

PERBANDINGAN ALGORITMA MACHINE LEARNING DALAM MEMPREDIKSI KELULUSAN MAHASISWA

Oleh :

M Cholis Afandi

Uce Indahyanti

Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

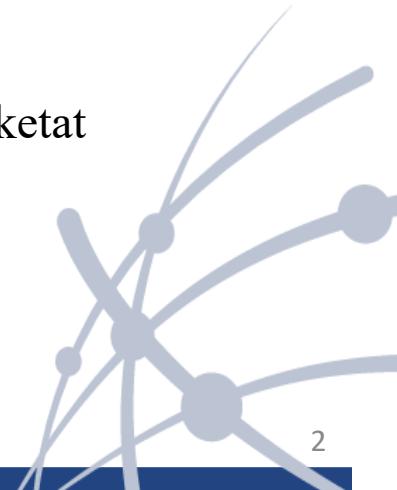
Juni, 2025



PENDAHULUAN

Latar Belakang

1. Faktor terpenting dalam pembangunan suatu bangsa adalahh pendidikan. Dengan memberikan pendidikan berkualitas tinggi kepada para mahasiswa.
2. Mahasiswa merupakan subjek yang sangat penting bagi keberhasilan akademik. Salah satu faktor utama yang mempengaruhi tingkat akademik adalah kemampuan siswa untuk menyelesaikan studi mereka tepat waktu
3. Machine learning digunakan untuk teknik data mining, yang merupakan metodologi yang ketat yang telah menunjukkan hasil dalam domain klasifikasi dan prediksi.



RUMUSAN MASALAH

1. Bagaimana membandingkan performa algoritma Naive Bayes, Decision Tree, dan Random Forest dalam memprediksi kelulusan mahasiswa berdasarkan data akademik?
2. Algoritma mana yang paling optimal digunakan pada data akademik mahasiswa Prodi Informatika UMSIDA?



www.umsida.ac.id



[umsida1912](https://www.instagram.com/umsida1912/)



[umsida1912](https://twitter.com/umsida1912)



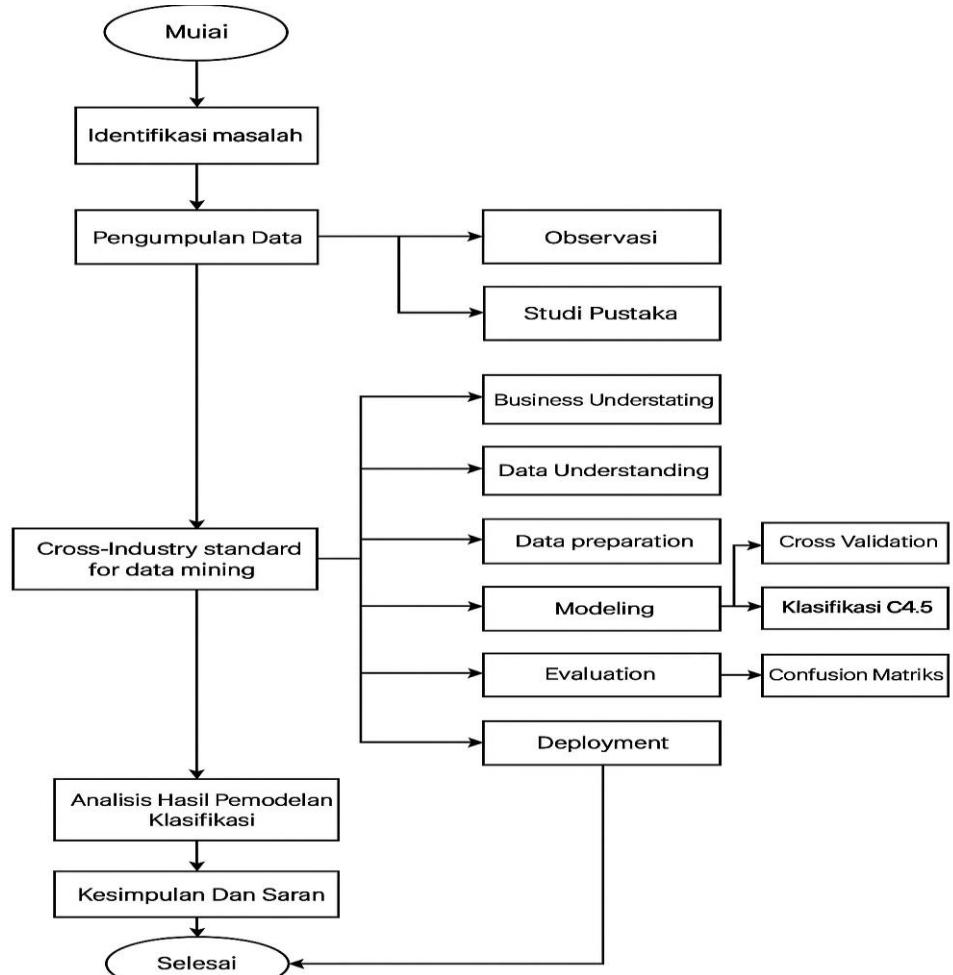
universitas
muhammadiyah
sidoarjo



[umsida1912](https://www.youtube.com/umsida1912)

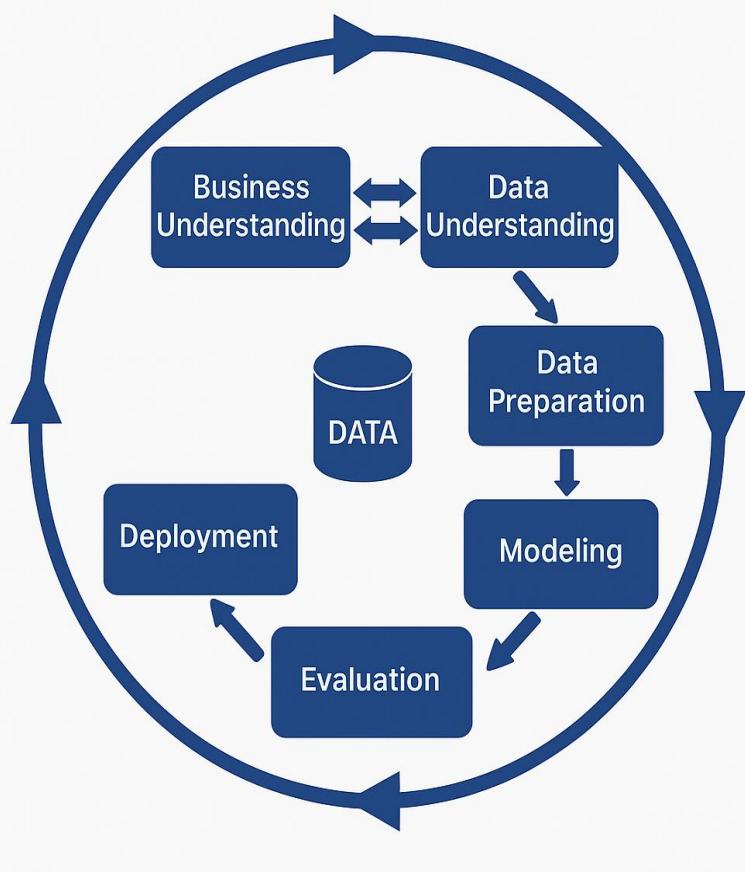


METODE



- **Metode:** CRISP-DM (Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, Deployment).
- **Data:** Mahasiswa Informatika angkatan 2020–2021 (IPS, jumlah SKS semester 1–6).
- **Algoritma:** Naive Bayes, Decision Tree, Random Forest.
- **Evaluasi:** Confusion Matrix (Akurasi, Presisi, Recall, F1-Score).
- **Skenario:** Pembagian data 80:20 dan 60:40.

Metode



1. Business Understanding

Memahami tujuan dan permasalahan bisnis yang ingin diselesaikan dengan data mining.

2. Data Understanding

Mengumpulkan, mengeksplorasi, dan mengenali struktur serta kualitas data yang tersedia.

3. Data Preparation

Membersihkan, mengolah, dan mentransformasi data agar siap digunakan dalam pemodelan.

4. Modeling

Membangun dan melatih model Machine Learning menggunakan data yang telah dipersiapkan.

5. Evaluation

Menilai performa model dengan metrik seperti akurasi, presisi, recall, dan F1-score untuk memastikan model sesuai dengan tujuan bisnis.

6. Deployment

Mengimplementasikan model ke dalam sistem nyata agar dapat digunakan oleh pengguna atau institusi terkait.

HASIL DAN PEMBAHASAN

Algoritma	Akurasi	Presisi	Recall	F1-Score
Naïve Bayes	86.25%	93%	86%	88%
Decision Tree	96.25%	96%	96%	96%
Random Forest	97.50%	98%	97%	97%

Berdasarkan hasil penelitian, algoritma Random Forest memberikan performa terbaik dalam memprediksi kelulusan mahasiswa dengan akurasi 97.50%, mengungguli Decision Tree (96.25%) dan Naïve Bayes (86.25%). Dengan nilai presisi, recall, dan F1-score yang tinggi, Random Forest dinilai paling akurat dan stabil, sehingga direkomendasikan sebagai model terbaik untuk sistem prediksi kelulusan mahasiswa.

HASIL DAN PEMBAHASAN

DATASET

Data penelitian ini diambil dari mahasiswa Program Studi Informatika angkatan 2020–2021. Atribut yang digunakan meliputi nilai Indeks Prestasi Semester (IPS) dari semester 1 hingga 6, serta jumlah Satuan Kredit Semester (SKS). Data tersebut diberi label status kelulusan mahasiswa, yaitu 'Lulus' atau 'Tidak Lulus'

NIM	NAMA	IPS_1	SKS_1	IPS_2	SKS_2	IPS_3	SKS_3	IPS_4	SKS_4	IPS_5	SKS_5	IPS_6	SKS_6
201080200001	MOKHAMAD DIKI ARMANDA	3.83	20	3.80	20	3.84	20	3.60	21	3.80	23	3.91	21
201080200002	MUHAMMAD ALAIKA ASYROF	3.78	20	3.30	20	3.35	23	3.55	21	2.67	23	3.60	15
201080200003	ARI TOPAN IQBAL MADAGASKAR	3.80	20	3.57	20	3.26	23	3.62	21	3.55	23	3.72	18
201080200004	FAHRI FAUZI HIDAYAT	3.72	20	3.78	18	3.47	23	3.65	18	3.60	23	3.70	23
201080200005	RENO FAIZAL FAHMI	2.53	20	3.13	18	3.47	20	3.52	20	3.60	23	3.76	21
201080200006	MUHAMMAD ALVIN AZZAMUL AZMI	3.78	20	3.82	20	3.70	23	3.78	21	3.86	23	3.83	18
201080200007	MOHAMMAD FADLI ZAKA	3.83	20	3.80	20	3.75	23	3.73	21	3.90	23	3.78	18
201080200008	EMIL MAULANA RAMADHAN	3.75	20	3.62	20	3.61	23	3.51	21	3.04	23	3.78	18
201080200009	DAVITO RASENDRIYA RIZQULLAH PUTRA	3.78	20	3.75	20	3.32	23	3.68	21	3.71	23	3.83	18

Pembahasan

1. PENERAPAN NAÏVE BAYES

Classification Report:

	precision	recall	f1-score	support
Lulus	1.00	0.84	0.91	70
Tidak Lulus	0.48	1.00	0.65	10
accuracy			0.86	80
macro avg	0.74	0.92	0.78	80
weighted avg	0.93	0.86	0.88	80
Akurasi NB :	86.25%			

Confusion Matrix:

[[59 11]
[0 10]]

- Pada algoritma **Naïve Bayes**, hasil akurasi yang diperoleh pada skenario pertama maupun skenario kedua adalah **86.25%**. Tidak adanya perubahan nilai akurasi ini menunjukkan bahwa performa Naïve Bayes tidak terlalu sensitif terhadap variasi proporsi data training dan testing. Hal ini sejalan dengan karakteristik algoritma Naïve Bayes yang berbasis probabilistik dan cenderung menghasilkan performa yang stabil pada dataset berukuran sedang.



Pembahasan

2. PENERAPAN DECISION TREE

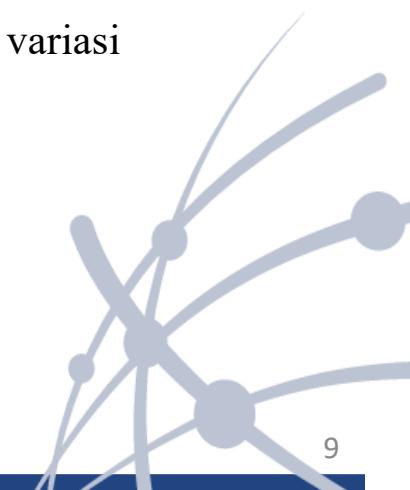
Classification Report:

	precision	recall	f1-score	support
Lulus	0.97	0.99	0.98	70
Tidak Lulus	0.89	0.80	0.84	10
accuracy			0.96	80
macro avg	0.93	0.89	0.91	80
<u>weighted</u> avg	0.96	0.96	0.96	80
Akurasi NB:	96.25%			

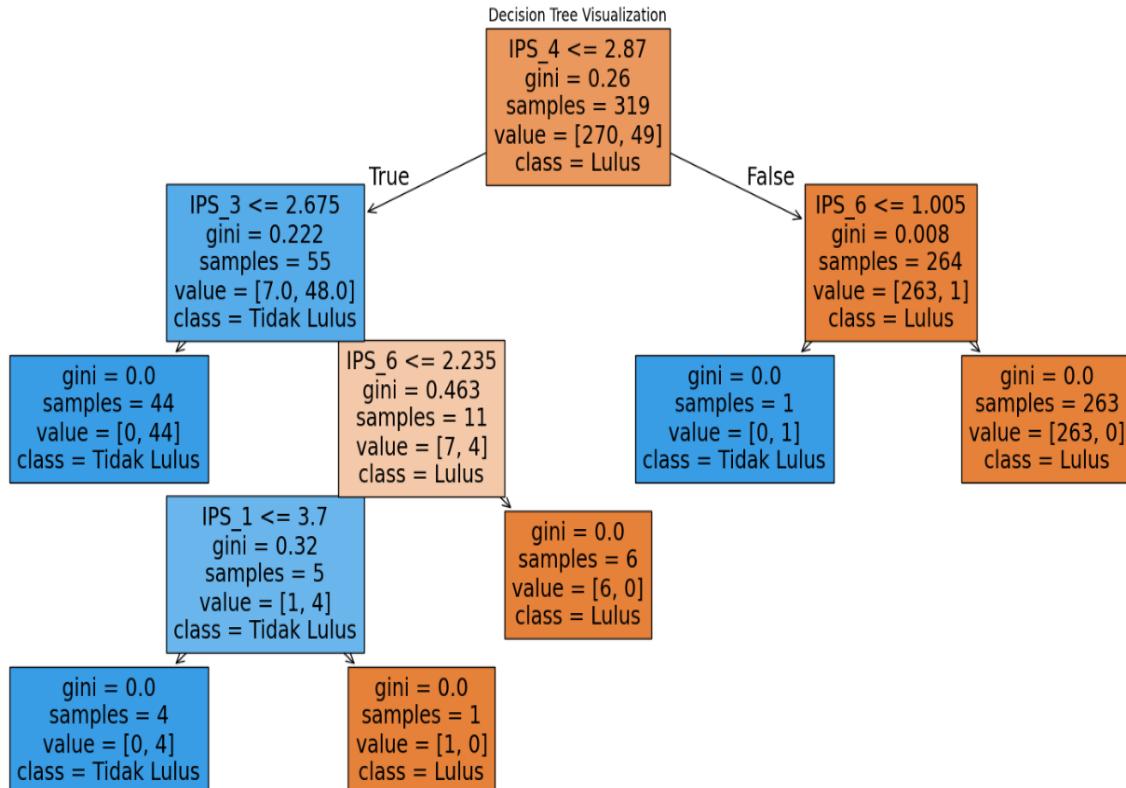
Confusion Matrix:

<u>69</u>	1
2	<u>8</u>

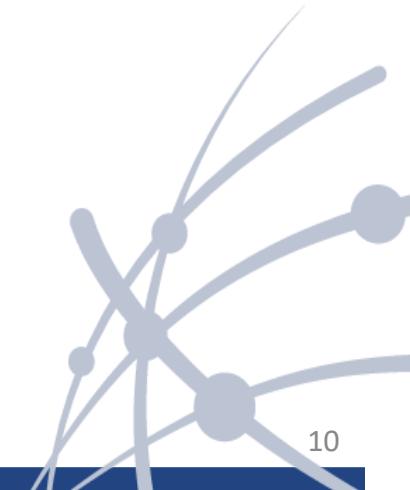
Pada algoritma **Decision Tree**, akurasi yang dihasilkan juga tetap sama pada kedua skenario, yaitu sebesar **96.25%**. Konsistensi nilai akurasi ini menandakan bahwa Decision Tree mampu membentuk aturan klasifikasi yang solid dan efektif, meskipun proporsi data latih dikurangi. Ketahanan model terhadap perbedaan proporsi data menunjukkan bahwa struktur pohon keputusan yang terbentuk tidak mengalami overfitting maupun underfitting. Dengan demikian, Decision Tree menjadi algoritma yang kuat dalam menangani prediksi kelulusan pada variasi jumlah data pelatihan.



Pohon Keputusan



Pohon keputusan ini mengelompokkan status kelulusan berdasarkan nilai IPS, dimana mahasiswa dengan $IPS_4 > 2,87$ cenderung lulus, sedangkan yang bernilai $\leq 2,87$ diperinci lebih lanjut dengan nilai IPS_3 dan IPS_1 sehingga mayoritasnya diprediksi tidak lulus; nilai IPS_6 juga digunakan untuk mendukung prediksi kelulusan pada kelompok dengan IPS_4 tinggi.



Pembahasan

3. PENERAPAN RANDOM FOREST

Classification Report:

	precision	recall	f1-score	support
Lulus	0.97	1.00	0.99	70
Tidak Lulus	1.00	0.80	0.89	10
accuracy			0.97	80
macro avg	0.99	0.90	0.94	80
<u>weighted</u> avg	0.98	0.97	0.97	80
Akurasi Random Forest:	97.50%			

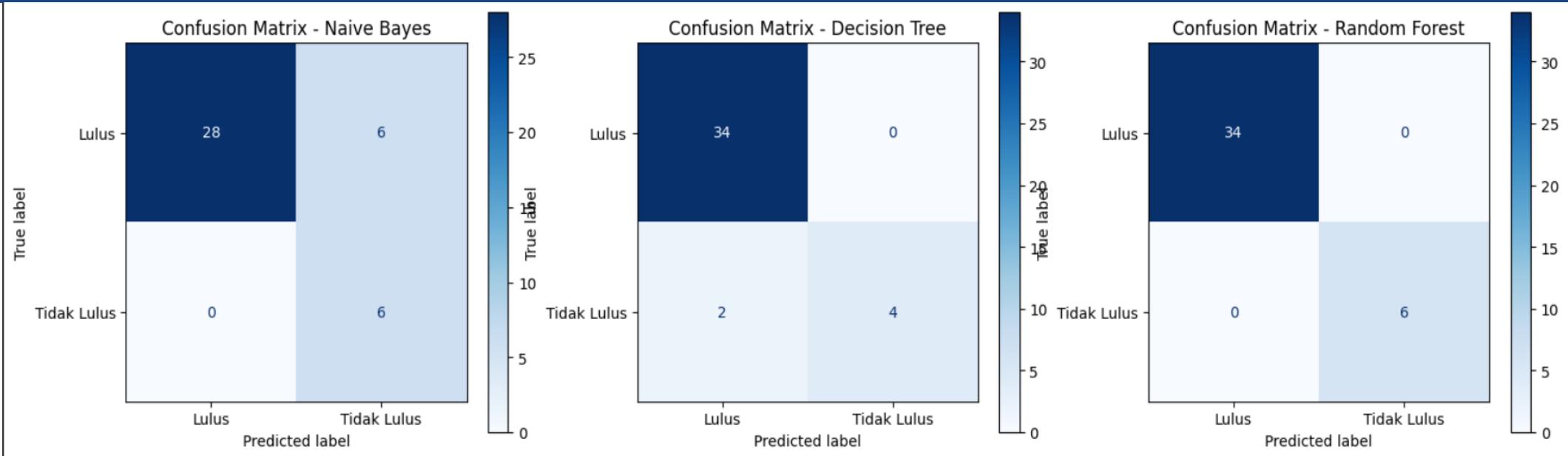
Confusion Matrix:

[[70 0]
[2 8]]

Berbeda halnya dengan algoritma **Random Forest**, di mana terjadi perbedaan akurasi antara kedua skenario. Pada skenario 80:20, Random Forest mencatat akurasi tertinggi yaitu **97.50%**, sedangkan pada skenario 60:40, akurasinya sedikit menurun menjadi **96.88%**. Meskipun terjadi penurunan, selisih akurasi ini relatif kecil dan tidak berdampak signifikan terhadap keandalan model. Penurunan ini dapat dijelaskan melalui karakteristik ensemble learning yang digunakan oleh Random Forest, di mana semakin banyak data pelatihan yang tersedia, semakin baik model dapat membangun variasi pohon keputusan untuk menghasilkan prediksi yang akurat.

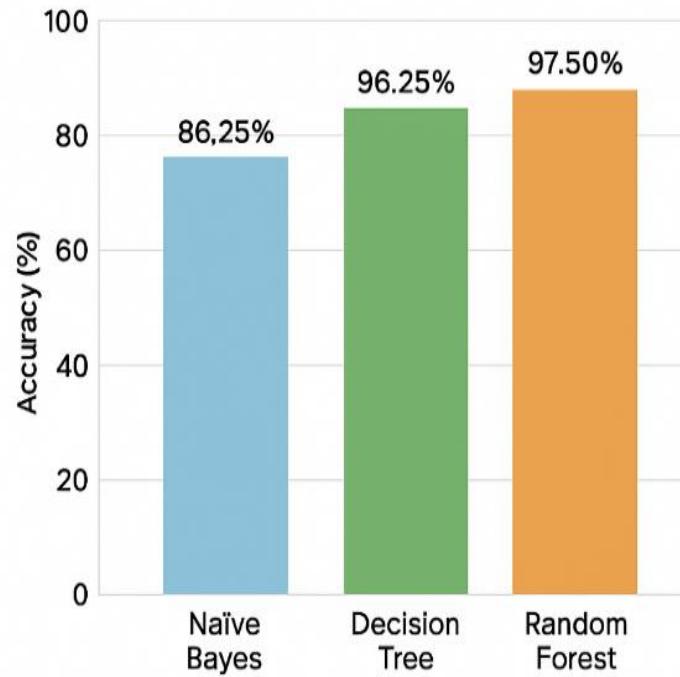


Visualisasi Confusion Matrix



confusion matrix dari tiga algoritma: Naive Bayes, Decision Tree, dan Random Forest. Naive Bayes memprediksi 28 "Lulus" dengan benar dan 6 salah, sedangkan semua "Tidak Lulus" diprediksi benar. Decision Tree memprediksi 34 "Lulus" dengan benar, 2 "Tidak Lulus" salah prediksi. Random Forest memberikan hasil terbaik dengan semua data diprediksi benar tanpa kesalahan.

Kesimpulan



Hasil penelitian menunjukkan bahwa algoritma **Random Forest** memiliki **performa terbaik** dalam memprediksi kelulusan mahasiswa, dengan akurasi tertinggi sebesar 97.50% pada skenario 80:20 dan 96.88% pada skenario 60:40. **Decision Tree** juga menunjukkan performa yang konsisten dengan akurasi 96.25% di kedua skenario, menandakan model yang stabil dan andal. Sementara itu, **Naïve Bayes** menghasilkan akurasi stabil sebesar 86.25%, namun masih berada di bawah dua algoritma lainnya. Hal ini menunjukkan bahwa meskipun Naïve Bayes cukup stabil, algoritma ini kurang optimal dalam menangani data akademik yang lebih kompleks.



TEMUAN PENTING

- Forest unggul dalam akurasi dan kestabilan model prediksi.
- Ketidakseimbangan data kelas “Lulus” dan “Tidak Lulus” memengaruhi nilai recall.
- Perlu teknik balancing data (oversampling/undersampling) agar hasil lebih baik.



MANFAAT PENELITIAN

- Membantu perguruan tinggi dalam **memantau dan memprediksi kelulusan mahasiswa.**
- Dapat menjadi dasar sistem pendukung keputusan akademik.
- Memberikan kontribusi untuk peningkatan mutu pendidikan berbasis data.



www.umsida.ac.id



[umsida1912](https://www.instagram.com/umsida1912/)



[umsida1912](https://twitter.com/umsida1912)



universitas
muhammadiyah
sidoarjo



[umsida1912](https://www.youtube.com/umsida1912)



Referensi

- U. Putra and I. Yptk, “Penerapan algoritma klasifikasi untuk prediksi tingkat kelulusan mahasiswa menggunakan rappidminer,” vol. 6, no. 1, pp. 376–388, 2025, doi: 10.46576/djtechno.
- D. Mardinah and M. Thoriq, “Algoritma Multi Layer Perceptron sebagai Prediksi Kelulusan Mahasiswa Universitas Adzkia Tepat Waktu berdasarkan jenis kelamin , Indeks Prestasi Semester , dan Jumlah SKS,” vol. 1, no. 2, pp. 26–35, 2024
- R. M. Ubaidilah, “Prediksi Kelulusan Mahasiswa Berdasarkan Data Kunjung dan Peminjaman Buku menggunakan Rapid Miner dengan Metode C.45 dan Random Forest,” Int. Res. Big-Data Comput. Technol. IRobot, vol. 7, no. 2, pp. 14–20, 2023, doi: 10.53514/ir.v7i2.410.
- A. Wahyudi, “Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Decision Tree Dan Naïve Bayes,” J. Permata Indones., vol. 14, no. 2, pp. 132–138, 2023, doi: 10.59737/jpi.v14i2.276.
- E. Ismanto and M. Novalia, “Komparasi Kinerja Algoritma C4.5, Random Forest, dan Gradient Boosting untuk Klasifikasi Komoditas,” Techno.Com, vol. 20, no. 3, pp. 400–410, 2021, doi: 10.33633/tc.v20i3.4576.
- T. Wahyuningsih, D. Manongga, I. Sembiring, and S. Wijono, “Comparison of Effectiveness of Logistic Regression, Naive Bayes, and Random Forest Algorithms in Predicting Student Arguments,” Procedia Comput. Sci., vol. 234, pp. 349–356, 2024, doi: 10.1016/j.procs.2024.03.014.
- J. M. Polgan et al., “Optimasi Prediksi Kelulusan Mahasiswa Menggunakan Random Forest untuk Meningkatkan Tingkat Retensi,” vol. 13, pp. 2364–2374, 2025.

Referensi

- M. A. Nurrohmat, “Aplikasi Pemrediksi Masa Studi dan Predikat Kelulusan Mahasiswa Informatika Universitas Muhammadiyah Surakarta Menggunakan Metode Naive Bayes,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 1, no. 1, pp. 29–34, 2015, doi: 10.23917/khif.v1i1.1179.
- Oon Wira Yuda, Darmawan Tuti, Lim Sheih Yee, and Susanti, “Penerapan Penerapan Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Random Forest,” *SATIN - Sains dan Teknol. Inf.*, vol. 8, no. 2, pp. 122–131, 2022, doi: 10.33372/stn.v8i2.885.
- A. Performance, E. Alshdaifat, D. Alshdaifat, A. Alsarhan, F. Hussein, and S. Moh, “The Effect of Preprocessing Techniques , Applied to Numeric,” *Data*, vol. 6, no. 11, 2021.
- H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, “Predicting Students Graduate on Time Using C4.5 Algorithm,” *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, p. 67, 2021, doi: 10.20473/jisebi.7.1.67-73.
- N. Sitohang, “Jurnal Sains Informatika Terapan (JSIT),” Penerapan Data Min. Untuk Peringatan Dini Banjir Menggunakan Metod. Klustering K-Means, vol. 2, no. 1, pp. 16–20, 2023
- R. Bakri, N. P. Astuti, and A. S. Ahmar, “Evaluating Random Forest Algorithm in Educational Data Mining: Optimizing Graduation on-time prediction using Imbalance Methods,” *ARRUS J. Soc. Sci. Humanit.*, vol. 4, no. 1, pp. 108–116, 2024, doi: 10.35877/soshum2449.
- U. Indahyanti, N. L. Azizah, and H. Setiawan, “Pendekatan Ensemble Learning Untuk Meningkatkan Akurasi Prediksi Kinerja Akademik Mahasiswa,” *J. Sains dan Inform.*, vol. 8, no. 2, pp. 160–169, 2022, doi: 10.34128/jsi.v8i2.459.

