



Similarity Report

Metadata

Name of the organization

Universitas Muhammadiyah Sidoarjo

Title

20250708_Artikel

Author(s) Coordinator

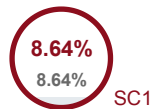
UmsidaUmsida

Organizational unit

FST

Record of similarities

SCs indicate the percentage of the number of words found in other texts compared to the total number of words in the analysed document. Please note that high coefficient values do not automatically mean plagiarism. The report must be analyzed by an authorized person.

**3968**

Length in words

28414

Length in characters

Alerts

In this section, you can find information regarding text modifications that may aim at temper with the analysis results. Invisible to the person evaluating the content of the document on a printout or in a file, they influence the phrases compared during text analysis (by causing intended misspellings) to conceal borrowings as well as to falsify values in the Similarity Report. It should be assessed whether the modifications are intentional or not.

Characters from another alphabet	ß	1
Spreads	A→	0
Micro spaces		3
Hidden characters	␣	0
Paraphrases (SmartMarks)	a	26

Active lists of similarities

This list of sources below contains sources from various databases. The color of the text indicates in which source it was found. These sources and Similarity Coefficient values do not reflect direct plagiarism. It is necessary to open each source, analyze the content and correctness of the source crediting.

The 10 longest fragments

Color of the text

NO	TITLE OR SOURCE URL (DATABASE)	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
1	Uji Akurasi K-Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan Indra Astutik Ika Ratna Universitas Muhammadiyah Sidoarjo, Fitriani Arif Senja Universitas Muhammadiyah Sidoarjo, Chulloh Dafid Mizta, Eviyanti Ade Universitas Muhammadiyah Sidoarjo;	52 1.31 %
2	https://teknosi.fti.unand.ac.id/index.php/teknosi/article/download/2280/pdf	25 0.63 %
3	Peran Komisi Pemilihan Umum Dalam Meningkatkan Partisipasi Generasi-Z Pada Pemilu Tahun 2024 Di Kabupaten Agam Diazzaki Yuris Fadila, Adil Mubarak;	22 0.55 %

4	http://digilib.isi.ac.id/14417/3/EMILLO%20IN%20ZIGHANA_2023_BAB%20V.pdf	18 0.45 %
5	Uji Akurasi K-Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan Indra Astutik Ika Ratna Universitas Muhammadiyah Sidoarjo, Fitrani Arif Senja Universitas Muhammadiyah Sidoarjo, Chulloh Dafid Mizta, Eviyanti Ade Universitas Muhammadiyah Sidoarjo;	15 0.38 %
6	https://teknosi.fti.unand.ac.id/index.php/teknosi/article/download/2280/pdf	14 0.35 %
7	Uji Akurasi K-Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan Indra Astutik Ika Ratna Universitas Muhammadiyah Sidoarjo, Fitrani Arif Senja Universitas Muhammadiyah Sidoarjo, Chulloh Dafid Mizta, Eviyanti Ade Universitas Muhammadiyah Sidoarjo;	13 0.33 %
8	Uji Akurasi K-Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan Indra Astutik Ika Ratna Universitas Muhammadiyah Sidoarjo, Fitrani Arif Senja Universitas Muhammadiyah Sidoarjo, Chulloh Dafid Mizta, Eviyanti Ade Universitas Muhammadiyah Sidoarjo;	13 0.33 %
9	Uji Akurasi K-Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan Indra Astutik Ika Ratna Universitas Muhammadiyah Sidoarjo, Fitrani Arif Senja Universitas Muhammadiyah Sidoarjo, Chulloh Dafid Mizta, Eviyanti Ade Universitas Muhammadiyah Sidoarjo;	13 0.33 %
10	https://dosenppkn.com/contoh-kata-pengantar/	13 0.33 %

from RefBooks database (5.54 %)

NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
Source: Paperity		
1	Uji Akurasi K-Means dalam Prediksi Partisipasi Pemilu pada Demografi Wilayah Kabupaten Pasuruan Indra Astutik Ika Ratna Universitas Muhammadiyah Sidoarjo, Fitrani Arif Senja Universitas Muhammadiyah Sidoarjo, Chulloh Dafid Mizta, Eviyanti Ade Universitas Muhammadiyah Sidoarjo;	125 (7) 3.15 %
2	Peran Komisi Pemilihan Umum Dalam Meningkatkan Partisipasi Generasi-Z Pada Pemilu Tahun 2024 Di Kabupaten Agam Diazzaki Yuris Fadila, Adil Mubarak;	22 (1) 0.55 %
3	PENERAPAN METODE NAIVE BAYES PADA ANALISIS SENTIMEN APLIKASI MCDONALDS DI GOOGLE PLAY STORE Fathwa Daud Dhyya' Rohannisa, Irawan Bambang, Agus Bahtiar;	12 (1) 0.30 %
4	Kombinasi Metode Multi-Attribute Utility Theory dan Pivot Pairwise Relative Criteria Importance Assessment Dalam Penentuan Lulusan Terbaik Maharani Ainaya Qintan, Temi Ardiansah;	11 (1) 0.28 %
5	KLASIFIKASI TINGKAT PARTISIPASI PEMILU BERDASARKAN SEKTOR INDUSTRI MENGGUNAKAN ALGORITMA NAÏVE BAYES Dijaya Rohman, Fitrani Arif Senja, Yusuf Raharja;	11 (1) 0.28 %
6	ANALISIS SENTIMEN MENGGUNAKAN METODE NAIVE BAYES DAN METODE SUPPORT VECTOR MACHINE PADA KASUS PELANTIKAN ARTIS SEBAGAI ANGGOTA ANGGOTA DPR RI TAHUN 2024 Tiwuk Widiastuti, Lete Patrisius Remby, Mola Sebastian Adi Santoso, Pa. Bernard Jose Adrian Junio Ajilo, Triyanto;	9 (1) 0.23 %
7	IMPLEMENTASI METODE KLASIFIKASI DENGAN ALGORITMA SUPPORT VECTOR MACHINE KERNEL GAUSSIAN RBF UNTUK PREDIKSI PARTISIPASI PEMILU TERHADAP DEMOGRAFI KOTA SURABAYA Fitrani Arif Senja Universitas Muhammadiyah Sidoarjo, Safitri Dewi Eka Universitas Muhammadiyah Sidoarjo;	8 (1) 0.20 %
8	Implementasi hukum lingkungan melalui penyuluhan pemilahan sampah di Desa Kesiman Kertalangu Indradana Gde Nararya, Sahri Aflah Ramadiansyah;	6 (1) 0.15 %

9	PREDIKSI KETERLAMBATAN SEBUAH MASKAPAI MENGGUNAKAN METODE K-NEAREST NEIGHBOR Ardiansyah Jaya Winata;	6 (1) 0.15 %
10	IMPLEMENTASI KLASIFIKASI KEHAMILAN BERESIKO DENGAN METODE NAIVE BAYES PADA PUSKESMAS KELURAHAN MALAKA JAYA Sri Lestari, Virginia Khoirunnisa;	5 (1) 0.13 %
11	Analisis Perbandingan Naïve Bayes dan Neural Network dalam Klasifikasi Minat Masyarakat pada Kursus Komputer Sudi Suryadi,Fitria Nabila Syah, Nasution Fitri Aini;	5 (1) 0.13 %
from the home database (0.00 %)		
NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
from the Database Exchange Program (0.00 %)		
NO	TITLE	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
from the Internet (3.10 %)		
NO	SOURCE URL	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
1	https://teknosi.fti.unand.ac.id/index.php/teknosi/article/download/2280/pdf	39 (2) 0.98 %
2	http://digilib.isi.ac.id/14417/3/EMILLO%20IN%20ZIGHANA_2023_BAB%20V.pdf	18 (1) 0.45 %
3	https://journal.unej.ac.id/JEI/article/download/1629/1028/10589	16 (2) 0.40 %
4	https://jurnalistiqomah.org/index.php/jemb/article/download/1842/1489	14 (2) 0.35 %
5	https://dosenppkn.com/contoh-kata-pengantar/	13 (1) 0.33 %
6	https://myskripsi.ums.ac.id/media/konsultasi/a510210013/Luaran_Artikel_Bahasa_Indonesia_EfMg29p.d ocx	12 (1) 0.30 %
7	http://repository.umsu.ac.id/bitstream/handle/123456789/25774/Skripsi%28SitiRaisha-2009020076%29.pdf?sequence=1	11 (1) 0.28 %

List of accepted fragments (no accepted fragments)

NO	CONTENTS	NUMBER OF IDENTICAL WORDS (FRAGMENTS)
----	----------	---------------------------------------

Prediction of Voter Participation Levels Using the Naive Bayes Algorithm Based on the Village Development Index Data in Sidoarjo Regency
[Prediksi Tingkat Partisipasi Pemilu Menggunakan Algoritma Naive Bayes Berdasarkan Data Indeks Desa Membangun di Kabupaten Sidoarjo]

Husnul Abidin1), Arif Senja Fitrani 2), Hamzah Setiawan3), Uce Indahyanti4).
1),2),3),4) Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia
Page | 1
2 | Page
Page | 3

Abstract. Electoral participation serves as a crucial indicator in assessing the success of democracy, particularly in rural areas that possess diverse social and economic characteristics. This study aims to predict the level of electoral participation in the villages of Sidoarjo Regency by utilizing the Village Development Index (Indeks Desa Membangun/IDM) indicators as predictor variables. The Naïve Bayes classification method is applied to process IDM data, which includes social, economic, and ecological aspects, in order to build a comprehensive predictive model. The test results show that the Naïve Bayes algorithm is capable of predicting electoral participation with an accuracy of 78.65%, a precision of 79%, a recall of 76%, and an F1-score of 77%. These metrics indicate that the model performs fairly well in classifying electoral participation based on IDM factors. Additionally, data visualization highlights the villages with the highest and lowest participation differences, offering further insights into the relationship between village conditions and levels of political engagement. This research contributes to the development of data mining-based information systems that can be used as strategic planning tools to enhance political participation in rural areas.

Abstrak. Partisipasi pemilu menjadi indikator penting dalam menilai keberhasilan demokrasi, terutama di wilayah pedesaan yang memiliki karakteristik sosial dan ekonomi yang beragam. Penelitian ini bertujuan untuk memprediksi tingkat partisipasi pemilu di desa-desa Kabupaten Sidoarjo dengan memanfaatkan indikator Indeks Desa Membangun (IDM) sebagai variabel prediktor. Metode klasifikasi Naïve Bayes diterapkan untuk mengolah data IDM yang mencakup aspek sosial, ekonomi, dan ekologi guna membangun model prediktif yang komprehensif. Hasil pengujian menunjukkan bahwa algoritma Naïve Bayes mampu memprediksi partisipasi pemilu dengan akurasi sebesar 78,65%, precision sebesar 79%, recall sebesar 76%, dan F1-score sebesar 77%. Nilai-nilai tersebut mengindikasikan performa model yang cukup baik dalam mengklasifikasikan partisipasi pemilu berdasarkan faktor-faktor IDM. Selain itu, visualisasi data turut menunjukkan desa-desa dengan selisih partisipasi tertinggi dan terendah, memberikan wawasan lebih lanjut mengenai keterkaitan kondisi desa dengan tingkat partisipasi politik. Penelitian ini memberikan kontribusi dalam pengembangan sistem informasi berbasis data mining yang dapat dimanfaatkan sebagai alat bantu perencanaan strategis dalam meningkatkan partisipasi politik masyarakat di wilayah pedesaan.

Kata Kunci - Pemilu, Partisipasi Politik, Indeks Desa Membangun, Naive Bayes, Data Mining

1. I. Pendahuluan

Pemilihan umum merupakan salah satu pilar utama demokrasi, yang mencerminkan tingkat partisipasi politik masyarakat sebagai indikator keberhasilan pemerintahan dalam membangun keterlibatan publik [1]. Di Indonesia, upaya untuk memahami faktor-faktor yang memengaruhi partisipasi pemilu menjadi semakin penting, khususnya dalam konteks pedesaan. Kabupaten Sidoarjo, sebagai wilayah yang memiliki karakteristik desa yang heterogen dari sisi sosial, ekonomi, dan infrastruktur, menjadi objek kajian yang relevan [2].

Untuk memetakan potensi desa dan tingkat kemandiriannya, pemerintah telah mengembangkan Indeks Desa Membangun (IDM) sebagai alat ukur komprehensif yang mencakup indikator layanan dasar, pendidikan, kesehatan, infrastruktur, dan partisipasi sosial [3]. Berdasarkan asumsi bahwa aspek-aspek ini turut membentuk kesadaran dan kepedulian politik masyarakat, penelitian ini bertujuan untuk menguji sejauh mana indikator IDM dapat digunakan sebagai variabel prediktor terhadap tingkat partisipasi pemilu masyarakat desa di Sidoarjo.

Penelitian ini mengimplementasikan algoritma Naïve Bayes untuk memprediksi partisipasi pemilu berdasarkan data IDM di wilayah Kabupaten Sidoarjo [4]. Tujuannya adalah untuk memetakan kecenderungan partisipasi politik masyarakat secara lebih terukur, dengan mempertimbangkan aspek sosial, ekonomi, dan infrastruktur yang tercakup dalam indikator IDM [5]. Pendekatan ini memberikan alternatif baru dalam memahami partisipasi politik di tingkat desa dengan dukungan teknologi data mining berbasis klasifikasi probabilistik [5].

Terdapat kesamaan antara penelitian ini dan studi sebelumnya yang berjudul “**Penerapan Data Mining Menggunakan Metode Klasifikasi Naïve Bayes untuk Memprediksi Partisipasi Pemilihan Gubernur,**” di mana keduanya memanfaatkan algoritma Naïve Bayes dalam konteks pemilihan [5]. Namun, penelitian tersebut hanya memanfaatkan variabel demografis seperti usia, jenis kelamin, dan jarak TPS, serta belum memanfaatkan data IDM secara menyeluruh sebagai landasan prediksi. Di sisi lain, penelitian ini menggabungkan indikator-indikator penting dari IDM, seperti ketahanan sosial, ekonomi, dan ekologi, untuk membangun model klasifikasi yang lebih menyeluruh. Selain itu, jika pada penelitian-penelitian sebelumnya konteks wilayah yang digunakan bersifat nasional atau umum, penelitian ini secara khusus difokuskan pada wilayah Kabupaten Sidoarjo, yang memiliki karakteristik desa dan dinamika sosial tersendiri. Fokus lokal ini menjadi keunggulan karena memungkinkan model yang dikembangkan lebih kontekstual dan sesuai dengan kondisi nyata di lapangan [3].

Lebih lanjut, penelitian ini juga menggunakan platform Google Colab berbasis Python, yang mendukung pengolahan data secara cepat, fleksibel, dan bersifat open source probabilistik [6]. Hal ini memberikan nilai tambah dari sisi transparansi, replikasi, dan efisiensi teknis, dibandingkan dengan penelitian terdahulu yang sebagian besar masih menggunakan perangkat lunak tertutup seperti Weka. Dengan demikian, penelitian ini hadir sebagai pendekatan baru yang lebih kontekstual, komprehensif, dan teknologis dalam memprediksi partisipasi pemilu di tingkat desa, serta diharapkan dapat memberikan gambaran awal yang kuat untuk strategi peningkatan partisipasi politik masyarakat di wilayah pedesaan, khususnya di Kabupaten Sidoarjo.

2. II. Metode

Tahapan Penelitian Tahapan penelitian merupakan serangkaian langkah sistematis yang dilakukan mulai dari perencanaan awal hingga evaluasi akhir guna mencapai tujuan penelitian. Pada penelitian ini, proses dimulai dari studi literatur, dilanjutkan dengan pengumpulan data, tahap pre-processing, pemrosesan data menggunakan algoritma Naïve Bayes, hingga tahap validasi model. Penelitian ini dilakukan dengan menggunakan IDM dan data partisipasi pemilu tahun 2024 di Kabupaten Sidoarjo. Rangkaian tahapan penelitian secara umum ditunjukkan pada Gambar 1 berikut:

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 dengan mengakses IDM melalui laman resmi Kemendesa PDTT, serta memperoleh data partisipasi pemilu tahun 2024 dari situs KPU. Kabupaten Sidoarjo secara keseluruhan, termasuk setiap kecamatan dan desanya, menjadi fokus area dalam pengumpulan data. Dataset yang diperoleh telah diintegrasikan dalam satu file dan mencakup 48 atribut, dengan informasi yang berkaitan dengan indikator sosial, ekonomi, ekologi, kependudukan, dan status perkembangan desa. Variabel target dalam penelitian ini adalah tingkat partisipasi pemilu, yang akan diklasifikasikan menggunakan algoritma Naïve Bayes [7]. Data ini dianalisis menggunakan bahasa pemrograman Python melalui platform Google Colab untuk membangun model klasifikasi prediktif seperti yang ditunjukkan pada Tabel 1.

Tabel 1. Atribut Data

No	Atribut Kelompok	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 Naïve Bayes pada data IDM yang mencakup proses penggabungan data IDM dan data partisipasi pemilu, pembersihan data dari nilai kosong atau tidak konsisten, transformasi nilai jika diperlukan, serta pengelompokan data ke dalam kategori partisipasi (tinggi, rendah) [6]. Tahapan ini penting dilakukan agar data siap digunakan dalam proses klasifikasi tingkat partisipasi pemilu di Kabupaten Sidoarjo.

1. Data Cleaning

Data cleaning dilakukan untuk membersihkan dataset dari nilai kosong, teks tidak konsisten, dan data yang kurang relevan [11]. Pada dataset ini ditemukan beberapa masalah, seperti nilai “-” (kosong), data kategorikal seperti “Ada” dan “Tidak Ada” yang perlu diubah ke bentuk numerik, serta kolom dengan nilai yang seragam di seluruh baris. Penanganan dilakukan dengan mengganti data kosong menggunakan nilai median atau modus, mengubah data teks menjadi angka, dan menghapus atribut yang tidak memberikan informasi penting agar data siap digunakan dalam proses klasifikasi [8].

2. Data Transformation

Pada tahap ini, data diubah ke dalam format numerik agar dapat diproses oleh algoritma Naïve Bayes [10]. Atribut yang semula berbentuk teks seperti “Ada” dan “Tidak Ada” dikonversi menjadi angka, misalnya “Ada” menjadi 1 dan “Tidak Ada” menjadi 0. Transformasi ini bertujuan untuk menyelaraskan format data dengan kebutuhan proses klasifikasi berbasis data mining [7].

3. Seleksi Fitur

Dataset yang digunakan terdiri dari 48 atribut prediktor yang mewakili indikator-indikator dalam data Indeks Desa Membangun (IDM). Untuk menyederhanakan proses klasifikasi dan meningkatkan akurasi model, dilakukan seleksi fitur dengan menghapus beberapa kelompok atribut yang dianggap kurang berpengaruh terhadap tingkat partisipasi pemilu [9]. Atribut yang dihilangkan antara lain kelompok Utilitas Dasar, Produksi Desa, Pendukung Ekonomi, Pengelolaan Lingkungan, Penanggulangan Bencana, dan Tata Kelola Keuangan Desa. Hasilnya, tersisa 27 atribut prediktor dari 7 kelompok indikator yang dinilai relevan, serta 1 atribut target yaitu tingkat partisipasi masyarakat dalam 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
---	--------	--------------------------------

4. Random dataset

Pengacakan data dilakukan untuk memastikan setiap baris memiliki peluang yang sama dalam mewakili seluruh atribut. Dengan mengacak urutan data, model Naïve Bayes dapat belajar dari data secara merata dan tidak terpengaruh oleh susunan awal dalam dataset. Proses ini dilakukan sebelum data dibagi menjadi data latih dan data uji [2].

5. Split data

Setelah melalui tahap pra-pemrosesan, data kemudian dibagi menjadi dua bagian, yaitu 70% sebagai **data latih (training)** dan 30% sebagai **data uji (testing)**. **Pembagian ini** dilakukan untuk melatih model Naïve Bayes menggunakan sebagian besar **data**, sementara sisanya **digunakan untuk menguji seberapa baik model dapat memprediksi data baru** [10].

3. Process

Gambar 1. Flowchart Pemrosesan Data

Flowchart pada Gambar 1 di atas menggambarkan tahapan pemrosesan data dalam implementasi algoritma Naïve Bayes pada data IDM untuk

memprediksi tingkat partisipasi pemilu di Kabupaten Sidoarjo. Proses dimulai dengan menyiapkan dataset berupa data IDM yang telah digabungkan dengan data partisipasi pemilu 2024. Data ini kemudian diproses dan dimasukkan sebagai data latih melalui tahap pre-processing, seperti pembersihan, transformasi format, dan pengkategorian kelas target (tinggi, rendah). Selanjutnya, algoritma Naïve Bayes dijalankan untuk menghitung probabilitas masing-masing kelas berdasarkan fitur yang tersedia, dengan asumsi independensi antar atribut. Model mempelajari distribusi data latih melalui perhitungan probabilitas prior dan likelihood, lalu diuji menggunakan data uji untuk mengevaluasi performa klasifikasinya. Hasil akhirnya adalah prediksi label partisipasi pemilu berdasarkan karakteristik desa, dan proses ditutup dengan evaluasi menggunakan metrik seperti akurasi, precision, **recall, dan F1-score** [11].

4. Output

Setelah proses pelatihan model selesai, algoritma Naïve Bayes menghasilkan output berupa prediksi tingkat partisipasi pemilu masyarakat desa yang terbagi ke dalam dua kategori, yaitu “tinggi” dan “rendah”, berdasarkan indikator-indikator dalam Indeks Desa Membangun.

5. Analisis / Evaluasi

Evaluasi merupakan tahap akhir dalam penelitian ini yang bertujuan untuk menilai sejauh mana model Naïve Bayes mampu memprediksi tingkat partisipasi pemilu dengan tepat. Pada tahap ini, performa model dianalisis menggunakan sejumlah **metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, yang** memberikan gambaran menyeluruh mengenai ketepatan prediksi model terhadap data yang diuji. Melalui evaluasi ini, peneliti dapat mengetahui kekuatan dan kelemahan model serta menentukan apakah model layak digunakan untuk prediksi lebih lanjut di masa mendatang.

3. III. Hasil dan Pembahasan

Classification Report

Pada Tabel 3 merupakan Classification Report yang menunjukkan bahwa performa model sangat baik dalam mengklasifikasikan desa dengan partisipasi tinggi, dengan nilai precision, recall, dan F1-score masing-masing 0.90, 0.82, dan 0.86. **Namun, model mengalami kesulitan dalam mengenali partisipasi rendah. terlihat dari nilai** precision hanya 0.27 dan recall sebesar 0.42. Hal ini mengindikasikan bahwa model cenderung salah dalam mengklasifikasikan desa yang sebenarnya memiliki partisipasi rendah sebagai tinggi [12]. Performa yang buruk pada kelas "rendah" kemungkinan besar disebabkan oleh ketimpangan jumlah data antara kelas tinggi (1268 data) dan kelas rendah (204 data), yang membuat model lebih banyak belajar dari pola kelas mayoritas.

Tabel 3. Classification Report

	Precision	Recall	F1-Score	Support
0	0.27	0.42	0.33	204
1	0.90	0.82	0.86	1268
Akurasi	0.76 1472			
Macro Avg	0.59	0.62	0.59	1472
Weighted Avg	0.81	0.76	0.78	1472

Confusion Matrix

Gambar 2 di bawah menyajikan Confusion Matrix dari hasil pengujian model terhadap data uji. Berdasarkan visualisasi tersebut, model mampu mengklasifikasikan dengan benar sebanyak 1.040 data pada kelas partisipasi tinggi, serta 86 data pada kelas partisipasi rendah. Sementara itu, terdapat sejumlah data yang belum dapat diklasifikasikan secara tepat, yaitu 118 data dari kelas rendah yang terprediksi sebagai tinggi dan 228 data dari kelas tinggi yang terklasifikasi sebagai rendah. Temuan ini menunjukkan bahwa model cenderung lebih akurat dalam mengenali pola dari kelas mayoritas (partisipasi tinggi), namun masih menghadapi tantangan dalam membedakan karakteristik desa dengan partisipasi rendah [13]. Analisis lanjutan terhadap fitur-fitur yang berkontribusi pada kelas rendah dapat menjadi strategi yang penting untuk meningkatkan performa model secara keseluruhan.

Gambar 2. Confusion Matrix

Presentase Tingkat Partisipasi

Gambar 3 menggambarkan distribusi kategori partisipasi pemilu pada desa-desa yang dianalisis. Terlihat bahwa sebagian besar desa, yaitu sebanyak 4.211 desa (85,9%), berada pada kategori partisipasi tinggi, sedangkan hanya 694 desa (14,1%) yang tergolong dalam kategori partisipasi rendah. Ketidakseimbangan ini merupakan hal yang umum dalam data sosial, di mana dominasi kelas mayoritas sering terjadi. Kondisi seperti ini perlu menjadi perhatian dalam proses pemodelan, karena dapat memengaruhi performa model, khususnya dalam mengenali pola dari kelas minoritas. Oleh karena itu, strategi seperti penyeimbangan data melalui resampling atau oversampling dapat dipertimbangkan dalam penelitian lanjutan agar setiap kelas memiliki peluang yang setara untuk dipelajari secara optimal oleh model [14].

Gambar 3. Presentase Tingkat Partisipasi di Sidoarjo

Desa dengan Partisipasi Tertinggi - Terendah

Pada Gambar 4 menunjukkan persentase partisipasi pemilih di beberapa desa yang tersebar di berbagai kecamatan di Kabupaten Sidoarjo. Desa Krembung di Kecamatan Krembung tercatat sebagai desa dengan partisipasi pemilu tertinggi, yaitu sebesar 95,97%, disusul oleh Desa Mojorangagung di Kecamatan Wonoayu (94,96%) dan Desa Gempolklutuk di Kecamatan Tarik (93,64%). Ketiga desa ini menunjukkan tingkat keterlibatan masyarakat yang sangat tinggi dalam pelaksanaan pemilu. Sementara itu, beberapa desa tercatat memiliki tingkat partisipasi yang cukup rendah. Desa Tambakoso di Kecamatan Waru merupakan desa dengan partisipasi terendah, yakni hanya 53,36%, diikuti oleh Wadungasri (68,03%) dan Betro (71,89%), yang juga berada di kawasan Sidoarjo bagian timur.

No Desa Kecamatan Partisipasi (%)

1	Krembung	Krembung	95.97
2	Mojorangagung	Wonoayu	94.96
3	Gempolklutuk	Tarik	93.64
4	Sumokembangsri	Balombangendo	93.61
5	Plumbon	Porong	92.72
...
6	Sawotratap	Gedangan	72.67
7	Gelang	Tulangan	72.31

8	Betro	Sedati	71.89
9	Wadungasri	Waru	68.03
10	Tambakoso	Waru	53.36

Gambar 4. Korelasi Atribut

Desa dengan Partisipasi Tertinggi - Terendah

Tabel tersebut menyajikan selisih berbagai indikator antara dua desa di Kabupaten Sidoarjo yang memiliki tingkat partisipasi pemilu paling tinggi dan paling rendah. Indikator partisipasi menunjukkan selisih sebesar 42,61%, yang mencerminkan perbedaan sangat mencolok dalam tingkat keterlibatan masyarakat dalam pemilu antar kedua wilayah tersebut. Beberapa indikator pelayanan dasar, seperti akses terhadap pendidikan tingkat menengah (SMP/MTs dan SMA/SMK) serta layanan dokter, masing-masing menunjukkan selisih 2.0 poin, mengindikasikan adanya perbedaan akses yang cukup nyata antara kedua desa tersebut. Sementara itu, sebagian besar indikator lainnya, seperti layanan telekomunikasi, kondisi jalan desa, keberadaan ruang publik, serta berbagai layanan sosial seperti posyandu dan musyawarah desa, memiliki nilai selisih 0.0, yang berarti tidak terdapat perbedaan dalam aspek tersebut antara desa dengan partisipasi tinggi dan rendah. Menariknya, beberapa indikator bahkan menunjukkan selisih negatif, seperti pada layanan tenaga kesehatan lainnya (-12.0), layanan bidan (-8.0), dan kearifan sosial/budaya (-8.0), yang menandakan bahwa desa dengan partisipasi rendah justru memiliki skor lebih tinggi pada indikator-indikator tersebut. Temuan ini menunjukkan bahwa tingginya partisipasi pemilu tidak semata-mata ditentukan oleh ketersediaan infrastruktur atau layanan publik, namun juga dipengaruhi oleh berbagai faktor sosial, budaya, dan tingkat kesadaran politik masyarakat di masing-masing wilayah.

No Indikator Selisih (Tertinggi - Terendah)

1	Partisipasi %	42.61
2	Akses terhadap SMP/ MTs/ Sederajat	2.0
3	Akses terhadap SMA/ SMK/ MA/ MAK/ Sederajat	2.0
4	Layanan Dokter	2.0
5	Akses terhadap PAUD/ TK/ Sederajat	1.0
6	Pemanfaatan Teknologi dalam Pelayanan Desa (SPBE)	0.0
7	Pelaksanaan Pelayanan dan Administrasi Desa	0.0
8	Layanan Telekomunikasi	0.0
9	Kondisi Jalan di desa	0.0
10	Keberadaan Ruang Publik Terbuka	0.0
11	Kegiatan Olahraga	0.0
12	Musyawarah Desa	0.0
13	Aktivitas Posyandu	0.0
14	Layanan Sarana Kesehatan	0.0
15	Akses terhadap SD/ MI/ Sederajat	0.0
16	Satkamling	-1.0
17	Taman Bacaan Masyarakat/ Perpustakaan Desa	-1.0
18	Fasilitas Olahraga	-1.0
19	Frekuensi Gotong Royong	-2.0
20	Fasilitas Kesehatan Poskesdes/ Polindes	-4.0
21	Jaminan Kesehatan Nasional	-5.0
22	Layanan Bidan	-8.0
23	Kearifan Sosial/ Budaya	-8.0
24	Layanan Tenaga Kesehatan Lainnya	-12.0

4. VII. Simpulan

Berdasarkan hasil analisis, mayoritas desa di Kabupaten Sidoarjo termasuk dalam kategori partisipasi tinggi, yakni sebesar 85,9%, sementara hanya 14,1% desa yang tergolong dalam partisipasi rendah. Ketimpangan distribusi ini menyebabkan model cenderung lebih akurat dalam mengklasifikasikan desa dengan partisipasi tinggi, sebagaimana terlihat pada nilai precision dan recall masing-masing 0.90 dan 0.82, dibandingkan dengan desa berpartisipasi rendah yang hanya mencapai precision 0.27 dan recall 0.42. Akurasi model secara keseluruhan tercatat sebesar 76%. Hasil confusion matrix juga menunjukkan bahwa model lebih sering salah dalam memprediksi partisipasi rendah sebagai tinggi, yang merupakan konsekuensi dari dominasi kelas mayoritas dalam data latih. Selain itu, analisis pada perbedaan indikator antara desa dengan partisipasi tertinggi dan terendah menunjukkan bahwa perbedaan signifikan hanya terdapat pada beberapa aspek, seperti akses pendidikan menengah dan layanan kesehatan dasar, sementara sebagian besar indikator lainnya memiliki selisih nol, bahkan beberapa bernilai negatif. Hal ini mengindikasikan bahwa tingkat partisipasi pemilu tidak hanya ditentukan oleh faktor layanan publik, tetapi juga oleh aspek sosial, budaya, serta kesadaran politik masyarakat. Oleh karena itu, pendekatan penyeimbangan data serta analisis fitur yang lebih mendalam menjadi penting untuk diterapkan pada penelitian selanjutnya guna meningkatkan performa model terhadap kelas minoritas.

5. Ucapan Terima Kasih

Peneliti mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam proses penyusunan penelitian ini. Ucapan terima kasih secara khusus disampaikan kepada Pemerintah Kabupaten Sidoarjo atas kerja sama dan dukungan data yang telah diberikan terkait partisipasi pemilu. Peneliti juga menyampaikan apresiasi kepada dosen pembimbing atas bimbingan dan masukan yang sangat berharga selama proses penelitian. Terima kasih juga disampaikan kepada seluruh dosen Program Studi Informatika Universitas Muhammadiyah Sidoarjo atas ilmu dan dukungannya selama masa studi. Tidak lupa, peneliti menyampaikan rasa terima kasih yang mendalam kepada keluarga dan teman-teman atas dukungan moral dan semangat yang diberikan. Semoga penelitian ini dapat memberikan manfaat dalam meningkatkan partisipasi politik masyarakat, khususnya di wilayah pedesaan.

6. 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. Djijaya, "Klasifikasi Tingkat Partisipasi Pemilu Berdasarkan Sektor Industri Menggunakan Algoritma Naive Bayes," J. Teknkom (Teknik Inf. dan Komputer), vol. 7, no. 1, pp. 135-143, 2024, doi: 10.37600/teknikom.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. Ili, "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.
- [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 Title ~~ELENH~~," *Ayay*, 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. Fitriani, 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.
- [12] F. Ratnawati, "file:///D:/Septian Dwi Hanggara/SEMANGAT SKRIPSI/Fix Jurnal/twitbab2-2.pdf," *INOVTEK Polbeng - Seri Inform.*, vol. 3, no. 1, p. 50, 2018.
- [13] 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.
- [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. 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.