**PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK IDENTIFIKASI SISWA BERPRESTASI DI KELAS UNGGULAN PADA SMP 28 SAROLANGUN**

**TUGAS AKHIR**

*Diajukan Sebagai Salah Satu Syarat Memperoleh Gelar Sarjana (S1) Pada Program Studi Informatika Departemen Teknik Elektronika Fakultas Teknik Universitas Negeri Padang*

****

**NUR CHINTIA NINGSIH**

**NIM. 21343011**

**PROGRAM STUDI INFORMATIKA**

**DEPARTEMEN TEKNIK ELEKTRONIKA**

**FAKULTAS TEKNIK UNIVERSITAS NEGERI PADANG**

**2024**

DAFTAR ISI

[BAB I](#_Toc180681485) [PENDAHULUAN 1](#_Toc180681486)

[A. Latar Belakang 1](#_Toc180681487)

[B. Identifikasi Masalah 5](#_Toc180681488)

[C. Batasan Masalah 6](#_Toc180681489)

[D. Rumusan Masalah 6](#_Toc180681490)

[E. Tujuan Penelitian 6](#_Toc180681491)

[F. Manfaat Penelitian 7](#_Toc180681492)

[BAB II](#_Toc180681493) [TINJAUAN PUSTAKA 8](#_Toc180681494)

[A. Data *Mining* 8](#_Toc180681495)

[B. *Clustering* 12](#_Toc180681496)

[C. K-Means 14](#_Toc180681497)

[D. Siswa Berprestasi 16](#_Toc180681498)

[E. Kelas Unggulan 20](#_Toc180681499)

[F. Kerangka Konseptual 20](#_Toc180681500)

[BAB III](#_Toc180681501) [METODE PENELITIAN 23](#_Toc180681502)

[A. Jenis Penelitian 23](#_Toc180681503)

[B. Desain Penelitian 23](#_Toc180681504)

[C. Lokasi Dan Waktu Penelitian 23](#_Toc180681505)

[D. Populasi dan Sampel 24](#_Toc180681506)

[E. Variabel Penelitian 25](#_Toc180681508)

[F. Instrumen Penelitian 25](#_Toc180681509)

[G. Prosedur Pengumpulan Data 25](#_Toc180681510)

[H. Prosedur Analisis Data 28](#_Toc180681511)

[I. Validitas dan Reliabilitas Data 29](#_Toc180681512)

[DAFTAR PUSTAKA 30](#_Toc180681513)

# BAB I

# PENDAHULUAN

## A. Latar Belakang

Peningkatan kualitas pendidikan menjadi salah satu prioritas utama dalam kebijakan pembangunan di Indonesia yang harus terus diprioritaskan. Salah satu faktor penting yang memengaruhi pencapaian tersebut adalah peran guru dalam memberikan pengajaran yang efektif dan berkualitas. Kemajuan kualitas serta keaktifan siswa dalam pembelajaran menjadi elemen dasar keberhasilan proses pendidikan, meskipun tidak semua siswa memiliki kecepatan yang sama dalam memahami materi yang diajarkan (Triandini, Defit, dan Nurcahyo, 2021). Hal ini tentu menjadi perhatian khusus bagi pihak sekolah dalam upaya mencapai keberhasilan pendidikan.

Setiap siswa memiliki potensi untuk mengembangkan hard skills dan soft skills yang dapat mendukung mereka dalam meraih kesuksesan di masa depan. Hard skills umumnya berkaitan dengan keahlian teknis atau keterampilan yang bisa dipelajari melalui pendidikan formal dan pelatihan (Budiningsih et al., 2020). Di sisi lain, soft skills mencakup kemampuan sosial, kepribadian, dan keterampilan interpersonal, seperti kemampuan berkomunikasi, bekerja sama, memimpin, serta beradaptasi (Ubaydillah, 2019).

Namun, tidak semua siswa mampu memaksimalkan potensi yang dimiliki sepenuhnya. Sebagian mungkin merasa kurang tertarik atau tidak nyaman dalam mengembangkan kemampuan tertentu, sementara yang lain mungkin kekurangan akses terhadap sumber daya atau peluang yang dibutuhkan untuk mengasah kemampuan tersebut. Oleh karena itu, sangat penting bagi siswa untuk mengembangkan dan memanfaatkan potensi hard skills dan soft skills mereka secara optimal. Hal ini tidak hanya membantu mereka meraih tujuan dalam karir dan kehidupan, tetapi juga memberikan kontribusi positif bagi masyarakat dan lingkungan sekitar (Siagian, 2020).

Saat ini, proses pemilihan siswa berprestasi di sekolah masih memiliki beberapa kelemahan. Salah satu masalah utamanya adalah pengolahan data yang memakan waktu lama dan rentan terhadap kesalahan manusia dalam pengambilan keputusan (Zulfahmi & Faradika, 2019). Meskipun pemilihan siswa berprestasi penting untuk memberikan penghargaan dan pengakuan atas prestasi siswa, metode yang digunakan saat ini oleh sekolah belum optimal. Proses tersebut seringkali memakan waktu lama dan membutuhkan banyak tenaga untuk mengumpulkan serta mengolah data siswa, yang pada akhirnya dapat mengganggu kegiatan operasional sekolah sehari-hari.

Selain itu, kesalahan manusia dalam pengambilan keputusan juga sering menjadi masalah dalam proses pemilihan siswa berprestasi. Hal ini dapat menyebabkan ketidakadilan, baik dengan memberikan penghargaan kepada siswa yang tidak pantas menerimanya, maupun mengabaikan siswa yang sebenarnya layak. Untuk mengatasi masalah ini, inovasi dan perbaikan dalam proses pemilihan siswa berprestasi sangat diperlukan. Diharapkan guru dapat memilih siswa dengan lebih adil dan tepat.

Setiap tahun, jumlah data siswa terus meningkat, yang menyebabkan akumulasi data yang belum dikelola dengan baik untuk mengungkap informasi dan pengetahuan baru melalui pola-pola yang terbentuk dari akumulasi tersebut. Peningkatan volume data ini menuntut penggunaan strategi dan metode yang tepat untuk mentransformasikan data tersebut menjadi informasi dan pengetahuan yang dapat digunakan oleh pendidik dalam proses pengambilan keputusan dan kebijakan. Guru dan wali kelas dihadapkan pada tantangan dalam menentukan tingkat prestasi siswa, apakah rendah, cukup, atau tinggi, serta mengidentifikasi murid dengan peringkat tertinggi dalam kelas. Dengan demikian, kelas yang ideal dapat terbentuk guna meningkatkan prestasi dan memotivasi siswa.

Salah satu solusi yang bisa diterapkan adalah penggunaan teknologi untuk mengotomatisasi dan meningkatkan akurasi pengolahan data, sehingga proses pengambilan keputusan menjadi lebih cepat dan efisien (Marius Robert Seran, 2016). Selain itu, sekolah juga harus memastikan bahwa aspek keadilan menjadi prioritas dalam pemilihan siswa berprestasi, agar siswa yang memang layak mendapatkan penghargaan bisa diakui dan diapresiasi dengan baik.

Penelitian ini dilakukan di SMP 28 Sarolangun dengan tujuan mengidentifikasi siswa berprestasi di kelas unggulan menggunakan metode clustering dengan algoritma *K-Means*. *Clustering* menurut Nainggolan dan Purba (2020), merupakan proses pengelompokan data ke dalam beberapa kelompok yang memiliki kesamaan maksimum. *K-Means* adalah salah satu algoritma yang populer digunakan untuk tujuan ini karena mampu mengelompokkan objek dalam jumlah besar berdasarkan kesamaan melalui prosedur partisi iteratif.

Proses ini sering dimanfaatkan dalam data *mining* untuk mengidentifikasi kelompok alami dari kasus yang serupa, yang kemudian dapat dianalisis lebih lanjut (Arofah dan Marisa, 2018). Dalam konteks pengembangan fitur *e-learning* untuk rekomendasi siswa berprestasi di SMP 28 Sarolangun, metode *K-Means Clustering* diterapkan untuk mengolah data berdasarkan beberapa kriteria, yaitu penilaian akademik (meliputi nilai tugas harian, UTS, dan UAS), penilaian dari wali kelas (terkait sikap sosial siswa), keaktifan dalam kegiatan ekstrakurikuler, riwayat prestasi, dan persentase kehadiran siswa.

Penelitian ini menetapkan tiga *cluster*, yaitu cluster 1 (C1) untuk siswa dengan nilai tinggi, cluster 2 (C2) untuk siswa dengan nilai sedang, dan cluster 3 (C3) untuk siswa dengan nilai rendah. Melalui pengembangan sistem ini, diharapkan pengelompokan siswa dapat dilakukan secara akurat sesuai dengan kemampuan dan penilaian mereka, sehingga membantu sekolah dalam memberikan perhatian yang sesuai berdasarkan tingkat prestasi masing-masing siswa.

Berdasarkan solusi yang telah dipaparkan, diharapkan pengembangan sistem ini mampu menghasilkan kelompok (cluster) siswa yang sesuai dengan kemampuan dan penilaian mereka. Berdasarkan latar belakang tersebut, peneliti memilih judul **"Penerapan Algoritma K-means Clustering untuk Identifikasi Siswa Berprestasi di Kelas Unggulan pada SMP 28 Sarolangun."**

## B. Identifikasi Masalah

Berdasarkan latar belakang di atas, masalah yang diidentifikasi dalam penelitian ini adalah sebagai berikut:

1. Waktu Pengolahan Data yang Lama: Proses pengolahan data siswa berprestasi dengan metode manual membutuhkan waktu yang lama, sehingga kurang efisien dalam penggunaannya.
2. Potensi Kesalahan Manusia: Penggunaan metode manual dalam mengelompokkan siswa rentan terhadap kesalahan manusia, yang dapat berdampak pada keakuratan hasil.
3. Ketidakadilan dalam Penilaian: Pengelompokan siswa secara manual berpotensi menimbulkan ketidakadilan, karena tidak adanya standar yang objektif dalam penentuan siswa berprestasi.
4. Kurangnya Efisiensi dalam Pengelompokan Siswa: Sistem yang ada tidak mendukung pengelompokan otomatis berdasarkan kriteria yang relevan, seperti nilai akademik, sikap sosial, keaktifan ekstrakurikuler, riwayat prestasi, dan tingkat absensi.
5. Kebutuhan Akan Sistem yang Lebih Akurat dan Efisien: Sekolah membutuhkan sistem yang dapat mengidentifikasi siswa berprestasi secara akurat dan efisien, dengan mempertimbangkan berbagai variabel terkait prestasi siswa.

## C. Batasan Masalah

Penelitian ini membatasi ruang lingkup pada penerapan algoritma *K-Means Clustering* untuk mengidentifikasi siswa berprestasi dengan menggunakan data nilai akademik, sikap sosial, keaktifan ekstrakurikuler, riwayat prestasi, dan absensi dari e-learning. Penelitian ini tidak mencakup metode pengelompokan lain, faktor personal eksternal, atau evaluasi dampak jangka panjang terhadap hasil belajar siswa. Pembatasan ini memastikan fokus penelitian pada aplikasi *K-Means Clustering* dan relevansi hasil dalam konteks yang diteliti.

## D. Rumusan Masalah

1. Bagaimana cara menerapkan algoritma *K-Means Clustering* untuk mengidentifikasi siswa berprestasi di kelas unggulan?
2. Variabel apa saja yang paling efektif untuk digunakan dalam algoritma *K-Means Clustering* dalam konteks pemilihan siswa berprestasi?
3. Sejauh mana algoritma *K-Means Clustering* dapat meningkatkan akurasi dan efisiensi dalam proses pemilihan siswa berprestasi dibandingkan dengan metode manual saat ini?

## E. Tujuan Penelitian

* 1. Mengimplementasikan algoritma *K-Means Clustering* dalam pengembangan sistem untuk identifikasi siswa berprestasi di kelas unggulan pada SMP 28 Sarolangun, sehingga memungkinkan pengelompokan siswa berdasarkan data yang objektif.
  2. Mengevaluasi variabel-variabel seperti nilai akademik, penilaian sikap sosial, keaktifan ekstrakurikuler, riwayat prestasi, dan persentase kehadiran guna menemukan variabel yang paling signifikan dalam mengoptimalkan algoritma *K-Means Clustering* untuk pemilihan siswa berprestasi.
  3. Meningkatkan akurasi dan efisiensi sistem pemilihan siswa berprestasi dengan cara membandingkan metode berbasis *K-Means Clustering* dengan metode manual yang saat ini digunakan, sehingga sistem yang dihasilkan mampu mengurangi potensi kesalahan dan ketidakadilan dalam proses pemilihan siswa berprestasi.

## F. Manfaat Penelitian

1. Teoritis

Penelitian ini memberikan kontribusi pada pemahaman tentang penerapan algoritma *K-Means Clustering* dalam konteks pendidikan, khususnya dalam mengidentifikasi siswa berprestasi. Hasilnya dapat memperkaya literatur mengenai penggunaan teknik data mining untuk analisis pendidikan dan pengelompokan siswa.

1. Praktis

Penelitian ini memberikan manfaat langsung kepada pihak sekolah dengan menyediakan metode yang lebih akurat dan efisien untuk mengidentifikasi siswa berprestasi. Dengan menggunakan algoritma *K-Means Clustering*, sekolah dapat meningkatkan proses pemilihan penghargaan, memotivasi siswa, dan menyusun strategi pengajaran yang lebih tepat sasaran.

# BAB II

# TINJAUAN PUSTAKA

## A. Data *Mining*

1. Pengertian Data *Mining*

Data *mining* secara sederhana adalah proses menemukan informasi baru dengan mencari pola atau aturan dari sejumlah data yang sangat besar (Davies, 2004). Proses ini juga didefinisikan sebagai upaya untuk mengekstrak pengetahuan yang sebelumnya tidak diketahui secara manual dari kumpulan data (Pramudiono, 2007). Data *mining* juga sering disebut sebagai Knowledge Discovery in Database (KDD), yaitu proses pengumpulan dan pemanfaatan data historis untuk menemukan keteraturan, pola, atau hubungan dalam data berukuran besar (Santoso, 2007).

Pada dasarnya, data *mining* bertujuan menemukan pola-pola yang signifikan dari data dalam jumlah besar, yang dapat disimpan dalam basis data, data warehouse, atau sumber penyimpanan lainnya. Bidang ini melibatkan beberapa disiplin ilmu seperti sistem basis data, statistik, machine learning, dan komputasi tingkat tinggi, serta didukung oleh disiplin lain seperti jaringan saraf tiruan, pengenalan pola, dan analisis data spasial (Han, 2006). Proses penemuan pola ini dapat berlangsung secara otomatis atau semi-otomatis, dengan pola yang dihasilkan harus bermakna dan menguntungkan, biasanya dalam konteks ekonomi (Witten, 2005).

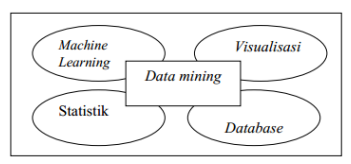
Karakteristik data *mining* mencakup penemuan pola tersembunyi yang sebelumnya tidak diketahui, penggunaan data dalam skala besar untuk hasil yang lebih terpercaya, serta penerapannya dalam pengambilan keputusan strategis yang kritis (Davies, 2004).

Berdasarkan berbagai definisi tersebut, dapat disimpulkan bahwa data *mining* adalah teknik untuk menggali informasi yang berharga dari kumpulan data yang sangat besar, sehingga dapat ditemukan pola menarik yang sebelumnya tidak diketahui. Istilah "*mining*" merujuk pada usaha untuk mendapatkan barang berharga dari bahan dasar dalam jumlah besar, yang menunjukkan bahwa data *mining* memiliki akar dalam kecerdasan buatan, machine learning, statistik, dan basis data. Beberapa metode yang umum digunakan dalam data *mining* meliputi clustering, classification, association rules mining, jaringan saraf tiruan, dan algoritma genetik (Pramudiono, 2007).

1. Pengenalan Pola, Data *Mining*, dan Machine *Learning*

Pengenalan pola adalah disiplin ilmu yang mempelajari cara mengklasifikasikan objek ke dalam berbagai kategori atau kelas, serta mengenali pola dan tren dari data. Objek-objek ini dapat berupa pasien, mahasiswa, pemohon kredit, gambar, sinyal, atau pengukuran lain yang perlu diklasifikasikan atau dicari fungsi regresinya, tergantung pada aplikasinya (Santoso, 2007). Data mining, yang juga dikenal sebagai *knowledge discovery in database* (KDD), melibatkan penggunaan data historis untuk menemukan pola, keteraturan, atau hubungan dalam dataset yang besar. Hasil dari proses data mining ini dapat digunakan untuk mendukung pengambilan keputusan di masa depan, sehingga istilah *pattern recognition* jarang digunakan karena dianggap sebagai bagian dari data mining itu sendiri (Santoso, 2007).

*Machine learning* adalah cabang dari kecerdasan buatan yang berfokus pada pengembangan teknik yang dapat diprogram untuk belajar dari data masa lalu. Pengenalan pola, data *mining*, dan *machine learning* sering digunakan secara bergantian karena memiliki tujuan yang mirip. Ketiga bidang ini berkaitan erat dengan probabilitas, statistik, dan kadang juga optimasi. Dalam konteks data *mining*, *machine learning* berfungsi sebagai alat analisis utama. Hubungan antara ketiga bidang ini dapat digambarkan dalam diagram interaksi yang menjelaskan keterkaitan mereka (Santoso, 2007).



Gambar 2.1 Hubungan Bidang Ilmu Data *mining*, *Manchine learning* dan Pengenalan Pola

1. Tahap-Tahap Data *Mining*

Sebagai suatu proses berurutan, data mining terdiri dari beberapa tahapan yang digambarkan secara interaktif, di mana pengguna terlibat langsung atau melalui bantuan *knowledge base* (Han, 2006). Ada enam tahap utama dalam data mining:

1. Pembersihan Data (*Data Cleaning*)

Tahap ini melibatkan penghilangan *noise* dan data yang tidak konsisten atau tidak relevan. Data yang diperoleh sering kali tidak lengkap, memiliki kesalahan, atau tidak sesuai dengan tujuan analisis. Proses ini membantu mengurangi jumlah dan kompleksitas data, sehingga meningkatkan performa teknik data mining.

1. Integrasi Data (*Data Integration*)

Pada tahap ini, data dari berbagai sumber digabungkan ke dalam satu database baru. Penggabungan atribut-atribut penting seperti nama, produk, atau nomor pelanggan dilakukan secara hati-hati untuk menghindari kesalahan yang bisa menyebabkan hasil analisis yang salah.

1. Seleksi Data (*Data Selection*)

Tidak semua data dalam database digunakan, hanya data yang relevan dengan analisis yang akan dipilih. Sebagai contoh, dalam *market basket analysis*, hanya ID pelanggan yang diperlukan, bukan nama pelanggan.

1. Transformasi Data (*Data Transformation*)

Data diubah ke dalam format yang sesuai untuk analisis. Beberapa metode data mining, seperti *clustering* dan analisis asosiasi, membutuhkan data dalam format tertentu, seperti data kategorikal. Oleh karena itu, data numerik sering kali perlu diubah atau dibagi menjadi interval.

1. Proses Mining

Ini adalah tahap utama di mana metode data mining diterapkan untuk menemukan pengetahuan tersembunyi dan berharga dari data.

1. Evaluasi Pola (*Pattern Evaluation*)

Pada tahap ini, pola yang ditemukan dievaluasi untuk menentukan apakah sesuai dengan hipotesis. Jika hasil tidak memadai, proses ini dapat diulang dengan metode yang berbeda atau hasil tersebut dapat diterima sebagai temuan yang tidak terduga namun berguna.

1. Presentasi Pengetahuan (*Knowledge Presentation*)

Tahap terakhir ini melibatkan visualisasi dan penyajian hasil dalam bentuk yang mudah dipahami oleh semua pihak. Hal ini penting terutama jika keputusan yang diambil melibatkan orang yang tidak memahami data mining. Visualisasi dapat membantu menyampaikan hasil analisis dengan lebih jelas (Han, 2006).

## B. *Clustering*

* 1. Pengertian *Clustering*

Pengelompokan atau *clustering* adalah proses mengkategorikan data ke dalam kelompok berdasarkan kemiripan yang dimiliki. Setiap kelompok atau *cluster* terdiri dari objek-objek yang memiliki karakteristik serupa satu sama lain, tetapi berbeda dengan objek-objek di *cluster* lain. Tujuan dari pengelompokan adalah membagi semua data menjadi kelompok-kelompok yang homogen di mana kemiripan antar objek di dalam satu *cluster* adalah yang tertinggi, sementara kemiripan dengan objek di *cluster* lain minimal (Nabila et al., 2021).

Dalam implementasinya, terdapat beberapa metode yang umum digunakan untuk *clustering*. Dua pendekatan utama adalah pendekatan partisi dan pendekatan hierarkis. Pendekatan partisi, juga dikenal sebagai metode partisi, adalah teknik pengelompokan di mana data dibagi ke dalam *cluster* yang telah ditentukan berdasarkan hasil analisis. Sedangkan pendekatan hierarkis adalah teknik pengelompokan yang menggunakan struktur hierarki berbasis dendrogram. Dendrogram adalah diagram yang menampilkan proses pengelompokan hierarkis, menggambarkan bagaimana data berevolusi dari kelompok kecil menjadi kelompok yang lebih besar. Dalam metode ini, data yang memiliki tingkat kemiripan tinggi akan dikelompokkan berdekatan, sementara data yang kurang mirip akan dikelompokkan lebih jauh (Karlina & Sanoyo, 2021).

* 1. Metode *Clustering*

Secara umum, terdapat berbagai metode untuk melakukan klasterisasi data, yang dipilih berdasarkan jenis data dan tujuan klasterisasi tersebut. Berikut adalah metode-metode klasterisasi beserta algoritma yang termasuk di dalamnya (Baskoro, 2010):

1. Metode *Partitioning*

Metode ini membentuk beberapa partisi dan mengevaluasi setiap partisi menggunakan kriteria tertentu. Algoritma yang termasuk dalam metode ini adalah K-Means, K-Medoid, Proclus, Clara, dan PAM.

1. Metode *Hierarki*

Metode ini membentuk pengelompokan secara hierarkis dari sekumpulan data berdasarkan kriteria tertentu. Ada dua jenis pendekatan dalam metode ini, yaitu *Agglomerative* (dari bawah ke atas) dan *Divisive* (dari atas ke bawah). Algoritma yang digunakan dalam metode ini antara lain Birch, Cure, dan Chameleon.

1. Metode Berbasis Densitas

Metode ini mengandalkan konektivitas dan fungsi densitas data. Algoritma yang digunakan dalam metode ini meliputi DBSCAN, OPTICS, dan DENCLU.

1. Metode Berbasis *Grid*

Metode ini menggunakan struktur *multi-level granularitas*. Algoritma yang termasuk dalam metode ini adalah STING, WaveCluster, dan CLIQUE.

1. Metode Berbasis Model

Dalam metode ini, setiap klaster diasumsikan memiliki model tertentu, dan tujuan utamanya adalah menemukan kecocokan terbaik untuk model tersebut. Algoritma yang termasuk dalam metode ini mencakup pendekatan statistik seperti COBWEB dan jaringan syaraf tiruan seperti SOM.

## C. K-Means

*K-Means Clustering* merupakan salah satu metode dalam analisis *Cluster* yang bertujuan untuk membagi objek ke dalam sejumlah *Cluster*, di mana setiap objek akan ditempatkan dalam *Cluster* yang rata-ratanya paling dekat dengan objek tersebut. Algoritma ini adalah salah satu algoritma yang cukup sederhana dan terkenal, serta mudah dipelajari sebagai cara untuk menyelesaikan permasalahan dalam pengelompokan data. *K-Means Clustering* bekerja dengan membagi data ke dalam k kelompok, di mana k merupakan parameter masukan. Setiap pengamatan atau data akan dikelompokkan ke dalam *Cluster* berdasarkan kedekatannya dengan rata-rata *Cluster* tersebut, yang dihitung secara berulang selama proses berlangsung (Kamila, Khairunnisa, dan Mustakim, 2019).

Langkah-langkah yang dilakukan dalam penerapan *K-Means Clustering* menurut Solichin dan Khairunnisa (2020) adalah sebagai berikut:

1. Menentukan nilai k, yang merupakan jumlah klaster yang diinginkan.
2. Memilih nilai k secara acak sebagai titik pusat awal untuk *Cluster*.
3. Setiap titik dalam kumpulan data kemudian dikelompokkan ke dalam salah satu klaster berdasarkan jarak Euclidean antara titik data dan pusat *Cluster* yang telah dipilih.

Berikut adalah rumus jarak Euclidean:



Dalam rumus ini, D(*i,j*) adalah jarak antara data ke-i dengan centroid ke-j. X\_ki ​merujuk pada data ke-i dalam cluster pada atribut data ke-k, sementara C*kj* ​ adalah centroid ke-j pada atribut data ke-k. n\_k adalah jumlah atribut data pada ke-k. Dengan rumus ini, jarak terdekat antara data dan centroid dapat dihitung untuk menentukan *Cluster* yang paling sesuai bagi data tersebut.

Selanjutnya, data dikelompokkan berdasarkan jarak terdekat antara data dengan pusat *Cluster*. Proses ini menentukan *Cluster* mana yang menjadi tempat bagi data tertentu. Setelah data dikelompokkan, nilai centroid akan diperbarui dengan menghitung rata-rata dari seluruh data dalam setiap *Cluster* berdasarkan anggota baru. Rumus yang digunakan untuk memperbarui centroid adalah:



Dalam rumus ini, C\_ki ​ adalah titik pusat suatu *Cluster* pada atribut ke-k dan *Cluster* ke-i. X\_kq ​ adalah data ke-q dalam *Cluster* ke-i pada atribut ke-k, dan n\_kin merupakan jumlah data dalam *Cluster* ke-i pada atribut ke-k.

Proses iterasi dilakukan dengan memperbarui centroid dan menentukan pengelompokan berdasarkan jarak terdekat menggunakan rumus jarak Euclidean. Iterasi ini berlanjut hingga tidak ada lagi perubahan dalam keanggotaan *Cluster* atau *Cluster* yang telah terbentuk. Ketika data tidak lagi berpindah *Cluster*, iterasi berhenti dan pembentukan *Cluster* selesai.

## D. Siswa Berprestasi

1. Pengertian Siswa

Siswa adalah individu yang sedang menempuh pendidikan, baik di tingkat sekolah dasar, sekolah menengah pertama (SMP), maupun sekolah menengah atas (SMA). Mereka belajar untuk memperoleh pengetahuan serta mencapai pemahaman dalam dunia pendidikan. Beberapa ahli memberikan definisi mengenai siswa, antara lain:

* + 1. Menurut Arifin (2000), siswa adalah individu yang sedang berada dalam proses pertumbuhan dan perkembangan sesuai dengan fitrahnya. Mereka membutuhkan bimbingan dan pengarahan yang konsisten untuk mencapai kemampuan optimal berdasarkan potensi fitrah yang dimiliki.
    2. Sarwono (2007) mendefinisikan siswa sebagai seseorang yang secara resmi terdaftar untuk mengikuti kegiatan pembelajaran di institusi pendidikan.

Dari definisi tersebut, siswa dapat dipahami sebagai status yang diberikan kepada seseorang dalam konteks pendidikan, di mana mereka diharapkan menjadi calon intelektual yang siap menjadi generasi penerus bangsa.

1. Pengertian Prestasi

Prestasi adalah hasil yang dicapai seseorang melalui berbagai aktivitas yang dilakukannya. Setiap kegiatan yang berhasil diselesaikan dapat dianggap sebagai suatu prestasi. Prestasi juga mencerminkan keterampilan atau hasil nyata yang diperoleh dalam jangka waktu tertentu. Dalam konteks penelitian ini, prestasi merujuk pada pencapaian siswa dalam proses pembelajaran.

Menurut Gagne (1985:40), prestasi belajar dibagi menjadi lima aspek, yaitu:

1. Kemampuan intelektual – mencakup kemampuan berpikir dan memecahkan masalah.
2. Strategi kognitif – melibatkan kemampuan mengelola dan mengatur proses berpikir.
3. Informasi verbal – berkaitan dengan kemampuan mengingat dan menyampaikan informasi secara lisan.
4. Sikap – meliputi pandangan atau perasaan siswa terhadap pembelajaran.
5. Keterampilan – merujuk pada kemampuan praktis yang diperoleh melalui latihan atau pengalaman.

Prestasi belajar ini mengukur berbagai kemampuan siswa yang berkembang selama proses pendidikan.

1. Faktor- Faktor Yang Mempengaruhi Prestasi

Faktor yang mempengaruhi prestasi siswa dapat dibagi menjadi dua kategori utama, yaitu faktor internal dan faktor eksternal:

* 1. Faktor Internal

Faktor internal adalah hal-hal yang berasal dari dalam diri siswa itu sendiri. Faktor-faktor ini meliputi:

* + 1. Faktor Jasmaniah: berkaitan dengan kesehatan fisik dan kondisi tubuh, termasuk adanya cacat tubuh yang bisa memengaruhi kemampuan belajar.
    2. Faktor Psikologis: mencakup berbagai aspek seperti tingkat kecerdasan (inteligensi), perhatian, minat, bakat, motif, kematangan, dan kesiapan mental siswa dalam menerima pelajaran.
    3. Faktor Kelelahan: rasa lelah, baik secara fisik maupun mental, dapat memengaruhi kinerja siswa dalam belajar.
  1. Faktor Eksternal

Faktor eksternal adalah pengaruh dari lingkungan luar yang memengaruhi prestasi siswa. Faktor ini terdiri dari:

* + 1. Faktor Keluarga: mencakup cara orang tua mendidik anak, hubungan antar anggota keluarga, suasana di rumah, keadaan ekonomi keluarga, pemahaman orang tua tentang pendidikan, dan latar belakang budaya keluarga.
    2. Faktor Sekolah: melibatkan hal-hal seperti metode mengajar guru, kurikulum, hubungan antara guru dan siswa, hubungan antar siswa, disiplin di sekolah, ketersediaan alat pembelajaran, waktu belajar, standar belajar yang ditetapkan, kondisi bangunan sekolah, metode belajar, dan tugas-tugas rumah.
    3. Faktor Masyarakat: termasuk kegiatan siswa dalam masyarakat, pengaruh media massa, lingkungan pergaulan siswa, dan bentuk kehidupan sosial di sekitarnya. Kedua faktor ini, baik internal maupun eksternal, memiliki peran penting dalam menentukan prestasi akademik seorang siswa.

## E. Kelas Unggulan

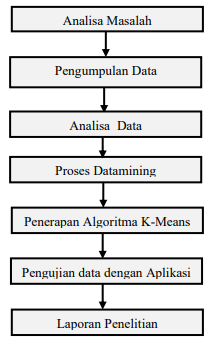
Kelas unggulan, sebagaimana dijelaskan dalam buku pedoman yang diterbitkan oleh Direktorat Pendidikan Dasar (1996:199), adalah sebuah kelompok siswa yang dipilih berdasarkan prestasi akademik yang luar biasa. Siswa-siswa ini ditempatkan dalam kelas tertentu dengan program pengajaran yang disesuaikan dengan kurikulum yang dikembangkan serta tambahan materi pelajaran di bidang tertentu.

Dasar hukum untuk penyelenggaraan kelas unggulan tercantum dalam Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional, yang menggantikan Undang-Undang Nomor 2 Tahun 1989. Pada Bab IV Pasal 5 Ayat 4, disebutkan bahwa warga negara dengan kecerdasan atau bakat istimewa berhak menerima pendidikan khusus. Selain itu, Bab V Pasal 12 Ayat 1 menegaskan bahwa setiap peserta didik di setiap jenjang pendidikan berhak mendapatkan layanan pendidikan yang sesuai dengan bakat, minat, dan kemampuannya.

SK Menteri Pendidikan dan Kebudayaan Nomor 054/U/1993 juga mengatur layanan pendidikan bagi siswa yang memiliki kecerdasan luar biasa atau bakat istimewa, dengan menyediakan program khusus dan kelas khusus sesuai dengan potensi mereka, seperti yang diatur dalam Pasal 15.

## F. Kerangka Konseptual

Penelitian ini menggunakan algoritma K-Means Clustering untuk mengidentifikasi siswa berprestasi di kelas unggulan pada SMP 28 Sarolangun. Gambar 2.1 menunjukkan kerangka konseptual yang menjadi panduan dalam melaksanakan penelitian ini secara terstruktur.

****

**Gambar 2.1 Kerangka Konseptual**

Diagram alir di Gambar 2.1 menggambarkan serangkaian langkah utama dalam penelitian ini:

* + - 1. Analisa Masalah: Mengidentifikasi dan menganalisis permasalahan yang ingin dipecahkan, yaitu bagaimana mengelompokkan siswa berdasarkan prestasi mereka di kelas unggulan.
      2. Pengumpulan Data: Data yang relevan dikumpulkan, termasuk nilai akademik, catatan kehadiran, partisipasi siswa dalam kegiatan ekstrakurikuler, serta data terkait lainnya yang memengaruhi prestasi siswa.
      3. *Preprocessing Data*: Pada tahap ini, data yang sudah dikumpulkan diproses melalui langkah pembersihan, normalisasi, dan transformasi untuk mempersiapkannya ke tahap analisis lebih lanjut.
      4. *K-Means Clustering*: Algoritma K-Means digunakan untuk mengelompokkan siswa berdasarkan kemiripan data. Ini dapat membantu mengidentifikasi kelompok siswa dengan tingkat prestasi tertentu (misalnya, siswa berprestasi tinggi, sedang, dan rendah).
      5. Pengujian Data dengan Aplikasi: Data yang sudah dikelompokkan melalui proses clustering diuji dengan perangkat lunak atau aplikasi tertentu untuk memastikan hasil yang diperoleh valid dan relevan.
      6. Interpretasi Hasil: Hasil clustering dianalisis untuk mengidentifikasi karakteristik masing-masing kelompok siswa dan menentukan faktor-faktor yang mempengaruhi keberhasilan mereka.
      7. Laporan Penelitian: Seluruh hasil penelitian, termasuk analisis clustering dan temuan terkait, dirangkum dalam laporan untuk memberikan rekomendasi terkait pengelolaan siswa di kelas unggulan.

# BAB III

# METODE PENELITIAN

## A. Jenis Penelitian

Penelitian ini bersifat kuantitatif karena menganalisis data berbentuk angka dan statistik untuk mengidentifikasi pola prestasi siswa. Pendekatan ini memungkinkan pengelompokan siswa berdasarkan kriteria akademik dan non-akademik secara mendalam dan sistematis, sehingga hasilnya dapat digunakan untuk mengoptimalkan pembinaan di kelas unggulan.

## B. Desain Penelitian

Penelitian ini menggunakan metode studi kasus yang berfokus pada analisis prestasi siswa di kelas unggulan SMP 28 Sarolangun. Studi kasus ini memungkinkan peneliti untuk mengeksplorasi secara mendalam fenomena prestasi akademik siswa dalam konteks tertentu. Data diolah menggunakan algoritma K-Means Clustering untuk menemukan pola dan kelompok siswa berdasarkan kriteria prestasi yang relevan.

## C. Lokasi Dan Waktu Penelitian

Tempat penelitian dilakukan di SMPN 28 Sarolangun, sebuah sekolah menengah pertama yang terletak di bawah naungan Kementerian Pendidikan dan Kebudayaan. Lokasi sekolah berada di Jln. Lintas Sumatera Simpang Nibung, Desa Simpang Nibung, Kecamatan Singkut, Kabupaten Sarolangun, Provinsi Jambi.

## D. Populasi dan Sampel

Populasi dalam penelitian ini mencakup seluruh data prestasi siswa di kelas unggulan SMP 28 Sarolangun. Data yang digunakan adalah nilai rata-rata akademik dan indikator prestasi lainnya dari tahun ajaran 2022 hingga 2023.

Sumber data secara umum dibagi menjadi dua, yaitu data primer dan data sekunder. Berikut penjelasannya:

1. Sumber data primer merujuk pada pengembangan fitur *e-learning* yang akan diterapkan pada SMPN 28 Sarolangun. Data primer adalah data penting dan utama yang dibutuhkan oleh sistem; jika tidak ada data tersebut, maka akan mempengaruhi kelancaran dalam pembangunan sistem.
2. Sedangkan data sekunder yang dimaksud adalah data penunjang yang berhubungan dengan *e-learning* di SMPN 28 Sarolangun. Data sekunder adalah data penunjang untuk data primer; jika data tersebut tidak ada, tidak akan berpengaruh pada kelanjutan pembangunan sistem.

Adapun data tersebut dapat dilihat pada Tabel 3.1:

Tabel 3.1 Sumber Data

|  |  |  |
| --- | --- | --- |
| **No** | Data | **Tipe Data** |
| 1 | Nilai Tugas Harian | Primer |
| 2 | Nilai Ujian Tengah Semester (UTS) | Primer |
| 3 | Nilai Ujian Akhir Semester (UAS) | Primer |
| 4 | Penilaian Wali Kelas (Sikap Sosial Siswa) | Primer |
| 5 | Keaktifan dalam Kegiatan Ekstrakurikuler | Primer |
| 6 | Riwayat Prestasi Siswa | Primer |
| 7 | Persentase Kehadiran Siswa | Primer |
| 8 | Data Siswa | Sekunder |
| 9 | Data Kelas | Sekunder |
| 10 | Notifikasi dan Pengumuman dari Sekolah | Sekunder |

## E. Variabel Penelitian

Variabel Independen: Metode K-Means Clustering: Teknik ini digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik prestasi siswa di kelas unggulan SMP 28 Sarolangun.

Variabel Dependen: Pola dan klaster prestasi siswa di kelas unggulan SMP 28 Sarolangun.

## F. Instrumen Penelitian

Instrumen penelitian yang digunakan adalah perangkat lunak **RapidMiner** dan **Tableau**. RapidMiner digunakan untuk menjalankan algoritma K-Means Clustering dalam menganalisis data prestasi siswa. Tableau digunakan untuk visualisasi data, membantu dalam interpretasi hasil analisis. Data yang digunakan dalam penelitian ini diambil dari laporan prestasi siswa di SMP 28 Sarolangun.

## G. Prosedur Pengumpulan Data

Untuk melaksanakan proses K-Means Clustering, diperlukan data yang cukup banyak dan relevan dengan kebutuhan penelitian. Dalam penelitian ini, peneliti menggunakan data siswa SMP 28 Sarolangun, termasuk penilaian akademik, penilaian dari wali kelas, keaktifan dalam kegiatan ekstrakurikuler, riwayat prestasi, dan persentase kehadiran siswa. Pengumpulan data dilakukan dengan mengajukan surat izin penelitian kepada pihak sekolah SMP 28 Sarolangun dan menyertakan proposal penelitian yang menjelaskan tujuan dan metodologi penelitian. Setelah menerima persetujuan dari pihak sekolah, data dapat diambil dari sistem informasi akademik sekolah dengan cara mengunduh data yang telah ditentukan dalam proposal penelitian.

1. *K-Means Clustering*

Setelah data melalui tahap pre-processing dan transformasi, data siap untuk diolah menggunakan metode K-Means Clustering. Metode ini digunakan untuk mengelompokkan data ke dalam beberapa cluster dengan titik pusat yang berbeda untuk setiap cluster. Dalam penelitian ini, peneliti menetapkan penggunaan tiga cluster. Proses K-Means Clustering terdiri dari lima langkah, yaitu:

1. Menentukan Titik Pusat Cluster

Langkah pertama dalam K-Means Clustering adalah menentukan titik pusat untuk setiap cluster. Proses ini bertujuan untuk menentukan titik awal sebagai referensi dalam mencari jarak antara data dan cluster yang telah ditentukan. Titik awal ini dikenal sebagai centroid. Penentuan titik pusat setiap cluster dapat dilakukan dengan menghitung rata-rata dari data yang akan diolah atau berdasarkan preferensi peneliti.

1. Menghitung Jarak Data ke Setiap Cluster

Setelah titik pusat setiap cluster ditentukan, langkah berikutnya adalah menghitung jarak antara data dan setiap cluster yang telah dibentuk. Rumus yang digunakan untuk menghitung jarak (distance) dari satu data ke cluster tertentu adalah sebagai berikut:

Distance = ( 𝑋 𝑖 − 𝑋 avg ) 2 + ( 𝑌 𝑖 − 𝑌 avg ) 2 + ( 𝑍 𝑖 − 𝑍 avg ) 2 ​ Di mana: 𝑋 𝑖 X i ​ , 𝑌 𝑖 Y i ​ , dan 𝑍 𝑖 Z i ​ adalah nilai data dari atribut pertama, kedua, dan ketiga.

𝑋 avg X avg ​ , 𝑌 avg Y avg ​ , dan 𝑍 avg Z avg ​ adalah centroid untuk atribut yang sama. Dalam penelitian ini, karena ada tiga cluster, perhitungan jarak dilakukan tiga kali untuk setiap cluster.

1. Mengalokasikan Data ke dalam Cluster

Setelah mendapatkan jarak antara setiap data dan setiap cluster, langkah selanjutnya adalah mengelompokkan dan mengalokasikan data ke dalam cluster. Proses ini dilakukan dengan memilih data yang memiliki jarak terdekat ke masing-masing cluster. Data akan dimasukkan ke dalam cluster berdasarkan urutan jarak terkecil yang diperoleh.

1. Menentukan Titik Pusat Cluster Baru

Setelah data dialokasikan ke dalam cluster, langkah berikutnya adalah menentukan titik pusat cluster baru. Proses ini dilakukan dengan cara menghitung rata-rata pada setiap atribut hanya untuk data yang masuk ke dalam cluster tersebut. Misalnya, jika data yang masuk ke cluster satu adalah data pertama, ketiga, kelima, dan keenam, maka rata-rata dihitung hanya menggunakan data tersebut.

1. Memverifikasi Titik Pusat Cluster

Setelah mendapatkan titik pusat cluster baru, langkah terakhir adalah memverifikasi apakah titik pusat yang baru sama dengan titik pusat yang lama. Jika hasil titik pusat cluster baru tidak berbeda dari titik pusat yang lama, maka proses K-Means dianggap selesai dan hasil pengelompokan data sudah final. Namun, jika hasilnya berbeda, proses akan diulang mulai dari langkah kedua untuk menghitung jarak data ke setiap cluster kembali.

## H. Prosedur Analisis Data

Tahap *pre-processing* data merupakan langkah penting di mana data yang telah dikumpulkan dipilah dan dipisahkan untuk mendapatkan informasi yang diperlukan dalam proses selanjutnya. Tahapan ini melibatkan beberapa proses yang saling terkait. Berikut adalah proses-proses dalam tahap *pre-processing*:

1. Data *Reduction*

Data *reduction* adalah proses untuk mengurangi dimensi, atribut, atau jumlah data yang tidak relevan dalam file data. Proses ini bertujuan untuk mendapatkan atribut dan data yang diperlukan dalam penelitian ini, sehingga mempermudah analisis.

1. Data *Cleaning*

Data *cleaning* merupakan proses untuk menangani data yang hilang atau tidak lengkap, menghapus duplikasi data, serta memperbaiki kesalahan pengetikan atau ketidaksesuaian lainnya. Dalam tahap ini, data akan disesuaikan agar konsisten, termasuk pengaturan penggunaan huruf besar dan kecil, serta format penulisan angka dan huruf sesuai dengan yang dibutuhkan.

1. Data *Transformation*

Mengingat bahwa metode *K-Means Clustering* membutuhkan data dalam bentuk numerik, tahap transformation ini sangat penting. Proses ini melibatkan pengubahan atribut non-numerik menjadi nilai numerik agar dapat diolah menggunakan algoritma *K-Means Clustering*.

1. Data *Integration*

Data *integration* adalah proses menggabungkan data dari beberapa sumber jika diperlukan. Proses ini meliputi pengintegrasian skema, identifikasi masalah entitas, serta deteksi dan penyelesaian konflik yang mungkin muncul pada nilai data.

## I. Validitas dan Reliabilitas Data

Validitas data dalam penelitian ini diukur dengan memastikan bahwa data yang digunakan adalah sah dan dapat dipercaya. Proses validasi melibatkan pemeriksaan sumber data, yaitu laporan prestasi siswa di SMP 28 Sarolangun yang resmi dan terverifikasi. Selain itu, dilakukan cross-check untuk memastikan data tersebut konsisten dan relevan dengan tujuan penelitian.

Reliabilitas data dijamin dengan menggunakan data dari periode yang konsisten, yaitu dari tahun ajaran 2022 hingga 2023. Data diambil dari sumber yang sama, yaitu laporan prestasi akademik SMP 28 Sarolangun, untuk memastikan bahwa metode pengumpulan dan pelaporan data konsisten.

# DAFTAR PUSTAKA

Arifin, M, (2000) Strategi Belajar Mengajar, Bandung: Jurusan Pendidikan Kimia FMIPA. UPI.

Baskoro, H., 2010, Implementasi Algoritma K-Means Menggunakan Data Penyewaan Alat Berat untuk Melakukan Estimasi Nilai Outcome, Universitas Pembangunan Veteran, Jakarta.

Davies, and Paul Beynon, 2004, “Database Systems Third Edition”, Palgrave Macmillan, New York.

Han, J. dan M. Kamber. 2006. Data Mining Concepts and Techniques Second Edition. San Francisco: Morgan Kaufmann.

Ian H. Witten, E. F. (2005). Data mining: Practical Machine Learning Tools and Techniques, Second Edition. San Francisco: Elsevier Inc.

Kamila, I., Khairunnisa, U. & Mustakim. 2019. Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. Jurnal Ilmiah Rekayasa dan Manajemen Sistem Informasi Vol. 5, No. 1, Februari 2019

Karlina, B., & Sanoyo, A. M. (2021). Pengaruh Cluster Emiten terhadap Return Saham JSX Berbasis Parameter Rasio Analisa Fundamental. Jurnal Akuntansi, Keuangan, Dan Manajemen, 2(4), 279–291. <https://doi.org/10.35912/jakman.v2i4.294>

M. M. K-means, A. Solichin, and K. Khairunnisa, “Klasterisasi Persebaran Virus Corona ( Covid-19 ) Di DKI Jakarta,” vol. 5, no. 2, 2020.

Nabila, Z., Rahman Isnain, A., & Abidin, Z. (2021). Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma KMeans. Jurnal Teknologi Dan Sistem Informasi (JTSI), 2(2), 100.

Santoso, Singgih. 2007. Statistik Deskriptif: Konsep dan Aplikasi dengan Microsoft Exel dan SPSS. Yogyakarta: ANDI.

Sarwono. (2007). Psikologi Remaja. Jakarta: Raja Grafindo Persada