

ANALISIS SENTIMEN PADA PLATFORM X TERHADAP LAYANAN PROVIDER TRI MENGGUNAKAN NAÏVE BAYES DAN SUPPORT VECTOR MACHINE

Adrian Nur Ihsan¹, Shandy Tresnawati²

^{1,2} Politeknik TEDC Bandung; Jl. Politeknik Jl. Pesantren No.2, Cibabat, Kec. Cimahi Utara, Kota Cimahi, Jawa Barat 40513; telp (022) 6645951

Received: 27 Agustus 2024

Accepted: 5 Oktober 2024

Published: 12 Oktober 2024

Keywords:

Analisis Sentimen, Naïve Bayes, Support Vector Machine, Platform X, Data Mining SEMMA.

Correspondent Email:

if.adrianihsn21@gmail.com

Abstrak. Perkembangan teknologi digital dan media sosial telah mengubah cara masyarakat berinteraksi dan menyampaikan opini. Platform X (sebelumnya dikenal sebagai Twitter) menjadi salah satu media yang banyak digunakan untuk menyuarakan pendapat terkait berbagai isu, termasuk layanan telekomunikasi. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap layanan Provider Tri yang diungkapkan melalui tweet di X, dengan menggunakan algoritma Naïve Bayes dan Support Vector Machine serta menerapkan metodologi data mining SEMMA. Data penelitian ini diperoleh melalui teknik crawling, menghasilkan 4.333 tweet yang dikumpulkan dari 1 Januari 2023 hingga 30 April 2024. Dataset ini kemudian dievaluasi dengan membaginya ke dalam berbagai rasio pelatihan dan pengujian. Analisis menggunakan rasio 90:10, yang dilakukan dengan 10-fold cross-validation, menunjukkan bahwa algoritma Naïve Bayes memperoleh akurasi sebesar 67,73%, presisi sebesar 72,30%, recall sebesar 67,73%, dan f1-score sebesar 58,66%. Sebaliknya, algoritma Support Vector Machine mencapai akurasi sebesar 76,10%, presisi sebesar 75,65%, recall sebesar 76,10%, dan f1-score sebesar 75,20%. Hasil menunjukkan bahwa algoritma Support Vector Machine lebih unggul dengan akurasi tertinggi sebesar 76,10%. Selain itu, analisis sentimen menunjukkan bahwa mayoritas sentimen pengguna terhadap layanan Provider Tri adalah positif.

Abstract. The advancement of digital technology and social media has transformed the way people interact and express opinions. Platform X (formerly known as Twitter) has become a widely used medium for voicing opinions on various issues, including telecommunication services. This study aims to analyze user sentiment toward Provider Tri's services expressed through tweets on X, using Naïve Bayes and Support Vector Machine algorithms and applying the SEMMA data mining methodology. The research data was obtained through crawling techniques, resulting in 4,333 tweets collected from January 1, 2023, to April 30, 2024. This dataset was then evaluated by splitting it into various training and testing ratios. Analysis using a 90:10 ratio, performed with 10-fold cross-validation, revealed that the Naïve Bayes algorithm achieved an accuracy of 67.73%, precision of 72.30%, recall of 67.73%, and an f1-score of 58.66%. In contrast, the Support Vector Machine algorithm reached an accuracy of 76.10%, precision of 75.65%, recall of 76.10%, and an f1-score of 75.20%. The results indicate that the Support Vector Machine algorithm outperforms with the highest accuracy of 76.10%. Additionally, sentiment analysis shows that the majority of user sentiment toward Provider Tri's services is positive.

1. PENDAHULUAN

Perkembangan dunia digital yang semakin pesat, membawa perubahan besar dalam berbagai aspek kehidupan. Akses terhadap informasi, hiburan, dan komunikasi menjadi lebih mudah dan cepat, berkat kemajuan teknologi dan internet. Media sosial menjadi platform untuk mengakses informasi, yang umumnya disajikan dalam bentuk teks dan dikategorikan berdasarkan isinya. Media sosial, seperti platform X tidak hanya menjadi sumber informasi, tetapi juga tempat bagi masyarakat untuk menyampaikan opini. Opini tersebut bisa berupa apresiasi, ujaran kebencian, atau bahkan hoaks yang berpotensi memicu perdebatan di ruang publik digital. Pengguna X biasanya berbagi pesan yang disebut tweet. Tweet ini berfungsi untuk menampung dan mengetahui sentimen dari opini masyarakat.

Analisis sentimen adalah metode untuk mengidentifikasi dan mengkategorikan opini berbentuk teks sebagai sentimen positif atau negatif. Penerapan analisis ini mencakup penilaian terhadap pendapat, sentimen, evaluasi, emosi, atau sikap terkait produk, individu, organisasi, layanan, isu, atau peristiwa dalam masyarakat. Selain itu, analisis sentimen akan selalu berhubungan dengan masyarakat karena sumber informasi yang didapat dari media sosial dimana masyarakat sebagai penggunaannya[1].

Sebagai salah satu metode penting dalam memahami opini publik, analisis sentimen dapat diterapkan untuk mengukur persepsi pengguna terhadap berbagai produk dan layanan. Dalam konteks industri telekomunikasi, analisis sentimen menjadi semakin relevan mengingat peran penting penyedia layanan seluler dalam kehidupan sehari-hari masyarakat. Salah satu operator yang menjadi perhatian adalah Provider Tri, yang sudah lama menjadi salah satu pemain utama di pasar telekomunikasi Indonesia.

Tri merupakan operator seluler di Indonesia yang dimiliki oleh PT Hutchison 3 Indonesia dan memulai operasinya di pasar Indonesia pada tahun 2005. Seiring berjalannya waktu, Tri semakin dikenal, dan perusahaan terus berinovasi dengan menawarkan berbagai paket data, panggilan, dan SMS kepada pelanggan. Pada tahun 2022, Tri bergabung dengan Indosat Ooredoo untuk membentuk Indosat Ooredoo Hutchison, namun tetap mempertahankan

merek "Tri". Meskipun begitu, kualitas layanan tetap menjadi perhatian, dengan masalah seperti hilangnya sinyal, kegagalan registrasi kartu, pemotongan pulsa secara otomatis, serta pengurangan paket data secara tiba-tiba yang masih sering dikeluhkan pelanggan. Oleh karena itu, analisis sentimen pada media sosial X bisa memberi wawasan tentang bagaimana masyarakat melihat layanan provider Tri. Opini tentang Provider Tri banyak ditemukan di X dengan kata kunci "Kartu Tri".

Penelitian sebelumnya tentang analisis sentimen telah dilakukan oleh beberapa peneliti, seperti Yuma Akbar dan Aldino Nur Ihsan dengan judul "Analisis Sentimen Twitter Terhadap Opini Masyarakat Pada Sea Games Kamboja 2023 Menggunakan Algoritma *Support Vector Machine*" memiliki nilai akurasi sebesar 92,08% [2]. Penelitian selanjutnya dilakukan oleh Samsir et al pada tahun 2021 berjudul "Analisis Sentimen Pembelajaran Daring Pada Twitter dimasa Pandemi COVID-19 Menggunakan Metode *Naïve Bayes*" dengan mendapatkan nilai akurasi 95% [3]. Penelitian lain dilakukan oleh Tanthy Tawaqalia Widowati dan Mujiono Sadikin berjudul "Analisis Sentimen Twitter Terhadap Tokoh Publik Dengan Algoritma *Naive Bayes* Dan *Support Vector Machine*" Hasilnya menunjukkan akurasi dari *Naive Bayes* 91.48%, dan SVM 85.47% [4].

Dalam penelitian ini, peneliti tertarik untuk menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* untuk menganalisis sentimen terhadap layanan Provider Tri menggunakan metode SEMMA. Dari referensi penelitian yang ada, Penelitian ini akan mengklasifikasikan sentimen menjadi kategori positif dan negatif. Kemudian hasil dari kedua metode algoritma akan dibandingkan untuk menentukan algoritma mana yang memberikan performa terbaik berdasarkan metrik akurasi, presisi, *recall*, dan *f1-score*.

Dengan alasan alasan diatas, penelitian ini akan dilakukan untuk menganalisis opini-opini sentimen pengguna pada platform X terhadap layanan Provider Tri menggunakan *Naive Bayes* dan *Support Vector Machine*.

2. TINJAUAN PUSTAKA

2.1 Analisis Sentimen

Analisis sentimen adalah salah satu metode untuk mengumpulkan pendapat publik

mengenai berbagai hal seperti layanan publik, isu, kinerja pemerintahan, atau topik lainnya. Metode ini dapat digunakan untuk mengevaluasi layanan yang telah diberikan. Analisis sentimen dapat dilakukan dengan berbagai cara, salah satunya adalah dengan mengumpulkan opini dari media sosial [5].

2.2 Media Sosial X

X adalah salah satu media sosial dan komunikasi yang paling banyak digunakan di seluruh dunia. Aplikasi ini memungkinkan pengguna untuk membuat pesan di halaman pribadi mereka dengan batas maksimal 280 karakter. Manfaat dari X ini selain sebagai alat komunikasi juga dapat berfungsi untuk media informasi, platform untuk berbisnis, menggerakkan opini publik, dan juga sebagai sarana hiburan [6].

2.3 Data Mining

Data mining adalah proses untuk mengumpulkan dan mengolah data dengan tujuan mengekstrak informasi penting dari data tersebut. Proses ini biasanya dilakukan menggunakan perangkat lunak yang memanfaatkan perhitungan matematika, atau kecerdasan buatan. Dalam data mining, terdapat berbagai metode yang dapat digunakan, salah satunya adalah SEMMA (*Sample, Explore, Modify, Model, dan Assess*) [7].

2.4 Google Colab

Google Colab adalah platform berbasis cloud computing yang disediakan oleh Google. Platform ini memungkinkan pengguna untuk menjalankan kode *Python* dalam lingkungan *cloud* tanpa perlu menginstal atau mengonfigurasi lingkungan lokal mereka. Google Colab banyak digunakan oleh ilmuwan data, peneliti, dan pengembang untuk berbagai tugas, seperti pemrosesan data, pengembangan model kecerdasan buatan, analisis data, dan pelatihan model mesin [8].

2.5 Naïve Bayes

Naive Bayes adalah metode klasifikasi yang sederhana namun efektif. Metode ini menerapkan teorema probabilitas untuk menentukan peluang terbaik dengan memperkirakan probabilitas masa depan berdasarkan data yang sudah ada sebelumnya. Klasifikasi Bayes sering diandalkan karena mampu memberikan akurasi yang baik dalam waktu yang cepat, terutama ketika diterapkan pada dataset yang besar [9].

2.6 Support Vector Machine

Support Vector Machine (SVM) adalah metode pembelajaran *linier* yang bertujuan menemukan *hyperplane* optimal untuk memisahkan dua kelas, yaitu positif dan negatif. Keunggulan SVM terletak pada kemampuannya untuk menangani pemisahan *input non-linier* dalam dimensi tinggi melalui penggunaan fungsi kernel [10].

2.7 Lexicon Based

Lexicon Based adalah pendekatan yang menggunakan kamus untuk menentukan bobot kalimat dalam sebuah dataset, sehingga label kelas sentimen pada dataset tersebut dapat diidentifikasi [11]. Kelebihan metode ini adalah kemampuan untuk secara otomatis memberi label pada kalimat, yang dapat menghemat waktu, terutama saat mengolah dataset berukuran besar [12].

2.8 Jupyter Notebook

Jupyter adalah singkatan dari tiga bahasa pemrograman, yaitu *Julia* (*Ju*), *Python* (*Py*), dan *notebook R* [13]. *Jupyter Notebook* adalah aplikasi web gratis yang memungkinkan pembuatan dan berbagi dokumen komputasi, menawarkan pengalaman yang sederhana, efisien, dan berfokus pada dokumen [14].

3. METODE PENELITIAN

Pada tahap ini, metode SEMMA dipilih sebagai pendekatan dalam pengembangan model data mining. SEMMA merupakan singkatan dari lima tahapan utama: *Sample, Explore, Modify, Model, dan Assess*. Berikut adalah penjelasan dari setiap tahapan SEMMA:

3.1. Sample

Pada tahap ini, data untuk penelitian dikumpulkan melalui metode *crawling* menggunakan Google Colab. Proses pengambilan data memerlukan *Auth-Token X*. Selama pengumpulan data, peneliti menggunakan kata kunci "Kartu Tri" untuk mengakses tweet pengguna X terkait layanan Provider Tri.

3.2. Explore

Pada tahap ini, setelah data tweet terkumpul, peneliti melakukan seleksi atribut yang akan digunakan dari banyaknya atribut yang diperoleh melalui *crawling* dengan kata kunci "Kartu Tri." Peneliti kemudian hanya akan menggunakan atribut *full_text*, yang berisi tweet pengguna X.

3.3. Modify

Pada tahap ini, data dari kolom teks yang telah dipilih akan diolah. Proses ini bertujuan untuk mempersiapkan data agar siap untuk diklasifikasikan pada tahap berikutnya. Tahapan yang dilakukan meliputi *Case folding*, *Cleaning*, *Tokenize*, *Normalize*, *Stopword Removal*, dan *Stemming*.

3.4. Model

Pada tahap ini, data dilabeli berdasarkan klasifikasi untuk mengidentifikasi opini positif atau negatif menggunakan metode *lexicon-based*. Setelah dataset dilabeli, kata-kata yang sering muncul dalam setiap klasifikasi akan divisualisasikan dalam bentuk *wordcloud*. Selanjutnya, dataset dibagi menjadi data latih dan data uji, yang kemudian akan diproses menggunakan algoritma *Naive Bayes* dan *Support Vector Machine*.

3.5. Asses

Pada tahap ini dilakukan evaluasi dari model. Pada metode *Naive bayes* dan *Support Vector Machine* digunakan *10-folds cross validation* serta *confussion matrix* untuk mengetahui akurasi, presisi, *recall* dan *f1-score* dari model yang digunakan.

4. HASIL DAN PEMBAHASAN

4.1. Sample

Data yang digunakan diambil dari media platform X mulai tanggal 1 Januari 2023 sampai 30 April 2024. Total data yang berhasil dikumpulkan mencapai 4.333 tweet di media sosial X terkait Layanan Provider Tri. Data yang didapatkan berupa file format csv. Gambar 1 menunjukkan hasil data yang diambil dari X.

conversation_id_str	favorite_count	full_text	id_str	image_url	in_reply_to_screen_name	in_reply_to_text	lang	location	quote_count	reply_count	retweet_count	tweet_url	user_id_str	username
1.780148	0	0	1.780148				id		0	0	0	https://t.co/...	1.780148	...

Gambar 1. Hasil Crawling Data

4.2. Explore

Hasil dari *crawling* data X berisi 15 kolom atribut yang merinci sebuah tweet, di antaranya, *conversation_id_str*, *created_at*, *favorite_count*, *full_text*, *id_str*, *image_url*, *in_reply_to_screen_name*, *in_reply_to_text*, *lang*, *location*,

quote_count, *reply_count*, *retweet_count*, *tweet_url*, *user_id_str*, dan *username*. Kemudian hanya atribut *full_text* saja yang digunakan, yang merupakan isi konten tweet pengguna X untuk memudahkan menganalisis data. Berikut adalah tabel hasil seleksi atribut.

Tabel 1. Hasil Seleksi Atribut

No	Full Text
1	woy kartu tri cukimay lu knp
2	sebenarnya sayang banget sama kartu tri ku yang 11 nomor ini tapi sinyalnya jelek banget kentut
3	Kartu tri lagi gangguan kah? Sinyal ada tapi ngak bisa digunakan @triindonesia
.	.
4333	Ini registrasi kartu tri kenapa gagal mulu sih bang TOLONGLAH CAPE NIH @triindonesia @3CareIndonesia

4.3. Modify

Tahap *Modify* adalah langkah persiapan data atau *pre-processing*. Tahapan yang akan dilakukan terdiri dari *case folding*, *cleaning*, *tokenize*, *normalize*, *stopword removal*, dan *stemming*.

4.3.1 Case Folding

Langkah pertama *Case Folding* dilakukan untuk mengubah semua karakter dalam dataset menjadi huruf kecil. Setelah proses ini, dataset yang dihasilkan tidak lagi mengandung huruf kapital. Tabel 2 berikut ini menunjukkan hasil dari *case folding*.

Tabel 2. Case Folding

Sebelum	Sesudah
@ultramenn__ Di aku sinyalnya sampe ilang ini bela2in ngisi kuota di kartu yg satunya tri yg masih ada sinyal tpi ttp ngelag https://t.co/JBxGfk0cp	@ultramenn__ di aku sinyalnya sampe ilang ini bela2in ngisi kuota di kartu yg satunya tri yg masih ada sinyal tpi ttp ngelag https://t.co/jbxgfk0cpd
INI KARTU TRI GW KENAPA DAH????????? KOK GAK MASUK AKSES INTERNET?????	ini kartu tri gw kenapa dah????????? kok gak masuk akses internet?????

4.3.2 Cleaning

Langkah selanjutnya adalah membersihkan data dari komponen yang tidak relevan dan tidak memiliki makna seperti karakter ASCII, angka, *link*, *hashtag*, URL, tanda baca, dan *whitespace*. Tabel 3 berikut ini menunjukkan hasil *cleaning*.

Tabel 3. *Cleaning*

Sebelum	Sesudah
@ultramenn__ di aku sinyalnya sampe ilang ini bela2in ngisi kuota di kartu yg satunya tri yg masih ada sinyal tpi ttp ngelag https://t.co/jbxgfk0cpd	di aku sinyalnya sampe ilang ini belain ngisi kuota di kartu yg satunya tri yg masih ada sinyal tpi ttp ngelag
ini kartu tri gw kenapa dah???????? kok gak masuk akses internet????	ini kartu tri gw kenapa dah kok gak masuk akses internet

4.3.3 Tokenize

Langkah berikutnya adalah *tokenize*. Proses ini membagi teks menjadi unit-unit seperti kata. Tujuannya adalah untuk menghasilkan token-token yang dapat dianalisis secara individu. Tabel 4 berikut ini menunjukkan hasil *tokenize*.

Tabel 4. *Tokenize*

Sebelum	Sesudah
di aku sinyalnya sampe ilang ini belain ngisi kuota di kartu yg satunya tri yg masih ada sinyal tpi ttp ngelag	['di', 'aku', 'sinyalnya', 'sampe', 'ilang', 'ini', 'belain', 'ngisi', 'kuota', 'di', 'kartu', 'yg', 'satunya', 'tri', 'yg', 'masih', 'ada', 'sinyal', 'tpi', 'ttp', 'ngelag']
ini kartu tri gw kenapa dah kok gak masuk akses internet	['ini', 'kartu', 'tri', 'gw', 'kenapa', 'dah', 'kok', 'gak', 'masuk', 'akses', 'internet']

4.3.4 Normalize

Normalize merupakan tahap dimana dilakukan standarisasi kata yang memiliki makna sama dengan melakukan perubahan penulisan pada suatu kata yang disingkat dan tidak baku agar memiliki arti kata yang seragam. Tabel 5 berikut ini menunjukkan hasil *normalize*.

Tabel 5. *Normalize*

Sebelum	Sesudah
---------	---------

['di', 'aku', 'sinyalnya', 'sampe', 'ilang', 'ini', 'belain', 'ngisi', 'kuota', 'di', 'kartu', 'yg', 'satunya', 'tri', 'yg', 'masih', 'ada', 'sinyal', 'tpi', 'ttp', 'ngelag']	['di', 'aku', 'sinyalnya', 'sampai', 'ilang', 'ini', 'belain', 'ngisi', 'kuota', 'di', 'kartu', 'yang', 'satunya', 'tri', 'yang', 'masih', 'ada', 'sinyal', 'tapi', 'ttp', 'ngelag']
['ini', 'kartu', 'tri', 'gw', 'kenapa', 'dah', 'kok', 'gak', 'masuk', 'akses', 'internet']	['ini', 'kartu', 'tri', 'aku', 'kenapa', 'dah', 'kok', 'tidak', 'masuk', 'akses', 'internet']

4.3.5 Stopword Removal

Stopword removal adalah kata-kata umum dalam bahasa yang sering muncul namun biasanya tidak memberikan makna penting dalam analisis teks. kata-kata yang tidak relevan atau tidak mengandung opini akan dihilangkan. Tabel 6 berikut ini menunjukkan hasil *stopword removal*.

Tabel 6. *Stopword Removal*

Sebelum	Sesudah
['di', 'aku', 'sinyalnya', 'sampai', 'ilang', 'ini', 'belain', 'ngisi', 'kuota', 'di', 'kartu', 'yang', 'satunya', 'tri', 'ada', 'sinyal', 'tapi', 'ngelag']	['aku', 'sinyalnya', 'ilang', 'belain', 'ngisi', 'kuota', 'kartu', 'satunya', 'tri', 'ada', 'sinyal', 'tapi', 'ngelag']
['ini', 'kartu', 'tri', 'aku', 'kenapa', 'dah', 'kok', 'tidak', 'masuk', 'akses', 'internet']	['kartu', 'tri', 'aku', 'dah', 'tidak', 'masuk', 'akses', 'internet']

4.3.6 Stemming

Stemming dilakukan untuk mendapatkan kata dasar dari sebuah kata dengan cara menghilangkan imbuhan dari kata. Tabel 7 berikut menampilkan hasil dari tahap *stemming*.

Tabel 7. *Stemming*

Sebelum	Sesudah
['aku', 'sinyalnya', 'ilang', 'belain', 'ngisi', 'kuota', 'kartu', 'satunya', 'tri', 'ada', 'sinyal', 'tapi', 'ngelag']	['aku', 'sinyal', 'ilang', 'bela', 'ngisi', 'kuota', 'kartu', 'satu', 'tri', 'ada', 'sinyal', 'tapi', 'lag']
['kartu', 'tri', 'aku', 'dah', 'tidak', 'masuk', 'akses', 'internet']	['kartu', 'tri', 'aku', 'dah', 'tidak', 'masuk', 'akses', 'internet']

4.4. Model

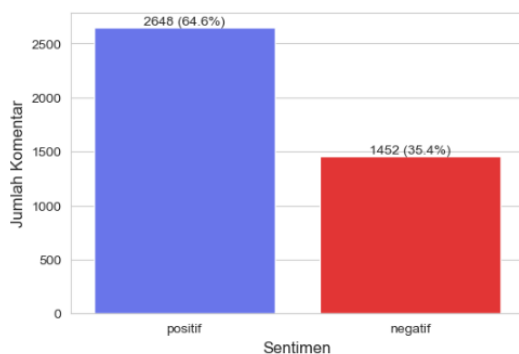
Pada tahap Model dalam penelitian ini, dilakukan pelabelan sentimen pada dataset menggunakan metode *lexicon-based* dan dua model klasifikasi dengan algoritma *Naive Bayes* dan *Support Vector Machine*.

4.4.1 Lexicon Based

Penelitian ini menggunakan metode berbasis leksikon untuk mengklasifikasikan dataset. Setiap tweet dalam dataset dikategorikan menjadi dua kelas yaitu, positif atau negatif. Kata-kata dalam tweet yang sesuai dengan kamus *lexicon* akan diberikan skor. Total skor pada suatu *tweet* kemudian menentukan apakah labelnya positif atau negatif. Tabel 8 menampilkan hasil dari dataset yang telah diberi label menggunakan metode berbasis *lexicon*. Gambar 2 menunjukkan jumlah dataset *tweet* yang telah diklasifikasikan menggunakan metode *Lexicon Based*.

Tabel 8. Hasil Pelabelan

No	Tweet	Label
1	aku make kartu tri kalo isi pulsa konter konter dekat ama rumah gampang	Positif
2	enak pake kartu tri isi pulsa suka dapet bonus kuota	Positif
3	min kartu tri saya tidak deteksi handphone saya baca kartu tidak valid	Negatif
4	saya pakai kartu tri semenjak merger indosat sinyal lelet banget parah kecewa berat	Negatif
5	kartu tri tidak aktif pas pasang handphone layar handphone tulis kartu tidak terdetek alias kartu sim tidak ada aktif september kartu ada ya	Negatif

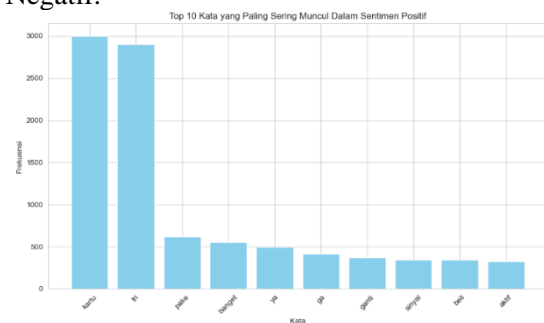


Gambar 2. Jumlah *Tweet* Setelah Dilakukan *Lexicon Based*

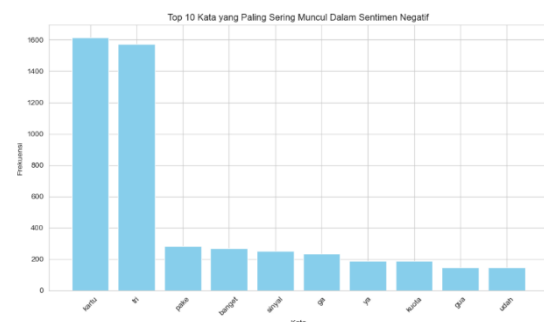
Hasil gambar menampilkan sentimen dari tweet-tweet terkait kartu Tri yang

diklasifikasikan menjadi dua kategori: positif dan negatif. Dari total tweet yang dianalisis, sebanyak 64.6% atau 2.648 tweet menunjukkan sentimen positif. Sementara itu, 35.4% atau 1.452 tweet mengindikasikan sentimen negatif.

Setelah data diberi label, tahap berikutnya adalah visualisasi dalam bentuk *Word Freq* Positif dan *Word Freq* Negatif. Gambar 3 menunjukkan hasil *Word Freq* Positif, sedangkan Gambar 4 menampilkan *Word Freq* Negatif.



Gambar 3. *Word Freq* Positif

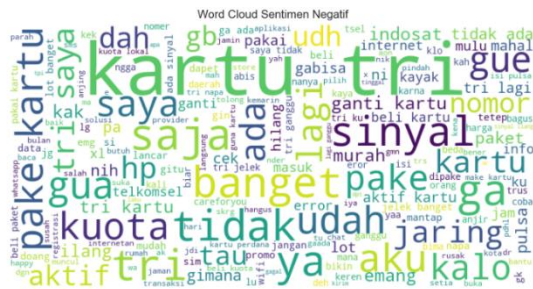


Gambar 4. *Word Freq* Negatif

Selanjutnya, data divisualisasikan dalam bentuk *Word Cloud* untuk masing-masing kelas, positif dan negatif. Gambar 5 menunjukkan *Word Cloud* Positif, sedangkan Gambar 6 menampilkan *Word Cloud* Negatif.



Gambar 5. *Word Cloud* Positif



Gambar 6. Word Cloud Negatif

4.4.2 Naïve Bayes

Pada tahap ini, peneliti akan membagi dataset menjadi data latih (*training*) dan data uji (*testing*). Tabel 8 menunjukkan rasio perbandingan antara data latih dan data uji yang akan digunakan dalam algoritma *Naive Bayes*.

Tabel 9. Perbandingan Data Latih dan Data Uji Naïve Bayes

Data Latih	Data Uji
70%	30%
80%	20%
90%	10%

4.4.3 Support Vector Machine

Pada tahap ini, peneliti akan membagi dataset menjadi data latih (*training*) dan data uji (*testing*). Tabel 9 menunjukkan rasio perbandingan antara data latih dan data uji yang akan digunakan dalam algoritma *Support Vector Machine*.

Tabel 10. Perbandingan Data Latih dan Data Uji Support Vector Machine

Data Latih	Data Uji
70%	30%
80%	20%
90%	10%

4.5. Asses

Tahap terakhir, yaitu *Asses*, merupakan proses evaluasi terhadap setiap model dalam penelitian ini. Hasil evaluasi mencakup nilai akurasi, presisi, *recall*, *F1-score*, *confusion matrix*, dan hasil dari *10-fold cross validation*.

4.5.1 Naïve Bayes

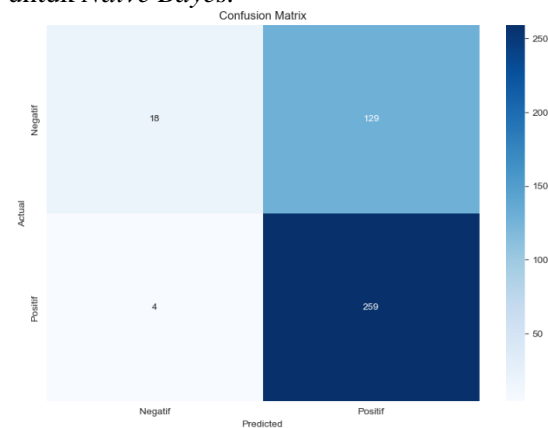
Dalam penelitian ini, metode *Naive Bayes* diterapkan dengan membagi dataset menjadi data latih dan data uji seperti yang dijelaskan pada tahap model. Data kemudian diproses menggunakan metode *Naive Bayes*. Tabel 11

menunjukkan hasil pengujian algoritma *Naive Bayes*.

Tabel 11. Hasil Akurasi, Presisi, *Recall*, *F1-Score* Naive Bayes

DL: DU	Akurasi	Presisi	<i>Recall</i>	<i>F1-Score</i>
70:30	65.69%	73.44%	65.69%	55.51%
80:20	65.98%	72.35%	65.98%	56.32%
90:10	67.56%	72.15%	67.56%	58.68%

Hasil dari *Confusion Matrix* untuk algoritma *Naive Bayes* dengan rasio 90%:10% menunjukkan bahwa terdapat 259 data yang benar terprediksi sebagai sentimen positif (*True Positive*) dan 18 data yang benar terprediksi sebagai sentimen negatif (*True Negative*). Di sisi lain, terdapat 129 data yang salah terprediksi sebagai sentimen positif padahal sebenarnya negatif (*False Positive*), dan 4 data yang salah terprediksi sebagai sentimen negatif padahal sebenarnya positif (*False Negative*). Gambar 7 menampilkan hasil *Confusion Matrix* untuk *Naive Bayes*.



Gambar 7. Confusion Matrix Naive Bayes

Untuk memastikan akurasi dan konsistensi performa algoritma *Naive Bayes*, dilakukan pengujian menggunakan *10-fold cross-validation*. Hasil dari pengujian ini menunjukkan bahwa akurasi mencapai 67.73%, presisi 72.30%, *recall* 67.73%, dan *f1-score* 58.66%.

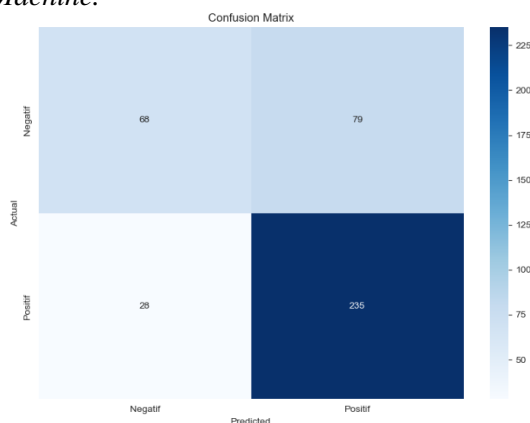
4.5.2 Support Vector Machine

Dalam penelitian ini, metode *Support Vector Machine* diterapkan dengan membagi dataset menjadi data latih dan data uji seperti yang dijelaskan pada tahap model. Data kemudian diproses menggunakan metode *Support Vector Machine*. Tabel 12 menunjukkan hasil pengujian algoritma *Support Vector Machine*.

Tabel 12. Hasil Akurasi, Presisi, Recall, F1-Score SVM

DL: DU	Akurasi	Presisi	Recall	F1-Score
70:30	72.93%	72.75%	72.93%	71.19%
80:20	73.17%	73.16%	73.17%	71.30%
90:10	73.90%	73.40%	73.90%	72.32%

Hasil dari *Confusion Matrix* untuk algoritma *Support Vector Machine* dengan rasio 90%:10% menunjukkan bahwa terdapat 235 data yang benar terprediksi sebagai sentimen positif (*True Positive*) dan 68 data yang benar terprediksi sebagai sentimen negatif (*True Negative*). Di sisi lain, terdapat 79 data yang salah terprediksi sebagai sentimen positif padahal sebenarnya negatif (*False Positive*), dan 28 data yang salah terprediksi sebagai sentimen negatif padahal sebenarnya positif (*False Negative*). Gambar 7 menampilkan hasil *Confusion Matrix* untuk *Support Vector Machine*.



Gambar 8. Confusion Matrix Support Vector Machine.

Untuk memastikan akurasi dan konsistensi performa algoritma *Support Vector Machine*, dilakukan pengujian menggunakan *10-fold cross-validation*. Hasil dari pengujian ini menunjukkan bahwa akurasi mencapai 76.10%, presisi 75.65%, recall 76.10%, dan *f1-score* 75.20%.

4.5.3 Hasil

Sebanyak 4.333 tweet dikumpulkan. Setelah melewati proses *case folding*, *cleaning*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*, jumlah data berkurang menjadi 4.100 karena hanya tweet dengan label sentimen positif dan negatif yang dipertahankan. Selanjutnya, pengujian dilakukan menggunakan algoritma *Naïve Bayes* dan *Support Vector Machine* untuk menilai performa masing-masing dalam

mengklasifikasikan sentimen positif dan negatif. Kinerja algoritma dievaluasi dengan metrik akurasi, presisi, recall, dan *f1-score*. Tabel 12 menunjukkan hasil pengujian dari algoritma *Naïve Bayes* dan *Support Vector Machine* setelah *10-fold cross validation*.

Tabel 13. Hasil 10-fold Cross Validation

10-fold Cross Validation				
Algoritma	Akurasi	Presisi	Recall	F1-Score
Naïve Bayes	67.73%	72.30%	67.73%	58.66%
SVM	76.10%	75.65%	76.10%	75.20%

Hasil dari pengujian *10-fold cross validation* menunjukkan bahwa algoritma *Naive Bayes* memiliki kinerja yang konsisten dengan akurasi rata-rata lebih dari 67%. Sementara itu, algoritma SVM menunjukkan performa yang lebih baik dengan akurasi rata-rata sekitar 76%. Metrik presisi, recall, dan *f1-score* juga menunjukkan hasil yang positif, menunjukkan bahwa kedua algoritma ini efektif dalam mengklasifikasikan sentimen pada tweet mengenai layanan Provider Tri.

5. KESIMPULAN

Berdasarkan hasil analisis sentimen pada X terhadap layanan provider Tri dengan menggunakan algoritma *Naive Bayes* dan *Support Vector Machine*, dapat disimpulkan sebagai berikut:

- Setelah memproses 4.333 data tweet mentah dan melalui tahap preprocessing, diperoleh 4.100 data tweet bersih. Dengan menggunakan metode pelabelan berbasis lexicon, sentimen dari tweet-tweet terkait provider Tri diklasifikasikan ke dalam dua kategori: positif dan negatif. Dari total 4.100 tweet yang dianalisis, sebesar 64,6% atau 2.648 tweet menunjukkan sentimen positif, sementara 35,4% atau 1.452 tweet mengindikasikan sentimen negatif. Hal ini menunjukkan bahwa mayoritas tweet tentang layanan provider Tri cenderung memiliki sentimen positif.
- Berdasarkan hasil evaluasi, algoritma *Naive Bayes* menunjukkan performa dengan akurasi sebesar 67.73%, presisi 72.30%, recall 67.73%, dan *f1-score* 58.66%.

Sementara itu, algoritma Support Vector Machine menghasilkan akurasi 76.10%, presisi 75.65%, recall 76.10%, dan f1-score 75.20%. Hasil perbandingan kinerja kedua algoritma didapatkan akurasi tertinggi pada algoritma *Support Vector Machine* dengan akurasi sebesar 76.10%.

DAFTAR PUSTAKA

- [1] D. Rusdianan and D. Rosiyadi, "Analisa Sentimen terhadap Tokoh Publik Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine," *CESS (Journal Comput. Eng. Syst. Sci.)*, vol. 4, no. 2, pp. 230–235, 2019.
- [2] Y. Akbar and A. N. Ihsan, "Analisis Sentimen Twitter Terhadap Opini Masyarakat Pada Sea Games Kamboja 2023 Menggunakan Algoritma Support Vector Machine," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 6, no. 2, pp. 814–821, 2023, doi: 10.31539/intecom.v6i2.7670.
- [3] W. A. Prabowo and C. Wiguna, "Sistem Informasi UMKM Bengkel Berbasis Web Menggunakan Metode SCRUM," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 149, 2021, doi: 10.30865/mib.v5i1.2604.
- [4] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.
- [5] S. Suryono, E. Utami, and E. T. Luthfi, "Analisis Sentiment Pada Twitter Dengan Menggunakan Mssetode Naive Bayes Classifier," *Proc. Semin. Nas. GEOTIK 2018*, pp. 9–15, 2018, [Online]. Available: <http://hdl.handle.net/11617/9777>
- [6] C. Jovanica, D. D. Rahmintaningrum, H. A. Nuradni, and A. Salsabila, "Analisis Pengaruh Aktor Pada Tagar #Roketchina Di Media Sosial Twitter Menggunakan Social Network Analysis (Sna)," *J. Ilm. Komun. Makna*, vol. 10, no. 1, p. 43, 2022, doi: 10.30659/jikm.v10i1.15644.
- [7] L. Nursinggah, T. Mufizar, and U. Perjuangan, "Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis Dengan Metode Naïve Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024.
- [8] R. A. A. Yanuar, "Jurnal Teknik Informatika, Vol. 16, No. 2, April 2024," vol. 16, no. 2, pp. 1–7, 2024.
- [9] B. Samodera, M. Muharrom, and A. Haromainy, "IMPLEMENTASI MAJORITY VOTE PADA METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE (STUDI KASUS : KENAIKAN PAJAK HIBURAN)," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, pp. 2525–2535, 2024.
- [10] F. F. Irfani, "Analisis Sentimen Review Aplikasi Ruangguru Menggunakan Algoritma Support Vector Machine," *JBMI (Jurnal Bisnis, Manajemen, dan Inform.)*, vol. 16, no. 3, pp. 258–266, 2020, doi: 10.26487/jbmi.v16i3.8607.
- [11] A. Herdhianto, *Sentiment Analysis Menggunakan Naïve Bayes Classifier (NBC) pada Tweet Tentang Zakat*. 2020.
- [12] Y. Azhar, "Metode Lexicon-Learning Based Untuk Identifikasi Tweet Opini Berbahasa Indonesia," *J. Nas. Pendidik. Tek. Inform.*, vol. 6, no. 3, p. 237, 2018, doi: 10.23887/janapati.v6i3.11739.
- [13] Asyrofi, "Jurnal Pendidikan dan Konseling," *J. Pendidik. dan Konseling*, vol. 4, pp. 1349–1358, 2022.
- [14] C. Debora Mait, J. Armando Watuseke, P. David Gibrael Saerang, S. Reynaldo Joshua, and U. Sam Ratulangi, "Sistem Pendukung Keputusan Menggunakan Fuzzy Logic Tahani Untuk Penentuan Golongan Obat Sesuai Dengan," *J. Media Infotama*, vol. 18, no. 2, pp. 344–353, 2022.