

Implementasi Algoritma Decision Trees C4.5, Naive Bayes, dan SVM Untuk Prediksi Partisipasi Pemilu Berdasarkan Data Indeks Desa Membangun Di Kawasan Mataraman

Oleh:

Mahmud Adi Yuwanto

Arif Senja Fitrani, S.Kom., M.Kom

Progam Studi Informatika

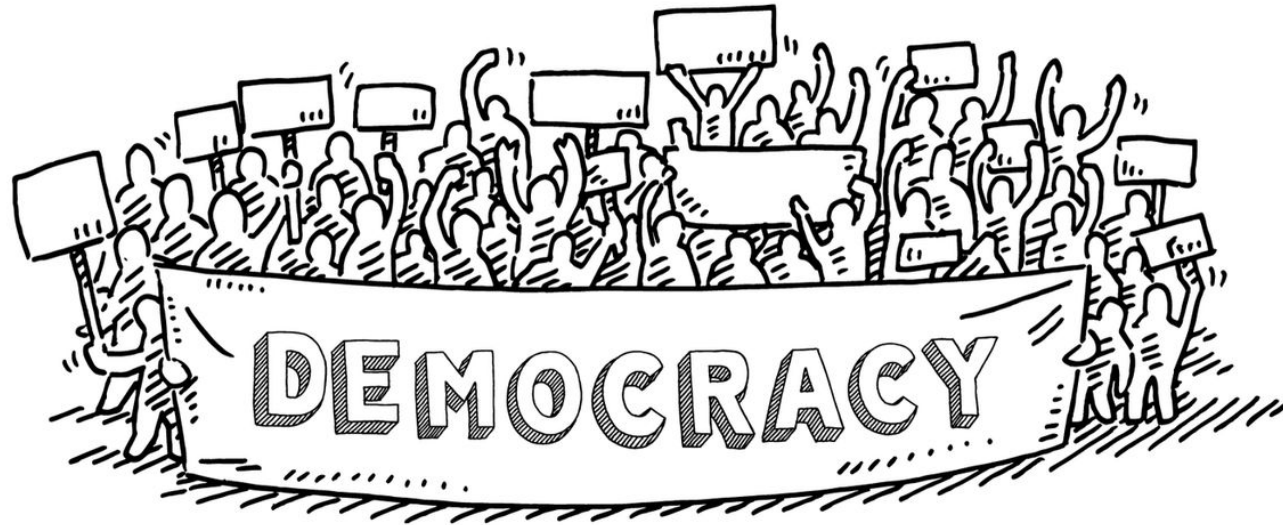
Universitas Muhammadiyah Sidoarjo

2025



Pendahuluan

Meskipun akses informasi politik semakin terbuka, tingkat partisipasi pemilu masih bervariasi antarwilayah, seperti di kawasan Mataraman. Faktor non-politis turut memengaruhi antusiasme masyarakat dalam menggunakan hak pilihnya. Penelitian ini bertujuan untuk memanfaatkan algoritma Decision Tree C4.5 dengan data Indeks Desa Membangun (IDM) untuk memprediksi partisipasi pemilu. Pendekatan ini diharapkan dapat menghasilkan model prediksi yang akurat, serta mendukung strategi peningkatan partisipasi pemilu berbasis data.



Rumusan Masalah

- Bagaimana implementasi algoritma Decision Trees dalam memprediksi tingkat partisipasi pemilu berdasarkan data Indeks Desa Membangun (IDM) di kawasan Mataraman?
- Bagaimana tingkat akurasi model prediksi yang dibangun menggunakan algoritma Decision Trees dalam mengklasifikasikan tingkat partisipasi pemilu?

Tujuan

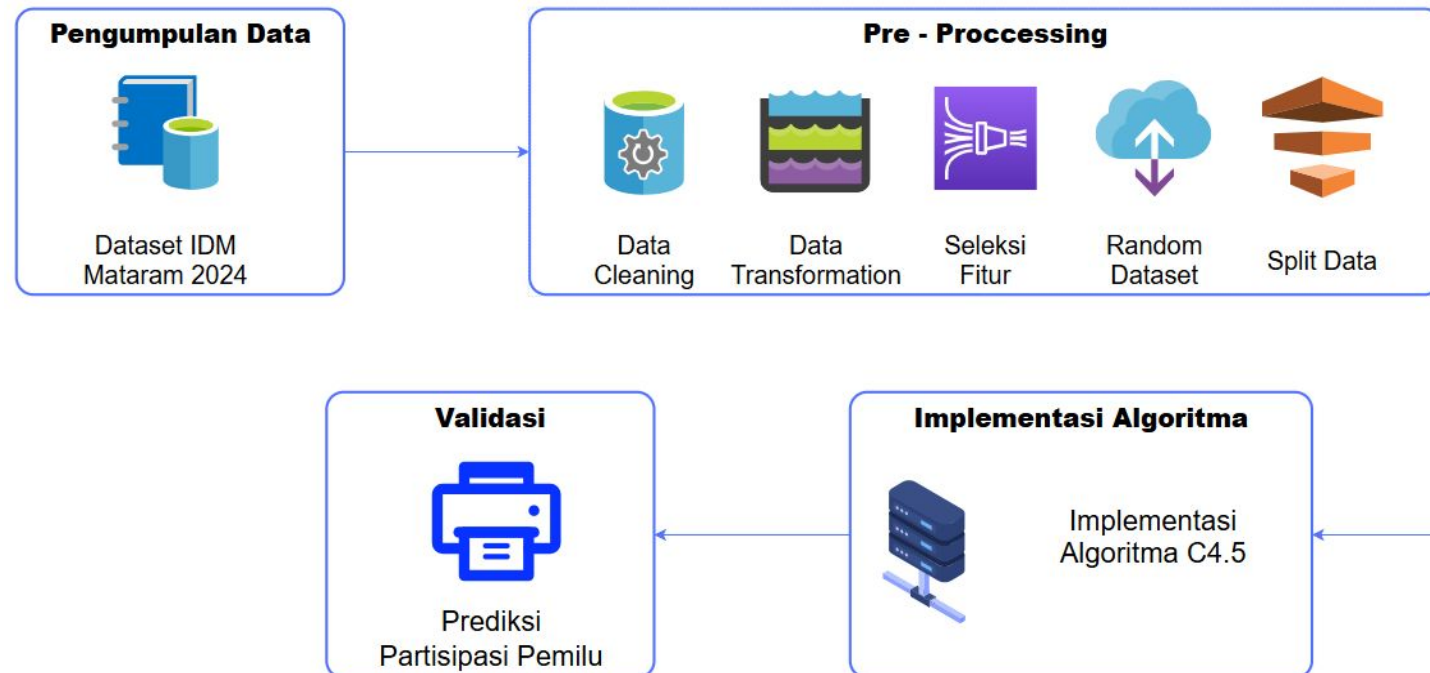
- Menerapkan algoritma Decision Trees untuk memprediksi tingkat partisipasi pemilu berdasarkan data IDM di kawasan Mataraman.
- Mengevaluasi tingkat akurasi model prediksi Decision Trees dalam mengklasifikasikan kategori partisipasi pemilu.

Batasan Masalah

1. **Data yang digunakan** adalah data Indeks Desa Membangun (IDM) tahun terakhir yang tersedia serta data partisipasi pemilu dari desa-desa di kawasan **Mataraman** (Ngawi, Magetan, Ponorogo, Pacitan, Nganjuk, dan Madiun).
2. **Periode waktu** penelitian dibatasi pada tahun **2024** sehingga tidak melibatkan data runtun waktu (time series) dari tahun-tahun sebelumnya.
3. **Jenis atribut yang digunakan** hanya mencakup **indikator kuantitatif** pada IDM, karena indikator tersebut dapat dioperasionalkan sebagai variabel independen (X) untuk menjelaskan partisipasi pemilu.
 - Kolom administratif (Provinsi, Kabupaten, Kecamatan, Desa, Kode Wilayah, Tahun) hanya digunakan untuk identifikasi dan penggabungan data, tidak dijadikan variabel analisis.
 - Kolom kategori/sub-dimensi yang berupa label teks tidak digunakan karena tidak bersifat kuantitatif.
 - Kolom **Total Skor** dan **Status IDM** tidak digunakan untuk menghindari bias, sebab keduanya merupakan hasil agregasi keseluruhan indikator.
4. **Seleksi atribut** dilakukan dari 48 indikator IDM menjadi **9 atribut utama** dengan pendekatan analisis korelasi. Proses reduksi dilakukan melalui:
 - Transformasi matriks korelasi ke bentuk pasangan atribut.
 - Seleksi atribut dengan korelasi positif kuat.
 - Eliminasi atribut dengan korelasi lemah atau redundan.
 - Hasil akhir atribut yang dipertahankan adalah: **X1, X2, X3, X4, X9, X10, X14, X18, dan X19**.

Metode Penelitian

Tahapan penelitian merupakan langkah-langkah sistematis yang dilakukan mulai dari perencanaan awal hingga evaluasi akhir guna mencapai tujuan penelitian [11]. Pada penelitian ini, proses dimulai dari studi literature, dilanjutkan dengan pengumpulan data, tahap pre-processing, pemrosesan data menggunakan algoritma C4.5, hingga tahap validasi model.



Metode Penelitian

2. Pre-Processing

Tahap awal penerapan algoritma C4.5 pada data IDM meliputi penggabungan, modifikasi, pembersihan, dan penyederhanaan data untuk prediksi partisipasi pemilu di wilayah Mataraman.

Data Cleaning:

Mengidentifikasi dan menangani kesalahan seperti duplikasi, ketidakkonsistenan, dan data tidak lengkap dengan cara menghapus atau memperbaikinya.

Data Transformation:

Mengubah data ke format numerik dan menyeragamkan skala agar sesuai dengan kebutuhan analisis C4.5.

Seleksi Fitur:

Dari 48 atribut, 6 kelompok indikator yang kurang relevan dihapus. Hasilnya, digunakan 27 atribut dari 7 kelompok indikator dan 1 atribut target, yaitu partisipasi pemilu.

Metode Penelitian

Random Dataset

Data diacak untuk memastikan distribusi atribut merata dan mencegah bias saat pelatihan dan pengujian model.

Split Data

Dataset dibagi menjadi 70% untuk training dan 30% untuk testing guna mengukur kemampuan generalisasi model.

3. Process

Data yang telah dibersihkan diuji dengan algoritma Decision Tree C4.5, menggunakan 72.311 entri yang dibagi 70% untuk pelatihan (50.618 data) dan 30% untuk pengujian (21.692 data). C4.5 dipilih karena mendukung atribut numerik dan kategorikal.

4. Output

Model memprediksi partisipasi pemilu dalam dua kategori, “Tinggi” dan “Rendah”, dan hasilnya divisualisasikan dalam bentuk pohon keputusan serta tabel evaluasi.

5. Analisis / Evaluasi

Kinerja model diukur menggunakan akurasi, presisi, recall, dan F1-score untuk menilai efektivitas klasifikasi partisipasi pemilu.

Hasil & Pembahasan

	Precisi on	Recall	F1-Score	Support
0	0.58	0.49	0.53	966
1	0.86	0.90	0.88	3405
Accuracy			0.81	4371
Macro Avg	0.72	0.69	0.70	4371
Weighted Avg	0.80	0.81	0.80	4371

Classification Report C4.5

	Precisi on	Recall	F1-Score	Support
0	0.00	0.00	0.00	966
1	0.78	1.00	0.88	3405
Accuracy			0.78	4371
Macro Avg	0.39	0.50	0.44	4371
Weighted Avg	0.61	0.78	0.68	4371

Classification Report Naïve Bayes

Hasil & Pembahasan

	Precisi on	Recall	F1-Score	Support
0	0.00	0.00	0.00	966
1	0.78	1.00	0.88	3405
Accuracy			0.78	4371
Macro Avg	0.39	0.50	0.44	4371
Weighted Avg	0.61	0.78	0.68	4371

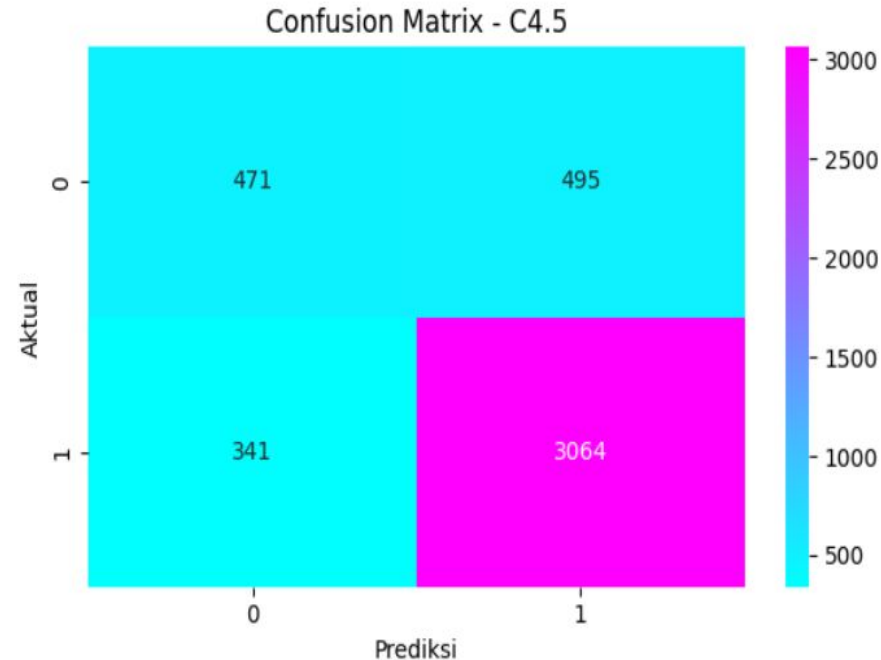
Classification Report SVM

3 Tabel diatas menyajikan hasil evaluasi tiga algoritma klasifikasi, yaitu C4.5 Decision Tree, Naive Bayes, dan Support Vector Machine (SVM), berdasarkan nilai akurasi, precision, recall, serta f1-score. Pada algoritma C4.5 Decision Tree yang ditunjukkan pada Tabel 2, model menghasilkan akurasi sebesar 80,87%. Model mampu mengenali kelas Tinggi dengan cukup baik, ditunjukkan oleh nilai recall sebesar 90% dan precision sebesar 0,86, yang menghasilkan f1-score 0,88. Namun, performa model terhadap kelas Rendah masih terbatas, dengan recall hanya 49% dan precision 0,58. Temuan ini mengindikasikan bahwa Decision Tree cenderung lebih optimal dalam mengenali pola pada kelas mayoritas (Tinggi), meskipun masih terdapat kesalahan klasifikasi pada kelas minoritas (Rendah).

Hasil & Pembahasan

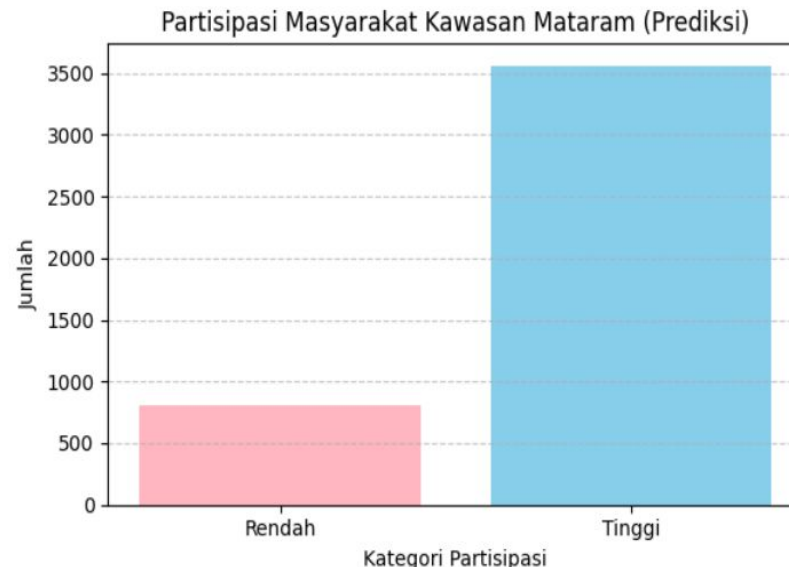
- Berbeda halnya dengan algoritma Naive Bayes yang ditunjukkan pada tabel 3, model hanya mencapai akurasi sebesar 77,90%. Pada kelas Tinggi, model memperlihatkan kinerja sangat baik dengan recall 100% dan precision 0,78, menghasilkan f1-score 0,88. Akan tetapi, model sama sekali gagal mengenali kelas Rendah, yang ditunjukkan oleh nilai precision, recall, dan f1-score 0,00. Kondisi ini memperlihatkan bahwa Naive Bayes tidak mampu mengklasifikasikan data minoritas, sehingga hasil prediksi menjadi sangat bias terhadap kelas mayoritas. Sementara itu, algoritma SVM pada Tabel 4, menunjukkan performa yang identik dengan Naive Bayes, dengan akurasi 77,90% dan pola evaluasi yang sama. Kelas Tinggi dikenali dengan sangat baik (recall 100%, f1-score 0,88), namun kelas Rendah sama sekali tidak terklasifikasi (precision, recall, dan f1-score 0,00). Hal ini menegaskan bahwa model SVM juga mengalami bias kuat terhadap kelas mayoritas akibat distribusi data yang tidak seimbang.
- Secara keseluruhan, dari ketiga algoritma yang diuji, C4.5 Decision Tree memberikan hasil yang paling seimbang dengan akurasi tinggi dan kemampuan yang lebih baik dalam mengenali kedua kelas, meskipun kinerja terhadap kelas Rendah masih perlu ditingkatkan.

Hasil & Pembahasan



Gambar 2 menampilkan **Confusion Matrix** hasil pengujian model. Model mampu mengenali 3.064 data kelas Tinggi dan 471 data kelas Rendah secara tepat, meskipun masih terjadi mis-klasifikasi sebanyak 341 data Tinggi dan 495 data Rendah. Kondisi ini menunjukkan kecenderungan model lebih optimal dalam mengidentifikasi kelas mayoritas, sementara akurasi terhadap kelas minoritas masih rendah, sehingga diperlukan analisis distribusi dan fitur untuk meningkatkan performa secara menyeluruh

Hasil & Pembahasan



Presentase Tingkat Partisipasi pada Gambar diatas menampilkan distribusi hasil prediksi partisipasi masyarakat di Mataram. Terlihat bahwa prediksi pada kategori Tinggi (± 3.500 data) jauh lebih dominan dibandingkan kategori Rendah (± 800 data). Pola ini konsisten dengan hasil Confusion Matrix yang menunjukkan model lebih akurat pada kelas mayoritas. Ketimpangan distribusi ini mengindikasikan adanya imbalanced dataset yang memicu bias model terhadap kelas Tinggi, sehingga diperlukan strategi lanjutan seperti resampling atau metode klasifikasi yang lebih adaptif terhadap kelas minoritas

Hasil & Pembahasan

Fitur	Rendah Tinggi Selisih		
Layanan Tenaga Kesehatan Lainnya	13,97	14,18	0,21
Akses terhadap SMA/SMK/MA/MAK/Sederajat	5,87	5,98	0,11
Akses terhadap SD/MI/Sederajat	7,29	7,39	0,10
Akses terhadap SMP/MTs/Sederajat	6,38	6,47	0,09
Keberadaan Ruang Publik Terbuka	2,63	2,70	0,08
Kegiatan Olahraga	1,98	2,05	0,07
Fasilitas Olahraga	2,71	2,77	0,06
Layanan Bidan	16,08	16,09	0,01
Akses terhadap PAUD/TK/Sederajat	10,73	10,61	-0,11

Tabel di atas menyajikan perbandingan rata-rata nilai tiap fitur berdasarkan kategori partisipasi (Rendah dan Tinggi). Secara umum, nilai rata-rata pada kategori partisipasi Tinggi menunjukkan angka yang lebih besar dibandingkan kategori Rendah, yang tercermin pada selisih positif di hampir semua fitur [15]. Perbedaan paling menonjol terlihat pada Layanan Tenaga Kesehatan Lainnya, dengan selisih 0,21, diikuti oleh Akses terhadap SMA/SMK/MA/MAK/Sederajat dan Akses terhadap SD/MI/Sederajat, masing-masing dengan selisih 0,11 dan 0,10. Hal ini menunjukkan bahwa desa dengan partisipasi tinggi umumnya memiliki akses pendidikan dan layanan kesehatan yang relatif lebih baik dibandingkan desa dengan partisipasi rendah [16]. Sebaliknya, pada fitur Akses terhadap PAUD/TK/Sederajat, nilai rata-rata pada partisipasi Rendah justru lebih tinggi dibandingkan partisipasi Tinggi dengan selisih -0,11. Temuan ini menunjukkan adanya pengecualian, di mana akses PAUD/TK tidak selalu berbanding lurus dengan tingkat partisipasi masyarakat. Secara keseluruhan, tabel ini mengindikasikan bahwa faktor layanan kesehatan dan akses pendidikan tingkat dasar hingga menengah berkontribusi positif terhadap tingginya partisipasi, sementara pada akses PAUD/TK terlihat adanya pola yang berbeda.

Kesimpulan

1. Algoritma **C4.5 Decision Tree** memberikan performa terbaik dengan akurasi **80,87%**, mampu mengenali kelas mayoritas (partisipasi Tinggi) meskipun masih lemah pada kelas minoritas (partisipasi Rendah).
2. Algoritma **Naïve Bayes** dan **SVM** cenderung bias pada kelas mayoritas dan gagal mengklasifikasikan kelas minoritas.
3. Permasalahan **imbalanced dataset** (partisipasi Tinggi lebih dominan dibanding Rendah) memengaruhi pola klasifikasi, sehingga model lebih sering memprediksi ke kelas mayoritas.
4. Hasil analisis fitur menunjukkan **akses pendidikan (SD, SMP, SMA/SMK)** dan **layanan kesehatan (tenaga kesehatan, bidan, posyandu)** berkontribusi positif terhadap meningkatnya partisipasi pemilu.
5. Akses **PAUD/TK** tidak berbanding lurus dengan tingkat partisipasi, berbeda dengan jenjang pendidikan lainnya.
6. Secara keseluruhan, **C4.5 Decision Tree** efektif untuk memahami hubungan antara indikator IDM dan partisipasi pemilu.
7. Diperlukan strategi lanjutan, seperti **resampling data** atau metode klasifikasi lain yang adaptif terhadap data tidak seimbang, agar prediksi kelas minoritas bisa lebih optimal.

Referensi

- [1] A. Sentimen, A. Pemilu, and A. N. Bayes, "SWADHARMA (JEIS)," vol. 05, pp. 131–139, 2025.
- [2] Y. Raharja, A. S. Fitriani, and R. Dijaya, "Klasifikasi Tingkat Partisipasi Pemilu Berdasarkan Sektor Industri Menggunakan Algoritma Naïve Bayes," J. Tekinkom (Teknik Inf. dan Komputer), vol. 7, no. 1, pp. 135–143, 2024, doi: 10.37600/tekinkom.v7i1.1204.
- [3] D. E. Safitri and A. S. Fitriani, "Implementasi Metode Klasifikasi Dengan Algoritma Support Vector Machine Kernel Gaussian Rbf Untuk Prediksi Partisipasi Pemilu Terhadap Demografi Kota Surabaya," Indones. J. Bus. Intell., vol. 5, no. 1, p. 36, 2022, doi: 10.21927/ijubi.v5i1.2259.
- [4] B. A. B. Iii, "digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id digilib.uinsby.ac.id," pp. 66–90.
- [5] T. dkk Madjid, "Peringkat Indeks Desa Membangun," p. 871, 2020.
- [6] H. Rusli, "PREDIKSI HASIL PEMILIHAN UMUM BERDASARKAN DATA MEDIA SOSIAL MENGGUNAKAN TEKNIK DATA MINING NAIVE BAYES," 2025. [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [7] A. S. Fitriani, "JTAM (Jurnal Teori dan Aplikasi Matematika) Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur," vol. 3, no. 2, pp. 98–104, 2019, doi: 10.31764/jtam.v3i2.995.

Referensi

- [8] M. N. Zarti, E. Sahputra,) ; Anisya Sonita, and Y. Apridiansyah, "Application Of Data Mining Using The Naïve Bayes Classification Method To Predict Public Interest Participation In The 2024 Elections Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes Untuk Memprediksi Partisipasi Minat Masyarakat Pada Pemilu 2024," J. Kom., vol. 3, no. 1, pp. 105–114, doi: 10.53697/jkomitek.v3i1.
- [9] S. Pokhrel, "No TitleEAENH," Ayan, vol. 15, no. 1, pp. 37–48, 2024.
- [10] D. Syahputra, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Menentukan Produk Laptop Terlaris," 2022.
- [11] A. W. Anggraeni, A. S. Fitrani, and A. Eviyanti, "Penerapan Algoritma Support Vector Machine untuk Memprediksi Tingkat Partisipasi Pemilu terhadap Kualitas Pendidikan," Edumatic J. Pendidik. Inform., vol. 8, no. 1, pp. 21–27, 2024, doi: 10.29408/edumatic.v8i1.24838.
- [13] D. M. Chulloh, A. S. Fitrani, I. R. Indra Astutik, and A. Eviyanti, "Uji Akurasi K-Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan," Jutisi : Jurnal Ilmiah Teknik Informatika dan Sistem Informasi, vol. 13, no. 1, p. 201, 2024, doi: 10.35889/jutisi.v13i1.1753.
- [14] D. Pilendia, S. Muhammadiyah, and S. Penuh, "PEMANFAATAN ADOBE FLASH SEBAGAI DASAR PENGEMBANGAN BAHAN AJAR FISIKA : STUDI LITERATUR," 2020. [Online]. Available: <http://ejournal.stkip-mmb.ac.id/index.php/pgsd/login>
- [15] W. Kurniawati, A. S. Fitrani, S. Aji, and S. Suprianto, "Analisis Kehadiran Pemilih di Tempat Pemungutan Suara Pada Pemilihan Presiden dengan Menggunakan Algoritma C4.5 (Studi Kasus di Desa Wonokasian Sidoarjo)," Journal of Electrical Engineering, vol. 1, no. 1, p. 18, 2024, doi: 10.47134/jte.v1i1.2475.
- [16] J. Riset and S. Informasi, "ANALISIS POLA KEHADIRAN MAHASISWA MENGGUNAKAN," vol. 2, no. 1, pp. 60–66, 2025.

Referensi

- [15] I. M. Widiarta, Y. Mulyanto, and A. Sutrianto, "Rancang Bangun Sistem Informasi Inventory Menggunakan Metode Agile Software Development (Studi Kasus Toko Nada)," Digital Transformation Technology (Digitech), vol. 3, no. Maret, p. 20, 2023.
- [16] M. A. Afandi and N. L. Azizah, "Digital Based Posyandu Service Design in Ganggang Panjang Village," Jul. 05, 2024. doi: 10.21070/ups.4924.
- [17] P. A. T. Subur et al., "Implementation of Agile Software Development in the Design of Management Information Systems Inventory and Finance at," 2025.

