BAB II
LANDASAN TEORI

2.1 Prestasi Nilai Akademik

Prestasi nilai akademik siswa merujuk pada hasil yang diperoleh siswa selama mengikuti kegiatan pembelajaran formal, yang umumnya diwujudkan dalam bentuk angka atau simbol tertentu. Capaian ini merupakan indikator dari proses belajar siswa yang mencakup aspek pemahaman, penerapan, analisis, hingga evaluasi materi. Nilai tersebut biasanya diberikan oleh pendidik melalui berbagai bentuk penilaian, baik berupa tes standar maupun penilaian langsung. Prestasi nilai akademik sering kali menjadi titik fokus utama dalam dunia pendidikan, karena dijadikan tolok ukur keberhasilan belajar siswa serta efektivitas sistem pendidikan itu sendiri. Dalam banyak sistem pendidikan, khususnya yang bersifat kompetitif, prestasi nilai akademik kerap dijadikan sebagai faktor penting yang menentukan akses siswa terhadap jenjang pendidikan selanjutnya maupun peluang karier di masa mendatang. Beberapa fungsi penting prestasi nilai akademik, antara lain :

1. Indikator Kualitas Pembelajaran: Menunjukkan sejauh mana siswa menguasai materi pelajaran
2. Motivasi Belajar: Dapat mendorong siswa untuk meningkatkan usaha belajar mereka
3. Umpan Balik untuk Institusi Pendidikan: Menjadi indikator relevansi kurikulum dan efektivitas metode pengajaran

Beberapa indikator yang digunakan untuk mengukur prestasi nilai akademik siswa meliputi :

1. Nilai Rapor: Dokumen yang mencatat hasil belajar siswa selama periode tertentu.
2. Indeks Prestasi Akademik (IPA): Merupakan angka atau simbol yang menggambarkan hasil belajar siswa secara keseluruhan, sering digunakan di perguruan tinggi.
3. Angka Kelulusan: Menunjukkan keberhasilan siswa dalam menyelesaikan pendidikan di institusi tertentu.

Prestasi nilai akademik memiliki karakteristik sebagai berikut :

1. Dapat Diukur: Melalui tes prestasi yang dirancang untuk menilai kemampuan siswa.
2. Hasil Individu: Prestasi tersebut merupakan hasil usaha individu, bukan hasil dari orang lain.

Secara keseluruhan, prestasi nilai akademik siswa merupakan ukuran penting dalam dunia pendidikan yang tidak hanya mencerminkan kemampuan akademis tetapi juga dapat berfungsi sebagai motivator dan indikator kualitas pendidikan.

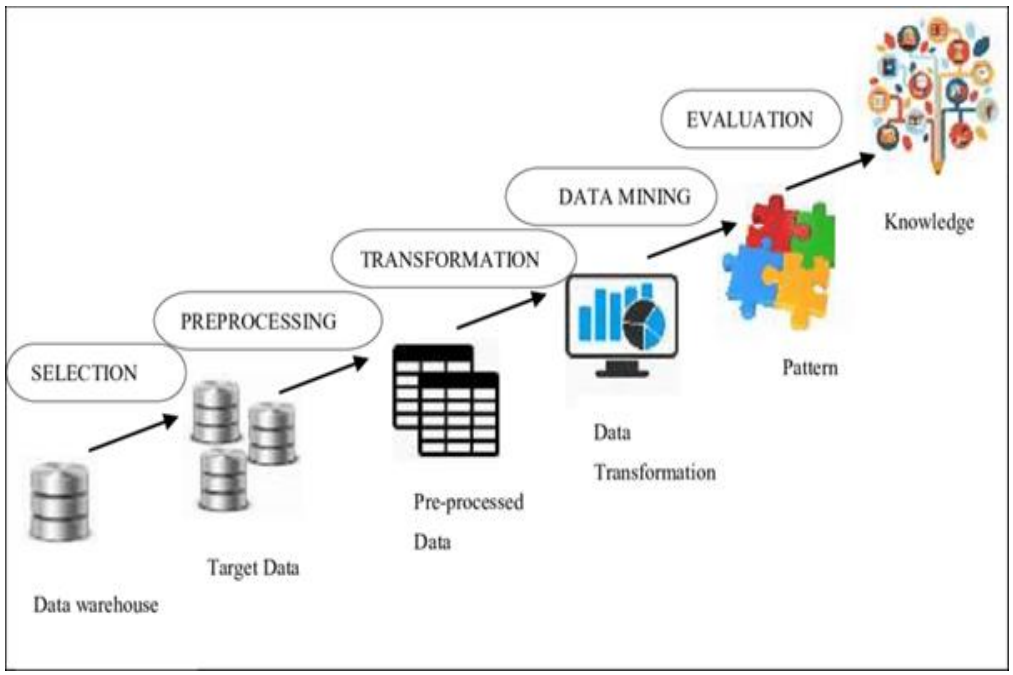
2.2 *Knowledge Discovery In Database* (KDD)

Knowledge Discovery in Database (KDD) merupakan suatu proses yang melibatkan pengumpulan dan pemanfaatan data historis untuk mengidentifikasi pola, keteraturan, atau hubungan dalam kumpulan data yang besar. KDD memiliki keterkaitan yang erat dengan *data mining*, karena bertujuan untuk

menggali informasi tersembunyi yang terdapat dalam basis data. Secara umum, KDD adalah metode untuk mengekstrak pengetahuan dari data berskala besar guna menemukan pola-pola baru yang kemudian dapat menghasilkan informasi serta wawasan yang bermanfaat [1].

Lebih lanjut, KDD merupakan bidang ilmu yang bersifat interdisipliner, melibatkan kontribusi dari statistik, basis data, kecerdasan buatan (artificial intelligence), visualisasi data, dan komputasi paralel. Proses ini memungkinkan penarikan kesimpulan berupa pola atau kecenderungan yang relevan, yang selanjutnya diolah menjadi informasi yang akurat, tepat, dan mudah dipahami oleh pengguna.

KDD adalah sebuah proses untuk mencari dan mengidentifikasi *pattern* dalam sebuah *database*. Pada sebuah *Knowledge Discovery in Database* atau KDD memiliki beberapa tahapan yaitu [2] :



Gambar 2.1 Proses *Knowledge Discovery In Database* (KDD)

1. *Data Selection* (pemilihan data)

   Tidak semua data yang tersedia akan digunakan dalam proses analisis. Oleh karena itu, dilakukan tahap pemilihan data, yaitu menyaring dan memilih data yang relevan sesuai dengan tujuan analisis yang ingin dicapai dari basis data yang tersedia.

2. Pembersihan data dan proses data (*cleaning and processing*)

   Pada tahap ini, data dibersihkan dari unsur-unsur yang mengganggu seperti *noise* atau *inkonsistensi*. Sebelum memasuki proses *data mining*, perlu dilakukan pembersihan data, termasuk menghapus duplikasi, memperbaiki kesalahan, serta menyelaraskan data yang tidak konsisten. Selain itu, dilakukan proses data *enrichment*, yaitu menambahkan informasi tambahan yang relevan untuk memperkaya data awal agar lebih siap digunakan dalam proses KDD.

3. Transformasi data (*transformation*)

   Transformasi merupakan tahapan penting yang bersifat kreatif dan bergantung pada jenis pola yang ingin ditemukan. Data dari basis data akan diubah dengan berbagai teknik agar sesuai dengan kebutuhan analisis. Tahap ini juga membantu menyederhanakan data dan meningkatkan efisiensi komputasi, terutama ketika berhadapan dengan dataset berskala besar.

4. Penambangan data (*data mining*)

   Ini adalah inti dari proses KDD, di mana berbagai pola dan informasi tersembunyi dalam data dicari menggunakan beragam metode atau

algoritma. Teknik yang digunakan dalam data mining sangat beragam dan pemilihannya disesuaikan dengan tujuan analisis dan konteks dari keseluruhan proses KDD.

5. Evaluasi pola dan presentasi pengetahuan (*knowledge extraction*)

   Tahap akhir ini bertujuan untuk menilai apakah pola yang ditemukan bertentangan dengan hipotesis atau fakta yang sudah ada sebelumnya. Di tahap ini, hasil dari data mining diterjemahkan ke dalam bentuk yang dapat dipahami dengan mudah oleh pengguna akhir, sehingga pengetahuan yang diperoleh bisa langsung dimanfaatkan.

2.3 *Data Mining*

Data mining merupakan salah satu tahapan penting dalam proses *Knowledge Discovery in Database* (KDD). Proses ini merujuk pada kegiatan pengolahan data dalam skala besar yang berasal dari rekaman data historis. Informasi yang tersimpan dari masa lalu ini kemudian dieksplorasi kembali untuk menemukan pola atau pengetahuan baru yang relevan dengan kebutuhan saat ini [3].

Saat ini, *data mining* telah menjadi bidang ilmu yang banyak dimanfaatkan karena kemampuannya dalam membantu menyelesaikan permasalahan yang berkaitan dengan data. Terdapat berbagai teknik dalam *data mining* seperti klasifikasi, klasterisasi, prediksi, estimasi, dan asosiasi yang digunakan dalam proses analisis data.

Data mining juga dapat diartikan sebagai proses analitis untuk mengevaluasi kumpulan data dalam rangka menemukan hubungan yang tidak

terduga serta merangkum data dengan cara baru yang mudah dipahami dan bermanfaat bagi pengguna [4]. Selain itu, *data mining* merupakan salah satu penerapan teknologi untuk menemukan model atau pola dalam data historis yang dapat digunakan untuk melakukan prediksi pada waktu yang akan datang. [5].

Berdasarkan penjelasan-penjelasan tersebut, dapat disimpulkan bahwa *data mining* adalah proses penggalian informasi dari sekumpulan data berukuran besar untuk memperoleh pengetahuan (*knowledge*) yang tersembunyi. Pengetahuan yang diperoleh dari proses ini sangat bermanfaat sebagai dasar dalam pengambilan keputusan.

2.4 *Clustering*

Clustering merupakan teknik pengelompokan data yang bertujuan untuk mengelompokkan objek-objek yang memiliki karakteristik serupa ke dalam satu grup atau *cluster*, sedangkan objek yang memiliki perbedaan signifikan ditempatkan pada *cluster* yang berbeda [6].

Secara umum, *clustering* adalah proses pengorganisasian data ke dalam sejumlah kelompok berdasarkan tingkat kemiripan antar data. Objek-objek yang memiliki kesamaan akan dikelompokkan dalam satu *cluster*, sementara objek yang berbeda akan dipisahkan ke *cluster* lainnya. Setiap *cluster* idealnya berisi data yang memiliki tingkat kemiripan yang tinggi. Tingkat kemiripan tersebut biasanya ditentukan berdasarkan perhitungan jarak antar data. Dalam proses ini, jarak antar data dalam satu *cluster* diusahakan sekecil mungkin, sementara jarak antar *cluster* dibuat sejauh mungkin, sehingga data dalam satu

kelompok tampak *homogen*, dan antar kelompok tampak *heterogen*. Pendekatan ini mengasumsikan bahwa terdapat sejumlah parameter penting yang dapat merepresentasikan tingkat kesamaan maupun perbedaan antar objek [7].

2.5 Algoritma *K-Means*

Algoritma *K-Means* merupakan metode pengelompokan data non-hierarki yang digunakan untuk membagi data ke dalam beberapa kelompok berdasarkan kemiripan karakteristik. Data yang memiliki kesamaan akan dimasukkan dalam kelompok (*cluster*) yang sama, sedangkan data yang memiliki perbedaan akan ditempatkan di kelompok yang berbeda. Tujuan utama dari proses klasterisasi ini adalah untuk meminimalkan nilai fungsi objektif, yaitu dengan mengurangi variasi dalam satu *cluster* serta memperbesar perbedaan antar *cluster* [8].

K-Means adalah algoritma klasterisasi interaktif yang membagi kumpulan data menjadi sejumlah *cluster* yang telah ditentukan sebelumnya. Algoritma ini hanya dapat diterapkan pada atribut yang bersifat numerik. Keunggulan *K-Means* terletak pada kesederhanaannya, kecepatan proses, kemudahan penerapan, serta fleksibilitas penggunaannya dalam berbagai aplikasi. Karena itulah, secara historis *K-Means* dianggap sebagai salah satu algoritma yang paling penting dan banyak digunakan dalam bidang *data mining* [9].

Prinsip kerja algoritma *K-Means* adalah mengelompokkan data berdasarkan jaraknya terhadap titik pusat *cluster* (*centroid*). Data akan dikelompokkan sedemikian rupa agar data yang memiliki karakteristik serupa berada dalam satu

cluster, dan data yang berbeda berada di *cluster* lainnya. Adapun langkah-langkah dasar dari algoritma *K-Means* adalah sebagai berikut [10].

1. Menentukan jumlah *Cluster* (*k*) yang diinginkan.
2. Melakukan inisialisasi awal terhadap *k centroid*, yang biasanya dipilih secara acak.
3. Mengelompokkan seluruh data ke dalam *cluster* berdasarkan kedekatan dengan *centroid*. Kedekatan ini dihitung berdasarkan jarak antara data dan pusat *cluster*. Salah satu metode pengukuran yang sering digunakan untuk menghitung jarak tersebut adalah jarak *Euclidean*, yang dirumuskan sebagai berikut:

   $$D(i,j) = \sqrt{(x1i - x1j)^2 \ + \ (x2i - x2j)^2 \ + \cdots + \ (xki - xkj)^2}$$

   Dimana :

   $D(i,j)$ = Jarak data ke i ke pusat Cluster j

   xki = Data ke i pada atribut data ke k

   xkj = Titik pusat ke j pada atribut ke k

4. Hitung kembali pusat *Cluster* dengan keanggotaan *Cluster* yang sekarang. Pusat *Cluster* adalah rata-rata dari semua data atau objek dalam *Cluster* tertentu yang dirumuskan sebagai persamaan berikut:

   $$C = \frac{\Sigma m}{n}$$

   Keterangan:

   C = Merupakan *Centroid* data

   m = Anggota data yang termasuk ke dalam *Centroid* tertentu

   n = Jumlah data yang menjadi anggota *Centroid* tertentu

5. Melakukan perulangan dari langkah ketiga dan keempat, sampai anggota tiap *Cluster* tidak ada yang berubah.

2.6 RapidMiner

RapidMiner merupakan perangkat lunak *open source* yang memberikan keleluasaan bagi pengguna untuk menyesuaikan dan memodifikasi fitur-fiturnya sesuai kebutuhan analisis yang spesifik. Hal ini memberikan fleksibilitas tinggi dalam proses pengolahan data dan memungkinkan pengguna untuk menyesuaikan proses analisis agar sesuai dengan tujuan tertentu [11]. Perangkat ini menawarkan lingkungan kerja yang terpadu dan mendukung berbagai metode analisis, seperti *data mining*, *text mining*, serta analisis prediktif dan deskriptif, yang membuatnya cocok untuk mengolah data dalam skenario yang kompleks secara lebih efektif [12].

Salah satu kelebihan utama RapidMiner terletak pada desain antarmukanya yang ramah pengguna, sehingga bahkan pengguna tanpa latar belakang teknis atau keterampilan pemrograman yang kuat tetap dapat mengoperasikan dan memanfaatkan alat analisis data ini. Tampilan antar muka yang intuitif memungkinkan pembuatan workflow analisis melalui fitur *drag-and-drop*, di mana pengguna dapat mengatur urutan proses seperti data preprocessing, penerapan algoritma *K-Means*, hingga evaluasi hasil *clustering* dengan mudah [13].

Selain itu, RapidMiner mendukung berbagai jenis format data, seperti CSV, file Excel, dan juga dari basis data relasional, sehingga proses impor dan manipulasi data menjadi lebih fleksibel. Fitur visualisasi yang disediakan juga

cukup beragam dan interaktif, mulai dari *scatter plot*, *heatmap*, hingga grafik dinamis lainnya, yang sangat membantu dalam memahami dan mengevaluasi hasil *clustering* secara visual [14].

2.7 Penelitian Terdahulu

Penelitian terdahulu telah banyak mengkaji penggunaan metode *K- Means Clustering* dalam mengelompokkan data besar dan kompleks menjadi kelompok-kelompok yang lebih kecil dan lebih mudah di interpretasikan, yang sangat berguna dalam analisis data. Beberapa penelitian menunjukkan bahwa *K-Means Clustering* mampu memberikan hasil yang memuaskan dalam pengelompokan data dengan tingkat akurasi yang cukup tinggi. Hasil-hasil tersebut memberikan landasan yang kuat untuk menerapkan metode ini untuk mengetahui prestasi nilai akademik siswa di SMKN 3 Rantau Utara.

Tabel 2.1 Penelitian Terdahulu

Penelitian	1
Judul	PENERAPAN *DATA MINING* UNTUK PENGELOMPOKAN SISWA BERDASARKAN NILAI AKADEMIK DENGAN ALGORITMA *K-MEANS*
Nama	Penda Sudarto Hasugian1, Jijon Raphita Sagala2,*
Tahun	2022
Hasil	Kesimpulan yang dapat diambil dari proses penelitian dengan menerapkan tahapan *data mining* dengan algoritma *k-means clustering* untuk melakukan pengelompokan data atau *clustering* terhadap data siswa yang dikelompokkan menjadi 3 *cluster* yaitu *cluster* siswa unggul, cluster sedang dan *cluster* rendah berdasarkan nilai akademik. Hasil penerapan metode *k-means* hitung manual dengan pengujian mengunakan aplikasi rapid miner dari data nilai siswa adalah sama dimana *Cluster 1* (C1) *cluster* unggul terdapat 2 siswa, Siswa002 dan siswa003. *Cluster 2* (C2) yang merupakakn cluster Sedang, dimana ada 4 data yang termasuk didalam *cluster* ini yaitu Siswa006, Siswa008, Siswa009, dan Siswa010. *Cluster 3* (C3) merupakan *cluster*

	rendah dimana ada 4 data yang tergolong kedalam *cluster* ini yaitu Siswa001, Siswa004, Siswa005 dan Siswa007. Proses klasterisasi memberikan hasil klasifikasi pengelompokan data yang efektif. Sehingga dapat menghemat waktu dalam melakukan klasterisari kelas siswa [15].
Penelitian	2
Judul	PENERAPAN ALGORITME *K-MEANS CLUSTERING* UNTUK MENGELOMPOKKAN SISWA BERDASARKAN NILAI AKADEMIK DI SMP NEGERI 207 SSN
Nama	Reza Pahlevi Kurniawan1*, Ferdiansyah2
Tahun	2023
Hasil	Setelah melakukan serangkaian pengujian dan evaluasi pada aplikasi yang telah dibangun menggunakan *Dataset* dan Algoritme yang diusulkan, maka dapat disimpulkan bahwa jumlah *Cluster* yang optimal untuk mengelompokkan data nilai siswa kelas 7 angkatan 2022 SMPN 207 SSN yaitu 3 *Cluster* yang menghasilkan 16 siswa dengan nilai terendah, 92 siswa dengan nilai menengah dan 141 siswa dengan nilai tertinggi. Dari hasil *Clustering* tersebut dievaluasi menggunakan metode *Davies Bouldin Index* yang menghasilkan nilai 1,12388. Penelitian ini juga mendapatkan bahwa *K-Means* memiliki kelemahan yang disebabkan oleh penentuan pusat *Cluster* awal atau *Centroid* yang dilakukan secara acak. Hasil *Clustering* yang diperoleh dari algoritme *K-Means* sangat bergantung pada inisialisasi nilai *Centroid* awal yang ditentukan, hal ini menyebabkan sulitnya mendapatkan nilai *Centroid* awal yang unik, sehingga peneliti menggunakan metode DBI untuk mengevaluasi hasil *Clustering* agar dapat menentukan jumlah *Cluster* yang optimal dan mengetahui kualitas dari hasil *Clustering* yang diperoleh. Dari kelemahan algoritme *K-Means* yang telah diketahui, diharapkan pada penelitian selanjutnya untuk menggunakan algoritme lainnya dan bandingkan hasilnya [16].
Penelitian	3
Judul	PENINGKATAN EFISIENSI PEMANTAUAN KEHADIRAN SISWA MELALUI ANALISIS *K-MEANS CLUSTERING* DI SEKOLAH MENENGAH PERTAMA NEGERI 3 RANCAEKEK, KABUPATEN BANDUNG
Nama	Fitriani Agustina1* , Rudi Kurniawan2 , Tati Suprapti3
Tahun	2024

Hasil	Hasil dari penelitian dan pengamatan yang telah dilakukan, peneliti mendapatkan kesimpulan yang dimana Metode Algoritma *K-Means Clustering* dapat diterapkan dalam melakukan penilaian terhadap Tingkat Kehadiran Siswa pada SMPN 3 Rancaekek Kabupaten Bandung. Penerapan Algoritma *K-Means Clustering* pada Rapidminer menampilkan hasil anggota *Cluster 0* sebanyak 113 anggota, *Cluster 1* sebanyak 43 anggota, *Cluster 2* sebanyak 37 anggota dan *Cluster 3* sebanyak 117 dengan total data 310 anggota. Hasil dari Performance Evaluasi *Davies Bouldin Index* (DBI) terbaik setelah dilakukan 7 percobaan *clustering* yang telah didapatkan pada penelitian ini 4 *Clustering* dengan Nilai DBI nya adalah 0.997 [17] .
Penelitian	4
Judul	JUMLAH *CLUSTER* OPTIMAL DALAM PENGELOMPOKAN SISWA SMK DENGAN METODE *ELBOW K-MEANS CLUSTERING*
Nama	Ninik Tri Hartanti1*, Erni Seniwati2, Rina Pramitasari3 , Irma Rofni Wulandari3
Tahun	2024
Hasil	Berdasarkan semua tahapan dalam penelitian tentang jumlah clusteroptimal untuk pengelompokan jumlah siswa SMK di Kabupaten Magelang diperoleh hasil antara lain:Hasil perhitungan dengan metode *Elbow* didapatkan bahwa banyaknya *cluster* optimal adalah 3. Keluaran dari proses perhitungan dengan menerapkan algoritma *K-Means* adalah terdapat 3 *cluster*, diperoleh 2 wilayah yaitu Mertoyudan, Muntilan di *cluster*1, di *cluster*2 adalah 10wilayah yaitu, Salaman, Tegalrejo, Secang,Salam, Borobudur, Tempuran,Windusari, Kaliangkrik,Kajoran dan Ngablak sedangkan di *cluster3* adalah 9wilayah yaitu Mungkid, Grabak, Bandongan, Sawangan, Pakis, Candimulyo, Dukun, Srumbung, dan Ngluwar.Pembagian kategori pada setiap *cluster* ditentukan berdasarkan kecenderungan siswa untuk memilih melanjutkan studi ke Sekolah Menengah Kejuruan (SMK). *Cluster*1 terdapat 2 wilayah yang terdefinisikan dengan rendahnya antusiasme siswa dalam pemilihan sekolah menengah kejuruan sebagai pilihannya, *cluster3* terdiri dari 9 wilayah terdefinisikan minat siswa yang cukup besar untuk melanjutkan ke SMK, sedangkan *cluster2* merupakan tingginya antusiasme siswa dalam pemilihan sekolah menengah kejuruan di 10 wilayah[18].

Penelitian	5
Judul	PENERAPAN ALGORITMA *K-MEANS CLUSTERING* PADA DATA NILAI SISWA UNTUK MENENTUKAN KELOMPOK PENERIMA BEASISWA
Nama	Sindrawati1), Dodi Syaripudin2), Agung Rachmat Raharja3)
Tahun	2024
Hasil	Hasil pengujian datasetyang berjumlah 109 data menggunakan *software* rapidminer9.0 terbentuk 3 *cluster*. Pada *cluster0* (*Cluster* pertama) terdapat 26 items, pada *cluster1* (*cluster* kedua) terdapat 46 items, dan pada *cluster2* (*cluster* ketiga) terdapat 37 *items*. Dari 3 *cluster* yang terbentuk kemudian di hasilkan *centroid* akhir untuk masing masing *cluster* yaitu *cluster 0* = 1.077, 79.577, 84.923, 81 *cluster 1* = 1.587, 76.717, 80.239, 77.457 dan *cluster 2* =1.135, 79.676, 80.081, 81.459. Dari 3 *cluster* yang terbentuk, *cluster 1* adalah *cluster* yang paling potensial untuk mendapatkan beasiswa dengan jumlah 26 siswa karena rata-rata nilainya yang tinggi. *Cluster 3* dengan jumlah 37 siswa juga berpotensi mendapatkan beasiswa untuk 50% siswa dengan nilai rata-rata tertinggi. Dan di *cluster 2* dengan jumlah 46 siswa tidak berpotensi untuk mendapatkan beasiswa karena nilai rata-rata yang rendah. Hasil dari penelitian ini dapat membantu pihak Sekolah menentukan siapa yang layak mendapatkan beasiswa lanjutan [19].