

BAB II

LANDASAN TEORI

2.1. Prestasi Belajar

Pendidikan adalah upaya individu untuk mengembangkan dan meningkatkan kemampuan fisik dan mental mereka sesuai dengan standar sosial-budaya yang ada. Kemajuan suatu bangsa terkait erat dengan faktor pendidikan karena pendidikan memainkan peran penting dalam meningkatkan sumber daya manusia, yang merupakan unsur fundamental keberhasilan nasional. Untuk mencapai tujuan pendidikan, lembaga pendidikan harus beradaptasi dengan perkembangan dan perubahan masyarakat yang dinamis (Lestari et al., 2024). Untuk membangun sistem pendidikan yang unggul dan berkualitas tinggi, kegiatan pengajaran sangat penting untuk memperluas wawasan siswa dan mengembangkan kepribadian mereka. Tolok ukur keberhasilan dalam proses pembelajaran adalah capaian akademik yang diukur melalui skor ujian siswa, yang biasanya disajikan dalam bentuk rapor. capaian siswa pada dasarnya merefleksikan seberapa besar usaha mereka dalam belajar. Belajar sendiri adalah upaya individu yang mendorong perkembangan tingkah sebagai akibat dari kemampuan pribadi serta interaksi dengan lingkungan di sekitarnya (Kristiani & Pahlevi, 2021).

Prestasi belajar merupakan hasil yang diperoleh siswa setelah menjalani proses pembelajaran tertentu. Pencapaian ini mencerminkan sejauh mana siswa telah mengerti, menguasai, dan menggunakan materi yang diajarkan oleh guru,

sesuai dengan sasaran pembelajaran yang telah ditentukan. Hasil ini umumnya disajikan dalam bentuk angka, skor, atau rangking, yang ditentukan melalui berbagai metode penilaian seperti tes, ujian, pekerjaan rumah (PR), atau penilaian harian.

Prestasi belajar tidak hanya menunjukkan kemampuan mengingat seperti pengetahuan dan keterampilan berpikir siswa, tetapi juga mencakup aspek efektif seperti sikap dan motivasi, serta aspek psikomotorik berupa keterampilan praktis. semua ini merupakan bagian penting dari tujuan Pendidikan. Sebagai contoh, sebuah penelitian yang mengkaji motivasi dan minat belajar terhadap prestasi dalam mata Pelajaran Pendidikan kewarganegaraan menemukan bahwa motivasi dan minat belajar yang tinggi berkaitan positif dengan prestasi akademik siswa. Oleh karena itu, kedua aspek tersebut berfungsi sebagai indikator utama dalam menilai prestasi belajar siswa secara menyeluruh (Zihura et al., 2024).

2.1.1 Indikator Prestasi Belajar

Indikator prestasi belajar siswa sekolah dasar adalah alat ukur atau standar yang digunakan untuk menentukan seberapa sukses siswa dalam meraih tujuan pembelajaran yang telah ditetapkan di tingkat pendidikan dasar. Indikator ini bertindak sebagai gambaran capaian belajar siswa yang dapat diukur secara teratur, sehingga mampu menggambarkan kemampuan akademik, partisipasi dalam proses belajar, serta ketekunan siswa. Di lingkungan pendidikan dasar, indikator prestasi belajar tidak hanya menunjukkan hasil mengingat, tetapi juga mencerminkan perkembangan sikap dan kebiasaan belajar siswa secara keseluruhan. Menurut Thahir (2014:62 dalam (Albarado et al., 2020) menyatakan

bahwa indikator prestasi belajar siswa biasanya menggunakan hasil belajar, yang mencakup ranah psikologis seperti pengetahuan dan proses belajar. Hasil yang diperoleh dari proses belajar terdiri dari poin nilai (skor), dari mana besarnya prestasi belajar siswa dapat diamati.

Di lingkungan pendidikan formal, prestasi belajar tidak semata-mata dianggap sebagai hasil akhir yang tercermin dalam angka atau nilai, melainkan juga sebagai cerminan dari kemajuan kemampuan akademik siswa yang dicapai melalui proses pembelajaran yang terus-menerus. Dengan demikian, untuk mendapatkan pandangan yang objektif dan dapat diukur tentang prestasi belajar tersebut, diperlukan indikator-indikator yang jelas dan mudah diamati. Indikator prestasi belajar ini berperan sebagai pedoman untuk mengevaluasi pencapaian hasil belajar siswa, serta sebagai landasan dalam membuat keputusan di bidang pendidikan, baik oleh para guru, institusi sekolah, maupun para peneliti pendidikan.

Indikator prestasi belajar dalam penelitian ini meliputi nilai rapor siswa dalam enam mata pelajaran, yaitu Pendidikan Agama, Pendidikan Pancasila, Bahasa Indonesia, Matematika, Seni Rupa, dan PJOK, tingkat penyelesaian pembelajaran mereka. berdasarkan Kriteria Ketuntasan Minimal (KKM) sebesar 75, serta klasifikasi prestasi siswa. Berdasarkan nilai rata-rata dari enam mata pelajaran, nilai dibagi menjadi dua kategori: "Tuntas" dan "Tidak Tuntas". Nilai rapor merupakan indikator resmi yang mencerminkan capaian belajar siswa, karena merangkum hasil evaluasi dari beberapa aspek penilaian dalam satu periode pembelajaran. Sifatnya yang menyeluruh menjadikan nilai rapor sebagai

gambaran yang relatif stabil dan representatif terhadap perkembangan kemampuan kognitif siswa dari waktu ke waktu. Sementara itu, hasil ujian atau tes pembelajaran berfungsi untuk mengukur pemahaman siswa terhadap kompetensi tertentu yang telah dirancang oleh guru sebagai bagian dari instrumen penilaian proses pembelajaran. Di bidang penelitian pendidikan, terutama dalam studi kuasi-eksperimental, metode *pretest–posttest* sering diterapkan untuk mengukur keefektifan pembelajaran dengan membandingkan kemampuan siswa sebelum dan setelah intervensi dilakukan. Ketuntasan belajar menandai pencapaian standar minimum yang ditentukan oleh sekolah atau kurikulum, sedangkan klasifikasi prestasi belajar membagi siswa ke dalam kelompok tinggi, sedang, dan rendah sesuai dengan rentang nilai yang mereka peroleh. Adapun kemajuan hasil belajar diukur dari perbedaan skor antara pretest dan posttest, yang menggambarkan pengaruh intervensi pembelajaran tersebut.

Secara umum, indikator ketuntasan, klasifikasi, dan kemajuan hasil belajar memberikan pandangan objektif tentang efektivitas pembelajaran pada topik tertentu serta tingkat penguasaan kompetensi siswa.

2.1.2 Faktor-Faktor yang Mempengaruhi Prestasi Belajar

Pencapaian belajar siswa pada hakikatnya dipengaruhi oleh berbagai faktor, baik yang bersumber dari pengaruh eksternal maupun internal. Berikut adalah Faktor-faktor yang mempengaruhi proses pembelajaran siswa:

A. Pengaruh Faktor Eksternal terhadap proses pembelajaran

1. Faktor Lingkungan

Lingkungan merupakan bagian integral dari kehidupan siswa; di sanalah mereka menerima pendidikan dan berinteraksi dalam ekosistem yang hidup, menumbuhkan saling ketergantungan. Faktor lingkungan mencakup faktor sosial dan non-sosial. Lingkungan sosial dapat dibagi lagi menjadi lingkungan sekolah, lingkungan keluarga, dan lingkungan masyarakat; sedangkan lingkungan non-sosial berkaitan dengan berbagai fenomena yang terjadi di alam. Baik lingkungan sosial maupun non-sosial dapat memiliki dampak positif atau negatif. Oleh karena itu, guru dan orang tua harus mempertimbangkan bagaimana mengatur lingkungan belajar anak untuk meminimalkan pengaruh negatif. Untuk meningkatkan partisipasi anak dalam proses belajar di sekolah, sekolah perlu bekerja sama erat dengan lingkungan keluarga dan masyarakat. Tanpa dukungan dari lingkungan keluarga, masyarakat, dan semua pemangku kepentingan terkait, sekolah tidak dapat berhasil mencapai visi dan misi pendidikannya.

2. Faktor Instrumental

Faktor instrumental mengacu pada fasilitas, infrastruktur, atau bukti nyata yang mendukung pembelajaran, seperti gedung sekolah, Ruang sholat, buku, alat bantu belajar, fasilitas belajar, lapangan olahraga, kurikulum sekolah, peraturan sekolah, buku panduan, silabus, bahan ajar, dll. Infrastruktur dan fasilitas belajar yang komprehensif dapat menciptakan lingkungan belajar yang baik, tetapi jika infrastruktur dan fasilitas sekolah tidak memadai. Kesempurnaan tidak selalu berarti proses pembelajaran yang buruk, karena ada faktor lain, seperti Dedikasi Guru, efektivitas

pengajaran, metode pengajaran, nilai-nilai sekolah, dan lingkungan belajar secara keseluruhan.

B. Pengaruh Faktor Internal terhadap proses pembelajaran

1. Faktor Fisiologis

Faktor fisiologis merupakan salah satu faktor yang memengaruhi pembelajaran karena berkaitan erat dengan kesehatan dan kondisi fisik. Kondisi fisik yang buruk dapat memengaruhi konsentrasi. Untuk menjaga kondisi fisik yang baik, kita harus memperhatikan pola makan dan menjaga jadwal tidur yang teratur. Masalah fisiologis atau kesehatan seperti kelelahan, penyakit, dan kekurangan gizi dapat menghambat konsentrasi, mengurangi motivasi, dan mengganggu kemampuan siswa untuk menyerap materi pembelajaran dengan optimal.

2. Faktor Psikologis

Faktor psikologis meliputi kecerdasan, minat, motivasi, bakat, sikap, kematangan, dan kesiapan. Kecerdasan adalah keterampilan yang mencakup tiga jenis: kemampuan untuk mengatasi dan beradaptasi dengan situasi secara cepat dan efektif, kemampuan untuk memahami atau menerapkan konsep abstrak secara efektif, dan kemampuan untuk memahami dan mempelajari hubungan dengan cepat. Perhatian adalah keadaan pikiran yang sangat aktif di mana pikiran terfokus pada satu atau sekelompok objek. Perhatian muncul dalam dua cara: satu adalah perhatian yang berasal dari keinginan (perhatian volisional), dan yang lainnya adalah perhatian yang tidak berasal dari keinginan (perhatian non-

volisional). Minat adalah kecenderungan untuk terus menerus fokus pada dan mengingat aktivitas tertentu. Minat sangat berkaitan dengan kepribadian. Orang yang tertarik pada bidang tertentu cenderung lebih fokus dan memiliki keinginan yang kuat untuk terus belajar.

Mari kita bahas topik ini lebih dalam. Secara umum, bakat adalah kemampuan laten yang dimiliki setiap orang, artinya potensi untuk berprestasi (sukses) berdasarkan kemampuan individu. Untuk mewujudkan potensi ini sepenuhnya, pendidikan dan pelatihan juga diperlukan untuk mempersiapkan kesuksesan di masa depan. Hal ini akan secara signifikan mempengaruhi tingkat keterlibatan mereka dalam menyelesaikan tugas-tugas yang diberikan oleh guru. Siswa dengan motivasi belajar yang tinggi akan bekerja dengan tekun. Motivasi memainkan peran penting dalam proses pembelajaran. Motivasi belajar membimbing siswa untuk mempersiapkan diri dan mengingatkan mereka tentang tujuan pembelajaran yang harus mereka capai. Kecenderungan motivasi siswa juga tercermin dalam kinerja mereka di kelas. Sikap positif terhadap suatu mata pelajaran bermanfaat bagi proses pembelajaran siswa.

Pandanglah tugas pembelajaran sebagai tantangan yang harus diselesaikan dengan sangat baik. Motivasi adalah. Variabel yang kurang stabil, artinya motivasi siswa dapat berubah dalam waktu singkat. Sikap Ini adalah gejala intrinsik dengan dimensi emosional, yaitu, respons yang relatif tetap terhadap hal-hal tertentu. Suatu tanda, dan sebaliknya. Kedewasaan adalah tahap pertumbuhan di mana seluruh fisiologi siswa.

Semua organ siap untuk melakukan keterampilan baru. Pada saat yang sama, kesiapan mengacu pada kemampuan. Respons atau reaksi. Keduanya saling terkait dan memengaruhi proses pembelajaran. Kedewasaan adalah Kesiapan fisiologis mengacu pada keadaan kesiapan anak untuk belajar. Misalnya, apakah tangan secara fisiologis mampu menulis. Kesiapan, di sisi lain, berkaitan dengan proses psikologis. Misalnya, jika seorang siswa merasa cemas sebelum kelas, itu menunjukkan bahwa ia belum siap untuk belajar (Sihombing et al., 2024).

Berdasarkan penjelasan faktor internal dan eksternal, penelitian ini tidak dapat secara langsung mengukur faktor fisiologis dan psikologis. Oleh karena itu, diperlukan beberapa indikator untuk mewakili faktor-faktor tersebut dalam bentuk data yang terukur. Indikator yang digunakan dalam penelitian ini meliputi nilai penilaian sumatif akhir semester siswa dalam berbagai mata pelajaran. Nilai-nilai tersebut meliputi nilai Pendidikan agama, Pendidikan Pancasila, Bahasa Indonesia, Matematika, Seni Rupa, PJOK. Prestasi akademik ini mencerminkan kemampuan kognitif, pemahaman, dan hasil belajar siswa, yang berkaitan erat dengan faktor psikologis seperti kecerdasan, minat, dan motivasi belajar (Zavy et al., 2025).

2.2. Data Mining

Data Mining mengacu pada proses menemukan pola atau informasi menarik dari sejumlah besar data yang tersimpan menggunakan teknik atau metode tertentu, dan kemudian menganalisisnya (Manik & Ariesta, 2023). Data Mining yang sering disebut sebagai penemuan pengetahuan basis data *knowledge*

discovery in database (KDD), melibatkan pengumpulan dan penggunaan data historis untuk menemukan pola dan hubungan dalam kumpulan data besar. Terdapat banyak teknik, metode, dan algoritma penambangan data (Gunawan & Purwayoga, 2022). Penerapan Data Mining dalam penelusuran pengetahuan dapat dilihat sebagai teknik penilaian formatif, yaitu teknik untuk mengevaluasi program pendidikan yang masih dalam pengembangan (Desiani et al., 2020).

Data Mining atau Penambangan data adalah serangkaian prosedur yang dirancang untuk menemukan informasi tambahan yang saat ini tidak ada dalam basis data. Proses ini melibatkan eksplorasi pola dan model data yang dapat mengubah data mentah menjadi wawasan yang berharga. Hal ini dicapai dengan mengekstrak dan menganalisis model yang berharga atau menarik dari basis data. Penambangan data merupakan bagian integral dari proses Penemuan Pengetahuan dalam Basis Data *Knowledge Discovery in Databases* (KDD), yang tujuannya adalah untuk mengeksplorasi basis data guna mengekstrak informasi yang berharga. Hal ini menekankan penggunaan teknologi terintegrasi dan prosedur eksplorasi ilmiah (Lestari et al., 2024).

Fungsi Data Mining adalah sebuah fungsi yang membantu kita memperoleh pemahaman yang lebih mendalam tentang data yang diamati. Melalui serangkaian proses, kita bertujuan untuk memahami pola perilaku data. Data ini kemudian dapat digunakan untuk menentukan karakteristiknya. Dengan menggunakan kemampuan penambangan data deskriptif, kita dapat menemukan pola tersembunyi tertentu di dalam data. Sementara itu, fungsi prediktif menjelaskan cara menemukan pola spesifik dalam data melalui serangkaian

proses. Pola-pola ini Hal ini dapat diidentifikasi dari berbagai variabel dalam data (Senubekti & Dewi, 2022).

2.3. *Knowledge Discovery In Database (KDD)*

Knowledge Discovery In Database (KDD) Merupakan Proses sistematis untuk mengekstrak pengetahuan yang valid, baru, berguna, dan mudah dipahami dari Kumpulan data yang besar. KDD tidak hanya berfokus pada algoritma penambahan data, tetapi juga mencakup seluruh proses mulai dari pemilihan data hingga interpretasi pengetahuan yang diperoleh. KDD adalah proses yang memungkinkan pendekatan komprehensif dalam pengolahan data, mulai dari pengumpulan data hingga memperoleh hasil (Glowania & Kozak, 2023). Menurut (Mughnyanti & Ginting, 2023) *knowledge discovery in database (KDD)* adalah serangkaian proses yang bertujuan untuk mengekstraksi pola-pola penting atau menarik dari sejumlah data berukuran sangat besar yang tidak dapat dikenali secara manual.

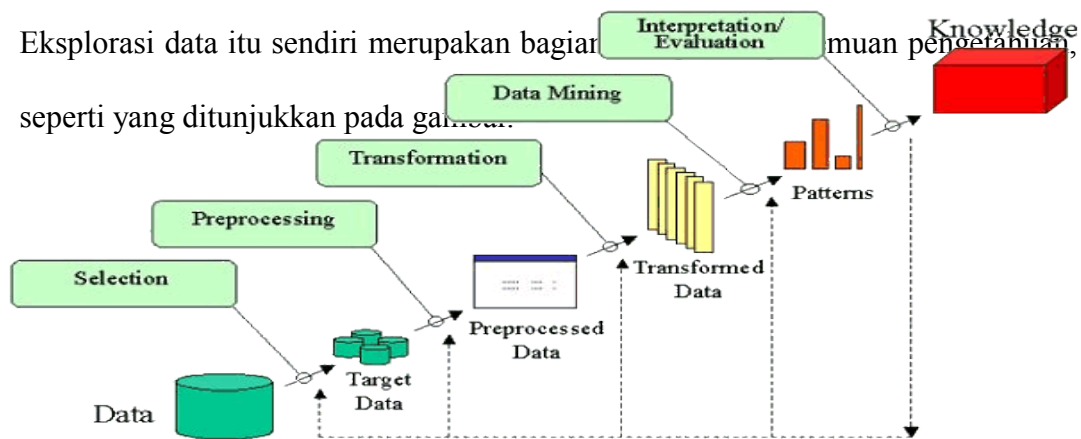
Proses dimulai dengan fase pengumpulan data, yaitu Data dikumpulkan dari sensor yang terpasang pada sistem dan disimpan dalam basis data di server (*internet of things*) (Gamaliel et al., 2022).

knowledge discovery in database (KDD) dan data mining merupakan istilah yang sering digunakan secara bergantian untuk menjelaskan proses penggalian informasi tersembunyi dalam suatu basis data yang besar. Sebenarnya kedua istilah tersebut memiliki konsep yang berbeda, tetapi berkaitan satu sama lain. dan salah satu tahapan dalam keseluruhan proses KDD adalah data mining (Sari et al., 2023).

2.3.1 Tahapan Knowledge Discovery In Database (KDD)

Knowledge discovery in database melibatkan teknik integrasi dan analisis pola di berbagai kumpulan data. Penemuan, interpretasi, dan visualisasi ilmiah.

Eksplorasi data itu sendiri merupakan bagian dari proses penemuan pengetahuan, seperti yang ditunjukkan pada gambar.



Gambar 2.1 tahapan *knowledge discovery in database*

Berdasarkan gambar diatas dapat dijelaskan bahwa tahapan *knowledge discovery in database* adalah sebagai berikut:

1. Data Selection

Data Selection (Pemilihan Data) merupakan proses penyiapan data dengan memilih sekumpulan data yang menjadi target/atribut/indikator dari dalam data. Pada penelitian ini, tahap seleksi mencakup penentuan atribut nilai akademik siswa per mata pelajaran yang akan digunakan sebagai variabel input model.

2. Pre-processing/Cleaning

Proses cleaning (pembersih data) meliputi koreksi kesalahan data, penghapusan data duplikat, dan pengecekan. Ketidakkonsistenan seperti

kesalahan ejaan. Kualitas data yang buruk pada tahap ini akan secara langsung menurunkan akurasi model yang dihasilkan (Kurniawan & Defit, 2020).

3. Data Transformasi

Transformasi data merujuk pada proses mengubah cara data diatur atau diproses. Pemrosesan fitur diterapkan pada data mentah untuk membuatnya lebih sesuai atau lebih mudah dipahami untuk metode pemodelan atau analisis data, seperti normalisasi numerik dan pengkodean variabel kategorikal.

4. Data Mining

Data mining adalah proses sistematis yang memanfaatkan kecerdasan buatan, teknik statistik dan matematika untuk mengekstrak pengetahuan dan informasi baru dari sejumlah besar data yang tersimpan dalam repositori data. Algoritma *decision tree* C4.5 diterapkan pada tahap ini untuk membangun model prediktif (Lestari et al., 2024).

5. Interpretation/ evaluation

Pola informasi yang dihasilkan data mining perlu disajikan dalam bentuk yang mudah dipahami oleh para pemangku kepentingan. Presentasi. Evaluasi dilakukan menggunakan *confusion matrix* dan *cross-validation* untuk memastikan validitas dan reliabilitas model yang dihasilkan (Sari et al., 2023).

2.4. *Decision Tree*

Decision tree merupakan metode terkenal buat menyusunkan bermacam perihai serta keluarannya lumayan gampang dimengerti serta dianalisis. Model ini nampak semacam tumbuhan ataupun hierarki serta digunakan buat membuat prediksi. Ide utama di balik pohon keputusan adalah mengubah data menjadi pohon keputusan, atau sekelompok aturan yang menampilkan gimana keputusan terbuat. Strategi ini digunakan buat mendekati nilai- nilai diskrit dari sesuatu guna sasaran (Mukminna et al., 2025).

Dalam bidang pendidikan, *decision tree* banyak dimanfaatkan untuk memprediksi prestasi belajar siswa berdasarkan data akademik yang tersedia. Dengan memahami kemampuan akademik siswa, sekolah dapat memanfaatkan sistem manajemen data terkomputerisasi, khususnya data yang berkaitan dengan kemampuan akademik siswa, untuk menyediakan solusi pembelajaran di masa mendatang. Hal ini sangat bermanfaat bagi sekolah. mengambil langkah yang tepat sejak dini sebelum prestasi siswa menurun lebih jauh (Lestari, & Lestari, 2024). Pohon keputusan umumnya digunakan untuk memperoleh informasi yang bertujuan untuk meningkatkan proses pengambilan keputusan. Pohon keputusan disusun berdasarkan simpul akar. (atau titik kontrol) dimulai, pengguna membuat keputusan melalui *leaf node*. Pengguna memecahkan masalah setiap *root node* berdasarkan node akar sesuai dengan algoritma *decision tree* (Sukri & Handrianto, 2024).

2.5. Algoritma (C4.5)

Algoritma C4.5 adalah metode komputasi untuk menghasilkan pohon keputusan. Pohon keputusan merupakan metode yang ampuh dan terkenal untuk klasifikasi dan prediksi. Algoritma pohon keputusan dapat mengubah dataset besar menjadi pohon keputusan. aturan-aturan ini. Aturan-aturan ini mudah dipahami dengan bahasa yang sederhana. Sederhana dan jelas (Lestari et al., 2024).

Algoritma C4.5 adalah salah satu algoritma eksplorasi data modern; algoritma ini merupakan pengembangan dari algoritma ID3. Algoritma ini merupakan penyempurnaan dari algoritma ID3. Penyempurnaan tersebut meliputi penanganan atribut numerik, nilai yang hilang, dan noise dalam dataset, serta pembuatan aturan berdasarkan model tanaman yang dibuat (Husein et al., 2022).

Algoritma C4.5 adalah salah satu metode klasifikasi yang paling sering digunakan dalam data mining, karena dapat memproses data dan memberikan informasi yang berguna untuk pengambilan keputusan. Algoritma ini membentuk model pohon keputusan yang dapat menggambarkan hubungan antar atribut secara sistematis. Berkat sifat-sifat ini, algoritma C4.5 menawarkan sejumlah kelebihan yang membuatnya efektif dan fleksibel untuk diterapkan pada berbagai masalah klasifikasi dan prediksi, terutama dalam pemrosesan data dalam jumlah besar. Berikut adalah kelebihan algoritma C4.5:

1. Mampu mengklasifikasikan data dalam bentuk pohon keputusan.

Algoritma C4.5 membentuk model klasifikasi berupa pohon keputusan

sehingga hasil klasifikasi mudah dipahami dan diinterpretasikan sebagai dasar pengambilan keputusan.

2. Memiliki kinerja yang baik dengan tingkat kesalahan rendah. Sistem yang dibangun menggunakan algoritma C4.5 menunjukkan hasil kinerja yang baik dan minim kesalahan, terutama dalam membedakan data anomali dengan data normal.
3. Efektif digunakan pada data berukuran besar. Algoritma C4.5 mampu mengolah dan mengeksplorasi data dalam jumlah besar untuk menemukan pola yang berguna dalam proses pengambilan keputusan.
4. Mendukung pengambilan keputusan strategis. Informasi yang dihasilkan dari klasifikasi C4.5 dapat digunakan sebagai solusi nyata bagi pengambilan keputusan, khususnya dalam perencanaan dan pengembangan strategi bisnis.
5. Bermanfaat dalam pengendalian persediaan . penerapan algoritma C4.5 dapat membantu perusahaan mengendalikan kelebihan dan kekurangan persediaan, sehingga mampu menekan biaya produksi dan meminimalkan risiko kerugian.
6. Mampu menggali pola perilaku konsumen. Melalui analisis database transaksi, algoritma C4.5 dapat mengidentifikasi kebiasaan dan perilaku konsumen, seperti produk yang sering atau jarang dibeli serta pola peningkatan pembelian.
7. Bersifat fleksibel dan multidisipliner. Algoritma C4.5 dapat diterapkan pada berbagai bidang, seperti analisis pasar, manajemen hubungan

pelanggan (CRM), manajemen risiko, peramalan, deteksi penipuan, serta penambangan teks dan web.

8. Merupakan metode klasifikasi dan prediksi yang kuat. Algoritma C4.5 dikenal sebagai metode yang andal dalam data mining karena mampu mengekstraksi informasi berharga dari database berukuran besar menggunakan teknologi pengenalan pola.

Berikut adalah kekurangan algoritma C4.5:

1. Sangat bergantung pada kualitas data. Algoritma C4.5 membutuhkan data yang telah melalui proses persiapan dengan baik. Data yang tidak lengkap, bising, atau tidak konsisten dapat menurunkan kualitas hasil klasifikasi .
2. Memerlukan data yang cukup kompleks untuk hasil optimal. Jurnal menyebutkan bahwa keakuratan prediksi dapat ditingkatkan apabila digunakan data yang lebih kompleks dan kaya atribut, sehingga pada data sederhana hasil klasifikasi kurang maksimal.
3. Pohon keputusan dapat menjadi terlalu besar. Pada data dengan banyak atribut dan variasi nilai, pohon keputusan yang dihasilkan dapat menjadi kompleks dan sulit untuk diinterpretasikan secara visual maupun konseptual.
4. Masih memerlukan pengembangan lanjutan. Dalam bagian saran, penulis menyatakan bahwa algoritma C4.5 masih perlu dikembangkan agar mampu menghasilkan prediksi yang lebih tepat, terukur, dan akurat (Senubekti & Dewi, 2022)

Pohon keputusan biasanya menangani data terstruktur, yang diorganisir dalam format tabel, berisi atribut dan catatan. Atribut adalah parameter yang digunakan untuk membangun pohon keputusan. Metode pohon keputusan C4.5 menggunakan recall sebagai kriteria pemisahan dan menyesuaikannya selama pemilihan atribut pemisahan. Fitur pemisahan dibangun menggunakan algoritma C4.5. Langkah pertama dalam pembuatan pohon keputusan, terdiri dari tahapan sebagai berikut:

1. Hitung nilai entropi
2. Tentukan nilai gain rasio untuk masing-masing atribut.
3. Atribut dengan gain ratio tertinggi dipilih sebagai root, sedangkan atribut dengan gain ratio lebih rendah dari root dipilih sebagai cabang.
4. Hitung ulang gain ratio untuk setiap atribut, tidak termasuk atribut yang dipilih sebagai root pada langkah sebelumnya.
5. Pilih atribut dengan gain ratio tertinggi sebagai cabang.
6. Ulangi langkah 4 dan 5 hingga nilai gain yang dihasilkan sama dengan nol untuk semua kualitas yang tersisa (Lestari et al., 2024).

Untuk menghitung nilai entropy dapat dihitung dengan persamaan:

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

s : himpunan data (dataset)

n : jumlah kelas

p_i : proporsi data pada kelas ke- i

$\log_2$: logaritma basis 2

Sedangkan untuk menghitung nilai information gain dapat dihitung dengan persamaan:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i)$$

S : seluruh data

A : atribut yang diuji

S_i : subset data ke- i hasil pembagian berdasarkan atribut A

$|S|$: jumlah data keseluruhan

$|S_i|$: jumlah data pada subset ke- i

2.6. Alat Bantu Aplikasi

Aplikasi Rapidminer dan Microsoft Excel adalah dua alat bantu *software* yang sering dipakai untuk mengolah data, melakukan analisis statistik, dan menerapkan tekni data mining. Kedua aplikasi ini sangat membantu dalam kegiatan penelitian, terutama di bidang Pendidikan, di mana data siswa dianalisis secara teratur dan berdasarkan fakta yang jelas.

2.6.1 RapidMiner

Rapidminer merupakan *software* yang bersifat *open source* (sumber terbuka). Rapidminer adalah solusi penelitian data mining, penelitian text mining, dan penelitian prediktif. Rapidminer menggunakan berbagai Teknik deskriptif dan prediktif untuk memberikan wawasan kepada pengguna yang membantu mereka membuat keputusan optimal. Rapidminer memiliki 500 operator penambangan data yang mencakup fungsi input, output, pra-pemrosesan data, dan visualisasi.

Rapidminer adalah perangkat lunak analisis data mandiri yang dapat digunakan sebagai mesin penambangan data atau diintegrasikan ke dalam produk lain (Ramdani et al., 2022).

2.6.2 Microsoft Excel

Microsoft Excel merupakan perangkat lunak pengolah data yang menyajikan data dalam format (*spreadsheet*). Perangkat lunak yang dikembangkan oleh Microsoft Corporation dan telah menjadi salah satu aplikasi paling banyak digunakan dalam dunia Pendidikan, bisnis, maupun penelitian ilmiah. Secara sederhana, *excel* bekerja layaknya sebuah table besar yang terdiri dari baris kolom, di mana setiap kotaknya disebut *cell* dan dapat diisi dengan teks, angka, maupun rumus perhitungan otomatis. Microsoft excel menjadi tempat pertama untuk merapikan, memilah, dan meringkas data tersebut sebelum dianalisis lebih lanjut.

2.7. Penelitian Terdahulu

Penelitian terdahulu merupakan kajian terhadap penelitian-penelitian sebelumnya yang berkaitan dengan topik penelitian ini, khususnya mengenai prediksi prestasi belajar siswa menggunakan algoritma *Decision Tree* C4.5. Studi ini bertujuan untuk menguraikan metode dan hasil penelitian yang relevan sebagai dasar dan pembandingan dalam penyusunan skripsi, serta sebagai acuan dalam pengembangan penelitian berjudul “Analisis Prediksi Prestasi Belajar Siswa SD Negeri 117854 Sikopi-Kopi Desa Pulo Dogom Menggunakan Algoritma Decision Tree (C4.5) pada RapidMiner.” Berikut adalah penelitian terdahulu pada Tabel 2.1 berikut:

Tabel 2.1 Penelitian Terdahulu

Judul Penelitian	Penulis dan Tahun	Hasil	Kelebihan	Kekurangan
Penerapan Algoritma C4.5 Dalam Menentukan Prediksi Prestasi Siswa Pada SMPN 51 Jakarta	Muhammad Hamdan Sukri, Yopi Handrianto (2024)	Model prediksi prestasi siswa berhasil dibangun sesuai implementasi, akurasi tidak disebutkan jelas	Fokus pada prediksi siswa melalui <i>machine learning</i> dengan rapidminer.	Tidak ada metrik performa lengkap di abstrak (Sukri & Handrianto, 2024).
Penerapan data mining untuk memprediksi prestasi siswa sekolah dasar menggunakan algoritma C4.5	Telutci, Rika Harman (2024).	Algoritma C4.5 berhasil mengklasifikasikan prestasi belajar siswa SD ke dalam kategori tertentu berdasarkan nilai akademik dan kehadiran dengan Tingkat akurasi sekitar 99%, serta menghasilkan pohon Keputusan yang mudah dipahami sebagai dasar pengambilan Keputusan sekolah.	Menghasilkan akurasi sangat tinggi, model mudah dipahami, relevan untuk SD, dan menggunakan atribut akademik yang umum tersedia.	Data hanya dari satu sekolah, menggunakan WEKA, tidak ada perbandingan algoritma, dan variabel penelitian terbatas (Telutci, Harman, 2024).
Analysis of C4.5 Algorithm Performance for Predicting	Dimas Lucky Satya Purnama, Untoro	Algoritma <i>decision tree</i> C4.5 memiliki kinerja yang sangat baik	Akurasi prediksi sangat tinggi, model mudah dipahami,	Penelitian ini dilakukan padaa Tingkat SMK, bukan SD, belum

Student Achievement Based on Socio-Economic Status, Motivation, Discipline, and Past Achievement	Apsiswanto (2025).	dalam memprediksi prestasi belajar siswa. Model yang dibangun menghasilkan Tingkat akurasi sekitar 98,02% dan AUC 0,96, sehingga model memiliki kinerja yang sangat baik	efektif untuk data presentasi akademik.	dibandingkan dengan algoritma lain (Lucky et al., 2025).
Implementation of Data Mining to Predict Student Study Period with Decision Tree Algorithm (C4.5)	Kirana Alyssa Putri, Dimas Febriawan, Firman Noor Hasan (2024).	Algoritma <i>decision tree</i> C4.5 dengan rapidminer untuk memprediksi periode studi mahasiswa, yaitu kelulusan tepat waktu atau tidak. Hasil pengujian menunjukkan Tingkat akurasi sebesar 83,87%, sehingga model dinilai cukup baik dalam melakukan klasifikasi.	Algoritma C4.5 yang mudah dipahami serta rapidminer sebagai alat bantu pengolahan data, evaluasi model dilakukan dengan metode yang jelas sehingga hasil penelitian dapat dipercaya.	Penelitian ini pada lama studi mahasiswa bukan pada prestasi belajar siswa (Putri et al., 2024).
Predicting Prospective Student Interests Using the C4.5 Algorithm and Naive Bayes	Annisa Amanda, Ali Akbar Ritonga, Elysa Rohayani Hasibun (2025).	Kedua metode mampu mengklasifikasi minat calon mahasiswa dengan akurasi tinggi 97%. C4.5 menghasilkan pola Keputusan	Model mudah dipahami, hasil akurasi tinggi, mampu menunjukkan faktor paling berpengaruh terhadap minat calon	Objek penelitian terbatas pada calon mahasiswa, bukan siswa sekolah dasar (Amanda et al., 2025)

		yang jelas, sedangkan naïve bayes memiliki nilai recall lebih tinggi.	mahasiswa.	
Penerapan Data Mining Klasifikasi Kepuasan Pelanggan Transportasi Online Menggunakan Algoritma C4.5	Ely Jannah, Volvo Sihombing, Masrizal (2023).	Penelitian menerapkan algoritma C4.5 untuk mengklasifikasikan tingkat kepuasan pelanggan transportasi online berdasarkan atribut harga, fasilitas, pelayanan, dan loyalitas. Data yang digunakan sebanyak 155 data, terdiri dari 100 data training dan 55 data testing. Hasil perhitungan menunjukkan atribut harga memiliki nilai gain ratio tertinggi (0,066) sehingga menjadi akar pohon keputusan. Sistem berhasil mengklasifikasikan pelanggan dengan hasil 91	Penelitian ini memiliki kelebihan karena menerapkan algoritma C4.5 secara sistematis mulai dari tahap seleksi data, preprocessing, transformasi data, hingga pembentukan pohon keputusan. Hasil klasifikasi yang diperoleh dapat divisualisasikan dalam bentuk pohon keputusan sehingga lebih mudah dipahami dan digunakan sebagai dasar pengambilan keputusan. Selain itu, penelitian ini juga mampu	Penelitian ini masih memiliki keterbatasan karena objek penelitian hanya berfokus pada pelanggan transportasi online sehingga hasil penelitian kurang dapat digeneralisasikan pada bidang lain. Selain itu, penelitian tidak menyajikan pengukuran performa model secara lengkap seperti nilai akurasi, <i>precision</i> , <i>recall</i> , atau <i>confusion matrix</i> sehingga tingkat keandalan model sulit dibandingkan dengan penelitian sejenis. Jumlah data yang digunakan juga relatif terbatas,

		pelanggan pada satu kategori dan 9 pelanggan pada kategori lainnya serta menghasilkan pohon keputusan yang dapat digunakan sebagai dasar pengambilan keputusan perusahaan.	mengidentifikasi atribut yang paling berpengaruh terhadap tingkat kepuasan pelanggan, yaitu atribut harga, melalui perhitungan gain ratio yang dilakukan secara rinci.	yaitu 155 data, sehingga kemampuan model dalam merepresentasikan kondisi yang lebih luas masih perlu ditingkatkan (Jannah et al., 2023).
Prediction Of Stunting in Toddlers Combining the Naïve Bayes Method and the C4.5 Algorithm	Sri Melyani, Syaiful Zuhri Harahap, Irmayanti (2024)	Kedua metode mencapai akurasi 98% dan presisi 100 dari 200 data uji balita.	Membandingkan dua metode sekaligus, akurasi sangat tinggi, menggunakan data rekam medis nyata.	Domain Kesehatan balita, bukan Pendidikan, sehingga tidak langsung relevan dengan prediksi prestasi belajar siswa.(Melyani et al., 2024)

Berdasarkan penelitian-penelitian terdahulu, algoritma *decision tree* C4.5 telah terbukti mampu memprediksi prestasi belajar dengan akurasi yang tinggi. Namun, penelitian yang ada masih banyak berfokus pada jenjang SMP, SMK, maupun perguruan tinggi, sehingga kajian pada tingkat sekolah dasar, khususnya di daerah pedesaan, masih sangat terbatas, penelitian ini melengkapi kekurangan tersebut dengan menerapkan algoritma C4.5 pada data siswa SD Negeri 117854 Sikopi-kopi menggunakan RapidMiner, melalui tahapan KDD yang lengkap mulai dari seleksi data hingga interpretasi hasil.