

BAB II

LANDASAN TEORI

2.1 Data

Data merupakan sekumpulan fakta, angka, simbol, teks, atau informasi lain yang menggambarkan suatu keadaan tertentu. Data dapat berasal dari peristiwa, objek, atau aktivitas yang terjadi dalam kehidupan sehari-hari. Pada dasarnya, data adalah bahan mentah yang belum memiliki makna ketika masih berdiri sendiri. Namun, setelah melalui proses pengolahan dan analisis, data dapat memberikan informasi dan pengetahuan yang berguna untuk memahami kondisi serta membantu pengambilan keputusan yang lebih tepat. Karena itu, data memiliki peran penting dalam berbagai bidang, seperti pendidikan, bisnis, kesehatan, serta pemerintahan. Tanpa data, keputusan yang diambil cenderung berdasarkan perkiraan atau asumsi semata, sehingga berpotensi menimbulkan kesalahan dalam proses perencanaan dan evaluasi.

Secara umum, data terbagi menjadi dua jenis, yaitu data kualitatif dan data kuantitatif. Data kualitatif adalah data yang berbentuk deskripsi atau kata-kata, biasanya digunakan untuk menggambarkan kualitas atau karakteristik suatu objek, seperti warna, jenis, atau kategori tertentu. Data ini tidak berbentuk angka, tetapi tetap memiliki nilai informatif. Sementara itu, data kuantitatif adalah data yang berbentuk angka dan dapat diukur secara matematis. Contoh data kuantitatif adalah jumlah siswa, nilai ujian, berat badan, atau suhu. Data kuantitatif juga terbagi menjadi dua, yaitu data diskrit yang dapat dihitung jumlahnya, dan data kontinu yang dapat diukur dalam satuan tertentu. Kedua jenis data ini sangat penting dalam penelitian karena membantu peneliti memahami fenomena

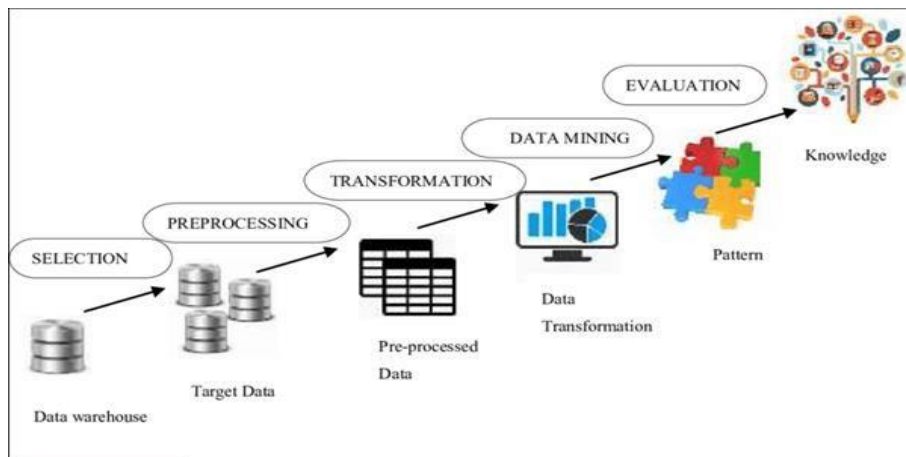
secara lebih terstruktur (Sutriyanti; Muspawi, 2024).

Data dapat diperoleh dari dua sumber utama, yaitu data primer dan data sekunder. Data primer diperoleh langsung dari objek penelitian, misalnya melalui wawancara, observasi, atau kuesioner. Data ini dianggap lebih akurat karena dikumpulkan secara langsung sesuai kebutuhan penelitian. Sementara itu, data sekunder berasal dari pihak lain, seperti jurnal, buku, laporan instansi, atau arsip yang sudah ada sebelumnya. Data sekunder memudahkan peneliti karena dapat digunakan sebagai referensi tambahan tanpa harus melakukan pengumpulan data langsung. Kedua sumber data tersebut memiliki kelebihan masing-masing dan dapat digunakan sesuai tujuan penelitian.

Agar data dapat digunakan secara efektif, diperlukan proses pengolahan yang meliputi pengumpulan, pembersihan, pengolahan, dan analisis. Pengumpulan data adalah tahap awal untuk memperoleh bahan penelitian. Setelah itu, data dibersihkan dari kesalahan, duplikasi, atau informasi yang tidak relevan. Selanjutnya, data diolah melalui perhitungan, pengelompokan, atau pengurutan. Tahap terakhir adalah analisis data untuk menemukan pola, hubungan, atau makna tertentu yang dapat digunakan sebagai dasar pengambilan keputusan. Dalam praktiknya, data juga perlu disajikan dalam bentuk tabel, grafik, atau diagram agar lebih mudah dipahami. Melalui pengolahan yang baik, data dapat memberikan gambaran yang jelas dan akurat sehingga bermanfaat untuk perencanaan, evaluasi, dan pengembangan di berbagai bidang.

2.2. Knowledge Discovery in Database

Knowledge Discovery in Database (KDD) merupakan suatu proses komprehensif yang bertujuan untuk menemukan pola, informasi, atau pengetahuan baru yang bernilai dari kumpulan data dalam jumlah besar. Proses ini tidak hanya sekadar mengekstraksi data, tetapi juga mencakup rangkaian langkah yang sistematis dan terstruktur, dimulai dari pemilihan data, pembersihan data, hingga interpretasi hasil analisis sehingga informasi yang dihasilkan dapat digunakan sebagai dasar pengambilan keputusan yang lebih akurat dan berbasis bukti.



Gambar 2.1. Knowledge Discovery in Database (KDD)

Proses *Knowledge Discovery in Database* (KDD) terdiri atas serangkaian tahapan yang sistematis, dimulai dari pengumpulan data, pembersihan, transformasi, sampai analisis dan interpretasi hasil. Di antaranya pengumpulan data merupakan langkah awal yang sangat krusial, karena kualitas dan kelengkapan data yang diperoleh akan menentukan validitas serta efektivitas hasil analisis yang dilakukan dalam penelitian maupun penerapan praktisnya di berbagai bidang (Khalish et al., 2025).

Dalam konteks pendidikan, KDD menjadi alat yang sangat penting untuk memahami berbagai aspek perilaku belajar siswa. Melalui analisis data seperti kehadiran, aktivitas belajar, nilai evaluasi, atau penggunaan media pembelajaran, KDD dapat membantu memprediksi kinerja akademik siswa, mengidentifikasi potensi permasalahan belajar lebih dini, serta memberikan dasar bagi pendidik dalam merancang strategi pembelajaran yang lebih efektif, adaptif, dan sesuai dengan kebutuhan masing-masing siswa.

2.2.1 Data Mining

Data mining adalah proses menemukan pola atau informasi menarik dalam data yang dipilih menggunakan teknik atau metode (Situmorang & Sirait, 2020). Dengan kata lain, *Data Mining* adalah proses pengolahan data dalam jumlah besar untuk menemukan pola, hubungan, serta informasi tersembunyi yang berguna bagi pengambilan keputusan. Proses ini melibatkan penerapan teknik statistik, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi pengetahuan dari data yang tersimpan dalam basis data atau sistem informasi.

Dalam bidang pendidikan, *Data Mining* digunakan untuk menganalisis data akademik seperti nilai siswa, kehadiran, dan aktivitas belajar, sehingga dapat membantu pihak sekolah dalam menentukan strategi peningkatan kualitas pembelajaran. Penerapan *Data Mining* juga memungkinkan pendidik mengenali pola perilaku belajar siswa yang berdampak pada hasil belajar. Melalui analisis tersebut, institusi pendidikan dapat melakukan evaluasi dan perbaikan terhadap metode pembelajaran yang digunakan.

Selain itu, algoritma yang sering digunakan dalam proses *Data Mining* adalah C4.5, yang termasuk dalam kategori algoritma decision tree. Algoritma ini berfungsi untuk mengklasifikasikan data berdasarkan atribut tertentu guna

menghasilkan keputusan yang logis dan terstruktur. Implementasi algoritma C4.5 terbukti efektif dalam membantu proses analisis dan pengambilan keputusan karena dapat mengidentifikasi pola dan faktor yang paling berpengaruh terhadap hasil suatu data.

2.3 Algoritma Decision Tree C4.5

Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang berfungsi untuk menghasilkan model klasifikasi dalam bentuk pohon keputusan (decision tree) (Huday & Zaehol Fatah, 2025). Algoritma C4.5 juga dikenal sebagai salah satu algoritma pembelajaran mesin yang efektif untuk membangun model keputusan berbasis pohon. Algoritma ini mampu menangani data kategorikal dan numerik, serta memiliki kemampuan untuk melakukan pemilihan atribut yang relevan untuk pengambilan keputusan (Suriani, 2023).

Proses utama dalam algoritma ini adalah melibatkan pemilihan atribut terbaik sebagai akar atau simpul berdasarkan nilai Gain Ratio, yaitu perbandingan antara information gain dan split information. Dengan pendekatan ini, algoritma C4.5 mampu mengatasi kelemahan ID3 dalam menghadapi atribut dengan banyak nilai (multi-valued attributes) dan juga dapat menangani data numerik maupun kategorikal.

Dalam konteks penelitian pendidikan, khususnya dalam analisis pola belajar siswa, algoritma C4.5 sangat bermanfaat karena mampu mengidentifikasi variabel-variabel yang memiliki pengaruh signifikan terhadap hasil belajar. Pohon keputusan yang dihasilkan tidak hanya memberikan hasil klasifikasi, tetapi juga menampilkan struktur hubungan antar atribut dalam bentuk visual yang intuitif dan mudah dipahami oleh guru maupun peneliti. Dengan demikian, pendidik

dapat memanfaatkan informasi tersebut untuk memahami faktor-faktor yang penting dalam proses belajar siswa, memperkirakan potensi keberhasilan atau kesulitan siswa, serta merancang strategi pembelajaran yang lebih efektif, terarah, dan sesuai dengan karakteristik siswa dalam upaya meningkatkan hasil belajar Bahasa Inggris.

Secara umum, tahapan algoritma C4.5 dimulai dari menentukan atribut terbaik menggunakan perhitungan entropy dan gain ratio, kemudian membentuk cabang untuk setiap nilai atribut, serta melakukan proses rekursif hingga semua data terklasifikasi dengan baik. Hasil akhirnya berupa struktur pohon yang terdiri dari simpul (nodes), cabang (branches), dan daun (leaves) yang mewakili keputusan akhir.

Dalam algoritma Decision Tree C4.5, salah satu konsep utama yang digunakan dalam proses pemilihan atribut adalah perhitungan entropi (*entropy*) (Justam et al., 2024). Entropi berfungsi untuk mengukur tingkat ketidakpastian atau ketidakteraturan dalam suatu dataset. Semakin tinggi nilai entropi, semakin besar ketidakpastian atau keragaman kelas dalam data tersebut, sedangkan nilai entropi yang rendah menunjukkan bahwa data lebih homogen atau lebih pasti dalam satu kelas. Entropi untuk suatu dataset S dihitung menggunakan rumus

$$\text{Entropi}(S) = - \sum_{i=1}^n p_i \log_2(p_i)$$

Di mana k merupakan jumlah kelas pada variabel target dan p_i menunjukkan proporsi data dalam kelas ke- i . Perhitungan ini menjadi dasar untuk menentukan tingkat keacakan dalam sebuah kumpulan data sebelum dilakukan pemisahan atau *splitting*.

Setelah nilai entropi dihitung, langkah berikutnya adalah menentukan Gain Informasi (*Information Gain*), yaitu ukuran yang digunakan untuk mengetahui seberapa besar penurunan ketidakpastian setelah dataset dibagi berdasarkan nilai suatu atribut. Information Gain menjadi indikator penting dalam menentukan apakah suatu atribut layak digunakan sebagai pemisah dalam membangun pohon keputusan. Rumus Information Gain adalah:

$$\text{Gain}(S, A) = \text{Entropi}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \cdot \text{Entropi}(S_v)$$

Di mana $\text{Entropi}(S)$ adalah entropi dataset sebelum pembagian, $\text{Values}(A)$ adalah himpunan nilai yang mungkin dari atribut A , $|S_v|$ adalah jumlah elemen dalam subset data saat atribut A memiliki nilai v , dan $\text{Entropi}(S_v)$ merupakan entropi dari subset tersebut. Semakin besar nilai Information Gain, semakin baik atribut tersebut dalam mengurangi ketidakpastian.

Namun, Information Gain memiliki kelemahan, yaitu cenderung bias terhadap atribut yang memiliki banyak nilai kategori. Untuk mengatasi masalah ini, C4.5 menggunakan Gain Ratio, yaitu ukuran yang mengombinasikan Information Gain dan Split Information agar proses pemilihan atribut menjadi lebih objektif. Gain Ratio dihitung dengan rumus:

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Entropi}(A)}$$

dengan

$$\text{Entropi}(A) = - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right)$$

Rumus ini menghitung seberapa besar informasi yang diperoleh dengan mempertimbangkan seberapa besar distribusi nilai atribut tersebut. Atribut yang memiliki nilai Gain Ratio tertinggi akan dipilih sebagai pemisah dalam pohon keputusan, sehingga menghasilkan model yang lebih akurat dan tidak bias.

Dalam praktiknya, algoritma C4.5 digunakan untuk membangun model klasifikasi dengan tahapan utama: memilih atribut terbaik sebagai akar atau simpul, melakukan pembagian (split) data ke cabang-cabang, dan menghentikan pembagian saat simpul sudah homogen atau tidak memungkinkan untuk membagi lebih lanjut.

Karena menghasilkan struktur pohon yang mudah diinterpretasikan (simpul, cabang, daun) dan memungkinkan penggunaan atribut numerik maupun kategorikal, algoritma C4.5 sangat cocok untuk aplikasi pengambilan keputusan berbasis data.

2.3.1 Langkah – Langkah Algoritma Decision Tree C4.5

Algoritma Decision Tree C4.5 membangun pohon keputusan melalui serangkaian langkah sistematis yang bertujuan memilih atribut terbaik untuk membagi data hingga menghasilkan model klasifikasi yang optimal. Proses ini diawali dengan menentukan atribut target yang akan diprediksi, misalnya kategori hasil belajar siswa. Setelah atribut target ditetapkan, langkah selanjutnya adalah menghitung nilai entropi (entropy) dari keseluruhan dataset. Entropi digunakan untuk mengukur tingkat ketidakpastian atau keragaman kelas dalam data sebelum dilakukan pembagian. Semakin tinggi nilai entropi, semakin besar ketidakpastian data tersebut.

Selanjutnya, algoritma melakukan perhitungan Information Gain untuk setiap atribut. Information Gain mengukur seberapa besar pengurangan

ketidakpastian yang terjadi setelah data dibagi berdasarkan atribut tertentu. Namun demikian, karena Information Gain cenderung bias terhadap atribut dengan banyak nilai kategori, algoritma C4.5 memperbaikinya melalui perhitungan Gain Ratio, yaitu rasio antara Information Gain dan nilai Split Information. Gain Ratio membantu memilih atribut yang paling efisien dalam memisahkan data tanpa dipengaruhi banyaknya kategori dalam atribut tersebut.

Setelah nilai Gain Ratio dihitung, atribut dengan Gain Ratio tertinggi dipilih sebagai node akar (root node) atau simpul utama dalam pohon keputusan. Proses ini dilanjutkan dengan membagi dataset ke dalam subset-subset baru berdasarkan nilai atribut tersebut. Kemudian, setiap subset diuji kembali untuk menentukan apakah perlu dilakukan pembagian lebih lanjut. Jika subset masih memiliki keragaman kelas ($\text{entropi} > 0$), maka algoritma mengulangi proses perhitungan entropi, Information Gain, dan Gain Ratio untuk subset tersebut guna menentukan atribut pemisah berikutnya. Sebaliknya, jika subset telah homogen atau tidak ada atribut lain yang dapat digunakan untuk pemisahan, maka subset tersebut menjadi leaf node yang merepresentasikan hasil klasifikasi.

Langkah berikutnya adalah menangani nilai atribut yang hilang (missing values). C4.5 mengakomodasi data yang tidak lengkap dengan memberikan bobot probabilistik terhadap nilai yang hilang, sehingga model tetap dapat dibangun secara optimal tanpa harus menghapus data yang tidak lengkap. Selain itu, C4.5 juga dapat menangani atribut kontinu dengan membagi nilai numerik tersebut ke dalam interval tertentu melalui proses pencarian threshold terbaik.

Tahapan terakhir dalam algoritma C4.5 adalah proses pruning atau pemangkasan pohon keputusan untuk mengurangi kompleksitas model dan

mencegah overfitting. Pruning dilakukan dengan menghilangkan cabang-cabang yang tidak memberikan kontribusi signifikan terhadap akurasi prediksi. Hasil dari seluruh proses ini adalah sebuah pohon keputusan yang ringkas, akurat, dan mudah dipahami, sehingga sangat efektif digunakan dalam pengambilan keputusan berbasis data di berbagai bidang termasuk pendidikan, khususnya untuk menganalisis pola belajar dan memprediksi hasil belajar siswa.

2.4 Alat Bantu Aplikasi

Aplikasi RapidMiner dan Microsoft Excel merupakan dua perangkat lunak yang banyak digunakan dalam proses pengolahan data, analisis statistik, serta penerapan teknik *Data Mining*. Kedua aplikasi ini berperan penting dalam mendukung kegiatan penelitian, khususnya pada bidang pendidikan yang memanfaatkan data siswa untuk dianalisis secara sistematis dan berbasis bukti.

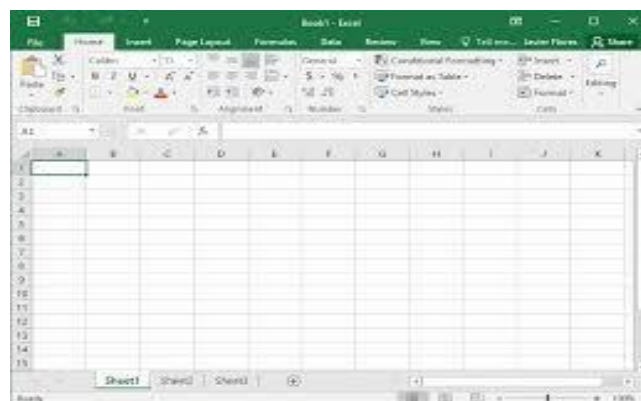
2.4.1 RapidMiner

RapidMiner adalah sebuah aplikasi data science yang menggunakan tampilan graphical user interface (GUI), sehingga pengguna dapat melakukan proses *Data Mining*, machine learning, dan analisis prediktif tanpa perlu menguasai bahasa pemrograman (Juanda Saputra & Izman Herdiansyah, 2022). Dengan tampilan yang intuitif dan cara kerja berbasis drag-and-drop, RapidMiner memudahkan pengguna menjalankan berbagai tahapan analisis data secara visual.

format data seperti CSV, Excel, dan SQL. Aplikasi ini juga memungkinkan pengguna menyimpan, mengatur, dan mengembangkan proyek analisis secara sistematis. Dengan berbagai keunggulan tersebut, RapidMiner menjadi salah satu alat yang efektif dan mudah digunakan bagi peneliti dalam melakukan analisis data, termasuk dalam penelitian pendidikan untuk memahami perilaku belajar siswa dan membantu proses pengambilan keputusan yang lebih tepat.

2.4.2 Microsoft Excel

Microsoft Excel atau biasa dikenal dengan Excel merupakan salah satu perangkat lunak (software) yang berada dalam Microsoft Office. Excel dibuat oleh Microsoft Corporation dan berjalan atau bisa digunakan pada sistem operasi Windows maupun Mac OS. Program aplikasi ini memiliki banyak fitur dan fungsi yang digunakan untuk mengolah angka. Fitur Fungsi dan Formula atau yang lebih dikenal dengan rumus Excel menjadikannya terkenal dan banyak digunakan dalam berbagai bidang dan persoalan seperti membuat, mengedit, mengurutkan, menganalisa, serta meringkas data (Odja et al., 2021). Selain itu, Microsoft Excel juga berfungsi untuk melakukan perhitungan aritmatika dan statistika sehingga dapat membantu menyelesaikan permasalahan-permasalahan yang berbaur logika dan matematika.



Gambar 2.3. Tampilan Awal Microsoft Excel

Microsoft Excel adalah aplikasi pengolah data yang sangat populer dan sering digunakan untuk berbagai kebutuhan, terutama dalam mengolah data awal (data preprocessing), melakukan analisis statistik sederhana, serta membuat visualisasi dalam bentuk tabel maupun grafik.

Excel menyediakan banyak fitur yang membantu pengguna, seperti filtering untuk menyaring data, sorting untuk mengurutkan data, pivot table untuk meringkas informasi, data cleaning untuk membersihkan data yang tidak sesuai, dan berbagai rumus matematis untuk melakukan perhitungan otomatis. Fitur-fitur tersebut menjadikan Excel sangat berguna dalam menyiapkan dataset sebelum dianalisis lebih lanjut menggunakan aplikasi *Data Mining* seperti RapidMiner.

Dalam penelitian pendidikan, Excel sering digunakan untuk mengecek kelengkapan data siswa, menghapus data yang ganda, memperbaiki bagian yang kosong atau tidak valid, serta mengatur format data agar cocok dengan kebutuhan analisis berikutnya. Selain itu, Excel memudahkan peneliti dalam melakukan perhitungan dasar, membuat rekap nilai siswa, serta menampilkan grafik distribusi data yang dapat memberikan gambaran awal mengenai pola atau kondisi data.

Dengan fungsi yang sederhana namun sangat fleksibel ini, Excel menjadi alat penting pada tahap awal penelitian berbasis data karena dapat mempercepat proses persiapan dan pemahaman data sebelum masuk ke tahap analisis yang lebih kompleks.

2.5 Pola Belajar

Dalam proses pembelajaran, pola belajar siswa berperan penting dalam menentukan aktivitas belajar siswa. Ada siswa yang mempunyai pola belajar yang

baik, tetapi ada juga siswa yang mempunyai pola belajar yang kurang baik. Pola belajar yang digunakan siswa juga berbeda-beda. Ada siswa yang berpikir secara cepat, dan ada pula cara berpikir secara lambat. Perbedaan kemampuan belajar merupakan ciri dari pola belajar (Andalangi et al., 2022).

Pola belajar merupakan cara, strategi, dan kebiasaan yang digunakan siswa untuk memahami, mengolah, dan menerapkan materi yang mencakup keterampilan listening, speaking, reading, dan writing. Dalam proses mempelajari Bahasa Inggris, siswa dituntut untuk tidak hanya menghafal kosakata dan struktur tata bahasa (*grammar*), tetapi juga mampu mempraktikkannya dalam konteks komunikasi nyata.

Pola belajar Bahasa Inggris dapat bervariasi sesuai dengan kemampuan, karakteristik, serta kebutuhan siswa. Sebagian siswa lebih efektif belajar melalui pendekatan visual seperti penggunaan gambar, video, flashcard, atau mind mapping. Siswa lainnya cenderung berhasil dengan pola belajar auditori melalui kegiatan mendengarkan percakapan, lagu, podcast, serta latihan pelafalan (*pronunciation*). Adapun siswa dengan pola belajar kinestetik akan lebih mudah memahami Bahasa Inggris melalui permainan peran simulasi percakapan, serta kegiatan praktik berbasis proyek.

Faktor keberhasilan pola belajar Bahasa Inggris tidak hanya ditentukan oleh preferensi gaya belajar, tetapi juga oleh frekuensi latihan, motivasi intrinsik, lingkungan bahasa dan kepercayaan diri dalam berkomunikasi. Semakin sering siswa berlatih melalui penggunaan media digital seperti *online dictionary*, *learning apps*, atau *e-learning platform*, semakin cepat kemampuan bahasa mereka berkembang. Selain itu, dukungan guru melalui pemilihan metode

pembelajaran yang interaktif serta penggunaan media yang menarik akan sangat membantu dalam peningkatan pemahaman konsep.

Pola belajar yang terencana, terstruktur, dan relevan dengan kebutuhan siswa akan menghasilkan peningkatan yang signifikan terhadap penguasaan kosakata, kemampuan memahami konteks, serta keterampilan berbicara dan menulis secara komunikatif. Dengan demikian, pola belajar yang mandiri, terbimbing, kelompok serta diskusi menjadi kunci dalam keberhasilan proses pembelajaran Bahasa Inggris di berbagai jenjang pendidikan.

2.6 Hasil Belajar

Hasil belajar merupakan capaian akhir yang diperoleh peserta didik setelah melalui rangkaian proses pembelajaran yang berlangsung secara sistematis, terencana, dan terarah, yang hasilnya dapat diukur melalui berbagai bentuk evaluasi seperti tes akademik, penilaian kinerja, observasi sikap, maupun penilaian keterampilan.

Hasil belajar pada hakekatnya adalah perubahan perilaku yang terjadi pada siswa setelah belajar dalam segi kognitif, afektif, dan psikomotor (Kamil & Safrul, 2023). Hasil belajar tidak hanya menunjukkan sejauh mana siswa mampu menguasai materi yang diberikan, tetapi juga mencerminkan perkembangan kemampuan intelektual, emosional, dan psikomotoriknya. Dengan demikian, hasil belajar menjadi salah satu indikator utama dalam menilai efektivitas proses pendidikan, kualitas pembelajaran, dan pencapaian tujuan kurikulum.

Faktor-faktor yang memengaruhi hasil belajar siswa sangat beragam, meliputi faktor internal seperti minat, motivasi, kesiapan belajar, kemampuan intelektual,

dan kondisi psikologis, serta faktor eksternal yang mencakup metode pembelajaran yang digunakan guru, ketersediaan media dan fasilitas belajar, dukungan keluarga, lingkungan sosial, serta pemanfaatan teknologi dalam proses pembelajaran.

Dalam konteks pembelajaran Bahasa Inggris, hasil belajar menggambarkan tingkat kemampuan kognitif siswa dalam memahami konsep bahasa, menguasai kosakata, menerapkan kaidah tata bahasa, serta kemampuan menggunakan bahasa tersebut dalam interaksi nyata, baik secara lisan maupun tulisan. Selain itu, hasil belajar Bahasa Inggris juga mengukur aspek keterampilan komunikatif yang mencakup *listening, speaking, reading, dan writing* sebagai empat keterampilan utama dalam pembelajaran bahasa.

Keberhasilan belajar Bahasa Inggris sangat dipengaruhi oleh lingkungan pembelajaran yang mendukung, terutama yang memberikan kesempatan praktik berbahasa secara intensif. Lingkungan yang kondusif, interaksi aktif antara guru dan siswa, serta kesempatan berdiskusi membuat proses belajar lebih bermakna. Penggunaan media pembelajaran seperti video, audio, aplikasi bahasa, dan simulasi percakapan juga membantu siswa memahami materi dengan lebih mudah karena dapat melihat contoh penggunaan bahasa dalam konteks nyata.

Selain itu, strategi pembelajaran yang melibatkan siswa secara aktif, seperti role play, presentasi, dan diskusi kelompok, berperan penting dalam meningkatkan kelancaran berbicara, keberanian, dan kepercayaan diri siswa. Hasil belajar Bahasa Inggris tidak hanya dilihat dari nilai akhir, tetapi juga dari kemampuan siswa menggabungkan pemahaman linguistik dengan keterampilan komunikasi yang efektif. Keberhasilan belajar mencakup aspek pengetahuan,

sikap, dan keterampilan, sehingga evaluasinya harus mempertimbangkan perkembangan siswa secara menyeluruh, bukan hanya melalui tes tertulis.

2.7 Penelitian Terdahulu

Tabel 2.1. Penelitian Terdahulu

Peneliti	Judul	Tahun	Data dan Metode yang Digunakan	Hasil
Nurmala, Eri Sasmita Susanto, I Made Widiarta	Implementasi Data Mining Untuk Memprediksi Kecocokan Gaya Belajar Siswa Sekolah Dasar	2024	Data; siswa SDN Sela sebanyak 136 siswa (kelas 1-6, 2023). Metode; algoritma C4.5 dengan pohon keputusan.	Prediksi gaya belajar siswa menghasilkan akurasi sebesar 86,76%, dengan menggunakan atribut kemampuan, hobi, dan metode pembelajaran."
Mustofa, Prasetyo, Supriadi, R. H. Yasri	Metode Decision Tree Dalam Pemilihan Gaya Belajar Pada Siswa Sekolah Dasar	2021	Data; siswa dari sekolah dasar. Metode; algoritma C4.5 untuk membangun pohon keputusan.	Algoritma C4.5 terbukti efektif dalam memodelkan gaya belajar siswa dengan akurasi yang tinggi, berdasarkan atribut-atribut yang tersedia
Suriani	Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma C4.5	2023	Data akademik mahasiswa; metode data mining C4.5	Algoritma C4.5 mampu memprediksi tingkat kelulusan dengan akurasi tinggi.
Endah Fauziningrum, Encis Indah Suryaningsih	Penerapan Data Mining Metode Decision Tree Untuk Mengukur Penguasaan Bahasa Inggris Maritim (Studi Kasus di	2021	Data; Taruna Universitas Maritim AMNI, menggunakan data nilai penilaian Bahasa Inggris Maritim pada mata kuliah	Menemukan bahwa kemampuan berbicara (speaking) memiliki dampak terbesar terhadap atribut lainnya, dan menunjukkan

	Universitas Maritim AMNI)		Bahasa Inggris Metode ; Algoritma Decision Tree (C4.5) untuk memprediksi kemampuan Bahasa Inggris Maritim berdasarkan empat atribut: speaking, listening, reading, dan writing.	bahwa ada kebutuhan untuk perbaikan dalam pembelajaran dan listening dan writing.
Justam et al.	Penerapan Algoritma C4.5 dan Random Forest untuk Pemetaan Kerusakan Jalan	2024	Data kerusakan jalan; metode C4.5 dan Random Forest	Random Forest menghasilkan akurasi lebih tinggi dibanding C4.5, namun C4.5 lebih mudah diinterpretasi.
Penelitian ini	Penerapan Teknik Data Mining dengan Algoritma Decision Tree C4.5 dalam Analisis Pola Belajar Siswa dalam Mata Pelajaran Bahasa Inggris	2025	Data: Data perilaku belajar siswa (absensi, kelas tambahan, penggunaan media digital dan nilai). Metode: Algoritma Decision Tree C4.5.	Masih dalam penelitian

Penelitian ini memiliki kebaruan yang terletak pada penerapan algoritma Decision Tree C4.5 secara lebih terfokus untuk menganalisis pola belajar siswa serta memprediksi hasil belajar Bahasa Inggris berdasarkan variabel yang dipilih secara spesifik sesuai karakteristik peserta didik. Berbeda dengan penelitian sebelumnya yang umumnya hanya meninjau hubungan antara pola belajar dan hasil belajar atau menggunakan algoritma C4.5 pada domain non-pendidikan,

penelitian ini mengintegrasikan proses KDD secara menyeluruh mulai dari preprocessing hingga interpretasi model. Keunggulan lainnya terletak pada pemilihan atribut yang lebih relevan terhadap konteks pembelajaran Bahasa Inggris, sehingga pohon keputusan yang dihasilkan lebih akurat dan mudah dipahami. Selain itu, penelitian ini menggunakan dataset aktual siswa dan memvalidasi model dengan pengukuran akurasi untuk memastikan reliabilitas hasil. Dengan demikian, penelitian ini memberikan kontribusi baru berupa model prediksi yang lebih terstruktur, terukur, dan aplikatif bagi guru dalam meningkatkan hasil belajar siswa.