



## **Analisis machine learning terhadap pengaruh keterlibatan keluarga dalam meningkatkan partisipasi pemilu**

### ***Machine learning analysis on the influence of family involvement in increasing election participation***

Arif Senja Fitriani\*, Andry Rachmadany

\*Universitas Muhammadiyah Sidoarjo, Jln. Raya Gelam No. 250, Candi, Sidoarjo, Indonesia, 61215

#### **INFORMASI ARTIKEL    ABSTRAK**

##### **Article History:**

*Submission: 28-05-2025*

*Revised: 20-06-2025*

*Accepted: 26-06-2025*

##### **Kata Kunci:**

Pemilu;  
partisipasi keluarga;  
model machine learning

##### **Keywords:**

*Election;*  
*family participation;*  
*machine learning models*

##### **\* Korespondensi:**

**Arif Senja Fitriani**  
asfjim@umsida.ac.id

Partisipasi masyarakat dalam pemilihan umum (pemilu) merupakan salah satu indikator penting dalam menilai kualitas demokrasi suatu negara. Tingkat partisipasi yang tinggi mencerminkan kesadaran politik dan keterlibatan aktif warga negara dalam menentukan arah kepemimpinan nasional dan daerah. Salah satu faktor yang diyakini berkontribusi terhadap peningkatan partisipasi pemilih adalah keterlibatan keluarga, terutama dalam membentuk sikap dan preferensi politik anggota rumah tangga. Penelitian ini bertujuan untuk menganalisis pengaruh keterlibatan keluarga terhadap tingkat partisipasi pemilih melalui metode *machine learning*, serta menguji efektivitas model klasifikasi dalam memprediksi partisipasi masyarakat berdasarkan data pemilu aktual. Penelitian ini menggunakan dataset yang diperoleh dari daftar hadir dan rekapitulasi hasil pemilu di Desa Wonokasian, Kecamatan Wonoayu, Kabupaten Sidoarjo. Dataset tersebut terdiri dari 5.500 instance dan 14 atribut, dengan kelas target berupa tingkat partisipasi "tinggi" dan "rendah". Metode klasifikasi yang digunakan adalah algoritma *Decision Tree C4.5* yang memiliki keunggulan dalam interpretabilitas dan efisiensi pengambilan keputusan berbasis data. Model klasifikasi yang dibangun mampu memprediksi tingkat partisipasi pemilih dengan tingkat akurasi sebesar 90,12%. Hasil tersebut menunjukkan bahwa atribut "hadir kel" dan "jml KK" merupakan bentuk pengaruh keterlibatan keluarga yang memiliki korelasi kuat terhadap keputusan untuk berpartisipasi dalam pemilu. Pendekatan *machine learning* berbasis klasifikasi *Decision Tree C4.5* terbukti efektif dalam memetakan pola partisipasi pemilih. Tingginya hasil prediksi dan keterlibatan dua atribut keluarga tersebut dapat digunakan sebagai dasar dalam merumuskan kebijakan untuk meningkatkan partisipasi politik, khususnya melalui penguatan peran keluarga dalam pendidikan demokrasi di tingkat daerah.

#### **ABSTRACT**

*Public participation in general elections is a key indicator for assessing the quality of democracy in a country. A high level of voter turnout reflects political awareness and the active involvement of citizens in shaping national and regional leadership. One factor believed to contribute significantly to increased voter participation is family involvement, particularly in shaping the political attitudes and preferences of household members. This study aims to analyze the influence of family involvement on voter participation levels using a machine learning approach, and to evaluate the effectiveness of a classification model in predicting community participation based on actual election data. The dataset used in this study was derived from voter attendance records and election result recapitulations in Wonokasian Village,*



*Wonoayu District, Sidoarjo Regency. It comprises 5,500 instances and 14 attributes, with the target class categorized as "high" and "low" voter participation. The classification method employed is the Decision Tree C4.5 algorithm, known for its interpretability and efficiency in data-driven decision-making. The resulting classification model achieved an accuracy rate of 90.12% in predicting voter participation levels. The findings indicate that the attributes "hadir kel" (family attendance) and "jml KK" (number of family households) serve as strong indicators of family involvement and show significant correlation with the decision to vote. The machine learning approach based on the Decision Tree C4.5 classification proves to be effective in mapping voter participation patterns. The high predictive performance and the relevance of family-related attributes provide a foundation for developing policy strategies to enhance political participation—particularly by strengthening the role of families in local democratic education.*

## 1. PENDAHULUAN

Pemilihan umum (pemilu) merupakan mekanisme utama dalam sistem demokrasi yang memberikan ruang bagi warga negara untuk secara langsung menentukan arah kebijakan politik dan pemerintahan. Melalui pemilu, masyarakat dapat menyalurkan aspirasi politiknya secara sah, menjadikan pemilu sebagai pilar fundamental dalam menjaga stabilitas dan legitimasi demokrasi [1]. Salah satu indikator utama dari keberhasilan pelaksanaan pemilu adalah tingkat partisipasi pemilih [2]. Tingginya partisipasi menunjukkan keterlibatan publik yang aktif dalam proses politik, sekaligus menjadi tolak ukur kepercayaan masyarakat terhadap sistem pemerintahan yang ada [3]. Oleh karena itu, memahami determinan yang mempengaruhi partisipasi pemilih menjadi penting untuk memastikan berjalannya demokrasi yang sehat dan inklusif [4].

Dalam struktur sosial masyarakat, keluarga memegang peranan strategis sebagai institusi sosial paling mendasar yang membentuk sikap dan perilaku politik individu. Sebagai lingkungan sosial pertama yang dijalani seseorang sejak lahir, keluarga menjadi tempat utama dalam proses sosialisasi nilai-nilai sosial dan politik [5]. Komunikasi yang terjalin antaranggota keluarga—baik secara formal maupun informal—berkontribusi besar dalam membentuk pandangan politik, kesadaran akan hak dan kewajiban sebagai warga negara, serta keputusan untuk berpartisipasi dalam pemilu [6]. Lebih jauh lagi, keluarga tidak hanya menjadi ruang pembelajaran sosial, tetapi juga memiliki pengaruh signifikan dalam keputusan kolektif yang berkaitan dengan partisipasi dalam ranah publik, termasuk di antaranya pemilihan umum [7].

Perkembangan teknologi informasi dan digitalisasi data membuka peluang baru dalam menganalisis partisipasi pemilu secara lebih komprehensif dan berbasis bukti. Berbagai sumber data, seperti data demografi, data administrasi kependudukan, serta data dari media sosial, kini menjadi sumber yang sangat potensial untuk memahami perilaku pemilih [8]. Informasi mengenai latar belakang individu seperti usia, jenis kelamin, tingkat pendidikan, serta keterhubungan sosial dalam keluarga dapat diintegrasikan untuk membangun model prediksi partisipasi yang lebih presisi [9]. Di sisi lain, data media sosial dapat menjadi indikator alternatif terhadap tingkat keterlibatan politik masyarakat, yang apabila dikombinasikan dengan data demografi mampu memperkaya pemetaan partisipasi politik [10].

Dalam menjawab kompleksitas dinamika perilaku pemilih, pendekatan berbasis pembelajaran mesin (*machine learning*) telah banyak digunakan dalam kajian ilmu sosial dan politik kontemporer. Metode ini memungkinkan pengolahan data berskala besar serta deteksi pola yang tidak selalu dapat diidentifikasi melalui analisis statistik konvensional [11]. Penerapan algoritma pembelajaran mesin dalam pemodelan telah menunjukkan keunggulan dalam hal akurasi prediksi, kemampuan generalisasi, serta kemampuannya dalam menangani data berbagai tipe dan membantu menemukan kombinasi terbaik antara deskriptor dan algoritma pembelajaran mesin [12]. Selain itu, model prediktif berbasis *machine learning* mampu mengidentifikasi kelompok yang berpotensi tidak hadir di TPS, sehingga dapat dimanfaatkan dalam penyusunan strategi peningkatan partisipasi secara lebih tepat sasaran [13].

Penelitian ini mengadopsi pendekatan klasifikasi berbasis algoritma pohon keputusan C4.5 yang dikenal memiliki interpretabilitas tinggi serta kemampuannya dalam menangani atribut kategorikal maupun numerik [14]. Implementasi penelitian dilakukan melalui dua model utama, yaitu model yang menggunakan semua atribut serta model yang telah diseleksi berdasarkan korelasi terhadap partisipasi. Proses penelitian dilakukan secara sistematis melalui tahapan pengumpulan data, prapemrosesan (meliputi pembersihan data, pengkodean fitur, korelasi atribut, randomisasi dataset, dan pembagian data latih dan uji), serta evaluasi model menggunakan metrik akurasi dan *confusion matrix*. Dataset yang digunakan dalam penelitian ini bersumber dari data riil di Desa Wonokasian, Kabupaten Sidoarjo, yang memuat informasi mengenai kehadiran dalam pemilu serta struktur keluarga, menjadikan penelitian ini unik karena memfokuskan analisis pada keterlibatan keluarga sebagai variabel utama yang mempengaruhi partisipasi pemilih.

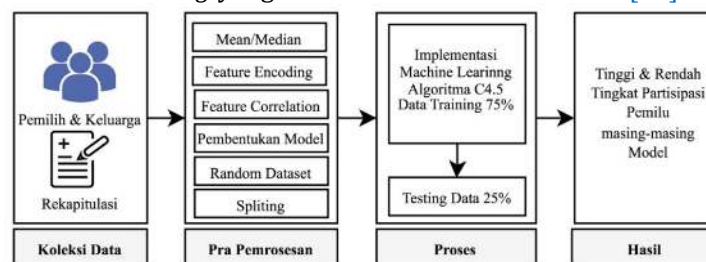
## 2. METODE

Penelitian ini mengimplementasikan pendekatan *machine learning* dengan fokus pada metode klasifikasi menggunakan algoritma *Decision Tree C4.5*, yang merupakan pengembangan dari algoritma ID3 [15]. Algoritma *Decision Tree C4.5* memiliki keunggulan dalam menangani data numerik dan kategorikal, serta menghasilkan pohon keputusan yang dapat diinterpretasikan secara visual [16]. Algoritma ini bekerja dengan memilih atribut terbaik berdasarkan nilai gain dan gain ratio, kemudian membentuk cabang-cabang pohon yang merepresentasikan keputusan klasifikasi. Dataset yang digunakan terdiri dari data pemilih, daftar hadir saat pemungutan suara, serta data rekapitulasi hasil pemilu yang diperoleh dari Desa Wonokasian. Total terdapat 4.249 instance dan 10 atribut yang mencakup informasi kependudukan, partisipasi anggota keluarga, dan data kehadiran, dengan label target berupa kategori tingkat partisipasi “Tinggi” dan “Rendah”.

Proses penelitian dimulai dari tahap pengumpulan dan integrasi data (*data collecting*), dilanjutkan dengan data preparation berupa pembersihan data dan transformasi atribut menjadi format yang dapat diproses oleh algoritma [17]. Setelah itu dilakukan pelatihan model menggunakan algoritma *Decision Tree C4.5*, dengan mempertimbangkan nilai informasi terbaik dari setiap atribut dalam membentuk pohon keputusan [16]. Model yang dihasilkan kemudian diuji dengan data uji untuk memperoleh metrik performa seperti akurasi, presisi, dan *recall*. Evaluasi dilakukan untuk menilai sejauh mana keterlibatan keluarga dapat menjadi prediktor dalam menentukan tingkat partisipasi pemilih.

### Kerangka Model

Kerangka model dalam penelitian ini terdiri dari beberapa tahap utama, yaitu pengumpulan data, praproses data, pembangunan model klasifikasi dengan algoritma C4.5, serta evaluasi performa model. Setiap tahapan dirancang untuk memastikan bahwa data yang digunakan memiliki kualitas yang baik dan hasil model dapat diinterpretasikan secara valid dalam konteks peningkatan partisipasi pemilu melibatkan empat tahapan utama, yaitu [Gambar 1](#): (1) pengumpulan data (*Data Collecting*), (2) pra pemrosesan (*Preprocessing*), (3) proses (*Processing*), dan (4) interpretasi hasil prediksi. Model dibangun untuk mengevaluasi pengaruh variabel-variabel keluarga terhadap tingkat partisipasi dalam pemilu [18]. Tujuan dari kerangka ini adalah menyediakan pendekatan sistematis dari hulu ke hilir dalam membangun sistem prediktif berbasis *machine learning* yang andal dan berbasis data lokal [19].



**Gambar 1.** Kerangka model dalam penerapan metode machine learning dengan algoritma Decision Tree C4.5 [11].

Koleksi data

Data dikumpulkan dari Desa Wonokasian, Kecamatan Wonoayu, Kabupaten Sidoarjo. Data utama yang digunakan mencakup daftar hadir pemilih saat hari pemungutan suara, data pemilih yang terdaftar, dan data rekapitulasi hasil pemilu. Ketiga sumber data tersebut digabungkan untuk membentuk dataset yang utuh dan representatif dengan total 4.249 *instance*. Proses pengumpulan data dilakukan melalui tiga sumber utama, yaitu: (1) data daftar hadir pada hari pemungutan suara yang mencerminkan partisipasi aktual pemilih, (2) data identitas pemilih dari DPT (Daftar Pemilih Tetap) Desa Wonokasian, Kecamatan Wonoayu, dan (3) data rekapitulasi hasil pemilu di tingkat desa yang diperoleh dari KPU Kabupaten Sidoarjo. Setiap *instance* dalam dataset mewakili satu individu dengan tambahan atribut mengenai jml KK, Asal Penduduk, Kategori Masy, kawin, jenis\_kelamin, alamat, rt, tps, lokasi, hadir kel, dan hadir (kelas prediksi) [Tabel 1](#). Data ini direkapitulasi secara longitudinal dari pemilu 2019.

**Tabel 1.** Dataset pemilih dan rekapitulasi partisipasi pemilu

No	jml KK	Asal	Kategori Masy	kawin	jenis_kelamin	alamat	rt	tps	lokasi	hadir kel	hadir
0	dua	lokal	Milenial	S	P	Wonokasian	rt01	TPS-1	dalam	TSH	TH
1	dua	lokal	Milenial	S	L	Wonokasian	rt01	TPS-1	dalam	TSH	H
...	...	...	...	...	...	...	...	...	...	...	...
4247	dua	lokal	Milenial	B	P	Klitih	rt21	TPS-16	dalam	HS	H
4248	dua	lokal	Milenial	B	L	Klitih	rt21	TPS-16	dalam	HS	H

Pra pemrosesan

Tahap ini dimulai dengan proses data *cleaning* untuk menghapus duplikasi dan menangani nilai yang hilang atau tidak relevan. Selanjutnya dilakukan proses *Encoding*, yang mana tipe atribut pada semua tipe atribut adalah *object* [Tabel 2](#), selanjutnya dilakukan proses perubahan menjadi tipe *numeric*. Detail pada perubahan masing-masing atribut ke *Label Encoding* pada atribut “Asal Penduduk”, “Kategori Masy”, “kawin”, “jenis\_kelamin”, “alamat”, “rt”, “tps” dan “lokasi”. Selanjutnya untuk *Ordinal Encoding* yaitu “jml KK” dan “hadir kel”. Langkah selanjutnya *Feature Correlation* yaitu, untuk mengukur hubungan atribut dengan kelas yang ditetapkan pada masing-masing atribut. Hasil [Tabel 3](#), menunjukkan kekuatan dan arah hubungan *linier* antara masing-masing atribut dengan label hadir. Dari evaluasi korelasi atribut maka dibentuk model baru yang tidak mengikutsertakan 3 (tiga) atribut terakhir yang mempunyai hasil korelasi negatif yaitu kawin, alamat, dan tps. Kemudian *random* dataset sebelum dilakukan tahap *split* dataset untuk *training* dan *testing*. Atribut-atribut dipilih berdasarkan relevansinya terhadap keterlibatan keluarga dan perilaku memilih. Data kemudian dibagi menjadi data latih dan data uji dengan rasio 75:25 agar sesuai dengan kebutuhan algoritma C4.5. Label target diklasifikasikan berdasarkan kehadiran pemilih yaitu “Hadir (H)” dan “Tidak Hadir (TH)”.

**Tabel 2.** Deskripsi dataset partisipasi

No	Column	Non-Null	Count	Dtype	Label
0	jml KK	4249	non-null	object	2 sampai dengan 7 orang
1	Asal Penduduk	4249	non-null	object	Lokal, Urban
2	Kategori Masy	4249	non-null	object	Pemula, Millenial, Pasca Millenial
3	kawin	4249	non-null	object	Belum, Sudah, Pernah Menikah
4	jenis_kelamin	4249	non-null	object	Laki-laki, Perempuan
5	alamat	4249	non-null	object	Desa/Dusun di Desa Wonokasian
6	rt	4249	non-null	object	No RT
7	tps	4249	non-null	object	id TPS
8	lokasi	4249	non-null	object	Tempat TPS di dalam/luar RT

No	Column	Non-Null	Count	Dtype	Label
9	hadir kel	4249	non-null	object	Jumlah Keluarga yang hadir di TPS
10	hadir	4249	non-null	object	Hadir, Tidak Hadir

Tabel 3. Korelasi atribut pada kelas prediksi hadir

Atribut	Korelasi	Rekomendasi
hadir kel	0,35	Diikuti. Masih memberikan kontribusi.
rt	0,19	Diikuti. Masih memberikan kontribusi.
lokasi	0,09	Diikuti. Masih memberikan kontribusi.
jml KK	0,03	Boleh diuji, tapi tidak signifikan
Asal Penduduk	0,02	Boleh diuji, tapi tidak signifikan
jenis_kelamin	-0,01	Hampir tidak ada pengaruh
Kategori Masy	-0,03	Hampir tidak ada pengaruh
kawin	-0,08	Hampir tidak ada pengaruh
alamat	-0,11	Hampir tidak ada pengaruh
tps	-0,15	Hampir tidak ada pengaruh

Hasil analisis korelasi antar atribut [Tabel 3](#) terhadap variabel target menunjukkan bahwa beberapa variabel masih memiliki kontribusi yang cukup dalam proses klasifikasi. Atribut hadir kel memiliki korelasi tertinggi sebesar 0,35, diikuti oleh rt 0,19, dan lokasi 0,09, yang ketiganya direkomendasikan untuk tetap disertakan dalam pemodelan karena masih memberikan kontribusi yang berarti. Sementara itu, atribut seperti jumlah KK 0,03 dan asal penduduk 0,02 memiliki korelasi yang sangat lemah namun masih dapat diuji lebih lanjut untuk melihat potensi pengaruhnya. Di sisi lain, atribut jenis kelamin, kategori masyarakat, status kawin, alamat, dan TPS menunjukkan korelasi negatif yang sangat rendah (berkisar antara -0,01 hingga -0,15), sehingga dianggap hampir tidak memberikan pengaruh terhadap hasil klasifikasi dan dapat dipertimbangkan untuk dieliminasi guna menyederhanakan model dan mengurangi kompleksitas fitur. Selanjutnya pada hasil korelasi atribut ditetapkan dua model dimana “model1” mengikutsertakan semua atribut dan “model2” tidak menghapus atribut ‘alamat’ dan ‘tps’.

#### Proses

Algoritma *Decision Tree C4.5* membentuk pohon keputusan dengan memilih atribut berdasarkan nilai Gain Ratio tertinggi. Sebelumnya, nilai Entropy dan Information Gain harus dihitung terlebih dahulu. Berikut adalah rumus-rumus penting:

- Algoritma *Decision Tree C4.5* menggunakan perhitungan entropi (1) untuk menentukan tingkat ketidakpastian suatu atribut:

$$Entropy(S) = -\sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

- Dimana  $p_i$  adalah proporsi kelas  $i$  dalam himpunan data  $S$ . Kemudian, nilai Information Gain (2) dihitung untuk mengukur seberapa besar pengurangan entropi setelah data dibagi oleh atribut tertentu:

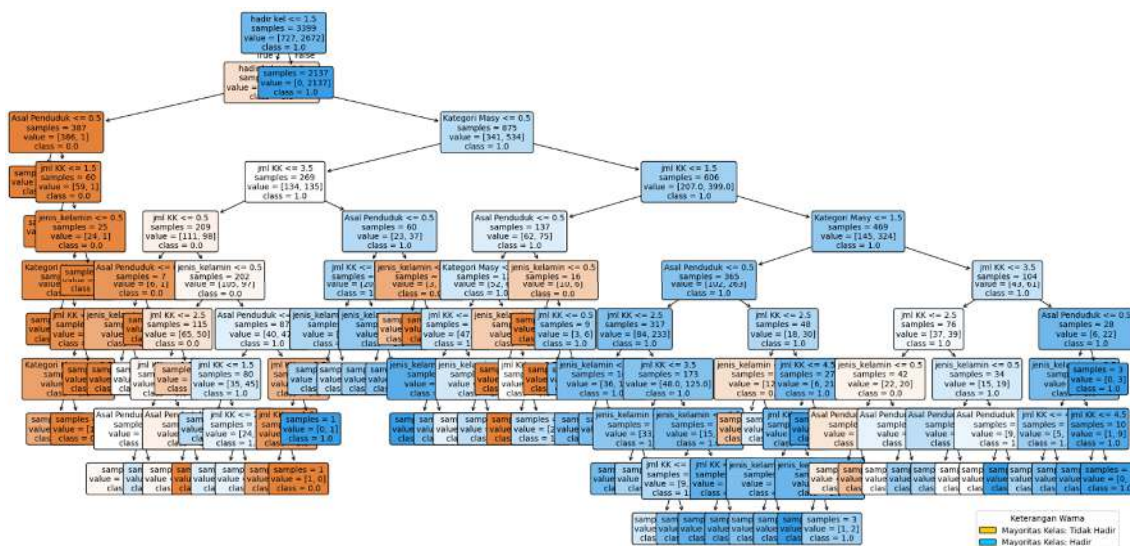
$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v) \quad (2)$$

- Namun, karena informasi gain cenderung bias terhadap atribut dengan banyak nilai, C4.5 menggunakan Gain Ratio (3) sebagai koreksi:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(A)} \quad (3)$$

Dengan:

$$SplitInfo(A) = -\sum_{v \in Values(A)} \frac{|S_v|}{|S|} \log_2 \left( \frac{|S_v|}{|S|} \right) \quad (4)$$



Gambar 2. Model pohon keputusan yang dihasilkan dari proses pelatihan

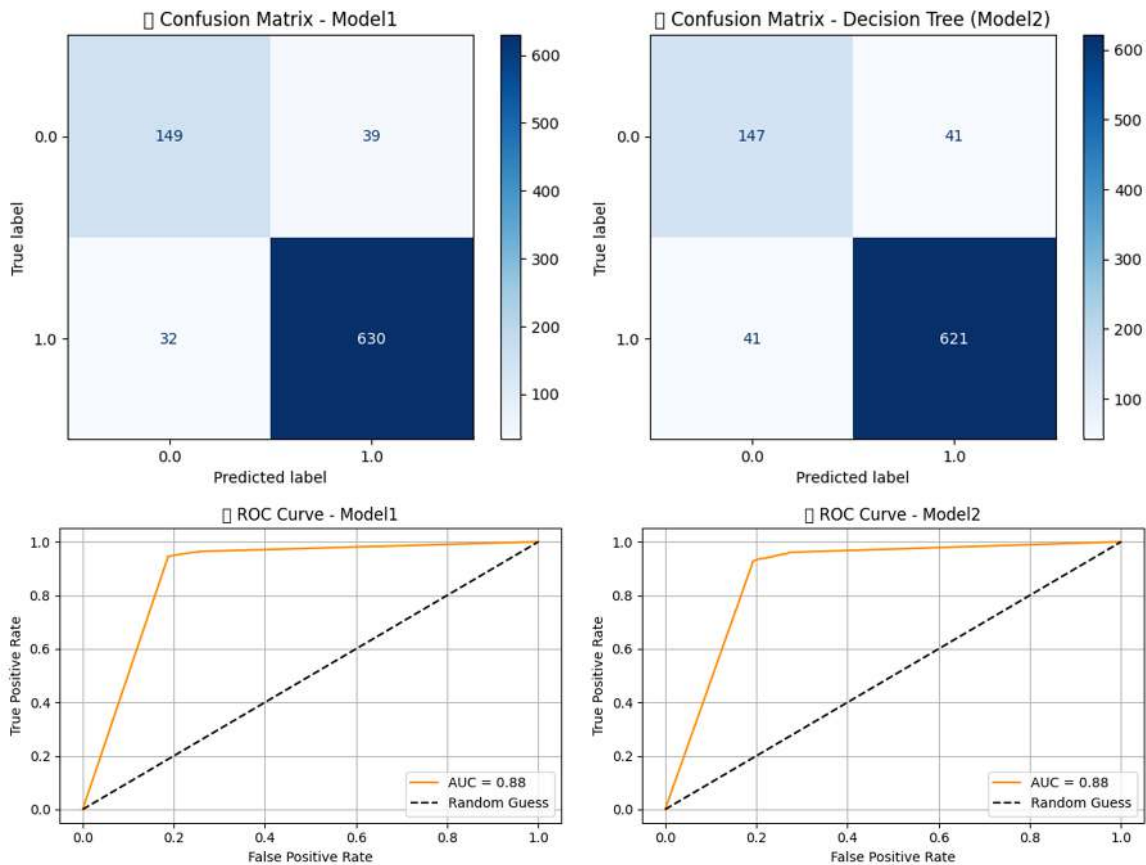
Split Info mengukur potensi informasi dari pembagian atribut, yaitu seberapa banyak data tersebar dalam partisi yang dibentuk oleh atribut. Proses penelitian dimulai dari tahap preprocessing data, termasuk pembersihan data dan transformasi atribut. Setelah itu dilakukan pelatihan model menggunakan algoritma C4.5 dengan kriteria pembentukan pohon berdasarkan nilai informasi terbaik (Gambar 2). Model kemudian diuji menggunakan data uji untuk memperoleh tingkat akurasi, presisi, dan recall. Evaluasi dilakukan untuk menilai performa model dalam mengklasifikasikan tingkat partisipasi berdasarkan keterlibatan keluarga.

### 3. HASIL DAN PEMBAHASAN

Perbandingan performa antara Model1 dan Model2 dilakukan dengan menggunakan metode Machine Learning dengan algoritma klasifikasi *Decision Tree C4.5*, yang diimplementasikan melalui pustaka scikit-learn dengan kriteria pemilihan node berbasis *entropy*. Sebelumnya, data diolah melalui tahap preprocessing yang mencakup pembersihan data, penerapan Label Encoding untuk fitur kategorikal, serta penghapusan atribut dengan korelasi rendah (pada Model2). Evaluasi model dilakukan menggunakan metrik *Accuracy*, *Recall* (Sensitivitas), *Precision*, *F1-Score*, dan *AUC (Area Under the Curve)* berdasarkan hasil *confusion matrix* dan kurva *ROC*.

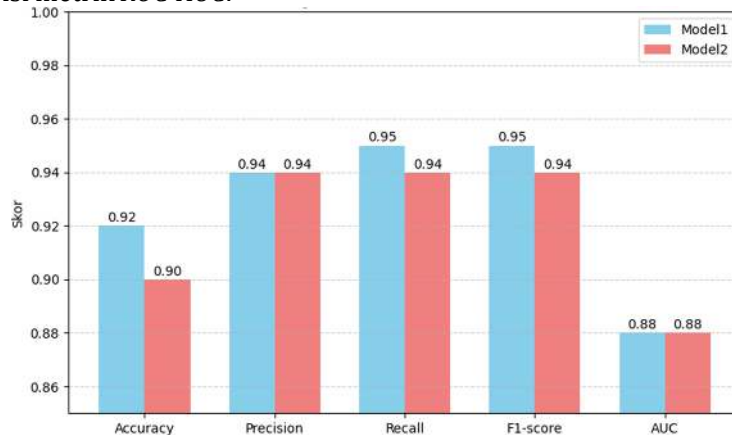
Model1 menunjukkan performa yang sangat baik, dengan nilai *F1-Score* dan *Recall* yang tinggi, menandakan kemampuan model dalam mengenali pemilih yang hadir secara akurat—hal yang krusial dalam konteks prediksi partisipasi pemilih. Tingkat kesalahan klasifikasi (*False Positive* dan *False Negative*) relatif kecil, dan nilai *AUC* sebesar 0,88 menunjukkan bahwa model mampu memisahkan kelas dengan cukup baik. Sementara itu, Model2, yang tidak menyertakan atribut 'lokasi' dan 'rt', menunjukkan penurunan kecil pada nilai *Recall* dan *Accuracy*, serta peningkatan jumlah FN dan FP. Hal ini mengindikasikan bahwa meskipun korelasi atribut 'lokasi' dan 'rt' terhadap target tergolong rendah, keduanya tetap memberikan kontribusi informasi yang relevan dalam proses klasifikasi.

Berdasarkan hasil *confusion matrix* (Gambar 4), "Model1" menunjukkan performa klasifikasi yang lebih baik dibandingkan "Model2". Pada "Model1", terdapat 149 data negatif dan 630 data positif yang berhasil diklasifikasikan dengan benar, sementara kesalahan prediksi terdiri dari 39 data negatif yang salah diklasifikasikan sebagai positif dan 32 data positif yang salah diklasifikasikan sebagai negatif. Di sisi lain, "Model2" menghasilkan 147 prediksi benar untuk data negatif dan 621 untuk data positif, dengan jumlah kesalahan prediksi yang lebih tinggi dan seimbang, yaitu 41 pada masing-masing kelas. Perbedaan ini menunjukkan bahwa Model1 lebih akurat dan sensitif dalam mengenali data positif, serta menghasilkan kesalahan yang lebih sedikit dibandingkan Model2, sehingga lebih direkomendasikan untuk digunakan dalam konteks klasifikasi ini.



Gambar 3. Model1 dan Model2 untuk Confusion Matrix dan ROC Curve

Berdasarkan grafik ROC (*Receiver Operating Characteristic*) untuk “Model1” dan “Model2” Gambar 4, kedua model menunjukkan performa klasifikasi yang sangat baik dengan nilai AUC (*Area Under Curve*) sebesar 0.88. Kurva ROC dari kedua model tampak hampir identik, menunjukkan bahwa keduanya memiliki kemampuan yang sama dalam membedakan antara kelas positif dan negatif. Nilai AUC sebesar 0.88 mengindikasikan bahwa terdapat peluang sebesar 88% bahwa model dapat membedakan secara benar antara sampel positif dan negatif secara acak. Kurva yang berada jauh di atas garis diagonal (random guess) juga memperkuat bahwa model ini bekerja jauh lebih baik dibandingkan dengan tebakan acak. Oleh karena itu, baik “Model1” maupun “Model2” dapat dikatakan memiliki performa klasifikasi yang sangat baik dan setara dari sisi metrik ROC-AUC.



Gambar 4. Perbandingan skor evaluasi Model1 dan Model2

Berdasarkan hasil evaluasi model klasifikasi menggunakan algoritma *Decision Tree C4.5* pada (Gambar 3), “Model1” menunjukkan performa yang lebih baik dibandingkan “Model2”. “Model1” mencatatkan akurasi sebesar 92%, lebih tinggi dibandingkan “Model2” yang hanya mencapai 90%. Dari segi sensitivitas (*recall*), “Model1” juga lebih unggul dengan nilai 95% dibandingkan 94% pada “Model2”, menunjukkan kemampuannya yang lebih baik dalam mendeteksi kelas positif, yaitu pemilih yang hadir. Meskipun nilai *precision* kedua model sama yaitu 94%, *F1-Score* “Model1” lebih tinggi (95%) dibandingkan “Model2” (94%), yang mengindikasikan keseimbangan yang lebih baik antara *precision* dan *recall*. Nilai *AUC* yang identik pada kedua model, yaitu 0,88, mengindikasikan bahwa keduanya memiliki kemampuan diskriminatif yang setara dalam membedakan antara kelas.

Secara implisit, hasil ini menunjukkan bahwa penghapusan atribut seperti lokasi dan RT pada “Model2” berdampak negatif terhadap performa model, meskipun atribut tersebut memiliki korelasi rendah secara statistik. Hal ini mengindikasikan bahwa informasi spasial dan struktural tetap memiliki peran penting dalam membentuk keputusan partisipasi pemilih. Dengan demikian, “Model1” dapat disimpulkan sebagai model yang lebih optimal dalam konteks prediksi partisipasi pemilu, dan mendukung hipotesis bahwa faktor-faktor keluarga dan lokasi tempat tinggal turut berkontribusi dalam keputusan seseorang untuk berpartisipasi dalam pemilu.

#### 4. SIMPULAN

Berdasarkan hasil pengujian model klasifikasi menggunakan algoritma *Decision Tree C4.5*, “Model1”—yang mengikutsertakan seluruh atribut—menunjukkan performa yang lebih unggul dibandingkan “Model2”, yang mengecualikan atribut “lokasi” dan “RT”. “Model1” berhasil mencapai akurasi sebesar 92%, dengan sensitivitas (*recall*) sebesar 95% dan *F1-Score* sebesar 95%. Sementara itu, “Model2” mencatatkan akurasi 90%, *recall* 94%, dan *F1-Score* 94%. Meskipun nilai *AUC* pada kedua model sama-sama tinggi, yaitu sebesar 0,88, “Model1” lebih efektif dalam mengidentifikasi kelas positif dan menghasilkan kesalahan prediksi yang lebih sedikit, sehingga lebih optimal dalam konteks klasifikasi tingkat partisipasi pemilu. Evaluasi ini menegaskan bahwa keterlibatan keluarga, yang diwakili oleh atribut seperti “hadir\_kel” dan “jml\_KK”, memiliki peranan signifikan dalam memprediksi partisipasi pemilu. Atribut “hadir\_kel” menunjukkan korelasi paling kuat terhadap kehadiran individu, yang mengindikasikan bahwa partisipasi dalam pemilu cenderung dipengaruhi oleh perilaku dan keputusan kolektif dalam keluarga. Temuan ini memperkuat argumen bahwa penguatan ikatan keluarga dan komunikasi antaranggota keluarga dapat menjadi strategi efektif dalam meningkatkan tingkat partisipasi pemilih, terutama di masyarakat dengan struktur sosial yang kuat seperti di Indonesia. Sebagai rekomendasi, hasil penelitian ini dapat dijadikan dasar untuk merancang program kampanye pemilu yang lebih terfokus pada keluarga sebagai unit sosial, bukan hanya individu. Penggunaan model klasifikasi yang mengikutsertakan atribut keterlibatan keluarga secara menyeluruh disarankan untuk menghasilkan prediksi yang lebih akurat dan relevan. Selain itu, penelitian lanjutan dapat mengkaji lebih dalam variabel-variabel sosial dan kultural lain yang turut mempengaruhi partisipasi pemilih, guna meningkatkan akurasi model serta memperkaya pemahaman tentang dinamika sosial dalam proses demokrasi.

#### REFERENSI

- [1] S. A. Hasibuan, “SISTEM PEMILIHAN UMUM DALAM KAITAN DENGAN NEGARA DEMOKRASI,” *War. Dharmawangsa*, vol. 17, no. 2, 2023, doi: 10.46576/wdw.v17i2.3174. <https://doi.org/10.46576/wdw.v17i2.3174>
- [2] Y. F. Basyari, D. Mutiarin, and Muhammad Noor Cahyadi Eko Saputro, “Implikasi penerapan sistem e-voting dalam pemilihan kepala desa melalui konsep agile government di kabupaten sleman,” *INFOTECH J. Inform. Teknol.*, vol. 4, no. 1, pp. 85–94, 2023, doi: 10.37373/infotech.v4i1.559. <https://doi.org/10.37373/infotech.v4i1.559>
- [3] A. F. Adam, S. H. Enala, F. Kontu, and M. N. Prasetya, “Sosialisasi dan Pelatihan

- Peningkatan Partisipasi Politik dan Kebijakan Publik,” *Abdimasku J. Pengabd. Masy.*, vol. 6, no. 1, p. 165, 2023, doi: 10.33633/ja.v6i1.1017. <https://doi.org/10.33633/ja.v6i1.1017>
- [4] A. Fitriyani, A. D. Rosadi, F. Laila, D. Puspita, and A. Imron, “Determinan Persepsi Perilaku Generasi Milenial Dengan Generasi Z Mempengaruhi Keputusan Dalam Pemilihan Umum Tahun 2024.,” *J. Educ. Sci.*, vol. 10, no. 1, pp. 14–24, 2024.
- [5] M. Z. Alfaruqy, “Generation Z and the Perceived Values From the Parents,” *Psyche (Stuttg.)*, vol. 4, no. 1, pp. 85–95, 2022, [Online]. Available: <http://journal.uml.ac.id/TIT>. <https://doi.org/10.36269/psyche.v4i1.658>
- [6] C. T. I. Lubis, “Penanaman Pendidikan Politik Dalam Meningkatkan Partisipasi Pemilih Pemula Pada Pemilihan Serentak 2024,” *J. Bakti Sos.*, vol. 2, no. 1, pp. 36–43, 2023. <https://doi.org/10.62668/attariiz.v2i01.533>
- [7] A. Sudarmaji, “Menggagas Pendidikan Berbasis Universal,” vol. 8, no. 2, pp. 91–110, 2024. <https://doi.org/10.32533/08201.2024>
- [8] E. Hassan, *Enhancing coffee bean classification: a comparative analysis of pre-trained deep learning models*, vol. 36, no. 16. Springer London, 2024. doi: 10.1007/s00521-024-09623-z. <https://doi.org/10.1007/s00521-024-09623-z>
- [9] 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 J. Ilm. Tek. Inform. dan Sist. Inf.*, vol. 13, no. 1, p. 201, 2024, doi: 10.35889/jutisi.v13i1.1753. <https://doi.org/10.35889/jutisi.v13i1.1753>
- [10] A. B. R. Arif senja Fitriani, Martin William, Dini Adni Navastara, Diana Purwitasari, Yudhi Purwananto, “Polarization-Affinity of Indonesian Political Figures: Insights from Network of Twitter Followers with Modularity Clustering and Linear Arrangement,” 2024 7th International Conference on Information and Communications Technology (ICOIACT), 2024. doi: 10.1109/ICOIACT64819.2024.10913292. <https://doi.org/10.1109/ICOIACT64819.2024.10913292>
- [11] A. S. Fitriani, N. E. Pratama, A. B. Raharjo, Y. Purwananto, and D. Purwitasari, “A Comparative Study on Machine Learning based Prediction Models for Public Participation Rate in an Election Voting,” in *Proceedings of the International Conference on Electrical Engineering and Informatics*, 2022. doi: 10.1109/IConEEI55709.2022.9972283. <https://doi.org/10.1109/IConEEI55709.2022.9972283>
- [12] A. Korotcov, V. Tkachenko, D. P. Russo, and S. Ekins, “Comparison of Deep Learning with Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Data Sets,” *Mol. Pharm.*, vol. 14, no. 12, 2017, doi: 10.1021/acs.molpharmaceut.7b00578. <https://doi.org/10.1021/acs.molpharmaceut.7b00578>
- [13] A. Korotcov, V. Tkachenko, D. P. Russo, S. Ekins, and M. P. Author, “Comparison of Deep Learning With Multiple Machine Learning Methods and Metrics Using Diverse Drug Discovery Datasets HHS Public Access Author manuscript Graphical abstract,” *Mol Pharm*, vol. 14, no. 12, 2017. <https://doi.org/10.1021/acs.molpharmaceut.7b00578>
- [14] A. S. Fitriani, M. A. Rosid, C. Taurusta, and I. Fauzia, “Classification Using C4.5 Algorithm in Election Participation Prediction,” in *IOP Conference Series: Materials Science and Engineering*, 2020. doi: 10.1088/1757-899X/874/1/012016. <https://doi.org/10.1088/1757-899X/874/1/012016>
- [15] S. Agarwal, “Data mining: Data mining concepts and techniques,” in *Proceedings - 2013 International Conference on Machine Intelligence Research and Advancement, ICMIRA 2013*, 2014. doi: 10.1109/ICMIRA.2013.45. <https://doi.org/10.1109/ICMIRA.2013.45>
- [16] E. Iwatani, “Overview of data mining’s potential benefits and limitations in education research,” *Pract. Assessment, Res. Eval.*, vol. 23, no. 15, 2018.
- [17] M. Ayub, “Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer,” no. January 2012, pp. 21–30, 2018.
- [18] P. Al Muqsith Prasetyo and A. Hermawan, “Analisis sentimen twitter terhadap pemilihan presiden menggunakan algoritma naïve bayes,” *INFOTECH J. Inform. Teknol.*, vol. 4, no. 2,

- pp. 224-233, 2023, doi: 10.37373/infotech.v4i2.863.  
<https://doi.org/10.37373/infotech.v4i2.863>
- [19] R. N. Sudardjat and V. S. Yosephine, "Integrasi machine learning dan sistem manajemen gudang untuk pemantauan inventory penjualan grosir minuman UMKM Integration of machine learning and warehouse management systems for inventory monitoring in UMKM wholesale beverage sales," vol. 5, pp. 121-130, 2024.  
<https://doi.org/10.37373/infotech.v5i1.1186>