

Performance of Machine Learning Algorithms on Imbalanced Sentiment Datasets Without Balancing Techniques

Dina Wulan Yekti Rahayu ^{1*}, Khothibul Umam ^{2*}, Maya Rini Handayani ^{3**}

^{*} Teknologi Informasi, Fakultas Sains dan Teknologi, UIN Walisongo Semarang

^{**} Komunikasi dan Penyiaran Islam, Fakultas Dakwah dan Komunikasi, UIN Walisongo Semarang
2208096081@student.walisongo.ac.id ¹, khothibul_umam@walisongo.ac.id ², mayarini@walisongo.ac.id ³

Article Info

Article history:

Received 2025-05-01

Revised 2025-06-15

Accepted 2025-06-20

Keyword:

*Machine Learning,
Imbalanced Datasets,
Sentiment Analysis,
SMOTE*

ABSTRACT

This study explores the performance of five sentiment classification algorithms—Naïve Bayes, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest—on an imbalanced sentiment dataset, with the SMOTE technique applied as a comparison. The research follows the Knowledge Discovery in Databases (KDD) framework, which includes data selection, preprocessing, transformation, data mining, and evaluation. The evaluation uses metrics such as accuracy, precision, recall, F1-score, and macro average F1-score. Initial results show that all five algorithms performed fairly well even without using a balancing technique, with Naïve Bayes achieving the highest F1-score of 0.84 and recall of 0.81. After applying SMOTE, only small improvements were observed in some models, such as Random Forest (F1-score increased from 0.81 to 0.85), while other models like Naïve Bayes experienced a decrease in performance, dropping to 0.77. This suggests that the effect of balancing techniques like SMOTE can vary depending on the algorithm. Thus, this study provides empirical contributions that highlight the importance of selecting appropriate approaches and the need for a deep understanding of each algorithm's behavior in the context of imbalanced data. Researchers are encouraged to carefully consider these aspects when designing experiments and interpreting results.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. PENDAHULUAN

Dataset dapat dikategorikan tidak seimbang apabila distribusi jumlah data pada masing-masing kelas kategori tidak proporsional. Dalam beberapa tahun terakhir, penggunaan algoritma machine learning untuk menyelesaikan permasalahan nyata yang kompleks semakin meningkat, dan banyak di antaranya dihadapkan pada permasalahan dataset yang tidak seimbang [1]. Salah satu bidang penerapannya adalah analisis sentimen, yaitu teknik untuk memahami pandangan dan sikap masyarakat terhadap isu tertentu [2]. Dalam penerapan machine learning untuk klasifikasi, terutama pada analisis sentimen, kondisi data yang tidak seimbang merupakan hal yang umum dijumpai. Ketidakseimbangan ini terjadi ketika satu kelas, seperti sentimen negatif, mendominasi secara signifikan dibandingkan kelas lainnya. Fenomena ini sering muncul

dalam isu-isu publik yang sensitif atau kontroversial di media sosial, di mana sentimen negatif cenderung lebih sering diekspresikan dibandingkan sentimen positif atau netral.

Untuk mengatasi masalah ini, sebagian besar penelitian menerapkan teknik balancing seperti oversampling, undersampling, atau metode sintetik seperti SMOTE. Teknik-teknik tersebut umumnya diterapkan hanya pada data latih, dengan tujuan agar model memiliki paparan yang lebih seimbang terhadap semua kelas selama proses pembelajaran. Meskipun data uji tetap dibiarkan dalam kondisi aslinya yang tidak seimbang, proses pelatihan model tetap didasarkan pada distribusi yang telah dimodifikasi secara buatan. Metode-metode tersebut masih memiliki kelemahan bawaan dan menghasilkan sampel sintetik yang kurang informatif [3]. Hal ini menimbulkan pertanyaan penting: sejauh mana performa model mencerminkan kemampuannya dalam mengenali kelas

minoritas saat dihadapkan pada data nyata yang tidak mengalami balancing sama sekali?.

Penelitian ini mengambil fokus pada isu revisi Undang-Undang TNI, yang saat ini ramai diperbincangkan di masyarakat, terutama terkait peran baru TNI dalam menghadapi “ancaman siber”. Isu ini memicu banyak respons publik yang terekam di media sosial, dan digunakan sebagai dasar untuk pengumpulan data. Teknik *web scraping* digunakan untuk mengumpulkan opini publik, menghasilkan dataset dengan distribusi sentimen yang tidak seimbang—dengan dominasi kuat pada sentimen negatif—serta jumlah data yang terbatas.

Selanjutnya, proses analisis dilakukan berdasarkan pendekatan Knowledge Discovery in Databases (KDD), yang mencakup tahapan selection, preprocessing, transformation, data mining, dan evaluation. Pada tahap data mining, penelitian ini menguji performa lima algoritma machine learning yang umum digunakan, yaitu Naïve Bayes, Logistic Regression, Support Vector Machine (SVM), Decision Tree, dan Random Forest.

Berbeda dari pendekatan umum yang sering melakukan balancing pada data latih, penelitian ini secara sengaja mempertahankan distribusi data dalam kondisi aslinya. Pendekatan ini dipilih untuk memperoleh pemahaman yang lebih jujur dan realistis mengenai kemampuan masing-masing model dalam mengenali kelas minoritas—khususnya sentimen positif—dalam kondisi distribusi data yang mencerminkan situasi nyata. Meskipun balancing data seperti SMOTE biasanya hanya dilakukan pada data latih, tetap terdapat risiko bahwa hasil pelatihan menciptakan ekspektasi performa yang tidak sepenuhnya mencerminkan kemampuan model dalam menghadapi data sebenarnya. Oleh karena itu, SMOTE tetap diterapkan dalam penelitian ini sebagai alat pembandingan, bukan sebagai pendekatan utama. Tujuannya adalah untuk melihat sejauh mana perubahan performa terjadi setelah dilakukan balancing pada data latih, dibandingkan dengan hasil model yang dilatih tanpa balancing sama sekali.

Penelitian sebelumnya oleh Pebrianti (2025) menguji penggunaan Support Vector Machine (SVM) dalam analisis sentimen terhadap kasus korupsi PT. Pertamina (Persero), dengan hasil yang menunjukkan bahwa ketidakseimbangan kelas mempengaruhi performa model. Dalam penelitian tersebut, SVM menghasilkan nilai akurasi 89% sebelum penerapan teknik penyeimbangan data menggunakan SMOTE. Ditemukan bahwa sentimen negatif mendominasi, dengan nilai precision negatif mencapai 89% dan precision positif 0%, recall untuk sentimen negatif 100% dan untuk positif 0%. [4].

Penelitian yang dilakukan oleh Stefanni dkk. (2025) dalam analisis sentimen terkait kebocoran data pribadi di platform X menggunakan algoritma Naïve Bayes Classifier menunjukkan hasil yang sangat baik dengan akurasi mencapai 98,92%. Data yang digunakan terdiri dari 433 data latih dan 186 data uji, dan hasil klasifikasi menunjukkan dominasi sentimen negatif (43,54%) diikuti oleh sentimen positif (28,50%) dan netral (27,96%). Model ini berhasil menghasilkan nilai presisi negatif dan netral sebesar 100%, sementara presisi positif mencapai 96,30%. Recall untuk sentimen negatif dan positif

masing-masing 100%, sedangkan untuk sentimen netral adalah 96,23%. Hasil ini menegaskan efektivitas Naïve Bayes dalam mengklasifikasikan sentimen terkait isu kebocoran data pribadi, meskipun sentimen negatif mendominasi[5].

Penelitian oleh Fannisa Salsabila Pratiwi dkk. (2025) ini mengangkat tantangan ketidakseimbangan kelas dalam model prediksi churn pelanggan pada industri perbankan. Hasil eksperimen menunjukkan bahwa meskipun akurasi tertinggi (87,13%) diperoleh tanpa menggunakan teknik oversampling dengan algoritma Random Forest, efektivitasnya dalam mendeteksi pelanggan churn terbatas. Penggunaan SMOTE dan Random Over-sampling meningkatkan kemampuan model dalam mendeteksi churn, meskipun menyebabkan penurunan akurasi, masing-masing menjadi 81,47% dan 86,20%. Algoritma Naive Bayes menunjukkan akurasi terbaik tanpa oversampling (79,20%), meskipun performanya dalam menangani churn tetap suboptimal [6].

Penelitian oleh Azwar Damari dkk. (2023) mengkaji upaya prediksi tanah longsor di Kota Samarinda, daerah yang rentan mengalami longsor setelah banjir. Mereka mengembangkan model prediksi dengan mengombinasikan algoritma Naïve Bayes, teknik penyeimbangan data SMOTE (Synthetic Minority Oversampling Technique), dan metode optimasi Particle Swarm Optimization (PSO). Hasil penelitian menunjukkan bahwa meskipun PSO mampu meningkatkan akurasi model, penggunaan SMOTE justru menurunkan akurasi pada beberapa skenario. Penurunan ini disebabkan oleh perubahan distribusi data akibat data sintesis dari SMOTE yang dapat memperkenalkan noise dan memengaruhi proses seleksi fitur. Temuan ini menekankan bahwa meskipun SMOTE sering digunakan untuk mengatasi ketidakseimbangan data, penerapannya harus disesuaikan dengan karakteristik data dan model yang digunakan. Sebaliknya, PSO dinilai lebih stabil dalam meningkatkan kinerja model prediksi[7].

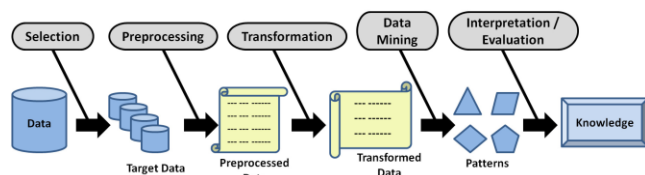
Penelitian I Gusti Ngurah Bagus Putra Adnyana dkk. (2025) yang dilakukan di Telkom University bertujuan untuk memprediksi mahasiswa yang mengundurkan diri, khususnya pada Program Studi S1 Sistem Informasi, sebagai bagian dari upaya menekan nilai Key Performance Indicator (KPI) terkait tingkat DO (Drop Out). Permasalahan utama dalam data adalah ketidakseimbangan kelas, di mana jumlah mahasiswa yang mengundurkan diri jauh lebih sedikit dibandingkan yang tidak. Penelitian ini menggunakan Support Vector Machine (SVM) sebagai algoritma klasifikasi. Model awal tanpa penanganan data imbalance menunjukkan akurasi tinggi sebesar 98,30%, namun memiliki keterbatasan dalam mengenali kelas minoritas (mahasiswa yang mengundurkan diri). Untuk mengatasi hal tersebut, diterapkan teknik SMOTE (Synthetic Minority Oversampling Technique) sebagai metode oversampling. Setelah penerapan SMOTE, akurasi model memang menurun menjadi 92,34%, namun terjadi peningkatan pada nilai recall, precision, dan F1-score, yang menandakan bahwa model menjadi lebih seimbang dan efektif dalam mengenali mahasiswa yang berisiko mengundurkan diri. Penelitian ini menyimpulkan bahwa meskipun penerapan SMOTE menurunkan akurasi secara

keseluruhan, metode tersebut meningkatkan kemampuan model dalam mendeteksi kelas minoritas[8].

Berdasarkan kondisi tersebut, penelitian ini bertujuan untuk mengevaluasi performa model klasifikasi sentimen pada dataset tidak seimbang dalam kondisi aslinya, tanpa manipulasi distribusi pada tahap pelatihan. Fokus utama diarahkan pada kemampuan model dalam mengenali sentimen minoritas dan dampak ketidakseimbangan terhadap metrik evaluasi, khususnya recall kelas positif. Hasil penelitian ini diharapkan memberikan kontribusi empiris terhadap pemahaman performa algoritma machine learning dalam skenario data nyata, serta menjadi pertimbangan bagi penelitian dan penerapan analisis sentimen berbasis data sosial di masa mendatang.

II. METODE

Metode penelitian ini menggunakan pendekatan *Knowledge Discovery in Databases (KDD)* [9].



Gambar 1. Metode Penelitian [9]

Gambar 1 merupakan alur dari metode KDD. Berikut adalah penjabaran langkah-langkah dari metode tersebut:

A. SELECTION

Selection merupakan proses pengumpulan data. Dalam penelitian ini, data dikumpulkan menggunakan *Tweet Harvest*, yaitu alat yang digunakan untuk mengambil *tweet* pada platform X berdasarkan kata kunci dan rentang waktu tertentu. Alat ini dikhususkan untuk keperluan akademik, seperti penelitian.

B. Preprocessing

Setelah data terkumpul kemudian dilakukan *preprocessing* data, yaitu serangkaian langkah untuk menyiapkan kualitas data teks sebelum ke tahap pengujian data mining. *Preprocessing* adalah langkah yang sangat penting dalam data mining, karena data di dunia nyata seringkali tidak lengkap, penuh gangguan (*noise*), dan tidak konsisten[5]. Adapun langkah dalam preprocessing sebagai berikut:

1) Cleansing

Pada tahap *cleansing*, dilakukan pembersihan teks tweet dari elemen-elemen non-informatif seperti *mention*, *hashtag*, tautan, simbol *retweet*, serta karakter non-alfanumerik. Proses ini bertujuan untuk menyederhanakan struktur teks sehingga lebih sesuai untuk tahap analisis lanjutan.

2) Case Folding

Pada tahap *Case folding*, dilakukan untuk menyamakan semua huruf dalam teks menjadi huruf kecil (*lowercase*).

Proses ini menghilangkan pengaruh perbedaan kapitalisasi pada kata-kata. Misalnya, kata "Negatif" dan "negatif" akan dianggap sama sehingga model data mining tidak terjebak dalam perbedaan huruf besar atau kecil.

3) Tokenization

Pada tahap *tokenization*, berfungsi untuk memecah teks menjadi unit-unit kecil yang disebut *token*. Token ini bisa berupa kata atau bahkan frasa yang lebih bermakna. Proses tokenisasi ini sangat penting karena memungkinkan kita untuk fokus pada bagian-bagian penting dari teks, seperti kata-kata kunci yang menunjukkan sentimen.

4) Stopwords Removal

Stopwords adalah kata-kata yang sering muncul dalam teks namun tidak memberikan banyak informasi, seperti "yang", "dan", "di", dan sebagainya. Penghapusan kata-kata ini membantu model data mining fokus hanya pada kata-kata yang relevan dan lebih berpengaruh terhadap analisis sentimen. Dalam dataset kecil, menghapus *stopwords* penting agar model data mining tidak terpengaruh oleh kata-kata yang tidak mendukung sentimen.

5) Stemming

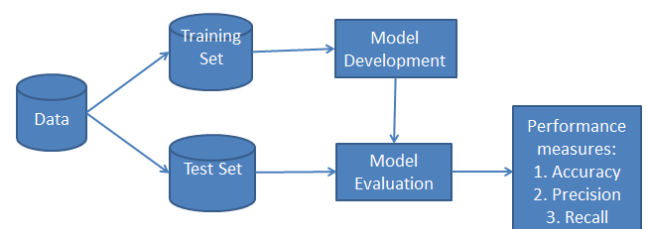
Stemming, yaitu proses mengubah kata-kata turunan menjadi bentuk dasar atau akar kata. Misalnya, kata-kata seperti "berbicara", "bicara", atau "pembicara" akan disamakan menjadi bentuk dasar "bicara". Dengan melakukan *stemming*, kita mengurangi jumlah variasi kata yang berbeda namun memiliki arti yang sama, sehingga meningkatkan kualitas analisis.

C. Transformation

Setelah data melewati tahap *Preprocessing*, tahap selanjutnya ialah tahap *Transformation*. Tahap ini dilakukan menggunakan metode pembobotan kata *Term Frequency-Inverse Document Frequency (TF-IDF)* yang digunakan untuk mengubah data teks menjadi bentuk numerik yang bisa dibaca oleh model data mining. Tujuan penggunaan metode ini adalah untuk menyoroti kata-kata yang penting dalam dokumen, sekaligus mengurangi pengaruh kata-kata umum yang muncul terlalu sering.

D. Data Mining

Proses data mining memainkan peran penting dalam mendukung pengambilan keputusan berbasis data dan telah menjadi bagian integral dari berbagai sektor, mulai dari e-commerce, pendidikan, hingga media sosial [6].



Gambar 2. Langkah-langkah Data Mining [18]

Gambar 2 merupakan visualisasi alur data mining. Proses data mining diawali dengan metode pembagian data train-test. Pembagian dilakukan secara acak dengan parameter `random_state`. Evaluasi model data mining dilakukan berdasarkan tiga metrik utama, yaitu akurasi, presisi, dan recall, untuk mengukur sejauh mana model data mining dapat mengklasifikasikan data baru, khususnya pada kondisi distribusi kelas yang tidak seimbang. Dalam studi ini, 5 algoritma machine learning yang umum digunakan sebagai model data mining diterapkan untuk klasifikasi sentimen, yakni Naïve Bayes, Logistic Regression, Support Vector Machine, Decision Tree, dan Random Forest.

E. Evaluation

Pada tahap ini menggunakan confusion matrix sebagai evaluasi model-model data mining.

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 3. Confusion Matrix

Gambar 3 merupakan confusion matrix, yang menyajikan tabulasi hasil prediksi dari model. Berikut keterangan yang ada pada confusion matrix:

- 1) True Positive (TP): Model memprediksi positif dan kenyataannya memang positif.
- 2) False Positive (FP): Model memprediksi positif, padahal kenyataannya negatif.
- 3) True Negative (TN): Model memprediksi negatif dan kenyataannya memang negatif.
- 4) False Negative (FN): Model memprediksi negatif, padahal kenyataannya positif.

III. HASIL DAN PEMBAHASAN

A. Selection

Pengambilan data pada platform X difokuskan pada tweet berbahasa Indonesia yang dipublikasikan antara 16 Maret 2025 hingga 27 Maret 2025, dengan menggunakan kata kunci "revisi uu tni" dan "ancaman siber". Sebanyak 1.000 tweet berhasil dikumpulkan. Atribut yang tidak relevan, seperti kolom waktu unggahan (timestamp), dihilangkan agar fokus analisis hanya tertuju pada konten teks tweet. Hasil dataset yang diperoleh ditunjukkan pada table 1.

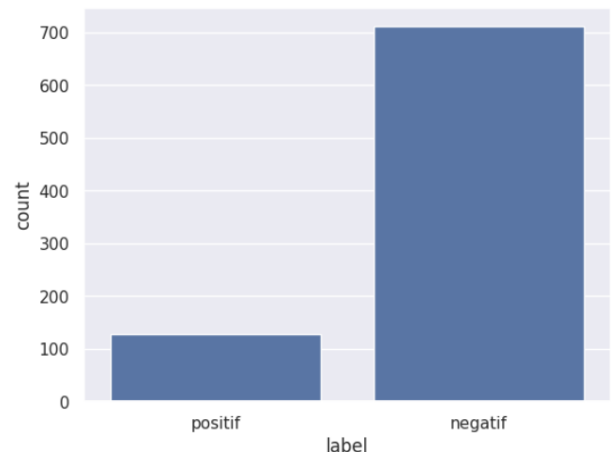
TABEL 1
HASIL SELECTION DATA

No	Full_text
1	Tapi ada sisi positif. Pemerintah bilang ini bantu keamanan apalagi ancaman siber lagi hot. Kalau sukses stabilitas bisa tarik investor tech. Cuma harus diawasi ketat!

2	@kamto_adi Ancaman siber itu karet pak. Dan sejauh ini aspirasi dan protes terhadap ketidakadilan sering dianggap sebagai ancaman. Kemarin bapak @prabowo sendiri yg berstatmen bahwa yg menyampaikan aspirasi dan kritik dianggap pengganggu.
---	--

Kemudian setelah diberi label sentimen, data menunjukkan persebaran kelas yang tidak seimbang.

<Axes: xlabel='label', ylabel='count'>



Gambar 4. Data hasil labelling

Berdasarkan gambar 4, dataset yang digunakan bersifat imbalanced, dengan proporsi data negatif sebanyak 84.74% dan data positif sebesar 15.26%, atau rasio kelas sekitar 85:15, menunjukkan kelas negatif lebih dominan.

B. Preprocessing

Hasil preprocessing data ditunjukkan pada table 2:

TABEL 2
HASIL PREPROCESSING DATA

Proses	Hasil
<i>Cleansing</i>	Ancaman siber itu karet pak. Dan sejauh ini aspirasi dan protes terhadap ketidakadilan sering dianggap sebagai ancaman. Kemarin bapak sendiri yg berstatmen bahwa yg menyampaikan aspirasi dan kritik dianggap pengganggu
<i>Case folding</i>	ancaman siber itu karet pak dan sejauh ini aspirasi dan protes terhadap ketidakadilan sering dianggap sebagai ancaman kemarin bapak sendiri yg berstatmen bahwa yg menyampaikan aspirasi dan kritik dianggap pengganggu
<i>Tokenization</i>	['ancaman', 'siber', 'itu', 'karet', 'pak', 'dan', 'sejauh', 'ini', 'aspirasi', 'dan', 'protes', 'terhadap', 'ketidakadilan', 'sering', 'dianggap', 'sebagai', 'ancaman', 'kemarin', 'bapak', 'sendiri', 'yg', 'berstatmen', 'bahwa', 'yg', 'menyampaikan', 'aspirasi', 'dan', 'kritik', 'dianggap', 'pengganggu']
<i>Stopwords Removal</i>	['ancaman', 'siber', 'karet', 'aspirasi', 'protes', 'ketidakadilan', 'dianggap', 'ancaman',

	'kemarin', 'yg', 'berstatmen', 'yg', 'aspirasi', 'kritik', 'dianggap', 'pengganggu']
<i>Stemming</i>	ancam siber karet aspirasi protes ketidakadilan anggap ancam kemarin yg berstatmen yg aspirasi kritik anggap ganggu

Pada tahap ini *Stemming* adalah hasil akhir data yang kemudian akan dipakai pada tahap *transformation*. Data setelah preprocessing tersisa 839 data dari yang awalnya 1000 data.

C. Transformation

Contoh Hasil TF-IDF (5 dokumen pertama, 10 fitur pertama):

	100m	2025	adaptasi	adaptif	aja	aktif	akun	alam	alat	aman
0	0.0	0.0	0.0	0.0	0.000	0.0	0.0	0.0	0.0	0.249
1	0.0	0.0	0.0	0.0	0.000	0.0	0.0	0.0	0.0	0.000
2	0.0	0.0	0.0	0.0	0.000	0.0	0.0	0.0	0.0	0.000
3	0.0	0.0	0.0	0.0	0.000	0.0	0.0	0.0	0.0	0.000
4	0.0	0.0	0.0	0.0	0.329	0.0	0.0	0.0	0.0	0.000

Gambar 5. Hasil TF-IDF

Gambar 5 menunjukkan hasil perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF) dari data teks, yang diambil dari 5 dokumen pertama dan hanya 10 fitur (*kata*) pertama. Setiap baris mewakili satu dokumen, sedangkan setiap kolom adalah skor TF-IDF untuk sebuah kata spesifik. Pada dokumen ke-0, kata "aman" memiliki skor TF-IDF sebesar 0.249, menunjukkan bahwa kata ini cukup relevan di dokumen itu, sementara kata-kata lain seperti "aktif", "alat", atau "akun" bernilai 0, artinya tidak muncul atau tidak penting di dokumen tersebut. Dengan memberikan bobot lebih tinggi pada kata-kata yang khas pada dokumen tertentu namun jarang di seluruh korpus, TF-IDF membantu model mengenali kata-kata yang paling berkontribusi dalam menentukan kelas (positif atau negatif). Hal ini sangat penting, terutama dalam konteks dataset yang tidak seimbang, karena TF-IDF dapat membantu menyoroti kata-kata minoritas yang relevan dan meningkatkan kemampuan model dalam mengenali sentimen positif yang jumlahnya lebih sedikit.

D. Data Mining

Pada skenario pertama, dilakukan pengujian terhadap lima model klasifikasi data mining—Naïve Bayes, Logistic Regression, Support Vector Machine, Decision Tree, dan Random Forest—menggunakan dataset yang tidak seimbang (dominan negatif). Pengujian model data mining dilakukan pada 839 tweet setelah dilakukan preprocessing data, dengan pembagian data train-test (80-20). Didapatkan hasil uji model data mining sebagaimana tercantum dalam Tabel 3.

Berdasarkan dari hasil table 3 tersebut, dapat disimpulkan bahwa:

- 1) Naïve Bayes menonjol sebagai model dengan performa terbaik pada penelitian ini. Untuk kelas positif, dengan F1-score mencapai 0.84, didukung oleh precision 0.88 dan recall 0.81. Hal ini menunjukkan keseimbangan model dalam mengenali kasus positif dengan kesalahan minimal. Karakteristik probabilistik dari Naïve Bayes tampaknya bekerja cukup efektif pada teks pendek seperti tweet, meskipun data tidak seimbang. Meskipun

algoritma ini biasanya sensitif terhadap data tidak seimbang, TF-IDF dan distribusi fitur yang mendukung independensi bersyarat membantu mengurangi pengaruh ketidakseimbangan.

TABEL 3
HASIL UJI MODEL DATA MINING

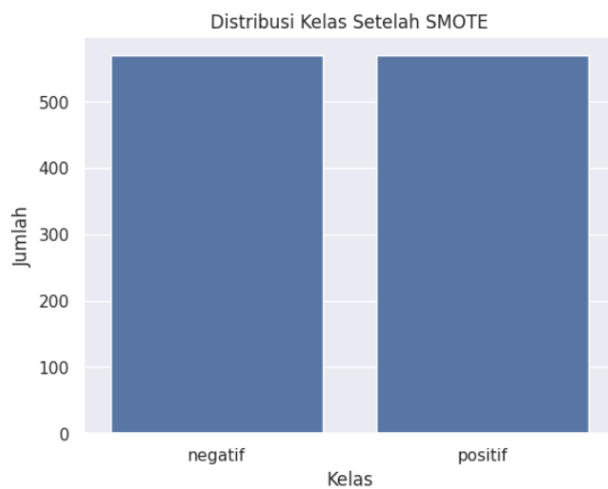
Algoritma	Akurasi	Precision	Recall	F1-score	Macro avg F1	Weighted avg
Kelas Positif						
Naïve B.	0.95	0.88	0.81	0.84	0.91	0.95
Logistic R.	0.94	0.86	0.73	0.79	0.88	0.94
SVM	0.94	0.86	0.73	0.79	0.88	0.94
Decision T.	0.93	0.78	0.81	0.79	0.88	0.94
Random F.	0.95	0.90	0.73	0.81	0.89	0.94
Kelas Negatif						
Naïve B.	0.95	0.97	0.98	0.97	0.91	0.95
Logistic R.	0.94	0.95	0.98	0.97	0.88	0.94
SVM	0.94	0.95	0.98	0.97	0.88	0.94
Decision T.	0.93	0.96	0.96	0.96	0.88	0.94
Random F.	0.95	0.95	0.99	0.97	0.89	0.94

- 2) Random Forest juga menunjukkan kinerja solid, terutama dari sisi precision (0.90), menandakan bahwa prediksi positifnya sangat akurat. Namun, nilai recall-nya lebih rendah (0.73), yang berarti masih terdapat sejumlah kasus positif yang gagal terdeteksi.
- 3) Logistic Regression dan SVM memperlihatkan performa yang identik, dengan F1-score 0.79. Kedua algoritma ini memiliki precision tinggi (0.86) namun recall lebih rendah (0.73), menandakan bahwa model cenderung lebih hati-hati dalam memprediksi kelas positif—cocok untuk menghindari false positive namun berisiko terhadap miss-classification kelas minoritas.
- 4) Sementara itu, Decision Tree menunjukkan performa terlemah, dengan precision hanya 0.78. Meskipun recall-nya cukup tinggi (0.81), F1-score tetap berada di angka 0.79. Model ini cenderung lebih permisif dalam memprediksi kelas positif, namun berisiko menghasilkan lebih banyak false positive.

Hasil ini mengindikasikan bahwa pada dataset tidak seimbang, performa keseluruhan model cenderung dipengaruhi oleh dominasi kelas mayoritas (negatif). Meskipun akurasi secara umum terlihat tinggi, metrik seperti F1-score kelas positif menjadi indikator penting untuk mengukur kemampuan model dalam menangani ketidakseimbangan data dengan mayoritas negatif. Secara keseluruhan, Naïve Bayes dan Random Forest menjadi kandidat yang cukup efektif pada kondisi imbalanced, masing-masing menawarkan trade-off antara keseimbangan recall dan precision.

Kemudian dilakukan perbandingan lebih lanjut pada skenario balanced dataset menggunakan teknik SMOTE (Synthetic Minority Over-sampling Technique). Penerapan SMOTE bertujuan untuk meningkatkan representasi kelas minoritas (positif) dengan cara menambahkan sampel sintetis,

sehingga model dapat dilatih pada dataset yang lebih seimbang.



Gambar 5. Hasil SMOTE

Gambar 5 ditunjukkan hasil SMOTE yaitu persebaran data yang seimbang antara label negatif dan positif yang masing-masing berjumlah 569 data. Kemudian dilakukan pengujian model data mining kembali dengan pembagian train-test (80-20). Hasil pengujian model data mining dengan SMOTE ditunjukkan pada table 4:

TABEL 4
HASIL UJI MODEL DATA MINING DENGAN SMOTE

Algoritma	Akurasi	Precision	Recall	F1-score	Macro avg F1	Weighted avg
Kelas Positif						
Naïve B.	0.93	0.77	0.77	0.77	0.86	0.93
Logistic R.	0.94	0.83	0.77	0.80	0.88	0.94
SVM	0.95	0.90	0.73	0.81	0.89	0.94
Decision T.	0.93	0.79	0.73	0.76	0.86	0.93
Random F.	0.96	0.95	0.77	0.85	0.91	0.96
Kelas Negatif						
Naïve B.	0.93	0.96	0.96	0.96	0.86	0.93
Logistic R.	0.94	0.96	0.97	0.97	0.88	0.94
SVM	0.95	0.99	0.97	0.97	0.89	0.94
Decision T.	0.93	0.95	0.96	0.96	0.86	0.93
Random F.	0.96	0.96	0.99	0.98	0.91	0.96

Dari hasil tersebut, dapat disimpulkan bahwa:

- 1) Secara umum, nilai F1-score untuk kelas positif tidak mengalami peningkatan yang signifikan. Bahkan, pada beberapa model seperti Naïve Bayes, performa justru mengalami penurunan dari F1-score 0.84 menjadi 0.77. Hal ini mengindikasikan bahwa teknik balancing seperti SMOTE tidak efektif pada kasus ini—terutama bagi model yang secara alami sensitif terhadap perubahan distribusi data seperti Naïve Bayes.
- 2) Sebaliknya, beberapa model lain mengalami perbaikan minor. Random Forest, misalnya, mengalami peningkatan F1-score kelas positif dari 0.81 menjadi 0.85, dan precision meningkat dari 0.90 menjadi 0.95,

dengan tetap mempertahankan recall yang lebih stabil (0.73 ke 0.77). Hal ini menunjukkan bahwa Random Forest bekerja lebih stabil pada kondisi ini.

- 3) Logistic Regression juga mengalami sedikit peningkatan recall kelas positif (dari 0.73 ke 0.77) namun tanpa perubahan signifikan pada F1-score (dari 0.79 ke 0.80). Model ini tetap stabil, meskipun peningkatan setelah SMOTE tidak terlalu besar.
- 4) SVM mencatat peningkatan F1-score dari 0.79 menjadi 0.81, terutama karena peningkatan precision yang cukup tajam (0.86 ke 0.90), meskipun recall (0.73 tetap). Sementara Decision Tree tidak memperlihatkan perubahan berarti, dengan F1-score yang justru sedikit menurun dari 0.79 ke 0.76, mengindikasikan bahwa model ini tidak terlalu terbantu oleh SMOTE pada kasus ini.

Jika dibandingkan dengan hasil sebelumnya (tanpa SMOTE), perubahan performa secara umum tergolong kecil dan tidak konsisten di seluruh model. Pada beberapa algoritma seperti Random Forest dan SVM, penerapan SMOTE memberikan peningkatan yang cukup baik dalam mengenali atau memprediksi kelas positif secara lebih akurat. Namun pada algoritma lain, terutama Naïve Bayes dan Decision Tree, SMOTE menyebabkan penurunan performa. Pada kasus ini mengonfirmasi bahwa balancing dengan SMOTE dan efektivitasnya sangat bergantung pada karakteristik masing-masing algoritma serta bagaimana algoritma tersebut menangani data sintetik. Di sisi lain, perubahan skor pada kelas negatif juga relatif stabil, menandakan bahwa SMOTE tidak berdampak besar terhadap kelas mayoritas yang sudah dominan sejak awal.

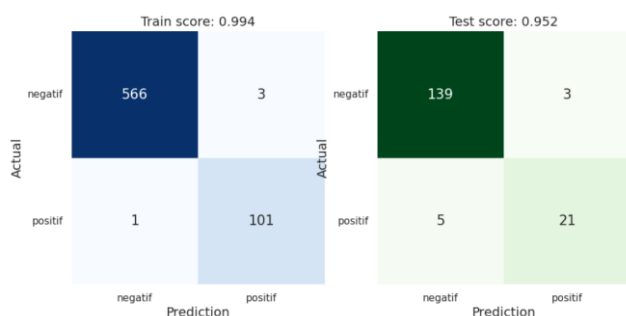
E. Evaluation

Pada tahap ini, dilakukan evaluasi pada ke-5 algoritma machine learning setelah dilakukan pengujian model tanpa menggunakan data SMOTE.

1) Naïve bayes

Gambar 6 merupakan hasil evaluasi algoritma Naïve Bayes. Meskipun akurasi secara keseluruhan cukup tinggi, hasil lebih mendalam pada confusion matrix menunjukkan adanya tantangan signifikan dalam mengidentifikasi kelas positif. Pada data pelatihan, model menghasilkan 566 true negative, 101 true positive, 3 false positive, dan 1 false negative.

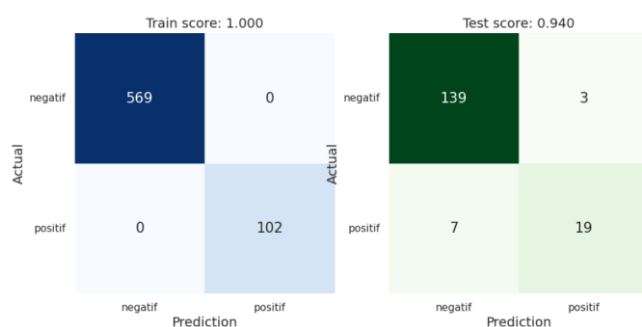
Sementara itu, pada data pengujian, model mencatat 139 true negative, 21 true positive, 3 false positive, dan 5 false negative. Naïve Bayes menunjukkan kinerja terbaik dalam mengenali kelas positif dengan keseimbangan antara recall dan precision.



Gambar 6. Evaluasi Naïve Bayes

2) Logistic Regression

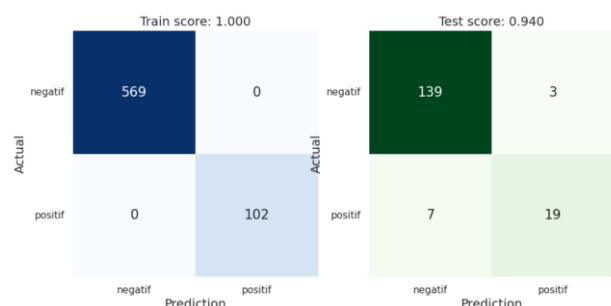
Berikut adalah hasil pengujian algoritma Logistic Regression. Gambar 7 merupakan hasil evaluasi algoritma Logistic Regression. Pada data pelatihan, model berhasil mengklasifikasikan 569 true negative, 102 true positive, 0 false positive, dan 0 false negative.



Gambar 7. Evaluasi Logistic Regression

Sementara pada data pengujian, model mencatat 139 true negative, 19 true positive, 3 false positive, dan 7 false negative. Sehingga model ini menghasilkan tingkat performa menengah dengan kecenderungan kehilangan lebih banyak data positif (false negative).

3) Support Vector Machine



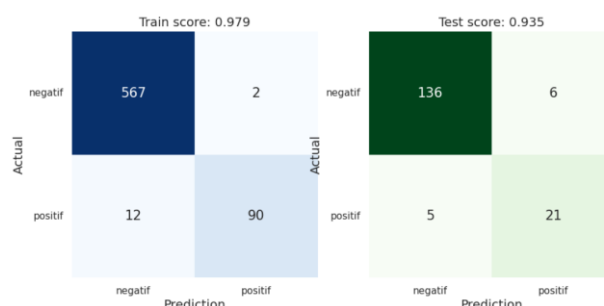
Gambar 8. Evaluasi Support Vector Machine

Gambar 8 merupakan hasil evaluasi algoritma Support Vector Machine. Pada data pelatihan, model berhasil mengklasifikasikan 569 true negative, 102 true positive, 0 false positive, dan 0 false negative. Sementara pada data pengujian, model mencatat 139 true negative, 19 true positive, 3 false positive, dan 7 false negative. Model ini juga berada

pada tingkat performa menengah dengan kecenderungan kehilangan lebih banyak data positif (false negative).

4) Decision Tree

Berikut adalah hasil pengujian metode Decision Tree. Gambar 9 merupakan hasil evaluasi algoritma Decision Tree. Pada data pelatihan, model berhasil mengklasifikasikan 567 true negative, 90 true positive, 2 false positive, dan 12 false negative.

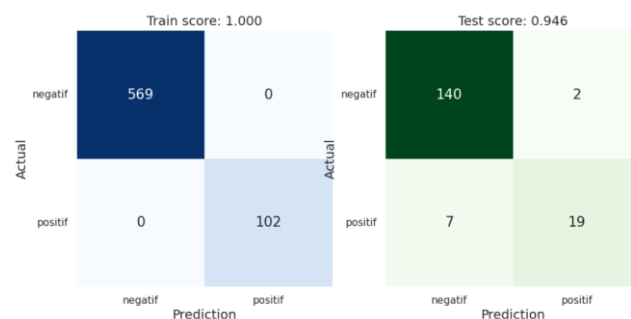


Gambar 9. Evaluasi Decision Tree

Sementara pada data pengujian, model mencatat 136 true negative, 21 true positive, 6 false positive, dan 5 false negative. Dengan demikian Decision Tree cenderung kurang presisi meskipun recall-nya tinggi.

5) Random Forest

Berikut adalah hasil pengujian metode Random Forest:



Gambar 10. Evaluasi Random Forest

Gambar 10 merupakan hasil evaluasi algoritma Random Forest. Pada data pelatihan, model berhasil mengklasifikasikan 569 true negative, 102 true positive, 0 false positive, dan hanya 0 false negative. Sementara pada data pengujian, model mencatat 140 true negative, 19 true positive, 2 false positive, dan 7 false negative. Random Forest berada pada tingkat performa menengah dengan kecenderungan kehilangan lebih banyak data positif (false negative).

IV. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa kelima algoritma data mining yang diuji mampu memberikan performa yang cukup baik dalam menangani data sentimen yang tidak seimbang (dominan negatif). Algoritma Naïve Bayes mencatat F1-score tertinggi sebesar 0.84 untuk kelas positif, disusul oleh Random Forest dengan F1-score sebesar 0.81. Karakteristik probabilistik dari Naïve Bayes bekerja cukup efektif pada kasus, meskipun algoritma ini biasanya sensitif terhadap data tidak seimbang, TF-IDF dan distribusi fitur yang mendukung independensi bersyarat membantu mengurangi pengaruh ketidakseimbangan.

Sebagai pembanding, penelitian ini juga menerapkan teknik balancing data menggunakan SMOTE. Hasil analisis menunjukkan bahwa peningkatan performa setelah balancing bersifat marginal, dan pada beberapa algoritma seperti Naïve Bayes, performa justru sedikit menurun (0,77). Temuan ini mengindikasikan bahwa efektivitas teknik balancing seperti SMOTE dapat bervariasi tergantung pada jenis algoritma yang digunakan dan karakteristik dataset dikasus ini.

Dengan demikian, hasil penelitian ini memberikan kontribusi empiris yang menekankan pentingnya pemilihan pendekatan yang sesuai, serta perlunya pemahaman mendalam terhadap perilaku masing-masing algoritma dalam konteks data tidak seimbang. Peneliti diharapkan mempertimbangkan aspek-aspek tersebut secara cermat dalam merancang eksperimen dan interpretasi hasil.

DAFTAR PUSTAKA

- [1] Chawla, N. V. (2010). Data Mining and Knowledge Discovery Handbook. Data Mining and Knowledge Discovery Handbook, 2003–2004. <https://doi.org/10.1007/978-0-387-09823-4>
- [2] Liawati, A., Narasati, R., Solihudin, D., Lukman Rohmat, C., & Eka Permana, S. (2024). Analisis Sentimen Komentar Politik Di Media Sosial X Dengan Pendekatan Deep Learning. JATI (Jurnal Mahasiswa Teknik Informatika), 7(6), 3557–3563.
- [3] Sun, H., Li, J., & Zhu, X. (2025). A novel expandable borderline SMOTE over-sampling method for class imbalance problem. IEEE Transactions on Knowledge and Data Engineering, 37(5), 2183–2199.
- [4] Pebrianti, R. D., & Karawang, U. S. (2025). Analisis Sentimen Masyarakat Platform X. 13(2).
- [5] Saputra, A. (2025). Analisis Sentimen Pengguna X Terhadap Kebocoran Data Pribadi Menggunakan Algoritma Naïve Bayes Classifier. 14(01), 32–40.
- [6] Pratiwi, F. S., Barata, M. A., Ardianti, A. D., Studi, P., Infomatika, T., Nahdlatul, U., Sunan, U., Studi, P., Mesin, T., Nahdlatul, U., & Sunan, U. (2025). Implementasi Metode Smote Dan Random Over- Sampling Pada Algoritma Machine Learning Untuk. 8(1), 87–98.
- [7] Damari, A., Azhima, T., Siswa, Y., Pranoto, W. J., Muhammadiyah, U., Timur, K., & Timur, K. (2025). Implementation Of The Pso-Smote Method On The Naive Bayes Algorithm To Address Class Imbalance In Landslide Disaster Data Penerapan Metode Pso-Smote Pada Algoritma Naive Bayes Untuk Mengatasi Class Imbalance. 10(1), 332–343.
- [8] Gusti, I., Bagus, N., & Adnyana, P. (2025). Prediksi Mahasiswa Mengundurkan Diri Menggunakan Metode Support Vector Machine. 12(1), 1793–1800.
- [9] Gullo, F. (2015). From patterns in data to knowledge discovery: What data mining can do. Physics Procedia, 62, 18–22. <https://doi.org/10.1016/j.phpro.2015.02.005>
- [10] Kotsiantis, S. B., Kanellopoulos, D., & Pintelas, P. E. (2006). Data Preprocessing for Supervised Learning. International Journal of Computer Science, 1(2), 111–117.
- [11] Kaur, H., & Chhabra, R. (2021). Data mining and its applications in various sectors: A comprehensive review. Materials Today: Proceedings, 47, 3435–3439.
- [12] Saputra, M. R. (2025). Analisis sentimen twitter terhadap konflik di papua menggunakan perbandingan naive bayes dan svm. 10(2), 1197–1208.
- [13] Shidiq, M. F. A., & Alita, D. (2025). Kasus Judi Online Menggunakan Data Dari Media Sosial X Pendekatan Naive Bayes Dan Svm. 8(1), 24–35.
- [14] No, V., Hal, J., Kumiasih, U., & Teguh, A. (2025). Analisis Sentimen Masyarakat Terhadap Isu Migrasi Rohingya Ke Indonesia. 7(1), 199–207.
- [15] Jhosefhin, N. V. R., & Dewi, C. (2025). Analisis Sentimen Crawling Data dari Sosial Media X tentang Gaza Menggunakan Metode SVM dan Decision Tree. 6(1), 427–437.
- [16] Fakultas Hukum UMSU. (2025, Februari 7). Isi dan makna RUU TNI terbaru yang sudah direvisi 2025. Universitas Muhammadiyah Sumatera Utara. <https://fahum.umsu.ac.id/berita/isi-dan-makna-ruu-tni-terbaru-yang-sudah-direvisi-2025/>
- [17] Tempo.co. (2019, Juli 12). Kontroversi pelibatan tentara hadapi ancaman siber dalam UU TNI. Tempo. <https://www.tempo.co/politik/kontroversi-pelibatan-tentara-hadapi-ancaman-siber-dalam-uu-tni-1224841>
- [18] Awan, A. A. (2023, Maret 3). Naive Bayes Classifier Tutorial: with Python Scikit-learn. DataCamp. <https://www.datacamp.com/tutorial/naive-bayes-scikit-learn>
- [19] Adib, K., Handayani, M. R., Yuniarti, W. D., & Umam, K. (2024). Opini Publik Pasca-Pemilihan Presiden: Eksplorasi Analisis Sentimen Media Sosial X Menggunakan SVM. SINTECH (Science and Information Technology) Journal, 7(2), 80–91. <https://doi.org/10.31598/sintechjournal.v7i2.1581>
- [20] Aufan, M. H., Handayani, M. R., Nurjanna, A. B., & Hendro, N. C. (2023). The Perceptions Of Semarang Five Star Hotel Tourists With Support Vector Machine On Google Reviews Persepsi Wisatawan Hotel Bintang Lima Semarang Dengan. x(December), 1–8.
- [21] No, V., & A, F. Y. (2025). Edumatic : Jurnal Pendidikan Informatika Optimasi Klasifikasi Sentimen Ulasan Game Berbahasa Indonesia : IndoBERT dan SMOTE untuk Menangani Ketidakseimbangan Kelas. 9(1), 256–265. <https://doi.org/10.29408/edumatic.v9i1.29666>
- [22] Hawari, F. A., Sholihati, I. D., Informasi, S., Nasional, U., Siswa, K., & Neighbors, K. (2025). Perbandingan Metode Naïve Bayes Dan K-Nearest Neighbors Dalam Klasifikasi Kepuasan Mahasiswa Terhadap Layanan Wi-Fi Di Universitas Nasional. 9(3), 5203–5208.
- [23] Dwilestari, G., & Afifah, T. A. (2025). Perbandingan Kinerja Algoritma Naive Bayes Dan Decision Tree Dalam Klasifikasi Kanker Paru-Paru. 9(1), 801–807.
- [24] Baja, P. R., & Ani, A. S. (2020). Jurnal Comasie. Comasie, 6(2), 107–118.
- [25] Fauzan, R., Vitianingsih, A. V., & Cahyono, D. (2025). Application of Classification Algorithms in Machine Learning for Phishing Detection Penerapan Algoritma Klasifikasi pada Machine Learning untuk Deteksi Phishing. 5(April), 531–540.