

Optimasi Machine Learning dalam Memprediksi Kelulusan Mahasiswa

Embun Fajar Wati¹, Luthfi Indriyani², Elvi Sunita³, Andi Diah Kuswanto⁴

^{1,2,3,4}Universitas Bina Sarana Informatika
Jakarta, Indonesia

e-mail: 1embun.efw@bsi.ac.id, 2luthfi.lfy@bsi.ac.id, 3elvi.evt@bsi.ac.id, 4andi.ahk@bsi.ac.id

Abstrak - Ketepatan kelulusan mahasiswa merupakan indikator penting dalam evaluasi kinerja perguruan tinggi. Keterlambatan kelulusan tidak hanya berdampak pada efektivitas proses akademik, tetapi juga pada efisiensi pengelolaan sumber daya dan reputasi institusi. Penelitian ini bertujuan untuk memprediksi ketepatan kelulusan mahasiswa menggunakan machine learning dengan jenis data mining, yaitu algoritma Decision Tree berdasarkan data kelulusan mahasiswa. Variabel yang digunakan antara lain jenis kelamin, pekerjaan, penghasilan, ipk, status pekerjaan. Metode pengujian dilakukan dengan optimasi PSO dan 10-fold cross-validation, sedangkan evaluasi model menggunakan nilai akurasi, presisi, recall, dan AUC. Hasil pengujian menunjukkan bahwa model Decision Tree memiliki kinerja yang sangat baik dalam memprediksi ketepatan kelulusan mahasiswa, dengan nilai akurasi sebesar 96.71%, presisi 92.06%, recall 99.09%, dan AUC 0.973. Temuan ini menunjukkan bahwa model efektif digunakan sebagai sistem peringatan dini untuk mengidentifikasi mahasiswa yang berpotensi terlambat lulus. Dengan demikian, pihak perguruan tinggi dapat memberikan intervensi akademik yang lebih tepat sasaran untuk meningkatkan tingkat kelulusan tepat waktu.

Kata Kunci: prediksi kelulusan, machine learning, decision tree, data mining, ketepatan kelulusan

PENDAHULUAN

Beberapa aspek yang mempengaruhi hasil akreditasi yang baik, salah satunya mahasiswa (Hendra et al., 2020). Keberhasilan seorang mahasiswa untuk menyelesaikan studinya tepat waktu merupakan faktor penting dalam menentukan pencapaian sebuah lembaga pendidikan tinggi (Rapel Aprilus Sigit et al., 2024). Ada banyak faktor yang bisa mempengaruhi keputusan kelulusan, seperti informasi pribadi (jenis kelamin, usia, status menikah, serta pekerjaan) dan hasil belajar (nilai ujian). Sebelum membuat keputusan, universitas perlu meninjau jumlah calon mahasiswa serta faktor-faktor yang terbesar dalam menentukan kelulusan mereka (Embun Fajar Wati, Budi Sudrajat, et al., 2024).

Untuk mengatasi masalah keterlambatan dalam proses kelulusan, diperlukan prediksi agar dapat mengetahui tingkat kelulusan mahasiswa dalam rentang waktu yang diharapkan (Teuku Afriliansyah, 2023). Diperlukan teknik yang dapat memprediksi apakah seseorang akan lulus atau tidak. Teknik yang biasanya digunakan adalah dengan menganalisis data melalui proses data mining (Azis Wahyudi et al., 2023).

Prediksi kelulusan bisa memberikan informasi yang sangat membantu dalam mengambil keputusan; karena itu, dilakukan penelitian mengenai kelulusan mahasiswa (Embun Fajar Wati et al., 2023). Mengenali secara tepat mahasiswa yang berpotensi mengalami keterlambatan kelulusan memungkinkan lembaga pendidikan untuk memberikan bantuan yang tepat dan tepat waktu, sehingga mahasiswa tetap bisa

berada di jalur yang benar untuk menyelesaikan studi mereka (Hadi Putra et al., 2025). Proses prediksi dimulai dengan membuat model berdasarkan dataset pelatihan, yaitu kumpulan data yang sudah memiliki label, lalu model tersebut digunakan untuk memprediksi kelas dari data baru yang belum memiliki label (Dihin Muriyatmoko et al., 2024). Tujuan penelitian ini adalah membuat model prediksi untuk mengetahui apakah mahasiswa akan lulus atau tidak menggunakan algoritma Decision Tree (C4.5), sehingga bisa mengenali lebih awal mahasiswa yang kemungkinan besar akan mengalami keterlambatan dalam studi mereka (Stefanus Alvian Setiono & Eko Purwanto, 2025).

Data mining dipilih sebagai cara untuk mengolah data supaya diperoleh informasi yang baru (Embun Fajar Wati, Elvi Sunita Perangin-Angin, & Anggi Puspita Sari, 2024). Algoritma Decision Tree adalah salah satu metode data mining untuk mengklasifikasikan data dengan membuat pohon keputusan yang sudah diperiksa ketepatannya (Moch.Ilham Maulana Noor & Mochamad Izzan Firasyansyah, 2024). Algoritma ini bisa memproses data berupa angka dan kategori, serta memiliki kemampuan memilih atribut yang penting untuk mengambil keputusan (Uci Suriani, 2023) (Embun Fajar Wati, Elvi Sunita Perangin-Angin, & Luthfi Indriyani, 2024). Pada penelitian ini akan digunakan metode decision tree dengan optimasi PSO.

PSO (Particle Swarm Optimization) digunakan untuk memilih atribut yang penting dan membantu mengurangi input yang tidak relevan (Embun Fajar Wati et al., 2025). Optimasi swarm partikel untuk meningkatkan tingkat keakuratan hasil

dari metode yang digunakan (Wati et al., 2021).

Beberapa penelitian terdahulu dengan kasus dan metode yang sama menghasilkan beberapa temuan. Hasil penelitian menunjukkan bahwa metode PSO mampu meningkatkan hasil klasifikasi C4.5 dalam hal memprediksi kelulusan. Model C4.5 yang di gabungkan dengan PSO mencapai akurasi sebesar 68,60%, yang lebih baik dibandingkan model C4.5 biasa yang hanya mencapai 64,45% (Fajar Haditio et al., 2025). Dari hasil evaluasi dan validasi menggunakan tools RapidMiner, diperoleh akurasi metode Decision Tree (C4.5) sebesar 98,04% pada pengujian ketiga. Akurasi metode Naïve Bayes mencapai 96,00% pada pengujian keempat, sedangkan metode K-Nearest Neighbor (K-NN) mencapai 90,00% pada pengujian kedua. Hasil ini menunjukkan bahwa metode Decision Tree memiliki akurasi lebih tinggi dibandingkan Naïve Bayes dan K-NN dalam menentukan kelulusan mahasiswa (Agus Budiyantra et al., 2020). Algoritma Decision Tree mencapai performa dengan akurasi 90,25% dalam memprediksi tingkat kelulusan mahasiswa (Umar Haji Mussa'id et al., 2024).

Penelitian ini dapat memberikan kontribusi bagi perguruan tinggi yaitu model dapat digunakan untuk early warning system yang bekerja dengan menganalisis IPK untuk mendeteksi mahasiswa berisiko tinggi terlambat lulus dan melakukan intervensi akademik.

METODE PENELITIAN

Penelitian ini dilakukan melalui beberapa tahapan untuk mengelompokkan data mengenai prediksi kelulusan mahasiswa secara tepat. Metode yang digunakan adalah algoritma Decision Tree sebagai cara untuk mengklasifikasikan data, sedangkan RapidMiner digunakan sebagai alat bantu dalam memproses data tersebut (Moch.Ilham Maulana Noor & Mochamad Izzan Firasyansyah, 2024).

Pengumpulan data dengan studi literatur dan data kelulusan mahasiswa di salah satu perguruan tinggi swasta. Metode yang digunakan untuk mengumpulkan data adalah metode dokumentasi. Data yang telah diproses adalah data sekunder dari perguruan tinggi dengan jumlah dataset sebanyak 301. Pekerjaan terdiri dari karyawan, wiraswasta, dan ngojek.

Beberapa tahapan sebagai berikut :

1. Pengumpulan dataset kelulusan mahasiswa yang akan diproses menggunakan model Decision Tree.
2. Pengolahan data dilakukan dengan metode Knowledge Discovery in Database (KDD) yang disebut preprocessing, yang meliputi proses membersihkan data agar data yang sudah didapat bisa digunakan dengan baik. Proses ini melibatkan penghapusan data yang sama, kesalahan, serta memastikan aturan

pengecekan yang benar. Selain itu, data yang tidak lengkap atau terdapat gangguan juga diperbaiki agar tidak menyebabkan masalah saat memproses data, sehingga hasil yang diperoleh lebih tepat dan akurat.

3. Pengujian model menggunakan decision tree dengan optimasi PSO dilakukan dengan teknik 10-Fold Cross Validation untuk memastikan hasil klasifikasi tetap akurat dan optimal.
4. Analisis dan pembahasan dari hasil pengujian diuraikan secara rinci, khususnya fokus pada tingkat akurasi klasifikasi yang diperoleh dari RapidMiner, serta membahas berbagai faktor yang mempengaruhi kemampuan prediksi kelulusan mahasiswa secara tepat waktu.
5. Evaluasi berdasarkan hasil analisis yang dilakukan, termasuk tingkat keberhasilan metode Decision Tree dalam memprediksi kelulusan mahasiswa serta beberapa rekomendasi yang diberikan kepada perguruan tinggi.

Langkah-langkah ini dibuat agar penelitian berjalan dengan teratur dan menghasilkan prediksi yang tepat. Atribut yang digunakan pada data kelulusan mahasiswa yaitu, jenis kelamin, pekerjaan, penghasilan, ipk, status pekerjaan. Objek penelitian ini adalah model prediksi yang digunakan untuk memprediksi status kelulusan mahasiswa ke dalam dua kategori, yaitu "tepat" dan "terlambat". Variabel dependen atau label yang digunakan sebagai target prediksi dalam model ini adalah "kelulusan".

HASIL DAN PEMBAHASAN

Pengolahan data dengan preprocessing dilakukan dengan cepat karena data sudah terisi semua dengan benar. Data sebanyak 301 terpakai semua. Preprocessing juga mencakup transformasi data dengan membagi data penghasilan dan ipk ke dalam beberapa kategori pada tabel 1 dan tabel 2.

Tabel 1. Transformasi Atribut Penghasilan

Penghasilan	Kategori
1-5 juta	Rendah
6-10 juta	Tinggi

Tabel 2. Transformasi Atribut IpK

IpK	Kategori
3-4	Muda
1-2.9	Tua

Pengujian dilakukan dengan menggunakan RapidMiner untuk mengetahui seberapa tepat algoritma Decision Tree dalam mengklasifikasikan data prediksi kelulusan mahasiswa secara tepat waktu. Berbagai operator digunakan dalam

RapidMiner untuk memproses data, melatih model, serta menguji kinerja klasifikasi, sehingga memastikan hasil yang akurat dan dapat diandalkan. Pemodelan machine learning dengan metode decision tree dengan optimasi Particle Swarm Optimization (PSO) dan teknik 10-Fold Cross Validation pada gambar 1.

Analisis dan pembahasan terkait dengan hasil pengujian. Hasil pengujian metode decision tree dengan PSO dan 10-Fold Cross Validation menunjukkan bahwa akurasi, presisi, recall, dan AUC bernilai sangat tinggi dan dapat dilihat pada gambar 2 performance vector. Pada gambar 2 terdapat nilai akurasi 96.71%, nilai presisi 92.06%, dan nilai recall 99.09% dengan keterangan :

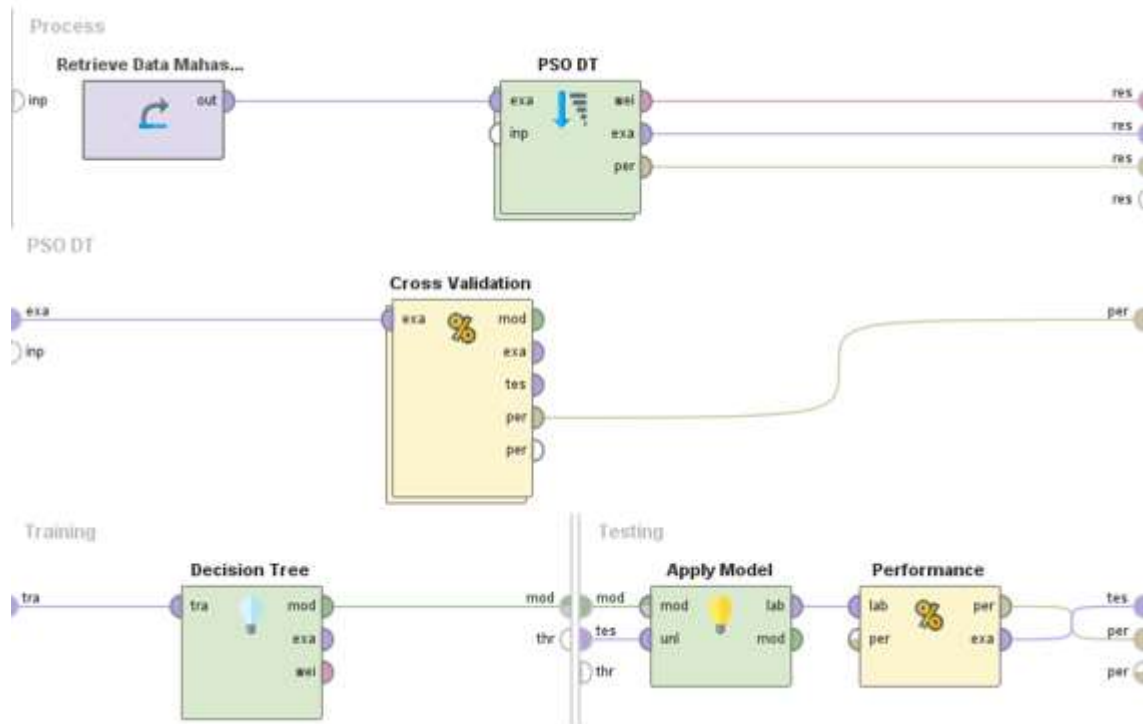
1. Prediksi tepat dan true tepat 194
2. Prediksi terlambat dan true tepat 9
3. Prediksi tepat dan true terlambat 1
4. Prediksi terlambat dan true terlambat 97

menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik karena mendekati angka 1 dan dapat diandalkan dalam membedakan kelas positif dan kelas negatif yang sangat baik.

Nilai atribut weight yang dihasilkan dari metode decision tree dengan PSO pada gambar 4 menunjukkan atribut yang paling berpengaruh dalam hasil prediksi kelulusan mahasiswa adalah jenis kelamin dan pekerjaan. Untuk atribut yang paling kecil nilainya yaitu penghasilan dapat dilakukan seleksi fitur untuk dihapus agar dapat mengurangi kompleksitas model, mempercepat komputasi, mengurangi noise, dan meningkatkan akurasi model.

Pada gambar 1 bagian atas terdapat proses yaitu model PSO (Particle Swarm Optimization) yang dihubungkan dengan data mahasiswa. Sedangkan bagian tengah terdapat jenis validasi yaitu 10-fold cross validation yaitu 10 kali proses validasi.

Nilai AUC 0.973 pada gambar 3 yang



Gambar 1. Pemodelan Decision Tree dengan PSO dan 10-Fold Cross Validation

Pada gambar 1 bagian bawah terdapat training dengan decision tree dan testing dengan apply model yang dihubungkan dengan performance untuk memberikan hasil dalam bentuk persentase.

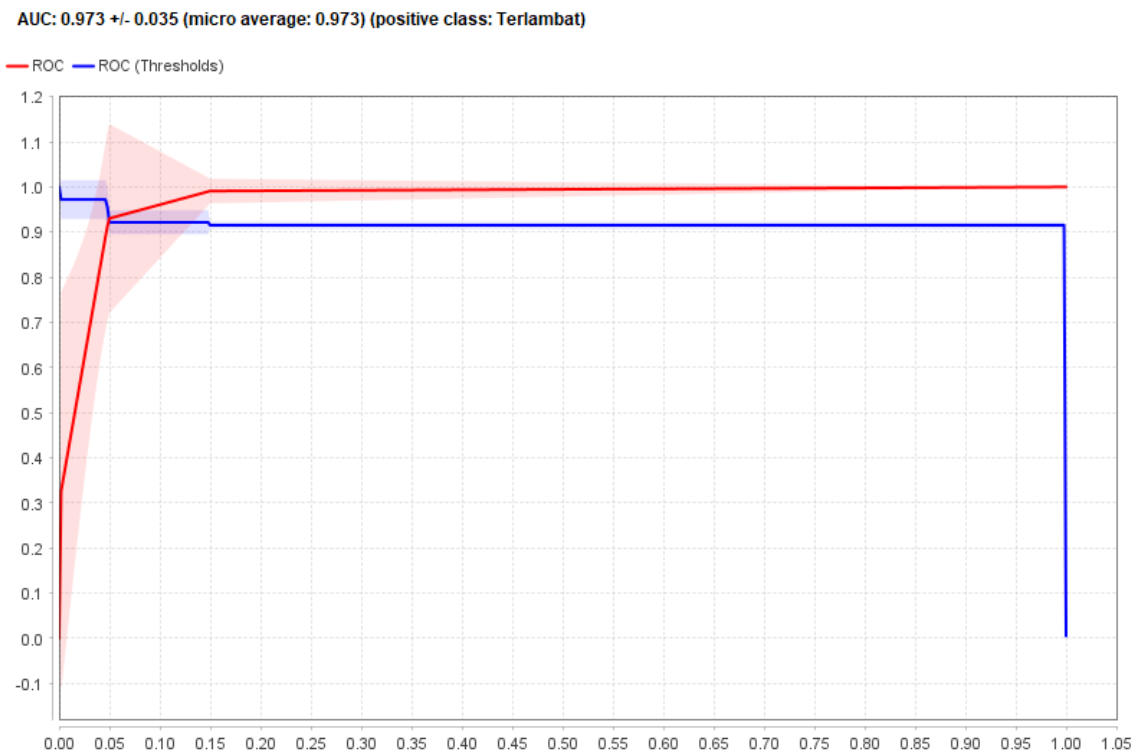
Pada gambar 2, performance vector menghasilkan beberapa nilai yaitu accuracy, precision, recall, dan AUC dengan hasil yang sangat tinggi yaitu 96.71%, 92.06%, 99.09%, dan 0.973.

Pada gambar 3, terdapat kurva ROC yang ditampilkan dengan nilai AUC yaitu 0.973. nilai AUC tersebut menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik dan dapat diandalkan dalam membedakan kelas positif dan negatif yang sangat baik.

PerformanceVector

```
PerformanceVector:  
accuracy: 96.71% +/- 3.73% (micro average: 96.68%)  
ConfusionMatrix:  
True:  Tepat  Terlambat  
Tepat: 194    1  
Terlambat: 9    97  
precision: 92.06% +/- 6.93% (micro average: 91.51%) (positive class: Terlambat)  
ConfusionMatrix:  
True:  Tepat  Terlambat  
Tepat: 194    1  
Terlambat: 9    97  
recall: 99.09% +/- 2.87% (micro average: 98.98%) (positive class: Terlambat)  
ConfusionMatrix:  
True:  Tepat  Terlambat  
Tepat: 194    1  
Terlambat: 9    97  
AUC (optimistic): 0.999 +/- 0.004 (micro average: 0.999) (positive class: Terlambat)  
AUC: 0.973 +/- 0.035 (micro average: 0.973) (positive class: Terlambat)  
AUC (pessimistic): 0.948 +/- 0.066 (micro average: 0.948) (positive class: Terlambat)
```

Gambar 2. Performance Vector



Gambar 3. Diagram AUC

attribute	weight
Jenis Kelamin	0.401
Pekerjaan	0.189
Penghasilan	0.029
IPK Terakhir	0.142
Status Pekerjaan	0.077

Gambar 4. Atribut Weight

Dari nilai-nilai yang sudah dihasilkan melalui pengujian model, terbukti bahwa model klasifikasi mempunyai tingkat ketepatan yang sangat tinggi dengan klasifikasi yang tepat berjumlah $194+97=291$ dari data kelulusan mahasiswa sebanyak 301. Jumlah error pada klasifikasi sangat kecil yaitu hanya ada 1 mahasiswa yang sebenarnya terlambat tetapi diprediksi tepat, dan 9 mahasiswa dengan kelulusan tepat yang diprediksi terlambat.

Hasil ini menunjukkan bahwa perguruan tinggi bisa menggunakan model ini untuk early warning system mahasiswa yang berpotensi terlambat lulus dan melakukan intervensi akademik.

KESIMPULAN

Akurasi sebesar 96.71%, presisi 92.06%, dan recall 99.09% menunjukkan bahwa model memiliki performa yang sangat baik. Nilai akurasi yang tinggi menandakan bahwa sebagian besar prediksi model sudah tepat. Nilai presisi sebesar 92.06% menunjukkan bahwa dari seluruh mahasiswa yang diprediksi tepat waktu, 92.06% benar-benar tepat waktu (kesalahan prediksi sangat kecil). Sementara itu, recall yang sangat tinggi yaitu 99.09% menunjukkan bahwa model mampu mendeteksi hampir seluruh mahasiswa yang benar-benar tepat waktu secara akurat.

Secara keseluruhan, model ini sangat andal dan efektif, terutama dalam mengidentifikasi mahasiswa yang benar-benar tepat waktu, sehingga cocok dijadikan sistem peringatan dini untuk mendeteksi risiko keterlambatan kelulusan mahasiswa. Pada penelitian selanjutnya dapat digunakan data yang sama tetapi dengan beberapa perbandingan metode dan optimasi yang berbeda.

REFERENSI

Agus Budiyantra, Irwansyah, Egi Prengki, Pandi Ahmad Pratama, & Ninuk Wiliani. (2020). KOMPARASI ALGORITMA DECISION TREE, NAIVE BAYES DAN K-NEAREST NEIGHBOR UNTUK MEMPREDIKSI

- MAHASISWA LULUS TEPAT WAKTU. JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER, 5(2), 265–270.
- Azis Wahyudi, Kusriani, & Ferry Wahyu Wibowo. (2023). PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN METODE DECISION TREE DAN NAIVE BAYES. JURNAL PERMATA INDONESIA, 14(2), 132–138.
- Dihin Muriyatmoko, Aziz Musthafa, & Muqoddam Husni Wijaya. (2024). Klasifikasi Profil Kelulusan Nilai AKPAM Dengan Metode Decision Tree C4.5. Seminar Nasional Sains Dan Teknologi, 1–9.
- Embun Fajar Wati, Budi Sudrajat, & Raudah Nasution. (2024). Modelling of C4.5 Algorithm for Graduation Classification. International Journal of Information System & Technology, 8(1), 40–46.
- Embun Fajar Wati, Elvi Sunita Perangin-Angin, & Anggi Puspita Sari. (2023). Prediction of Student Graduation using the K-Nearest Neighbors Method. International Journal of Information System & Technology, 7(3), 211–216.
- Embun Fajar Wati, Elvi Sunita Perangin-Angin, & Anggi Puspita Sari. (2024). Improved Naive Bayes Algorithm with Particle Swarm Optimization to Predict Student Graduation. International Journal of Information System & Technology, 7(6), 386–391.
- Embun Fajar Wati, Elvi Sunita Perangin-Angin, & Luthfi Indriyani. (2024). Customer Loyalty Classification with Comparison of Naive Bayes, C4.5, and KNN Methods. International Journal of Information System & Technology, 8(3), 177–185.
- Embun Fajar Wati, Elvi Sunita Perangin-Angin, & Luthfi Indriyani. (2025). Comparison of Naive Bayes and C4.5 Methods with Particle Swarm Optimization on Customer Loyalty Classification. International Journal of Information System & Technology, 8(6), 338–344.
- Fajar Haditio, M. Irham, & Andi. (2025). PREDIKSI KELULUSAN TEPAT WAKTU MAHASISWA PROGRAM STUDI MANAJEMEN UNIVERSITAS PAMULANG MENGGUNAKAN ALGORITMA C4.5 DAN PARTICAL SWARM OPTIMIZATION (PSO). Journal of Scientech Research and Development, 7(1), 1016–1027.
- Hadi Putra, Khairunnisa Nasution, & Elkin Rilvani. (2025). PREDIKSI KELULUSAN MAHASISWA TEPAT WAKTU MENGGUNAKAN DECISION TREE: STUDI PERBANDINGAN ALGORITMA ID3 DAN C4.5. JURNAL MEDIA AKADEMIK (JMA), 3(7), 1–15.
- Hendra, Mochammad Abdul Azis, & Suhardjono. (2020). Analisis Prediksi Kelulusan Mahasiswa Menggunakan Decision Tree Berbasis Particle Swarm Optimization. Jurnal SISFOKOM (Sistem Informasi Dan Komputer), 9(1), 102–107.
- Moch. Ilham Maulana Noor, & Mochamad Izzan Firasyansyah. (2024). Optimalisasi Klasifikasi

- Prediksi Kelulusan Mahasiswa Kota Subang Menggunakan Algoritma Decision Tree. *Jurnal Jawara Sistem Informasi*, 2(1), 1–5.
- Rapel Aprilius Sigit, Zuprizal Kurniawan, & Rahmadden. (2024). KOMPARASI ALGORITMA MACHINE LEARNING UNTUK KLASIFIKASI KELULUSAN MAHASISWA. *JURNAL JARINGAN SISTEM INFORMASI ROBOTIK (JSR)*, 108(113), 8–1.
- Stefanus Alvian Setiono, & Eko Purwanto. (2025). Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree. *Seminar Nasional Teknologi Informasi Dan Bisnis (SENATIB)*, 401–406.
- Teuku Afriliansyah. (2023). ALGORITMA ANN, KNN, NAIVE BAYES, SVM DAN DECISION TREE DALAM MENENTUKAN KELULUSAN MAHASISWA. *Jurnal Pendidikan Bumi Persada*, 2(2), 42–54.
- Uci Suriani. (2023). Penerapan Data Mining untuk Memprediksi Tingkat Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree C4.5. *Journal of Computer and Information Systems Ampera*, 3(2), 55–66.
- Umar Haji Mussa'id, Agus Susilo Nugroho, & Rahmawan Bagus Trianto. (2024). PREDIKSI TINGKAT KELULUSAN MAHASISWA S1 UNIVERSITAS AN NUUR DENGAN METODE DECISION TREE C4.5. *Julia: Jurnal Ilmu Komputer An Nuur*, 4(1), 21–29.
- Wati, E. F., Sari, A. P., Alawiah, E. T., Siregar, M. H., & Rudianto, B. (2021). Particle Swarm Optimization Comparison on Decision Tree and Naive Bayes for Pandemic Graduation Classification. *2nd International Conference on Advanced Information Scientific Development (ICAISD)*, 1–11.