

# PREDIKSI KINERJA AKADEMIK MAHASISWA MENGGUNAKAN PENDEKATAN MACHINE LEARNING

Oleh :

Berliana Aulia Agustin

211080200163

Program Studi Informatika  
Universitas Muhammadiyah Sidoarjo  
2025



# Latar Belakang

1. Faktor terpenting dalam pembangunan suatu bangsa adalah 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

Rumusan masalah pada penelitian ini adalah bagaimana membangun pemodelan klasifikasi untuk prediksi kinerja akademik mahasiswa menggunakan algoritma Naive Bayes, Decision Tree, dan Support Vector Machine untuk mencari akurasi terbaik. Klasifikasi yang dihasilkan berupa 3 kelas lulus, tidak lulus, dan remidi.



# Tujuan

Tujuan yang ingin dicapai pada penelitian ini adalah :

1. Menggunakan machine learning untuk mengevaluasi kebiasaan kerja mahasiswa
2. Membandingkan performa tiga algoritma klasifikasi : Naïve Bayes, Decision Tree, dan Support Vector Machine (SVM)
3. Menentukan algoritma dengan akurasi terbaik untuk prediksi akademik



[www.umsida.ac.id](http://www.umsida.ac.id)



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



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



universitas  
muhammadiyah  
sidoarjo

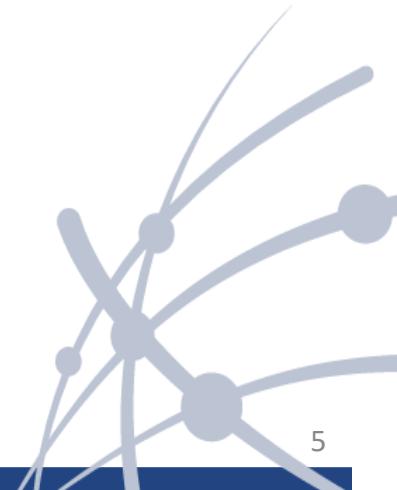


[umsida1912](https://www.youtube.com/umsida1912)



# Analisis GAP

Pembaruan yang akan diciptakan pada penelitian ini adalah menggunakan tiga algoritma yang berbeda dan membandingkan akurasi dengan menghasilkan pembagian data terbaik (40 : 60) untuk optimasi akurasi dan memanfaatkan Confusion Matrix dan F1 Score untuk evaluasi model yang menggunakan dataset spesifik dari e-learning program studi informatika.



# Metode Penelitian

•Waktu dan Lokasi: Semester 7, 2024-2025, Universitas Muhammadiyah Sidoarjo

•Alat dan Bahan:

- Data akademik mahasiswa (UTS, UAS, tugas, presensi, dll.)
- Ms. Excel untuk pengolahan data awal
- Pemrograman Python untuk analisis



[www.umsida.ac.id](http://www.umsida.ac.id)



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



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



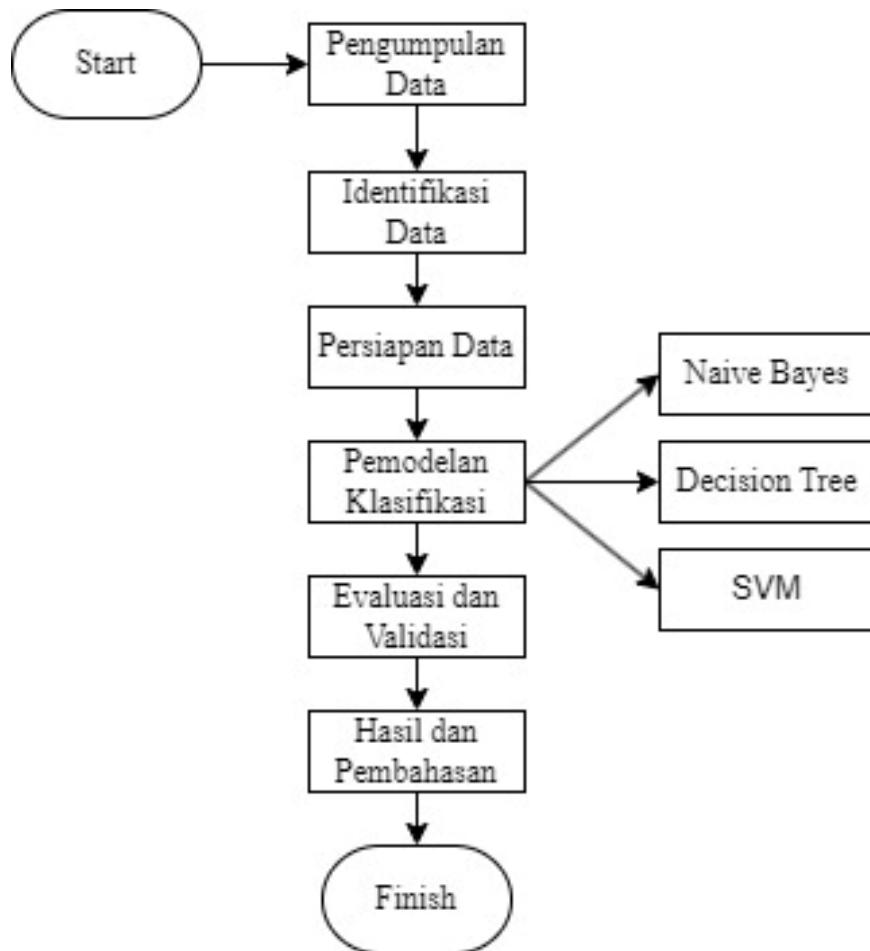
universitas  
muhammadiyah  
sidoarjo



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



# Metode Penelitian



- **Pengumpulan Data:** Data dari log aktivitas e-learning
- **Identifikasi Data:** Memastikan data sesuai dan lengkap
- **Persiapan Data:** Konversi data ke format numerik
- **Pemodelan Klasifikasi:** Naive Bayes, Decision Tree, SVM
- **Evaluasi dan Validasi:** Confusion Matrix & F1 Score



[www.umsida.ac.id](http://www.umsida.ac.id)



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



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



[universitas  
muhammadiyah  
sidoarjo](https://www.facebook.com/universitasmuhammadiyahsidoarjo)



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

# Hasil

Algoritma	Waktu Proses	Ketelitian	Pembagian Data Terbaik
Naive Bayes	18.47 ms	95.99%	40 : 60
Decission Tree	9.58 ms	98.51%	40 : 60
Support Vector Machine	0.05 ms	97.01%	40 : 60

Hasil pemodelan klasifikasi menggunakan algoritma Naive Bayes, Decision Tree, dan Support Vector Machine pada penelitian ini, menunjukkan bahwa pembagian data yang paling terbaik terdapat pada pembagian data 40 : 60 dimana pada algoritma Decision Tree memiliki tingkat akurasi tertinggi yaitu 98,51% diikuti oleh SVM 97,01% dan Naive Bayes 95,99%

# Pembahasan

## DATASET

Data yang diperoleh dari e-learning program studi informatika mata kuliah algoritma pemrograman data dan algoritma struktur dapat dilihat pada gambar dibawah ini, Terdapat 1122 dataset.

No	Presensi	Keaktifan	TUGAS	UTS	UAS	NA	Aslab	UKM	Kelulusan
1	75	75	70	75	75	73,8	Tidak	Tidak	Remidi
2	75	75	81	70	75	75,3	Tidak	Tidak	Lulus
3	94	85	83	90	75	84,4	Tidak	Tidak	Lulus
4	100	85	79	80	75	81,6	Tidak	Tidak	Lulus
5	100	85	85	83	83	85,9	Tidak	Tidak	Lulus
6	100	85	81	85	78	84,1	Tidak	Tidak	Lulus
7	100	85	81	85	83	85,3	Tidak	Tidak	Lulus
8	100	85	83	75	78	82,1	Tidak	Tidak	Lulus
9	100	85	83	90	78	85,9	Tidak	Ya	Lulus
10	88	85	83	83	80	83,1	Tidak	Ya	Lulus
11	94	85	85	83	83	85,1	Tidak	Ya	Lulus
12	100	85	83	83	78	84,1	Tidak	Ya	Lulus
13	100	85	83	83	83	85,4	Tidak	Ya	Lulus
14	94	85	83	88	83	85,9	Tidak	Ya	Lulus
15	0	0	0	0	0	0	Tidak	Tidak	Tidak Lulus
16	88	85	83	85	75	82,4	Ya	Ya	Lulus



# Pembahasan

## DATA CLEANING

Setelah tahap seleksi data selesai dilakukan, Langkah selanjutnya adalah proses data cleaning dari hasil seleksi sebelumnya. Proses ini bertujuan untuk membersihkan data dari duplikasi, data kosong, dan nilai yang tidak valid.

```
▶ df.duplicated().sum()  
→ 956  
[ ] df = df.drop_duplicates()  
df.shape  
→ (166, 9)
```



# Pembahasan

## TRANSFORMASI DATA

Langkah selanjutnya yakni transformasi data. Transformasi data merupakan langkah penting dalam proses analisis, dimana data diubah menjadi format yang sesuai dari kategorikal menjadi numerik agar dapat diolah dengan baik. Berikut tahapan transformasi data dari data yang berbentuk kategorikal hingga menjadi numerik.

No	Presensi	Keaktifan	TUGAS	UTS	UAS	NA	Aslab	UKM	Kelulusan	
0	1	75	75	70	75	75	73,8	Tidak	Tidak	Remidi
1	2	75	75	81	70	75	75,3	Tidak	Tidak	Lulus
2	3	94	85	83	90	75	84,4	Tidak	Tidak	Lulus
3	4	100	85	79	80	75	81,6	Tidak	Tidak	Lulus
4	5	100	85	85	83	83	85,9	Tidak	Tidak	Lulus

```
if 'No' in df.columns:  
    df.drop('No', axis=1, inplace=True)  
df.head()
```

	Presensi	Keaktifan	TUGAS	UTS	UAS	NA	Aslab	UKM	Kelulusan
0	75	75	70	75	75	73,8	0	0	1
1	75	75	81	70	75	75,3	0	0	0
2	94	85	83	90	75	84,4	0	0	0
3	100	85	79	80	75	81,6	0	0	0
4	100	85	85	83	83	85,9	0	0	0



# Pembahasan

## PENERAPAN NAÏVE BAYES

Berdasarkan Teorema Bayes untuk menghitung probabilitas yang digunakan untuk klasifikasi berbasis probabilitas. Akurasi terbaik 95.99% dengan pembagian data 40:60. dengan proses yang cepat dalam komputasi namun kurang akurat dibandingkan Decision Tree.

	precision	recall	f1-score	support
0	1.00	0.95	0.98	400
1	0.75	1.00	0.86	12
2	0.73	1.00	0.84	37
accuracy			0.96	449
macro avg	0.83	0.98	0.89	449
weighted avg	0.97	0.96	0.96	449

Akurasi NB: 95.99%  
Training Time: 71.33 ms  
Prediction Time: 35.54 ms



# Pembahasan

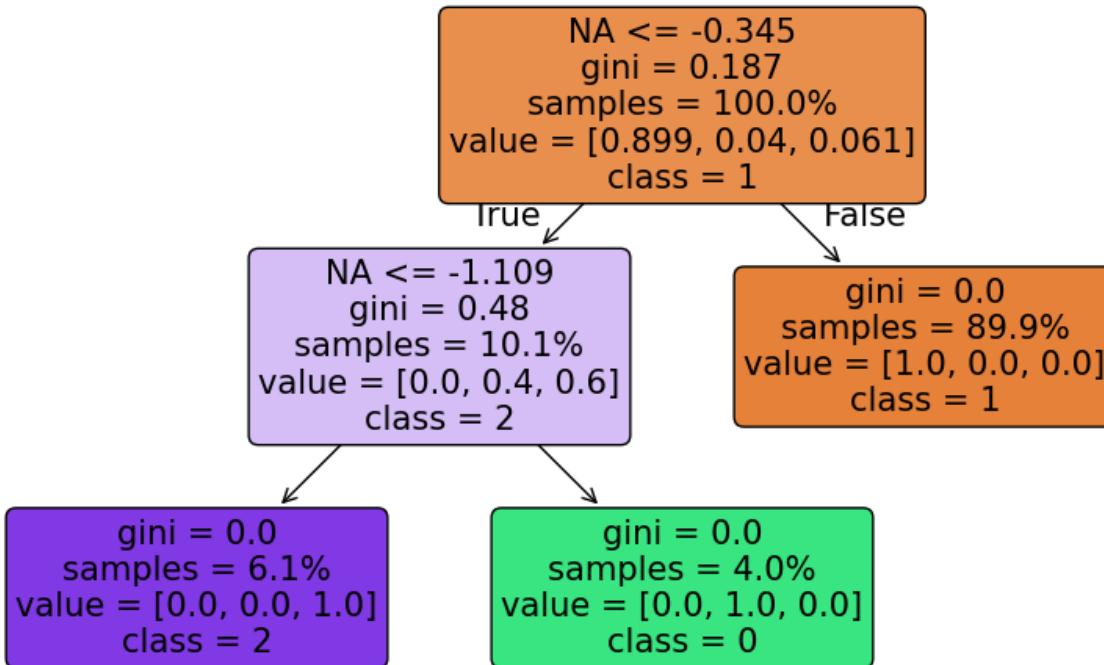
## PENERAPAN DECISION TREE

Menggunakan pohon keputusan untuk klasifikasi dengan memilih fitur terbaik untuk membagi data yang mendapatkan nilai akurasi terbaik 98.51% dengan pembagian data 40:60, dimana algoritma decision tree paling akurat dalam penelitian ini.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	60
1	0.50	1.00	0.67	1
2	1.00	0.83	0.91	6
accuracy			0.99	67
macro avg	0.83	0.94	0.86	67
weighted avg	0.99	0.99	0.99	67
Akurasi Model Decision Tree: 98.51%				
Training Time: 8.42 ms				
Prediction Time: 7.41 ms				



# Pohon Keputusan



Berikut gambar mengenai keputusan terkait penelitian ini. Diagram pohon keputusan tersebut diperoleh dari hasil proses cleaning dan data transformasi :

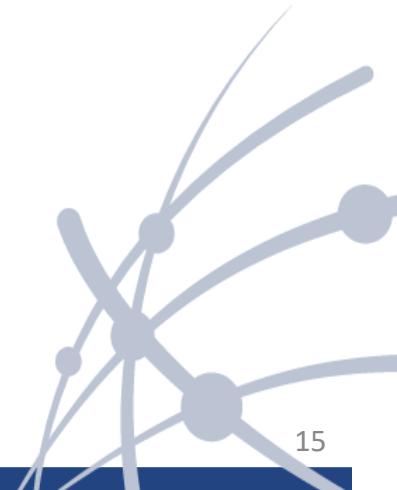
- NA adalah fitur yang memiliki pengaruh besar terhadap keputusan kelulusan.
- Jika  $NA \leq -0.345$ , maka ada kemungkinan lebih besar bahwa mahasiswa tidak lulus (kelas 2).
- Jika  $NA > -0.345$ , maka hampir semua mahasiswa lulus (kelas 1).
- Semakin kecil nilai NA, semakin besar kemungkinan mahasiswa tidak lulus.

# Pembahasan

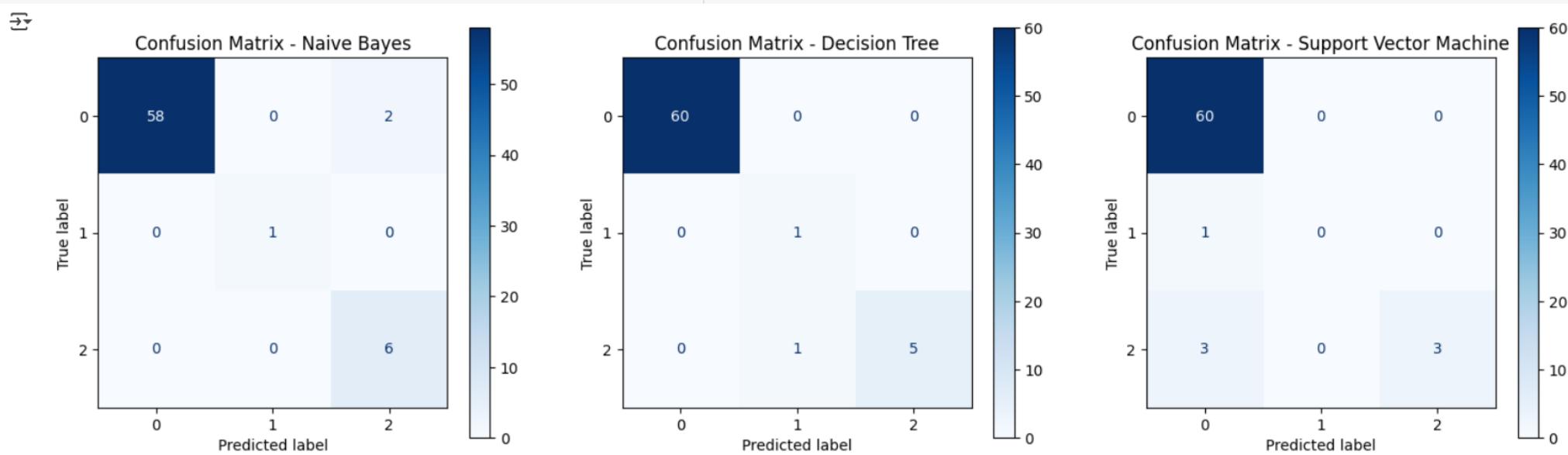
## PENERAPAN SUPPORT VECTOR MACHINE

Algoritma ini menggunakan hyperplane untuk memisahkan kelas data. Cocok untuk data dengan dimensi tinggi yang mendapatkan nilai akurasi terbaik: 97.01% dengan pembagian data 40:60 dengan waktu proses tercepat (0.05 ms).

	precision	recall	f1-score	support
0	0.97	1.00	0.98	60
1	0.00	0.00	0.00	1
2	1.00	0.83	0.91	6
accuracy			0.97	67
macro avg		0.66	0.61	67
weighted avg		0.96	0.97	67
Akurasi SVM : 97.01%				
Training Time: 9.58 ms				
Prediction Time: 0.05 ms				



# Visualisasi Confusion Matrix



Gambar menunjukkan tiga matriks konfusi dari model Naïve Bayes, Decision Tree, dan Support Vector Machine (SVM). Model Decision Tree dan SVM memiliki performa terbaik dalam mengklasifikasikan kelas mayoritas (label 0), dengan Decision Tree memiliki sedikit kesalahan pada kelas 2. Naïve Bayes menunjukkan beberapa kesalahan klasifikasi pada kelas 0 dan 2. Secara keseluruhan, Decision Tree tampaknya memiliki kinerja yang lebih stabil dibandingkan model lainnya dalam dataset ini.

# Kesimpulan

**Decision Tree** memiliki akurasi terbaik (**98.51%**), meskipun SVM lebih cepat dalam pemrosesan dengan menggunakan pendekatan Machine learning terbukti efektif dalam mengevaluasi kinerja akademik mahasiswa. Hasil penelitian dapat digunakan untuk sistem eringatan dini di perguruan tinggi



# Referensi

- [1] R.- Annisa and A.- Sasongko, “Prediksi Nilai Akademik Mahasiswa Menggunakan Algoritma Naïve Bayes,” JST (Jurnal Sains dan Teknol., vol. 9, no. 1, pp. 1–10, 2020, doi: 10.23887/jstundiksha.v9i1.19488.
- [2] A. Putri et al., “Komparasi Algoritma K-NN, Naive Bayes dan SVM untuk Prediksi Kelulusan Mahasiswa Tingkat Akhir,” MALCOM Indones. J. Mach. Learn. Comput. Sci., vol. 3, no. 1, pp. 20–26, 2023, doi: 10.57152/malcom.v3i1.610.
- [3] 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.
- [4] M. Windarti and A. Suradi, “Perbandingan Kinerja 6 Algoritme Klasifikasi Data Mining untuk Prediksi Masa Studi Mahasiswa,” Telematika, vol. 12, no. 1, p. 14, 2019, doi: 10.35671/telematika.v12i1.778.
- [5] M. P. A. Ariawan, I. B. A. Peling, and G. B. Subiksa, “Prediksi Nilai Akhir Mata Kuliah Mahasiswa Menggunakan Metode K-Means Clustering (Studi Kasus : Mata Kuliah Pemrograman Dasar),” J. Nas. Teknol. dan Sist. Inf., vol. 9, no. 2, pp. 122–131, 2023, doi: 10.25077/teknosi.v9i2.2023.122-131.
- [6] A. Rohman and A. Rufiyanto, “Implementasi Data Mining Dengan Algoritma Decision Tree C4 . 5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandaran,” Proceeding SINTAK 2019, pp. 134–139, 2019.



[www.umsida.ac.id](http://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

- [7] E. Haryatmi and S. Pramita Hervianti, “Penerapan Algoritma Support Vector Machine Untuk Model Prediksi Kelulusan Mahasiswa Tepat Waktu,” *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 2, pp. 386–392, 2021, doi: 10.29207/resti.v5i2.3007.
- [8] Satrio Junaidi, R. Valicia Anggela, and D. Kariman, “Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Network (ANN),” *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 109–119, 2024, doi: 10.52158/jacost.v5i1.489.
- [9] T. Algorithm, “Penerapan Machine Learning Untuk Memprediksi Kelulusan Mahasiswa Menggunakan Algoritma Decision Tree Application of Machine Learning to Predict Student Graduation Using the Decision,” pp. 1–6, 2024.
- [10] Z. Saputra, D. Sartika, and M. Haviz Irfani, “Prediksi Calon Mahasiswa Penerima KIP Pada Universitas Indo Global Mandiri menggunakan Algoritma Decision Tree,” *J. Rekayasa Tek. Inform. dan Inf.*, vol. 43, no. 3, pp. 231–240, 2024, [Online]. Available: <https://docs.python.org/3.13/tutorial/index.html>



www.umsida.ac.id



umsida1912



umsida1912



universitas  
muhammadiyah  
sidoarjo



umsida1912



DARI SINI PENCERAHAN BERSEMI