

ANALISIS SENTIMEN PENGGUNA TIKTOKSHOP MENGGUNAKAN METODE *NAÏVE BAYES* DAN *SUPPORT VECTOR MACHINE (SVM)*

Fadhlan Ahmad Ghiffary^{1*}, Mohammad Syafrullah²

^{1,2}Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

E-mail : ¹*2011501117@student.budiluhur.ac.id , ²mohammad.syafrullah@budiluhur.ac.id²

(* : corresponding author)

Abstrak- Penelitian ini bertujuan untuk menganalisis sentimen tanggapan masyarakat terhadap aplikasi TikTokShop menggunakan metode *Naïve Bayes* dan *Support Vector Machine (SVM)*. Kontribusi unik dari penelitian ini adalah memberikan perbandingan empiris antara dua metode *machine learning* dalam konteks *e-commerce* yang belum banyak dieksplorasi dalam penelitian sebelumnya. Hasil dari penelitian ini diharapkan dapat memberikan panduan yang lebih baik dalam memahami persepsi pengguna serta meningkatkan strategi pemasaran dan keamanan di *platform TikTokShop*. Data yang digunakan berasal dari berbagai sumber seperti *review*, komentar, dan postingan di media sosial yang relevan dengan *TikTokShop*. Fokus analisis diarahkan pada *klasifikasi* menjadi positif atau negatif, serta kemungkinan untuk mempertimbangkan subkategori atau nuansa yang lebih halus. Studi ini membandingkan efektivitas dua metode pembelajaran mesin, yaitu *Naïve Bayes* dan *SVM*, untuk mengidentifikasi metode yang lebih optimal dalam *klasifikasi* sentimen pengguna *TikTokShop*. Metode *Naïve Bayes* adalah pendekatan sederhana yang mengasumsikan independensi fitur, cocok untuk *klasifikasi* sentimen berdasarkan tokenisasi, penghapusan *stopwords*, dan perhitungan probabilitas kelas sentimen. *SVM*, sebagai metode *supervised learning*, memanfaatkan *hyperplane* optimal untuk memisahkan data menjadi kelas sentimen positif dan negatif. Kedua metode dievaluasi dengan metrik seperti akurasi, presisi, *recall*, dan *F1-score* menggunakan *dataset* terbagi 80:20 untuk data latih dan data uji. Hasil pengujian menunjukkan bahwa metode *Naïve Bayes* memperoleh akurasi 63.08%, presisi 82.35%, *recall* 40%, dan skor *F1* 53.84%. Sementara hasil pengujian *SVM* menunjukkan akurasi 92.30%, presisi 89.47%, *recall* sebesar 97.14% dan skor *F1* 95.58%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa metode *Support Vector Machine (SVM)* lebih cocok untuk *klasifikasi* sentimen pada ulasan pengguna *TikTokShop*. Penelitian ini dapat memberikan wawasan bagi pengembang aplikasi *TikTokShop* dalam memahami persepsi dan tanggapan masyarakat, serta sebagai referensi bagi penelitian selanjutnya dalam *analisis sentimen* menggunakan metode *machine learning*.

Kata Kunci: Analisis Sentimen, *TikTokShop*, *Naïve Bayes*, *Support Vector Machine (SVM)*, *Klasifikasi*

TIKTOKSHOP USER SENTIMENT ANALYSIS USING THE *NAÏVE BAYES* METHOD AND *SUPPORT VECTOR MACHINE (SVM)*

Abstract- This study aims to analyze the sentiment of public responses to the *TikTokShop* application using the *Naive Bayes* and *Support Vector Machine (SVM)* methods. The unique contribution of this study is to provide an empirical comparison between two machine learning methods in the context of *e-commerce* that has not been widely explored in previous studies. The results of this study are expected to provide better guidance in understanding user perceptions and improving marketing and security strategies on the *TikTokShop* platform. The data used comes from various sources such as reviews, comments and posts on social media that are relevant to *TikTokShop*. The focus of the analysis is directed at classification into positive or negative, as well as the possibility to consider subcategories or more subtle nuances. This study compares the effectiveness of two machine learning methods, namely *Naive Bayes* and *SVM*, to identify a more optimal method for classifying *TikTokShop* user sentiment. The *Naive Bayes* method is a simple approach that assumes feature independence, suitable for sentiment classification based on tokenization, removal of *stopwords*, and calculation of sentiment class probabilities. *SVM*, as a supervised learning method, utilizes an optimal *hyperplane* to separate data into positive and negative sentiment classes. Both methods are evaluated with metrics such as accuracy, precision, *recall*, and *F1-score* using a dataset split 80:20 for training data and test data. The test results show that the *Naive Bayes* method obtained 63.08% accuracy, 82.35% precision, 40% *recall*, and 53.84% *F1* score. Meanwhile, the *SVM* test results show an accuracy of 92.30%, precision of 89.47%, *recall* of 97.14% and *F1* score of 95.58%. Based on these results, it can be concluded that the *Support Vector Machine (SVM)* method is more suitable for classifying sentiment in *TikTokShop* user reviews. This research can provide insight for *TikTokShop* application developers in understanding public perceptions and responses, as well as as a reference for further research in sentiment analysis using machine learning methods.

Keyword: Sentiment Analysis, *TikTokShop*, *Naive Bayes*, *Support Vector Machine (SVM)*, Classification

1. PENDAHULUAN

Penelitian ini memberikan kontribusi unik dalam menganalisis sentimen pengguna *TikTokShop* dengan membandingkan dua metode *machine learning* yang berbeda, *Naïve Bayes* dan *Support Vector Machine (SVM)*. Perbandingan ini memberikan wawasan yang lebih mendalam mengenai metode yang lebih efektif dalam *klasifikasi* sentimen dalam konteks *e-commerce*, yang belum banyak dibahas dalam literatur sebelumnya. Temuan ini diharapkan dapat membantu pengembang dalam merancang strategi yang lebih tepat untuk meningkatkan pengalaman pengguna dan memitigasi dampak negatif di *platform TikTokShop*, sebagai fitur *e-commerce* yang terintegrasi dengan aplikasi *TikTok*, telah menawarkan pengalaman belanja yang inovatif dengan memadukan konten *video* dan transaksi *online*. Dalam periode enam bulan terakhir, *platform* ini menunjukkan pertumbuhan yang signifikan dengan peningkatan jumlah pengguna aktif harian sebesar 25% dan *volume* transaksi yang hampir dua kali lipat. Kategori produk seperti fashion, kecantikan, dan gadget mendominasi penjualan, didukung oleh pengaruh *influencer* dan konten kreator yang memainkan peran penting dalam mendorong penjualan. Namun, meskipun *TikTokShop* berkembang pesat, *TikTok* sebagai *platform* utama juga menghadapi berbagai tantangan, terutama terkait dengan penggunaannya yang banyak berada di bawah umur, yang rentan terhadap dampak negatif seperti *bullying* dan perilaku *narcissistic*.

Kolom komentar yang digunakan untuk memberikan penilaian terhadap *video* sering kali dipenuhi dengan kata-kata kasar dan menyerang, yang dapat menyebabkan terjadinya *bullying*. Selain itu, tombol "like" mendorong perilaku *narcissistic* di kalangan pengguna, yang berupaya mendapatkan lebih banyak like sebagai bentuk validasi sosial. Banyaknya *review* negatif tentang *TikTok* di *Google Play Store* menunjukkan adanya kekhawatiran terkait dampak

negatif tersebut, dan ini menekankan pentingnya dilakukan analisis sentimen untuk mengelompokkan *reviews* tersebut menjadi opini negatif atau positif.

Permasalahan ini menjadi kritis karena kurangnya pemahaman mendalam tentang sentimen pengguna dapat berdampak negatif pada pengalaman berbelanja dan mempengaruhi tingkat kepuasan pelanggan, baik di *TikTokShop* maupun di *platform TikTok* secara umum. Oleh karena itu, sangat penting bagi *TikTokShop* dan *TikTok* untuk dapat mengidentifikasi persepsi pengguna secara tepat guna mengembangkan strategi pemasaran, pelayanan, serta kebijakan keamanan yang lebih efektif. Dalam konteks ini, sistem informasi memainkan peran penting sebagai alat yang membantu menyampaikan informasi yang relevan secara efisien, mengelola data secara efektif, dan menghasilkan informasi berkualitas yang dapat mendukung keputusan organisasi [1].

Saat ini, analisis data merupakan topik yang banyak digunakan oleh para peneliti karena melibatkan penggalian informasi dari kumpulan data yang tidak terstruktur. Analisis ini tidak terbatas pada konteks aplikasi *TikTok* dan *TikTokShop*, namun juga dapat diterapkan pada berbagai bidang seperti ekonomi, politik, masyarakat, dan hukum. Dengan mampu mengklasifikasikan polaritas teks, analisis sentimen dapat menentukan apakah isi teks tersebut positif atau *negative* [2]. Tujuan survei ini adalah untuk membantu memahami opini masyarakat terhadap penutupan *TikTokShop* oleh pemerintah. Penelitian ini diharapkan dapat memberikan informasi bagi para pemangku kepentingan dan bermanfaat bagi para sarjana dan peneliti yang ingin menganalisis emosi terkait isu serupa. Perubahan dalam penggunaan internet didorong oleh perkembangan teknologi komunikasi, memungkinkan informasi menyebar dengan cepat melalui media sosial. *TikTok*, aplikasi populer di Indonesia, pernah diblokir pemerintah karena tuduhan pelanggaran seperti pornografi dan pelecehan agama [3].

Melalui penelitian ini, analisis sentimen terhadap interaksi pengguna *TikTokShop* dan ulasan *TikTok* di *Google Play Store* akan dilakukan dengan menggunakan metode *Naïve Bayes* dan *Support Vector Machine (SVM)*. Penelitian ini bertujuan untuk menganalisis sentimen pengguna berdasarkan komentar dan ulasan, membandingkan kinerja kedua metode tersebut dalam klasifikasi sentimen, serta memberikan rekomendasi praktis untuk meningkatkan pengalaman pengguna, strategi pemasaran, dan langkah-langkah untuk memitigasi dampak negatif di *TikTok* dan *TikTokShop*. Dengan demikian, hasil penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai sentimen pengguna, yang pada akhirnya akan membantu *TikTok* dalam meningkatkan kepuasan pelanggan, keamanan pengguna, serta daya saing di pasar *e-commerce*, sekaligus memperkuat sistem informasi yang mendukungnya [4].

Beberapa penelitian sebelumnya telah mengeksplorasi analisis sentimen menggunakan berbagai metode *machine learning*, namun sedikit yang secara khusus memfokuskan pada perbandingan antara *Naïve Bayes* dan *SVM* dalam konteks aplikasi *e-commerce* seperti *TikTokShop*. Sebagai contoh, penelitian lebih fokus pada analisis umum tanpa melihat spesifik konteks atau aplikasi tertentu [9]. Selain itu, metode *Naïve Bayes* dan *SVM* menawarkan kelebihan masing-masing dalam klasifikasi sentimen, namun belum ada studi komprehensif yang membandingkan efektivitas keduanya di bidang ini. Oleh karena itu, penelitian ini berupaya untuk mengisi gap tersebut dengan menawarkan perbandingan empiris antara kedua metode dalam analisis sentimen pengguna *TikTokShop*.

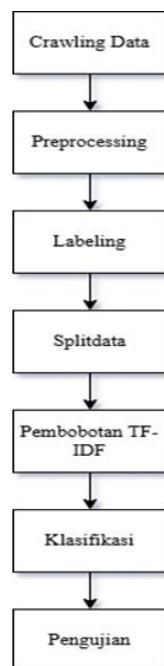
2. METODE PENELITIAN

2.1 Data Penelitian

Data yang digunakan dalam penelitian ini diperoleh melalui proses pengindeksan, yaitu dengan mencari informasi yang tersedia di *internet* atau dengan mengirimkan informasi yang diperoleh berdasarkan perintah tertentu, dan *file* tersebut disimpan sebagai *file CSV*. Kumpulan data ini terdiri dari 406 wawasan yang dikumpulkan dari *Twitter*. Dan penelitian ini menggunakan teknik pengindeksan media sosial di *Twitter* untuk mendapatkan opini pengguna *Twitter*.

2.2 Penerapan Model

Penelitian ini menggunakan beberapa langkah atau proses yang harus dilalui selama penelitian ini. Tujuannya adalah agar tahapan-tahapan tersebut dibuat agar penelitian ini dapat berjalan sesuai rencana.



Gambar 1. Alur Penelitian

Pada Gambar 1 menunjukkan langkah awal pengumpulan data dilakukan dengan mengumpulkan informasi tentang situs jejaring sosial *Twitter* dan diimpor ke *database* untuk diproses melalui sistem analisis sentimen *online*. Pra-pemrosesan melibatkan perubahan *font*, penghapusan *noise*, penggantian kata-kata yang tidak sesuai, dan pemberian label kelas sentimen secara manual. Data yang diberi label diberi pembobotan menggunakan TF-IDF dan dibagi menjadi data latih dan data uji (80:20). Data latih digunakan untuk proses klasifikasi dan pemodelan, sedangkan data uji digunakan untuk menguji keakuratan hasil. Hasil penerapan metode ditampilkan dalam bentuk grafik untuk menunjukkan presisi, akurasi, dan *recall*.

2.3 Pengumpulan Data

Tahap pengumpulan data penelitian ini dilakukan dengan teknik *scraping* dengan menggunakan bahasa pemrograman *Python*. *Dataset* di *scraping* menggunakan kata kunci *auth_token* "*tiktoshop*". Setelah itu, hasil *scraping* disimpan dalam format CSV dan diimpor ke *database* untuk diproses lebih lanjut.

2.4 Pre-Processing

Pada tahap ini, data *Twitter* dikumpulkan dan diproses untuk mendapatkan data yang bersih agar hasil klasifikasi sentimen lebih akurat. *Preprocessing* dilakukan agar proses pengujian tepat. [5]

a. Case Folding

Case Folding adalah proses mengubah kata dari huruf besar/kapital menjadi huruf kecil/non-kapital, seperti "*TIKTOK*" menjadi "*tiktok*".

b. Cleansing

Cleansing adalah dengan membersihkan kata atau menghilangkan karakter pada tweet yang mengandung URL, tag, hashtag dan simbol lainnya, seperti simbol URL <https://t.co/YnI3Mj7Dk6?>, hashtag *#moneyzone* dan simbol lainnya [6].

c. Normalisasi

Normalisasi dilakukan setelah tahap *cleansing* untuk mengubah kata-kata slang, tidak baku, atau singkatan menjadi kata baku. Proses ini melibatkan kamus bahasa gaul dan kata-kata yang diteliti secara manual terkait *TikTokShop*, seperti mengubah "u" menjadi "kamu" [7].

d. Stopword

Stopword adalah menghapus kata dari dokumen dengan bantuan daftar kamus kata bahasa Indonesia. Contohnya, kata "ada" dihapus karena termasuk *stop word* yang sering muncul [8].

e. Stemming

Stemming adalah mengubah kata menjadi bentuk sebenarnya dengan menghilangkan semua akhiran dan awalan, seperti menyelam menjadi selam dan pembayaran menjadi bayar.

2.5 Labeling

Pada langkah pelabelan ini, tag atau kategori ditetapkan berdasarkan fitur-fitur yang terdapat dalam suatu dokumen atau kalimat. Label positif dan negatif diberikan pada tweet yang diperoleh pada fase *scraping*, dengan tag positif untuk *tweet* yang mendukung *TikTokShop* dan tag negatif untuk *tweet* yang tidak mendukung. Proses pelabelan dilakukan secara manual oleh pemilik toko sebagai ahli, yang menentukan kategori positif atau negatif untuk setiap *tweet*. Ini merupakan proses kategorisasi berdasarkan subjektivitas manusia sesuai dengan teori yang dijelaskan sebelumnya.

2.6 Split Data

Pembagian data menjadi 80:20 dalam metode *Naïve Bayes* dan *Support Vector Machine* (SVM) adalah praktik standar dalam pembelajaran mesin. Rasio ini mengalokasikan 80% data untuk pelatihan dan 20% untuk pengujian, sehingga model dapat belajar dengan baik tanpa risiko *overfitting* atau *underfitting*. Dengan menggunakan sebagian besar data untuk pelatihan, model dapat memahami pola dengan baik, sementara data uji yang cukup besar memberikan evaluasi yang akurat. Praktik ini menghasilkan hasil evaluasi yang konsisten, efisien secara komputasi, dan memastikan kemampuan generalisasi model yang baik. Sehingga, pembagian data ini penting untuk memastikan validitas dan kinerja model yang baik dalam pembelajaran mesin.

2.7 Pembobotan TF-IDF

Proses pembobotan TF-IDF bertujuan untuk mengetahui hubungan antara setiap *tweet* dalam *dataset*. Setelah melalui tahap pra-pemrosesan, kata-kata unik dari setiap *tweet* dihitung TF-nya. Selanjutnya, nilai IDF dihitung dengan membagi jumlah total *tweet* dengan jumlah *tweet* yang mengandung kata tersebut. Akhirnya, nilai TF-IDF dihitung dengan hasil penjumlahan TF dan IDF, untuk kemudian digunakan dalam data latih dan data uji [9]. Formula ini dapat ditunjukkan pada persamaan 1

$$W_{ij} = tf_{ij} \times \log \frac{D}{idf_i} \quad (1)$$

Keterangan

W_{ij} = Bobot TF-IDF

idf_i = *Inverse Dokumen Frequency*

tf_{ij} = Frekuensi Suatu Kata

2.8 Klasifikasi Naïve Bayes

Naïve Bayes Classifier (NBC) adalah algoritma klasifikasi berdasarkan teorema *Bayes*. Pengklasifikasi *Naïve Bayes* berkinerja sangat baik dibandingkan dengan model klasifikasi lain seperti pohon keputusan atau jaringan saraf. Kelebihan metode ini adalah hanya diperlukan sedikit data latih untuk menentukan parameter yang dibutuhkan dalam proses klasifikasi. Formula ini dapat ditunjukkan pada persamaan 2

Rumus :

$$P(A | B) = P(B | A) P(A) P(B) \quad (2)$$

Keterangan:

- A = Data dengan kelas yang belum diketahui
- B = Hipotesis bahwa data A adalah kelas khusus
- P(A|B) = Probabilitas hipotesis A berdasarkan kondisi B (probabilitas posterior)
- P(B|A) = Probabilitas hipotesis B (probabilitas prior)
- AP(B) = Probabilitas hipotesis B berdasarkan kondisi pada hipotesis

2.9 Klasifikasi Support Vector Machine

Support Vector Machine (SVM) merupakan metode untuk menganalisis data dan mengidentifikasi pola yang sesuai untuk klasifikasi. Penelitian ini berfokus pada klasifikasi biner yang terdiri dari dua kelas positif atau negatif. SVM merupakan metode klasifikasi yang memisahkan data secara optimal dengan membangun *hyperplane* dalam ruang input multidimensi. *Hyperplane* terbaik dapat ditentukan dengan mengukur anggota batas kelas. Awalnya digunakan untuk mengklasifikasikan data numerik, SVM juga efisien dan cepat dalam menangani data teks. Data teks cocok untuk klasifikasi dengan algoritma SVM karena kernel teks biasanya besar dan beberapa fitur mungkin tidak penting, tetapi cenderung berkorelasi dan biasanya dibagi ke dalam kelas yang berbeda secara terpisah [10].

Formula Hyperlane klasifikasi linear pada *Support Vector Machine* ditunjukkan pada persamaan 3 sebagai berikut:

$$f(x) = w \cdot x + b \quad (3)$$

Dari persamaan tersebut, diperoleh pertidaksamaan untuk kelas +1 (positif) ditunjukkan pada persamaan 4 sebagai berikut:

$$w \cdot x + b \leq +1 \quad (4)$$

Sementara itu, pertidaksamaan untuk kelas -1 ditunjukkan pada persamaan 5 sebagai berikut:

$$w \cdot x + b \leq -1 \quad (5)$$

Di sini, w adalah bobot, b adalah bias, dan x adalah jumlah data.

2.10 Pengujian Evaluasi

Evaluasi Pengujian pada penelitian ini untuk mengevaluasi hasil dari algoritma yang digunakan. Metode evaluasi yang digunakan *confusion matrix* metode ini cukup untuk membantu proses analisis *classifie*. Matriks ini memberikan gambaran detail performa model dengan membandingkan prediksi model dengan nilai sebenarnya. *Confusion matrix* terdiri dari empat komponen utama yaitu *True Positif* (TP), *False Positive* (FP), *True Negative* (TN) dan *False Negative* (FN). Berikut ini penjelasan detail masing-masing komponen beserta penafsirannya:

- a. *True Positive* (TP): Jumlah prediksi positif yang benar.
- b. *False Positive* (FP): Jumlah prediksi positif yang salah.
- c. *True Negative* (TN): Jumlah prediksi negatif yang benar.
- d. *False Negative* (FN): Jumlah prediksi negatif yang salah.

Rumus untuk menghitung skor presisi, *recall*, presisi dan *F1-score* pada *confusion matrix* adalah sebagai berikut:

Formula *accuracy* dapat ditunjukkan pada persamaan 6 (6)

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

Formula *recall* dapat ditunjukkan pada persamaan 7 (7)

$$Recall = \frac{Tren\ Positive}{(True\ Positive + False\ Negative)}$$

Formula *precision* dapat ditunjukkan pada persamaan 8 (8)

$$Precision = \frac{True\ Positives}{(True\ Positives + False\ Positives)}$$

Formula *F1-Score* dapat ditunjukkan pada persamaan 9 (9)

$$F1 - Score = \frac{2 \times (Precision \times Recall)}{(Precision + Recall)}$$

3. HASIL DAN PEMBAHASAN

Pada penelitian ini menentukan dataset yang digunakan, yang diambil dengan scraping menggunakan kata kunci "auth_token tiktokshop". Hasilnya disimpan dalam format CSV dan diimpor ke database untuk diproses lebih lanjut.

Tabel 1. Tabel Data Hasil *Crawling* Data

	<i>full text</i>	<i>Username</i>
0	Kemaren ada liat frame bagus tapi mahal...tap...	Ehemajadeh

1	WTS ACC TIKTOK F23.7K LIKES 272.7K https://t.c...	Fazcys
2	siang <i>ready</i> akun <i>tiktok</i> murah udah bisa <i>live</i> d...	Pibeystore
3	siang <i>ready</i> akun <i>tiktok</i> murah udah bisa <i>live</i> d...	Definnisi
4	Coba cek keranjang kuningnya di tiktokshop kap...	LeikaSkincare

Tabel 1 menunjukkan data hasil *crawling* data. Data yang baru dikumpulkan dari *twitter* sering kali masih mengandung simbol, tautan, dan elemen lainnya yang tidak relevan. Oleh sebab itu, proses *preprocessing* diperlukan untuk membersihkan data dan memastikan konsistensinya sebelum analisis lebih lanjut. Hasil dari *preprocessing* ini dapat dilihat pada Gambar 2, yang memperlihatkan data setelah dibersihkan dan siap untuk dianalisis.

Tweet sebelum	Tweet sesudah
Kemaren ada liat frame bagus tapi mahal... tapi bagus bgt naksir. Tapi sadar diri kalo kacamata ga oernah awet. Akhirnya ke tiktokshop beli yg 100rb an. Bagus juga lumayan. Lensanya juga enak.. https://t.co/cc94Y2VZxU	kemarin lihat frame bagus mahal bagus banget naksir sadar diri kalau kacamata oernah awet akhir tiktokshop beli 100rb an bagus lumayan lensa enak
WTS ACC TIKTOK F23.7K LIKES 272.7K https://t.co/Y4DXvSrCo6 - bind email no google - bisa live livestudio tiktokshop tiktok affiliate - bekas konten game mobile legend - cocok untuk akun apa aja 6xxk #zonauang https://t.co/NvZFbnHYC	wts sesuai tiktok f237k likes 2727k bind email nomor google live livestudio tiktokshop tiktok affiliate bekas konten game mobile legend cocok akun apa 6xxk
siang ready akun tiktok murah udah bisa live dan tiktokshop wts wtb need acc jual beli akun tiktok #zonauang https://t.co/XiDkCO43Ya	siang ready akun tiktok murah bisa live tiktokshop wts wtb need sesuai jual beli akun tiktok

Gambar 2. Hasil *Preprocessing*

Setelah proses *preprocessing*, dilakukan penandaan manual untuk memastikan data bersih dan siap untuk diberi tag. Tag positif menunjukkan dukungan dan kepercayaan terhadap *TikTokshop*, sedangkan tag negatif mencerminkan penolakan atau ketidakpercayaan. Proses ini penting untuk mengklasifikasikan *tweet* ke dalam kategori yang sesuai, dan hasil manual dari langkah ini disajikan dalam tabel 2.

Tabel 2. Pelabelan Data *Tweet*

No	Username	Tweet	Label
1	_icecreamcorn	kirain aku kak pas pesan <i>tiktokshop</i> lewat waktu estimasi aku gapernah chat seller lama <i>online shop</i> akhir gedeg buat kabarin sellernya	Negatif
2	Slypzier	enakan <i>tiktokshop</i> mana banyak <i>disc</i> nya lg	Positif
3	Holistifx	gembel deh <i>tiktokshop</i> gada gratis ongkos kirim	Negatif
4	Ausswie	ya allah ni murah bgt sunscreen spray jinot dapet 2 tp aku udah pesen duluan dan udah dikirim aaakk (di <i>tiktokshop</i>)	Positif

Setelah tahap labelling selesai, proses selanjutnya adalah *split data*. Dari total 406 *tweet* yang dilabeli, data ini dibagi dengan rasio 80:20, menghasilkan 324 *tweet* sebagai data training dan 82 *tweet* sebagai data testing. Sebagai ilustrasi, Tabel 3 di bawah ini menyajikan contoh data dari *data training* dan *data testing* yang telah dihasilkan.

Tabel 3. Dataset

Data Training	Data Testing
pertama kali jajan tiktokshop moga enak tidak nyangka weh diskon lumayan mah coba cek keranjang kuning <i>tiktokshop</i> kapan harga segitu dapat 2 produk beli sekarang klik sini	pagi peeps oreo jual akun <i>tiktok followers</i> indo sudah <i>live tiktokshop</i> yaaa wts wtb need sesuai jual beli akun <i>tiktok</i> butuh saldo mandiri buat bayar <i>tiktokshop</i> nyata bisa ubah metode bayar 189600 <i>drop fee</i>

Dataset ini akan dianalisis menggunakan algoritma *Naïve Bayes Classifier* dan *Support Vector Machine* untuk mengklasifikasikan sentimen, dengan data pelatihan yang disimpan menggunakan metode TF-IDF sebagai bobot untuk setiap kata, dapat dilihat pada tabel 4. Setelah *preprocessing*, pelabelan dan *splidata*, data yang discrap menjadi terorganisir. Kata-kata diubah menjadi angka melalui TF-IDF untuk memudahkan proses klasifikasi, ditunjukkan dalam tabel 5.

Tabel 4. Data Sampel

No.	Tweet Data Sampel
1.	Pertama kali jajan tiktokshop moga enak tidak nyangka weh diskon lumayan mah
2.	Cob acek keranjang kuning tiktokshop kapan harga segitu dapat 2 produk beli sekarang klik sini

Langkah pertama dalam pembobotan kata adalah menghitung bobot TF (*term* frekuensi) untuk menentukan kemunculan setiap kata dalam dokumen, yang kemudian dibagi dengan jumlah kata total di dokumen, ditunjukkan dalam tabel 5.

Tabel 5. Perhitungan TF

Kata	Frekuensi Kemunculan	Jumlah Total Kata	TF
Pertama	1	12	0.08333
Kali	1	12	0.08333
Jajan	1	12	0.08333
Tiktokshop	1	12	0.08333
Moga	1	12	0.08333
Enak	1	12	0.08333
Tidak	1	12	0.08333
Nyangka	1	12	0.08333
Weh	1	12	0.08333
Diskon	1	12	0.08333
Lumayan	1	12	0.08333
Mah	1	12	0.08333
Coba	0.8	12	0.06667
Cek	0.8	12	0.06667
Keranjang	0.8	12	0.06667
Kuning	0.8	12	0.06667
Tiktokshop	0.8	12	0.06667
Kapan	0.8	12	0.06667
Harga	0.8	12	0.06667
Segitu	0.8	12	0.06667
Dapat	0.8	12	0.06667
2	0.8	12	0.06667
Produk	0.8	12	0.06667
Beli	0.8	12	0.06667
Sekarang	0.8	12	0.06667
Klik	0.8	12	0.06667
Sini	0.8	12	0.06667

Formula Perhitungan manual TF ditunjukkan pada persamaan 10:

$$TF(w) = \text{Frekuensi kemunculan kata } w / \text{Total jumlah kata} \quad (10)$$

Contoh perhitungan untuk kata "pertama":

Frekuensi kemunculan kata "pertama" = 1

Total jumlah kata = 12

$TF(\text{pertama}) = 1 / 12 = 0.08333$

Setelah menghitung TF, langkah selanjutnya adalah menghitung bobot IDF untuk mengidentifikasi kata-kata penting yang membedakan dokumen. Ini bertujuan meningkatkan kinerja sistem dalam temu kembali informasi secara efektif, dapat dilihat pada tabel 6.

Tabel 6. Perhitungan IDF

Word	n(w)	IDF (manual)
Pertama	1	5.78074
Kali	1	4.39445
Jajan	1	4.17131
Tiktokshop	3	0.0803
Moga	1	4.39445
Enak	1	5.0876
Tidak	1	2.89037
Nyangka	1	5.78074
Weh	1	5.78074
Diskon	1	3.58352
Lumayan	1	4.39445
Mah	1	4.39445
Coba	1	3.83483
Cek	1	3.7013
Keranjang	1	3.98898
Kuning	1	4.39445
Kapan	1	4.68213
Harga	1	3.38285
Segitu	1	5.78074
Dapat	1	4.17131
2	1	4.39445
Produk	1	4.68213
Beli	2	1.24814
Sekarang	1	3.83483
Klik	1	5.0876
Sini	1	5.0876

Formula Perhitungan manual IDF ditunjukkan pada persamaan 11:

$$IDF(w) = \log(N/n(w)) \quad (11)$$

Dimana:

N = Total jumlah dokumen (baris)

n(w) = Jumlah dokumen yang mengandung kata w

Contoh perhitungan untuk kata "pertama":

$N = 84$ (total jumlah dokumen)

$n(w) = 1$ (kata "pertama" muncul di 1 dokumen)

$IDF(\text{pertama}) = \log(84/1) = 5.78074$

Nilai IDF digunakan untuk menghitung bobot kata (TF-IDF) dengan mengalikan nilai TF dan IDF, terlihat pada tabel 7.

Tabel 7. Perhitungan TF-IDF

Kata	TF	IDF	TF-IDF
Pertama	0.08333	5.78074	0.48173
Kali	0.08333	4.39445	0.3662
Jajan	0.08333	4.17131	0.34761
Tiktokshop	0.08333	0.0803	0.00669
Moga	0.08333	4.39445	0.3662
Enak	0.08333	5.0876	0.42397
Tidak	0.08333	2.89037	0.24086
Nyangka	0.08333	5.78074	0.48173
Weh	0.08333	5.78074	0.48173
Diskon	0.08333	3.58352	0.29863
Lumayan	0.08333	4.39445	0.3662
Mah	0.08333	4.39445	0.3662
Coba	0.06667	3.83483	0.25566
Cek	0.06667	3.7013	0.24675
Keranjang	0.06667	3.98898	0.26593
Kuning	0.06667	4.39445	0.29296
Tiktokshop	0.06667	0.0803	0.00535
Kapan	0.06667	4.68213	0.31214
Harga	0.06667	3.38285	0.22552
Segitu	0.06667	5.78074	0.38538
Dapat	0.06667	4.17131	0.27809
2	0.06667	4.39445	0.29296
Produk	0.06667	4.68213	0.31214
Beli	0.06667	1.24814	0.08321
Sekarang	0.06667	3.83483	0.25566
Klik	0.06667	5.0876	0.33917
Sini	0.06667	5.0876	0.33917

Formula Rumus perhitungan TF-IDF ditunjukkan pada persamaan 12:

$$TF\text{-}IDF = TF \times IDF \quad (12)$$

Dimana:

TF (*Term Frequency*) = Frekuensi kemunculan suatu kata dalam dokumen.

IDF (*Inverse Document Frequency*) = Alogaritma natural dimana jumlah dokumen dibagi dengan jumlah dokumen yang mengandung kata tersebut

Dari *dataset* ini dianalisis menggunakan algoritma *naïve bayes* melibatkan penggunaan model untuk memprediksi kelas data baru berdasarkan pembelajaran dari data latih. Untuk menghitung probabilitas prior dan kondisional, diperlukan informasi dari *dataset*. Evaluasi *naïve bayes* didasarkan pada akurasi yang dicapai, yang ditentukan melalui *confusion matrix*. Pengujian *naïve bayes* mengukur efektivitas model dalam memprediksi kelas dari data yang tidak terlihat sebelumnya, dengan hasil disajikan dalam matriks konfusi 2×2 untuk analisis lebih lanjut, dapat dilihat pada gambar 3.

```

Evaluation Metrics for Naive Bayes on Test Data:
Accuracy: 63.08%
Average Precision: 82.35%
Average Recall: 40%
Average F1 Score: 53.84%

Confusion Matrix for Naive Bayes on Test Data:
Predicted Labels
Actual -1      27      3
Labels 1      21      14
  
```

Gambar 3. *Naïve Bayes*

Berdasarkan *confusion matrix* yang ditampilkan *Naïve Bayes Classifier*, *Naïve Bayes Classifier* berhasil mencapai akurasi 63.08%, presisi 82.35%, recall 40%, dan skor F1 53.84%.

Dari *dataset* ini akan dianalisis menggunakan algoritma *Support Vector Machine* (SVM) dengan model pembelajaran mesin yang mencari *hyperplane* dalam ruang N-dimensi untuk memisahkan kelas data. SVM efektif untuk klasifikasi linier dan non-linier. Pengujian SVM mengevaluasi performa model dengan menyajikan hasil dalam tabel *confusion matrix* 2×2 dapat dilihat pada gambar 4.

Evaluation Metrics for SVM on Test Data: Accuracy: 92.30% Average Precision: 89.47% Average Recall: 97.14% Average F1 Score: 95.58%
Confusion Matrix for SVM on Test Data: Actual/Predicted - 1 1 0 26 4 1 1 34

Gambar 4. Support Vector Machine

Berdasarkan *confusion matrix* yang ditampilkan, *Support Vector Machine*. *Support Vector Machine* berhasil mencapai akurasi 92.30%, presisi 89.47%, recall 97.14%, dan skor F1 95.58%.

4. KESIMPULAN

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa metode *Support Vector Machine* (SVM) lebih akurat dalam mengklasifikasikan emosi dan tanggapan manusia terhadap aplikasi TikTokShop daripada metode *Naïve Bayes*. Hasil analisis sentimen juga memberikan informasi berharga kepada pengembang untuk meningkatkan kualitas layanan dan strategi pemasaran yang lebih efektif. Ini menunjukkan pentingnya penggunaan data dan algoritma dalam mengoptimalkan pengalaman pengguna. Untuk penelitian selanjutnya, penulis memberikan saran untuk mengintegrasikan hasil analisis sentimen dengan strategi bisnis dan pengembangan produk *TikTokShop* dan menerapkan teknik *ensemble learning* dengan mengkombinasikan beberapa metode untuk meningkatkan akurasi.

DAFTAR PUSTAKA

- [1] Rahel Lina Simanjuntak, Theresia Romauli Siagian, Vina Anggriani, and Arnita Arnita, "Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Shopee Dengan Menggunakan Algoritma Naïve Bayes," *Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 3, no. 3, pp. 23–39, 2023, doi: 10.55606/teknik.v3i3.2411.
- [2] C. Dwi Angelina, "Penerapan Algoritma Naive Bayes Classifier Pada Analisis Sentimen Masyarakat Indonesia Terhadap Childfree Pada Twitter," *Prosiding Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI) Ke-3*, 30 Agustus 2023-Jakarta, vol. 2, no. 2, 2023, pp. 398-407.
- [3] S. Fide, S. Suparti, and S. Sudarno, "Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Play Menggunakan Metode Support Vector Machine (Svm) Dan Asosiasi," *Jurnal Gaussian*, vol. 10, no. 3, pp. 346–358, 2021, doi: 10.14710/j.gauss.v10i3.32786.
- [4] J. Alfiah Zulqornain and P. Pandu Adikara, "Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 7, pp. 2886–2890, 2021.
- [5] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, "Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM)," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 375–384, 2024, doi: 10.57152/malcom.v4i2.1206.
- [6] R. Sanjaya, E. Tohidi, E. Wahyudi, and K. Kaslani, "Analisis Sentimen Terhadap Berhentinya Tiktokshop Pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 507–514, 2024, doi: 10.36040/jati.v8i1.8443.
- [7] S. S. Hasibuan, Angraini, E. Saputra, and Megawati, "Jurnal Media Informatika Budidarma Sentimen Analisis Terhadap Fitur Tiktok Shop Menggunakan Naïve Bayes dan K-Nearest Neighbor," *Jurnal Media Informatika Budidarma*, vol. 8, no. 1, pp. 303–311, 2024, doi: 10.30865/mib.v8i1.7238.
- [8] Friska Aditia Indriyani, Ahmad Fauzi, and Sutan Faisal, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 176–184, 2023, doi: 10.37373/tekn.v10i2.419.
- [9] S. Berliani and S. Lestari, "Analisis Sentimen Masyarakat Terhadap Isu Pecat Sri Mulyani Pada Twitter Menggunakan Metode Naive Bayes Dan Support Vector Machine," *Jurnal Sains dan Teknologi*, vol. 5, no. 3, pp. 951–960, 2024, doi: 10.55338/saintek.v5i3.2746.
- [10] T. T. Widowati and M. Sadikin, "Analisis Sentimen Twitter terhadap Tokoh Publik dengan Algoritma Naive Bayes dan Support Vector Machine," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 11, no. 2, pp. 626–636, 2021, doi: 10.24176/simet.v11i2.4568.