

Klasifikasi Profil Kelulusan Nilai AKPAM Dengan Metode Decision Tree C4.5

Dihin Muriyatmoko¹, Aziz Musthafa¹, Muqoddam Husni Wijaya^{1*}

¹Teknik Informatika, Universitas Darussalam Gontor, Ponorogo

Jl. Raya Siman, Dusun I, Demangan, Kec. Siman, Kabupaten Ponorogo, Jawa Timur

*Corresponding Author : mqwijaya48@gmail.com

Abstrak

Universitas Darussalam (UNIDA) Gontor, yang dikelola oleh Pondok Modern Darussalam Gontor, menyajikan pendidikan universitas dengan pendekatan unik. Mahasiswa UNIDA, berinteraksi intensif dengan dosen dalam suasana pesantren, di mana nilai-nilai keilmuan ditanamkan 24 jam sehari. UNIDA Gontor juga menawarkan berbagai organisasi dan kegiatan kemahasiswaan, menciptakan lingkungan dinamis. Angka Kumulatif Penunjang Akademik (AKPAM) digunakan sebagai indikator kinerja mahasiswa, tetapi beberapa siswa mengalami kesulitan mencapai batas kelulusan AKPAM. Pada tahun ajaran 2022-2023, sekitar 700 siswa di semester genap dan 379 siswa di semester ganjil tidak memenuhi syarat AKPAM minimal. Karena tidak ada model kelulusan mahasiswa AKPAM berbasis data, penelitian ini menerapkan metode data mining, terutama Decision Tree C4.5. Penelitian ini bertujuan memudahkan mahasiswa mencapai nilai AKPAM dengan saran kriteria kinerja. Dengan menggunakan Decision Tree C4.5, penelitian melibatkan pengumpulan data, preprocessing data, pembagian data, pembuatan model, uji coba, dan validasi. Dengan Cross Validation sebagai metode pengujian, penelitian mencapai akurasi 99.58%. Dengan proses yang melibatkan preprocessing, split data, pemodelan, dan validasi menggunakan Cross Validation, penelitian ini berhasil mengimplementasikan seluruh proses dari pengumpulan data hingga pembuatan model. Hasil akurasi tinggi membuat model ini dapat diandalkan untuk prediksi atau pengambilan keputusan berdasarkan data yang diolah.

Kata kunci: AKPAM, UNIDA, Data Mining, Decision Tree C4.5

Abstract

Darussalam Gontor University (UNIDA) managed by Pondok Modern Darussalam Gontor, provides university education with a unique approach. UNIDA students, interact intensively with lecturers in a pesantren atmosphere, where scholarly values are instilled 24 hours a day. UNIDA Gontor also offers various student organisations and activities, creating a dynamic environment. The Academic Support Cumulative Score (AKPAM) is used as an indicator of student performance, but some students have difficulty reaching the AKPAM pass mark. In the 2022-2023 academic year, about 700 students in even semesters and 379 students in odd semesters did not meet the minimum AKPAM requirement. Since there is no data-based AKPAM student graduation model, this research applies data mining methods, especially Decision Tree C4.5. This research aims to make it easier for students to achieve AKPAM grades with performance criteria suggestions. Using Decision Tree C4.5, the research involves data collection, data preprocessing, data sharing, model building, testing, and validation. With Cross Validation as the testing method, the research achieved 99.58% accuracy. With a process involving preprocessing, data splitting, modelling, and validation using Cross Validation, this research successfully implemented the entire process from data collection to model building. The high accuracy results make this model reliable for prediction or decision-making based on the processed data.

Keywords : AKPAM,UNIDA, Data Mining, Decision Tree C4.5

PENDAHULUAN

Universitas Darussalam (UNIDA) Gontor, sebuah institusi yang dikelola oleh Pondok Modern Darussalam Gontor, menyajikan pendidikan tinggi dengan pendekatan unik. Mahasiswa UNIDA, atau "Mahasiswa Santri," terlibat dalam interaksi intensif dengan dosen dalam suasana pesantren, di mana nilai-nilai keilmuan disematkan selama 24 jam sehari. Selain kegiatan kursus, UNIDA Gontor menawarkan berbagai organisasi dan kegiatan kemahasiswaan, menciptakan lingkungan yang dinamis. Angka Kumulatif Penunjang Akademik (AKPAM) menjadi alat pemantauan kinerja mahasiswa, tetapi data AKPAM menunjukkan bahwa beberapa siswa mengalami kesulitan mencapai batas kelulusan.

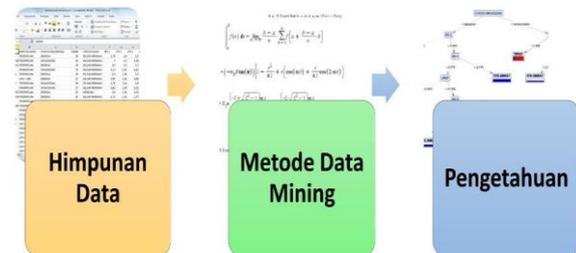
Selain kegiatan kursus, UNIDA Gontor juga menyediakan banyak organisasi dan unit kegiatan kemahasiswaan, menciptakan lingkungan yang dinamis. Angka Kumulatif Penunjang Akademik (AKPAM) merupakan alat pemantauan kinerja yang menjadi tolak ukur kinerja mahasiswa. Sayangnya, data AKPAM menunjukkan masih ada beberapa siswa yang kesulitan mencapai nilai tersebut.

Pada tahun ajaran 2022-2023, sekitar 700 siswa berada di semester genap dan 379 siswa di semester ganjil tidak memenuhi syarat minimal AKPAM. Dengan tidak adanya model kelulusan mahasiswa AKPAM yang berbasis data, maka diperlukan pendekatan baru. Oleh karena itu, penelitian ini akan menggunakan metode data mining khususnya *Decision Tree C4.5* untuk melatih model kelulusan siswa berbasis AKPAM.

Data mining adalah proses otomatis dari data yang ada. Data tersebut akan diolah menjadi bentuk data yang sangat besar, mencari pola atau informasi yang menarik pada data yang dipilih dengan menggunakan teknik atau metode tertentu. Pemilihan metode atau algoritma yang tepat sangat bergantung pada proses dan tujuan *Knowledge Discovery In Database* (KDD) secara keseluruhan (Rahman et al. 2019).

Di balik konsep *Data Mining* adalah pertumbuhan luar biasa dalam volume data yang

dihasilkan oleh aktivitas sehari-hari manusia dan perangkat yang terhubung di seluruh dunia. Data dihasilkan oleh bisnis, sains, media sosial, sensor, dan seringkali mencakup jutaan atau bahkan miliaran entri. Tantangannya adalah bagaimana mengelola, mengatur, dan mengekstraksi wawasan berharga dari data ini.



Gambar 1. Konsep Data Mining

beberapa konsep utama dalam data mining:

1. Pengumpulan Data, Proses data mining dimulai dengan mengumpulkan data dari berbagai sumber seperti database, data warehousing, data streaming, dan lainnya. Data ini bisa terstruktur (misalnya, database relasional) atau tak terstruktur (misalnya, teks, gambar, atau suara).
2. Pemahaman Data, Tahap ini melibatkan pemahaman mendalam terhadap data yang akan dianalisis. Ini mencakup pemahaman terhadap struktur data, karakteristik variabel, dan jenis relasi antar variabel.
3. Pre-processing Data, Data mining seringkali melibatkan pre-processing data untuk membersihkan data dari noise, mengisi nilai yang hilang, dan mengatasi masalah kualitas data. Proses ini memastikan bahwa data yang digunakan untuk analisis adalah data yang berkualitas dan dapat diandalkan.
4. Pemilihan dan Transformasi Variabel, Pada tahap ini, variabel yang paling relevan dan berpengaruh dalam analisis dipilih. Transformasi variabel juga dapat dilakukan untuk mengubah bentuk data atau meningkatkan fitur-fitur tertentu.
5. Pemilihan Model, Memilih model atau algoritma yang sesuai dengan jenis

analisis yang akan dilakukan. Beberapa teknik data mining yang umum digunakan termasuk regresi, clustering, klasifikasi, asosiasi, dan lainnya.

6. Pelatihan Model, Model atau algoritma yang dipilih dilatih menggunakan data yang telah disiapkan. Ini melibatkan penyesuaian parameter model agar dapat menghasilkan prediksi atau hasil yang optimal.
7. Evaluasi Model, Setelah model dilatih, evaluasi dilakukan untuk menilai sejauh mana model dapat menghasilkan prediksi yang akurat. Berbagai metrik evaluasi digunakan tergantung pada jenis masalah data mining yang sedang dihadapi.
8. Penerapan Model, Setelah evaluasi dan penyesuaian, model dapat diterapkan pada data baru untuk menghasilkan prediksi atau menemukan pola baru.
9. Interpretasi dan Visualisasi, Hasil data mining diinterpretasikan dan divisualisasikan agar mudah dimengerti oleh pengambil keputusan. Visualisasi seringkali digunakan untuk menggambarkan pola atau tren dalam data.

Data mining digunakan dalam berbagai bidang, termasuk bisnis, ilmu pengetahuan, kesehatan, keamanan, dan lainnya. Hal ini membantu organisasi untuk mengoptimalkan proses bisnis, meningkatkan pengambilan keputusan, dan mengidentifikasi peluang atau tantangan yang mungkin muncul.

Data Mining melibatkan penerapan berbagai teknik analisis statistik, matematika, dan kecerdasan buatan untuk mengungkap pola dan tren yang mungkin tidak langsung terlihat dalam data. Ini mencakup langkah-langkah seperti pemrosesan data, transformasi data, pemilihan fitur, dan penggunaan algoritma dan model yang sesuai, beberapa teknik yang umum digunakan salah satunya adalah Klasifikasi.

Klasifikasi dalam *Data Mining* adalah suatu teknik yang digunakan untuk mengkategorikan data ke dalam kelas-kelas yang berbeda berdasarkan atribut atau karakteristiknya (Suhendra, Wiminata, and Harmaja 2022). Ini adalah proses menemukan

model atau fungsi yang dapat digunakan untuk memprediksi kelas suatu instance baru berdasarkan atribut dari instance tersebut.

Proses klasifikasi melibatkan pembuatan model menggunakan dataset pelatihan, yaitu kumpulan data berlabel, dan kemudian menggunakan model tersebut untuk memprediksi kelas data baru yang tidak berlabel. Model klasifikasi dapat dibangun dengan menggunakan berbagai algoritma, seperti pohon keputusan, k-nearest neighbours, support vector machine, dan naïve Bayes. Keakuratan model klasifikasi bergantung pada kualitas dataset pelatihan, pilihan algoritma, dan parameter yang digunakan dalam algoritma. Klasifikasi banyak digunakan di berbagai bidang, seperti keuangan, kesehatan, dan pendidikan, untuk membuat prediksi dan mengklasifikasikan data.

Decision Tree merupakan sebuah *Knowledge Discovery In Database* (KDD). Hasil dari metode ini akan mendapatkan sebuah pengetahuan dari database yang diolah (Delisah et al. 2022). Database akan diekstraksi sehingga menghasilkan informasi atau pengetahuan yang berguna. *Decision tree* adalah sebuah model prediktif yang digunakan dalam data mining dan machine learning. Model ini berbentuk seperti pohon, di mana setiap simpul pada pohon mewakili sebuah keputusan berdasarkan nilai dari suatu atribut, dan setiap cabang pada simpul mewakili kemungkinan nilai dari atribut tersebut. Pada akhirnya, model ini menghasilkan prediksi atau klasifikasi berdasarkan serangkaian keputusan yang diambil pada setiap simpul (Amri, Kusri, and Kusnawi 2023).

Decision tree dapat digunakan untuk berbagai macam masalah, seperti klasifikasi, regresi, dan segmentasi data. Beberapa kelebihan dari *decision tree* adalah mudah dipahami dan diinterpretasikan, dapat menangani data yang kompleks, dan dapat digunakan untuk memilih atribut yang paling penting dalam suatu masalah. Namun, *decision tree* juga memiliki kelemahan, seperti rentan terhadap overfitting dan tidak dapat menangani data yang tidak seimbang.

Algoritma C4.5 merupakan algoritma yang digunakan untuk membentuk pohon keputusan dan aturan – aturan (*rule*). Algoritma C4.5 memetakan nilai atribut menjadi kelas yang

dapat diterapkan untuk klasifikasi baru. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 yang digunakan untuk menghasilkan pohon keputusan dari dataset. Algoritma C4.5 dapat mengatasi atribut yang bersifat kontinu dan diskrit, mengatasi data yang hilang, dan melakukan *pruning* pada pohon untuk menghindari *overfitting* (Eirlangga and Syaputra 2022). Dengan menggunakan algoritma *Decision Tree C4.5*, kita dapat membangun model pohon keputusan yang dapat digunakan untuk mengklasifikasikan data berdasarkan atribut-atributnya. Algoritma ini memiliki kelebihan dalam mengatasi atribut kontinu dan diskrit, mengatasi data yang hilang, dan melakukan *pruning* untuk menghindari *overfitting* (Ahadi et al. 2022).

Beberapa rumus yang umum digunakan dalam algoritma ini antara lain:

$$\text{Entropy} = -\sum p_i \log_2(p_i)$$

- a) Rumus Entropy: Entropy digunakan untuk mengukur ketidakpastian atau keberagaman data pada setiap simpul dalam pohon keputusan. di mana p_i adalah probabilitas munculnya nilai i dalam dataset, dan \log_2 adalah logaritma basis 2. Entropy digunakan untuk mengukur seberapa tidak pastinya suatu dataset (Yulianto 2020).

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

- b) Rumus Gain: Gain digunakan untuk memilih atribut terbaik yang akan digunakan sebagai pemisah pada setiap simpul dalam pohon keputusan. Secara ringkas, gain merupakan ukuran seberapa banyak entropi berkurang (informasi yang diperoleh) dengan membagi data berdasarkan atribut tertentu. Atribut dengan gain informasi tertinggi dipilih sebagai atribut pemisah dalam algoritma pohon keputusan C4.5 (Arifianto et al. 2022).

Berdasarkan landasan teori diatas dapat menyimpulkan rumusan masalah Pola pada profil nilai AKPAM mahasiswa Universitas Darussalam Gontor belum pernah diteliti sebelumnya, hal ini menyebabkan ketidaktahuan

kita terhadap profiling mahasiswa yang lulus dan tidak lulus.

Tujuan dari penelitian ini adalah memperoleh hasil klasifikasi profil kelulusan nilai AKPAM dengan metode *Decision Tree C4.5*

METODE PENELITIAN

Penelitian ini memiliki beberapa tahapan metode, layaknya metode pada *Data Mining* berikut tahapan tersebut.

Pengumpulan Data

Untuk data dalam penelitian ini diambil dari sistem Siakad UNIDA Gontor dan data terkait rekapitulasi AKPAM sudah tersedia sehingga hanya perlu mengunduhnya dari sistem Siakad tersebut. Data yang diambil adalah data perolehan nilai di 2 tahun terakhir yaitu tahun 2020 sampai dengan 2022 dengan jumlah total data 10.000 data, data tersebut merupakan data keseluruhan dari semua program studi di Universitas Darussalam Gontor Kampus Siman, dengan beberapa kriteria seperti Kelimuan, Kerohanian, Keorganisasian, Olahraga dan Kesenian, Kemasysrakatan, Pengalaman dan Lomba.

Preprocessing Data

Preprocessing data pada penelitian ini merupakan proses Dimana data diolah terlebih dahulu sebelum di proses dan pemodelan, sehingga butuh beberapa tahapan agar data dapat diolah ke tahapan selanjutnya dengan tujuan data lebih terstruktur (Sugiyono 2016) Adapun tahap-tahap tersebut.

- a) Data Cleaning Tahap ini dilakukan untuk membersihkan data dari noise, missing value, dan inkonsistensi data seperti duplikasi dalam format yang berbeda, kesalahan pada kode nama, dan lain sebagainya (Budimulia and Ridho 2020).
- b) Data Transformation Tahap ini dilakukan untuk mengubah data menjadi "bentuk" yang sesuai dengan metode analisis. Ada beberapa cara untuk melakukan transformasi data, yaitu *normalization*, *attribute selection*,

discretization, dan konsep hierarchy generation.

- c) Data Reduction Tahap ini dilakukan untuk mengurangi jumlah data yang tidak relevan atau tidak diperlukan.

Pembagian Data

Memisahkan dataset menjadi dua subset: data pelatihan (*training data*) dan data pengujian (*testing data*). Data pelatihan digunakan untuk melatih model, sementara data pengujian digunakan untuk menguji performa model.

Pembuatan Model

Pemodelan Decision Tre menggunakan algoritma C4.5 untuk membangun decision tree dari data pelatihan. Algoritma ini bekerja dengan memilih atribut yang paling informatif untuk dijadikan node pada tree dan melakukan pemisahan berdasarkan nilai-nilai atribut tersebut.

Uji Coba dan Validasi

Untuk hasil evaluasi, dapat menggunakan data pelatihan yang sebelumnya sudah dibagi sebelumnya untuk mengevaluasi kinerja model decision tree. Metrik evaluasi yang umum digunakan meliputi akurasi, presisi, recall , F1-score, dan area di bawah kurva ROC (AUC-ROC). Dan untuk penelitian ini saya akan menggunakan *Cross Validation* untuk pengujian.

Cross validation adalah teknik yang umum digunakan dalam machine learning dan data mining untuk mengevaluasi kinerja model(Bertolini, Finch, and Nehm 2021). Ini melibatkan membagi dataset menjadi subset pelatihan dan pengujian secara berulang, sehingga setiap titik data diuji dan dilatih. Salah satu pendekatan cross-validation yang umum adalah 10-fold cross-validation, di mana data dibagi menjadi 10 bagian yang sama besar, di mana model dilatih pada 9 bagian dan diuji pada bagian yang tersisa. Proses ini diulang kembali dengan setiap bagian yang berbeda sebagai data pengujian. Tujuan dari cross-validation adalah untuk memastikan bahwa model yang dilatih tidak hanya baik pada subset tertentu dari data, tetapi umumnya baik pada semua data.

HASIL DAN PEMBAHASAN
Pengumpulan Data

Data yang sudah saya kumpulkan adalah data yang saya peroleh dari sistem Siacad UNIDA Gontor, dengan jumlah total data 10.000 data, yang terdiri dari 16 program studi dengan kriteria data NIM, Nama, Prodi, Keilmuan, Kerohanian, Kesenian dan Olahraga, Keorganisasian, Kemasyarakatan, Pengalaman dan lomba, total serta Keterangan. Dengan kriteria inilah penelitian ini dilakukan.

Gambar 2. Data

Preprocessing Data

Preprocessing data merupakan tahap kritis dalam penelitian data mining yang bertujuan untuk mempersiapkan data sebelum dilakukan analisis lebih lanjut. Dalam, proses preprocessing data seringkali dijelaskan secara rinci untuk memastikan kejelasan dan reproduktibilitas penelitian.dalam penelitian ini saya menggunakan *tools rapidminer* agar memudahkan saya menyelesaikan tahapan proses tersebut. *preprocessing* pada penelitian ini menerapkan beberapa Langkah diantaranya sebagai berikut

- 1) Data Cleaning Tahap ini dilakukan untuk membersihkan data. Untuk itu perlu mengecek agar data tidak ada *data missing* dan untuk hasilnya ialah nihil. Tidak ada *data missing* dalam data yang telah saya kumpulkan.

Gambar 3. Hasil *Data Cleaning*

2) Data Transformation Tahap ini dilakukan untuk mengubah data menjadi "bentuk" yang sesuai dengan metode analisis. Data yang saya ubah diantaranya NIM diubah menjadi id dan bertipe data real, keterangan diubah menjadi label untuk data ini yang bertipe data binominal.

Row No.	Nim (real) id	Keterangan (binominal) label	Prodi (polynomial) regular	Keilmuan (integer) regular	Kerohanian (integer) regular	Kesenian dan Keolahragaan (integer) regular
1	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
2	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
3	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
4	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
5	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
6	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
7	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
8	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
9	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	45	15
10	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
11	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
12	3720161110	LULUS	Pendidikan Agama Isl.	60	45	15
13	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
14	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
15	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0
16	3720161110	BELUM LULUS	Pendidikan Agama Isl.	0	0	0

Gambar 4. Data transformation

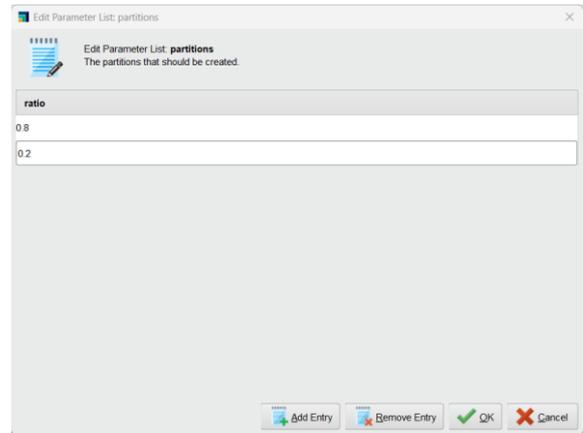
3) Data Reduction Tahap ini dilakukan untuk mengurangi jumlah data yang tidak relevan atau tidak diperlukan. Tahapan yang saya lakukan adalah membuang atribut yang tidak diperlukan seperti Nama, Prodi, dan Total, yang tidak relevan pada penelitian ini. Selanjutnya adalah proses balancing data agar data yang akan kita proses seimbang dan adil.

class	size
LULUS	4291
BELUM LULUS	4291

Gambar 5. Data Balancing

Pembagian Data

Di tahap ini, data yang sebelumnya telah di preprocessing dapat diolah dan dibuat model, hal pertama dilakukan adalah *split data* yaitu pembagian data menjadi 2, *data training* dan *data testing*. Rasio pembagian pada *data training* 80% dan *data testing* 20%.

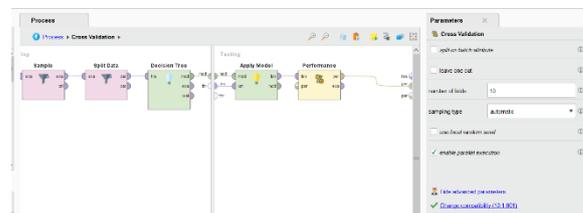


Gambar 6. Pembagian data

Pembuatan Model

Selanjutnya menuju ke tahap modeling *decision tree C4.5*, dengan kriteria *gain ratio* yaitu memilih atribut terbaik dalam proses pengambilan keputusan atau konstruksi pohon keputusan. Gain Ratio mengukur seberapa baik suatu atribut dapat memisahkan data menjadi kelas-kelas yang berbeda. Gain Ratio memperhitungkan besarnya informasi yang diberikan oleh suatu atribut terhadap pemisahan data dan mengkompensasi terhadap atribut yang memiliki banyak nilai yang mungkin.

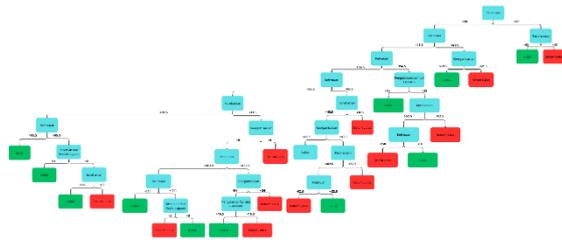
Penelitian ini menggunakan *maximal depth* mencapai 10 dan pemilihan kedalam kriteria *gain ratio* di dalam model *Decision Tree C4.5*



Gambar 7. Model

Uji Coba dan Validasi

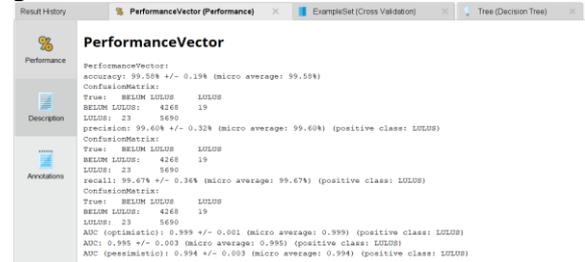
Dengan rangkaian proses yang sudah dilakukan seperti preprocessing data, split data, pemodelan, dan validasi menggunakan *Cross Validation*, uji coba pada penelitian ini mencapai akurasi yang baik dengan angka 99.58%. dan untuk hasil Tree seperti berikut.

Gambar 8. Hasil *Tree*

Tree yang disajikan merupakan representasi dari suatu sistem keputusan atau model klasifikasi yang digunakan untuk menentukan apakah seseorang lulus atau tidak lulus berdasarkan beberapa variabel. Pohon keputusan ini memiliki cabang-cabang yang menggambarkan aturan-aturan berdasarkan nilai variabel-variabel seperti Keilmuan, Kerohanian, dan Keorganisasian. Pada setiap level cabang,

terdapat keputusan atau label yang menunjukkan apakah seseorang dianggap lulus atau tidak lulus.

Lalu untuk performance vector ada pada gambar berikut



Gambar 9. Hasil Uji coba dan Validasi

Performance vector ini memberikan gambaran komprehensif tentang seberapa baik model klasifikasi telah berhasil dalam melakukan prediksi pada dataset yang digunakan.

Tabel 1. Hasil *Cross Validation*

Performance	Score
Accuracy	99.58% +/- 0.19% (micro average: 99.46%)
Precision	99.60% +/- 0.28% (micro average: 99.58%)
Recall	99.67% +/- 0.36% (micro average: 99.54%)

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian ini dapat disimpulkan bahwa Hasil uji coba dan validasi menggunakan Cross Validation menunjukkan bahwa model decision tree C4.5 mencapai akurasi tinggi sebesar 99.58%. Dengan demikian, penelitian ini berhasil mengimplementasikan seluruh proses dari pengumpulan data hingga pembuatan model, dengan hasil akurasi yang tinggi, membuat model ini dapat diandalkan untuk prediksi atau pengambilan keputusan berdasarkan data yang telah diolah.

Berdasarkan kesimpulan penelitian yang telah disampaikan, kita dapat melihat variabel-variabel yang paling berpengaruh terhadap hasil prediksi (lulus atau tidak lulus) adalah sebagai berikut:

1. Keilmuan: Variabel ini memiliki pengaruh yang paling signifikan pada hasil prediksi. *Decision tree*

membagi data pertama kali berdasarkan nilai Keilmuan, menunjukkan bahwa tingkat keilmuan sangat mempengaruhi keputusan apakah seseorang akan lulus atau tidak.

2. Kerohanian: Variabel ini juga memiliki pengaruh yang cukup besar pada hasil prediksi. *Decision tree* membagi cabang-cabang berdasarkan nilai "Kerohanian", menunjukkan bahwa aspek keagamaan atau kerohanian juga penting dalam menentukan apakah seseorang akan lulus atau tidak.
3. Keorganisasian: Meskipun tidak sekuat Keilmuan dan Kerohanian, variabel "Keorganisasian" juga memiliki pengaruh yang cukup besar pada hasil prediksi. Terdapat beberapa pembagian di dalam

decision tree yang bergantung pada nilai "Keorganisasian", menunjukkan bahwa partisipasi dalam kegiatan organisasi atau kegiatan ekstrakurikuler memiliki dampak pada prediksi lulus atau tidak lulus.

Saran untuk pengembangan penelitian ini lebih lanjut dapat mencakup beberapa aspek. Pertama, penting untuk memperluas sampel dengan menambah jumlah data dari periode yang lebih panjang. Hal ini akan meningkatkan generalisabilitas hasil penelitian. Selanjutnya, perlu dipertimbangkan untuk menggunakan metode klasifikasi lain selain *decision tree C4.5*, seperti *Support Vector Machine* atau *Neural Networks*, untuk membandingkan kinerja model secara menyeluruh. Selain itu, penelitian ini telah memberikan pola yang cukup baik terhadap atribut atribut nilai yang ada di penilaian AKPAM, sehingga mahasiswa bisa mencari kegiatan yang sesuai dengan kriteria atribut yang dibutuhkan. Misal, pada nilai keilmuan mahasiswa bisa mengikuti seperti kegiatan seminar dan workshop. Kemudian kerohanian bisa mengikuti kajian atau puasa sunnah seperti senin-kamis atau puasa daud, dan begitu pula untuk atribut nilai lainnya.

DAFTAR PUSTAKA

- Ahadi, Eko, Indra Gunawan, Ika Okta Kirana, Dedy Hartama, and Muhammad Ridwan Lubis. 2022. "Penentuan Keberhasilan Pembelajaran Daring Pada Masa Pandemi Covid-19 Dengan Menggunakan Algoritma C4.5 Di Stikom Tunas Bangsa." *Jurnal Komputer Dan Informatika* 10 (1): 78–85. <https://doi.org/10.35508/jicon.v10i1.6446>.
- Amri, Zaenul, Kusri, and Kusnawi. 2023. "Edumatic : Jurnal Pendidikan Informatika Prediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Naïve" 7 (2): 187–96. <https://doi.org/10.29408/edumatic.v7i2.18620>.
- Arifianto, Aji Seto, Kursita Dewi Safitri, Khafidurrohman Agustianto, and I Gede Wiryawan. 2022. "Pengaruh Prediksi Missing Value Pada Klasifikasi Decision Tree C4.5." *Jurnal Teknologi Informasi Dan Ilmu Komputer* 9 (4): 779–86. <https://doi.org/10.25126/jtiik.2022944778>.
- Bertolini, Roberto, Stephen J. Finch, and Ross H. Nehm. 2021. "Enhancing Data Pipelines for Forecasting Student Performance: Integrating Feature Selection with Cross-Validation." *International Journal of Educational Technology in Higher Education* 18 (1). <https://doi.org/10.1186/s41239-021-00279-6>.
- Budimulia, Wisnu Damar, and Farid Ridho. 2020. "Penerapan Komputasi Paralel Pada Aplikasi Data Cleaning Multiple Data Edit." *Seminar Nasional Official Statistics* 2019 (1): 7–14. <https://doi.org/10.34123/semnasoffstat.v2019i1.120>.
- Delisah, Nining Rahaningsih, Fadhil Muhammad Basysyar, Kaslani, and Abdul Ajiz. 2022. "Pengelompokan Nasabah Surety Bond Menggunakan Algoritma K-Means." *KOPERTIP : Jurnal Ilmiah Manajemen Informatika Dan Komputer* 5 (1): 17–22. <https://doi.org/10.32485/kopertip.v5i1.145>.
- Eirlangga, Yofhanda Septi, and Aldo Eko Syaputra. 2022. "Klasifikasi Penjurusan Pada Sekolah Menengah Atas (SMA) Dengan Metode Algoritma C4.5." *Jurnal Informasi Dan Teknologi* 4 (3): 160–65. <https://doi.org/10.37034/jidt.v4i3.235>.
- Rahman, Fauziah Abdul, Rahimah Kassim, Zirawani Baharum, Helmi Adly Mohd Noor, and Norhaidah Abu Haris. 2019. "Data Cleaning in Knowledge Discovery Database-Data Mining (KDD-DM)." *International Journal of Engineering and Advanced Technology* 8 (6s3): 2196–99. <https://doi.org/10.35940/ijeat.f1100.0986s319>.
- Sugiyono. 2016. "Dokumen Karya Ilmiah | Skripsi | Prodi Teknik Informatika - S1 | FIK | UDINUS | 2016." *Fik* 1 (1): 1–2.
- Suhendra, Fadjar, Flencia Wiminata, and Okta Jaya Harmaja. 2022. "Implementasi Data Mining Dalam Pengkategorian Barang Expedisi Muatan Kapal Laut Pada Pt Maritim Era Sukses Menggunakan

Metode Apriori.” *Jurnal Sistem Informasi Dan Ilmu Komputer Prima (JUSIKOM PRIMA)* 5 (2): 120–26.
<https://doi.org/10.34012/jurnalsisteminformasidanilmukomputer.v5i2.2415>.

Yulianto, Muhamad Arief. 2020.

“Implementasi FIS Sugeno Pada Algoritma C4. 5 Berbasis Particle Swarm Optimization (PSO) Untuk Prediksi Prestasi Siswam.” *JOAIIA: Journal of Artificial Intelligence and Innovative Applications* 1 (1): 12–22.
<http://www.openjournal.unpam.ac.id/index.php/JOAIIA/article/view/4272>.