

Teknik Data Mining Dengan Menggunakan Algoritma Decision Tree Untuk Mengetahui Pola Pemahaman Mahasiswa Pada Matakuliah Pemrograman

Sri Novida Sari^{1,*}, Putri Annisa², An Nisa Dian Rahma³, Rama Prameswara Ritonga², Dito Putro Utomo²

¹Program Studi Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan, Medan, Indonesia

²Program Studi Teknologi Rekayasa Multimedia Grafis, Politeknik Negeri Medan, Medan, Indonesia

³Jurusan Administrasi Niaga, Program Studi MICE, Politeknik Negeri Medan, Medan, Indonesia

Email: ¹,*srinovida@polmed.ac.id, ²putriannisa@polmed.ac.id, ³annisadianrahma@polmed.ac.id,

⁴ramaprameswararitonga@polmed.ac.id, ⁵ditoputrutomo@polmed.ac.id

(* : coressponding author)

Abstrak-Politeknik Negeri Medan sebagai salah satu perguruan tinggi vokasi terkemuka di Kota Medan memiliki peran penting dalam mencetak lulusan yang siap kerja dan memiliki kompetensi terapan sesuai kebutuhan industri. Salah satu jurusan yang berperan strategis adalah Jurusan Teknik Komputer dan Informatika, yang berfokus pada pengembangan kemampuan mahasiswa dalam bidang teknologi dan pemrograman. Mata kuliah pemrograman menjadi dasar penting dalam pembentukan kemampuan analisis dan logika mahasiswa. Namun, pada kenyataannya masih banyak mahasiswa yang mengalami kesulitan dalam memahami konsep-konsep dasar pemrograman, yang berdampak pada rendahnya capaian akademik dan motivasi belajar. Penelitian ini bertujuan untuk mengidentifikasi pola pemahaman mahasiswa dalam mata kuliah pemrograman menggunakan algoritma Decision Tree sebagai metode klasifikasi. Melalui pendekatan data mining, penelitian ini berupaya mengekstraksi pola-pola tersembunyi dari data akademik mahasiswa untuk mengetahui faktor-faktor yang memengaruhi tingkat pemahaman mereka. Algoritma Decision Tree dipilih karena mampu menghasilkan model klasifikasi yang mudah dipahami dan diinterpretasikan, serta efektif dalam menangani data kategorikal maupun numerik. Data penelitian diolah menggunakan Google Colaboratory dengan bantuan pustaka scikit-learn. Proses pengujian dilakukan melalui pembentukan model klasifikasi, visualisasi pohon keputusan, serta analisis confusion matrix untuk mengukur performa model. Berdasarkan hasil pengujian, diperoleh nilai akurasi sebesar 50% dan F1-score sebesar 51,68%, yang menunjukkan bahwa model Decision Tree memiliki kemampuan yang baik dalam memprediksi dan mengklasifikasikan tingkat pemahaman mahasiswa terhadap mata kuliah pemrograman. Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam pengembangan strategi pembelajaran berbasis data di lingkungan pendidikan vokasi. Melalui hasil yang diperoleh, dosen diharapkan dapat menyesuaikan metode pengajaran sesuai dengan karakteristik dan kemampuan mahasiswa, sehingga proses pembelajaran menjadi lebih adaptif, efektif, dan berdampak positif terhadap peningkatan pemahaman mahasiswa pada mata kuliah pemrograman.

Kata Kunci: Data Mining; Pola; Pemahaman Mahasiswa; Algoritma Decision Tree; Matakuliah Pemrograman

Abstract-Medan State Polytechnic, as one of the leading vocational universities in Medan City, plays a crucial role in producing graduates who are ready to work and possess applied competencies according to industry needs. One of the strategic departments is the Computer Engineering and Informatics Department, which focuses on developing students' abilities in technology and programming. Programming courses are an important foundation in developing students' analytical and logical skills. However, many students still experience difficulties in understanding basic programming concepts, which results in low academic achievement and learning motivation. This study aims to identify patterns of student understanding in programming courses using the Decision Tree algorithm as a classification method. Through a data mining approach, this study attempts to extract hidden patterns from students' academic data to identify factors that influence their level of understanding. The Decision Tree algorithm was chosen because it is able to produce classification models that are easy to understand and interpret, and is effective in handling both categorical and numerical data. The research data was processed using Google Colaboratory with the help of the scikit-learn library. The testing process was carried out through the formation of a classification model, decision tree visualization, and confusion matrix analysis to measure model performance. Based on the test results, an accuracy value of 50% and an F1-score of 51.68% were obtained, indicating that the Decision Tree model has a good ability to predict and classify students' level of understanding of programming courses. Overall, this research provides an important contribution to the development of data-based learning strategies in vocational education environments. Through the results obtained, lecturers are expected to be able to adjust teaching methods according to student characteristics and abilities, so that the learning process becomes more adaptive, effective, and has a positive impact on improving student understanding of programming courses.

Keywords: Data Mining; Patterns; Student Understanding; Decision Tree Algorithm; Programming Course

1. PENDAHULUAN

Politeknik Negeri Medan merupakan salah satu perguruan tinggi negeri vokasi terkemuka di Kota Medan yang berfokus pada pengembangan keterampilan praktis mahasiswa. Sebagai institusi pendidikan vokasi, Politeknik Negeri Medan memiliki tanggung jawab untuk menghasilkan lulusan yang siap kerja dan memiliki kompetensi terapan yang kuat sesuai dunia kerja. Sebagai lembaga vokasi, Politeknik Negeri Medan mengedepankan pendekatan pembelajaran berbasis praktik dibandingkan dengan pembelajaran teori. Model pembelajaran ini bertujuan untuk memperkuat pemahaman konseptual mahasiswa melalui penerapan langsung di laboratorium atau dalam bentuk proyek-proyek terapan yang relevan dengan kebutuhan di dunia kerja.

Jurusan Teknik Komputer dan Informatika merupakan salah satu jurusan di Politeknik Negeri Medan yang berperan dalam mencetak tenaga profesional di bidang penerapan teknologi. Jurusan ini memiliki berbagai mata kuliah inti yang

mendukung kemampuan mahasiswa, termasuk mata kuliah pemrograman sebagai dasar utama dalam meningkatkan cara berfikir dan proses analisis terhadap permasalahan yang terjadi.

Mata kuliah pemrograman menjadi fondasi penting bagi mahasiswa di Jurusan Teknik Komputer dan Informatika. Penguasaan terhadap pemrograman menjadi indikator utama untuk menilai kesiapan mahasiswa dalam mengikuti perkuliahan serta kemampuan ketika mahasiswa menyelesaikan perkuliahan. Namun, kenyataannya masih banyak mahasiswa yang mengalami kesulitan dalam memahami dan mengikuti proses pembelajaran pada matakuliah pemrograman. Dimana hal ini dapat dilihat dari nilai akhir matakuliah yang didapatkan oleh mahasiswa ataupun terkendalanya mahasiswa dalam menyelesaikan tugas, ujian ataupun proyek yang diberikan kepada mahasiswa.

Permasalahan yang sering muncul di lapangan adalah minimnya kemampuan dan pemahaman mahasiswa dalam menyelesaikan tugas-tugas pemrograman. Hal ini dapat disebabkan oleh berbagai faktor, seperti perbedaan gaya belajar, ataupun maupun keterbatasan waktu untuk praktik mandiri. Faktor lainnya bisa dilihat dari bagaimana dosen memberikan, mengarahkan dan menyampaikan materi didalam perkuliahan. Baik perkuliahan yang dilakukan dengan praktik langsung didalam kelas ataupun menggunakan video tutorial ataupun model pembelajaran lainnya.

Dampak dari permasalahan ini cukup signifikan, antara lain rendahnya capaian nilai mahasiswa dalam mata kuliah pemrograman dan menurunnya motivasi belajar. Jika tidak diatasi, hal ini dapat memengaruhi kualitas lulusan serta reputasi program studi secara keseluruhan. Sehingga permasalahan tersebut harus diselesaikan agar kiranya terpenuhi capaian lulusan dari setiap mahasiswa ketika menyelesaikan perkuliahan nantinya.

Salah satu pendekatan yang dapat dilakukan untuk menyelesaikan permasalahan ini adalah dengan memahami pola pemahaman mahasiswa atau pola pengajaran yang diterapkan dosen di kelas. Dengan mengetahui pola-pola tersebut, institusi pendidikan dapat merancang pembelajaran yang lebih efektif dan adaptif sesuai kebutuhan mahasiswa[1].

Data mining merupakan salah satu pendekatan dalam ilmu komputer yang dapat digunakan untuk mengekstraksi pola-pola tersembunyi dari sekumpulan data akademik mahasiswa. Dengan memanfaatkan teknik ini, dapat menemukan hubungan antara variabel - variabel dengan tingkat pemahaman mahasiswa terhadap materi matakuliah pemrograman[2]. Di antara berbagai metode dalam data mining, algoritma Decision Tree menjadi pilihan yang tepat karena mampu menyajikan model klasifikasi dalam bentuk struktur pohon yang mudah dipahami dan diinterpretasikan. Algoritma ini juga efektif dalam menangani data kategorikal dan numerik serta memberikan hasil klasifikasi yang jelas dan logis[3]. Secara teknis, algoritma Decision Tree bekerja dengan memecah dataset berdasarkan atribut-atribut terbaik yang memberikan informasi paling besar, hingga membentuk pohon keputusan[4]. Setiap cabang dalam pohon menggambarkan kondisi atau keputusan, dan setiap daun mewakili hasil klasifikasi akhir. Model ini sangat sesuai digunakan untuk memetakan pola pemahaman mahasiswa berdasarkan data akademik yang tersedia. Proses penyelesaian pada Decision Tree berdasarkan dengan nilai Entropy dan Gain. Nilai Entropy dan Gain nantinya dipergunakan untuk pembentukan struktur pada pohon keputusan[5].

Dalam mendukung proses penelitian, maka dibutuhkan beberapa sumber sebagai referensi atau masukan pada penelitian seperti yang dilakukan oleh [6] dimana pada penelitian yang dilakukan berfokus pada klasifikasi tingkat kesejahteraan keluarga. Hasil pada penelitian bahwasannya algoritma Decision Tree efektif dalam membentuk pola untuk melihat tingkat kesejahteraan pada keluarga

Selain itu, juga dilakukan penelitian oleh [7] penelitian yang dilakukan berfokus untuk memprediksi mahasiswa Drop Out. Dari proses yang telah dilakukan pada penelitian hasil yang diperoleh bahwasannya Algoritma Decision Tree teruji efektif dalam melakukan prediksi dimana didapatkan tingkat akurasi sebesar 81,82%.

Berikutnya penelitian juga dilakukan oleh [8] dimana penelitian yang dilakukan terfokus pada memprediksi terhadap persediaan obat di apotek. Proses penelitian dengan algoritma decision tree dilakukan pengujian dengan menggunakan rapidminer dengan hasil dari proses bahwasannya tingkat akurasi sebesar 80%. Hal ini menandakan bahwasannya algoritma decision tree efektif dalam membantu memprediksi ketersediaan obat diapotek.

Terakhir penelitian yang dilakukan oleh [9] dimana penelitian memiliki fokus pada prediksi tingkat kelulusan mahasiswa. Temuan yang didapatkan menegaskan bahwa model dapat memprediksi tingkat kelulusan mahasiswa dengan akurat dengan nilai akurasi sebesar 99,64%.

Penelitian ini memiliki perbedaan dengan penelitian terdahulu karena secara spesifik menyoroti pemahaman mahasiswa dalam mata kuliah pemrograman serta mengaitkannya dengan pola-pola pembelajaran dan faktor akademik lain yang relevan. Selain itu, penelitian ini bertujuan untuk menghasilkan model klasifikasi yang dapat digunakan sebagai alat bantu evaluasi dan pengambilan keputusan oleh dosen.

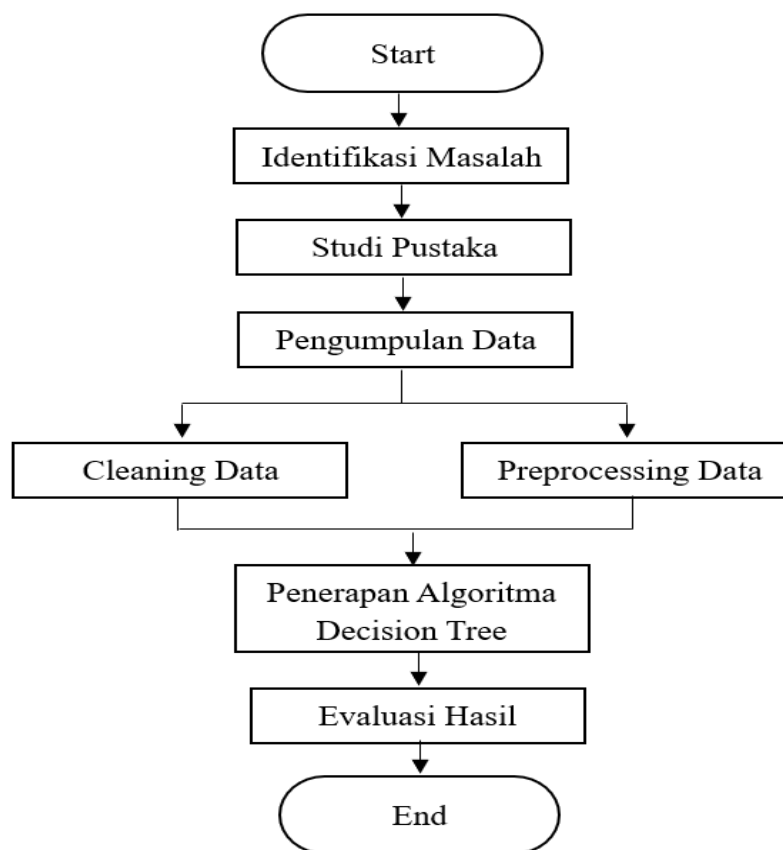
Adapun tujuan dari penelitian ini adalah untuk mengidentifikasi pola pemahaman mahasiswa dalam mata kuliah pemrograman dengan memanfaatkan algoritma Decision Tree, serta memberikan rekomendasi strategis bagi pengembangan proses pembelajaran yang lebih efektif.

Diharapkan dari hasil penelitian ini, dosen dapat memahami lebih baik karakteristik pembelajaran mahasiswa, sehingga strategi pengajaran dapat disesuaikan dengan kebutuhan dan potensi masing-masing individu. Pada akhirnya, penelitian ini bertujuan untuk meningkatkan kualitas pembelajaran dan mendorong mahasiswa agar lebih memahami dan menguasai mata kuliah pemrograman secara optimal.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan alur proses yang dilakukan pada penelitian, adapun tahapan penelitian dapat dilihat pada gambar berikut:



Gambar 1. Flowchart Tahapan Penelitian

Pada gambar 1 dapat dilihat flowchart atau alur proses yang dilakukan pada penelitian. Adapun penjelasan dari setiap proses dapat dilihat pada tabel berikut:

Tabel 1. Tahapan Penelitian

No	Tahapan	Keterangan
1	Identifikasi Masalah	Identifikasi masalah merupakan tahapan penemuan dan perumusan masalah yang terjadi pada penelitian sehingga layak untuk diangkat menjadi penelitian
2	Studi Pustaka	Studi pustaka merupakan tahapan mengulas kembali terhadap buku, artikel atau publikasi yang sejenis dengan penelitian
3	Pengumpulan Data	Pengumpulan data dilakukan dengan memberikan kuseioner kepada mahasiswa untuk melihat apakah mahasiswa memahami atau tidak
4	Cleaning Data	Cleaning data merupakan proses peninjauan kembali terhadap data. Dimana data yang tidak layak ataupun data duplikat pada penelitian akan dihapuskan
5	Preprocessing Data	Pada tahapan ini dilakukan penyesuaian terhadap data dengan algoritma yang akan digunakan pada proses penyelesaian
6	Penerapan Algoritma	Tahapan ini merupakan tahapan pengolahan data dengan menggunakan Algoritma Decision Tree untuk mendapatkan hasil berupa pola terhadap pemahaman mahasiswa kepada matakuliah Pemrograman
7	Evaluasi Hasil	Tahapan akhir evaluasi terhadap akhir dengan melakukan pengujian terhadap pola untuk melihat seberapa besar tingkat akurasi ataupun kepercayaan terhadap pola yang didapatkan

2.2 Data Mining

Data mining merupakan proses penting dalam mengekstraksi informasi yang berguna dan tersembunyi dari kumpulan data yang besar. Data mining adalah proses penemuan pola yang menarik dan pengetahuan dari data dalam jumlah besar. Proses ini melibatkan berbagai teknik statistik, matematika, dan kecerdasan buatan (AI) untuk menemukan hubungan, pola, atau tren yang tidak dapat terlihat secara eksplisit. Data mining biasanya diterapkan pada berbagai bidang seperti

kesehatan, pendidikan, keuangan, dan pemasaran, di mana pengambilan keputusan yang tepat berdasarkan data sangat dibutuhkan[10].

Dalam praktiknya, data mining terdiri dari beberapa tahapan utama, di antaranya adalah pengumpulan data, pembersihan data (data cleaning), transformasi data, pemodelan (modeling), evaluasi, dan implementasi hasil. Proses ini dikenal secara luas dengan istilah Knowledge Discovery in Database (KDD). Salah satu metode populer dalam data mining adalah klasifikasi, yaitu proses membagi data ke dalam kelompok berdasarkan atribut-atribut tertentu. Metode klasifikasi yang umum digunakan antara lain adalah Decision Tree, Naïve Bayes, dan K-Nearest Neighbor. Decision Tree menjadi pilihan yang banyak digunakan karena mampu menghasilkan model yang mudah dipahami dan diinterpretasikan[11].

Dalam bidang pendidikan, penerapan data mining berkembang pesat seiring ketersediaan data akademik yang besar. Data mining digunakan untuk menganalisis performa mahasiswa, prediksi kelulusan, deteksi siswa berisiko, hingga pemetaan gaya belajar. Dengan menganalisis data nilai, kehadiran, partisipasi kelas, dan variabel lainnya, institusi pendidikan dapat mengambil kebijakan yang lebih efektif dan personal. Oleh karena itu, penelitian-penelitian terbaru mulai mengintegrasikan data mining sebagai alat bantu utama dalam pengambilan keputusan berbasis data (data-driven decision making)[12]. Pada data mining sendiri terdapat 5 (lima) teknik didalamnya yang mengikuti bagaimana proses tahapan penyelesaiannya seperti berikut[13]:

1. Asosiasi (Association)

Asosiasi adalah teknik data mining yang digunakan untuk menemukan pola hubungan atau keterkaitan antar item dalam sebuah dataset. Teknik ini sangat populer dalam analisis keranjang belanja (market basket analysis). Tujuannya adalah untuk menemukan aturan asosiasi seperti “jika membeli A maka cenderung membeli B”. Beberapa contoh algoritma yaitu Apriori, FP-Growth, Eclat

2. Prediksi (Prediction)

Prediksi adalah proses memperkirakan nilai masa depan berdasarkan pola data historis. Meskipun mirip dengan klasifikasi, prediksi biasanya digunakan untuk data berkelanjutan (continuous data), bukan label diskrit. Beberapa contoh algoritma yaitu Linear Regression, Decision Tree Regression, ARIMA

3. Estimasi (Estimation)

Estimasi merupakan proses memperkirakan nilai numerik dari suatu variabel berdasarkan nilai dari variabel lainnya. Estimasi hampir serupa dengan prediksi, tetapi lebih fokus pada penilaian nilai sekarang atau yang belum diketahui daripada masa depan. Beberapa contoh algoritma yaitu Multiple Regression, KNN Regression, SVR

4. Klasifikasi (Classification)

Klasifikasi adalah proses mengelompokkan data ke dalam kelas-kelas yang telah ditentukan berdasarkan atribut tertentu. Ini merupakan salah satu teknik paling umum dalam data mining dan termasuk dalam supervised learning. Beberapa contoh algoritma yaitu Decision Tree, Naïve Bayes, SVM, KNN, Random Forest

5. Klasterisasi (Clustering)

Klasterisasi adalah teknik unsupervised learning yang bertujuan untuk mengelompokkan data ke dalam beberapa kelompok (cluster) berdasarkan kemiripan. Berbeda dengan klasifikasi, klasterisasi tidak memerlukan label atau kelas yang sudah ada sebelumnya. Beberapa contoh algoritma yaitu K-Means, DBSCAN, Hierarchical

2.3 Algoritma Decision Tree

Algoritma Decision Tree adalah salah satu teknik klasifikasi dalam data mining yang digunakan untuk memodelkan keputusan dan konsekuensinya dalam bentuk struktur pohon. Metode ini bekerja dengan membagi data berdasarkan atribut-atribut tertentu menggunakan ukuran statistik seperti entropy, information gain, dan gain ratio. Salah satu keunggulan Decision Tree adalah modelnya yang mudah dipahami karena hasil klasifikasinya dapat divisualisasikan dalam bentuk cabang dan simpul seperti diagram alur. Algoritma ini banyak digunakan dalam berbagai bidang seperti kesehatan, keuangan, pendidikan, dan pemasaran[14].

Salah satu algoritma Decision Tree yang paling populer adalah C4.5, yang merupakan pengembangan dari algoritma ID3. C4.5 memperbaiki beberapa kekurangan ID3, seperti kemampuannya menangani nilai kontinu, missing values, serta melakukan proses pruning untuk menghindari overfitting. Dalam proses klasifikasi, algoritma Decision Tree memilih atribut terbaik sebagai akar pohon berdasarkan perhitungan information gain atau gain ratio. Atribut tersebut kemudian digunakan untuk membagi dataset hingga diperoleh leaf nodes yang menyatakan hasil klasifikasi. Dalam prosesnya algoritma Decision Tree memiliki beberapa tahap sebagai berikut[15]:

1. Inisialisasi Data

Dataset harus dipersiapkan, biasanya dalam bentuk tabel yang terdiri dari atribut-atribut dan label kelas.

2. Perhitungan Entropy

Entropy digunakan untuk mengukur ketidakpastian atau impurity dalam suatu himpunan data. Rumus entropy:

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

3. Perhitungan Information Gain

Information gain digunakan untuk menentukan seberapa besar pengurangan entropy setelah suatu atribut digunakan untuk membagi data:

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_v|}{|S|} * Entropy(S_v) \quad (2)$$

4. Perhitungan Gain Ratio (untuk C4.5)

C4.5 menggunakan gain ratio untuk mengatasi bias terhadap atribut dengan banyak nilai unik:

$$GainRatio(A) = \frac{Gain(S, A)}{SplitInfo(A)} \quad (3)$$

$$SplitInfo(A) = - \sum \frac{|S_v|}{|S|} \log_2 \left(\frac{|S_v|}{|S|} \right) \quad (4)$$

5. Pemilihan Atribut Terbaik

Atribut dengan gain ratio tertinggi dipilih sebagai simpul akar (root) pohon, dan proses dilanjutkan secara rekursif.

6. Pembuatan Cabang dan Daun

Data dibagi berdasarkan nilai atribut, dan proses perhitungan diulangi hingga setiap node menjadi leaf node (kelas akhir), atau tidak ada atribut lagi yang bisa dipilih.

7. Pruning (Pemangkasan)

Tahapan akhir adalah pemangkasan pohon (optional), untuk menghindari overfitting. Pruning menghapus node yang tidak memberikan kontribusi signifikan terhadap akurasi klasifikasi.

Parameter pengukuran pada penelitian berkaitan dengan data yang digunakan pada penelitian. Pada data mining data berkaitan dengan atribut, adapun atribut data yang digunakan pada penelitian dapat dilihat berikut:

Tabel 2. Atribut Pengukuran Data

No	Nama Atribut
1	Jumlah Frekuensi Tugas
2	Jenis Penugasan
3	Modul / Media Pembelajaran
4	Monitoring Tugas / Project
5	Sifat Ujian Akhir

Dalam penelitian ini, evaluasi performa algoritma Decision Tree dilakukan dengan menggunakan metode confusion matrix sebagai alat utama pengukuran, dan F1-Score sebagai ukuran keseimbangan antara presisi dan recall. Confusion matrix sangat berguna untuk mengevaluasi hasil klasifikasi karena dapat menunjukkan distribusi prediksi yang benar dan salah terhadap masing-masing kelas.

1. Confusion Matrix

Confusion matrix adalah tabel evaluasi yang menunjukkan performa model klasifikasi dengan membandingkan hasil prediksi dan nilai aktual. Untuk kasus klasifikasi biner, confusion matrix terdiri dari empat komponen:

- True Positive (TP): jumlah data yang diprediksi benar sebagai kelas positif
- False Positive (FP): jumlah data yang diprediksi salah sebagai kelas positif
- True Negative (TN): jumlah data yang diprediksi benar sebagai kelas negatif
- False Negative (FN): jumlah data yang diprediksi salah sebagai kelas negatif

Dari komponen-komponen tersebut, dihitung metrik evaluasi utama:

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

2. F1-Score

F1-Score adalah ukuran harmonik antara precision dan recall. Metrik ini penting ketika terjadi ketidakseimbangan kelas (class imbalance), karena mampu memberikan gambaran lebih adil terhadap performa model.

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (8)$$

2.4 Penelitian Terdahulu

Penelitian terdahulu merupakan dasar ide terhadap proses penelitian yang dilakukan. Penelitian terdahulu berkaitan dengan Judul, Metode yang dipakai, Temuan dan Keterbatasan dari penelitian yang telah dilakukan. Adapun penelitian terdahulu dapat dilihat berikut:

Tabel 3. Penelitian Terdahulu

No	Penulis, Judul dan Tahun	Metode/Temuan Utama/Keterbatasan/Perbedaan
1	[16] Implementasi Data Mining Algoritma Decision Tree Untuk	Metode: Pendekatan data mining dengan algoritma Decision Tree C4.5 untuk

No	Penulis, Judul dan Tahun	Metode/Temuan Utama/Keterbatasan/Perbedaan
	Klasifikasi Status Gizi Balita Di Kecamatan Ciledug	Temuan Utama: Hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 sangat efektif dalam mengklasifikasikan status gizi balita dengan akurasi tertinggi mencapai 99,18%. Model berhasil membedakan antara kategori gizi normal dan tidak normal berdasarkan atribut seperti usia, berat badan, dan tinggi badan. Evaluasi dengan confusion matrix menunjukkan bahwa sebagian besar prediksi tepat sasaran. Keterbatasan: Data yang digunakan hanya berasal dari satu wilayah administratif (Kecamatan Ciledug), sehingga generalisasi ke wilayah lain masih perlu diuji. Penelitian juga belum menjelaskan secara rinci pengaruh faktor eksternal lain seperti lingkungan, tingkat pendidikan orang tua, atau pola konsumsi harian balita. Perbedaan: Perbedaan utama antara penelitian ini dengan penelitian sebelumnya terletak pada cakupan data dan pengujian parameter model yang lebih lengkap.
2	[17] Prediksi Perilaku Pelanggan Pada Produk UMKM Batik Dengan Menggunakan Algoritma Decision Tree	Metode: Pendekatan data mining dengan algoritma Decision Tree C4.5 Temuan Utama: Hasil penelitian menunjukkan bahwa algoritma Decision Tree C4.5 mampu membentuk pohon keputusan yang memberikan wawasan terhadap perilaku pelanggan, dengan tingkat akurasi mencapai 81,25%. Keterbatasan: Penelitian ini memiliki keterbatasan dalam jumlah data yang relatif kecil, yaitu hanya 68 record setelah pemrosesan, yang mungkin kurang merepresentasikan keseluruhan pola perilaku pelanggan dalam jangka panjang. Perbedaan: Perbedaan utama penelitian ini dibandingkan penelitian sejenis terletak pada fokus implementasi Decision Tree untuk prediksi perilaku pelanggan dalam lingkup UMKM batik lokal
3	[18] Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa	Metode: Metode klasifikasi berbasis algoritma Decision Tree (C4.5) Temuan Utama: Hasil utama dari penelitian ini menunjukkan bahwa algoritma Decision Tree mampu menghasilkan klasifikasi yang cukup baik dalam memprediksi kelulusan mahasiswa tepat waktu, dengan akurasi tertinggi sebesar 87,93% pada skenario pembagian data 90% training dan 10% testing. Precision dan recall juga menunjukkan performa yang memadai, masing-masing sebesar 94,87% dan 81,32% untuk kelas “tepat waktu”. Keterbatasan: Meskipun hasil akurasi cukup tinggi, penelitian ini masih memiliki beberapa keterbatasan. Salah satunya adalah ketergantungan terhadap data alumni sebagai acuan prediksi, yang mungkin tidak mencerminkan kondisi real-time mahasiswa aktif saat ini. Perbedaan: Jumlah data yang besar dan representatif (1.739 record) serta penggunaan teknik validasi berulang untuk memastikan stabilitas hasil klasifikasi.
4	[19] Penerapan Algoritma Decision Tree dalam Klasifikasi Data “Framingham” Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang	Metode: Algoritma Decision Tree dengan pendekatan machine learning Temuan Utama: Model Decision Tree yang dibangun dari dataset Framingham menghasilkan akurasi sebesar 73,36%, menunjukkan kinerja klasifikasi yang cukup baik dalam memprediksi risiko penyakit jantung. Atribut-atribut yang paling signifikan dalam klasifikasi adalah usia, tekanan darah, dan riwayat merokok. Keterbatasan: Keterbatasan utama dari penelitian ini terletak pada ketidakseimbangan data (class imbalance) antara kelas 0 dan kelas 1, yang menyebabkan rendahnya nilai precision dan recall pada kelas risiko tinggi. Perbedaan: Penelitian ini berbeda dari studi sebelumnya karena secara spesifik mengimplementasikan Decision Tree dengan pemrograman Python menggunakan pustaka scikit-learn
5	[20] Prediksi Siswa Putus Sekolah Menggunakan Algoritma Decision Tree C4.5	Metode: Penelitian ini menggunakan metode Decision Tree C4.5 Temuan Utama: Dengan penerapan algoritma Decision Tree C4.5, model mampu menghasilkan akurasi sebesar 92,20%, precision sebesar 84%, dan recall sebesar 50%. Keterbatasan: Keterbatasan utama penelitian ini adalah terbatasnya sumber dan jumlah data, karena data yang digunakan adalah data umum dan tidak berasal dari institusi pendidikan tertentu.

No	Penulis, Judul dan Tahun	Metode/Temuan Utama/Keterbatasan/Perbedaan
		Perbedaan: Pendekatan kombinasi penghitungan manual dan implementasi program Python menjadi keunggulan tersendiri, karena mampu memverifikasi keakuratan hasil klasifikasi dari dua sisi (teoritis dan praktis).

Dari penelitian terdahulu yang telah dilakukan maka dapat dilihat untuk State of The Art sebagai berikut penelitian berbasis data mining dengan algoritma Decision Tree terus berkembang dan telah banyak diaplikasikan dalam berbagai bidang, mulai dari kesehatan, pendidikan, hingga bisnis. Kelima artikel yang dikaji dalam penelitian ini menunjukkan pendekatan dan konteks aplikasi yang berbeda, namun memiliki benang merah dalam penggunaan algoritma Decision Tree (khususnya C4.5) sebagai metode klasifikasi utama. Dengan demikian, penelitian ini akan berkontribusi lebih lanjut dalam ranah pendidikan vokasi, khususnya untuk mengidentifikasi pola pemahaman mahasiswa pada mata kuliah pemrograman. Pendekatan ini belum banyak dijelajahi secara khusus dalam artikel-artikel sebelumnya, sehingga menawarkan novelty dalam konteks domain dan integrasi antara data akademik dengan evaluasi berbasis Decision Tree.

3. HASIL DAN PEMBAHASAN

3.1 Penerapan Algoritma Decision Tree

Sebelum dilakukan proses penyelesaian penelitian, terlebih dahulu dilakukan proses pengumpulan data terhadap respon dari mahasiswa di Jurusan Teknik Komputer dan Informatika terhadap pemahaman pada matakuliah komputer. Pengumpulan data terhadap mahasiswa meliputi respon terhadap atribut yang akan digunakan pada penelitian. Dalam hal ini, atribut yang digunakan pada penelitian meliputi dari Jumlah Frekuensi Tugas, Jenis Penugasan, Modul / Media Pembelajaran, Monitoring Tugas / Project dan Sifat Ujian Akhir. Selain itu, juga dilakukan pengumpulan data terhadap pemahaman mahasiswa terhadap matakuliah Pemrograman berdasarkan dengan atribut – atribut tersebut. Adapun isian nilai terhadap masing – masing atribut dapat dilihat berikut:

Tabel 4. Jenis Penugasan

No	Jenis Penugasan
1	Kelompok
2	Mandiri

Pada tabel 4 diatas, dapat dilihat bahwasannya untuk atribut Jenis Penugasan terdapat 2 (dua) isian yaitu Kelompok dan Mandiri. Pada atribut ini berkaitan dengan jenis tugas yang diberikan kepada mahasiswa apakah bersifat kelompok ataupun bersifat mandiri. Selanjutnya, dapat diketahui terhadap isian pada atribut Modul / Media Pembelajaran. Adapun isian tersebut dapat dilihat berikut:

Tabel 5. Modul / Media Pembelajaran

No	Modul / Media Pembelajaran
1	Praktik Dikelas
2	Video Tutorial

Pada tabel 5, berkaitan dengan bagaimana proses pembelajaran yang dilakukan oleh Dosen pengampu pada matakuliah Pemrograman. Pada tabel tersebut dapat dilihat bahwasannya untuk perkuliahan dilakukan Praktik Didalam Kelas ataupun terdapat Video Tutorial pada pembelajaran. Berikutnya merupakan data terhadap atribut Monitoring Tugas / Project, adapun dapat terlihat pada tabel berikut:

Tabel 6. Monitoring Tugas / Project

No	Monitoring Tugas / Project
1	Ada
2	Tidak Ada

Pada tabel 6 diatas dapat dilihat terhap atribut selanjutnya yang berkaitan dengan proses monitoring terhap Tugas ataupun Project. Pada atribut tersebut berkaitan dengan apakah terdapat atau tidak proses monitoring yang dilakukan oleh Dosen Pengampu Matakuliah terhadap Tugas ataupun Project yang diberikan. Atribut berikutnya yaitu Sifat Ujian Akhir yang dapat dilihat berikut ini:

Tabel 7. Sifat Ujian Akhir

No	Sifat Ujian Akhir
1	Ujian Kelas
2	Project

Pada tabel 7 diatas, dapat dilihat atribut yang berkaitan dengan bagaimana Ujian Akhir dilakukan. Pada tabel tersebut dapat dilihat bahwasannya Sifat Ujian Akhir terbagi atas Ujian Kelas ataupun Project. Terakhir yaitu respon terhadap hasil pemahaman mahasiswa yang nantinya menjadi kelas pada proses penelitian. Adapun hasil pemahaman dapat dilihat berikut:

Tabel 8. Hasil Pemahaman Mahassiswa

No	Hasil Pemahaman
1	Paham
2	Cukup Paham
3	Tidak Paham

Pada tabel 8 diatas dapat dilihat terhadap pengukuran pemahaman mahasiswa terhadap Matakuliah Pemrograman. Pengukuran pemahaman berdasarkan dengan proses pembelajaran yang telah dilakukan sesuai dengan atribut-atribut yang digunakan pada penelitian. Pada tabel tersebut dapat dilihat bahwasannya terdapat 3 (tiga) tingkat pemahaman pada mahasiswa yaitu Paham, Cukup Paham dan Tidak Paham. Setelah diketahui untuk seluruh Atribut, Nilai Atribut dan Hasil Pemahaman, maka selanjutnya dapat dilakukan proses pengumpulan data berdasarkan dengan respon dari mahasiswa di Jurusan Teknik Komputer dan Informatika terhadap matakuliah Pemrograman. Adapun data tersebut dapat dilihat berikut:

Tabel 9. Hasil Pengumpulan Data Mahasiswa

N o	Jumlah Frekuensi Tugas	Jenis Penugasan	Modul / Media Pembelajaran	Monitoring Pertemuan	Sifat Ujian Akhir	Hasil
1	3	Kelompok	Video Tutorial	Ada	Ujian Kelas	Cukup Paham
2	2	Mandiri	Praktik Dikelas	Ada	Project	Tidak Paham
3	1	Mandiri	Video Tutorial	Ada	Project	Cukup Paham
4	2	Mandiri	Video Tutorial	Ada	Ujian Kelas	Cukup Paham
5	5	Mandiri	Praktik Dikelas	Ada	Project	Cukup Paham
...	...	...	...	...	...	...
7	5	Kelompok	Video Tutorial	Tidak Ada	Ujian Kelas	Tidak Paham

Pada tabel 9 dapat dilihat secara keseluruhan terhadap sampel data yang akan digunakan pada penelitian. Pada penelitian, jumlah responden atau data terdapat 75 orang mahasiswa yang memebrikan respon. Berdasarkan tabel tersebut, dilakukan terlebih dahulu tahapan preprocessing terhadap Atribut Jumlah Frekuensi Tugas. Preprocessing dilakukan berdasarkan dengan penyesuaian terhadap Algoritma Decision Tree yang merupakan proses pengelompokan terhadap data. Maka dari itu, isian dari atribut tersebut tersebut yang bersifat Numerik / Angka harus dilakukan proses pengelompokan terlebih dahulu. Proses pengelompokan yang dilakukan berdasarkan dengan berikut:

Tabel 10. Preprocessing Atribut Jumlah Frekuensi Tugas

No	Jumlah Frekuensi Tugas	Hasil Pengelompokan
1	0-2	Sedikit
2	3-4	Sedang
3	>4	Banyak

Pada tabel 10 dapat dilihat terhadap tahapan preprocessing terhadap isian data pada Atribut Jumlah Frekuensi Tugas. Pada tabel tersebut dapat dilihat bahwasannya jumlah tugas 0 – 2 dikategorikan dalam Tugas Sedikit, untuk jumlah tugas 3-4 dikategorikan dalam Tugas Sedang dan jika tugas lebih dari 4 dikategorikan terhadap Tugas Banyak. Berdasarkan dengan proses preprocessing maka dapat dilihat terhadap data sebagai berikut:

Tabel 11. Data Hasil Preprocessing

N o	Jumlah Frekuensi Tugas	Jenis Penugasan	Modul / Media Pembelajaran	Monitoring Pertemuan	Sifat Ujian Akhir	Hasil
1	Sedang	Kelompok	Video Tutorial	Ada	Ujian Kelas	Cukup Paham

2	Sedikit	Mandiri	Praktik Dikelas	Ada	Project	Tidak Paham
3	Sedikit	Mandiri	Video Tutorial	Ada	Project	Cukup Paham
4	Sedikit	Mandiri	Video Tutorial	Ada	Ujian Kelas	Cukup Paham
5	Banyak	Mandiri	Praktik Dikelas	Ada	Project	Cukup Paham
...	...	...	...	...	...	...
7	Banyak	Kelompok	Video Tutorial	Tidak Ada	Ujian Kelas	Tidak Paham

Berdasarkan pada tabel 10, maka selanjutnya dapat dilakukan proses penyelesaian penelitian dengan menggunakan Algoritma Decision Tree dimana proses penyelesaian dengan melakukan pengujian menggunakan Bahasa Python pada Google Colaboration. Adapun tahapan proses pengujian dapat dilihat berikut:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier, plot_tree, export_text
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report, f1_score
from sklearn.preprocessing import LabelEncoder
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files
```

Pada coding program diatas digunakan untuk melakukan Import Library yang akan digunakan pada pengujian dimana Library yg digunakan seperti Pandas, Numpy, Confusion Matriks dan Pohon Keputusan.

```
uploaded = files.upload()
```

Proses selanjutnya adalah melakukan Import data pada Google Colaboration, dalam hal ini data yang digunakan merupakan data Excel.

```
data = pd.read_excel("DataPengujian.xlsx")
print("==== Contoh Data =====")
print(data.head())
```

```
==== Contoh Data =====
Jumlah Frekuensi Tugas Jenis Penugasan Modul / Media Pembelajaran \
0          Sedang          Kelompok          Video Tutorial
1          Sedikit         Mandiri          Praktik Dikelas
2          Sedikit         Mandiri          Video Tutorial
3          Sedikit         Mandiri          Video Tutorial
4          Banyak          Mandiri          Praktik Dikelas

Monitoring Pertemuan Sifat Ujian Akhir          Hasil
0          Ada          Ujian Kelas          Cukup Paham
1          Ada          Project          Tidak Paham
2          Ada          Project          Cukup Paham
3          Ada          Ujian Kelas          Cukup Paham
4          Ada          Project          Cukup Paham
```

Gambar 4. Hasil Pembacaan Data dan Menampilkan Data

Setelah memasukkan data pada Google Colaboration seperti gambar 4, selanjutnya melakukan pembacaan data dimana untuk nama file yang digunakan pada penelitian yaitu DataPengujian.xlsx.

```
le = LabelEncoder()
for col in data.columns:
    if data[col].dtype == 'object':
        data[col]= le.fit_transform(data[col].astype(str))
print("\n==== Data Setelah Encoding =====")
print(data.head())
```

Setelah data berhasil dibaca, maka dengan menggunakan syntax diatas dilakukan proses encoding data, dimana data yang bersifat kategorikal selanjutnya dirubah menjadi data numerikal.

==== Data Setelah Encoding ====

	Jumlah Frekuensi Tugas	Jenis Penugasan	Modul / Media Pembelajaran \
0	1	0	1
1	2	1	0
2	2	1	1
3	2	1	1
4	0	1	0

	Monitoring Pertemuan	Sifat Ujian Akhir	Hasil
0	0	1	0
1	0	0	2
2	0	0	0
3	0	1	0
4	0	0	0

Gambar 5. Hasil Encoding Data

Setelah dilakukan proses encoding data seperti tampak pada gambar 5 diatas, maka selanjutnya dilakukan proses Split data dan Pembentukan Atribut Pohon keputusan dengan menggunakan coding

```
target_column = 'Hasil'
X = data.drop(target_column, axis=1)
y = data[target_column]
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.4, random_state=42
)
model = DecisionTreeClassifier(criterion='entropy', random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)
```

Pada proses Split Data, data yang digunakan untuk proses pengujian sebesar 40% atau 0,4 dari total data keseluruhan. Selanjutnya, dilakukan proses untuk menghitung nilai Akurasi dan F1-Score.

```
akurasi = accuracy_score(y_test, y_pred)
f1 = f1_score(y_test, y_pred, average='weighted')
print("\n==== HASIL EVALUASI MODEL =====")
print(f"Akurasi Model : {akurasi * 100:.2f}%")
print(f"F1-Score (Weighted): {f1:.4f}")
print("\nClassification Report:\n", classification_report(y_test, y_pred))
```

Dengan menggunakan syntax diatas dapat dipergunakan untuk menghitung nilai akurasi dan f1-score. Adapun hasil dapat dilihat pada gambar berikut

```
==== HASIL EVALUASI MODEL =====
Akurasi Model : 50.00%
F1-Score (Weighted): 0.5168

Classification Report:
              precision    recall  f1-score   support

    0               0.64         0.47         0.54         15
    1               0.22         0.40         0.29          5
    2               0.60         0.60         0.60         10

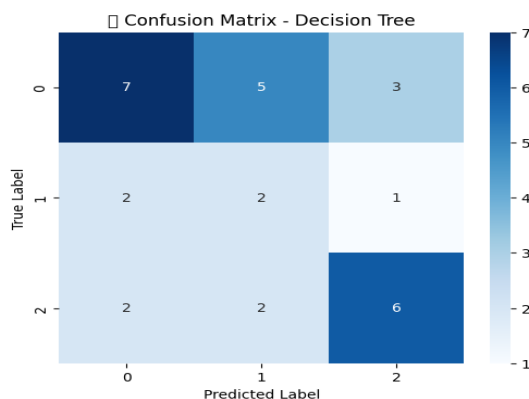
 accuracy               0.49
 macro avg              0.49
 weighted avg           0.56
```

Gambar 6. Nilai Akurasi dan F1-Score

Dari hasil pengujian yang dilakukan pada gambar 6, dapat dilihat bahwasannya untuk nilai akurasi yang didapatkan dengan Data Uji sebesar 40% adalah 50% dan nilai F1-Score sebesar 51,68%.

```
cm = confusion_matrix(y_test, y_pred)
plt.figure(figsize=(6, 5))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=np.unique(y),
            yticklabels=np.unique(y))
plt.xlabel('Predicted Label')
plt.ylabel('True Label')
plt.title('Confusion Matrix - Decision Tree')
plt.show()
```

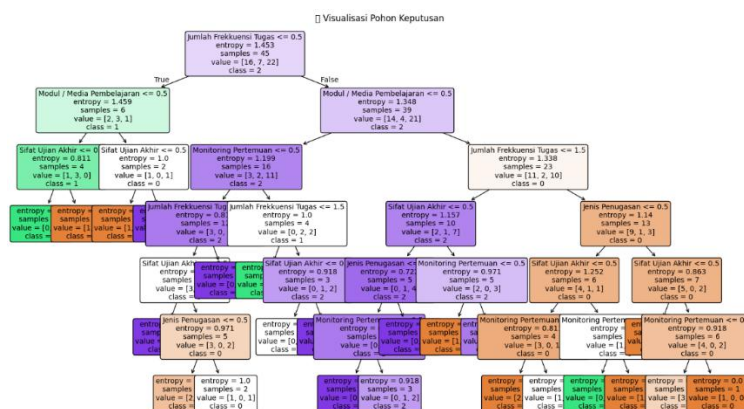
Selanjutnya, setelah didapatkan nilai akurasi, dengan menggunakan coding tersebut menampilkan untuk gambar Confusion Matrix. Adapun hasil Confusion Matrix dapat dilihat berikut:



Gambar 7. Confusion Matrix

Pada gambar 7 diatas dapat dilihat untuk nilai Confusion Matrix pada setiap Hasil Kelas. Selanjutnya dilakukan proses untuk menampilkan hasil pohon keputusan pada algoritma decision tree.

```
plt.figure(figsize=(18, 10))
plot_tree(
    model,
    feature_names=X.columns,
    class_names=[str(c) for c in np.unique(y)],
    filled=True,
    rounded=True,
    fontsize=10
)
plt.title("Visualisasi Pohon Keputusan")
plt.show()
```



Gambar 8. Pohon Keputusan

Pada gambar 8 diatas dapat dilihat untuk hasil pembentukan pohon keputusan dengan menggunakan algoritma Decision Tree, dari pohon tersebut nantinya akan disusun menjadi sebuah aturan Rule.

```
print("\n==== ATURAN (RULES) POHON KEPUTUSAN =====")
rules = export_text(model, feature_names=list(X.columns))
print(rules)
```

Syntax coding diatas dipergunakan untuk membuat aturan rule berdasarkan pembentukan pohon keputusan. Adapun hasil pembentukan rule pohon keputusan dapat dilihat berikut

```
==== ATURAN (RULES) POHON KEPUTUSAN =====
|--- Jumlah Frekuensi Tugas <= 0.50
|   |--- Modul / Media Pembelajaran <= 0.50
|   |   |--- Sifat Ujian Akhir <= 0.50
|   |   |   |--- class: 1
|   |   |   |--- Sifat Ujian Akhir > 0.50
|   |   |   |--- class: 0
|   |--- Modul / Media Pembelajaran > 0.50
|   |--- Sifat Ujian Akhir <= 0.50
```

```

| | | |--- class: 0
| | | |--- Sifat Ujian Akhir > 0.50
| | | |--- class: 2
|--- Jumlah Frekkuensi Tugas > 0.50
| |--- Modul / Media Pembelajaran <= 0.50
| | |--- Monitoring Pertemuan <= 0.50
| | | |--- Jumlah Frekkuensi Tugas <= 1.50
| | | |--- Sifat Ujian Akhir <= 0.50
| | | |--- class: 2
| | | |--- Sifat Ujian Akhir > 0.50
| | | |--- Jenis Penugasan <= 0.50
| | | |--- class: 0
| | | |--- Jenis Penugasan > 0.50
| | | |--- class: 0
| | | |--- Jumlah Frekkuensi Tugas > 1.50
| | | |--- class: 2
| | |--- Monitoring Pertemuan > 0.50
| | | |--- Jumlah Frekkuensi Tugas <= 1.50
| | | |--- class: 1
| | | |--- Jumlah Frekkuensi Tugas > 1.50
| | | |--- Sifat Ujian Akhir <= 0.50
| | | |--- class: 1
| | | |--- Sifat Ujian Akhir > 0.50
| | | |--- class: 2
| |--- Modul / Media Pembelajaran > 0.50
| | |--- Jumlah Frekkuensi Tugas <= 1.50
| | |--- Sifat Ujian Akhir <= 0.50
| | | |--- Jenis Penugasan <= 0.50
| | | |--- Monitoring Pertemuan <= 0.50
| | | |--- class: 2
| | | |--- Monitoring Pertemuan > 0.50
| | | |--- class: 2
| | | |--- Jenis Penugasan > 0.50
| | | |--- class: 2
| | |--- Sifat Ujian Akhir > 0.50
| | | |--- Monitoring Pertemuan <= 0.50
| | | |--- class: 0
| | | |--- Monitoring Pertemuan > 0.50
| | | |--- class: 2
| | |--- Jumlah Frekkuensi Tugas > 1.50
| | | |--- Jenis Penugasan <= 0.50
| | | |--- Sifat Ujian Akhir <= 0.50
| | | |--- Monitoring Pertemuan <= 0.50
| | | |--- class: 0
| | | |--- Monitoring Pertemuan > 0.50
| | | |--- class: 0
| | | |--- Sifat Ujian Akhir > 0.50
| | | |--- Monitoring Pertemuan <= 0.50
| | | |--- class: 1
| | | |--- Monitoring Pertemuan > 0.50
| | | |--- class: 0
| | |--- Jenis Penugasan > 0.50
| | | |--- Sifat Ujian Akhir <= 0.50
| | | |--- class: 0
| | | |--- Sifat Ujian Akhir > 0.50
| | | |--- Monitoring Pertemuan <= 0.50
| | | |--- class: 0
| | | |--- Monitoring Pertemuan > 0.50
| | | |--- class: 0

```

Setelah didapatkan aturan rule berdasarkan dengan pohon keputusan, nantinya aturan ini dipergunakan proses pengambilan keputusan dalam penentuan strategi terhadap pembelajaran terhadap matakuliah Pemrograman.

3.2 Pembahasan

Pengujian dilakukan menggunakan algoritma Decision Tree dengan pendekatan supervised learning. Dataset yang digunakan berasal dari file DataPengujian.xlsx yang berisi sekumpulan atribut (fitur) sebagai variabel input dan satu kolom target (label) sebagai variabel output. Sebelum dilakukan pelatihan model, data terlebih dahulu melalui tahap pra-pemrosesan, yaitu proses Label Encoding untuk mengubah data kategorikal seperti “Sedikit”, “Cukup”, atau “Banyak” menjadi bentuk numerik agar dapat diolah oleh algoritma Decision Tree. Data kemudian dibagi menjadi dua bagian, yaitu data latih sebesar 60% dan data uji sebesar 40% dengan random state yang tetap agar hasil dapat direplikasi. Hasil pengujian model menunjukkan nilai Akurasi sebesar 50% dan nilai F1-Score sebesar 51,68%.

Nilai akurasi menunjukkan bahwa model mampu memprediksi kelas target dengan tingkat kesalahan yang rendah. Nilai F1-Score juga memperlihatkan keseimbangan antara tingkat presisi dan recall, yang menandakan bahwa model tidak hanya akurat tetapi juga cukup stabil terhadap perbedaan jumlah data pada tiap kelas.

Selain itu, dilakukan pengujian dengan Confusion Matrix yang divisualisasikan dalam bentuk heatmap. Hasil matriks tersebut memperlihatkan bahwa sebagian besar data uji berhasil diklasifikasikan dengan benar ke dalam kelas yang sesuai. Hanya sebagian kecil data yang mengalami kesalahan klasifikasi, yang dapat disebabkan oleh kemiripan nilai antar fitur atau ketidakseimbangan jumlah data antar kelas.

Dari hasil visualisasi pohon keputusan, diperoleh struktur Decision Tree yang memperlihatkan proses pengambilan keputusan secara hierarkis. Setiap simpul (node) pada pohon menunjukkan atribut yang digunakan untuk membagi data berdasarkan nilai information gain tertinggi.

Pada cabang utama pohon, atribut dengan pengaruh paling besar terhadap hasil klasifikasi adalah Jumlah Frekuensi Tugas. Nilai atribut ini membagi data ke beberapa cabang keputusan berikutnya, hingga mencapai simpul akhir (leaf node) yang menunjukkan hasil klasifikasi akhir.

Hasil pengujian ini menunjukkan bahwa algoritma Decision Tree dapat digunakan secara efektif untuk mengklasifikasikan data dan mengidentifikasi pola hubungan antar atribut. Keunggulan utama Decision Tree terletak pada kemampuannya untuk menghasilkan model yang dapat dijelaskan secara visual dan logis, sehingga mudah diinterpretasikan baik oleh peneliti maupun pengguna non-teknis.

Meskipun model memberikan akurasi tinggi, tetap perlu diperhatikan potensi overfitting, terutama jika jumlah atribut dan kedalaman pohon terlalu besar. Oleh karena itu, tahap selanjutnya dapat dilakukan optimasi dengan metode pruning atau perbandingan dengan algoritma lain seperti Random Forest dan Naïve Bayes untuk menguji konsistensi performa model. Penelitian ini membuka peluang untuk dilakukan pengembangan lebih lanjut di masa mendatang. Pada tahap berikutnya, penelitian dapat difokuskan pada perluasan dan pendalaman analisis data dengan menambahkan variabel-variabel baru yang lebih beragam, seperti gaya belajar, motivasi, tingkat partisipasi di kelas, serta penggunaan media pembelajaran digital. Dengan memperluas atribut yang dianalisis, diharapkan pola pemahaman mahasiswa dapat digambarkan secara lebih komprehensif dan akurat.

Tahapan pengembangan berikutnya juga dapat diarahkan pada penerapan algoritma data mining yang lebih kompleks atau kombinasi metode (hybrid), seperti Random Forest, Gradient Boosting, atau Neural Network, untuk meningkatkan akurasi prediksi dan mengatasi keterbatasan Decision Tree yang mungkin muncul pada dataset besar. Pendekatan komparatif antar algoritma ini akan memberikan wawasan baru mengenai model klasifikasi terbaik yang dapat diterapkan dalam konteks pendidikan, khususnya untuk menganalisis kinerja belajar mahasiswa.

Selain itu, hasil penelitian ini dapat dikembangkan menjadi sistem pendukung keputusan (SPK) berbasis web atau aplikasi analitik pendidikan. Sistem ini akan mengintegrasikan hasil proses data mining sehingga dosen dan pihak kampus dapat secara interaktif memantau, menganalisis, dan mengevaluasi tingkat pemahaman mahasiswa secara real-time. Dengan adanya sistem tersebut, hasil penelitian tidak hanya berhenti pada analisis akademik, tetapi juga dapat dimanfaatkan langsung sebagai alat bantu evaluasi dan perencanaan pembelajaran.

Tahapan lanjutan lainnya adalah melakukan implementasi dan pengujian di berbagai mata kuliah lain atau institusi berbeda untuk menguji keandalan model yang telah dikembangkan. Dengan memperluas ruang lingkup penelitian, dapat diketahui sejauh mana model dan metode yang digunakan bersifat general dan mampu diadaptasi pada konteks pembelajaran lain. Langkah ini juga akan memperkuat kontribusi penelitian dalam bidang educational data mining dan membuka peluang kolaborasi lintas perguruan tinggi dalam pengembangan analitik pembelajaran berbasis data.

Dalam jangka waktu 3–5 tahun ke depan, penelitian ini akan diarahkan menuju pengembangan yang lebih luas dan aplikatif dengan fokus pada pemanfaatan data mining untuk peningkatan kualitas pembelajaran secara menyeluruh. Tahap awal dari pengembangan jangka panjang adalah integrasi model analisis ke dalam sistem informasi akademik. Dengan memanfaatkan data historis mahasiswa yang tersimpan dalam sistem kampus, hasil model Decision Tree dapat digunakan secara otomatis untuk memberikan rekomendasi terhadap proses pembelajaran, misalnya dalam penentuan strategi pengajaran, deteksi dini mahasiswa berisiko rendah pemahaman, dan pemberian bimbingan belajar yang terarah.

Pada fase selanjutnya, penelitian ini akan dikembangkan menuju implementasi sistem pendukung keputusan adaptif (adaptive learning analytics system). Sistem ini akan mampu mempelajari perilaku belajar mahasiswa secara dinamis dan menyesuaikan rekomendasi pembelajaran berdasarkan pola yang terus diperbarui. Penggunaan algoritma lanjutan seperti Random Forest, XGBoost, atau Deep Learning dapat diterapkan untuk meningkatkan akurasi prediksi dan memperkaya kemampuan analisis terhadap data dalam skala besar.

Dalam jangka menengah, penelitian akan diperluas pada lingkup lintas mata kuliah dan lintas institusi untuk membangun model pembelajaran berbasis data nasional. Kolaborasi dengan perguruan tinggi lain diharapkan dapat menghasilkan basis data yang lebih besar dan beragam, sehingga model yang dikembangkan menjadi lebih general dan representatif. Hal ini juga akan membuka peluang untuk membangun framework analitik pendidikan berbasis machine learning yang dapat digunakan oleh berbagai institusi sebagai alat bantu evaluasi akademik.

Sementara itu, dalam jangka panjang, arah penelitian akan difokuskan pada pengembangan sistem cerdas pembelajaran berbasis kecerdasan buatan (AI-Based Learning Assistant) yang terintegrasi dengan platform e-learning. Sistem ini akan mampu memberikan umpan balik langsung kepada mahasiswa berdasarkan pola interaksi dan hasil belajar mereka, sekaligus membantu dosen dalam merancang pendekatan pembelajaran yang personal (personalized learning). Dengan demikian, hasil penelitian ini tidak hanya berkontribusi pada pengembangan ilmu di bidang data mining dan pendidikan, tetapi juga menjadi fondasi bagi implementasi pembelajaran cerdas di era transformasi digital pendidikan tinggi.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan menggunakan algoritma Decision Tree, dapat disimpulkan bahwa metode ini mampu menghasilkan model klasifikasi yang memiliki tingkat akurasi tinggi serta interpretasi yang mudah dipahami. Hasil pengujian menunjukkan nilai akurasi sebesar 50,00% dan F1-Score sebesar 51,69%, yang menandakan bahwa model memiliki kemampuan yang baik dalam mengenali pola data dan memberikan hasil prediksi yang akurat. Melalui proses pembentukan pohon keputusan, diperoleh aturan-aturan (*rules*) yang menggambarkan hubungan antara atribut-atribut dalam dataset dengan hasil klasifikasinya. Aturan tersebut memberikan pengetahuan baru tentang faktor-faktor yang paling berpengaruh terhadap hasil keputusan. Hal ini membuktikan bahwa algoritma Decision Tree tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai alat analisis yang mampu menjelaskan alur pengambilan keputusan secara logis dan sistematis. Secara keseluruhan, penerapan algoritma Decision Tree dalam penelitian ini terbukti efektif dalam mengidentifikasi pola dan menghasilkan keputusan yang akurat berdasarkan data yang tersedia. Dengan tingkat akurasi yang tinggi, metode ini dapat dijadikan acuan dalam pengembangan sistem pendukung keputusan yang berbasis data mining pada domain penelitian yang relevan, seperti evaluasi kinerja, prediksi hasil belajar, atau analisis perilaku pengguna.

UCAPAN TERIMA KASIH

Terima kasih kepada Politeknik Negeri Medan atas pendanaan yang diberikan melalui Kontrak : B/345/PL5.PT.01.05/2025 yang berasal dari dana DIPA POLMED tahun 2025.

REFERENCES

- [1] S. P. Dewi, N. Nurwati, and E. Rahayu, "Penerapan Data Mining Untuk Prediksi Penjualan Produk Terlaris Menggunakan Metode K-Nearest Neighbor," *Build. Informatics, Technol. Sci.*, vol. 3, no. 4, pp. 639–648, 2022, doi: 10.47065/bits.v3i4.1408.
- [2] M. Fajar, S. Adam, B. Putra, S. I. Puteri, A. Fajrissiddiq, and L. Sani, "Eksplorasi dan Analisis Data Mining untuk Prediksi Pola Konsumen Menggunakan Teknik Klasifikasi dan Clustering," in *SENTIMETER (Seminar Nasional Teknologi Informasi, Mekatronika dan Ilmu Komputer) Universitas Nusa Putra, 17 Mei 2025 Eksplorasi*, 2025.
- [3] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, and Rahmaddeni, "Implementation of Decision Tree Algorithm C4.5 and Support Vector Regression for Stroke Disease Prediction," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 61–67, 2022, doi: 10.1088/1742-6596/1641/1/012025.
- [4] K. Khotimah, "Teknik Data Mining menggunakan Algoritma Decision Tree (C4.5) untuk Prediksi Seleksi Beasiswa Jalur KIP pada Universitas Muhammadiyah Kotabumi," *J. SIMADA (Sistem Inf. dan Manaj. Basis Data)*, vol. 4, no. 2, pp. 145–152, 2022, doi: 10.30873/simada.v4i2.3064.
- [5] F. Faisal, H. Dhika, and H. Veris, "Penerapan Algoritma Decision Tree Dalam Penjualan Handphone," *JRKT (Jurnal Rekayasa Komputasi Ter.)*, vol. 1, no. 04, pp. 239–246, 2021, doi: 10.30998/jrkt.v1i04.6157.
- [6] S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok," *J. Krisnadana*, vol. 1, no. 2, pp. 21–32, 2022, doi: 10.58982/krisnadana.v1i2.108.
- [7] I. Iddrus and D. W. Sari, "Penerapan Data Mining Menggunakan Algoritma Decision Tree C4.5 Untuk Memprediksi Mahasiswa Drop Out Di Universitas Wiraraja," *J. Adv. Res. Inform.*, vol. 1, no. 02, pp. 1–7, 2023, doi: 10.24929/jars.v1i02.2684.
- [8] Dwita Elisa Sinaga, Agus Perdana Windarto, and Rizki Alfadillah Nasution, "Analisis Data Mining Algoritma Decision Tree Pada Prediksi Persediaan Obat (Studi Kasus : Apotek Franch Farma)," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 4, pp. 123–131, 2022, doi: 10.30865/klik.v2i4.328.
- [9] U. Suriani, "Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5," *Journalcisa*, vol. 3, no. 2, pp. 55–66, 2023, [Online]. Available: <http://jesik.web.id/index.php/jesik/article/view/91>
- [10] D. P. Indini, Mesran, and Dito Putro Utomo, "Penerapan Data Mining Dalam Pengelompokan Data Reseller di Telkomsel Authorized Partner (TAP) Deli Tua Dengan Algoritma K-Means," *J. Ilm. Media Siso*, vol. 17, no. 2, pp. 189–202, 2023, doi: 10.33998/mediasisfo.2023.17.2.1391.
- [11] D. P. Indini, S. R. Siburian, Nurhasanah, and D. P. Utomo, "Implementasi Algoritma DBSCAN untuk Clustering Seleksi Penentuan Mahasiswa yang Berhak Menerima Beasiswa Yayasan," in *Prosiding Seminar Nasional Sosial, Humaniora, dan Teknologi*, 2022, pp. 325–331.

- [12] Mesran, M. Syahrizal, Sarwandi, S. Aripin, D. P. Utomo, and A. Karim, "A comparison of the performance of data mining classification algorithms on medical datasets with the application of data normalization," *AIP Conf. Proc.*, vol. 3048, no. 1, 2024.
- [13] Zahrudin, A. I. Purnamasari, and I. Ali, "Analisis Tren Penjualan Fashion Import Menggunakan Algoritma Fp-Growth Pada Toko Air Gaul," *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 3, no. 2, pp. 75–84, 2023, doi: 10.69916/jkbti.v3i2.127.
- [14] W. A. Firmansyach, U. Hayati, and Y. Arie Wijaya, "Analisa Terjadinya Overfitting Dan Underfitting Pada Algoritma Naive Bayes Dan Decision Tree Dengan Teknik Cross Validation," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 262–269, 2023, doi: 10.36040/jati.v7i1.6329.
- [15] H. P. Herlambang, F. Saputra, M. H. Prasetyo, D. Puspitasari, and D. Nurlaela, "Perbandingan Klasifikasi Tingkat Penjualan Buah di Supermarket dengan Pendekatan Algoritma Decision Tree, Naive Bayes dan K-Nearest Neighbor," *J. Insa. - J. Inf. Syst. Manag. Innov.*, vol. 3, no. 1, pp. 21–28, 2023, doi: 10.31294/jinsan.v3i1.2097.
- [16] S. B. Bulkisah, R. Astuti, and A. Bahtiar, "Implementasi Data Mining Algoritma Decision Tree Untuk Klasifikasi Status Gizi Balita Di Kecamatan Ciledug," *J. Ilm. Inform. Komput.*, vol. 29, no. 1, pp. 1–12, 2024, doi: 10.35760/ik.2024.v29i1.10346.
- [17] M. Rizaludin and F. Fikriah, "Prediksi Perilaku Pelanggan Pada Produk UMKM Batik Dengan Menggunakan Algoritma Decision Tree," *Teknomatika*, vol. 13, no. 02, pp. 8–16, 2023.
- [18] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, 2023, doi: 10.33379/gtech.v7i1.1850.
- [19] A. R. Raharja, Jayadi, A. Pramudianto, and Y. Muchsam, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data 'Framingham' Untuk Menunjukkan Risiko Seseorang Terkena Penyakit Jantung dalam 10 Tahun Mendatang," *Technol. J.*, vol. 1, no. 1, 2024, doi: 10.62872/cwgzp962.
- [20] Y. L. Fatma and N. Rochmawati, "Prediksi Siswa Putus Sekolah Menggunakan Algoritma Decision Tree C4.5," *J. Informatics Comput. Sci.*, vol. 5, no. 04, pp. 486–493, 2024, doi: 10.26740/jinacs.v5n04.p486-493.