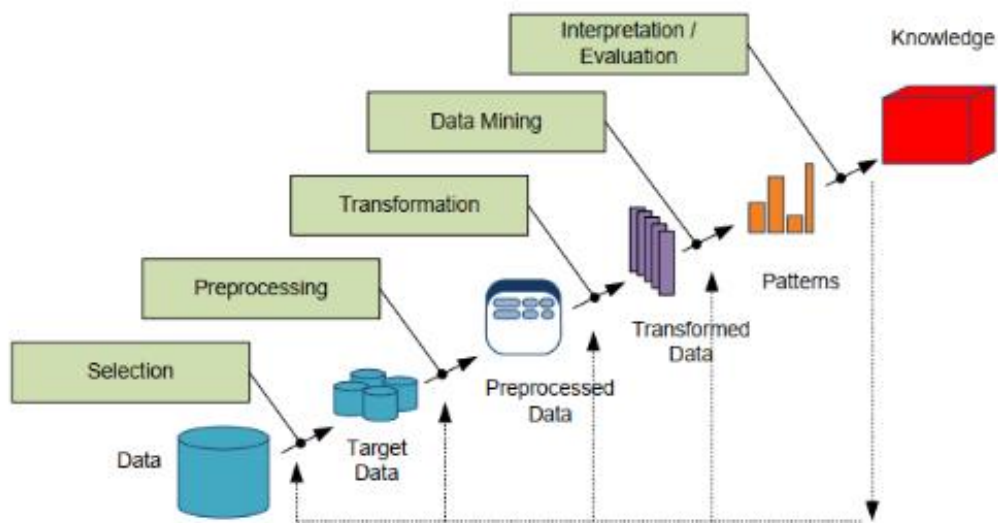


## BAB II

### LANDASAN TEORI

#### 2.1. *Knowledge Discovery in Database*

*Knowledge Discovery in Database* (KDD) merupakan salah satu pendekatan umum dalam analisis data besar. KDD terdiri dari beberapa tahapan yang saling terkait untuk mengolah dan menemukan pola dalam data (Zadia et al., 2025). *Knowledge Discovery in Database* (KDD) adalah proses sistematis untuk mengekstraksi pola, informasi penting, atau hubungan tersembunyi di dalam data, sehingga menghasilkan pengetahuan yang dapat digunakan untuk pengambilan keputusan.



Gambar 2.1 Tahapan Dalam *Knowledge Discovery in Database*

Sumber: Sutoyo & Almaarif (2020)

Salah satu langkah penting dalam rangkaian proses KDD adalah data mining, sebagaimana terlihat pada alur proses berikut:

1. *Selection* (Seleksi Data)

Pada tahap ini, peneliti menentukan data mana yang relevan untuk dianalisis sesuai tujuan penelitian. Pemilihan dilakukan untuk menghindari data redundan sehingga proses analisis lebih efektif.

2. *Preprocessing/Cleaning* (Pembersihan Data)

Tahap ini bertujuan menghilangkan kesalahan seperti data hilang, duplikasi, atau inkonsistensi. Data yang bersih menghasilkan pemodelan yang lebih akurat.

3. *Transformation* (Transformasi)

Data kemudian diubah ke format yang sesuai, misalnya melalui normalisasi, pengkodean kategori, atau penyusunan kembali struktur tabel agar siap untuk dianalisis dengan algoritma tertentu.

4. Data Mining

Pada tahap Ini merupakan inti dari KDD, yaitu penerapan metode atau algoritma seperti klasifikasi, klusterisasi, atau prediksi untuk menemukan pola terstruktur dalam data.

5. *Interpretasi dan Evaluasi*

Hasil yang diperoleh harus dipahami, divalidasi, dan dievaluasi agar dapat disimpulkan menjadi pengetahuan baru yang bermakna bagi pengambil keputusan.

## 2.2. Data Mining

Data mining atau penambangan data merupakan rangkaian proses untuk menemukan pengetahuan atau pola tertentu dari suatu kumpulan data. Melalui proses ini, berbagai masalah dapat diselesaikan dengan menganalisis data yang telah tersedia di dalam database(Sutoyo & Almaarif, 2020).

Data mining yang secara khusus diterapkan pada data pendidikan disebut *Education Data Mining (EDM)*. EDM merupakan bidang ilmu yang berkembang untuk menemukan informasi penting yang sebelumnya tersembunyi dari berbagai kumpulan data pendidikan. Informasi yang ditemukan tersebut kemudian dianalisis agar dapat membantu menyelesaikan berbagai permasalahan di dunia pendidikan, seperti pengembangan e-learning, penilaian kinerja mahasiswa, analisis jejaring sosial, prediksi risiko kegagalan studi, pengembangan sistem pembelajaran cerdas, analisis teks, dan sebagainya. Hal ini menjadi penting karena jumlah data di institusi pendidikan terus meningkat seiring bertambahnya aktivitas akademik maupun non-akademik.

Ada beberapa definisi data mining yang dikenal dari berbagai sumber, diantaranya adalah:

1. Data mining merupakan sebuah usaha untuk mengumpulkan data yang kemudian akan digunakan sebagai pola untuk mengetahui hubungan antara Kumpulan data secara historis yang diharapkan dapat memberikan hasil untuk digunakan sebagai jawaban untuk memutuskan sebuah jawaban dari permasalahan yang ada dan yang akan datang(Amrullah, Dimas Agung, 2024).

2. Data mining atau penambangan kata adalah teknik yang relatif cepat dan mudah untuk menemukan pola dan atau relasi antar data, secara otomatis. Dengan menggabungkan empat disiplin ilmu computer seperti definisi diatas, pengetahuan bisa ditemukan dalam lima proses berurutan yaitu seleksi, pemrosesan, transformasi, data mining, dan interpretasi/evaluasi (Suliman, 2021).
3. Data mining adalah suatu sistem yang membantu individu atau organisasi dalam mengambil keputusan dengan menggunakan data, analisis, dan teknologi. Karena dalam Data Mining terdapat teknik dan metode untuk memenuhi kebutuhan informasi yang luas (Lubis et al., 2025).

Berdasarkan beberapa definisi tersebut, dapat disimpulkan bahwa Data mining adalah proses untuk menggali, mengolah, dan menganalisis data secara otomatis guna menemukan pola, hubungan, atau informasi penting yang dapat digunakan sebagai dasar pengambilan keputusan. Proses ini melibatkan beberapa tahapan seperti seleksi, pemrosesan, transformasi, analisis data, hingga evaluasi, dan bermanfaat bagi individu maupun organisasi untuk memahami data historis serta memprediksi kebutuhan di masa mendatang

### **2.3. Pengertian Prestasi Akademik**

Prestasi akademik merupakan salah satu tolok ukur kemajuan pendidikan yaitu dengan melihat pada hasil belajar yang dicapai oleh siswa. Untuk menciptakan sumber daya manusia yang berkualitas membutuhkan suatu pendidikan, karena pendidikan sangat penting untuk menyempurnakan perkembangan individu

dalam menguasai pengetahuan, sosial, sikap dan sebagainya (Malik & Simatupang, 2020).

Kegiatan belajar mengajar pada institusi pendidikan akan menghasilkan prestasi akademik yang ditandai dengan perubahan pengetahuan dan keterampilan peserta didik. Melalui capaian akademik, dapat terlihat bagaimana kemampuan peserta didik berkembang sepanjang proses pendidikan. Prestasi tersebut juga mencerminkan keberhasilan berbagai program pembelajaran yang diterapkan oleh lembaga pendidikan.

Prestasi akademik siswa kategori masih kurang. Tinggi rendahnya capaian prestasi akademik tersebut bergantung pada beberapa faktor, diantaranya adalah kehadiran, kurus bahasa Inggris, dan minat bahasa Inggris. Keterkaitan antara kurus dan minat dengan prestasi akademik adalah sangat penting. Oleh karena itu pentingnya siswa harus memahami dan melek digital agar mahir dan terampil dalam digitalisasi belajar (Wulandari et al., 2022).

Dalam dunia pendidikan, kegiatan belajar mengajar memiliki peran yang sangat penting karena menjadi sarana utama untuk membentuk pengetahuan, sikap, dan keterampilan siswa. Proses ini melibatkan interaksi antara pendidik, peserta didik, serta lingkungan belajar yang saling mendukung. Interaksi yang berkualitas akan berkontribusi pada peningkatan kemampuan akademik siswa secara berkelanjutan.

Hasil dari kegiatan pembelajaran tersebut tampak melalui perubahan positif dalam pemahaman konsep, kemampuan berpikir kritis, dan penguasaan keterampilan yang relevan dengan kebutuhan masa kini. Dengan demikian,

prestasi akademik tidak hanya menjadi bukti keberhasilan siswa dalam memahami materi pelajaran, tetapi juga menunjukkan kesiapan mereka untuk menghadapi tantangan di tingkat pendidikan selanjutnya maupun dalam kehidupan sehari-hari.

### **2.3.1 Faktor yang Mempengaruhi Prestasi Akademik *Internal* dan *Eksternal***

Faktor internal yang mempengaruhi hasil belajar siswa meliputi kurus, minat belajar berperan penting dalam meningkatkan usaha dan ketekunan siswa dalam menghadapi tugas akademik. Self-efficacy atau keyakinan terhadap kemampuan diri dalam menghadapi tugas, berperan besar dalam menentukan seberapa besar siswa mampu mengatasi hambatan dalam pembelajaran.

gaya belajar yang sesuai dengan preferensi individu dapat meningkatkan pemahaman siswa terhadap materi yang diajarkan. Sebaliknya, prokrastinasi akademik yang tinggi sering kali menghambat proses belajar dan mengurangi kualitas hasil belajar siswa. Regulasi diri juga sangat penting dalam membantu siswa mengelola waktu dan emosi mereka, yang berkontribusi pada pencapaian tujuan akademik.

Selain faktor internal, faktor eksternal juga mempengaruhi hasil belajar siswa, seperti dukungan orang tua, kualitas pengajaran guru, lingkungan sekolah, pengaruh teman sebaya, dan akses belajar di rumah. Dukungan orang tua yang positif dapat meningkatkan motivasi siswa dan memperbaiki kinerja akademik mereka. Kualitas pengajaran yang efektif berdampak langsung pada pemahaman siswa terhadap materi dan partisipasi mereka dalam pembelajaran. Lingkungan sekolah yang kondusif, fasilitas yang memadai, dan

hubungan sosial yang positif dengan teman sebaya turut mendukung pencapaian hasil belajar yang lebih baik. Akses belajar yang baik di rumah, termasuk fasilitas internet dan ruang belajar yang nyaman, juga mempengaruhi keberhasilan akademik siswa (Wayan, 2025).

Faktor internal adalah aspek yang berasal dari diri siswa yang memengaruhi keberhasilan belajar, seperti motivasi, kepercayaan diri, gaya belajar, serta kemampuan mengatur waktu dan kebiasaan belajar. Siswa yang memiliki motivasi tinggi dan keyakinan terhadap kemampuannya biasanya lebih bersungguh-sungguh dalam menyelesaikan tugas dan memahami materi. Sebaliknya, kebiasaan menunda serta kurangnya kemampuan mengelola strategi belajar dapat menghambat pencapaian hasil akademik secara optimal.

Di samping itu, faktor eksternal juga berperan penting dalam menentukan prestasi siswa. Lingkungan keluarga yang mendukung, metode pengajaran guru yang efektif, serta suasana sekolah yang kondusif dapat meningkatkan kualitas proses belajar. Pengaruh teman sebaya dan ketersediaan fasilitas belajar di rumah, seperti internet atau ruang belajar yang nyaman, turut memperkuat kemampuan siswa dalam memahami materi. Kombinasi faktor internal dan eksternal inilah yang pada akhirnya memengaruhi tinggi rendahnya prestasi akademik siswa.

#### **2.4. Algoritma *K-means***

Algoritma *K-Means* merupakan salah satu metode analisis *unsupervised* yang digunakan untuk mengelompokkan sejumlah data berdasarkan kemiripan karakteristiknya. Proses ini dilakukan dengan membagi data ke dalam beberapa kelompok atau *cluster* sehingga setiap data yang memiliki kesamaan ciri akan

berada pada *cluster* yang sama. Sebaliknya, data yang memiliki perbedaan karakteristik akan ditempatkan pada *cluster* yang berbeda. Dengan demikian, K-Means membantu memetakan data ke dalam kelompok yang lebih terstruktur sehingga pola atau informasi tertentu dapat terlihat lebih jelas (Haviyola et al., 2024).

Dalam penerapannya, algoritma ini bekerja dengan menentukan jumlah *cluster* terlebih dahulu, kemudian menghitung jarak setiap data terhadap pusat *cluster* (centroid). Data kemudian dialokasikan ke pusat *cluster* yang jaraknya paling dekat. Proses pembaruan pusat *cluster* ini dilakukan secara berulang hingga posisi centroid stabil dan tidak berubah lagi. Hasil akhirnya berupa pengelompokan data yang optimal sesuai kemiripan masing-masing objek.

#### **2.4.1. Clustering**

*Clustering* merupakan suatu proses dimana data dikelompokkan ke dalam berbagai kelompok yang berbeda, dengan tujuan agar objek-objek serupa tergabung dalam satu kelompok atau cluster, sedangkan objek-objek yang berbeda ditempatkan dalam kelompok yang berbeda pula. Setiap cluster berisi data yang memiliki kesamaan sebanyak mungkin. *Clustering* merupakan salah satu metode analisis data yang digunakan untuk mengelompokkan data menjadi beberapa kelompok (cluster) berdasarkan kesamaan fitur atau atribut yang dimiliki oleh setiap data (Yamin Nurzaman & Nurina Sari, 2023).

*Clustering* merupakan salah satu model dalam teknik data mining yang merupakan inti dari knowledge discovery in database. K-Means clustering merupakan metode pengelompokan data ke dalam cluster berdasarkan variabel,

dengan tujuan utama untuk meminimalkan jarak antara titik data dan clusterterkait. Teknik lain yang bisa digunakan dengan memanfaatkan variasi dari data adalah dengan membuat prediksi hasil panen pada dimasa yang akan datang. Prediksi adalah strategi yang tepat untuk mengetahui hasil pertanian dengan mempertimbangkan berbagai macam factor.

Data cluster yang diperoleh dari perhitungan clustering K-Means tersebut kemudian dilakukan perbandingan perhitungan K-Means pada aplikasi data mining untuk melihat kesamaan hasil clustering, dan setelah itu data tersebut dianalisis untuk selanjutnya digunakan sebagai dataset pada proses peramalan atau prediksi menggunakan aplikasi data mining berdasarkan nilai error terkecil (Farismana, 2024). Kemudian data tersebut dianalisis secara mendalam untuk selanjutnya digunakan sebagai dataset pada proses peramalan atau prediksi menggunakan aplikasi data mining berdasarkan nilai error terkecil agar mendapatkan hasil prediksi yang optimal.

#### 2.4.2. Rumus Umum Algoritma K-Means

Rumus utama dalam algoritma K-Means adalah perhitungan jarak Euclidean yang digunakan untuk menentukan kedekatan data dengan centroid:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2}$$

Keterangan.

$d(x_i, c_j)$  = jarak antara data ke-i dengan centroid ke-j

$x_{ik}$  = nilai data ke-i pada atribut ke-k

$c_{jk}$  = nilai data ke-i pada atribut ke-k

$n$  = jumlah atribut data

Rumus pembaruan centroid Digunakan setelah semua data masuk ke cluster untuk menghitung pusat cluster baru.

$$c_j = \frac{1}{|C_j|} \sum_{x_i \in C_j} x_i$$

Keterangan:

$c_j$  = centroid baru cluster j

$C_j$  = himpunan data pada cluster j

$|C_j|$  = jumlah data dalam cluster j

## 2.5. Algoritma *Naive Bayes*

Metode klasifikasi *Naive Bayes* merupakan salah satu algoritma dalam teknik klasifikasi yang dikenal sederhana, mudah diterapkan, serta memiliki kecepatan pemrosesan yang tinggi. Algoritma ini bekerja berdasarkan pendekatan statistik yang menggunakan peluang atau probabilitas untuk menentukan kelas suatu data. Dalam prosesnya, *Naive Bayes* menganalisis data uji dengan membandingkan pola yang terdapat pada data-data sebelumnya sehingga prediksi yang dihasilkan lebih akurat dan terukur (Rahman, 2023).

Pendekatan ini menghitung kemungkinan suatu data termasuk ke dalam kelas tertentu dengan memanfaatkan informasi dari atribut-atribut yang dimilikinya. Meskipun disebut “*naive*” karena mengasumsikan bahwa setiap

atribut bersifat saling independen, metode ini tetap terbukti efektif dalam berbagai jenis permasalahan klasifikasi. Dengan memanfaatkan probabilitas bersyarat, *Naive Bayes* dapat menentukan kelas yang paling mungkin berdasarkan data yang telah terjadi sebelumnya.

Dalam proses perhitungannya, *Naive Bayes* menggunakan Teorema Bayes untuk memperoleh nilai probabilitas posterior, yaitu peluang suatu hipotesis benar berdasarkan evidence yang diamati. Nilai probabilitas ini diperoleh dari kombinasi peluang awal suatu kelas (prior probability) dan peluang kemunculan evidence terhadap kelas tersebut (likelihood). Hasil klasifikasi ditentukan dari nilai probabilitas terbesar di antara semua kelas yang mungkin.

Keunggulan utama *Naive Bayes* terletak pada kemampuannya menghasilkan prediksi yang stabil meskipun menggunakan data yang tidak terlalu besar atau tidak sepenuhnya bersih. Algoritma ini dapat berfungsi dengan baik pada dataset yang memiliki banyak atribut berbeda karena proses perhitungannya sederhana dan tidak memerlukan pemodelan kompleks. Selain itu, hasil klasifikasi yang dihasilkan *Naive Bayes* mudah dipahami sehingga peneliti dapat mengetahui alasan sebuah data dikategorikan ke dalam kelas tertentu berdasarkan nilai probabilitas yang dihasilkan. Karakteristik ini sangat bermanfaat dalam penelitian akademik karena mempermudah sekolah ataupun guru menganalisis faktor-faktor yang memengaruhi prestasi siswa.

### **2.5.1. Rumus Utama *Naive Bayes***

Algoritma *Naive Bayes* ditemukan oleh Thomas Bayes dimana algoritma ini merupakan metode yang prosesnya menilai dari probabilitas sederhana dan

data statistik yang digunakan dalam melakukan prediksi peluang dilakukan sebelumnya. Naive Bayes adalah algoritma pembelajaran mesin berbasis probabilitas yang menggunakan Teorema Bayes untuk menghitung probabilitas posterior dari suatu kelas berdasarkan fitur-fitur yang ada. Dalam formula Teorema Bayes, probabilitas suatu kelas  $P(H|X)$  dihitung berdasarkan probabilitas awal kelas  $P(H)$ , probabilitas fitur diberikan kelas  $P(X|H)$ , dan probabilitas fitur  $P(X)$ . Berikut Rumus teorema Bayes (Aqillah Fadia Haya et al., 2022).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$  = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$  = Probabilitas hipotesis H (prior prob.)

$P(X|H)$  = Probabilitas X berdasarkan kondisi tersebut

$P(X)$  = Probabilitas dari X

Metode Naive Bayes menggunakan model statistik untuk melakukan proses klasifikasi data. Metode ini menghitung nilai probabilitas data uji berdasarkan data kasus yang sudah pernah terjadi. Berikut adalah simulasi perhitungan pada probabilitas naive bayes yang bisa dilihat pada Persamaan 1 berikut (Rahman, 2023).

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Keterangan:

$P(A|B)$  adalah peluang A jika diketahui keadaan B

$P(B|A)$  adalah peluang evidence B jika diketahui hipotesis A

$P(B)$  adalah probabilitas B tanpa melihat evidence apapun

$P(A)$  adalah peluang evidence A

## **2.6. Penerapan Naive Bayes dan K-means dalam Analisis Faktor Penentu Prestasi Akademik**

Dalam analisis data mining, terdapat beberapa teknik seperti klasifikasi, regresi, klastering, asosiasi, dan lain sebagainya. Salah satu teknik yang paling umum digunakan adalah klastering. Klastering adalah teknik untuk mengelompokkan objek-objek dalam sebuah data berdasarkan kesamaan atau kemiripan antara objek-objek tersebut. Salah satu metode klastering yang paling umum digunakan adalah metode K-Means. Metode K-Means adalah teknik untuk mengelompokkan data dengan cara membagi data kedalam beberapa kelompok atau klaster, di mana setiap data atau objek dalam kelompok tersebut memiliki kesamaan atau kemiripan tertentu. Metode ini dilakukan dengan cara menghitung jarak antara setiap data dengan pusat klaster atau centroid, dan membagi data kedalam kelompok yang memiliki pusat klaster terdekat. Penerapan metode K-Means dalam analisis data mining dapat digunakan untuk memecahkan berbagai masalah di berbagai bidang, seperti bisnis, kesehatan, pariwisata, dan

Pendidikan. Dalam pendidikan, metode K-Means dapat digunakan untuk mengelompokkan siswa berdasarkan kinerja akademiknya, sehingga dapat membantu Sekolah dalam memberikan layanan terbaik bagi siswa dan meningkatkan kualitas pendidikan (Umbu et al., 2023).

Metode Naive Bayes sangat efektif dalam pembelajaran berdasarkan data pelatihan, karena mengandalkan probabilitas bersyarat. Namun, metode Naive Bayes juga memiliki kekurangan, yakni hanya dapat digunakan dalam konteks klasifikasi dengan pembelajaran terawasi (supervised learning) dan hanya cocok untuk data yang bersifat kategorikal. Dari segi statistik terdapat suatu teknik yang disebut Naive Bayes yang bisa digunakan untuk memproyeksikan probabilitas kemunculan suatu kategori. Penerapan Naive Bayes sering kali dilakukan untuk mengelompokkan data yang telah dikumpulkan dari proses pelatihan yang intensif. Selain itu juga dapat untuk klasifikasi peluang yang efektif yang menurunkan kemungkinan jumlah dengan menyatukan frekuensi dan nilai numerik dari gabungan data yang sudah ada (Sholihin Fauzan et al., 2024).

Analisis faktor penentu prestasi akademik dapat dilakukan dengan memanfaatkan algoritma *data mining* seperti Naive Bayes dan K-Means. Kedua metode ini memberikan pendekatan yang berbeda sehingga mampu memberikan gambaran yang lebih komprehensif terkait pola dan faktor yang memengaruhi keberhasilan belajar siswa. Penggunaan kedua algoritma ini sangat relevan untuk mengolah data siswa yang beragam, mulai dari nilai akademik, kehadiran, kursus, hingga minat.

Naive Bayes merupakan metode klasifikasi berbasis probabilitas yang digunakan untuk memprediksi kategori atau kelas tertentu, seperti apakah siswa memiliki prestasi akademik tinggi, sedang, atau rendah. Metode ini menggunakan prinsip Teorema Bayes, yaitu menghitung probabilitas suatu kategori berdasarkan data sebelumnya. Dengan menganalisis atribut seperti kehadiran, nilai tugas, latar belakang keluarga, hingga motivasi belajar, Naive Bayes dapat memberikan prediksi mengenai prestasi siswa secara cepat dan akurat. Penggunaan metode ini memungkinkan peneliti mengetahui faktor mana yang lebih berpeluang menyebabkan perbedaan prestasi akademik antar siswa.

Sementara itu, K-Means berfungsi untuk mengelompokkan siswa ke dalam beberapa *cluster* berdasarkan kemiripan karakteristik tertentu, misalnya kelompok siswa berprestasi tinggi, sedang, dan rendah. Melalui proses pengelompokan ini, pola-pola penting seperti faktor dukungan keluarga, motivasi belajar, atau kebiasaan belajar siswa dapat terlihat dengan lebih jelas. Kombinasi kedua metode ini memberikan hasil analisis yang lebih komprehensif. K-Means membantu mengidentifikasi kelompok siswa berdasarkan karakteristik mereka, sedangkan Naive Bayes memprediksi prestasi akademik serta menentukan faktor dominan yang mempengaruhinya. Dengan demikian, kedua algoritma ini dapat menjadi alat yang efektif untuk membantu sekolah dalam perencanaan pembelajaran dan peningkatan prestasi siswa.

## 2.7. Pengertian Siswa

siswa atau murid adalah salah satu komponen dalam pengajaran, disamping faktor guru, tujuan dan metode pengajaran Siswa merupakan salah satu elemen penting dalam proses pembelajaran karena keberadaan mereka menentukan arah, tujuan, serta keberhasilan proses pengajaran. Hal ini dikarenakan seluruh aktivitas pendidikan pada dasarnya dirancang untuk membantu siswa mencapai perkembangan pengetahuan dan keterampilan yang optimal. Dengan demikian, kedudukan siswa menjadi faktor utama yang harus diperhatikan dalam merancang dan melaksanakan suatu kegiatan pengajaran (Alramadhani & Febrianto, 2023).

Setiap siswa dipandang sebagai individu yang unik, memiliki potensi berbeda, serta mengalami perkembangan sesuai dengan kemampuan masing-masing. Perbedaan karakter, minat, dan gaya belajar membuat setiap peserta didik membutuhkan pendekatan yang sesuai agar potensi yang dimiliki dapat berkembang secara maksimal. Oleh sebab itu, keberagaman yang ada pada diri siswa harus menjadi perhatian penting dalam penyusunan strategi dan pelaksanaan proses pembelajaran.

Peserta didik dapat dipahami sebagai individu yang terlibat dalam proses pendidikan baik secara formal maupun sepanjang hidupnya. Dalam konteks sekolah, siswa adalah anak yang mengikuti kegiatan belajar sebagai upaya untuk memperoleh ilmu pengetahuan, keterampilan, pengalaman, dan pembentukan karakter. Melalui pendidikan di sekolah, diharapkan siswa dapat tumbuh menjadi

pribadi yang berpengetahuan, terampil, berakhlak, serta mampu berdiri sendiri dalam menjalani kehidupan.

Siswa adalah individu yang sedang berada dalam tahap perkembangan, baik secara fisik, kognitif, sosial maupun emosional, dan mengikuti proses pendidikan formal pada jenjang tertentu. Dalam konteks pendidikan, siswa dipandang sebagai subjek yang aktif dalam kegiatan pembelajaran, bukan hanya penerima pengetahuan, tetapi juga sebagai pihak yang berinteraksi dengan guru, lingkungan belajar, serta berbagai sumber pembelajaran lainnya.

Dalam konteks penelitian ini, siswa dipandang sebagai individu yang memiliki potensi akademik berbeda khususnya pada mata pelajaran Bahasa Inggris. Oleh karena itu, penting untuk memahami karakteristik siswa secara menyeluruh, meliputi kemampuan kognitif, motivasi belajar, tingkat kedisiplinan, dan lingkungan sosial. Pemahaman tersebut diperlukan untuk mengidentifikasi faktor-faktor yang memengaruhi prestasi akademik dan untuk mengelompokkan siswa berdasarkan pola belajar mereka menggunakan algoritma K-Means, serta memprediksi tingkat prestasi mereka menggunakan algoritma Naive Bayes. Dengan demikian, siswa tidak hanya diposisikan sebagai objek analisis, tetapi juga sebagai pusat perhatian dalam upaya meningkatkan mutu pembelajaran dan prestasi akademik di sekolah.

## **2.8. Prestasi Akademik Bahasa Inggris**

Prestasi akademik Bahasa Inggris merupakan gambaran kemampuan siswa dalam memahami, menguasai, serta mengaplikasikan berbagai keterampilan berbahasa meliputi *listening*, *speaking*, *reading*, dan *writing*. Mata pelajaran

Bahasa Inggris memiliki karakteristik yang berbeda dibanding mata pelajaran lain karena menekankan pada kemampuan komunikasi dan pemahaman konteks. Oleh sebab itu, prestasi akademik dalam Bahasa Inggris tidak hanya mencerminkan kemampuan kognitif, tetapi juga kemampuan linguistik, motivasi, dan penguasaan kosakata siswa.

Keberhasilan akademik Bahasa Inggris dipengaruhi oleh kualitas proses belajar, kemampuan pemahaman teks, serta kesiapan siswa dalam mengikuti pembelajaran berbasis keterampilan. Prestasi dalam mata pelajaran ini juga berkaitan erat dengan minat belajar, lingkungan bahasa, serta metode pembelajaran yang digunakan guru. Dengan memahami faktor-faktor tersebut, sekolah dapat mengembangkan strategi pembelajaran yang lebih efektif untuk meningkatkan kemampuan berbahasa siswa.

Dalam konteks pembelajaran Bahasa Inggris, kebiasaan menggunakan bahasa ini dalam kehidupan sehari-hari terbukti dapat memperkuat penguasaan kosa kata, struktur kalimat, serta meningkatkan rasa percaya diri dalam berkomunikasi. Selain itu, rencana atau motivasi untuk bekerja di luar negeri juga diyakini dapat meningkatkan semangat belajar mahasiswa, khususnya dalam penguasaan Bahasa Inggris yang merupakan salah satu prasyarat utama untuk bersaing secara global (Maramis, 2025)

## **2.9. Alat Bantu Pemrograman untuk Implementasi**

Dalam penelitian ini, berbagai perangkat lunak digunakan untuk menganalisis dan membandingkan metode K-means dan Naive Bayes. Berikut adalah beberapa perangkat lunak yang digunakan.

### 2.9.1. Microsoft Excel

Microsoft Excel merupakan sebuah aplikasi spreadsheet yang memiliki kemampuan komprehensif untuk menampilkan dan mengolah berbagai jenis data. Aplikasi ini banyak dimanfaatkan untuk melakukan perhitungan numerik, menyusun tabel, membuat grafik, serta menghasilkan laporan yang sistematis, terutama ketika berhubungan dengan data yang melibatkan angka dalam jumlah besar. Fleksibilitas dan kelengkapan fitur yang dimiliki menjadikan Microsoft Excel sebagai alat yang sangat efektif dalam mendukung aktivitas pengolahan data (Korib et al., 2025).



**Gambar 2.1 Logo Excel**

**Sumber: Microsoft Excel**

Aplikasi ini tidak hanya digunakan dalam dunia bisnis, tetapi juga sangat bermanfaat di bidang pendidikan, penelitian, dan aktivitas administratif lainnya. Dengan fitur yang lengkap dan mudah digunakan, Excel menjadi perangkat penting dalam mendukung berbagai kebutuhan yang berkaitan dengan analisis dan manajemen data.

## 2.9.2 RapidMiner

RapidMiner adalah sebuah aplikasi yang digunakan untuk mengolah data menggunakan berbagai teknik dan metode dalam data mining, sehingga data mentah dapat diubah menjadi informasi yang lebih bermakna dan bermanfaat. Sebagai perangkat lunak *open-source*, RapidMiner menyediakan beragam fitur yang membantu pengguna dalam melakukan proses analisis data secara menyeluruh dan efisien (Fatah et al., 2025).

RapidMiner bekerja dengan mengekstraksi pola-pola penting dari kumpulan data (*dataset*) dan mengombinasikannya dengan pendekatan statistika, kecerdasan buatan, serta pemanfaatan basis data.



**Gambar 2.2 Logo RapidMiner**

**Sumber: RapidMiner**

Melalui proses ini, RapidMiner mampu menghasilkan model prediksi maupun klasifikasi yang akurat berdasarkan pola yang ditemukan dalam data. Perangkat ini juga mendukung berbagai algoritma *Data mining*, sehingga pengguna dapat memilih metode yang paling sesuai dengan kebutuhan analisis. Aplikasi ini sering dimanfaatkan untuk menguji, memodelkan, dan

mengevaluasi data penelitian sehingga peneliti dapat menemukan insight yang tepat dan relevan.

Dalam konteks penelitian pendidikan, RapidMiner mempermudah peneliti dalam melakukan eksplorasi data siswa dan menguji berbagai skenario model tanpa harus memiliki kemampuan pemrograman lanjutan. Data mentah dapat diolah menjadi bentuk yang lebih siap analisis melalui fitur seperti normalisasi, penghapusan duplikasi, penanganan missing value, dan pengaturan tipe atribut. Setelah itu, peneliti dapat langsung menerapkan algoritma tertentu dan mendapatkan keluaran berupa hasil pengelompokan atau klasifikasi yang mudah dipahami. RapidMiner juga menyediakan tampilan grafik dan tabel yang membantu dalam interpretasi hasil, sehingga peneliti atau pihak sekolah dapat memahami pola siswa secara lebih komprehensif. Dengan kemampuannya tersebut, RapidMiner menjadi alat yang efektif untuk mendukung proses penelitian yang membutuhkan akurasi, kecepatan, dan kejelasan visual dalam analisis data.

#### **2.10. Penelitian Terdahulu**

Pada penelitian ini, penulis menuliskan beberapa referensi yang digunakan dalam penelitian ini sebagai pendukung dan landasan dalam pengerjaan proposal ini, yaitu sebagai berikut:

Tabel 2.1 Penelitian Terdahulu

Peneliti	Tahun	Judul	Metode	Hasil
Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Metode Data Mining dengan Algoritma Naive Bayes	2023	Prediksi kelulusan tepat waktu mahasiswa menggunakan Naive Bayes.	Naive Bayes.	Akurasi $\sim$ 0.88; precision & recall kelas “on-time” dan “late” dilaporkan (Junaidi et al., 2023).
Wahyudi, Rahmawati & Supriyanto	2024	<i>Implementasi RapidMiner untuk Menentukan Siswa Unggulan Menggunakan Metode K-Means</i>	K-Means (RapidMiner)	Dataset 31 siswa, 24 atribut; terbentuk 2 cluster, membantu pemilihan siswa unggulan. (Wahyudi et al., 2024)
Firmansyah, Fatchan &	2024	<i>Analisa Klasifikasi</i>	Naive Bayes (RapidMiner)	Akurasi 93,33 %, precision 100 %,

Rifa'i		<i>Tingkat Kelulusan Mahasiswa Metode Algoritma Naïve Bayes Menggunakan RapidMiner</i>		recall 66,67 %.(Firmansyah et al., 2024)
Musa, Mohammed Ahmadu, Asabe Sandra Williams	2024	<i>Comparative Analysis of K-Means and Naïve Bayes Algorithms for Predicting Students' Academic Performance</i>	Perbandingan K-Means vs Naïve Bayes	Naïve Bayes akurasi 92,6 %, K-Means akurasi 38,9 %. (Musa et al., 2024)
Penelitian ini	2025	<i>Penerapan algoritma K-Means dan Naive Bayes untuk menganalisis</i>	Kombinasi K-Means + Naïve Bayes	Masih dalam tahap penelitian

		<i>faktor penentu prestasi akademik bahasa Inggris siswa SMA swasta di Kabupaten Labuhanbatu</i>		
--	--	--	--	--