

PENERAPAN DATA MINING MENGGUNAKAN ALGORITMA C4.5 UNTUK PREDIKSI PRESTASI BELAJAR SISWA DI MTS YAMUALLIM PANONGAN

Hasanah Lestari, Ade Irma Purnamasari, Tati Suprapti,
Teknik Informatika, STMIK IKMI Cirebon
Jalan Perjuangan No.10B, Karyamulya, Kota Cirebon, Indonesia
hasanahlestari2000@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk meningkatkan akurasi prediksi prestasi belajar siswa di MTs Yamu'allim Panongan dengan mengoptimalkan penggunaan algoritma C4.5. Algoritma ini dipilih karena dapat menangani kompleksitas data dan menghasilkan hasil yang mudah diinterpretasi. Dataset yang mencakup prestasi belajar siswa dari semester sebelumnya digunakan untuk melatih dan menguji model, menciptakan dasar yang kuat untuk meningkatkan kualitas prediksi prestasi belajar. Penelitian juga melibatkan analisis mendalam terhadap berbagai aspek, termasuk aspek sosial dan lingkungan, karena pemahaman terhadap faktor-faktor ini dianggap penting dalam memengaruhi prestasi belajar siswa. Dengan implementasi C4.5 yang dioptimalkan, penelitian ini tidak hanya bertujuan meningkatkan akurasi prediksi, tetapi juga memberikan gambaran komprehensif mengenai faktor-faktor yang berkontribusi terhadap prestasi belajar siswa. Hasil penelitian menunjukkan bahwa optimasi algoritma C4.5 secara signifikan meningkatkan tingkat akurasi prediksi prestasi belajar siswa. Temuan ini dapat membantu pihak sekolah dalam mengidentifikasi siswa yang memerlukan perhatian khusus atau bantuan tambahan untuk meningkatkan prestasinya. Model klasifikasi menunjukkan akurasi yang tinggi dalam memprediksi kategori prestasi belajar, memberikan manfaat konkret dengan memberikan pandangan yang lebih tepat mengenai potensi dan kebutuhan masing-masing siswa.

Kata kunci : *Prediksi Prestasi Belajar, Algoritma C4.5, Optimasi, Akurasi Prediksi Tinggi, MTs Yamu'allim Panongan.*

1. PENDAHULUAN

Pendidikan adalah suatu usaha yang bertujuan yang dilakukan oleh individu untuk mengembangkan dan meningkatkan bakat jasmani dan rohaninya, sesuai dengan standar masyarakat dan budaya yang ada. Kemajuan suatu negara sangat erat kaitannya dengan faktor pendidikan, karena pendidikan mempunyai peran yang sangat penting dalam meningkatkan sumber daya manusia yang merupakan elemen fundamental dalam keberhasilan suatu negara. Untuk mencapai tujuan pendidikan, penyelenggaraan pendidikan harus disesuaikan dengan kemajuan dan perubahan masyarakat yang dinamis. Sekolah sebagai satu-satunya lembaga pendidikan resmi yang dikelola pemerintah mempunyai peranan penting dalam mencapai tujuan pendidikan nasional.[1]

Tesis merupakan tonggak penting bagi siswa di tahun terakhir mereka. Baik Perguruan Tinggi Negeri (PTN) maupun Perguruan Tinggi Swasta (PTS) mewajibkan mahasiswanya memenuhi prasyarat untuk memperoleh gelar sarjana dengan berhasil menyelesaikan tesis pada bidang keilmuan spesifiknya. Pada tahap awal penulisan skripsi, seringkali mahasiswa dihadapkan pada ketidakpastian dalam memilih tema atau judul skripsi. Ketidakpastian ini dapat menghambat penyerahan skripsi tepat waktu. Langkah pertama dalam memilih judul skripsi adalah mengevaluasi sejauh mana pendekatan tersebut dapat berdampak pada proses penulisan skripsi. Perhatian segera harus diberikan untuk menyelesaikan masalah yang

berkaitan dengan tesis, dan salah satu strategi potensial adalah dengan memanfaatkan sistem teknologi informasi itu sendiri. Oleh karena itu, penulis mengembangkan sistem rekomendasi dengan memanfaatkan algoritma Decision Tree untuk membantu mahasiswa dalam memilih judul skripsi yang sesuai dengan kemampuannya.[2]

Penelitian ini berfokus untuk menyelidiki dampak perbaikan di bidang Informatika, yaitu peningkatan algoritma C4.5, terhadap prediksi prestasi akademik siswa di MTs Yamu'allim Panongan dengan tepat. Fokus utama berpusat pada kemajuan teknologi yaitu penerapan teknologi digital dan pendidikan online. Proyek ini bertujuan untuk menyelidiki bagaimana peningkatan di bidang Informatika dapat meningkatkan keakuratan prediksi keberhasilan siswa, khususnya dengan mengatasi hambatan dalam memahami berbagai faktor yang mempengaruhi prestasi, seperti perbedaan individu dan kompleksitas sosial-lingkungan.

Studi ini menyelidiki penelitian sebelumnya, khususnya penggunaan algoritma C4.5, yang telah menunjukkan peningkatan presisi. Namun, hal ini juga menggarisbawahi perlunya optimasi lebih lanjut untuk mengakomodasi variasi atribut individu siswa. Perhatian khusus difokuskan pada keterbatasan penelitian sebelumnya yang belum cukup menyelidiki variabel khas siswa. Tujuan utama penelitian ini adalah untuk memperbaiki dan menyelidiki kekurangan ini secara lebih mendalam dengan menganalisis ciri-ciri individualitas.

Tujuan utama dari proyek ini adalah untuk meningkatkan pemahaman tentang faktor-faktor yang mempengaruhi prestasi akademik siswa di MTs Yamu'allim Panongan dengan mengoptimalkan efisiensi algoritma C4.5. Tujuan dari penelitian ini adalah untuk meningkatkan akurasi dan presisi dalam meramalkan prestasi akademik siswa. Hal ini akan menjadi landasan bagi pengembangan strategi pembelajaran yang sangat fleksibel dan mampu beradaptasi dengan cepat terhadap perubahan keadaan. Penelitian bertujuan untuk mengisi kesenjangan pengetahuan, yaitu dengan mengintegrasikan faktor sosial dan lingkungan ke dalam model yang digunakan untuk memprediksi kinerja akademik.

Metodologi penelitian melibatkan penerapan teknik analisis data dan pemanfaatan algoritma C4.5 yang ditingkatkan. Pendekatan tersebut melibatkan penggunaan data prestasi belajar siswa sebagai dataset untuk melatih dan menguji model, yang mencakup serangkaian uji coba dan validasi. Analisis data melibatkan penerapan pendekatan statistik dan evaluasi kemandirian model. Tujuan dari proyek ini adalah untuk meningkatkan akurasi prediksi hasil belajar siswa dengan mengoptimalkan algoritma C4.5, guna mendapatkan pemahaman menyeluruh tentang cara kerjanya.

Capaian penelitian tersebut diharapkan dapat memberikan implikasi yang signifikan dalam memahami prediksi hasil belajar siswa di era digital. Prediksi implementasi algoritma C4.5 yang dioptimalkan diharapkan dapat memberikan kontribusi yang berharga bagi bidang Informatika, khususnya domain pendidikan. Antisipasi dampak positif dari penelitian ini siap untuk mempercepat kemajuan masa depan dalam penggunaan teknologi untuk memberikan pengalaman pembelajaran yang semakin inklusif, mudah beradaptasi, dan selaras dengan kebutuhan siswa di era digital.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Penambangan data adalah proses mengekstraksi dan mengidentifikasi pola informasi penting untuk disimpan dalam database. Penambangan data adalah serangkaian prosedur yang dirancang untuk mengungkap informasi tambahan yang saat ini tidak ada dalam database. Prosesnya memerlukan eksplorasi pola dan model data yang memiliki kemampuan untuk mengubah data yang belum diproses menjadi wawasan yang berharga. Hal ini dicapai dengan mengekstraksi dan menganalisis model yang berharga atau menarik dari database. Penambangan data merupakan bagian integral dari proses Penemuan Pengetahuan Dalam Basis Data (KDD), dimana tujuannya adalah untuk mengeksplorasi basis data untuk menemukan informasi berharga. KDD menekankan

pada pemanfaatan teknik integrasi dan prosedur eksplorasi ilmiah.

Proses menafsirkan dan memvisualisasikan pola yang muncul dari pengumpulan data KDD sangatlah rumit dan memerlukan identifikasi pola yang valid, relevan, dan dapat dipahami dalam data. Rangkaian tindakan ini terdiri dari fase terpisah dari pemurnian data dan integrasi data. Pembersihan data digunakan untuk menghapus data yang tidak kompatibel dan mengganggu dari beberapa database yang mungkin memiliki format atau platform berbeda, setelah dikonsolidasikan dalam gudang data. Seleksi dan transformasi Data disimpan di gudang data dan kemudian direduksi menggunakan berbagai pendekatan.

Transformasi data adalah langkah pertama dalam proses yang bertujuan untuk mengubah data yang diproses menjadi data keluaran. Di bawah ini adalah teknik untuk mengubah data. A. Data atribut standar dapat diakses dan diturunkan melalui cantering. B. Normalisasi adalah pendekatan sistematis untuk memisahkan data atribut dengan cara yang terstandarisasi. C. Penskalaan memerlukan penyesuaian data hingga mencapai ukuran atau skala yang telah ditentukan. Operasi ini memerlukan perolehan informasi yang ditargetkan (penambangan data) dan memanipulasinya dalam database menggunakan berbagai metode dalam pemrosesan penambangan data. Ada beberapa algoritma yang tersedia di bidang data mining, sehingga memerlukan pilihan salah satu yang cocok untuk tugas tertentu yang ada. Melakukan pencarian model dan mempresentasikan ilmu yang diperoleh. Bagian ini bertanggung jawab atas representasi pengetahuan, yang melibatkan memastikan bahwa model informasi yang diperoleh konsisten dengan realitas yang ada atau aturan yang ditetapkan. Tahap terakhir adalah Knowledge Discovery in Databases (KDD) yang berupaya mempraktekkan pengetahuan dengan memanfaatkan model yang mudah dipahami oleh peserta. Penambangan data memenuhi berbagai tujuan, meliputi:

- Model ini menawarkan deskripsi yang komprehensif dan memfasilitasi identifikasi kumpulan item dengan definisi yang tepat.
- Estimasi dan pengelompokan memiliki kesamaan, meskipun estimasi membutuhkan perhitungan yang lebih rumit dibandingkan pengelompokan.
- Pemisahan adalah teknik yang digunakan untuk membedakan pola atau menetapkan perbedaan di antara kumpulan item atau prinsip panduan untuk mengantisipasi kumpulan item yang tidak terkategori.
- Segregasi adalah tindakan memisahkan orang-orang yang mempunyai ciri-ciri yang sama untuk mengumpulkan data, mengawasi, atau mengamati mereka. Cluster mengacu pada tindakan mengelompokkan catatan.[3]

2.2. Decision Tree

Pendekatan pohon keputusan merupakan salah satu dari berbagai metode klasifikasi. Metode pohon keputusan adalah metodologi yang digunakan untuk mengubah sejumlah besar data menjadi pohon keputusan, yang mewakili kumpulan aturan. Memahami aturan dapat dengan mudah dicapai dengan memanfaatkan bahasa alami. Pohon keputusan adalah kerangka hierarki yang memungkinkan pembagian kumpulan data besar menjadi subkumpulan lebih kecil menggunakan serangkaian aturan keputusan. Setiap himpunan pembagi menghasilkan himpunan konsekuen yang anggota-anggotanya menunjukkan kemiripan satu sama lain.

Pohon keputusan sering kali menawarkan data secara terstruktur, disusun sebagai tabel dengan atribut dan catatan. Atribut adalah parameter yang berfungsi sebagai kriteria dalam konstruksi sebuah pohon. Saat mengambil keputusan untuk bermain tenis, penting untuk mempertimbangkan kondisi cuaca, yang mencakup unsur-unsur seperti angin dan suhu. Atribut target digunakan untuk memberikan hasil atau hasil yang diinginkan untuk setiap item data individual. Instance adalah nilai yang dialokasikan ke atribut. Misalnya atribut cuaca terdiri dari contoh seperti cerah, mendung, dan hujan. Pendekatan pohon keputusan memerlukan konversi data tabular menjadi model pohon, diikuti dengan transformasi model pohon menjadi seperangkat aturan, yang kemudian disederhanakan.[4]

2.3. Algoritma C4.5

Algoritma C4.5 adalah metode komputasi yang digunakan untuk menghasilkan pohon keputusan. Pohon keputusan adalah metode yang ampuh dan dikenal baik untuk mengklasifikasikan dan memprediksi. Algoritme pohon keputusan mengubah kumpulan data besar menjadi pohon keputusan yang mewujudkan aturan. Aturan dapat dengan mudah dipahami dengan bahasa yang sederhana dan lugas. Selanjutnya dapat digambarkan menggunakan bahasa database seperti Structured Query Language (SQL) untuk mengambil record tertentu tergantung pada kategori tertentu [5].

Algoritma yang umum digunakan untuk menghasilkan pohon keputusan terdiri dari ID3, C4.5, dan CART. Algoritma C4.5 merupakan turunan dari algoritma ID3. Prosesnya mencakup memasukkan data ke dalam struktur pohon hierarki, kemudian mengubah struktur pohon menjadi seperangkat aturan, dan terakhir menyederhanakan aturan. Algoritma C4.5 digunakan untuk membangun Pohon Keputusan dengan terlebih dahulu memilih variabel atau atribut sebagai akar, menghasilkan cabang untuk setiap nilai, membagi kasus ke dalam cabang-cabang tersebut, dan kemudian mengulangi proses

untuk setiap cabang secara berulang hingga semua kasus berada dalam a cabang memiliki kelas yang sama. Algoritma C4.5 terdiri dari beberapa langkah:

- a. Atur data ke dalam kategori tertentu.
- b. Identifikasi simpul akar yang akan dibuat di pohon keputusan. Akarnya diperoleh dengan menentukan Gain setiap karakteristik berdasarkan faktor-faktor yang diberikan.
- c. Selanjutnya atribut dengan nilai Gain tertinggi dipilih setelah diperoleh nilai Gain pada masing-masing variabel.
- d. Ulangi tahapan tersebut hingga semua rekaman telah ditetapkan ke kelas yang sama.[6]

Algoritma Pohon Keputusan C4.5, biasa disebut dengan Klasifikasi versi 4.5, merupakan versi perbaikan dari algoritma ID3. Algoritme C4.5 berbagi prinsip operasional penting dengan algoritma ID3 sebagai hasil dari penyempurnaan ini. Algoritme C4.5 mematuhi prosedur yang tepat untuk menghasilkan pohon keputusan.

- a. Pilih atribut yang akan ditetapkan sebagai root.
- b. Membangun cabang terpisah untuk setiap nilai.
- c. Pisahkan instance dalam beberapa bagian
- d. Ulangi prosedur untuk setiap cabang hingga semua instance di cabang tersebut memiliki klasifikasi yang sama.

Metode Pohon Keputusan C4.5 menggunakan Rasio Perolehan sebagai kriteria pemisahan yang dimodifikasi selama proses pemilihan atribut pemisahan. Karakteristik terpisah adalah langkah utama dalam membangun pohon keputusan menggunakan algoritma C4.5. Algoritma C4.5 terdiri dari tahapan sebagai berikut:

- a. Hitung nilai entropi.
- b. Tentukan nilai gain rasio untuk masing-masing atribut.
- c. Atribut dengan Gain Ratio tertinggi dipilih sebagai root, sedangkan atribut dengan Gain Ratio lebih rendah dari root dipilih sebagai cabang.
- d. Hitung ulang Gain Ratio untuk setiap atribut, tidak termasuk atribut yang dipilih sebagai root pada langkah sebelumnya.
- e. Pilih atribut dengan Gain Ratio tertinggi sebagai cabang.
- f. Ulangi langkah 4 dan 5 hingga nilai Gain yang dihasilkan sama dengan nol untuk semua kualitas yang tersisa.

Untuk menghitung nilai Entropy dapat dihitung dengan persamaan :

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

S = himpunan kasus

A = fitur

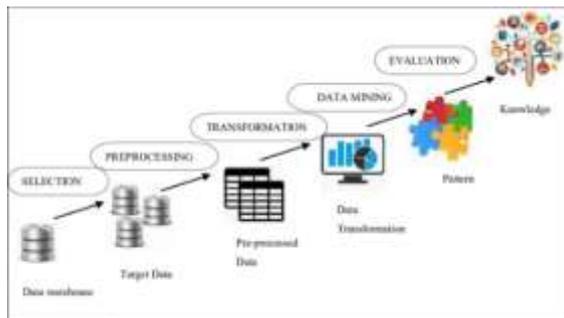
n = jumlah partisi

S = proporsi dari terhadap S [7]

3. METODE PENELITIAN

Penelitian ini menggunakan teknik kuantitatif, yang diakui secara luas di bidang penelitian karena efektivitasnya dalam mengumpulkan dan menafsirkan data numerik. Pendekatan ini memungkinkan peneliti untuk mengukur dan menentukan hubungan antara berbagai komponen dengan menggunakan metodologi statistik. Sebagaimana yang dijelaskan [8] Penelitian kuantitatif berpusat pada pemeriksaan dan evaluasi hubungan antar variabel untuk menguji teori tertentu.

Dalam lingkup penelitian ini, variabel-variabel terkait yang berkaitan dengan prestasi akademik siswa di MTS Yamuallim Panongan dinilai dan diuji dengan menggunakan instrumen penelitian. Peneliti bertujuan untuk meningkatkan akurasi dalam memprediksi prestasi belajar siswa dengan memanfaatkan algoritma C4.5 yang dioptimalkan dan menganalisis data kuantitatif yang dikumpulkan. Tindakan selanjutnya yang harus dilakukan dalam tugas akhir ini adalah sebagai berikut. Prosedurnya digambarkan secara visual pada gambar berikutnya.



Gambar 1. Alur penelitian KDD

3.1. Data Selection

Bagian ini mengimplementasikan pemilihan data, yang melibatkan pemilihan dan kombinasi banyak kumpulan data untuk digunakan dalam proses prediksi menggunakan alat RapidMiner. Tujuan dari tahap pemilihan data ini adalah untuk mengidentifikasi atribut-atribut yang diyakini berdampak pada pemilihan penerima Prediksi Prestasi Siswa yang dinilai. Tujuan dari metode ini adalah untuk meningkatkan efisiensi algoritma C4.5 dalam meramalkan kinerja akademik siswa kelas VII sampai IX di MTs Yamu'allim Panongan. Hal ini dicapai dengan memperhatikan beberapa ciri seperti kemampuan belajar, kemampuan membaca, interpretasi, kemampuan berbicara, dan kemampuan bertindak.

Selanjutnya, data kategorikal akan diproses, dan jika dianggap perlu, tindakan akan diambil untuk mengatasi ketidakseimbangan data dan melakukan pembersihan data awal untuk menyelesaikan masalah seperti data yang hilang atau hasil yang tidak wajar. Tujuan dari semua proses ini adalah untuk memastikan bahwa data

yang disiapkan optimal, sehingga memungkinkan algoritma C4.5 menghasilkan prediksi pencapaian belajar siswa yang lebih tepat dengan mempertimbangkan parameter penting ini. Di bawah ini adalah tabel seleksi wajib :

No. No.	NAME	ALAM	JENIS KEL.	BEKAS	KEHAMIL	KEPERKAW.	ALAM	KEPERKAW.	BERKAW.	BERKAW.	TANG.	PEKER.
1	INDAH P.	0	PERempuan	4	0	0	0	0	0	0	11	Siswa
2	TRISNATI.	0	Laki-laki	4	4	0	0	0	0	0	23	Tidak
3	HERMANS.	0	Laki-laki	0	0	0	0	0	0	0	0	Siswa
4	SITI RAH.	0	Laki-laki	0	0	0	0	0	0	0	0	Siswa
5	LALUPTI.	0	Perempuan	0	0	0	0	0	0	0	0	Siswa
6	DEWANTI.	0	Perempuan	0	0	0	0	0	0	0	0	Tidak
7	ADIANI.	0	Perempuan	0	0	0	0	0	0	0	0	Siswa
8	HELITATI.	0	Perempuan	0	0	0	0	0	0	0	0	Siswa
9	LADY S.	0	Perempuan	0	0	0	0	0	0	0	0	Siswa
10	WANI S.	0	Perempuan	4	4	4	4	4	4	4	24	Tidak
11	SARANI.	0	Laki-laki	0	0	0	0	0	0	0	0	Siswa
12	HERZEL.	0	Laki-laki	0	0	0	0	0	0	0	0	Siswa
13	HERNATI.	0	Laki-laki	0	0	0	0	0	0	0	0	Siswa
14	HERNANI.	0	Laki-laki	0	0	0	0	0	0	0	0	Tidak
15	HERNANI.	0	Perempuan	0	0	0	0	0	0	0	0	Siswa

Gambar 2. Data Selection

3.2. Data Pre-processing

Pada tahap ini, data yang telah diimpor akan melalui serangkaian proses pengolahan. Langkah-langkah melibatkan seleksi fitur untuk memilih variabel-variabel yang relevan seperti belajar, membaca, interpretasi, bicara, bertindak, dan menghafal. Selain itu, nilai yang hilang akan ditangani menggunakan metode yang sesuai, dan data kategorikal dapat diubah ke dalam format numerik. Proses selanjutnya mencakup pembagian data menjadi dua bagian, yakni untuk pelatihan dan pengujian. Jika ada atribut kategorikal, langkah encoding akan diaplikasikan untuk mengonversi data menjadi format yang dapat diolah oleh algoritma.

Selain itu, penanganan ketidakseimbangan data dan pembersihan lanjutan juga dapat diterapkan sesuai kebutuhan. Hasil dari tahapan ini adalah data yang sudah diproses dan siap digunakan dalam proses selanjutnya, seperti pemodelan menggunakan algoritma C4.5 di RapidMiner. Untuk lebih jelas terdapat pada table dibawah ini :

NAME	ALAM	JENIS KEL.	KEHAMIL	KEPERKAW.	ALAM	KEPERKAW.	TANG.	PEKER.	SEKOL.
INDAH P.	0	0	0	0	0	0	0	Siswa	0
TRISNATI.	4	4	0	0	0	0	0	Siswa	0
HERMANS.	0	0	0	0	0	0	0	Siswa	0
SITI RAH.	0	0	0	0	0	0	0	Siswa	0
LALUPTI.	0	0	0	0	0	0	0	Tidak	0
DEWANTI.	0	0	0	0	0	0	0	Siswa	0
ADIANI.	0	0	0	0	0	0	0	Siswa	0
HELITATI.	0	0	0	0	0	0	0	Siswa	0
LADY S.	0	0	0	0	0	0	0	Siswa	0
WANI S.	4	4	4	4	4	4	4	Tidak	0
SARANI.	0	0	0	0	0	0	0	Tidak	0
HERZEL.	0	0	0	0	0	0	0	Siswa	0
HERNATI.	0	0	0	0	0	0	0	Siswa	0
HERNANI.	0	0	0	0	0	0	0	Tidak	0
HERNANI.	0	0	0	0	0	0	0	Siswa	0
SITI RAH.	0	0	0	0	0	0	0	Siswa	0
HERNANI.	0	0	0	0	0	0	0	Siswa	0

Gambar 3. Data Pre-processing

3.3. Data Transformation

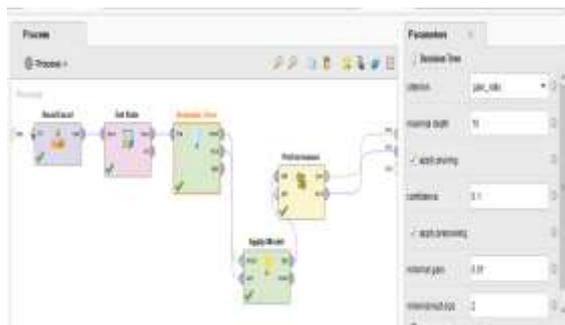
Transformasi data mengacu pada proses mengubah susunan atau karakteristik data asli agar lebih sesuai atau lebih mudah dipahami untuk metode pemodelan atau analisis data. Transformasi data berupaya mengoptimalkan kualitas data, mengurangi gangguan, dan menyelaraskannya

dengan asumsi tertentu yang diperlukan agar teknik analisis atau pemodelan dapat digunakan.[9]

Selama tahap transformasi data, data impor yang telah dipilih untuk dianalisis lebih lanjut akan melalui serangkaian prosedur untuk meningkatkan kualitasnya dan memfasilitasi interpretasinya. Untuk memulai, penting untuk mengubah variabel kategori ke dalam format numerik untuk memungkinkan pemrosesan dengan metode pemodelan. Selain itu, sangat penting untuk memanfaatkan operator pengganti nilai yang hilang untuk mengelola data yang tidak lengkap. Selanjutnya dapat dilakukan prosedur normalisasi atau standarisasi untuk menjamin bahwa data mempunyai skala yang konsisten, sehingga mencegah kelaziman variabel tertentu dalam analisis. Prosedur ini memfasilitasi penyesuaian struktur data untuk menghasilkan hasil yang lebih efisien ketika digunakan pada tahapan berikut, seperti pemodelan dengan algoritma C4.5 di RapidMiner.[10]

3.4. Data Mining

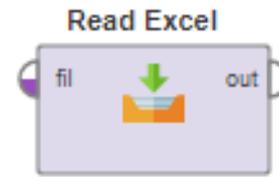
Data Mining adalah prosedur sistematis yang mengekstraksi pengetahuan dan informasi baru dari sejumlah besar data yang disimpan di gudang data, menggunakan teknik statistik dan matematika dari kecerdasan buatan. Inti dari penelitian ini terletak pada eksplorasi pola atau informasi yang menarik melalui penggunaan prosedur, metode, atau algoritma yang telah ditetapkan. Memprediksi prestasi belajar siswa di Mts Yamuallim Panongan menggunakan Algoritma Decision Tree. Urutan proses data mining melibatkan penggunaan metode Decision Tree:



Gambar 1. Proses Rapid mainer

3.5. Read Excel

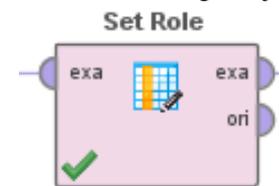
Buka dan akses file Excel. Manfaatkan operasi ini untuk mengambil data dari Microsoft Excel. Pengguna harus menentukan lembar kerja yang ingin diperoleh sebagai tabel data. Tabel harus disusun sedemikian rupa sehingga setiap baris berhubungan dengan sebuah instance, dan setiap kolom berhubungan dengan karakteristik yang berbeda. Fungsi awal “Baca Excel” meliputi kumpulan data instruksional, yaitu data kinerja siswa di Mts Yamuallim Panongan.



Gambar 6. Read Excel

3.6. Set Role

Operator "Set Role" bertanggung jawab untuk menetapkan fungsi atau peran suatu atribut dalam suatu dataset. Atribut di dalam kumpulan data dapat memenuhi banyak tugas, seperti berfungsi sebagai label target untuk prediksi, atribut masukan, atau atribut lain dengan tujuan berbeda.



Gambar 7. Set Role

Tabel 1. Parameter dan Atribut operator Set Role

No	Parameter	Isi
1.	Nama	id
2.	Prestasi	Label

Tabel di atas memiliki atribut atau parameter bernama “Id” dengan nilai atau isi “Id”, yang berfungsi sebagai pengenalan atau nomor identifikasi unik dalam suatu sistem atau dataset. Contoh parameter ID yang digunakan dalam penelitian ini adalah ID identitas siswa, yang berfungsi untuk mengidentifikasi secara unik setiap siswa dalam sistem sekolah. Baris kedua pada tabel di atas, berlabel "i", menawarkan penjelasan yang lebih komprehensif tentang atribut "Pencapaian" dan nilai "Label" yang terkait. Hal ini menunjukkan bahwa parameter "Prestasi" digunakan untuk memberi label atau kategori pada data dalam penelitian ini.

3.7. Decision Tree

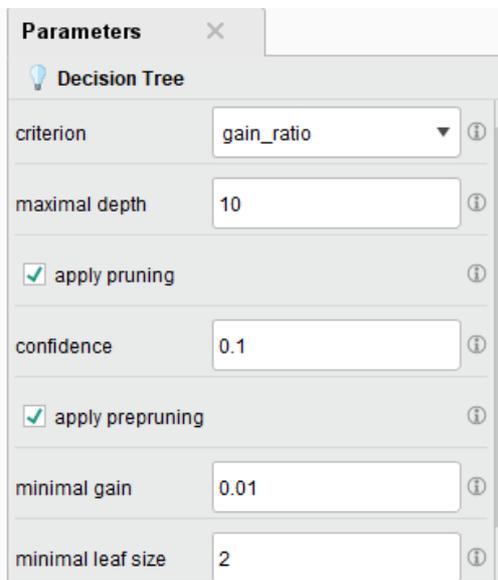
Konfigurasi Operator Decision Tree Setelah mengakses operator "Decision Tree", ubah konfigurasi operator agar selaras dengan kebutuhan analisis data studi. Identifikasi kualitas target (label) dan atribut prediktor yang diinginkan sesuai kebutuhan.



Gambar 8. Decision Tree

Penggabungan aturan parameter kriteria, khususnya Gain Ratio, dalam Decision Tree dapat berdampak besar pada struktur Decision Tree dan keunggulan model yang dihasilkan secara keseluruhan. Rasio Keuntungan adalah metrik yang digunakan untuk mengukur tingkat ketidakmurnian data pada node tertentu selama pembuatan Decision Tree. Ini memfasilitasi proses pemilihan kualitas. Algoritma Gain Ratio menghasilkan pohon keputusan yang lebih jelas dan mudah dipahami dengan mengurangi proses pemilihan atribut. Untuk mengoptimalkan "kedalaman maksimal", seseorang harus secara hati-hati menyeimbangkan kemampuan model untuk menemukan pola yang signifikan sambil menghindari kompleksitas yang berlebihan.

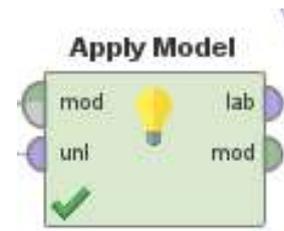
Menggunakan berbagai nilai dalam uji coba dan mengevaluasi kinerja model adalah metode yang efektif untuk memastikan parameter optimal. Pilihan ambang batas kepercayaan pada Pohon Keputusan, seperti nilai presisi 0,1, memiliki pengaruh besar baik terhadap keputusan yang dibuat oleh model maupun arsitektur Decision Tree yang dibangun secara keseluruhan. Nilai keyakinan berkorelasi erat dengan ambang batas keyakinan yang diperlukan agar suatu keputusan dianggap sah.



Gambar 9. Parameters Decision Tree

3.8. Apply Model

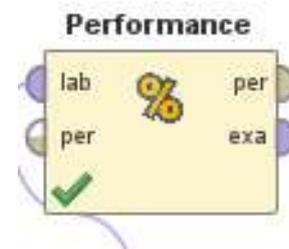
Pendekatan ini melibatkan penggunaan model ahli untuk menghasilkan prediksi atau mengklasifikasikan data yang tidak dikenal. Gunakan fungsi "Apply Model" untuk menghasilkan prediksi untuk setiap kumpulan data baru. Proses ini melibatkan navigasi melalui pohon keputusan dengan mengevaluasi karakteristik data untuk mencapai node terminal yang sesuai.



Gambar 10. Apply Model

3.9. Performance

Fungsi "performance" mengkuantifikasi kinerja model Decision Tree dalam penelitian ini dengan mengevaluasi metrik akurasi dan kesalahan klasifikasi..



Gambar 11. Performance

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan sumber data sekunder yang diperoleh melalui kesepakatan formal antara penulis dan masyarakat. Data yang diperoleh diperoleh dari komunitas yang secara aktif memberikan persetujuannya untuk membagikan informasinya. Data yang digunakan termasuk dalam kategori data sekunder dan terdiri dari kumpulan 410 sampel data, terutama terkait Data Prestasi Belajar Siswa. Data di atas mencakup prestasi akademik siswa MTs Yamu'allim Panongan dalam bidang akademik terkini. Ini mencakup serangkaian evaluasi, termasuk nilai ujian, tugas, dan penilaian lainnya, yang berfungsi sebagai indikasi pencapaian akademik.

Fase analisis yang sedang berlangsung melibatkan tabel akurasi yang menyajikan hasil berikut: 258 Prakiraan sedang dengan presisi kelas 100%, 85 Prediksi rendah dengan presisi kelas 100%, dan 67 Prediksi tinggi dengan presisi kelas 100%.

	Hasil Sengah	Hasil Tinggi	Hasil Rendah	Akurasi persentase
pred Sengah	258	0	0	100.00%
pred Tinggi	0	67	0	100.00%
pred Rendah	0	0	85	100.00%
Overall	100.00%	100.00%	100.00%	

Gambar 12. Gambar Tabel Accuracy pada nilai indeks

Bagian ini memperkenalkan operator pohon keputusan yang mengekstrak nilai dari informasi

pelabelan yang diberikan sebelumnya, seperti yang ditunjukkan pada gambar di atas. Jika jumlah seluruh skor kurang dari atau sama dengan 12,5, maka skor pembelajaran yang diantisipasi dianggap rendah. Jika skor kumulatif keseluruhan melebihi 12,5, ada dua kemungkinan hasil. Jika jumlah seluruh skor di bawah 18,5, maka skor pembelajaran yang diantisipasi dianggap sedang. Skor kumulatif sebesar 18,5 menunjukkan pencapaian yang sangat dinantikan.



Gambar 13. Pohon Keputusan

Pada bagian ini, kami menyajikan pendekatan baru terhadap pelabelan yang menggabungkan nilai confidence. Nilai ini berfungsi untuk mengukur tingkat kepastian yang kita miliki mengenai keakuratan aturan asosiasi atau prediksi model.

The screenshot shows a data table with the following columns: Date, Status, Age, Gender, Marital Status, Education, Income, Spending Score, and Loyalty Points. The table contains multiple rows of data, with some cells highlighted in yellow and green.

Gambar 14. Nilai Confidence Interval

Dengan menggunakan prosedur penilaian ini, Anda dapat menilai keefektifan model Pohon Keputusan atau C4.5 yang telah Anda buat dalam analisis data Anda. Pastikan konfigurasi dan pengaturan evaluasi disesuaikan dengan kebutuhan dan tujuan individu. Dalam bidang data mining, pengetahuan berkaitan dengan pemahaman yang diperoleh melalui proses menganalisis dan menyelidiki data. Penambangan data adalah teknik sistematis yang berupaya mengidentifikasi pola, tren, atau wawasan berharga dari kumpulan data yang rumit. Pemahaman ini selanjutnya dapat digunakan untuk meningkatkan pengambilan keputusan, memahami hubungan antar faktor, dan mendukung pemahaman yang lebih mendalam tentang suatu peristiwa atau proses.

Misalnya, dalam upaya penambangan data yang menggunakan algoritme C4.5 untuk memperkirakan kinerja siswa, pengetahuan yang diperoleh dapat mencakup korelasi antara faktor-faktor seperti durasi belajar, ketidakhadiran,

bantuan orang tua, dan prestasi siswa. Investigasi seperti ini dapat menghasilkan wawasan berharga mengenai elemen-elemen utama yang mempengaruhi pencapaian pembelajaran, sehingga membantu pengambilan keputusan dalam bidang pendidikan. Kemahiran dalam kerangka penambangan data juga mencakup kemampuan untuk menganalisis hasil model dan mengevaluasi efektivitasnya. Penggunaan algoritma C4.5 untuk menghasilkan pohon keputusan menghasilkan pengetahuan tentang bagaimana pohon memprediksi hasil dengan memanfaatkan aturan yang diturunkan dari data.

Penting untuk diketahui bahwa pengetahuan yang diperoleh melalui data mining bersifat dinamis dan terus berkembang seiring dengan semakin banyaknya data yang dikumpulkan dan pemahaman yang lebih dalam mengenai area permasalahan telah tercapai. Oleh karena itu, tahap awal dalam memperoleh pengetahuan penting adalah proses penambangan data, namun praktik pemantauan, evaluasi, dan revisi model yang berkelanjutan sangat penting untuk menjaga ketepatan waktu dan pentingnya pengetahuan dalam jangka panjang.

5. KESIMPULAN DAN SARAN

kemampuan untuk meningkatkan ketepatan peramalan kinerja akademik siswa di MTs Yamu'allim Panongan dengan meningkatkan algoritma C4.5. Pemilihan yang cermat dan penyesuaian parameter yang tepat dapat meningkatkan ketepatan model prediksi, sehingga memberikan kontribusi yang besar terhadap pemantauan dan evaluasi kinerja siswa.

Integrasi faktor sosial sebagai elemen penting. Temuan penelitian menyoroti perlunya mengintegrasikan faktor sosial ke dalam model prediksi. Penelitian ini meningkatkan pemahaman tentang faktor-faktor yang mempengaruhi prestasi siswa dengan mengintegrasikan variabel sosial, seperti lingkungan keluarga dan dukungan sosial. Fitur ini menawarkan sudut pandang yang lebih luas, meningkatkan pengetahuan yang diperoleh melalui model prediksi.

Pentingnya Penerapan Temuan Penelitian untuk Meningkatkan Pembelajaran. Penerapan temuan penelitian ini diharapkan memberikan efek positif pada metode pengajaran. Guru dan staf sekolah dapat memanfaatkan data prediktif untuk mengembangkan strategi pembelajaran terfokus, mendorong kemajuan individu siswa, dan mengoptimalkan efektivitas proses pembelajaran.

Disarankan untuk berkolaborasi dengan ahli statistik atau profesional ilmu data selama fase pengoptimalan algoritme. Dengan memilih dan menyempurnakan parameter secara cermat, model prediksi dapat dibangun berdasarkan basis statistik yang kuat dan andal. Pemeriksaan Menyeluruh Terhadap Faktor-Faktor Sosial Lakukan lebih

banyak survei dengan fokus yang lebih besar pada faktor-faktor sosial yang mempengaruhi prestasi akademik siswa. Dengan memperoleh pemahaman yang mendalam, model prediktif dapat ditingkatkan dengan memasukkan data yang lebih akurat, sehingga meningkatkan akurasi dan relevansinya.

Pelatihan Pemanfaatan Hasil Peramalan Memberikan panduan kepada pendidik dan staf sekolah dalam menggabungkan hasil prediksi dengan benar ke dalam tugas profesional rutin mereka. Meningkatkan kapasitas mereka untuk menilai dan melaksanakan temuan penelitian

DAFTAR PUSTAKA

- [1] P. Ekayani, "PENTINGNYA PENGGUNAAN MEDIA PEMBELAJARAN UNTUK MENINGKATKAN PRESTASI BELAJAR SISWA." [Online]. Available: <https://www.researchgate.net/publication/315105651>
- [2] A. Sofalul Khazari, F. Marisa, and I. Dharma Wijaya, "SISTEM REKOMENDASI PENENTUAN JUDUL SKRIPSI MENGGUNAKAN ALGORITMA DECISION TREE."
- [3] I. Zulfa, R. Rayuwati, and K. Koko, "Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode apriori," *Teknika: Jurnal Sains dan Teknologi*, vol. 16, no. 1, p. 69, Jun. 2020, doi: 10.36055/tjst.v16i1.7601.
- [4] P. Bidang Komputer Sains dan Pendidikan Informatika, D. Akademi Perekam dan Informasi Kesehatan Iris Padang Jl Gajah Mada No, and S. Barat, "Jurnal Edik Informatika Data Mining : Klasifikasi Menggunakan Algoritma C4.5 Yuli Mardi".
- [5] "jurnal juna eska".
- [6] V. S. Ginting, K. Kusri, and E. Taufiq, "Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python," *Inspiration: Jurnal Teknologi Informasi dan Komunikasi*, vol. 10, no. 1, Jun. 2020, doi: 10.35585/inspir.v10i1.2535.
- [7]] Jurusan, S. Komputer, F. Mipa, U. Tanjungpura, J. Prof, and H. H. Nawawi, "IMPLEMENTASI DATA MINING DENGAN ALGORITMA C4.5 UNTUK PENJURUSAN SISWA (STUDI KASUS: SMA NEGERI 1 PONTIANAK) [1] Beti Novianti, [2] Tedy Rismawan [3] Syamsul Bahri," 2016.
- [8] F. Faisal Nugraha, I. Sunandar, and C. Juliane, "Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 9, no. 4, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [9] A. Novandya and I. Oktria, "Penerapan Algoritma Klasifikasi Data Mining C4.5 Pada Dataset Cuaca Wilayah Bekasi," 2017.
- [10] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Faktor Exacta*, vol. 11, no. 3, Oct. 2018, doi: 10.30998/faktorexacta.v11i3.2777.