasarkan latar belakang tersebut, penelitian ini dilakukan untuk menganalisis pengelompokan data nilai siswa menggunakan metode Clustering K-Means sebagai upaya menentukan siswa berprestasi secara sistematis, objektif, dan berbasis data. [5]

2. LANDASAN TEORI

2.1. Data Mining

Data mining adalah proses yang digunakan untuk menemukan pola dan informasi berguna dari kumpulan data besar. Menurut Nandang Iriadi (2021), data mining adalah proses iteratif yang ditujukan untuk analisis basis data yang bertujuan menyaring informasi dan pengetahuan yang dapat membuktikan keakuratan data, serta berpotensi bagi para ahli ilmiah dalam pengambilan keputusan dan pemecahan masalah. Data mining menggabungkan teknik dari statistik, kecerdasan buatan, dan machine learning untuk mengekstraksi informasi yang relevan dari berbagai database besar. [6]

2.2. Clustering

Clustering adalah metode yang digunakan untuk mengelompokkan data ke dalam beberapa grup berdasarkan kesamaan yang ada di antara data tersebut. Menurut Widodo (2013), clustering dapat diartikan sebagai proses pengelompokan objek-objek data yang memiliki kesamaan tinggi satu sama lain dalam satu cluster dan berbeda dengan objek-objek di cluster lainnya. Prinsip dasar dari clustering adalah memaksimalkan kesamaan dalam satu cluster dan meminimalkan kesamaan antar cluster, sehingga objek-objek data dapat dipresentasikan sebagai titik dalam ruang multidimensi [7]

2.2.1. K-Means

K-Means adalah metode clustering yang digunakan untuk mengelompokkan data ke dalam beberapa cluster berdasarkan karakteristik yang sama. Menurut G. Abdillah et al. (2016), K-Means adalah algoritma yang berusaha mempartisi data ke dalam k kelompok, di mana setiap kelompok memiliki karakteristik yang serupa dan berbeda dengan kelompok lainnya. Proses ini melibatkan langkah-langkah seperti menentukan jumlah cluster, inisialisasi centroid, penempatan data ke cluster terdekat, dan pembaruan centroid hingga konvergensi tercapai. [8] adapun langkah-langkah k-means dapat dilihat pada flowchart berikut:



Gambar 1 Flowchart K- Means

Dari gambar 1 dapat dilihat diagram alir-Means yang dimulai dengan menentukan banyaknya cluster. Setelah nilai ditentukan, langkah selanjutnya adalah menentukan pusat cluster, dan dilanjutkan dengan menghitung jarak setiap objek setiap pusat cluster. Jarak antara tiap titik data yang ada dan tiap centroid akan dihitung menggunakan analisis Euclidean hingga ditentukan jarak terpendek antara setiap titik data dan centroid. Selanjutnya dilakukan pengelompokan objek berdasarkan jarak minimum terhadap pusat cluster. Pusat cluster tersebut kemudian untuk sementara dijadikan pusat cluster, atau centroid, mean. Jika masih terdapat objek yang harus dipindah ke cluster yang lain, maka proses diulang kembali, tetapi jika tidak, maka proses selesai [9]

Setiap cluster yang terbentuk akan meningkatkan kriteria partisi, seperti fungsi perbedaan berdasarkan jarak, sehingga objek-objek di dalam cluster menjadi mirip, dan objek-objek pada cluster yang berbeda ditemukan tidak mirip dalam hal atribut dataset. Jarak Euclidean digunakan sebagai ukuran jarak dalam pendekatan K-Means untuk menyoroti kesamaan antara setiap cluster dengan jarak terkecil dan kesamaan tertinggi. Jarak Euclidean antara titik $a = (a_1, a_2, \dots, a_K)$ dan titik $b = (b_1, b_2, \dots, b_n)$ dapat dihitung dengan Rumus dibawah ini:

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

dimana:

d = jarak antara nilai data dan nilai pusat cluster

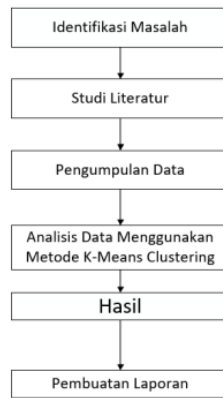
b_i = nilai data, $i = 1, 2, \dots, n$, n = jumlah data

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

l = jumlah atribut atau dimensi [10]

3. METODOLOGI PENELITIAN

Struktur langkah-langkah yang jelas diperlukan untuk memberikan pedoman dalam pengembangan penelitian ini. Kerangka ini menguraikan langkah-langkah yang diambil dalam penelitian ini. Kerangka kerja terdiri dari pedoman yang digambarkan pada Gambar



Gambar 2 Metode Penelitian

3.1 Identifikasi masalah

Tahap awal dalam penyelidikan ini adalah untuk identifikasi masalah. Identifikasi masalah bertujuan untuk menjabarkan dan menjelaskan problem mengenai masalah yang akan diteliti dan berkembang pada objek yang diteliti sehingga latar belakang penelitian ini dapat diidentifikasi dengan bagaimana mengatasi pengelompokan prestasi murid berdasarkan nilai mata pelajaran.

3.2 Studi Literatur

Untuk mengumpulkan data dan informasi untuk penelitian, dilakukan studi pustaka dengan memanfaatkan studi ilmiah maupun literatur yang berkaitan dengan penelitian ini.

3.3 Pengumpulan Data

Data diperoleh dari data online dari website Kaggle.

3.4 Metode K-Means digunakan untuk menganalisis data

Analisis data dilakukan pada tahap ini dengan menggunakan survei yang lalu dan temuan wawancara. Data tersebut akan dianalisis dengan pendekatan K-Means clustering untuk mengetahui apakah prestasi belajar murid tergolong rendah, cukup, atau tinggi

3.5 Hasil

Data yang telah dianalisis dengan metode K-Means clustering menghasilkan penentuan pengelompokan prestasi murid rendah, cukup, dan tinggi.

3.6 Pembuatan Laporan

Pembuatan laporan berdasarkan hasil penelitian yang kemudian ditransformasikan menjadi laporan penelitian yang memuat penjabaran komprehensif.

4. HASIL DAN PEMBAHASAN

Software RStudio digunakan dalam penelitian ini untuk membahas pengelompokan menggunakan algoritma K-Means dengan menggunakan 10 data nilai Siswa

Tabel 1 Data Nilai Siswa

Nama	Matematika	Bahasa Indonesia	Bahasa Inggris	IPA	IPS
Ali	85	90	88	92	87
Budi	78	85	80	75	82
Cici	92	95	93	90	91
Dedi	70	65	75	80	72
Evi	88	90	85	87	90
Fani	76	78	80	72	74
Gita	95	92	94	90	89
Hadi	82	88	86	84	83
Intan	89	91	88	88	85
Joko	60	70	65	62	68

Gambar 3 merupakan data nilai dan absensi murid yang sudah dimasukkan dalam bentuk Excel dan dipanggil ke Rapid Miner Studio.

	A	B	C	D	E	F
1	Nama	Matematika	Bahasa Indonesia	Bahasa Inggris	IPA	IPS
2	Ali	85.000	90.000	88.000	92.000	87.000
3	Budi	78.000	85.000	80.000	75.000	82.000
4	Cici	92.000	95.000	93.000	90.000	91.000
5	Dedi	70.000	65.000	75.000	80.000	72.000
6	Evi	88.000	90.000	85.000	87.000	90.000
7	Fani	76.000	78.000	80.000	72.000	74.000
8	Gita	95.000	92.000	94.000	90.000	89.000
9	Hadi	82.000	88.000	86.000	84.000	83.000
10	Intan	89.000	91.000	92.000	88.000	85.000
11	Joko	60.000	70.000	65.000	62.000	68.000

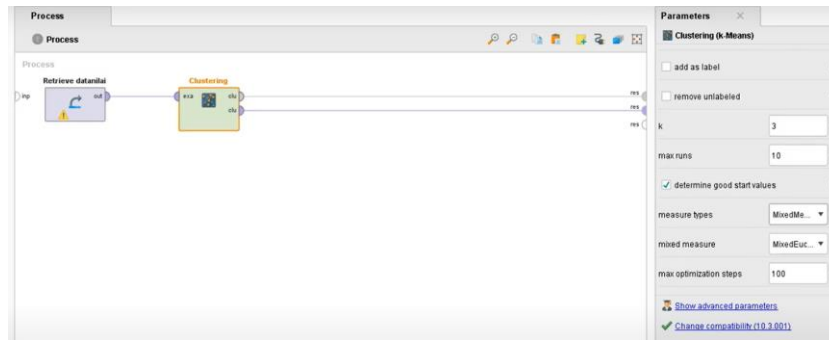
Gambar 3 Data Nilai Siswa

Gambar 4 merupakan Tabel yang sudah dirubah sesuai dengan tipe data nya

	Nama	Matematika	Bahasa Indonesia	Bahasa Inggris	IPA	IPS
	polynomial label	integer	integer	integer	integer	integer
1	Ali	85	90	88	92	87
2	Budi	78	85	80	75	82
3	Cici	92	95	93	90	91
4	Dedi	70	65	75	80	72
5	Evi	88	90	85	87	90
6	Fani	76	78	80	72	74
7	Gita	95	92	94	90	89
8	Hadi	82	88	86	84	83
9	Intan	89	91	92	88	85
10	Joko	60	70	65	62	68

Gambar 4 Format Kolom

Gambar 5 merupakan gambar proses clustering pada aplikasi Rapid Miner Studio



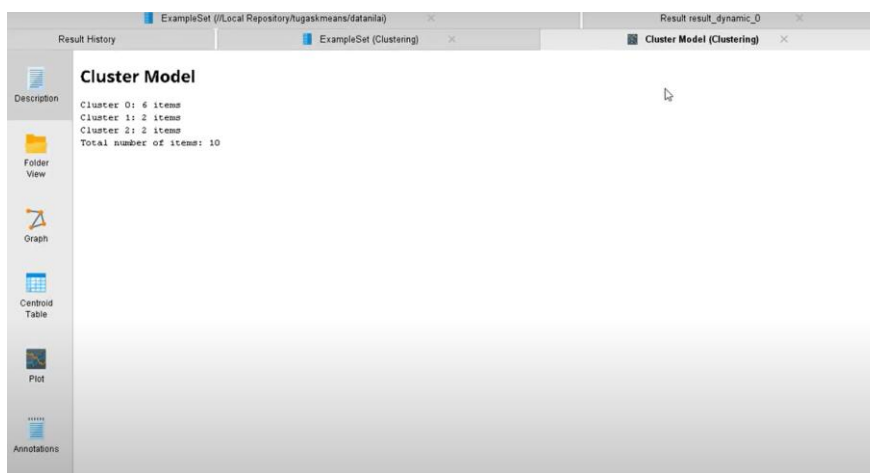
Gambar 5 Process

Gambar 6 Merupakan hasil clusterisasi dari data nilai siswa datanya sudah terkumpul data rata-rata nilai dari paling besar sampai paling kecil tertuang dalam tabel pada gambar berikut:

Flow No.	id	Nama	cluster	Matematika	Bahasa Indo...	Bahasa Ingg...	IPA	IPS
1	1	Ali	cluster_0	85	90	88	92	87
2	2	Budi	cluster_1	78	85	80	75	82
3	3	Cici	cluster_0	92	95	93	90	91
4	4	Dedi	cluster_2	70	65	75	80	72
5	5	Evi	cluster_0	88	90	85	87	90
6	6	Fani	cluster_1	76	78	80	72	74
7	7	Gita	cluster_0	95	92	94	90	89
8	8	Haadi	cluster_0	82	88	86	84	83
9	9	Intan	cluster_0	89	91	92	88	85
10	10	Joko	cluster_2	60	70	65	62	68

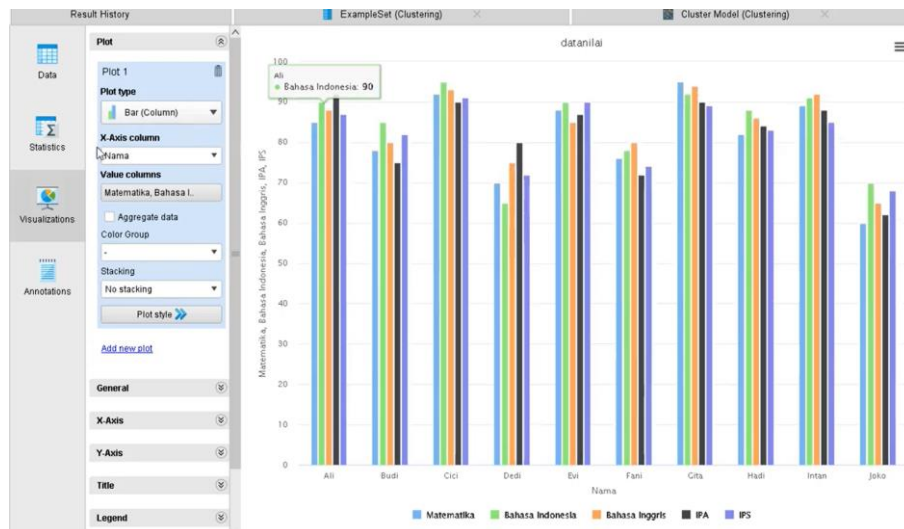
Gambar 6 Hasil clusterisasi tabel

Gambar 7 merupakan gambar hasil clusterisasi data nilai siswa yang berbentuk cluster model disana terdapat cluster 0 terdapat 6 items artinya 6 siswa ini adalah rata-rata nilai tertinggi, dan cluster 0 terdapat 1 items ini merupakan nilai tertinggi ke-2, dan selanjutnya cluster 2 terdapat 2 items pula, ini termasuk cluster nilai rata-rata siswa terendah jadi total keseluruhan adalah 10 data nilai rata-rata siswa 6 data nilai rata-rata siswa tertinggi pertama, data nilai 2 nilai siswa tertinggi ke-2 dan 2 data nilai rata-rata siswa tertinggi ke-3.



Gambar 7 Cluster Model

Gambar 8 Merupakan Visualisasi Data dari nilai rata-rata Siswa dari paling tinggi hingga paling rendah menggunakan Rapid Miner Studio.



Gambar 8 Visualisasi Data

5. KESIMPULAN DAN SARAN

Metode *K-Means* dapat mengelompokkan data nilai murid dalam kumpulan kecil maupun besar dan efektif dalam mengelompokkan dokumen dengan tepat, seperti yang terlihat di atas. Penentuan centroid (titik pusat) pada tahap awal algoritma *K-Means* memiliki pengaruh besar pada penemuan cluster, seperti yang ditunjukkan dengan pengujian memanfaatkan 10 *record* yang *centroid*-nya berbeda menghasilkan *cluster* yang bervariasi. Berdasarkan hasil penelitian ini didapat tiga cluster yaitu *cluster 0* yang merupakan kategori tinggi sebanyak 6 anggota, *cluster 1* yang merupakan kategori cukup sebanyak 2 anggota dan *cluster 2* yang merupakan kategori rendah sebanyak 2 anggota. Dari cluster tinggi, dapat dilanjutkan dengan mencari siswa unggulan (top rank). Dengan metode SAW, didapatkan Alfadhea yang menduduki top rank. Hasil ini dapat dimanfaatkan sekolah untuk menganalisis prestasi dari murid juga bisa membantu guru maupun wali kelas dalam membentuk kelas yang ideal dalam arti memotivasi murid untuk belajar lebih giat lagi.

Penggunaan metode lain untuk menghitung data yang dijelaskan di atas, atau memperluasnya dengan lebih banyak data, diharapkan dapat diaplikasikan dalam penelitian selanjutnya. Dapat diprediksi bahwa penelitian ini akan mengarah pada aplikasi maupun tools tambahan yang tetap digunakan dalam menyelesaikan permasalahan data murid di sekolah.

DAFTAR PUSTAKA

- [1] D .T. Larose and C. D. Larose, *Discovering Knowledge in Data : An Introduction to Data Mining:Second edition*, vol. 978047098. 2014.
- [2] A. & R.Andrian, "Penerapan Metode *K-Means* Untuk *Clustering* Mahasiswa Berdasarkan Nilai Akademik Dengan *Weka Interface* Studi Kasus Pada Jurusan Teknik Informatika UMM Magelang," *J. Ilm. Tek.*, vol. 18, no. 1, hal. 76–82, 2015.
- [3] A. Ramadhan dan Z. Efendi, "Perbandingan *K-Means* dan *Fuzzy C-Means* untuk Pengelompokan *Data User Knowledge Modeling*," *Semin. Nas. Teknol. Informasi, Komun. dan Ind.* 9, hal. 18–19, 2017.
- [4] P. A. Jusia dan F. M. Irfan, "*Clustering* Data Untuk Rekomendasi Penentuan Jurusan Perguruan Tinggi Menggunakan Metode *K-Means*," *J. IKRA-ITH Inform.*, vol. 3, no. 3, hal. 75–84, 2019.
- [5] Netriwati, "*K-Means – Penerapan, Permasalahan dan Metode Terkait*," *J. Pendidik. Mat.*, vol. 3, no. Pebruari, hal. 47–60, 2017
- [6] N. Iriadi, *Data Mining: Konsep dan Aplikasi*, Yogyakarta: Andi Offset, 2021.
- [7] A. B. Widodo, *Pengantar Data Mining*, Edisi 2, Yogyakarta: Andi Offset, 2013.
- [8] G. Abdillah et al., "Penerapan Data Mining Pemakaian Air Pelanggan Untuk Menentukan *Klasifikasi* Potensi Pemakaian Air Pelanggan Baru Di PDAM Tirta Raharja Menggunakan

Algoritma *K-Means*,” *Sentika*, vol. 2016, no. Sentika, pp. 18–19, 2016.

- [9] Y. Darmi dan A. Setiawan, “Penerapan Metode *Clustering K-Means* Dalam Pengelompokan Penjualan Produk,” *J. Media Infotama*, vol. 12, no. 2, hal. 148–157, 2016.
- [10] D.H.Margareth and H.Precise, *Data Mining introductory and advanced topicData Mining introductory and advanced Topics*, vol. 9780130888. 2020