Jurnal Komtika (Komputasi dan Informatika),
Vol. 8 No. 2 | Nov 2024

Klasifikasi Sentimen Terhadap Kebijakan Tapera Menggunakan Komparasi Machine Learning dan SMOTE

Henny Leidiyana[1], Titik Misriati[2*], Riska Aryanti[3]
[1,2,3]Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika
*email: titik.tmi@bsi.ac.id

DOI: https://doi.org/10.31603/komtika.v8i2.12595

Received: 01-11-2024, Revised: 14-11-2024, Accepted: 15-11-2024

ABSTRACT

The Indonesian government's Public Housing Savings Program (Tapera) aims to help low- and middle-income persons get housing financing. Although the initiative strives to satisfy housing requirements, the public has responded in a variety of ways, as evidenced by social media posts. The goal of this study is to examine public sentiment towards the Tapera policy using YouTube comment data to provide an overview of popular perspective. This study combines sentiment analysis with machine learning algorithms, including K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Multinomial Naïve Bayes (NB), and Decision Tree. The data is divided into three scenarios, namely 70% training data and 30% test data, 80% training data and 20% test data, and 90% training data and 10% test data. Data balancing is also performed with SMOTE. The performance evaluation is based on each algorithm's accuracy, precision, recall, and F1 Score values. The results showed that the SVM algorithm performed the best in all circumstances, with the greatest accuracy of 88% and an F1 score of 85%. The multinomial Naïve Bayes algorithm ranked second with steady accuracy, whereas KNN and Decision Tree had poorer performance. This suggests that SVM is the most effective method for processing sentiment data regarding Tapera policy.
Keywords: *sentiment analysis, classification, machine learning, smote, tapera.*

ABSTRAK

Program Tabungan Perumahan Rakyat (Tapera) merupakan inisiatif pemerintah Indonesia untuk membantu masyarakat berpenghasilan rendah dan menengah dalam mengakses pembiayaan perumahan. Meskipun bertujuan untuk memenuhi kebutuhan perumahan, program ini mendapatkan berbagai tanggapan dari masyarakat yang dapat dilihat melalui media sosial. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kebijakan Tapera berdasarkan data komentar di platform YouTube, sehingga dapat memberikan gambaran tentang persepsi publik. Penelitian ini menggunakan metode analisis sentimen dengan beberapa algoritma pembelajaran mesin, yaitu K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Multinomial Naïve Bayes (NB), dan Decision Tree. Data dibagi ke dalam tiga skenario, yaitu 70% data latih dan 30% data uji, 80% data latih dan 20% data uji, serta 90% data latih dan 10% data uji. Selain itu, dilakukan juga penyeimbangan data dengan SMOTE. Proses evaluasi kinerja dilakukan berdasarkan nilai akurasi, precision, recall, dan F1 Score dari masing-masing algoritma. Hasil penelitian menunjukkan bahwa algoritma SVM memberikan performa terbaik dalam semua skenario dengan akurasi tertinggi mencapai 88% dan F1 Score 85%. Algoritma Multinomial Naïve Bayes berada di posisi kedua dengan akurasi yang cukup stabil, sedangkan KNN dan Decision Tree menunjukkan kinerja yang lebih rendah. Hal ini mengindikasikan bahwa SVM adalah metode yang paling efektif untuk mengolah data sentimen terkait kebijakan Tapera.
Kata kunci: analisis sentimen, klasifikasi, machine learning, smote, tapera

PENDAHULUAN

Tabungan Perumahan Rakyat (Tapera) adalah program yang dibuat oleh pemerintah Indonesia untuk pemecahan masalah kebutuhan perumahan masyarakat. Program ini memberikan fasilitas bagi masyarakat agar bisa menabung untuk memperoleh kemudahan

dalam pembiayaan perumahan. Program ini ditujukan khususnya untuk kelompok masyarakat yang memiliki penghasilan rendah dan menengah. Badan Pengelola Tapera dibentuk sesuai dengan Undang-Undang Nomor 4 Tahun 2016 tentang Tabungan Perumahan Rakyat, dan kemudian Pemerintah mengeluarkan Peraturan Pemerintah Nomor 25 Tahun 2020 tentang Penyelenggaraan Tabungan Perumahan Rakyat yang berlaku efektif pada 20 Mei 2020 [1]. Namun hal ini menimbulkan berbagai reaksi di kalangan masyarakat. Reaksi tersebut dapat dilihat pada berbagai *platform*, seperti media sosial, berita daring, dan lain-lain. Pemerintah dapat menggunakan analisis sentimen untuk menilai reaksi masyarakat terhadap kebijakan Tapera. Hal ini dapat memberikan informasi dalam pengambilan keputusan dan membantu menyesuaikan kebijakan agar lebih selaras dengan kebutuhan publik.

Saat ini dengan melimpahnya informasi secara tidak langsung mempelajari banyak hal dari berbagai platform media sosial meskipun tidak semua masyarakat mengerti di banyak bidang. Misalkan dalam bidang politik, bagi masyarakat yang tertarik ilmu politik. Media sosial kini dijadikan *tools* untuk memperoleh berita politik. [2]. Saat ini, internet dipenuhi informasi dalam berbagai bentuk seperti situs web, media sosial, forum diskusi, dan blog. Informasi teks merupakan data tidak terstruktur sehingga dikonversi kedalam data terstruktur melalui analisis sentimen. Topik penelitian tentang analisis sentimen berdasarkan data yang diperoleh dari media sosial sudah banyak dilakukan di berbagai bidang seperti review produk [3], *fashion* [4], kesehatan [5], pendidikan [6], politik [7], kebijakan pemerintah [8], keuangan [9][10], sumber daya manusia [11], game [12], dan lain-lain.

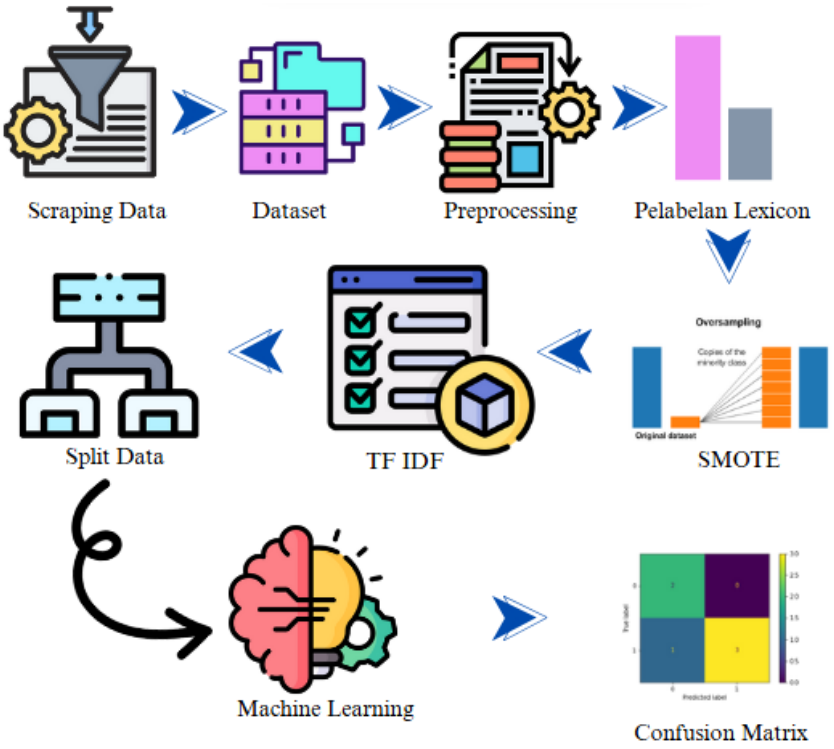
Penelitian yang memanfaatkan data yang bersumber dari media sosial juga banyak dilakukan seperti aplikasi Twitter [13], Facebook [14], Youtube [15], Instagram [16] menggunakan metode pembelajaran mesin untuk tugas analisa sentimen. Penelitian-penelitian tersebut menggunakan algoritma pembelajaran mesin dan membandingkan beberapa algoritma. Seperti penelitian yang dilakukan oleh Aravind Veluchamy yang melakukan perbandingan penerapan metode pembelajaran mesin Logistic Regression, Support Vector Machine, dan Gradient Boosting dengan penerapan metode berbasis Lexicon Valence Aware Dictionary and Sentiment Reasoner (VADER), *Pattern,* dan SentiWordNe*t* dalam tugas analisa sentimen untuk *review* produk Amazon dimana hasil tertinggi diperoleh pada model Logistic Regression dengan akurasi 90% [17]. Penelitian lain yang melakukan perbandingan penerapan metode pembelajaran mesin Logistic Regression, Multinomial Naïve Bayes (MNB), Support Vector Classifier (SVC), Extreme Gradient Boosting Classifier (XGB), Multi-layer Perceptron Classifier (MLP) dalam tugas analisa sentimen untuk *review game* yang diperoleh dari Amazon dan Twitter. Penelitian ini juga menggunakan metode SMOTE (Synthetic Minority Over-sampling Technique) karena data berisi jumlah ulasan positif yang jauh lebih besar daripada ulasan netral dan negatif. Hasilnya secara signifikan meningkatkan performa sebagian besar model dan penggunaan TF-IDF (Term Frequency-Inverse Document Frequency) menghasilkan kinerja yang lebih baik. SVC menjadi model dengan performa terbaik di antara kelima model yang dipakai [18]. Penelitian yang juga melakukan penerapan TF-IDF dan SMOTE yaitu analisis sentimen dengan data berupa ulasan aplikasi sister bagi mahasiswa UNEJ dengan hasil model Random Forest menggunakan SMOTE yang paling unggul [19].

Topik Tapera menjadi tren di media sosial khususnya pada aplikasi X [20] dan menarik perhatian para peneliti tentang topik ini karena membuat heboh masyarakat Indonesia, terbukti dengan adanya penelitian tentang Tapera berdasarkan *mentions* dan *reach* di beragam media

menggunakan *tools* Brand24 [21] dan analisis sentimen komentar mengenai Tapera pada aplikasi X menggunakan Naïve Bayes [22]. Tujuan penelitian yaitu melakukan analisa sentimen masyarakat terhadap Program Tapera. Data yang digunakan adalah komentar video yang diambil dari platform media sosial YouTube. Dengan serangkaian tahap pembelajaran mesin, diharapkan agar hasilnya dapat memberikan deskripsi tentang persepsi masyarakat terhadap program ini.

METODE

Penelitian ini dilakukan dalam beberapa tahapan yang ditunjukkan pada Gambar 1 dengan rincian sebagai berikut:



Gambar 1. Tahapan Penelitian

1. *Scrapping* Data
   Scrapping data dilakukan untuk mengumpulkan dataset berupa komentar di Youtube yang membahas tentang Tapera.
2. Dataset
   Dataset yang diperoleh berupa 2963 komentar dari Youtube yang membahas tentang kebijakan Tapera. Data ini mencakup berbagai pendapat, reaksi, dan pandangan terhadap kebijakan Tapera baik yang mendukung maupun yang memberikan kritik.
3. *Preprocessing* Data
   Dataset tersebut dilakukan *preprocessing* untuk untuk mempersiapkan data teks sebelum dilakukan analisis. Tahap preprocessing ini menghasilkan 2495 data dari 2963 komentar dengan tahapan sebagai berikut:
   a. *Case Folding*
      Case folding dilakukan dengan cara mengubah semua huruf dalam teks menjadi bentuk huruf kecil (*lowercase*) untuk memastikan konsistensi dalam analisis teks. Tabel 1 menunjukkan data sebelum dan sesudah *case folding*.

Tabel 1. *Case Folding*

Sebelum	Sesudah
Ini tapera malah jadi lahan korupsi berkedok tabungan yg di raup dari potongan gajih	ini tapera malah jadi lahan korupsi berkedok tabungan yg di raup dari potongan gajih
kl menurut aku bukan cuman menyengsarakan bisa2 rakyat mampus buat makan sja gk cukup mau d potong lagi gila namanya	kl menurut aku bukan cuman menyengsarakan bisa2 rakyat mampus buat makan sja gk cukup mau d potong lagi gila namanya
TAPERA ladang korupsi baru ini di tahun 2024 😂😂😂😂	tapera ladang korupsi baru ini di tahun 2024 😂😂😂😂

b. Pembersihan Data

Membersihkan ulasan yang tidak memiliki arti seperti menghapus hashtag dari ulasan yang biasanya digunakan oleh pengguna untuk memberikan tagar, menghapus angka dalam ulasan, menghapus tanda baca dalam ulasan, menghapus emoji dalam ulasan, repetisi (pengulangan kata) dengan melakukan pembatasan jumlah huruf menjadi dua untuk mengembalikan kata ke bentuk awal dan menghindari terjadinya kata ganda yang memiliki arti sama tetapi berbeda penulisan, menghapus spasi ganda dalam ulasan, menormalisasikan kata singkat dan tidak baku menjadi kata baku sesuai dengan KBBI, menghapus kata yang terdiri dari 3 huruf, seperti oh, iya, ini, itu, dll, dan tidak memberikan informasi penting bagi model saat melakukan prediksi. Tabel 2 menunjukkan data sebelum dan sesudah pembersihan data.

Tabel 2. Pembersihan Data

Sebelum	Sesudah
ini tapera malah jadi lahan korupsi berkedok tabungan yg di raup dari potongan gajih	tapera bahkan jadi lahan korupsi berkedok tabungan yang raup dari potongan gajih
kl menurut aku bukan cuman menyengsarakan bisa2 rakyat mampus buat makan sja gk cukup mau d potong lagi gila namanya	kalau menurut bukan cuman menyengsarakan bisa rakyat mampus buat makan tidak cukup potong lagi gila namanya
tapera ladang korupsi baru ini di tahun 2024 😂😂😂😂	tapera ladang korupsi baru tahun

c. *Tokenizing*

Memecah teks menjadi unit-unit yang lebih kecil berupa kata. Tabel 3 menunjukkan data sebelum dan sesudah *tokenizing*.

d. *Filtering*

Menghapus kata yang tidak relevan dalam ulasan. Tabel 4 menunjukkan data sebelum dan sesudah *filtering*.

e. *Stemming*

Menghilangkan imbuhan pada suatu kata menjadi bentuk dasarnya sesuai dengan KBBI. Tabel 5 menunjukkan data sebelum dan sesudah *stemming*.

Tabel 3. *Tokenizing*

Sebelum	Sesudah
tapera bahkan jadi lahan korupsi berkedok tabungan yang raup dari potongan gajih	['tapera', 'bahkan', 'jadi', 'lahan', 'korupsi', 'berkedok', 'tabungan', 'yang', 'raup', 'dari', 'potongan', 'gajih']
kalau menurut bukan cuman menyengsarakan bisa rakyat mampus buat makan tidak cukup potong lagi gila namanya	['kalau', 'menurut', 'bukan', 'cuman', 'menyengsarakan', 'bisa', 'rakyat', 'mampus', 'buat', 'makan', 'tidak', 'cukup', 'potong', 'lagi', 'gila', 'namanya']
tapera ladang korupsi baru tahun	['tapera', 'ladang', 'korupsi', 'baru', 'tahun']

Tabel 4. *Filtering*

Sebelum	Sesudah
['tapera ', 'bahkan', 'jadi', 'lahan', 'korupsi', 'berkedok', 'tabungan', 'yang', 'raup', 'dari', 'potongan', 'gajih']	['tapera', 'lahan', 'korupsi', 'berkedok', 'tabungan', 'raup', 'potongan', 'gajih']
['kalau', 'menurut', 'bukan', 'cuman', 'menyengsarakan', 'bisa', 'rakyat', 'mampus', 'buat', 'makan', 'tidak', 'cukup', 'potong', 'lagi', 'gila', 'namanya']	['cuman', 'menyengsarakan', 'rakyat', 'mampus', 'makan', 'potong', 'gila', 'namanya']
['tapera', 'ladang', 'korupsi', 'baru', 'tahun']	['tapera', 'ladang', 'korupsi']

Tabel 5. *Stemming*

Sebelum	Sesudah
['tapera', 'lahan', 'korupsi', 'berkedok', 'tabungan', 'raup', 'potongan', 'gajih']	tapera lahan korupsi kedok tabung raup potong gajih
['cuman', 'menyengsarakan', 'rakyat', 'mampus', 'makan', 'potong', 'gila', 'namanya']	cuman sengsara rakyat mampus makan potong gila nama
['tapera', 'ladang', 'korupsi']	tapera ladang korupsi

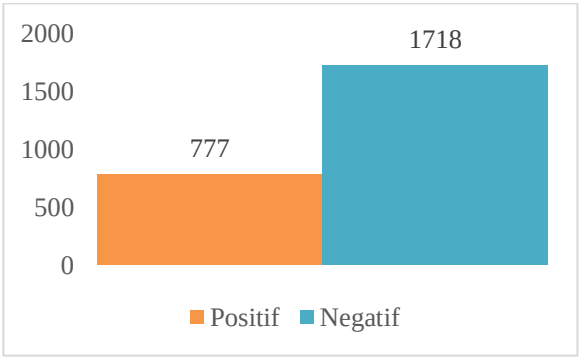
4. Pelabelan Lexicon
   Pelabelan secara otomatis menggunakan lexicon berdasarkan kata-kata atau frasa dalam suatu lexicon sehingga menghasilkan label positif, negatif, dan netral. Label netral akan dihilangkan karena penelitian ini hanya menggunakan dua kategori label, yaitu positif dan negatif
5. SMOTE
   Synthetic Minority Over-sampling Technique (SMOTE) digunakan untuk menangani masalah ketidakseimbangan kelas pada klasifikasi sentimen terhadap kebijakan Tapera yang dapat menyebabkan kinerja model yang buruk, terutama dalam memprediksi kelas minoritas.
6. Pembobotan TF-IDF
   Pembobotan *Term Frequency - Inverse Document Frequency* (TF IDF) untuk mengetahui frekuensi kemunculan setiap kata. TF-IDF adalah metode untuk mengonversi teks menjadi fitur numerik berdasarkan jumlah kemunculan kata dalam dokumen dibandingkan dengan kemunculannya di seluruh dokumen. Teknik ini berguna untuk menentukan kata yang paling sesuai dan mempunyai arti dalam sentimen teks. Dengan TF-IDF, model dapat mengenali kata yang lebih banyak muncul dalam konteks tertentu dan mengabaikan kata-kata yang umum muncul di semua dokumen. Hasil dari proses ini adalah gambaran

numerik dari teks yang dapat digunakan dalam model machine learning untuk mengklasifikasikan sentimen

7. Split Data
   Pembagian data latih dan data uji dilakukan dengan tiga skenario yaitu 1) 70% data latih dan 30% data uji, 2) 80% data latih dan 20% data uji, 3) 90% data latih dan 10% data uji.
8. *Machine Learning*
   Model klasifikasi menggunakan empat model *machine learning*, yaitu K-Nearest Neighbor (KNN), Support Vector Machine (SVM), MultinomialNB, dan Decision Tree. Algoritma ini dilatih menggunakan data latih untuk mengenali pola-pola dalam teks yang dapat digunakan untuk memprediksi sentimen. Pemilihan algoritma yang tepat akan sangat mempengaruhi hasil akhir dari model ini. Model ini dilatih untuk memahami karakteristik data sentimen sehingga dapat mengklasifikasikan teks baru dengan akurasi yang tinggi.
9. *Confusion Matrix*
   Evaluasi model dilakukan dengan *confusion matrix*. Matriks ini menampilkan jumlah prediksi yang benar dan salah yang dilakukan oleh model untuk setiap kategori sentimen. *Confusion matrix* terdiri dari beberapa metrik seperti *True Positive* (TP), *False Positive* (FP), *True Negative* (TN), dan *False Negative* (FN), yang semuanya digunakan untuk menghitung akurasi, presisi, *recall* dan f1 *score* dari model.

HASIL DAN PEMBAHASAN

Data penelitian ini berupa data dari Youtube yang berisi komentar-komentar mengenai diskusi tentang Tapera. Data yang sudah dibersihkan diberikan label otomatis menggunakan lexicon dan menghasilkan dua label yaitu positif dan negatif seperti pada gambar 2.



Gambar 2. Label Sentimen

Frekuensi kata yang sering muncul ditunjukkan pada Tabel 6. Kata yang paling sering muncul adalah "rakyat" dengan 1.088 kali kemunculan, menunjukkan bahwa komentar ini sangat berfokus pada masyarakat atau kepentingan publik yang berkaitan dengan kesejahteraan rakyat atau kebijakan tapera. Kata "rumah" yang muncul 853 kali dan "tapera" yang muncul 811 kali juga sangat sering ditemukan. Ini menunjukkan bahwa komentar tentang perumahan atau program Tapera, sebuah inisiatif pemerintah untuk membantu masyarakat memiliki rumah. Selain itu, kata "uang" dan "gaji" yang muncul masing-masing 546 dan 344 kali, serta kata "tabung" yang muncul 412 kali, menunjukkan adanya pembahasan yang cukup intens tentang masalah keuangan yang berkaitan dengan pengelolaan tabungan, penghasilan, serta bagaimana gaji atau pendapatan dikelola dalam kaitannya dengan program tapera. Kata "korupsi" yang

muncul 488 kali dan "perintah" yang muncul 499 kali mengindikasikan adanya kritik terhadap pemerintahan. Akhirnya, kata "kerja" yang muncul 279 kali menyoroti pekerjaan terkait dengan pengaturan kerja, penghasilan, atau kesejahteraan pekerja.

Tabel 6. Frekuensi Kemunculan Kata

index	words	count
0	rakyat	1088
1	rumah	853
2	tapera	811
3	uang	546
4	perintah	499
5	korupsi	488
6	tabung	412
7	gaji	344
8	potong	302
9	kerja	279

Visualisasi kata yang sering muncul ditampilkan dalam bentuk *word cloud* pada gambar 3 untuk sentimen positif dan Gambar 4 untuk sentimen negatif.

Gambar 3. Word Cloud Sentimen Positif

Gambar 4. Word Cloud Sentimen Negatif

Penelitian dilakukan dengan menggunakan algoritma klasifikasi yang berbeda, yaitu KNN, SVM, MultinomialNB, dan Decision Tree. Pengujian algoritma klasifikasi dilakukan dengan menggunakan tiga skenario pembagian data pelatihan dan data pengujian dengan perbandingan 70%:30%, 80%:20%, dan 90%:10% yang dapat dilihat pada Tabel 7. Pembobotan dilakukan dengan fitur TF-IDF karena fitur ini banyak digunakan untuk klasifikasi teks dan menunjukkan hasil yang lebih baik.

Tabel 7. Pembagian Data

Data Latih	Data Uji	Label		Total
		Positif	Negatif	
70%	30%	548	1198	1746
80%	20%	622	1374	1996
90%	10%	702	1543	2245

Ketidakseimbangan data kelas diatasi dengan menggunakan SMOTE *Oversampling* seperti yang ditampilkan pada Tabel 8.

Tabel 8. Pembagian Data Menggunakan SMOTE

Data Latih	Data Uji	Label		Total
		Positif	Negatif	
70%	30%	1198	1198	2396
80%	20%	1374	1374	2748
90%	10%	1543	1543	3086

Berdasarkan Tabel 8, pembagian data menggunakan metode SMOTE dilakukan dalam tiga scenario yaitu 70% data latih - 30% data uji, 80% data latih - 20% data uji, dan 90% data latih - 10% data uji. Pada setiap skenario, jumlah data positif dan negatif dibuat seimbang. Pada pembagian 70% data latih dan 30% data uji, terdapat 1.198 data positif dan 1.198 data negatif, sehingga total data yang digunakan adalah 2.396. sedangkan untuk skenario 80% data latih dan 20% data uji, terdapat 1.374 data positif dan 1.374 data negatif, dengan total data sebanyak 2.748. Pada pembagian 90% data latih dan 10% data uji, terdapat 1.543 data positif dan 1.543 data negatif, sehingga total data mencapai 3.086. Dengan penggunaan teknik SMOTE, sehingga data positif dan negatif dalam setiap pembagian menjadi seimbang, yang bertujuan untuk meningkatkan performa model dalam mengatasi masalah ketidakseimbangan kelas.

Perbandingan *confusion matrix* pada Tabel 9 dan dengan menggunakan metode SMOTE pada Tabel 10 menunjukkan performa beberapa algoritma klasifikasi pada tiga skenario pembagian data latih dan data uji. Tabel 10 pada pembagian 70% data latih dan 30% data uji, algoritma SVM memiliki akurasi tertinggi yaitu 86% dengan precision 84%, recall 82%, dan F1 Score 83%. Sementara itu, Multinomial NB juga menunjukkan performa yang baik dengan akurasi 82% dan F1 Score 80%.

Tabel 9. Perbandingan *Confusion Matrix*

Data Latih	**Data Uji**	**Algoritma**	**Akurasi (%)**	***Precision* (%)**	***Recall* (%)**	***F1 Score* (%)**
70%	30%	KNN	77	72	71	71
		SVM	86	85	80	82
		Multinomial NB	71	77	53	47
		Decision Tree	76	72	72	72
80%	20%	KNN	75	70	70	71
		SVM	87	86	83	84
		Multinomial NB	71	77	53	47
		Decision Tree	77	73	73	73
90%	10%	KNN	76	71	70	71
		SVM	87	86	82	83
		Multinomial NB	72	86	54	49
		Decision Tree	76	71	71	71

Tabel 10. Perbandingan *Confusion Matrix* Menggunakan SMOTE

Data Latih	**Data Uji**	**Algoritma**	**Akurasi (%)**	***Precision* (%)**	***Recall* (%)**	***F1 Score* (%)**
70%	30%	KNN	52	64	63	52
		SVM	86	84	82	83
		Multinomial NB	82	79	82	80
		Decision Tree	74	70	71	72
80%	20%	KNN	53	65	64	53
		SVM	86	84	83	83
		Multinomial NB	79	76	80	77
		Decision Tree	75	71	72	71
90%	10%	KNN	50	63	62	50
		SVM	88	87	84	85
		Multinomial NB	77	73	75	74
		Decision Tree	72	68	69	68

Pada pembagian 80% data latih dan 20% data uji, SVM kembali mencatatkan akurasi tertinggi dengan nilai 86% dan F1 Score 83%. Multinomial NB mengikuti dengan akurasi 79% dan F1 Score 77%. Pada skenario 90% data latih dan 10% data uji, SVM tetap unggul dengan akurasi 88% dan F1 Score 85%, diikuti oleh Multinomial NB dengan akurasi 77% dan F1 Score 75%. Algoritma KNN memiliki performa yang lebih rendah pada semua skenario, dengan akurasi sebesar 53% pada skenario data latih 80% dan data uji 20%. Sehingga secara keseluruhan, SVM menunjukkan performa terbaik pada semua skenario pembagian data, diikuti oleh Multinomial NB, sementara KNN dan Decision Tree memiliki performa yang lebih rendah dalam hal akurasi dan F1 Score.

KESIMPULAN

Berdasarkan hasil penelitian menunjukkan bahwa, algoritma SVM terbukti memiliki kinerja terbaik dalam menganalisis sentimen, dengan akurasi mencapai 88% dan F1 Score 85% pada berbagai skenario pembagian data latih dan uji. Sementara itu, algoritma Multinomial Naïve Bayes juga memberikan hasil yang cukup baik, sedangkan KNN dan Decision Tree kurang optimal karena akurasi yang lebih rendah. Algoritma SVM lebih unggul dibandingkan dengan algoritma lain dalam menganalisis data komentar dari YouTube tentang Program Tapera. Ini berarti bahwa SVM lebih mampu menangkap dan mengolah berbagai sentimen masyarakat, baik ulasan positif, negatif, maupun netral, secara lebih tepat. Dengan menggunakan SVM, pemerintah bisa lebih memahami pendapat masyarakat dan menyesuaikan kebijakan agar lebih sesuai dengan kebutuhan dan harapan masyarakat. Penelitian ini berkontribusi pada pengembangan cara yang efektif untuk menganalisis sentimen terhadap kebijakan pemerintah menggunakan data dari media sosial. Ini penting di masa sekarang, di mana pendapat masyarakat seringkali dibentuk dan disebarkan melalui *platform* online.

UCAPAN TERIMA KASIH

Penelitian ini dapat berjalan atas dukungan pendanaan dari Universitas Bina Sarana Informatika. Kami juga berterima kasih kepada Lembaga Penelitian dan Pengabdian kepada Masyarakat atas dukungan yang diberikan, sehingga penelitian ini dapat diselesaikan tepat waktu.

DAFTAR PUSTAKA

[1] Kompas.com, “Mengenal Lebih Dekat Tapera, Tabungan yang Mudahkan Masyarakat untuk Punya Rumah,” *Kompas.com*, 2023.

[2] Y. Matalon, O. Magdaci, A. Almozlino, and D. Yamin, “Using sentiment analysis to predict opinion inversion in Tweets of political communication,” *Scientific Reports*, vol. 11, no. 1, p. 7250, Mar. 2021, doi: 10.1038/s41598-021-86510-w.

[3] A. R. M. Fikri, J. Jondri, and W. Astuti, “Sentiment Analysis Against IndiHome and First Media Internet Providers Using Ensemble Stacking Method,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 2, Sep. 2022, doi: 10.47065/bits.v4i2.1969.

[4] A. Rasool, R. Tao, K. Marjan, and T. Naveed, “Twitter Sentiment Analysis: A Case Study for Apparel Brands,” *Journal of Physics: Conference Series*, vol. 1176, p. 022015, Mar. 2019, doi: 10.1088/1742-6596/1176/2/022015.

[5] M. T. J. Ansari and N. A. Khan, “Worldwide COVID-19 Vaccines Sentiment Analysis Through Twitter Content,” *Electronic Journal of General Medicine*, vol. 18, no. 6, p. em329, Nov. 2021, doi: 10.29333/ejgm/11316.

[6] R. Firdaus, I. Asror, and A. Herdiani, “Lexicon-Based Sentiment Analysis of Indonesian Language Student Feedback Evaluation,” *Indonesia Journal of Computing*, vol. 6, no. 1, pp. 1–12, 2021.

[7] D. Röchert, G. Neubaum, and S. Stieglitz, “Identifying Political Sentiments on YouTube: A Systematic Comparison Regarding the Accuracy of Recurrent Neural Network and Machine Learning Models,” 2020, pp. 107–121. doi: 10.1007/978-3-030-61841-4_8.

[8] B. Laurensz and E. Sediyono, “Analisis Sentimen Masyarakat terhadap Tindakan Vaksinasi dalam Upaya Mengatasi Pandemi Covid-19,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 2, pp. 118–123, May 2021, doi: 10.22146/jnteti.v10i2.1421.

[9] A. Yadav, C. K. Jha, A. Sharan, and V. Vaish, "Sentiment analysis of financial news using unsupervised approach," *Procedia Computer Science*, vol. 167, pp. 589–598, 2020, doi: 10.1016/j.procs.2020.03.325.
[10] D. Gurkhe, N. Pal, and R. Bhatia, "Effective Sentiment Analysis of Social Media Datasets using Naive Bayesian Classification," *International Journal of Computer Applications*, vol. 99, no. 13, pp. 1–4, Aug. 2014, doi: 10.5120/17430-8274.
[11] P. Hinge, A. Thakur, and H. Salunkhe, "Analysis of Human Resources Attrition: A Thematic and Sentiment Analysis Approach," pp. 820–828, 2023, doi: 10.2991/978-94-6463-136-4_72.
[12] J. Y. Tan and A. S. K. Chow, "Sentiment Analysis on Game Reviews: A Comparative Study of Machine Learning Approaches," in *Conference Proceedings: International Conference on Digital Transformation and Applications (ICDXA 2021)*, Tunku Abdul Rahman University College, 2021, pp. 209–216. doi: 10.56453/icdxa.2021.1023.
[13] K. Suppala and N. Rao, "Sentiment Analysis Using Naïve Bayes Classifier," *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, vol. 8, no. 8, 2019.
[14] A. Rianto and A. R. Pratama, "Sentiment Analysis of COVID-19 Vaccination Posts on Facebook in Indonesia with CrowdTangle," *Jurnal Riset Informatika*, vol. 3, no. 4, pp. 353–362, 2021.
[15] A. Baravkar, R. Jaiswal, J. Chhoriya, and Prof. B. Tekwani, "Sentimental Analysis of YouTube Videos," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 7, no. 2, pp. 554–562, Apr. 2021, doi: 10.32628/CSEIT2172112.
[16] M. Z. Naf'an, A. A. Bimantara, A. Larasati, E. M. Risondang, and N. A. S. Nugraha, "Sentiment Analysis of Cyberbullying on Instagram User Comments," *Journal of Data Science and Its Applications*, vol. 2, no. 1, pp. 38–48, 2019.
[17] H. Nguyen, A. Veluchamy, M. Diop, and R. Iqbal, "Comparative Study of Sentiment Analysis with Product Reviews Using Machine Learning and Lexicon-Based Approaches," *SMU Data Science Review*, vol. 1, no. 4, 2018.
[18] D. Sigmawaty and M. Adriani, "Learning Word Relatedness Over Time for Temporal Ranking," *Jurnal Ilmu Komputer dan Informasi*, vol. 12, no. 2, p. 91, Jul. 2019, doi: 10.21609/jiki.v12i2.745.
[19] A. F. Anjani, D. Anggraeni, and I. M. Tirta, "Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 163–172, Sep. 2023, doi: 10.25077/TEKNOSI.v9i2.2023.163-172.
[20] S. Reza, "Trending Topik di X Tentang Tapera di Mata Warganet," 2024. [Online]. Available: https://www.rri.co.id/samarinda/lain-lain/719937/trending-topik-di-x-tentang-tapera-di-mata-warganet
[21] R. Abidin and A. Herawati, "Analisis Sentimen Publik Terhadap Kebijakan Program Tabungan Perumahan Rakyat (Tapera)," *Journal of Information System and Computer*, vol. 4, no. 1, pp. 13–19, 2024.
[22] E. D. Harahap and R. Kurniawan, "Analisis Sentimen Komentar Terhadap Kebijakan Pemerintah Mengenai Tabungan Perumahan Rakyat (TAPERA) Pada Aplikasi X Menggunakan Metode Naïve Bayes," *Jurnal Teknik Informatika Unika ST. Thomas (JTIUST)*, vol. 9, no. 1, pp. 166–175, 2024.