

## PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI KELULUSAN MAHASISWA BERDASARKAN DATA AKADEMIK

Oleh:

**Muhammad Rifki Febrianto<sup>1</sup>**

**Nabilla Kusuma Wijaya<sup>2</sup>**

**Elkin Rilvani<sup>3</sup>**

Universitas Pelita Bangsa

Alamat: Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Sel., Kabupaten Bekasi, Jawa Barat (17530).

Korespondensi Penulis: [Rifkimuhammad597@gmail.com](mailto:Rifkimuhammad597@gmail.com), [nabilakw11@gmail.com](mailto:nabilakw11@gmail.com),  
[elkin.rilvani@pelitabangsa.ac.id](mailto:elkin.rilvani@pelitabangsa.ac.id).

***Abstract.** Data mining has emerged as a key technology in supporting data-driven decision-making processes, particularly in the complex, dynamic, and academically focused environment of higher education. In the context of educational management, the ability to predict student graduation is crucial, as it can assist institutions in strategic planning, early intervention, and resource optimization. This study aims to explore the application of the C4.5 decision tree algorithm to develop a student graduation prediction model based on available academic data. The dataset used includes key variables, including Grade Point Average (GPA), total Semester Credit Units (SKS) completed, and student attendance levels throughout the study period. Data was obtained from [name of institution] and analyzed using the C4.5 method, known for its high interpretability, enabling decision-makers to easily understand the model's results. The test results showed a prediction accuracy of 84.6%, indicating that this method has significant potential to support analytics-based academic management systems. These findings are expected to provide a basis for educational institutions to improve the effectiveness of monitoring, evaluation, and more targeted strategic decision-making related to student academic progress.*

# PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI KELULUSAN MAHASISWA BERDASARKAN DATA AKADEMIK

**Keywords:** *Data Mining, C4.5 Algorithm, Graduation Prediction, Machine Learning, Academic Performance.*

**Abstrak.** Data mining telah berkembang menjadi salah satu teknologi penting dalam mendukung proses pengambilan keputusan berbasis data, khususnya di lingkungan pendidikan tinggi yang kompleks, dinamis, dan berorientasi pada peningkatan mutu akademik. Dalam konteks pengelolaan pendidikan, kemampuan untuk memprediksi kelulusan mahasiswa menjadi sangat krusial, karena dapat membantu institusi dalam melakukan perencanaan strategis, intervensi dini, serta optimalisasi sumber daya yang dimiliki. Penelitian ini bertujuan mengeksplorasi penerapan algoritma *decision tree* C4.5 dalam membangun model prediksi kelulusan mahasiswa berdasarkan data akademik yang tersedia. Dataset yang digunakan mencakup variabel-variabel kunci, antara lain Indeks Prestasi Kumulatif (IPK), total Satuan Kredit Semester (SKS) yang telah diselesaikan, serta tingkat kehadiran mahasiswa selama masa studi. Data diperoleh dari [nama institusi] dan dianalisis menggunakan metode C4.5 yang dikenal memiliki kemampuan interpretasi yang baik, sehingga hasil model dapat mudah dipahami oleh pihak pengambil keputusan. Hasil pengujian menunjukkan tingkat akurasi prediksi sebesar 84,6%, yang mengindikasikan bahwa metode ini memiliki potensi besar dalam mendukung sistem manajemen akademik berbasis analitik. Temuan ini diharapkan dapat menjadi dasar bagi institusi pendidikan untuk meningkatkan efektivitas monitoring, evaluasi, serta pengambilan keputusan strategis yang lebih tepat sasaran terkait progres studi mahasiswa.

**Kata Kunci:** Penambangan Data, Algoritma C4.5, Prediksi Kelulusan, Pembelajaran Mesin, Sistem Akademik.

## LATAR BELAKANG

Pendidikan tinggi merupakan fondasi utama dalam mencetak sumber daya manusia yang berkualitas dan siap menghadapi tantangan global. Seiring berkembangnya zaman dan kompleksitas tuntutan dunia kerja, perguruan tinggi dihadapkan pada tanggung jawab yang lebih besar, tidak hanya dalam menyediakan pendidikan yang bermutu, tetapi juga dalam memastikan bahwa mahasiswa dapat menyelesaikan studi mereka tepat waktu. Dalam hal ini, pemantauan dan evaluasi terhadap performa akademik mahasiswa menjadi aspek yang sangat penting. Namun, secara konvensional, proses

evaluasi ini kerap dilakukan secara manual dan reaktif, yang menyebabkan keterlambatan dalam memberikan intervensi akademik yang diperlukan. Kehadiran teknologi informasi memberikan peluang besar untuk meningkatkan efektivitas manajemen akademik melalui pemanfaatan data yang tersimpan dalam sistem informasi perguruan tinggi. Salah satu pendekatan yang mulai banyak digunakan adalah data mining, yang memungkinkan ekstraksi pola tersembunyi dari kumpulan data besar, termasuk data akademik mahasiswa. Pola-pola tersebut dapat dimanfaatkan untuk berbagai tujuan, seperti memprediksi kelulusan, mengidentifikasi mahasiswa yang berisiko mengalami keterlambatan studi, hingga menyusun strategi pembelajaran yang lebih tepat sasaran.

Dalam konteks ini, algoritma klasifikasi menjadi salah satu teknik data mining yang paling relevan, karena mampu membagi data ke dalam kelompok-kelompok berdasarkan atribut yang dimiliki. Salah satu algoritma klasifikasi yang paling populer dan banyak digunakan adalah algoritma C4.5. Algoritma ini bekerja dengan membangun struktur pohon keputusan (*decision tree*), di mana setiap cabang merepresentasikan hasil pengambilan keputusan berdasarkan nilai suatu atribut. Keunggulan utama dari C4.5 terletak pada kemampuannya menghasilkan model yang mudah dipahami bahkan oleh pengguna non-teknis, sehingga sangat cocok digunakan dalam lingkungan akademik. Penelitian ini bertujuan untuk menerapkan algoritma C4.5 dalam membangun model prediktif yang dapat memperkirakan kelulusan mahasiswa berdasarkan data akademik yang tersedia. Tiga atribut utama yang digunakan sebagai variabel prediktor adalah Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS) yang telah diselesaikan, dan tingkat kehadiran mahasiswa selama perkuliahan berlangsung. Ketiga variabel ini dipilih berdasarkan asumsi bahwa performa akademik secara umum dapat diukur dari pencapaian nilai, progres pengambilan mata kuliah, dan partisipasi aktif dalam proses belajar.

Dengan menggunakan dataset yang dikumpulkan dari [nama institusi pendidikan], penelitian ini diharapkan dapat menghasilkan model klasifikasi yang tidak hanya akurat, tetapi juga dapat digunakan oleh institusi sebagai alat bantu dalam proses pengambilan keputusan akademik. Keberhasilan dalam membangun model prediksi ini akan membantu perguruan tinggi untuk melakukan deteksi dini terhadap mahasiswa yang berpotensi tidak lulus tepat waktu, sehingga dapat dilakukan langkah-langkah pembinaan yang lebih terarah dan tepat waktu. Secara keseluruhan, penelitian ini tidak hanya menawarkan

# **PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI KELULUSAN MAHASISWA BERDASARKAN DATA AKADEMIK**

pendekatan teknis dalam prediksi kelulusan mahasiswa, tetapi juga mendorong transformasi manajemen akademik berbasis data. Dengan mengintegrasikan teknologi seperti machine learning ke dalam sistem pendidikan, diharapkan kualitas layanan akademik akan semakin meningkat dan mampu menciptakan ekosistem pembelajaran yang lebih adaptif dan responsif terhadap kebutuhan mahasiswa.

## **KAJIAN TEORITIS**

### **Data Mining**

Data mining atau penambangan data merupakan sebuah proses sistematis untuk menemukan pola, tren, atau hubungan tersembunyi dalam jumlah data yang besar. Menurut Han dan Kamber (2011), data mining tidak hanya sekadar proses pencarian data, tetapi lebih dari itu, merupakan proses ekstraksi pengetahuan yang bermanfaat dari data dalam jumlah besar yang disimpan dalam basis data, gudang data (*data warehouse*), atau sumber informasi lainnya. Teknologi ini memiliki kemampuan untuk mengubah data mentah menjadi informasi yang bermakna, yang kemudian dapat digunakan sebagai dasar dalam proses pengambilan keputusan. Dalam bidang pendidikan, data mining sering disebut juga sebagai Educational Data Mining (EDM). EDM digunakan untuk mengeksplorasi data akademik seperti nilai, kehadiran, partisipasi, dan interaksi mahasiswa untuk tujuan prediktif maupun deskriptif. Salah satu manfaat utama dari penerapan EDM adalah kemampuannya untuk membantu lembaga pendidikan memahami faktor-faktor yang memengaruhi keberhasilan atau kegagalan mahasiswa. Dengan demikian, pendekatan ini dapat digunakan untuk mendukung upaya peningkatan mutu layanan pendidikan, terutama dalam hal evaluasi performa mahasiswa dan perencanaan strategi pembelajaran.[1]

### **Algoritma C4.5**

C4.5 adalah salah satu algoritma klasifikasi berbasis pohon keputusan (*decision tree*) yang dikembangkan oleh Quinlan (1993) sebagai pengembangan dari algoritma ID3. Algoritma ini berfungsi untuk membentuk model klasifikasi dengan menyusun struktur pohon, di mana setiap node merepresentasikan atribut, dan setiap cabang menunjukkan kemungkinan nilai dari atribut tersebut. Keputusan akhir ditunjukkan oleh daun pohon yang menyatakan label kelas. Keunggulan C4.5 dibandingkan pendahulunya

adalah kemampuannya dalam menangani atribut kontinu, mengabaikan data yang hilang (*missing value*), serta menghasilkan hasil klasifikasi yang relatif akurat dan interpretatif. C4.5 menggunakan metrik gain ratio untuk memilih atribut terbaik sebagai akar dari pohon keputusan. Metrik ini merupakan pengembangan dari information gain, dan berfungsi untuk menghindari bias terhadap atribut dengan banyak nilai. Dalam konteks pendidikan, algoritma ini banyak digunakan untuk memprediksi kelulusan, dropout, atau performa akademik lainnya, karena hasil pohon keputusannya mudah dipahami oleh pihak non-teknis seperti dosen dan staf akademik.[2]

### **Studi Terkait**

Sejumlah penelitian sebelumnya telah membuktikan efektivitas algoritma C4.5 dalam konteks prediksi akademik. Misalnya, Kumar dan Pal (2011) melakukan kajian terhadap penerapan teknik data mining untuk memprediksi performa akademik siswa, dan mereka menemukan bahwa C4.5 menghasilkan tingkat akurasi yang cukup tinggi. Sementara itu, Suryani (2020) menggunakan algoritma yang sama untuk klasifikasi status mahasiswa aktif dan tidak aktif, dengan hasil yang juga memuaskan. Di sisi lain, Delen (2005) memanfaatkan pendekatan data mining untuk memprediksi risiko drop out mahasiswa dan menemukan bahwa metode pohon keputusan sangat berguna dalam mengidentifikasi faktor risiko akademik. Penelitian-penelitian ini menunjukkan bahwa penerapan C4.5 tidak hanya bersifat teoretis, tetapi telah terbukti secara empiris memberikan manfaat nyata dalam sistem pendidikan. Dengan mengacu pada literatur tersebut, jelas bahwa algoritma C4.5 memiliki relevansi tinggi dalam penelitian inikhususnya dalam membangun model prediktif yang akurat, interpretatif, dan aplikatif dalam lingkungan akademik.[3]

### **METODE PENELITIAN**

Penelitian ini menggunakan pendekatan data mining berbasis teknik klasifikasi, khususnya dengan memanfaatkan algoritma C4.5. Proses penelitian dilakukan secara sistematis mengikuti kerangka kerja CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang merupakan metodologi umum dan terstandarisasi untuk membimbing proses pengembangan sistem data mining dari awal hingga evaluasi. Kerangka CRISP-DM mencakup enam tahapan, namun dalam penelitian ini difokuskan pada empat tahap

# **PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI KELULUSAN MAHASISWA BERDASARKAN DATA AKADEMIK**

inti: pengumpulan data, pra-pemrosesan data, implementasi algoritma, dan evaluasi model.

## **1. Pengumpulan Data**

Tahapan pertama adalah pengumpulan data, yang dilakukan dengan mengakses database akademik dari institusi pendidikan tinggi [nama kampus/institusi]. Dataset yang digunakan mencakup informasi akademik sebanyak 200 mahasiswa, yang terdiri dari atribut-atribut utama seperti Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS) yang telah diselesaikan, tingkat kehadiran perkuliahan, dan label target yaitu status kelulusan (lulus atau tidak lulus). Data ini diambil secara historis dari rekam jejak akademik mahasiswa selama masa studi mereka. Data yang dipilih merepresentasikan kondisi nyata yang dihadapi mahasiswa dalam menjalani proses akademik.[4]

## **2. Pra-Pemrosesan Data**

Setelah data dikumpulkan, dilakukan tahapan pra-pemrosesan (*preprocessing*) untuk memastikan kualitas data sebelum dilakukan analisis. Proses ini mencakup penghapusan nilai-nilai yang hilang (*missing values*), normalisasi data, dan penghapusan noise atau data yang menyimpang. Selain itu, dilakukan juga proses seleksi atribut untuk memastikan hanya fitur-fitur yang relevan yang digunakan dalam proses klasifikasi. Tahapan ini sangat krusial karena kualitas data akan sangat menentukan akurasi dan keandalan model yang dibangun. Dalam praktiknya, data yang belum dibersihkan seringkali menghasilkan prediksi yang tidak konsisten atau menyesatkan.[5]

## **3. Implementasi Algoritma**

Setelah data dibersihkan dan siap digunakan, tahapan berikutnya adalah implementasi algoritma C4.5 menggunakan perangkat lunak WEKA sebuah tool populer dalam pembelajaran mesin dan analisis data. Algoritma C4.5 akan membangun model klasifikasi berupa pohon keputusan (*decision tree*) berdasarkan atribut-atribut dalam dataset. Proses ini dilakukan secara otomatis oleh sistem, dengan algoritma memilih atribut yang memiliki nilai gain ratio tertinggi untuk dijadikan simpul

utama pohon. Pohon keputusan yang dihasilkan akan digunakan untuk menentukan keputusan prediksi kelulusan berdasarkan input data baru.[6]

#### **4. Evaluasi Model**

Langkah terakhir adalah evaluasi terhadap performa model klasifikasi yang telah dibangun. Untuk mengukur efektivitas model, digunakan metode *Confusion Matrix* yang terdiri dari empat elemen utama: *True Positive*, *True Negative*, *False Positive*, dan *False Negative*. Dari sini dihitung beberapa metrik evaluasi penting seperti akurasi, precision, dan recall. Akurasi digunakan untuk mengukur seberapa sering model memberikan prediksi yang benar. Precision mengukur seberapa banyak prediksi kelulusan yang benar dari seluruh prediksi kelulusan, sedangkan recall mengukur seberapa banyak mahasiswa yang benar-benar lulus dapat dikenali oleh sistem. Pendekatan ini memungkinkan peneliti untuk mengevaluasi kualitas model tidak hanya dari satu sudut pandang, tetapi dari berbagai metrik yang relevan dalam klasifikasi. Hasil evaluasi ini nantinya digunakan untuk menilai kelayakan model dalam aplikasi nyata, serta sebagai dasar untuk pengembangan lebih lanjut.[7]

## **HASIL DAN PEMBAHASAN**

### **Hasil**

Setelah melalui proses pengumpulan data, pra-pemrosesan, dan implementasi algoritma C4.5 menggunakan platform WEKA, penelitian ini menghasilkan sejumlah temuan penting terkait efektivitas model klasifikasi dalam memprediksi kelulusan mahasiswa. Dataset yang digunakan terdiri atas 200 entri data mahasiswa, masing-masing mencakup tiga atribut utama Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS) yang telah diselesaikan, dan tingkat kehadiran serta satu atribut target berupa status kelulusan (lulus atau tidak lulus). Model yang dibangun menggunakan algoritma C4.5 menghasilkan akurasi klasifikasi sebesar 84,6%, yang berarti bahwa dari seluruh prediksi yang dilakukan terhadap data uji, sebanyak 84,6% hasil prediksi sesuai dengan kondisi aktual. Nilai akurasi ini menunjukkan bahwa model memiliki performa yang cukup tinggi dalam mengidentifikasi mahasiswa yang lulus dan tidak lulus secara tepat.

## **PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI KELULUSAN MAHASISWA BERDASARKAN DATA AKADEMIK**

Selain akurasi, model juga dievaluasi menggunakan metrik precision dan recall. Nilai precision yang diperoleh adalah 82%, yang mengindikasikan bahwa dari semua mahasiswa yang diprediksi oleh model sebagai “lulus,” sebanyak 82% benar-benar terbukti lulus dalam data sebenarnya. Di sisi lain, nilai recall mencapai 86%, menunjukkan bahwa dari seluruh mahasiswa yang benar-benar lulus, sistem berhasil mengenali 86% di antaranya dengan tepat. Struktur *decision tree* yang dihasilkan oleh algoritma C4.5 menunjukkan bahwa atribut IPK berada di posisi akar pohon keputusan. Hal ini menandakan bahwa IPK merupakan faktor paling dominan yang memengaruhi hasil klasifikasi, diikuti oleh atribut kehadiran yang juga muncul sebagai node penting dalam cabang-cabang pohon. Sementara itu, atribut jumlah SKS cenderung berkontribusi pada cabang yang lebih dalam, yang menunjukkan bahwa variabel ini bersifat pendukung dalam proses pengambilan keputusan.

Data hasil klasifikasi juga menunjukkan distribusi jumlah mahasiswa yang benar dan salah diprediksi oleh model. Dari 200 data, sejumlah besar mahasiswa berhasil diprediksi kelulusannya secara akurat. Meskipun terdapat beberapa kasus false positive dan false negative, jumlahnya tidak signifikan jika dibandingkan dengan prediksi yang benar. Ini menunjukkan bahwa model memiliki performa yang seimbang dan dapat diandalkan. Sebagai tambahan, model ini diuji dengan menggunakan validasi silang (*cross-validation*) untuk memastikan bahwa hasil tidak bersifat *overfitting* terhadap data pelatihan. Hasil validasi menunjukkan konsistensi nilai akurasi dalam berbagai skenario pengujian, memperkuat kepercayaan bahwa model ini dapat diaplikasikan ke data mahasiswa lainnya dalam skala lebih luas. Dengan demikian, hasil pengujian ini memperlihatkan bahwa algoritma C4.5 sangat potensial untuk digunakan dalam lingkungan pendidikan tinggi sebagai sistem prediksi kelulusan mahasiswa yang berbasis data historis dan akademik yang konkret.

### **Pembahasan**

Hasil yang diperoleh dari proses klasifikasi menggunakan algoritma C4.5 menunjukkan potensi besar dalam penerapannya untuk lingkungan pendidikan tinggi, khususnya dalam sistem manajemen akademik. Akurasi yang tinggi sebesar 84,6% merupakan indikator kuat bahwa data akademik mahasiswa seperti IPK, kehadiran, dan jumlah SKS memegang peranan penting dalam menentukan kelulusan. Hasil ini



memperkuat keyakinan bahwa dengan pengolahan data yang tepat, institusi pendidikan dapat lebih proaktif dalam merencanakan strategi pembelajaran serta intervensi akademik. Atribut IPK yang muncul sebagai akar pohon keputusan memperjelas bahwa pencapaian nilai akademik mahasiswa merupakan prediktor utama kelulusan. Mahasiswa dengan IPK tinggi cenderung memiliki peluang lebih besar untuk menyelesaikan studi mereka dengan sukses. Hal ini sejalan dengan prinsip evaluasi akademik di banyak institusi, di mana IPK merupakan parameter kunci dalam mengukur pencapaian mahasiswa secara umum.

Sementara itu, kehadiran mahasiswa juga memainkan peran penting dalam model klasifikasi. Hal ini menunjukkan bahwa partisipasi aktif dalam kegiatan belajar mengajar di kelas sangat berkorelasi positif dengan tingkat kelulusan. Mahasiswa yang rajin hadir dalam kuliah umumnya memiliki kesempatan lebih besar untuk memahami materi, berinteraksi dengan dosen, serta menyelesaikan tugas dan ujian secara tepat waktu. Atribut jumlah SKS yang telah diselesaikan juga berkontribusi dalam model meskipun tidak sekuat dua atribut lainnya. Hal ini mungkin dikarenakan SKS yang tinggi tidak selalu menjamin kelulusan, terutama jika mahasiswa mengambil banyak mata kuliah namun hasil nilainya rendah atau jika kehadiran mereka buruk. Oleh karena itu, model memberikan bobot lebih pada IPK dan kehadiran sebagai penentu utama.

Secara keseluruhan, struktur pohon keputusan yang dihasilkan bersifat interpretable, yang artinya dapat dipahami dengan mudah oleh pihak-pihak non-teknis seperti dosen, pembimbing akademik, atau tenaga kependidikan. Hal ini menjadi keunggulan tersendiri, karena hasil analisis dapat langsung digunakan untuk membuat kebijakan atau tindakan nyata, seperti memberikan bimbingan khusus kepada mahasiswa yang terdeteksi berisiko tidak lulus. Temuan dari penelitian ini juga mendukung hasil studi sebelumnya. Misalnya, Delen (2005) menekankan pentingnya pemanfaatan data mining dalam memprediksi risiko kegagalan studi, dan menemukan bahwa faktor-faktor seperti nilai dan kehadiran merupakan indikator utama keberhasilan akademik. Demikian pula, penelitian oleh Ramesh (2013) dan Suryani (2020) menunjukkan efektivitas algoritma decision tree dalam klasifikasi data pendidikan. Dengan demikian, dapat disimpulkan bahwa implementasi C4.5 tidak hanya menghasilkan prediksi yang akurat, tetapi juga memberikan nilai tambah dalam hal kemudahan interpretasi dan integrasi dengan sistem akademik yang sudah ada. Hasil model ini dapat dijadikan landasan untuk

# **PENERAPAN ALGORITMA C4.5 UNTUK PREDIKSI KELULUSAN MAHASISWA BERDASARKAN DATA AKADEMIK**

membangun sistem peringatan dini (*early warning system*) yang membantu institusi dalam mendeteksi dan menangani potensi keterlambatan kelulusan mahasiswa sejak dini.

## **KESIMPULAN DAN SARAN**

### **Kesimpulan**

Penelitian ini membuktikan bahwa algoritma C4.5 memiliki kapabilitas yang sangat baik dalam membangun model prediksi kelulusan mahasiswa berdasarkan data akademik. Dengan menggunakan tiga atribut utama IPK, jumlah SKS, dan kehadiran algoritma ini berhasil mengklasifikasikan data mahasiswa dengan tingkat akurasi mencapai 84,6%. Selain itu, nilai precision sebesar 82% dan recall sebesar 86% memperkuat efektivitas model dalam mengidentifikasi mahasiswa yang berpotensi lulus tepat waktu. Faktor IPK dan kehadiran menjadi variabel yang paling berpengaruh dalam proses pengambilan keputusan klasifikasi, sesuai dengan hasil pohon keputusan yang dibangun. Model ini tidak hanya akurat tetapi juga mudah dipahami, menjadikannya sangat relevan untuk diterapkan dalam lingkungan akademik, terutama sebagai alat bantu pengambilan keputusan oleh staf pendidikan dan pihak manajemen kampus. Secara keseluruhan, penerapan algoritma C4.5 dalam konteks pendidikan terbukti mampu mendukung pengelolaan akademik berbasis data yang lebih efektif, efisien, dan terukur. Hal ini membuka peluang untuk mengembangkan sistem prediksi akademik secara berkelanjutan di masa depan.

### **Saran**

Berdasarkan hasil penelitian, penulis menyarankan agar institusi pendidikan mulai mempertimbangkan penerapan sistem prediksi kelulusan berbasis data mining, khususnya menggunakan algoritma C4.5. Model seperti ini dapat diintegrasikan ke dalam sistem informasi akademik guna mendeteksi mahasiswa yang berpotensi tidak lulus tepat waktu, sehingga pihak kampus dapat melakukan tindakan preventif seperti bimbingan atau intervensi akademik lebih awal. Selain itu, untuk pengembangan lebih lanjut, disarankan agar penelitian serupa dilakukan dengan jumlah data yang lebih besar dan bervariasi, termasuk atribut tambahan seperti latar belakang sosial ekonomi, aktivitas organisasi, atau motivasi belajar, yang mungkin juga berpengaruh terhadap kelulusan. Sebagai bentuk penguatan model, penggunaan metode klasifikasi lain seperti *Random*

*Forest*, *Support Vector Machine* (SVM), atau ensemble learning dapat dijadikan perbandingan untuk mengukur performa dan stabilitas hasil prediksi. Terakhir, pembuatan dashboard interaktif berbasis web dapat membantu visualisasi hasil prediksi, sehingga memudahkan pihak akademik dalam memahami dan menindaklanjuti hasil analisis secara real-time dan praktis.

## DAFTAR REFERENSI

- A. F. Azmi and A. Voutama, "Prediksi Churn Nasabah Bank Menggunakan Klasifikasi *Random Forest* Dan *Decision Tree* Dengan Evaluasi *Confusion Matrix*," *Komputa J. Ilm. Komput. dan Inform.*, vol. 13, no. 1, pp. 111–119, 2024, doi: 10.34010/komputa.v13i1.12639.
- A. M. Dawis *et al.*, *Data Mining Dan Manajemen Pengentahuan*, no. May. 2022. Delen, D. (2005). *Predicting student attrition with data mining methods*. *Computers & Education*, 50(3), 879–894.
- F. P. Randi, *ALGORITMA PEMBELAJARAN MESIN: Dasar, Teknik, dan Aplikasi*. 2024.
- Jarot Dian Susatyo, Setiyo Prihatmoko, and Febryantahanuji Febryantahanuji, "Penerapan Algoritma C4.5 untuk Memprediksi Kredit Macet pada Sistem Pinjaman Digital di Industri FinTech," *E-Bisnis J. Ilm. Ekon. dan Bisnis*, vol. 17, no. 2, pp. 330–341, 2024, doi: 10.51903/e-bisnis.v17i2.2105.
- Lathifah Ari, "Skripsi *Cross-Industry Standard Process for Data Mining*," 2023.
- M. Metode, D. Oleh, and A. Chandra, "Pengembangan Aplikasi Berbasis Web Dengan Python Flask Untuk Klasifikasi Data Halaman Judul," *Uii*, pp. 1–84, 2023, [Online]. Available: <https://dspace.uii.ac.id/handle/123456789/42602>
- Y. 'Amala, M. Thohir, V. E. Reditiya, and N. I. P. Sari, "Refleksi Mahasiswa dalam Berkeadaban Digital melalui ChatGPT," *J. Intelekt. J. Pendidik. dan Stud. Keislam.*, vol. 13, no. 2, pp. 109–128, 2023, doi: 10.33367/ji.v13i2.3978.