
Analisa Sentimen Pengguna Sosial Media Twitter Terhadap Perokok di Indonesia**Dewi Setiyawati¹, Nuri Cahyono²**

dewi.setiyawati@students.amikom.ac.id, nuricahyono@amikom.ac.id

Universitas Amikom Yogyakarta

Informasi Artikel

Diterima : 4 Feb 2023
Direview : 11 Feb 2023
Disetujui : 26 Feb 2023

Kata Kunci

Analisis Sentimen,
Perokok, Naive Bayes,
Support Vector Machine,
Decision Tree

Abstrak

Salah satu alat yang digunakan pengguna web untuk mengakses, berbagi, dan mendiskusikan subjek yang sedang diminati adalah media sosial. Salah satu situs jejaring sosial, Twitter, sering digunakan secara real time untuk berkomunikasi ini. Karena dampak negatifnya yang signifikan baik bagi kesehatan maupun ekonomi, merokok masih menjadi topik perdebatan dan perdebatan rutin di Indonesia. Penelitian ini dilakukan untuk menilai sentimen terhadap perokok dan membedakan antara emosi positif dan negatif. Data yang digunakan dalam penelitian ini diperoleh dengan melakukan crawling pada jaringan media sosial Twitter. Tiga teknik, yaitu Naive Bayes (NB), Support Vector Machine (SVM), dan Logistic Regression digunakan dalam penelitian ini. Dalam penelitian ini, 40,25% pengguna Twitter setuju dengan keberadaan perokok di Indonesia, sedangkan 59,74% tidak setuju. Metode Naive Bayes digunakan dalam penelitian ini, memberikan nilai akurasi tertinggi 62,1% dengan menggunakan 60% data latih dan 40% data uji.

Keywords

*Sentiment Analysis, Smokers,
Naive Bayes, Support Vector
Machine, Decision Tree*

Abstrak

One of the tools web users use to access, share, and discuss subjects of interest is social media. One social networking site, Twitter, is often used in real time to communicate this. Due to its significant negative impacts on both health and the economy, smoking is still a topic of regular debate and debate in Indonesia. This research was conducted to assess sentiment towards smokers and differentiate between positive and negative emotions. The data used in this study were obtained by crawling the Twitter social media network. Three techniques, namely Naive Bayes (NB), Support Vector Machine (SVM), and Logistic Regression are used in this study. In this study, 40.25% of Twitter users agreed with the existence of smokers in Indonesia, while 59.74% disagreed. The Naive Bayes method was used in this study, giving the highest accuracy value 62.1% using 60% training data and 40% test data.

A. Pendahuluan

Teknologi saat ini berkembang karena konsekuensi dari pertumbuhan internet yang meluas di seluruh dunia. Selaku konsekuensi dari meningkatnya kebutuhan warga dengan bermacam tipe informasi yang nantinya akan digunakan untuk kebutuhan yang berbeda - beda. Dengan timbulnya media sosial di Indonesia, menjadi menarik untuk mengklasifikasikan opini publik tentang perokok yang senantiasa jadi bahan perdebatan. Penggunaan metodologi serta perlengkapan analisis yang cocok untuk membuat analisis media lebih mudah serta lebih efektif. Utamanya media sosial mempunyai akibat yang signifikan pada cara bangsa beroperasi. Setiap orang mempunyai hak untuk membaca serta mengatakan pendapatnya secara online, baik positif ataupun negatif.

Nilai data yang dimiliki pengguna platform media sosial ini sangat berarti. Orang - orang dari segala dunia dengan leluasa mengatakan komentar mereka di web media sosial semacam YouTube, Twitter, Weibo, ataupun WeChat, dan lewat toko online ataupun web individu. Aspek politik serta ekonomi warga mereka menjadi bahan perdebatan di antara banyak pengguna. Web media sosial yang kerap digunakan untuk memperdebatkan isu - isu politik serta sosial adalah Twitter [1]. Twitter merupakan website microblogging yang terkenal di mana pengguna bisa mempublikasikan pembaruan status (diucap "tweet"). Banyak suka, tidak suka, serta kontribusinya pada bermacam topik tercantum dalam tweet ini [2].

Orang-orang di Indonesia dan banyak negara lain menggunakan Twitter sebagai bagian dari platform media sosial utama. Yang mengejutkan 73% hingga 87% pengguna internet yang membaca ulasan online mengklaim bahwa mereka terpengaruh ulasan dalam memutuskan apakah akan membeli produk atau menggunakan layanan. Dari 81% pengguna internet telah melakukan setidaknya 1 pencarian produk [3]. Konsekuensinya, Twitter bertindak sebagai tempat mengumpulkan informasi yang diungkapkan secara publik melalui tweet yang diposting ke timeline [4]. Istilah "tweet" berkaitan dengan setiap tulisan atau teks yang dikirim ke layanan Twitter. Twitter sering digunakan untuk bercerita tentang peristiwa selain berbagi informasi. Orang sering mengungkapkan emosinya saat bercerita [5].

Trending Topics merupakan fitur di Twitter yang mempermudah pengguna dalam mencari data tentang topik tertentu yang sedang trending. Tetapi, sebagian pengguna bisa jadi kesusahan menguasai topik pendek yang sedang tren di Twitter [6]. Sebagian riset lebih dahulu menggunakan informasi yang dikumpulkan dari Twitter untuk melaksanakan riset tentang topik yang berkaitan dengan tim. World Health Organization menegaskan bila tren merokok dikala ini terus bersinambung, hendak menyebabkan korban 10 juta orang pada tahun 2030, 70% di antara lain terletak di negeri berkembang seperti Indonesia. Meningkatnya jumlah perokok berakibat negatif terhadap kesehatan secara signifikan [7].

Selain itu, sebagian besar perokok penduduk Indonesia, memberikan kontribusi signifikan terhadap tingkat merokok terbesar di dunia. Aliansi Pengendalian Tembakau Asia Tenggara memperkirakan Indonesia memiliki 65,19 juta perokok pada tahun 2019, menjadikannya negara dengan prevalensi rokok tertinggi di kawasan ini [8]. Menurut World Health Organization pada tahun 2017, prevalensi perokok muda di Indonesia sebesar 11,5%; pria memiliki persentase lebih tinggi daripada wanita (1,5% vs 21,4%) [9]. Prevalensi merokok pada

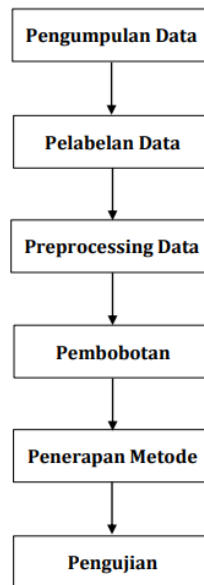
penduduk usia 10 hingga 18 tahun meningkat dari 7,2% menjadi 9,1% sejak tahun 2016, menunjukkan kecenderungan peningkatan prevalensi merokok secara nasional lebih besar pada kelompok anak-anak dan remaja [10]. Di Provinsi Jawa Barat, jumlah penduduk berusia di atas 15 tahun yang merokok meningkat dari 33,19% pada tahun 2017 menjadi 35,78% pada tahun 2018 [11].

Ada berbagai metode pengumpulan data melalui Twitter, dan masing-masing memiliki kelebihan dan kekurangan. Dalam penelitian ini, penulis mencoba melakukan analisis komparatif tentang kemampuan metode Twitter untuk pengumpulan data. Data dari hasil laporan tersebut kemudian dapat digunakan untuk komunitas atau lembaga yang terkait dengannya sebagai alat penilaian. Ini akan mengenali pendapatan atau pola yang akan digunakan untuk proses pengujian SVM, yang bertujuan untuk memberi label sentimen pada tweet, dengan melakukan pelatihan tentang klasifikasi SVM. Untuk mendapatkan informasi sentimen yang terkandung dalam kalimat opini, analisis sentimen merupakan teknik pemahaman, penggalian informasi subjektif dari sumber, dan secara otomatis mengolah data kontekstual. Asumsi kuat digunakan dalam algoritme pembelajaran Nave Bayes, yang akan didasarkan pada teori Bayesian.

Teori Bayes adalah teori yang menggunakan data yang dikumpulkan sebelumnya sehingga memutuskan apa yang memiliki probabilitas tertinggi [12]. Pohon Keputusan, sebaliknya, adalah algoritma prediksi yang menggunakan data yang dikumpulkan sebelumnya [13]. Analisis sentimen adalah proses penambahan data yang digunakan untuk menemukan dan mengekstraksi data dari teks dengan tujuan memahami sentimen sosial. Untuk lebih memahami tentang sikap, pendapat, dan pendekatan dalam teks informasi, analisis sentimen juga dilakukan [14]. Python adalah bahasa pemrograman yang digunakan oleh penulis penelitian ini untuk menganalisis dan menerapkan metode yang akan digunakan. Python lebih seperti bahasa manusia daripada bahasa mesin karena merupakan bahasa pemrograman tingkat tinggi. Python dirancang untuk memberikan kenyamanan programmer dalam hal efisiensi waktu, kemudahan pengembangan, dan kompatibilitas sistem [15].

B. Metode Penelitian

Metode penelitian ini meliputi pengumpulan data, pelabelan data, pra-pemrosesan, pembobotan, penerapan metode, dan pengujian. Proses analisis dimulai dengan merakit data Twitter menggunakan twints, melabeli data, dan menyiapkan data melalui preprocessing yang meliputi beberapa langkah seperti cleaning, case folding, tokenizing, stopwords, dan stemming, proses pembobotan kata TF-IDF, proses penerapan metode, dan proses pelaksanaan ujian pengujian menggunakan training set. Bahkan, untuk mendapatkan presisi yang lebih tinggi, evaluasi matriks konfusi harus digunakan. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

1. Pengumpulan Data

Kumpulan data yang digunakan dalam penelitian ini dikumpulkan dari media sosial melalui Twitter menggunakan twint serta terminal Gitbash. Twint adalah alat untuk mengambil informasi dari Twitter. Python adalah bahasa pemrograman untuk membuat Twint, dan pip atau conda dapat digunakan untuk menginstal python [16]. Library Twint dipilih karena dapat mengambil data tweet pada twitter tanpa memerlukan Twitter API [17]. Jalankan perintah untuk mengumpulkan data Twitter menggunakan terminal git bash. Data yang dikumpulkan terdiri dari pandangan perokok yang telah dikumpulkan selama 25 Oktober dan 3 November 2022 menggunakan frase pencarian "perokok" dan batas "3000." Pemilihan data adalah proses menghilangkan data tweet yang tidak terkait dengan perokok dari informasi mentah yang telah dikumpulkan, seperti tweet yang berisi promosi penjualan atau tweet kosong yang hanya berisi kata "perokok", dan data duplikat juga dibersihkan.

2. Pelabelan Data

Pelabelan data, juga disebut sebagai pelabelan, adalah tindakan untuk secara manual menentukan sentimen tweet dalam kumpulan data. Sentimen positif dan negatif akan digunakan untuk mengklasifikasikan opini atau sentimen tentang dataset. Sentimen positif berhubungan dengan komentar yang benar-benar menyemangati atau menghargai perokok. Pendapat yang mengandung kata-kata hinaan, pernyataan ketidaksetujuan, atau bahkan sentimen negatif lainnya akan dicatat sebagai sentimen negatif. Sebanyak 2.772 tweet frasa ditemukan selama proses pengumpulan data, 1.116 di antaranya positif dan 1.656 di antaranya negatif.

3. Preprocessing Data

Pada tahap preprocessing yang merupakan tahapan penting untuk tahap selanjutnya, ada beberapa langkah yang harus dikerjakan. Tahap preprocessing nantinya berguna untuk membersihkan kata sehingga memudahkan dalam proses Analisa. Beberapa tahapan preprocessing antara lain sebagai berikut :

- a. **Cleansing**
Cleansing adalah proses penghapusan karakter khusus dalam tweet seperti tanda baca (contohnya : koma (,), titik(.) tanda tanya (?), tanda seru (!) dan sebagainya), angka numerik (0 - 9), dan karakter lainnya (contohnya : &, %, *, dan sebagainya). Proses ini memanfaatkan library string dari python.
- b. **Case Folding**
Case folding adalah mengubah semua huruf dalam data menjadi huruf kecil semua. Proses ini menggunakan banuan syntax dari python yaitu `str.lower()`.
- c. **Tokenizing**
Tokenizing berfungsi untuk memisahkan kalimat menjadi kata-kata yang mempunyai makna sehingga bisa meningkatkan keakuratan data. Pada prosesini, menggunakan library python yaitu `nlk.tokenize`.
- d. **Stopword**
Stopword merupakan proses filtering untuk menghilangkan kata - kata yang tidak digunakan dalam proses analisis sentimen. Contohnya adalah “yang”, “dan”, “di”, “dari”, dll.
- e. **Stemming**
Stemming merupakan proses menghilangkan semua imbuhan yang ada pada awal dan akhir kata.

4. Pembobotan

Metode untuk mendeteksi keunikan teks disebut TF-IDF [18]. Efisiensi dan akurasi metode ini di atas normal. Untuk setiap term (kata) pada setiap dokumen, metode ini menentukan jumlah Term Frequency (TF) dan Inverse Document Frequency (IDF) [19]. Data yang telah mendapatkan preprocessing akan diberi bobot pada proses ini menggunakan TF-IDF. Data pengujian dan data pelatihan memang merupakan dua jenis informasi di mana hasil pembobotan akan disimpan.

5. Penerapan Metode dan Pengujian Akurasi

Algoritma klasifikasi studi ini dimaksudkan untuk menganalisis kumpulan data menggunakan tiga pendekatan berbeda. *Naive Bayes (NB)*, *Support Vector Machine (SVM)*, dan *Decision Tree* adalah algoritma pengujiannya. Naïve Bayes adalah cara untuk menggunakan probabilitas dan statistik. Algoritma ini ditemukan oleh seorang ilmuwan Inggris bernama Thomas Bayes. Dampak opini terhadap independensi setiap kondisi adalah salah satu ciri utama metode ini [20]. Teknologi pembelajaran terawasi bernama Support Vector Machine (SVM) berupaya mengidentifikasi pola dan mengatur informasi [21]. Pada Decision Tree, output mudah dibaca dan diinterpretasikan tanpa adanya pengetahuan statistic, namun kurang stabil karena adanya perubahan data skala kecil akan menyebabkan perubahan hasil yang cukup signifikan[22]. Untuk menghitung tingkat akurasi, dalam penelitian ini menggunakan Confusion Matrix dengan menghitung jumlah prediksi benar dan salah dari sebuah metode klasifikasi berbanding dengan data sesungguhnya atau prediksi target [23].

Tabel 1. Confusion Matrix

		Kelas Prediksi	
		Positif	Negatif
Kelas Sebenarnya	Positif	TP	FP
	Negatif	FN	TN

Keterangan :

- TP (True Positive) = jumlah dokumen dari kelas positif yang benar diklasifikasikan sebagai kelas positif
- TN (True Negative) = jumlah dokumen dari kelas negatif yang benar diklasifikasikan sebagai kelas negatif
- FP (False Positive) = jumlah dokumen dari kelas negatif yang salah diklasifikasikan sebagai kelas positif
- FN (False Negative) = jumlah dokumen dari kelas positif yang salah diklasifikasikan sebagai kelas negatif

Rumus confusion matrix untuk perhitungan accuracy, precision, dan recall seperti berikut [24].

$$accuracy = \frac{TP + TN}{Total}$$

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

C. Hasil dan Pembahasan

Data set yang digunakan dalam penelitian ini berasal dari social media twitter berupa tweet pengguna yang mengandung kata 'perokok'. Pengumpulan data dilakukan dengan menggunakan terminal gitbash dengan library twint. Pengumpulan data twitter melalui terminal git bash dengan menjalankan command. Data yang diambil berisikan pendapat masyarakat mengenai perokok yang diambil pada tanggal 25 Oktober 2022 sampai 3 November 2022 dengan kata pencarian "perokok" dan limit "3000". Pengimplementasian library twint dengan menjalankan command pada gitbash seperti dibawah ini :

```
$ pipenv Shell
$ twint -s "perokok" -o perokok.csv --limit 3000 --since 2022-10-25 --until
2022-11-03 -- csv
```

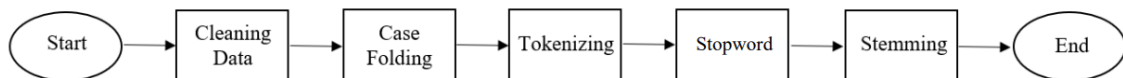
Untuk menyiapkan informasi sebelum dilabeli dengan sentimen positif atau negatif, data mentah yang sudah didapatkan akan diseleksi terlebih dahulu. Pemilihan data adalah proses menghilangkan data duplikasi dan quasi tweet. Pelabelan data, juga disebut pelabelan, adalah tindakan menentukan secara manual sentimen tweet dalam kumpulan data. Sentimen positif dan negatif akan digunakan untuk mengklasifikasikan opini atau sentimen tentang dataset. Sentimen positif berhubungan dengan komentar yang benar-benar mendukung atau menghargai perokok. Setiap kurangnya kosa kata persetujuan dalam pendapat selain bahasa yang kasar akan diklasifikasikan sebagai sentimen negatif. Karena signifikansi setiap sentimen sangat mempengaruhi keakuratan hasil, maka proses pelabelan data harus dilakukan secara objektif. Ditemukan 2.772 tweet bersih dengan emosi selama proses pengumpulan data dengan 1.116 tweet yang menunjukkan sentimen positif dan 1.656 tweet yang menunjukkan sentimen negatif (40,25% positif dan

59,74% negatif). Pelabelan data ini bertujuan untuk menyiapkan informasi untuk penanganan pada proses selanjutnya. Contoh dari hasil pelabelan data dapat dilihat pada Tabel 2.

Tabel 2. Hasil Pelabelan Data

No	Text	Sentimen
1	Anda perokok berat pengen cari istri nggak ngelarang gampang waktu pacaran minta belikan sebungkus rokok kalo dibelikan berarti dia mencintai mu apa adanya	Positif
2	asli cowo perokok itu setia korek nya aja ilang di cariin ampe ketemu lagi apa lagi pacar nya hahah	Positif
3	Saya malahan suka kalo cewek saya perokok keren aja gituu	Positif
4	Perokok memiliki sepuluh kali lebih banyak keriput daripada non-perokok	Negatif
5	mimpi david ngerokok hehh kalo smpe beneran dia jd perokok gua musuhiiiiinn	Negatif

Tahap penting dari proses penambangan adalah pra-pemrosesan. Penanganan data awal yang akan digunakan dalam proses penambangan tidak selalu sederhana. Oleh karena itu, menyelesaikan tahap preprocessing menjadi penting karena nantinya akan berguna untuk penjernihan kata dan mempermudah proses analisis. Beberapa tahapan preprocessing antara lain sebagai berikut :



Gambar 2. Tahapan Pre-processing Data

Pada pre-processing data menggunakan library yang telah disediakan oleh Bahasa pemrograman python. Contoh dari hasil preprocessing data dapat dilihat pada Tabel 3.

Tabel 3. Hasil Pre-processing Data

Tahapan	Hasil
Data Awal	Anda perokok berat pengen cari istri yang nggak ngelarang, gampang waktu pacaran minta belikan sebungkus rokok kalo dibelikan berarti dia mencintai mu apa adanya.
Cleaning Data	Anda perokok berat pengen cari istri yang nggak ngelarang gampang waktu pacaran minta belikan sebungkus rokok kalo dibelikan berarti dia mencintai mu apa adanya
Case Folding	anda perokok berat pengen cari istri yang nggak ngelarang gampang waktu pacaran minta belikan sebungkus rokok kalo dibelikan berarti dia mencintai mu apa adanya
Tokenizing	'anda', 'perokok', 'berat', 'pengen', 'cari', 'istri', 'nggak', 'ngelarang', 'gampang', 'waktu', 'pacaran', 'minta', 'belikan', 'sebungkus', 'rokok', 'kalo', 'dibelian', 'berarti', 'dia', 'mencintai', 'mu', 'apa', 'adanya'
Stopword	", 'perokok', 'berat', 'pengen', 'cari', 'istri', '', 'ngelarang', 'gampang', 'waktu', 'pacaran', 'minta', 'belikan', 'sebungkus', 'rokok', 'kalo', 'dibelian', 'berarti', '', 'mencintai', 'mu', 'apa', 'adanya'

Stemming ", 'okok', 'berat', 'ken', 'cari', 'istri', ", 'ngelarang', 'gampang', 'waktu', 'pacar',
'minta', 'belikan', 'bungkus', 'rokok', 'kalo', 'beli', 'arti', ", 'cinta', 'mu', 'apa', 'ada'

Adapun hasil visualisasi dari keseluruhan data yang bersentimen negatif maupun positif atau bisa disebut Wordcloud yang merupakan representasi visual dari semua data dengan melakukan plotting kata sering muncul pada data tersebut pada ruang dua dimensi dapat dilihat pada Gambar 3.



Gambar 3. Visualisasi Data Sentimen Negatif dan Positif

Menurut studi tentang sentimen, sentimen negatif mengkompensasi 59,74% dari semua sentimen, sedangkan sentimen positif mengkompensasi sekitar 40% dari semua tweet. Untuk menetapkan cara memperkirakan sentimen, perhitungan akurasi berdasarkan data di atas harus dilakukan. Tingkat akurasi ini ditentukan dengan menggunakan algoritma Nave Bayes, Support Vector Machines, dan Decision Trees. Kesesuaian metode analisis sentimen dipengaruhi oleh jumlah data pelatihan dan pengujian, keragaman set data yang digunakan, dan rasio data positif terhadap data negatif. Pengujian kali ini akan diberikan sebanyak 5 kali, dengan data pengujian dan pelatihan tersebar dengan berbagai cara seperti pada tabel 4.

Tabel 4. Pengujian Pemecahan Data

Pengujian	Data Testing	Data Training
1	20%	80%
2	40%	60%
3	50%	50%
4	60%	40%
5	80%	20%

Berikut hasil dari salah satu perhitungan tingkat akurasi menggunakan Confusion Matrix dengan menghitung jumlah prediksi benar dan salah dari sebuah metode klasifikasi berbanding dengan data sesungguhnya atau prediksi target.

Tabel 5. Confusion Matrix Naïve Bayes

	Kelas Prediksi	
	Positif	Negatif

Kelas	Positif	907	93
Sebenarnya	Negatif	537	127

Dari data dalam tabel confusion matrix didapatkan hasil precision sebesar 0,608, recall sebesar 0,621, dan accuracy sebanyak 0,621.

Hasil dari pengujian akurasi dengan ketiga algoritma dapat dilihat pada tabel 6.

Tabel 6. Pengujian Pemecahan Data

Pengujian	Data Testing	Akurasi Naïve Bayes	Akurasi SVM	Akurasi Decision Tree
1	20%	59,3%	57,5%	57,3%
2	40%	61,2%	60,0%	55,9%
3	50%	61,4%	59,8%	56,0%
4	60%	62,1%	60,8%	56,5%
5	80%	61,2%	60,6%	56,1%

Temuan dari lima tes memberikan data tentang perbedaan dari masing – masing tingkat akurasi. Metode Naive Bayes yang menggunakan 60% data uji dan 40% data latih memberikan hasil uji dengan akurasi tertinggi yaitu 62,1%. Kemudian pengujian yang menggunakan metode Support Vector Machines memiliki akurasi tertinggi sebesar 60% dan membutuhkan 60% data pengujian dan 40% data pelatihan. Decision Tree memiliki akurasi rata-rata 57,3% menggunakan 20% data uji dan 80% data pelatihan. Setelah menganalisis ketiga metode ini, ditemukan bahwa metode Nave Bayes, yang menggunakan 60% data uji memiliki akurasi tertinggi.

D. Simpulan

Berdasarkan hasil analisis sentimen terhadap perokok di Indonesia pada twitter dapat ditarik kesimpulan sebagai berikut :

1. Berdasarkan 2.772 tweet yang telah diposting oleh masyarakat ke Twitter di Indonesia terkait perokok, studi sentimen mengungkapkan bahwa 1.116 tweet tersebut menyatakan opini yang bernilai positif dan 1.656 tweet tersebut menyatakan opini yang bernilai negaif.
2. Jelas bahwa 40,25 % pengguna Twitter menyukai rokok di Indonesia sedangkan 59,74% menolaknya. Dengan kata lain, masyarakat masih ragu-ragu untuk menerima perokok.
3. Dari kelima pengujian yang dilakukan pada ketiga metode tersebut, metode Naive Bayes terlihat memiliki akurasi paling besar yaitu 62,1%, dengan sebaran data training sebesar 40% dan data testing sebesar 60%.
4. Terbukti dari hasil pengujian dan analisis bahwa distribusi data pelatihan dan pengujian berpengaruh besar terhadap tingkat kinerja sistem.
5. Metode pelabelan data memiliki pengaruh besar hanya pada nilai akurasi metode.

E. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Universitas Amikom Yogyakarta atas seluruh dukungan dalam penelitian ini.

F. Referensi

- [1] M. Kamyab, R. Tao, M. H. Mohammadi, dan A. Rasool, "Sentiment analysis on Twitter: A text mining approach to the Afghanistan status reviews," dalam *ACM International Conference Proceeding Series*, Nov 2018, hlm. 14–19. doi: 10.1145/3293663.3293687.
- [2] Amrita Shelar dan Ching-yu Huang, "Sentiment Analysis of Twitter Data," *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*, hlm. 1–2, 2018.
- [3] B. Pang dan L. Lee, "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval*, vol. 2, 2008.
- [4] M. K. Anam, B. N. Pikir, dan M. B. Firdaus, "Penerapan Naïve Bayes Classifier, K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen danPemerintah," *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 21, no. 1, hlm. 139–150, Nov 2021, doi: 10.30812/matrik.v21i1.1092.
- [5] F. Syah, H. Fajrin, A. N. Afif, R. Saeputra, D. Mirrantly, dan D. D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 7, no. 1, 2023, doi: 10.35870/jti.
- [6] K. Lee, D. Palsetia, R. Narayanan, Md. M. A. Patwary, A. Agrawal, dan A. Choudhary, "Twitter Trending Topic Classification," dalam *2011 IEEE 11th International Conference on Data Mining Workshops*, Des 2011, hlm. 251–258. doi: 10.1109/ICDMW.2011.171.
- [7] P2PTM Kemenkes RI, "Merokok, Tak Ada Untung Banyak Sengsaranya," *p2ptm.kemkes.go.id*, 2018.
- [8] Ade Miranti Karunia, "Perokok di Indonesia Tembus 65 Juta, Pemerintah RI Perlu Belajar dari Jepang, Inggris dan Swedia," *kompas.com*, hlm. 1–1, 2023.
- [9] M. Ibrahim, C. A. Fatmawati, dan W. Hadiyani, "Pengaruh Empowerment Melalui Fotografi Terhadap Proactive Coping Pada Remaja Perokok," *Risenologi*, vol. 7, no. 1a, hlm. 14–20, Agu 2022, doi: 10.47028/j.risenologi.2022.71a.326.
- [10] K. K. R. Biro Komunikasi dan Pelayanan Masyarakat, "Peringati Hari Tanpa Tembakau Sedunia, Kemenkes Targetkan 5 Juta Masyarakat Berhenti Merokok," *kemkes.go.id*, 2021.
- [11] Badan Pusat Statistik, "Persentase Merokok Pada Penduduk Umur ≥ 15 Tahun Menurut Provinsi," 2019.
- [12] B. Pratama dkk., "Sentiment Analysis of the Indonesian Police Mobile Brigade Corps Based on Twitter Posts Using the SVM and NB Methods," dalam *Journal of Physics: Conference Series*, Mei 2019, vol. 1201, no. 1. doi: 10.1088/1742-6596/1201/1/012038.
- [13] T. Asih, Q. Putri, A. Triayudi, dan R. T. Aldisa, "Sawo Manila No.61, RW.7, Pejaten Bar," *Journal of Information System Research*, vol. 4, no. 2, 2023, doi: 10.47065/josh.v4i2.2949.
- [14] D. Duei Putri, G. F. Nama, dan W. E. Sulistiono, "Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive

- Bayes Classifier," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 10, no. 1, Jan 2022, doi: 10.23960/jitet.v10i1.2262.
- [15] D. Aji Imadudin dan N. Cahyana Aminuallah, "Sistem Pengolahan Citra Digital Untuk Menentukan Bobot Kambing Menggunakan Metode Titik Berat," *Jurnal Teknologi Terkini*, vol. 3, hlm. 1–11, 2023.
- [16] M. Iqbal Habibie dan T. Widiaputra, "Web Scraping Of Disease Information From Social Media Twitter," 2022. [Daring]. Tersedia pada: <https://ejurnal.teknokrat.ac.id/index.php/teknoinfo/index>
- [17] A. Syaifuddin dan M. Muslimin, "Analisis Sentimen Pada Sosial Media Tentang Implementasi Kebijakan Pse Kominfo Menggunakan Algoritme Lexicon Based," *Smart City And Sustainable Development Goals*, vol. 1, no. 1, 2022.
- [18] P. H. Saputro, M. Aristian, dan D. ListianingTyas, "Klasifikasilagu Daerah Indonesia Berdasarkanlirikmenggunakanmetode Tf-Idf Dan Naïve Bayes," *Jurnal Teknologi Informasi dan Terapan*, vol. 4, no. 1, hlm. 47–52, Mar 2019, doi: 10.25047/jtit.v4i1.20.
- [19] Muljono, D. Putri Artanti, A. Syukur, A. Prihandono, dan de Rosal I. Moses Setiadi, "Analisa Sentimen Untuk Penilaian Pelayanan Situs Belanja Online Menggunakan Algoritma Naïve Bayes," *Konferensi Nasional Sistem Informasi 2018*, hlm. 1–6, 2018.
- [20] M. I. Petiwi, A. Triayudi, dan I. D. Sholihati, "Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine," *Jurnal Media Informatika Budidarma*, vol. 6, no. 1, hlm. 542, Jan 2022, doi: 10.30865/mib.v6i1.3530.
- [21] M. Akmal Iftikar dan Y. Sibaroni, "Analisis Sentimen Twitter: Penanganan Pandemi Covid-19 Menggunakan Metode Hybrid Naïve Bayes, Decision Tree, dan Support Vector Machine," vol. 9, hlm. 1–8, 2022.
- [22] A. Rozaq, Y. Yunitasari, K. Sussolaikah, E. R. N. Sari, dan R. I. Syahputra, "Analisis Sentimen Terhadap Implementasi Program Merdeka Belajar Kampus Merdeka Menggunakan Naïve Bayes, K-Nearest Neighbors Dan Decision Tree," *Jurnal Media Informatika Budidarma*, vol. 6, no. 2, hlm. 746, Apr 2022, doi: 10.30865/mib.v6i2.3554.
- [23] H. S. Utama, D. Rosiyadi, D. Aridarma, dan B. S. Prakoso, "Sentimen Analisis Kebijakan Ganjil Genap Di Tol Bekasi Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information," *Jurnal Pilar Nusa Mandiri*, vol. 15, no. 2, hlm. 247–254, Sep 2019, doi: 10.33480/pilar.v15i2.705.
- [24] D. Normawati dan S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," 2021.