

Implementation of C4.5 Decision Tree, Naïve Bayes, and SVM Algorithms for Predicting Election Participation Based on the Village Development Index Data in the Mataraman Region

[Implementasi Algoritma Decision Trees C4.5, Naïve Bayes, dan SVM untuk Prediksi Partisipasi Pemilu Berdasarkan Data Indeks Desa Membangun di Kawasan Mataraman]

Mahmud Adi Yuwanto¹⁾, Arif Senja Fitrani²⁾, Rohman Dijaya³⁾, Uce Indahyanti⁴⁾

¹⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

³⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

⁴⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: asfjim@umsida.ac.id

Abstract. Voter participation in general elections is an important indicator of democratic quality; however, participation rates often vary across regions. The Mataraman area in East Java illustrates this phenomenon, where some villages show high levels of participation while others remain relatively low. This study aims to predict voter participation levels based on the Village Development Index (Indeks Desa Membangun/IDM) using classification algorithms. The dataset used was the 2024 IDM publication, consisting of 72,311 entries with 48 initial indicators. After feature selection, 27 relevant attributes were obtained and analyzed using three algorithms: Decision Tree C4.5, Naïve Bayes, and Support Vector Machine (SVM). The results indicate that the C4.5 algorithm achieved the best performance, with an accuracy of 80.87% and an F1-score of 0.88 in the majority class (High participation). Meanwhile, Naïve Bayes and SVM tended to be biased toward the majority class and failed to classify the minority class (Low participation). Feature analysis further revealed that access to basic and secondary education, as well as healthcare services, positively contributed to higher voter participation. These findings highlight the potential of combining Decision Tree algorithms with IDM data to support more data-driven strategies for increasing voter turnout.

Keywords - Voter Participation; Village Development Index (IDM); Decision Tree C4.5; Machine Learning; Mataraman.

Abstrak. Partisipasi masyarakat dalam Pemilihan Umum (Pemilu) merupakan indikator penting bagi kualitas demokrasi, namun tingkat partisipasi seringkali bervariasi antarwilayah. Kawasan Mataraman di Jawa Timur menunjukkan fenomena tersebut, di mana sebagian desa memiliki tingkat partisipasi tinggi, sementara desa lain justru rendah. Penelitian ini bertujuan untuk memprediksi tingkat partisipasi pemilu berbasis Indeks Desa Membangun (IDM) dengan pendekatan algoritma klasifikasi. Data yang digunakan merupakan publikasi IDM tahun 2024 dengan 72.311 entri yang mencakup 48 indikator awal. Setelah melalui tahap seleksi fitur, diperoleh 27 atribut relevan yang kemudian dianalisis menggunakan tiga algoritma: Decision Tree C4.5, Naïve Bayes, dan Support Vector Machine (SVM). Hasil penelitian menunjukkan bahwa algoritma C4.5 memberikan performa terbaik dengan akurasi 80,87% dan F1-score 0,88 pada kelas mayoritas (partisipasi Tinggi). Sementara itu, Naïve Bayes dan SVM cenderung bias terhadap kelas mayoritas dan gagal mengklasifikasikan kelas minoritas (partisipasi Rendah). Analisis fitur mengindikasikan bahwa akses pendidikan tingkat dasar hingga menengah serta layanan kesehatan berkontribusi positif terhadap tingginya partisipasi pemilu. Temuan ini menegaskan potensi pemanfaatan algoritma Decision Tree dan data IDM untuk mendukung strategi peningkatan partisipasi pemilu secara lebih berbasis data.

Kata Kunci - Partisipasi Pemilu; Indeks Desa Membangun; Decision Tree C4.5; Machine Learning; Mataraman.

I. PENDAHULUAN

Seiring dengan meningkatnya kompleksitas kehidupan berbangsa dan bernegara, kesadaran masyarakat terhadap pentingnya partisipasi dalam proses demokrasi juga turut berkembang [1]. Pemilihan Umum (Pemilu) hadir sebagai sarana bagi warga negara untuk menyalurkan aspirasi dan menentukan arah kebijakan negara [2]. Dalam beberapa tahun terakhir, akses terhadap informasi politik semakin terbuka lebar melalui berbagai kanal digital, dan literasi pemilu pun menunjukkan tren yang positif di sejumlah wilayah [3].

Meskipun demikian, tingkat keikutsertaan masyarakat dalam pemilu masih menunjukkan variasi yang cukup besar antarwilayah. Di kawasan Mataraman, yang meliputi Kabupaten Ngawi, Magetan, Ponorogo, Pacitan, Nganjuk, dan Madiun, terdapat desa-desa dengan partisipasi pemilu yang tinggi, sementara desa lainnya justru mengalami

penurunan. Hal ini mengindikasikan bahwa selain faktor politik, terdapat pula faktor-faktor non-politis yang berkontribusi terhadap antusiasme masyarakat dalam menggunakan hak pilihnya [4].

Salah satu pendekatan yang dapat digunakan untuk menelusuri fenomena ini adalah dengan mengkaji karakteristik objektif tiap desa melalui Indeks Desa Membangun (IDM). IDM merupakan indikator yang dikembangkan oleh Kementerian Desa untuk mengukur tingkat kemandirian desa berdasarkan aspek sosial, ekonomi, dan ekologi [5]. Selama ini, IDM telah banyak dimanfaatkan untuk pemetaan status desa dan pengambilan kebijakan pembangunan. Namun, pemanfaatannya dalam konteks perilaku partisipatif masyarakat, khususnya dalam pemilu, masih sangat terbatas.

Penelitian ini bertujuan untuk mengisi kekosongan studi terkait pemanfaatan algoritma machine learning, khususnya Decision Tree, dalam menganalisis dan memprediksi tingkat partisipasi pemilu berbasis data IDM. Hingga saat ini, belum terdapat penelitian yang secara spesifik menggabungkan pendekatan machine learning dengan data IDM untuk memahami perilaku politik masyarakat desa maupun kawasan semi-perkotaan seperti Mataraman, Jawa Timur.

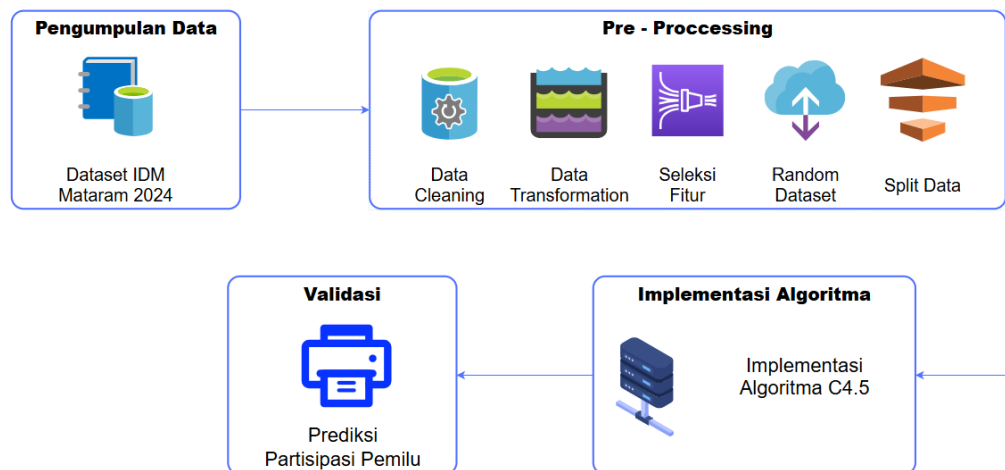
Studi terdahulu yang relevan, seperti “Peningkatan Model Klasifikasi pada Kab/Kota di Indonesia Menggunakan Metode Decision Tree”, hanya berfokus pada klasifikasi tingkat kemiskinan menggunakan data administratif tingkat Kabupaten, tanpa menyinggung aspek partisipasi politik atau pembangunan desa. Oleh karena itu, penelitian ini menawarkan kontribusi baru dengan mengeksplorasi integrasi metode Decision Tree dan data IDM dalam konteks politik elektoral di tingkat desa. Pendekatan ini diharapkan mampu menghasilkan model prediksi partisipasi pemilu yang akurat dan aplikatif, serta mendukung pengambilan keputusan berbasis data oleh pemerintah dan penyelenggara pemilu.

Dengan memanfaatkan data IDM yang bersifat kuantitatif dan terstruktur, serta algoritma Decision Tree C4.5 yang dikenal unggul dalam menyusun model klasifikasi dengan visualisasi yang mudah dipahami, penelitian ini diharapkan dapat mengungkap pola-pola tersembunyi antara karakteristik pembangunan desa dan tingkat partisipasi pemilih. Hasil penelitian ini diharapkan tidak hanya memperkaya wawasan akademik, tetapi juga memberikan kontribusi nyata dalam merancang strategi peningkatan partisipasi pemilu secara lebih terarah dan berbasis data.

II. METODE

Tahapan Penelitian

Dalam penelitian ini, pengembangan sistem informasi inventaris barang berbasis web dilakukan dengan menggunakan metode Agile. Pemilihan metode ini didasarkan pada kebutuhan akan proses pengembangan yang fleksibel, responsif terhadap perubahan, dan berorientasi pada kolaborasi dengan pengguna [5]. Agile adalah salah satu pendekatan dalam rekayasa perangkat lunak yang bersifat iteratif dan inkremental, di mana pengembangan dilakukan dalam siklus pendek (iteration) yang memungkinkan evaluasi dan perbaikan sistem secara terus-menerus [6].



Gambar 1. Tahapan Penelitian.

Rangkaian tahapan penelitian ditunjukkan pada Gambar 1 berikut, yang menjelaskan alur penelitian dari tahap awal hingga hasil :

1. Pengumpulan Data

Pengumpulan data dilakukan pada tahap awal penelitian dengan fokus pada wilayah Mataraman di Provinsi Jawa Timur, yang mencakup enam kabupaten: Ngawi, Magetan, Ponorogo, Pacitan, Nganjuk, dan Madiun. Wilayah ini dipilih karena memiliki karakteristik sosial dan budaya yang relatif serupa serta dianggap relevan untuk dianalisis dalam konteks prediksi partisipasi pemilu. Data yang digunakan merupakan data publikasi tahun 2024, yang terdiri dari berbagai indikator pembangunan desa pada Indeks Desa Membangun (IDM) serta data partisipasi pemilu.

Tahap seleksi atribut dilakukan untuk memastikan variabel yang digunakan benar-benar relevan. Pada tahap pertama, hanya indikator kuantitatif yang digunakan karena dapat dioperasionalkan sebagai variabel independen (X). Sementara itu, kolom administratif (Provinsi, Kabupaten, Kecamatan, Desa, Kode Wilayah, Tahun) hanya dipakai untuk keperluan identifikasi dan penggabungan data, sehingga tidak dijadikan variabel analisis. Kolom kategori atau sub-dimensi (misalnya Pendidikan atau Sub-dimensi Pendidikan) merupakan label teks sehingga tidak memiliki nilai kuantitatif, sedangkan variabel Total Skor dan Status ID dikeluarkan dari analisis untuk menghindari bias karena bersifat agregasi keseluruhan IDM [7]. Dari tahap seleksi ini, diperoleh 48 indikator kuantitatif yang dapat digunakan sebagai variabel independen (X1–X48). Adapun persentase partisipasi pemilu (%) dijadikan sebagai variabel dependen (Y).

Selanjutnya, pada tahap preprocessing dilakukan seleksi lanjutan untuk mereduksi jumlah variabel dan menghindari redundansi data. Dari 48 indikator awal, dipilih kembali sebanyak 27 atribut yang dianggap paling relevan dalam memengaruhi tingkat partisipasi pemilu. Proses ini dilakukan agar model klasifikasi yang dibangun tidak hanya lebih efisien, tetapi juga memiliki kemampuan generalisasi yang lebih baik. Rincian atribut akhir yang digunakan dalam penelitian ini dapat dilihat pada Tabel 1.

Tabel 1. Atribut Data

| No Atribut | Kelompok Atribut | Atribut |
|------------|----------------------|---|
| X1 – X4 | Pendidikan | Akses terhadap PAUD/ TK/ Sederajat (X1), Akses terhadap SD/ MI/ Sederajat (X2), Akses terhadap SMP/ MTs/ Sederajat (X3), Akses terhadap SMA/ SMK/ MA/ MAK/ Sederajat (X4) |
| X5 – X11 | Kesehatan | Layanan Sarana Kesehatan (X5), Fasilitas Kesehatan Poskesdes/ Polindes (X6), Aktivitas Posyandu (X7), Layanan Dokter (X8), Layanan Bidan (X9), Layanan Tenaga Kesehatan Lainnya (X10), Jaminan Kesehatan Nasional (X11) |
| X12 - X13 | Utilitas dasar | Air Minum (X12), Persentase Rumah Tidak Layak Huni (X13) |
| X14 – X18 | Aktivitas | Kearifan Sosial/ Budaya (X14), Frekuensi Gotong Royong (X15), Kegiatan Olahraga (X16), Mitigasi dan Penanganan Konflik Sosial (X17), Satkamling (X18) |
| X19 - X21 | Fasilitas masyarakat | Taman Bacaan Masyarakat/ Perpustakaan Desa (X19), Fasilitas Olahraga (X20), Keberadaan Ruang Publik Terbuka (X21) |
| X22 – X25 | Produksi desa | Keragaman Aktivitas Ekonomi (X22), Produk Unggulan Desa (X23), Ekonomi Kreatif (X24), Kerjasama Desa (X25) |
| X26 – X33 | Pendukung ekonomi | Akses Terhadap Pendidikan Non-formal/ Pusat Keterampilan/ Kursus (X26), Pasar Rakyat (X27), Toko/ Pertokoan (X28), Kedai/ Rumah Makan (X29), Penginapan (X30), Layanan |

| | | |
|-----------|--------------------------------|---|
| | | Pos dan/ Logistik (X31), Lembaga Ekonomi (X32), Layanan Keuangan (X33) |
| X34 – X37 | Pengelolaan lingkungan | Kearifan Lingkungan (X34), Sistem Pengelolaan Sampah (X35), Tingkat Pencemaran Lingkungan (X36), Sistem Pembuangan Air Limbah Domestik (Rumah Tangga) (X37) |
| X38 | Penanggulangan bencana | Penanggulangan Bencana (X38) |
| X39 – X40 | Kondisi akses jalan | Kondisi Jalan di desa (X39), Kondisi Penerangan Jalan Utama Desa (X40) |
| X41 – X43 | Kemudahan akses | Keberadaan Angkutan Perdesaan/ Angkutan Lokal/ Sejenis (X41), Akses Listrik (X42), Layanan Telekomunikasi (X43) |
| X44 – X46 | Kelembagaan dan pelayanan desa | Pelaksanaan Pelayanan dan Administrasi Desa (X44), Pemanfaatan Teknologi dalam Pelayanan Desa (SPBE) (X45), Musyawarah Desa (X46) |
| X47 - X48 | Tata kelola keuangan desa | Pendapatan Asli Desa (PADes) dan Dana Desa (X47), Jumlah Kepemilikan dan Produktivitas Aset Desa (X48) |
| Y | Pemilu | Tingkat Partisipasi Masyarakat |

2. Pre-Processing

Pre-processing data merupakan tahap awal dalam implementasi algoritma C4.5 pada data Indeks Desa Membangun, yang meliputi penggabungan data, modifikasi, pembersihan, dan penyederhanaan data untuk mendukung prediksi partisipasi pemilu di wilayah Mataraman.

a Data Cleaning

Pada tahap ini, data dianalisis untuk menemukan berbagai kesalahan seperti inkonsistensi, data duplikat, maupun data yang tidak lengkap. Setelah kesalahan teridentifikasi, dilakukan keputusan untuk menangani data yang tidak sesuai, baik dengan menghapus maupun memperbaikinya agar kualitas data tetap terjaga [8].

b Data Transformation

Transformasi data dilakukan untuk mengubah data ke dalam format yang lebih sesuai dengan kebutuhan analisis data mining. Proses ini mencakup konversi data ke bentuk numerik agar dapat dikenali oleh algoritma C4.5. Transformasi juga bertujuan untuk menyeragamkan skala data dan memudahkan proses pelatihan model.

c Seleksi Fitur

Data yang digunakan dalam penelitian ini terdiri dari 48 atribut prediktor yang dikelompokkan ke dalam 13 indikator berdasarkan data Indeks Desa Membangun (IDM). Untuk menyederhanakan model dan meningkatkan akurasi klasifikasi, dilakukan seleksi fitur dengan menghapus 6 kelompok atribut yang dinilai kurang relevan, yaitu : Utilitas dasar, Produksi desa ,Pendukung ekonomi, Pengelolaan lingkungan, Penanggulangan bencana, Tata kelola keuangan desa. Setelah proses seleksi, tersisa 7 kelompok atribut prediktor dengan total 27 atribut yang digunakan bersama 1 atribut target, yaitu partisipasi pemilu.

Tabel 2. Tabel Atribut Hasil Seleksi Fitur

| No Atribut | Kelompok Atribut | Atribut |
|------------|--------------------------------|---|
| X1 – X4 | Pendidikan | Akses terhadap PAUD/ TK/ Sederajat (X1), Akses terhadap SD/ MI/ Sederajat (X2), Akses terhadap SMP/ MTs/ Sederajat (X3), Akses terhadap SMA/ SMK/ MA/ MAK/ Sederajat (X4) |
| X5 – X11 | Kesehatan | Layanan Sarana Kesehatan (X5), Fasilitas Kesehatan Poskesdes/ Polindes (X6), Aktivitas Posyandu (X7), Layanan Dokter (X8), Layanan Bidan (X9), Layanan Tenaga Kesehatan Lainnya (X10), Jaminan Kesehatan Nasional (X11) |
| X12 – X16 | Aktivitas | Kearifan Sosial/ Budaya (X12), Frekuensi Gotong Royong (X13), Kegiatan Olahraga (X14), Mitigasi dan Penanganan Konflik Sosial (X15), Satkamling (X16) |
| X17 – X19 | Fasilitas masyarakat | Taman Bacaan Masyarakat/ Perpustakaan Desa (X17), Fasilitas Olahraga (X18), Keberadaan Ruang Publik Terbuka (X19) |
| X20 – X21 | Kondisi akses jalan | Kondisi Jalan di desa (X20), Kondisi Penerangan Jalan Utama Desa (X21) |
| X22 – X24 | Kemudahan akses | Keberadaan Angkutan Perdesaan/ Angkutan Lokal/ Sejenis (X22), Akses Listrik (X23), Layanan Telekomunikasi (X24) |
| X25 – X27 | Kelembagaan dan pelayanan desa | Pelaksanaan Pelayanan dan Administrasi Desa (X25), Pemanfaatan Teknologi dalam Pelayanan Desa (SPBE) (X26), Musyawarah Desa (X27) |
| Y | Pemilu | Tingkat Partisipasi Masyarakat |

d Random dataset

Setelah data dipersiapkan, dilakukan proses pengacakan terhadap dataset untuk memastikan keterwakilan yang adil dari setiap atribut dalam distribusi data. Hal ini penting guna menghindari bias dalam proses pelatihan dan pengujian model [9].

e Split data

Dataset yang telah diproses kemudian dibagi menjadi dua bagian:

- 70% untuk data training, yang digunakan dalam pelatihan model
- 30% untuk data testing, yang digunakan untuk evaluasi performa model

Pembagian ini bertujuan untuk mengukur seberapa baik model dapat menggeneralisasi pada data yang belum pernah dilihat sebelumnya.

3. Process

Tahap ini merupakan lanjutan dari pre-processing yang telah dilakukan sebelumnya. Dataset yang telah dibersihkan dan diseleksi kemudian diuji menggunakan algoritma klasifikasi Decision Tree C4.5. Dalam proses ini, data dibagi menjadi dua bagian menggunakan rasio 75% untuk pelatihan (training) dan 25% untuk pengujian (testing). Dari total data sebanyak 72.311 entri, diperoleh 54.233 baris untuk proses pelatihan dan 18.078 baris untuk pengujian. Algoritma C4.5 dipilih karena kemampuannya dalam menangani atribut numerik dan kategorikal.

4. Output

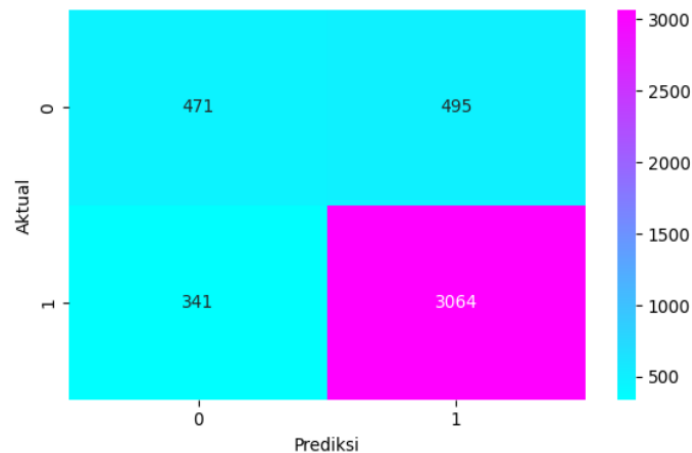
Setelah model dilatih, tahap berikutnya adalah menghasilkan prediksi tingkat partisipasi masyarakat dalam pemilu, yang dikategorikan ke dalam dua kelas yaitu “Tinggi” dan “Rendah”. Model C4.5 menghasilkan klasifikasi berdasarkan pola dari atribut-atribut yang telah ditentukan sebelumnya.

5. Analisis / Evaluasi

Evaluasi dilakukan untuk menilai seberapa baik performa dari model klasifikasi yang telah dibangun. Pengukuran dilakukan menggunakan beberapa metrik evaluasi seperti: Akurasi, Presisi, Recall, F1-Score. Evaluasi ini bertujuan untuk mengetahui seberapa efektif model dalam membedakan antara partisipasi tinggi dan rendah [10]. Metrik-metrik tersebut memberikan gambaran menyeluruh tentang kinerja algoritma C4.5 dalam konteks klasifikasi tingkat partisipasi masyarakat terhadap pemilu.

III. HASIL DAN PEMBAHASAN

Akurasi dan Presisi



Gambar 2. Confusion Matrix

Gambar 2 menyajikan Confusion Matrix dari hasil pengujian model terhadap data uji. Berdasarkan visualisasi tersebut, model mampu mengklasifikasikan dengan benar sebanyak 3.064 data pada kelas Tinggi, serta 471 data pada kelas Rendah. Sementara itu, terdapat sejumlah data yang belum dapat diklasifikasikan secara tepat, yaitu 341 data dari kelas Tinggi yang terprediksi sebagai Rendah dan 495 data dari kelas Rendah yang terklasifikasi sebagai Tinggi. Temuan ini menunjukkan bahwa model cenderung lebih akurat dalam mengenali pola dari kelas mayoritas (Tinggi), namun masih menghadapi kesulitan dalam membedakan karakteristik data dari kelas minoritas (Rendah). Analisis lanjutan terhadap distribusi dan fitur yang berkontribusi pada kelas Rendah dapat menjadi strategi penting untuk meningkatkan performa model secara keseluruhan [11].

Classification Report

Tabel 2. Classification Report C4.5

| Kelas | Precision | Recall | F1-Score | Support |
|-----------------|-----------|--------|-------------|---------|
| Rendah | 0.58 | 0.49 | 0.53 | 966 |
| Tinggi | 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 |

Tabel 3. Classification Report Naïve Bayes

| Kelas | Precision | Recall | F1-Score | Support |
|-----------------|-----------|--------|-------------|---------|
| Rendah | 0.00 | 0.00 | 0.00 | 966 |
| Tinggi | 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 |

Tabel 4. Classification Report SVM

| Kelas | Precision | Recall | F1-Score | Support |
|-----------------|-----------|--------|-------------|---------|
| Rendah | 0.00 | 0.00 | 0.00 | 966 |
| Tinggi | 0.78 | 1.00 | 0.88 | 3405 |
| Accuracy | | | 0.78 | 4371 |
| Macro Avg | 0.39 | 0.50 | 0.44 | 4371 |

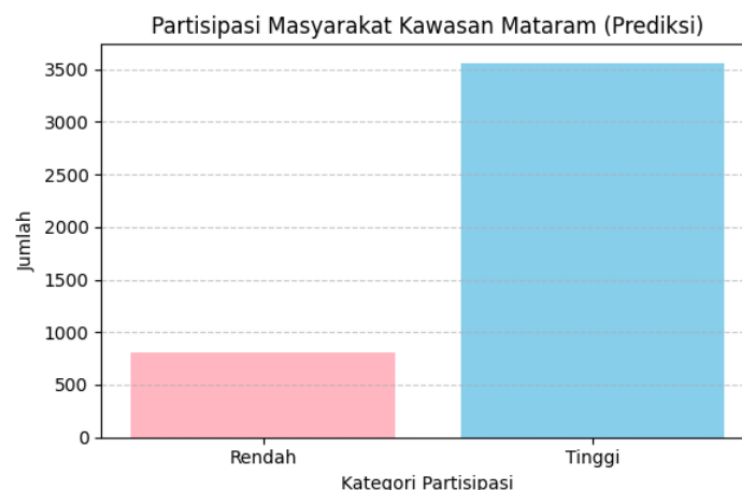
| Kelas | Precision | Recall | F1-Score | Support |
|--------------|-----------|--------|----------|---------|
| Weighted Avg | 0.61 | 0.78 | 0.68 | 4371 |

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).

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.

Presentase Tingkat Partisipasi



Gambar 3. Confusion Matrix C4.5

Gambar di atas menyajikan distribusi hasil prediksi tingkat partisipasi masyarakat pada kawasan Mataram. Grafik memperlihatkan bahwa jumlah data pada kategori partisipasi Tinggi jauh lebih besar dibandingkan kategori partisipasi Rendah. Secara visual, terlihat bahwa prediksi partisipasi Tinggi mencapai lebih dari 3.500 data, sedangkan partisipasi Rendah hanya sekitar 800 data. Perbedaan distribusi ini menunjukkan bahwa model cenderung mengklasifikasikan sebagian besar data ke dalam kategori Tinggi [12]. Hal ini sejalan dengan temuan sebelumnya pada Confusion Matrix, di mana model memiliki kemampuan yang lebih baik dalam mengenali kelas mayoritas (Tinggi) dibandingkan kelas minoritas (Rendah). Kondisi ini juga mengindikasikan adanya ketidakseimbangan data (imbalanced dataset) yang memengaruhi hasil prediksi [13]. Dominasi jumlah prediksi pada kelas Tinggi memperlihatkan bias model terhadap kelas mayoritas, sehingga diperlukan strategi lanjutan seperti resampling data atau penerapan metode klasifikasi yang lebih sensitif terhadap kelas minoritas [14].

Nilai Tiap Fitur Berdasarkan Kategori Partisipasi

Tabel 5. Nilai Tiap Fitur Berdasarkan Kategori Partisipasi

| 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.

VII. SIMPULAN

Berdasarkan hasil pengujian tiga algoritma klasifikasi, yaitu Decision Tree C4.5, Naïve Bayes, dan Support Vector Machine (SVM), diperoleh bahwa algoritma C4.5 Decision Tree memberikan performa terbaik dalam memprediksi tingkat partisipasi masyarakat pada Pemilu di wilayah Mataraman. Model C4.5 mencapai akurasi 80,87%, dengan kemampuan yang cukup baik dalam mengenali kelas mayoritas (partisipasi Tinggi) meskipun masih menghadapi keterbatasan pada kelas minoritas (partisipasi Rendah). Sementara itu, algoritma Naïve Bayes dan SVM menunjukkan kecenderungan bias yang kuat terhadap kelas mayoritas, bahkan gagal mengklasifikasikan kelas minoritas.

Temuan ini juga menegaskan adanya permasalahan imbalanced dataset, di mana jumlah data dengan kategori partisipasi Tinggi jauh lebih dominan dibandingkan kategori Rendah. Hal tersebut memengaruhi pola klasifikasi model sehingga prediksi lebih banyak diarahkan pada kelas mayoritas. Selain itu, hasil analisis fitur menunjukkan bahwa akses pendidikan (SD, SMP, SMA/SMK) dan layanan kesehatan (tenaga kesehatan, bidan, posyandu) memiliki kontribusi positif terhadap meningkatnya partisipasi masyarakat. Desa dengan akses pendidikan dan layanan kesehatan yang lebih baik cenderung memiliki tingkat partisipasi pemilu yang lebih tinggi. Namun, terdapat pengecualian pada variabel akses PAUD/TK yang justru tidak berbanding lurus dengan tingkat partisipasi.

Secara keseluruhan, Bab IV memperlihatkan bahwa model C4.5 Decision Tree dapat menjadi pendekatan yang efektif untuk memahami hubungan antara indikator pembangunan desa (IDM) dengan partisipasi pemilu. Namun, diperlukan strategi lanjutan, seperti resampling data atau penggunaan metode klasifikasi yang lebih adaptif terhadap data tidak seimbang, agar prediksi terhadap kelas minoritas dapat lebih optimal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada berbagai pihak yang telah memberikan dukungan dalam penyusunan penelitian ini. Ucapan terima kasih khusus disampaikan kepada Pemerintah Kabupaten Sidoarjo yang telah menyediakan data serta memberikan dukungan terkait partisipasi pemilu. Terima kasih juga ditujukan kepada dosen pembimbing atas arahan dan masukan yang sangat membantu dalam proses penelitian ini. Selain itu, penulis menghargai dukungan dari seluruh dosen Program Studi Informatika Universitas Muhammadiyah Sidoarjo, serta

keluarga dan teman-teman yang selalu memberi semangat. Semoga penelitian ini bermanfaat bagi pengembangan ilmu pengetahuan dan dapat memberikan kontribusi dalam meningkatkan partisipasi politik masyarakat.

REFERENSI

- [1] 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," *Indonesian Journal of Business Intelligence (IJUBI)*, vol. 5, no. 1, p. 36, 2022, doi: 10.21927/ijubi.v5i1.2259.
- [2] A. W. Anggraeni, A. S. Fitriani, and A. Eviyanti, "Penerapan Algoritma Support Vector Machine untuk Memprediksi Tingkat Partisipasi Pemilu terhadap Kualitas Pendidikan," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 1, pp. 21–27, 2024, doi: 10.29408/edumatic.v8i1.24838.
- [3] A. S. Kirana, R. Roeswidiah, and A. Pudoli, "ANALISIS SENTIMEN PADA MEDIA SOSIAL TERHADAP LAYANAN SAMSAT DIGITAL NASIONAL," vol. 8, pp. 53–63, 2025.
- [4] M. J. Siddiq *et al.*, "MASYARAKAT TERHADAP PILKADA 2024 DI MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE," vol. 13, no. 2, pp. 609–622, 2025.
- [5] Y. Izamahendra and M. Subhan, "DESA PARIAMAN TIMUR BERDASARKAN INDEKS DESA MEMBANGUN 2023," vol. 13, no. 2, pp. 204–211, 2025.
- [6] A. Akram *et al.*, "MINI-BATCH K-MEANS CLUSTERING UNTUK PENGELOMPOKAN," pp. 235–244, 2024.
- [7] E. F. Laili *et al.*, "KOMPARASI ALGORITMA DECISION TREE DAN SUPPORT VECTOR MACHINE (SVM) DALAM," vol. 8, no. 1, pp. 67–76, 2025.
- [8] F. Teknik *et al.*, "Metode Decision Trees untuk Evaluasi dan Supervisi Pembentukan Karakter Siswa di Sekolah Kejuruan," vol. 9, no. 1, pp. 8–12, 2025.
- [9] S. Danil, N. Rahaningsih, R. D. Dana, and K. Cirebon, "MODEL PADA KAB / KOTA DI INDONESIA MENGGUNAKAN METODE," vol. 13, no. 2, pp. 829–835, 2025.
- [10] J. Riset and S. Informasi, "ANALISIS POLA KEHADIRAN MAHASISWA MENGGUNAKAN," vol. 2, no. 1, pp. 60–66, 2025.
- [11] D. M. Chulloh, A. S. Fitriani, 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.
- [12] 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>
- [13] W. Kurniawati, A. S. Fitriani, 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.
- [14] D. M. Chulloh, A. S. Fitriani, 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.

- [15] W. Kurniawati, A. S. Fitriani, 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.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.