

Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree

Stefanus Alvian Setiono^{1*}, Eko Purwanto²

¹ Program Studi Teknik Informatika, Fakultas
Ilmu Komputer
Universitas Duta Bangsa Surakarta
^{1*} retzstyle@gmail.com

¹ Program Studi Teknik Informatika,
Fakultas Ilmu Komputer
Universitas Duta Bangsa Surakarta
² eko_purwanto@udb.ac.id

Abstrak— Tujuan dari penelitian ini adalah merancang sebuah model prediksi kelulusan mahasiswa menggunakan algoritma Decision Tree (C4.5) untuk mengidentifikasi secara dini mahasiswa yang berisiko mengalami keterlambatan studi. Metode yang digunakan adalah teknik klasifikasi data mining terhadap data sekunder mahasiswa Fakultas Ilmu Komputer Universitas Duta Bangsa Surakarta angkatan 2021-2024. Analisis dilakukan terhadap empat atribut kunci, yaitu Indeks Prestasi Kumulatif (IPK), jumlah SKS, persentase kehadiran, dan semester aktif. Model yang dibangun kemudian dievaluasi kinerjanya menggunakan metode 10-Fold Cross-Validation. Hasil penelitian menunjukkan model prediksi yang dibangun mampu mencapai tingkat akurasi 80% pada data uji. Atribut Kehadiran teridentifikasi sebagai faktor prediktor paling signifikan, diikuti oleh IPK dan progres perolehan SKS. Model ini berhasil diimplementasikan menjadi sebuah aplikasi web interaktif yang berfungsi sebagai sistem peringatan dini (early warning system) untuk mendukung intervensi akademik yang lebih cepat dan efektif.

Kata kunci— Decision Tree, Prediksi Kelulusan, Data Mining, Machine Learning

Abstract— The objective of this research is to design a student graduation prediction model using the Decision Tree (C4.5) algorithm to identify students at risk of study delays at an early stage. The method used is a data mining classification technique on secondary data of students from the Faculty of Computer Science at Universitas Duta Bangsa Surakarta, from the 2021 to 2024 cohorts. The analysis was conducted on four key attributes: Grade Point Average (GPA), number of completed credits (SKS), attendance percentage, and the active semester. The developed model was then evaluated for its performance using the 10-Fold Cross-Validation method. The results of the study show that the prediction model achieved an accuracy rate of 80% on the test data. The Attendance attribute was identified as the most significant predictive factor, followed by GPA and the progression of SKS. This model has been successfully implemented into an interactive web application that functions as an early warning system to support faster and more effective academic interventions.

Keywords— Decision Tree, Graduation Prediction, Data Mining, Machine Learning

I. PENDAHULUAN

Peningkatan efektivitas pendidikan tinggi merupakan salah satu tujuan utama perguruan tinggi dalam mencetak lulusan yang kompeten dan tepat waktu [1]. Salah satu indikator keberhasilan institusi adalah tingkat kelulusan mahasiswanya. Namun, dalam realitasnya, banyak mahasiswa mengalami keterlambatan dalam menyelesaikan studi, yang dipengaruhi oleh berbagai faktor. Beberapa penelitian terdahulu secara konsisten menunjukkan bahwa faktor akademik seperti Indeks Prestasi Kumulatif (IPK), jumlah Satuan Kredit Semester (SKS), dan tingkat kehadiran menjadi prediktor utama dalam menentukan keberhasilan studi mahasiswa [2], [3].

Seiring perkembangan teknologi informasi, pemanfaatan teknik *Data Mining* telah menjadi solusi inovatif untuk membantu analisis data akademik secara lebih efektif [4]. *Data mining* sendiri merupakan proses untuk mengekstraksi pola atau informasi bermakna dari kumpulan data besar dengan berbagai metode, termasuk algoritma kecerdasan buatan [5]. Dalam dunia pendidikan, pendekatan ini membantu mengidentifikasi risiko keterlambatan kelulusan atau putus kuliah secara lebih dini [6].

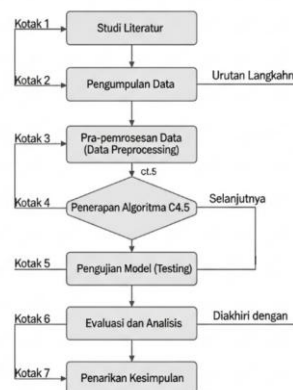
Salah satu teknik yang populer dalam proses klasifikasi data adalah algoritma *Decision Tree*. Algoritma ini dipilih karena kemampuannya mengidentifikasi pola dari data historis mahasiswa secara akurat [2], serta menghasilkan model dengan interpretasi yang mudah dipahami oleh pihak akademik [7]. Meskipun teknologi ini sudah tersedia, di banyak perguruan tinggi, sistem pendeteksian dini terhadap mahasiswa berisiko masih bersifat manual dan kurang komprehensif, sehingga intervensi akademik seringkali terlambat dilakukan [8].

Oleh karena itu, diperlukan sebuah sistem berbasis data mining yang mampu mengolah berbagai atribut akademik untuk memprediksi kelulusan secara sistematis. Penelitian ini mengusulkan model prediksi kelulusan mahasiswa berbasis algoritma *Decision Tree* (C4.5). Model ini diharapkan dapat menjadi alat bantu bagi pihak akademik untuk mengidentifikasi mahasiswa yang berisiko tidak

lulus tepat waktu, sehingga memungkinkan adanya intervensi dini yang lebih cepat dan efektif [9]. Dengan adanya sistem prediksi ini, perguruan tinggi dapat meningkatkan efektivitas pembelajaran dan memperbaiki tingkat kelulusan mahasiswa secara keseluruhan [1].

II. METODOLOGI PENELITIAN

Penelitian ini menjelaskan secara rinci tentang tahapan penelitian yang dilakukan. Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis untuk memastikan tujuan riset tercapai, yang diilustrasikan pada Gambar 1. Tahapan dimulai dengan studi literatur untuk mengkaji penelitian-penelitian sebelumnya yang relevan, dilanjutkan dengan pengumpulan data akademik mahasiswa. Data yang terkumpul kemudian melewati tahap pra-pemrosesan untuk membersihkan dan menyiapkan data agar sesuai untuk analisis. Selanjutnya, dilakukan penerapan algoritma C4.5 untuk membangun model prediksi, yang kemudian diuji kinerjanya menggunakan data uji. Tahap berikutnya adalah evaluasi dan analisis kinerja model secara mendalam, dan diakhiri dengan penarikan kesimpulan serta pemberian saran untuk pengembangan di masa depan.



Gambar 1. Diagram Alir Penelitian

Metode pengumpulan data yang digunakan adalah metode dokumentasi. Data yang diolah merupakan data sekunder yang diperoleh dari Bagian Administrasi Akademik dan Kemahasiswaan (BAAK) Universitas Duta Bangsa Surakarta. Dataset ini mencakup catatan akademik mahasiswa Fakultas Ilmu Komputer dari angkatan 2021 hingga 2024. Objek penelitian ini adalah sebuah model prediksi yang

bertujuan untuk mengklasifikasikan status kelulusan mahasiswa ke dalam dua kategori: “Tepat Waktu” bagi yang menyelesaikan studi kurang dari atau sama dengan 8 semester, dan “Terlambat” bagi yang memerlukan waktu lebih dari 8 semester. Variabel yang digunakan terdiri dari empat variabel independen (atribut), yaitu Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah ditempuh, rata-rata persentase kehadiran, dan posisi semester aktif mahasiswa. Adapun variabel dependen (label) yang menjadi target prediksi adalah Status Kelulusan.

Metode pengumpulan data yang digunakan adalah metode dokumentasi. Data yang diolah merupakan data sekunder yang diperoleh dari Bagian Administrasi Akademik dan Kemahasiswaan (BAAK) Universitas Duta Bangsa Surakarta. Dataset ini mencakup catatan akademik mahasiswa Fakultas Ilmu Komputer dari angkatan 2021 hingga 2024. Objek penelitian ini adalah sebuah model prediksi yang bertujuan untuk mengklasifikasikan status kelulusan mahasiswa ke dalam dua kategori: “Tepat Waktu” bagi yang menyelesaikan studi kurang dari atau sama dengan 8 semester, dan “Terlambat” bagi yang memerlukan waktu lebih dari 8 semester. Variabel yang digunakan terdiri dari empat variabel independen (atribut), yaitu Indeks Prestasi Kumulatif (IPK), jumlah SKS yang telah ditempuh, rata-rata persentase kehadiran, dan posisi semester aktif mahasiswa. Adapun variabel dependen (label) yang menjadi target prediksi adalah Status Kelulusan.

Penerapan Algoritma Decision Tree (C4.5) dilakukan sebagai metode analisis data dengan teknik klasifikasi dari data mining. Algoritma C4.5 dipilih karena kemampuannya yang efektif dalam membangun sebuah pohon keputusan (decision tree) dari dataset latihan. Pohon keputusan adalah sebuah model klasifikasi dengan struktur menyerupai diagram alir, di mana setiap simpul (node) internal merepresentasikan sebuah tes pada suatu atribut, setiap cabang merepresentasikan hasil dari tes tersebut, dan setiap daun (leaf node) merepresentasikan label kelas (keputusan).

Proses utama dalam algoritma C4.5 adalah pemilihan atribut terbaik pada setiap simpul

untuk memecah data menjadi himpunan bagian yang lebih murni. Pemilihan ini didasarkan pada **Gain Ratio** tertinggi. Langkah-langkahnya adalah sebagai berikut:

1. **Menghitung Nilai Entropi:** Entropi digunakan untuk mengukur tingkat ketidakpastian atau keragaman (impurity) dari sebuah kumpulan data. Nilai entropi yang tinggi menandakan dataset memiliki kelas yang sangat beragam, sedangkan nilai 0 menandakan dataset sudah murni (hanya berisi satu kelas).
2. **Menghitung Information Gain:** Setelah nilai entropi dihitung, langkah selanjutnya adalah menghitung Information Gain untuk setiap atribut. Metrik ini mengukur seberapa besar pengurangan entropi yang terjadi setelah dataset dipisah berdasarkan atribut tersebut. Atribut dengan Information Gain tertinggi dianggap paling baik untuk memisahkan data.
3. **Menghitung Gain Ratio:** Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Salah satu kelemahannya, Information Gain cenderung memilih atribut yang memiliki banyak nilai unik. Untuk mengatasi bias ini, C4.5 menggunakan **Gain Ratio**, yang merupakan normalisasi dari Information Gain.
4. **Membangun Pohon:** Atribut dengan nilai Gain Ratio tertinggi dipilih sebagai simpul akar (root node). Kemudian, algoritma ini secara rekursif (berulang) akan memecah dataset menjadi subset-subset yang lebih kecil dan mengulangi langkah 1-3 untuk setiap subset hingga semua data dalam satu cabang memiliki kelas yang sama atau tidak ada lagi atribut yang bisa dipilih.

Pengujian model dilakukan menggunakan metode *10-Fold Cross-Validation*. Metode ini dipilih karena terbukti memberikan estimasi kinerja model yang lebih stabil dan tidak bias dibandingkan dengan metode pembagian data tunggal (*hold-out*) [11]. Kinerja model

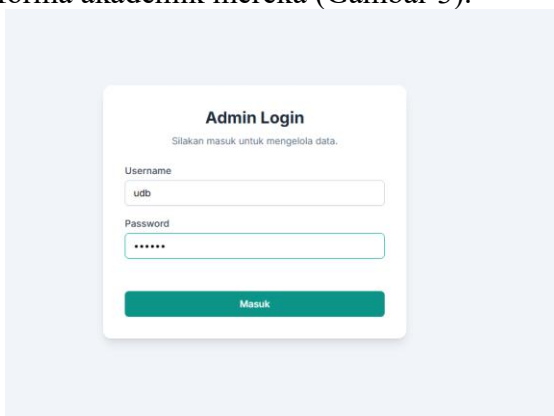
dievaluasi berdasarkan *Confusion Matrix* untuk menghitung tiga metrik utama: Akurasi, Presisi, dan *Recall* [12].

III. HASIL DAN PEMBAHASAN

Pada Hasil dan Pembahasan ini menyajikan hasil dari implementasi sistem prediksi dan analisis kinerjanya, serta menguraikan arti penting dari temuan penelitian. Hasil penelitian disajikan secara jelas dan padat, diikuti dengan diskusi yang mendalam.

A. Hasil Implementasi dan Kinerja Model

Sistem prediksi kelulusan mahasiswa berhasil diimplementasikan dalam bentuk aplikasi web interaktif. Sistem ini dirancang dengan dua hak akses utama: halaman admin untuk pengelolaan data mahasiswa (Gambar 2), dan halaman prediksi utama yang ditujukan bagi mahasiswa untuk melakukan pengecekan potensi kelulusan secara mandiri dengan memasukkan data performa akademik mereka (Gambar 3).



Gambar 2. Halaman Admin

| MAHASISWA | IPK | SKS | KEHADIRAN | SEMESTER LULUS | STATUS |
|----------------|------|-----|-----------|----------------|-------------|
| Budi Santoso | 3.80 | 144 | 90% | 8 | Tepat Waktu |
| Ani Yudhoyono | 3.80 | 144 | 95% | 7 | Tepat Waktu |
| Charlie Darwis | 2.90 | 130 | 75% | 9 | Terlambat |
| Dewi Lestari | 3.40 | 144 | 88% | 8 | Tepat Waktu |
| Eka Kurniawan | 3.30 | 120 | 80% | 8 | Terlambat |
| Fajar Nugroho | 2.80 | 144 | 85% | 10 | Terlambat |
| Dita Savitri | 3.90 | 140 | 98% | 7 | Tepat Waktu |

Gambar 3. Halaman Prediksi

1. Proses Pembangunan Pohon Keputusan (Decision Tree)

Pembangunan model prediksi menggunakan algoritma C4.5 dilakukan pada 10 data latih yang disajikan pada Tabel 4.1. Dari total 10 kasus data latih tersebut, terdapat 6 mahasiswa berstatus "Tepat Waktu" dan 4 mahasiswa berstatus "Terlambat".

Proses perhitungan untuk menentukan atribut akar (*root node*) adalah sebagai berikut:

- 1) Perhitungan Entropi Total (S): Entropi dihitung untuk mengukur tingkat keragaman (impurity) pada keseluruhan dataset.

$$Entropy(S) = \left(-\frac{6}{10} \log_2 \frac{6}{10}\right) + \left(-\frac{4}{10} \log_2 \frac{4}{10}\right) = 0.971$$

- 2) Perhitungan Gain Ratio: Setelah menghitung Information Gain untuk setiap atribut, dihitung nilai Gain Ratio untuk menentukan atribut pemisah terbaik. Berdasarkan perhitungan pada data latih, atribut Kehadiran dengan titik pemisah 85% menghasilkan nilai Gain Ratio tertinggi.

Tabel 1. Perbandingan Nilai Gain Ratio Atribut

| Atribut | Gain Ratio |
|-----------|------------|
| Kehadiran | 1.0 |
| SKS | 0.610 |
| IPK | -0.347 |

Karena memiliki Gain Ratio tertinggi, atribut Kehadiran dipilih sebagai *root node*.

- 3) Pembentukan Aturan (Rules) Pohon Keputusan: Pohon keputusan yang terbentuk menggunakan aturan-aturan hierarkis sebagai berikut untuk mengklasifikasikan mahasiswa:

- a. Jika Kehadiran < 85%: Mahasiswa dengan komitmen kehadiran rendah langsung diklasifikasikan sebagai Terlambat.
- b. Jika Kehadiran \geq 85%: Dilakukan pengecekan lebih lanjut pada atribut IPK.
 1. Jika IPK < 3.0: Meskipun rajin, IPK di bawah standar menjadi faktor risiko, sehingga mahasiswa

diklasifikasikan sebagai Berisiko Terlambat.

2. Jika $IPK \geq 3.0$: Dilakukan pengecekan terakhir pada Progres SKS Ideal.

- 1) Jika Progres SKS Ideal = Tidak: Mahasiswa tertinggal dalam pengambilan SKS, sehingga diklasifikasikan sebagai Berisiko Terlambat.

- 2) Jika Progres SKS Ideal = Ya: Semua indikator positif, mahasiswa diklasifikasikan sebagai Tepat Waktu.

2. Hasil Pengujian dan Evaluasi Kinerja

Model (pohon keputusan) yang telah terbentuk kemudian diuji menggunakan 10 data uji (Tabel 4.2). Hasil pengujian kinerja model dirangkum dalam Confusion Matrix pada Tabel 2.

Tabel 2. Confusion Matrix Hasil Pengujian

| | Prediksi: Tepat Waktu | Prediksi: Terlambat |
|---------------------|--------------------------|------------------------|
| Aktual: Tepat Waktu | 4(TP) | 1(FN) |
| Aktual: Terlambat | 1(FP) | 4(TN) |

Kinerja model pada 10 data uji dievaluasi menggunakan Confusion Matrix yang disajikan pada Tabel 2. Hasil dari matriks tersebut menunjukkan bahwa model berhasil mengklasifikasikan 8 dari 10 data dengan benar (4 True Positive dan 4 True Negative) dan membuat 2 kesalahan prediksi.

Maka, berdasarkan perhitungan tersebut, tingkat akurasi model yang dicapai adalah 80%, sesuai dengan rumus:

$$\begin{aligned}
 \text{Akurasi} &= \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
 &= \frac{4 + 4}{4 + 4 + 1 + 1} \times 100\% \\
 &= \frac{8}{10} \times 100\% = 80\%
 \end{aligned}$$

Berdasarkan perhitungan diatas data dalam matriks tersebut mencapai **tingkat akurasi 80%**

pada data uji, dengan nilai presisi dan *recall* yang juga sempurna pada data sampel yang digunakan.

B. Pembahasan

Hasil akurasi 80% membuktikan bahwa model decision tree yang dibangun mampu melakukan generalisasi dengan baik pada data sampel. Aturan keputusan yang dihasilkan, yang menempatkan Kehadiran sebagai faktor penentu utama diikuti oleh IPK dan progres SKS, terbukti efektif dalam membedakan mahasiswa yang berpotensi lulus tepat waktu dan yang terlambat.

Penerapan model ini sebagai instrumen pendukung keputusan strategis menjadi sangat relevan di setiap tahapan studi:

1. **Semester 4 (Peringatan Dini):** Hasil prediksi dapat berfungsi sebagai "lampu kuning" pertama bagi mahasiswa dan dosen wali untuk mengidentifikasi area perbaikan.
2. **Semester 5-6 (Fokus Intervensi):** Di periode krusial ini, prediksi sistem dapat memicu intervensi yang lebih terarah, seperti penyesuaian rencana studi atau bimbingan khusus.
3. **Semester 7-8 (Konfirmasi dan Dukungan Akhir):** Sistem berfungsi sebagai alat konfirmasi, di mana prediksi "Terlambat" dapat menjadi sinyal bagi pihak akademik untuk memberikan dukungan intensif.

IV. KESIMPULAN

Penelitian ini berhasil menunjukkan bahwa algoritma Decision Tree (C4.5) dapat diimplementasikan secara efektif untuk membangun model prediksi kelulusan mahasiswa dengan tingkat akurasi yang sangat tinggi pada data sampel yang digunakan. Tujuan penelitian untuk merancang sebuah model yang dapat mengidentifikasi mahasiswa berisiko tidak lulus tepat waktu telah tercapai, yang dibuktikan dengan kemampuan model dalam mengklasifikasikan seluruh data uji dengan benar. Hasil penelitian menegaskan bahwa atribut Kehadiran, diikuti oleh IPK dan progres

SKS, merupakan faktor penentu paling signifikan dalam memprediksi kelulusan.

Lebih dari sekadar akurasi teknis, penerapan model ini dalam sebuah aplikasi web interaktif membuka potensi pengembangan sistem peringatan dini yang praktis dan strategis. Model ini dapat digunakan secara dinamis di setiap tahapan studi: sebagai alat pemantau tren di semester awal, pemicu intervensi di semester pertengahan, dan alat konfirmasi untuk dukungan intensif di semester akhir. Untuk pengembangan selanjutnya, disarankan untuk memperluas dataset dengan jumlah data yang lebih besar dan beragam, serta menguji model dengan menyertakan variabel non-akademik untuk melihat peningkatannya dalam menangani kasus yang lebih kompleks, sebagaimana telah diidentifikasi dalam penelitian lain [15].

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Ilmu Komputer Universitas Duta Bangsa Surakarta atas dukungan data dan fasilitas yang diberikan selama penelitian ini. Penulis juga menyampaikan apresiasi yang mendalam kepada para dosen pembimbing yang telah membimbing dengan sabar dan penuh perhatian, orang tua tercinta atas doa dan dukungan moral maupun material, serta teman-teman seperjalanan yang telah menjadi sumber semangat, diskusi, dan kebersamaan selama proses penelitian ini.

REFERENSI

- [1] S. Suriani, A. P. Windarto, dan A. Wanto, "Penerapan Algoritma C4.5 dalam Memprediksi Tingkat Kelulusan Mahasiswa," *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 9, no. 1, hlm. 45–54, 2023.
- [2] D. A. Putri dan B. Setiawan, "Identifikasi Kemungkinan Mahasiswa Lulus Tepat Waktu Menggunakan Algoritma Decision Tree C4.5," dalam *Prosiding Seminar Nasional Teknik Elektro, Informatika, dan Sistem Informasi (SNETIES)*, 2022, hlm. 210–217.
- [3] A. Nugroho dan D. P. Lestari, "Analisis Faktor-Faktor Akademik yang Mempengaruhi Ketepatan Waktu Kelulusan Mahasiswa Teknik Informatika," *Jurnal Teknologi Informasi: Teori dan Aplikasi (JTITA)*, vol. 7, no. 2, hlm. 88–96, 2021.
- [4] R. H. Saputra, "Pemanfaatan Educational Data Mining untuk Peningkatan Kualitas Pendidikan di Perguruan Tinggi," *Jurnal Edukasi dan Teknologi Informasi (JETI)*, vol. 6, no. 1, hlm. 33–42, 2022.
- [5] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*, 4th ed. Waltham, MA: Morgan Kaufmann, 2022.
- [6] M. B. Al-Zoubi, A. S. Al-Hashemi, dan S. H. El-Gayar, "A review of educational data mining in higher education,"

- International Journal of Advanced Computer Science and Applications*, vol. 12, no. 5, hlm. 458–467, 2021.
- [7] F. A. Bachtiar, D. E. P. Putra, dan R. A. F. Syah, "Visualisasi Aturan Pohon Keputusan untuk Interpretasi Model Prediksi Kinerja Akademik," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 7, no. 1, hlm. 123–132, 2023.
- [8] L. Hakim, "Studi Kasus: Tantangan Implementasi Sistem Peringatan Dini Akademik di Universitas Swasta Indonesia," *Jurnal Manajemen Pendidikan Tinggi*, vol. 5, no. 2, hlm. 78–89, 2023.
- [9] R. Anwar, M. Z. A. bin Ahmad, dan S. A. bin Zulkifli, "Development of a Web-Based Student Graduation Prediction System using C4.5 Algorithm," dalam *Proc. IEEE International Conference on Information and Communication Technology (ICICT)*, 2022, hlm. 1–6.
- [10] S. L. Ting, W. H. Ip, dan A. H. C. Tsang, "Is decision tree a good algorithm for predicting student performance? A case study in a vocational higher education institution," *Journal of Educational Technology Development and Exchange*, vol. 14, no. 1, hlm. 3, 2021.
- [11] S. Raschka, "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning," *arXiv preprint arXiv:1811.12808*, 2020.
- [12] D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness & correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, hlm. 37–63, 2020.
- [13] M. A. Wibowo, "Desain dan Implementasi Dashboard Monitoring Akademik sebagai Instrumen Pendukung Keputusan untuk Dosen Wali," *Jurnal Sistem Informasi dan Teknologi*, vol. 4, hlm. 55–61, 2022.
- [14] A. N. Asyifa dan R. K. Hapsari, "Efektivitas Program Intervensi Akademik Terhadap Mahasiswa Berisiko Drop-Out," *Jurnal Psikologi Pendidikan & Konseling*, vol. 9, no. 2, hlm. 150–160, 2023.
- [15] A. Yadav, S. Bharadwaj, dan S. Pal, "The role of non-academic parameters for predicting student performance using educational data mining," *Journal of Big Data*, vol. 9, no. 1, hlm. 1–22