ANALISIS SENTIMEN KUALITAS BAHAN BAKAR PERTAMINA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE PADA PLATFORM X

Julius Omegatha Lamanda[1], Faisal Raihan Ibnu Nasrulloh[2], Jefri Setyawan[3], Shofyan Nur Addin[4], Achmad Yogi Pamungkas[5], Fuad Nur Hasan[6]

[1,2,3,4,5,6]Universitas Bina Sarana Informatika
Jalan Kramat Raya No.98, Senen, Jakarta Pusat

Email : [1] juliuslamanda@gmail.com, [2] faisalraihan108@gmail.com,[3] jefristywn195@gmail.com, [4] shofyanaddin17@gmail.com, [5] achmadyogi04@gmail.com, [6] fuad.fnu@bsi.ac.id

ABSTRAK

Kualitas bahan bakar merupakan faktor penting yang memengaruhi performa mesin dan kepuasan pelanggan. Sebagai penyedia utama bahan bakar di Indonesia, Pertamina perlu memastikan bahwa produk yang dipasarkan memenuhi harapan masyarakat. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap kualitas bahan bakar Pertamina berdasarkan data yang diperoleh dari Platform X (Twitter) dengan menggunakan algoritma Support Vector Machine (SVM) sebagai metode klasifikasi utama. Data dikumpulkan melalui proses crawling dengan menggunakan pustaka tweet-harvest versi 2.6.1 pada rentang waktu 1 April 2025 hingga 30 September 2025, menghasilkan 1.596 komentar. Data kemudian melalui tahap preprocessing meliputi cleaning, case folding, normalisasi, tokenizing, stopword removal, dan stemming. Pelabelan sentimen dilakukan secara otomatis menggunakan pendekatan lexicon-based dengan dua kategori, yaitu positif dan negatif, dengan hasil 799 data positif dan 795 data negatif. Proses klasifikasi menggunakan SVM dengan rasio data latih 80% dan data uji 20%. Berdasarkan hasil pengujian, model SVM menunjukkan performa yang efektif dan stabil dalam mengklasifikasikan sentimen teks berbahasa Indonesia pada konteks opini publik terhadap produk Pertamina.

Kata Kunci*:* *Analisis Sentimen, Support Vector Machine, Bahan Bakar Pertamina, Media Sosial X.*

ABSTRACT

Fuel quality is an important factor that influences engine performance and customer satisfaction. As the main fuel provider in Indonesia, Pertamina must ensure that its products meet public expectations. This study aims to analyze public sentiment toward the quality of Pertamina's fuel products based on data obtained from Platform X (Twitter), using the Support Vector Machine (SVM) algorithm as the primary classification method. Data were collected through a crawling process using the tweet-harvest library version 2.6.1 from April 1, 2025, to September 30, 2025, resulting in 1,596 comments. The data then underwent preprocessing, including cleaning, case folding, normalization, tokenizing, stopword removal, and stemming. Sentiment labeling was performed automatically using a lexicon-based approach with two categories—positive and negative—resulting in 799 positive and 795 negative data points. The classification process was carried out using SVM with an 80% training and 20% testing data split. Based on the evaluation results, the SVM model demonstrated effective and stable performance in classifying Indonesian-language text sentiment in the context of public opinion toward Pertamina's fuel products.

Keywords: *Sentiment Analysis, Support Vector Machine, Pertamina Fuel Quality, Social Media X.*

1. PENDAHULUAN

Industri bahan bakar memegang peranan penting dalam mendukung pertumbuhan ekonomi nasional, karena kualitas bahan bakar berpengaruh langsung terhadap kinerja mesin dan kepuasan pengguna. Saat ini, BBM menjadi kebutuhan utama masyarakat untuk menunjang aktivitas sehari-hari, khususnya

sebagai sumber energi bagi transportasi. Pertamina, sebagai perusahaan BUMN yang berfokus pada pengelolaan minyak dan gas bumi, berperan memproduksi sekaligus mendistribusikan bahan bakar yang dibutuhkan oleh masyarakat Indonesia [1].

Sebagai perusahaan BUMN utama penyedia bahan bakar di Indonesia, Pertamina menghadapi tantangan besar dalam pemantauan dan pemeliharaan kualitas produknya agar sesuai dengan harapan masyarakat. Kemajuan teknologi digital dan peningkatan penggunaan media sosial membuka kesempatan untuk menerapkan analisis sentimen guna mengumpulkan langsung opini konsumen tentang kualitas bahan bakar melalui Platform X.

Sebelum resmi menjadi PT Pertamina (Persero), pada sekitar tahun 1950, Pemerintah Republik Indonesia menunjuk Angkatan Darat untuk mendirikan PT Eksploitasi Tambang Minyak Sumatera Utara yang mengelola ladang minyak di wilayah Sumatera. Pada 10 Desember 1957, perusahaan ini berganti nama menjadi PT Perusahaan Minyak Nasional atau PERMINA, dan tanggal tersebut diperingati sebagai hari lahir Pertamina hingga kini. Kemudian, pada 1 Juli 1961, PT Pertamina berubah menjadi PN Permina, dan pada 20 Agustus 1968, PN Permina dan PN Pertamina digabung menjadi PN Pertamina [2].

Dalam bidang pembelajaran mesin, algoritma Support Vector Machine (SVM) sering digunakan untuk tugas klasifikasi teks, termasuk dalam analisis sentimen, karena kemampuannya dalam memisahkan kelas secara optimal. Penelitian ini bertujuan untuk membandingkan kinerja dua algoritma dalam menganalisis sentimen terkait kualitas bahan bakar Pertamina di Platform X, guna memberikan pemahaman yang komprehensif tentang persepsi masyarakat sekaligus menentukan metode pengolahan data sentimen yang paling efektif untuk menunjang peningkatan produk dan layanan Pertamina.

Aplikasi X merupakan salah satu platform media sosial yang digunakan untuk berkirim pesan, mengekspresikan pendapat, serta berbagi informasi terkini, mulai dari tren, isu, hingga berbagai wacana lainnya. Aplikasi ini sebelumnya dikenal dengan nama Twitter sebelum berganti merek menjadi X [3].

2. TINJAUAN PUSTAKA

Analisis sentimen merupakan bagian dari pemrosesan bahasa alami (NLP) yang fokus pada pendeteksian, pengambilan, dan pengelompokan opini dalam teks ke dalam kategori sentimen positif, negatif, atau netral. Metode ini mengalami perkembangan pesat terutama seiring dengan bertambahnya data dari media sosial dan platform ulasan, sehingga organisasi dapat memperoleh wawasan mendalam tentang pandangan publik terhadap produk dan layanan mereka. Dalam konteks aplikasi mobile, analisis sentimen berperan penting untuk menilai kinerja dan popularitas aplikasi berdasarkan ulasan dari para penggunanya [4].

2.1 Platform X sebagai Sumber Data

X merupakan salah satu media sosial dan platform komunikasi yang paling populer di dunia. Aplikasi ini memungkinkan pengguna untuk mengunggah pesan singkat di halaman pribadi dengan batas maksimal 280 karakter. Selain berfungsi sebagai alat komunikasi, X juga digunakan sebagai media informasi, platform bisnis, sarana penggerak opini publik, serta media hiburan [5].

2.2 Google Colab

Google Colab merupakan platform komputasi berbasis cloud yang dikembangkan oleh Google. Platform ini memungkinkan pengguna menjalankan kode Python langsung di lingkungan cloud tanpa instalasi atau konfigurasi pada perangkat lokal mereka. Google Colab sering dimanfaatkan oleh ilmuwan data, peneliti, serta pengembang untuk tugas seperti pemrosesan data, pengembangan model kecerdasan buatan, analisis data, dan pelatihan model machine learning [5].

2.3 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah teknik dalam pembelajaran mesin yang didasarkan pada teori pembelajaran statistik struktural. Algoritma ini umumnya menunjukkan performa lebih unggul dibandingkan metode pembelajaran mesin lainnya. Proses utama dalam SVM adalah menemukan

hyperplane atau batas yang memisahkan setiap kelas data secara optimal. SVM beroperasi dengan mencari *hyperplane* terbaik yang dapat memisahkan data ke dalam kelas tertentu dengan margin maksimal, menggunakan vektor dukungan (*support vectors*) untuk meningkatkan akurasi prediksi [6].

2.4 Analisis Sentimen

Analisis sentimen adalah teknik yang digunakan untuk mengenali dan mengelompokkan opini dalam teks ke dalam kategori sentimen positif atau negative. Salah satu tantangan utama dalam analisis sentimen adalah tahap *preprocessing* data teks, yang mencakup pembersihan teks, normalisasi, serta transformasi teks mentah menjadi format yang dapat diproses oleh mesin. Tahap *preprocessing* ini sangat krusial karena kualitasnya secara langsung memengaruhi tingkat akurasi model analisis sentimen [7].

2.5 Data Mining

Data mining merupakan proses penggalian pola, hubungan, atau informasi bernilai dari kumpulan data besar menggunakan teknik analitik, metode statistik, serta algoritma kecerdasan buatan. Tahapan utamanya mencakup pengumpulan data relevan, pembersihan atau preprocessing untuk mengatasi inkonsistensi dan nilai hilang, transformasi data sesuai kebutuhan analisis, penerapan algoritma untuk mengidentifikasi pola signifikan, serta interpretasi hasil guna mendukung pengambilan keputusan yang lebih baik [8].

2.6 Penelitian Terdahulu

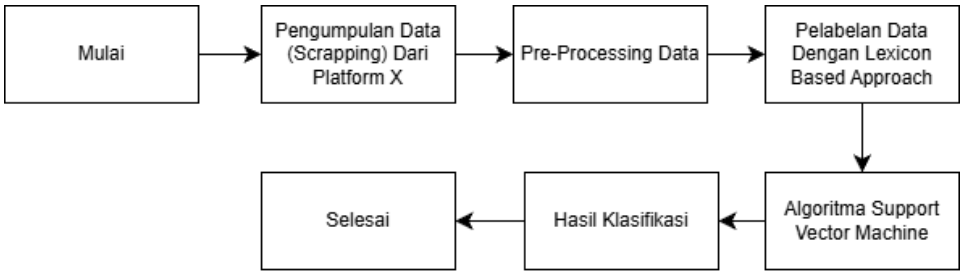
1. Penelitian [9] Menganalisis sentimen masyarakat terhadap kenaikan harga bahan bakar minyak menggunakan Support Vector Machine (SVM) dengan teknik SMOTE untuk menangani ketidakseimbangan data. Hasil menunjukkan akurasi 94% untuk sentimen negatif sebelum SMOTE dan meningkat menjadi 84% secara keseluruhan setelah SMOTE diterapkan. Penelitian ini membuktikan efektivitas kombinasi SVM dan SMOTE dalam mengklasifikasikan sentimen yang didominasi oleh sentimen negatif.

2. Penelitian [5] Menganalisis sentimen tweet di platform X terhadap layanan Provider Tri menggunakan Naive Bayes dan SVM dengan metodologi SEMMA pada 4.333 data (2023-2024). SVM unggul dengan akurasi 76,10%, presisi 75,65%, recall 76,10%, dan F1-score 75,20 (rasio 90:10, 10-fold CV), sementara mayoritas sentimen positif (64,6%).

3. Penelitian [10] Melakukan analisis sentimen pada 3000 ulasan aplikasi Shopee menggunakan algoritma Support Vector Machine (SVM). Setelah melalui tahapan preprocessing dan pengujian beberapa nilai parameter, model terbaik diperoleh pada nilai C = 0.75, dengan akurasi sebesar 98% dan f1-score 0.98, menunjukkan performa SVM yang sangat baik dalam mengklasifikasikan sentimen positif dan negatif pada ulasan pengguna aplikasi Shopee.

3. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan eksperimen komputasional yang bertujuan untuk menganalisis dan mengevaluasi kinerja algoritma Support Vector Machine (SVM) dalam mengklasifikasikan sentimen terkait kualitas bahan bakar Pertamina berdasarkan data dari Platform X. Algoritma SVM berfokus pada pencarian fungsi pemisah terbaik di antara berbagai fungsi yang ada untuk membedakan dua jenis objek. Dalam analisis sentimen, SVM dianggap efektif karena kemampuannya mengelompokkan data teks ke dalam kategori secara optimal. Oleh sebab itu, penelitian ini menggunakan algoritma SVM sebagai metode utama dalam proses klasifikasi sentimen [11].



Gambar 1. Diagram Alur Penelitian

Pada gambar 1 menunjukkan tahapan analisis sentimen yang dimulai dengan proses pengambilan data melalui teknik *web scraping* dari Platform X. Data yang terkumpul kemudian memasuki tahap *pre-processing* untuk dibersihkan dan dipersiapkan, termasuk penghapusan elemen yang tidak diperlukan, tokenisasi, penghilangan *stopword*, serta proses *stemming*. Setelah tahap pembersihan selesai, data diberi label sentimen menggunakan metode *lexicon*, yang menentukan apakah suatu teks bersentimen positif atau negatif berdasarkan skor kata dalam kamus sentimen. Data berlabel ini kemudian digunakan untuk melatih model klasifikasi dengan algoritma Support Vector Machine (SVM). Model tersebut selanjutnya menghasilkan prediksi sentimen pada data uji maupun data baru. Ketika hasil klasifikasi telah diperoleh, seluruh rangkaian penelitian dianggap telah selesai.

4. PEMBAHASAN

4.1 Pengumpulan Data

Pengumpulan data ini bersumber dari patform X. Data diekstraksi menggunakan teknik *Scrapping* dengan library tweet-harvest versi 2.6.1, objek data adalah komentar publik terhadap kualitas bahan bakar pertamina, dengan *keyword* "pertamina, pertamax, pertalite, kualitas" dengan rentang waktu 1 april 2025 sampai dengan 30 september 2025. Data yang di *crawling* berjumlah 1596 komentar. Data tersebut disimpan dalam format .csv, dengan kolom conversation_id, created_at, username, user_id dan metadata lain seperti gambar dibawah ini.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	conversati	created_a	favorite_c	full_text	id_str	image_url	in_reply_t	lang	location	quote_cou	reply_cour	retweet_c	tweet_url	user_id_st	username
2	1,97E+18	Mon Sep 2	0	begini bgt	1,97E+18			in		0	0	0	https://x.c	8,67E+17	
3	1,97E+18	Mon Sep 2	1	Ini namany	1,97E+18			in		0	0	0	https://x.c	1,89E+18	
4	1,97E+18	Mon Sep 2	2	@KitaKaw	1,97E+18		KitaKawan	in		0	0	2	https://x.c	1,95E+18	
5	1,97E+18	Mon Sep 2	0	Bahlil tolol	1,97E+18			in		0	0	0	https://x.c	71693376	
6	1,97E+18	Mon Sep 2	0	DAFTAR Ke	1,97E+18			in		0	0	0	https://x.c	1,16E+08	
7	1,97E+18	Mon Sep 2	0	@sunghon	1,97E+18		sunghome	in		0	1	0	https://x.c	1,87E+18	
8	1,97E+18	Mon Sep 2	0	@happyde	1,97E+18		happydeva	in		0	0	0	https://x.c	1,72E+18	
9	1,97E+18	Mon Sep 2	1	Pertamina	1,97E+18			in		0	0	0	https://x.c	1,73E+18	
10	1,97E+18	Mon Sep 2	1	Daily life Petugas Quality &amp											
11	1,97E+18	Mon Sep 2	0	@Selebtwi	1,97E+18		SelebtwitN	in		0	0	0	https://x.c	1,76E+18	
12	1,97E+18	Mon Sep 2	0	@ARSIPAJ	1,97E+18		ARSIPAJA	in		0	0	0	https://x.c	1,57E+18	
13	1,97E+18	Mon Sep 2	0	@tvOneNe	1,97E+18		tvOneNew	in		0	0	0	https://x.c	1,84E+18	
14	1,97E+18	Mon Sep 2	0	@ARSIPAJ	1,97E+18		ARSIPAJA	in		0	1	0	https://x.c	1,84E+18	
15	1,97E+18	Mon Sep 2	0	@ARSIPAJ	1,97E+18		ARSIPAJA	in		0	0	0	https://x.c	1,38E+18	

Gambar 2. Dataset hasil *Scrapping*

4.2 PreProcessing

Preprocessing adalah proses pengolahan data yang melibatkan serangkaian langkah pengumpulan, pembersihan, transformasi, dan analisis data mentah sehingga menjadi informasi yang bermakna dan dapat digunakan untuk penelitian. Data mentah dari platfrom X masih mengandung *noise* yang tidak relevan yang dapat menghambat kinerja proses *machine learning*. Untuk mengatasi hal ini, dilakukan serangkaian proses *preprocessing* yang terdiri dari *case folding, cleaning, normalization, tokenizing, stopword removal,* dan *stemming* untuk mendapatkan data yang lebih bersih dan mudah untuk dibaca

oleh algoritma. Berikut merupakan hasil dari tahapan *preprocessing* menggunakan 2 baris pada kolom *full_text*. Tahapan *preprocessing* yang dilakukan dalam penelitian ini:

A. Case Folding

Case folding adalah proses mengubah semua huruf dalam data tweet menjadi huruf kecil. Tujuan dari langkah ini adalah untuk menyederhanakan proses klasifikasi dan memudahkan pemrosesan teks sehingga algoritma dapat mengenali kata-kata tanpa mempedulikan perbedaan huruf besar dan kecil. Hal ini membantu meningkatkan konsistensi data input dalam analisis teks [12].

Tabel 1. Tahapan *Case Folding*

Sebelum	Sesudah
@happydevante @littlesemongko @tempo_english Ya Pertamina monopoli tapi harga BBM ditentukan pemerintah untuk tetap terjangkau bagi masyarakat. Mereka jual di bawah biaya produksi (khususnya subsidi seperti Pertalite) kerugian diganti subsidi negara	@happydevante @littlesemongko @tempo_english ya pertamina monopoli tapi harga bbm ditentukan pemerintah untuk tetap terjangkau bagi masyarakat. mereka jual di bawah biaya produksi (khususnya subsidi seperti pertalite) kerugian diganti subsidi negara
@ARSIPAJA di satu sisi mendukung demi produk lokal. Di sisi lain ga percaya lg sm pertamina yg isinya korup semua udh gitu bbm dpt yg kualitas tempe. Dahlah dukung bbm swasta aja yg bagus setidaknya berdampak ke masyarakat dg harga bersaing tp dpt kualitas	@arsipaja di satu sisi mendukung demi produk lokal. di sisi lain ga percaya lg sm pertamina yg isinya korup semua udh gitu bbm dpt yg kualitas tempe. dahlah dukung bbm swasta aja yg bagus setidaknya berdampak ke masyarakat dg harga bersaing tp dpt kualitas

Tabel 1 menunjukkan tahapan *Case Folding*, merupakan contoh sebelum dan sesudah dari tahapan tersebut dalam dataset, contohnya ada pada huruf kapital yang menjadi huruf kecil seperti Ya Pertamina menjadi ya pertamina.

B. Cleaning

Cleaning merupakan tahap pembersihan data yang bertujuan untuk mendeteksi, memperbaiki, atau menghapus data tidak valid, tidak berguna, atau berupa noise dari kumpulan data. Proses ini sangat penting dalam preprocessing untuk mengatasi inkonsistensi, nilai hilang, serta elemen pengganggu seperti simbol dan tanda baca. Dengan demikian, *cleaning* memastikan kualitas data optimal sebelum analisis lebih lanjut [13].

Tabel 2. Tahapan *Cleaning*

Sebelum	Sesudah
@happydevante @littlesemongko @tempo_english ya pertamina monopoli tapi harga bbm ditentukan pemerintah untuk tetap terjangkau bagi masyarakat. mereka jual di bawah biaya produksi (khususnya subsidi seperti pertalite) kerugian diganti subsidi negara	ya pertamina monopoli tapi harga bbm ditentukan pemerintah untuk tetap terjangkau bagi masyarakat. mereka jual di bawah biaya produksi (khususnya subsidi seperti pertalite) kerugian diganti subsidi negara
@arsipaja di satu sisi mendukung demi produk lokal. di sisi lain ga percaya lg sm pertamina yg isinya korup semua udh gitu bbm dpt yg kualitas tempe. dahlah dukung bbm swasta aja yg bagus	di satu sisi mendukung demi produk lokal. di sisi lain ga percaya lg sm pertamina yg isinya korup semua udh gitu bbm dpt yg kualitas tempe. dahlah dukung bbm swasta aja yg bagus

setidaknya berdampak ke masyarakat dg harga bersaing tp dpt kualitas	setidaknya berdampak ke masyarakat dg harga bersaing tp dpt kualitas

Tabel 2 menunjukkan tahapan *Cleaning* merupakan contoh dari melakukan *cleaning* pada dataset, contohnya pada penghilangan @happydevante @littlesemongko @tempo_english dan jika ada url seperti www atau https atau amp maka akan hilang seperti @ tersebut.

C. Tokenizing

Tokenizing adalah proses memecah teks menjadi bagian-bagian kecil seperti kata atau frasa agar dapat dianalisis lebih lanjut. Proses ini penting dalam pemrosesan bahasa alami untuk memudahkan interpretasi teks oleh komputer dan meningkatkan akurasi analisis seperti klasifikasi sentimen [14].

Tabel 3. Tahapan *Tokenizing*

Sebelum	Sesudah
ya pertamina monopoli tapi harga bbm ditentukan pemerintah untuk tetap terjangkau bagi masyarakat. mereka jual di bawah biaya produksi (khususnya subsidi seperti pertalite) kerugian diganti subsidi negara.	ya, pertamina, monopoli, tapi, harga, bbm, ditentukan, pemerintah, untuk, tetap, terjangkau, bagi, masyarakat, mereka, jual, di, bawah, biaya, produksi, (khususnya, subsidi, seperti, pertalite,), kerugian, diganti, subsidi, negara,
di satu sisi mendukung demi produk lokal. di sisi lain ga percaya lg sm pertamina yg isinya korup semua udh gitu bbm dpt yg kualitas tempe. dahlah dukung bbm swasta aja yg bagus setidaknya berdampak ke masyarakat dg harga bersaing tp dpt kualitas	di, satu, sisi, mendukung, demi, produk, lokal, di, sisi, lain, ga, percaya, lg, sm, pertamina, yg, isinya, korup, semua, udh, gitu, bbm, dpt, yg, kualitas, tempe, dahlah, dukung, bbm, swasta, aja, yg, bagus, setidaknya, berdampak, ke, masyarakat dg, harga, bersaing, tp, dpt, kualitas,

Tabel 3 menunjukkan tahapan *Tokenizing* merupakan contoh dari pemisahan kalimat menjadi frasa/kata.

D. Normalization

Normalization adalah tahap di mana dilakukan standarisasi kata-kata yang memiliki makna serupa dengan cara mengubah penulisan kata-kata yang disingkat atau tidak baku menjadi bentuk yang konsisten dan seragam. Proses ini bertujuan agar kata-kata tersebut memiliki arti yang sama dan mudah dipahami dalam analisis teks [5].

Tabel 4. Tahapan *Normalization*

Sebelum	Sesudah
ya, pertamina, monopoli, tapi, harga, bbm, ditentukan, pemerintah, untuk, tetap, terjangkau, bagi, masyarakat, mereka, jual, di, bawah, biaya, produksi, (khususnya, subsidi, seperti, pertalite,), kerugian, diganti, subsidi, negara,	ya, pertamina, monopoli, tapi, harga, bbm, ditentukan, pemerintah, untuk, tetap, terjangkau, bagi, masyarakat, mereka, jual, di, bawah, biaya, produksi, (khususnya, subsidi, seperti, pertalite,), kerugian, diganti, subsidi, negara,
di, satu, sisi, mendukung, demi, produk, lokal, di, sisi, lain, ga, percaya, lg, sm, pertamina, yg, isinya, korup, semua, udh, gitu, bbm, dpt, yg, kualitas, tempe, dahlah, dukung, bbm, swasta,	di, satu, sisi, mendukung, demi, produk, lokal, di, sisi, lain, tidak, percaya, lagi, sm, pertamina, yang, isinya, korup, semua, udah, gitu, bbm, dpt, yang, kualitas, tempe, dahlah, dukung,

aja, yg, bagus, setidaknya, berdampak, ke, masyarakat dg, harga, bersaing, tp, dpt, kualitas,	bbm, swasta, aja, yang, bagus, setidaknya, berdampak, ke, masyarakat dengan, harga, bersaing, tapi, dapat, kualitas,

Tabel 4 menunjukkan ahapan Normalization merupakan contoh dari perubahan kata tidak baku atau *slang* ke kata baku.

E. Menghapus Stopword

Stopword adalah kata-kata yang tidak memiliki makna penting dan tidak mengandung informasi relevan untuk analisis, biasanya berupa kata depan atau kata penghubung. Penting untuk membuat daftar khusus kata-kata stopword yang bisa diperbarui sesuai dengan data yang dianalisis. Contoh kata *stopword* dalam Bahasa Indonesia meliputi "di", "yang", "dan", "ke", "dari", "pada", "untuk", dan lain sebagainya [10].

Tabel 5. Tahapan Menghapus *Stopwords*

Sebelum	Sesudah
ya, pertamina, monopoli, tapi, harga, bbm, ditentukan, pemerintah, untuk, tetap, terjangkau, bagi, masyarakat, mereka, jual, di, bawah, biaya, produksi, (khususnya, subsidi, seperti, pertalite,), kerugian, diganti, subsidi, negara,	monopoli, harga, ditentukan, pemerintah, terjangkau, bagi, masyarakat, mereka, jual, bawah, biaya, produksi, khususnya, seperti, kerugian, diganti, negara,
di, satu, sisi, mendukung, demi, produk, lokal, di, sisi, lain, tidak, percaya, lagi, sm, pertamina, yang, isinya, korup, semua, udah, gitu, bbm, dpt, yang, kualitas, tempe, dahlah, dukung, bbm, swasta, aja, yang, bagus, setidaknya, berdampak, ke, masyarakat dengan, harga, bersaing, tapi, dapat, kualitas,	satu, sisi, mendukung, demi, produk, lokal, sisi, lain, tidak, percaya, isinya, korup, semua, tempe, dahlah, dukung, swasta, bagus, setidaknya, berdampak, masyarakat, harga, bersaing,

Tabel 5 menunjukkan tahapan Menghapus *Stopwords* merupakan contoh dari penghapusan kata hubung atau kata depan atau kata kata yang maknanya tidak berdampak pada pelabelan positif dan negatif. Contohnya ada kata yang, di, tapi, kalau, dan ada kata yang menjurus seperti pertamina, pertalite, pertamax sering keluar namun tidak berdampak pada pelabelan nantinya.

F. Stemming

Stemming adalah proses penghilang imbuhan dari kata untuk mengubahnya menjadi bentuk dasarnya, sehingga kata-kata yang memiliki variasi bentuk tetapi makna sama dapat disatukan. Contoh stemming misalnya mengubah kata membaca, dibaca, dan membacakan menjadi kata dasar baca [9].

Tabel 6. Tahapan *Stemming*

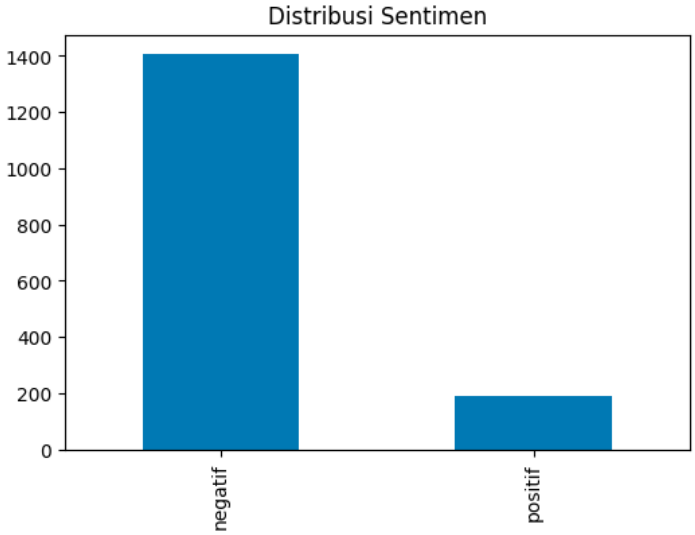
Sebelum	Sesudah
monopoli, harga, ditentukan, pemerintah, terjangkau, bagi, masyarakat, mereka, jual, bawah, biaya, produksi, khususnya, seperti, kerugian, diganti, negara,	monopoli, harga, ditentukan, pemerintah, jangkau, bagi, masyarakat, mereka, jual, bawah, biaya, produksi, khusu, seperti, rugi, ganti, negara,
satu, sisi, mendukung, demi, produk, lokal, sisi, lain, ga, percaya, isinya, korup, semua, tempe,	satu, sisi, mendukung, demi, produk, lokal, sisi, lain, ga, percaya, isi, korup, semua, tempe,

dahlah, dukung, swasta, bagus, setidaknya, berdampak, masyarakat, harga, bersaing,	dahlah, dukung, swasta, bagus, tidak, dampak, masyarakat, harga, saing,

Tabel 6 menunjukkan tahapan *Stemming* merupakan contoh dari penghilangan kata imbuhan seperti me-, di-, -kan.

4.3 Pelabelan Data

Pada tahap pelabelan, setiap data diberikan label atau kategori berdasarkan ciri-ciri kalimat dalam dataset. Proses labeling dilakukan menggunakan metode Lexicon-Based, di mana data yang telah melalui preprocessing diklasifikasikan menjadi kelas positif atau negatif berdasarkan nilai *polarity score* dari *lexicon* tersebut [15]. Gambar 3 dibawah ini merupakan hasil dari distribusi sentimen menggunakan *lexicon-Based* yang menyatakan bahwa label positif didekat 1400 dan label negatif di dekat 200.



Gambar 3. Diagram Batang Hasil Pelabelan

4.4 Hasil Support Vector Machine

Dataset yang telah diolah dari preprocessing dan pelabelan data otomatis menggunakan lexicon-Based akan lanjut diolah menggunakan algoritma yang sudah ditentukan, namun sebelum itu kita akan memberikan membagi data menjadi data latih dan data uji. Pembagian data menggunakan metode 90% dan 10% artinya dataset dibagi 90% untuk dilatih dan 10% data uji seperti gambar 4. di bawah ini.

```
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.1, random_state=42
)
```

Gambar 4. Pembagian Data 90:10

Setelah itu dilanjutkan pengubahan teks menjadi angka atau bisa disebut juga TF-IDF (Term Frequency – Inverse Document Frequency). TF-IDF akan mencari kata yang sering dan jarang keluar dalam data tersebut. Gambar 5. merupakan contoh dari TF-IDF bekerja dengan melakukan pembobotan setiap kata untuk memberikan hasil dalam kalimat tersebut apakah TF atau IDF yang lebih besar.

```
vectorizer = TfidfVectorizer(ngram_range=(1,3), max_features=5000)

X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
```

Gambar 5. Tahapan TF-IDF

Setelah banyaknya tahapan maka akan masuk ke penggunaan algoritma yaitu Support Vector Machine, pada gambar 6. merupakan cara kerja SVM menggunakan model linear kernel. Namun, karena data yang diperoleh mempunyai label yang tidak seimbang maka diberikan *class_weight='balanced'* agar Support Vector Machine dapat memberikan bobot lebih besar pada kelas minor yaitu kelas positif.

```
svm = LinearSVC(class_weight='balanced')
svm.fit(X_train_tfidf, y_train)
y_pred = svm.predict(X_test_tfidf)
```

Gambar 6. Support Vector Machine dengan Model Linear Kernel

Hasil dari semua tahapan dengan algoritma Support Vector Machine mendapatkan akurasi sebesar 0,94 atau 94% seperti pada gambar 7. Hasil yang diperoleh sangat tinggi pada prediksi negatif yaitu 0,96 atau 96% dengan presisi 96%, recall 97% , dan f1-score 96%. Sedangkan pada prediksi negatif mendapatkan 81% presisi, 74% recall dan 77% f1-score merupakan hasil yang cukup baik.

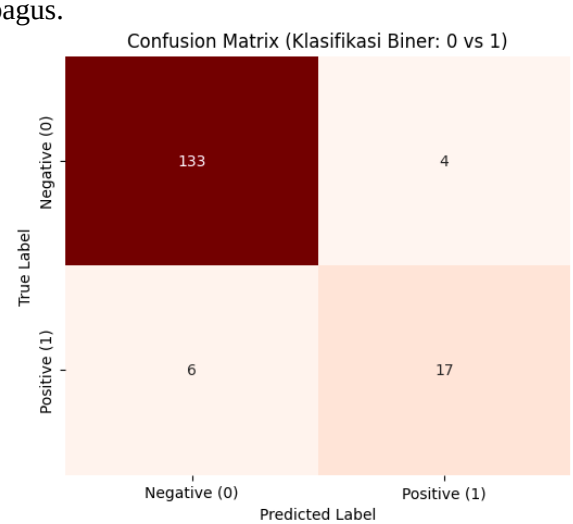
```
Classification Report:
              precision    recall  f1-score   support

     negatif       0.96      0.97      0.96       137
     positif       0.81      0.74      0.77        23

    accuracy                           0.94       160
   macro avg       0.88      0.85      0.87       160
weighted avg       0.94      0.94      0.94       160
```

Gambar 7. Hasil Algoritma Support Vector Machine

Gambar 8. Merupakan visualisasi dari hasil algoritma Support Vector Machine dari 160 data diprediksi dengan 133 komentar negatif diprediksi secara benar bahwa itu negatif dan 4 komentar negatif diprediksi secara salah bahwa itu negatif. 6 komentar positif diprediksi secara salah bahwa itu positif dan 17 komentar positif diprediksi benar bahwa itu positif. Jika diagram utama lebih gelap dari gambar maka matiks tersebut termasuk bagus.



Gambar 8. Confusion Matrix

5. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, diperoleh kesimpulan bahwa proses preprocessing memiliki peran penting dalam meningkatkan akurasi analisis sentimen terhadap komentar publik dari Platform X. Hasil pengujian menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki performa yang sangat bagus yaitu 94% dalam akurasi. Hal ini dikarenakan preprocessing yang

cukup baik dan SVM yang mampu memisahkan data dua kelas secara optimal meskipun dengan fitur teks yang terbatas. Dengan demikian, untuk analisis sentimen teks berbahasa Indonesia dalam konteks opini masyarakat terhadap kualitas bahan bakar Pertamina, algoritma SVM sangat baik dan cepat dalam pengolahannya.

Sebagai tindak lanjut, penelitian berikutnya dapat dikembangkan dengan menggunakan data dalam skala lebih besar, menambahkan label netral, mencoba metode pelabelan dengan 2 orang atau lebih sebagai annotator serta menerapkan model deep learning berbasis pre-trained embedding Bahasa Indonesia seperti IndoBERT untuk meningkatkan akurasi klasifikasi dan pemahaman konteks semantik.

DAFTAR PUSTAKA

[1] R. Maria *et al.*, "Analisis Sentimen Persepsi Masyarakat Terhadap Penggunaan Aplikasi My Pertamina Pada Media Sosial Twitter Menggunakan Metode Naïve Bayes Classifier," *Jurnal Komputer Antartika*, vol. 1, p. 2023, [Online]. Available: https://ejournal.mediaantartika.id/index.php/jka

[2] T. Wiryanti, D. Bambang, and B. Sulistiyono, "Fakultas Ekonomi-Universitas Dirgantara Marsekal Suryadarma."

[3] S. Amara, M. Khadapi, and K. Penulis, "Analisis Sentimen Masyarakat terhadap Program Makan Siang Gratis di Indonesia Tahun 2024 Menggunakan Long Short-Term Memory (LSTM)," *Jurnal Riset Sistem Informasi dan Teknik Informatika*, vol. 3, no. 4, pp. 150–160, doi: 10.61132/merkurius.v3i4.930.

[4] M. I. Burhan, "Analisis Sentimen Publik Terhadap Aplikasi Mypertamina Pada Google Playstore", doi: 10.56858/jmpkn.v8i1.390.

[5] A. N. Ihsan and S. Tresnawati, "Analisis Sentimen Pada Platform X Terhadap Layanan Provider Tri Menggunakan Naïve Bayes Dan Support Vector Machine," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Oct. 2024, doi: 10.23960/jitet.v12i3s1.5264.

[6] R. Budiarti, A. I. Purnamasari, A. Bahtiar, and E. Wahyudin, "Peningkatan Model Prediksi Konten Cyberbullying Pada Media Sosial Instagram Menggunakan Algoritma Support Vector Machine," *Jurnal Informasi Interaktif*, vol. 10, no. 1, 2025.

[7] N. Annisa Murnastiti and T. Noor Fatyanosa, "Analisis Sentimen Terhadap Makanan Manis di Platform X Menggunakan TF-IDF dan Naive Bayes," 2025. [Online]. Available: http://j-ptiik.ub.ac.id

[8] I. Agustina, G. Dwilestari, and A. R. Rinaldi, "Implementasi Data Mining Pada Proses Seleksi Beasiswa Menggunakan Naive Bayes Dan Backward Elimination".

[9] A. A. Nurrahman, M. Mauladi, and A. Rahman, "Analisis Sentimen Masyarakat terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Support Vector Machine dan SMOTE," *sudo Jurnal Teknik Informatika*, vol. 4, no. 2, pp. 50–56, Jun. 2025, doi: 10.56211/sudo.v4i2.908.

[10] Irma Surya Kumala Idris, Yasin Aril Mustofa, and Irvan Abraham Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)," *Jambura Journal of Electrical and Electronics Engineering*, vol. 5, pp. 823–848, Jan. 2023, doi: 10.1177/0165551510388123.

[11] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," vol. 6, no. 3, 2021, [Online]. Available: http://situs.com

[12] M. Nouval, D. Ramadhan, and A. Gunaryati, "Klasifikasi Sentimen Publik Terhadap Kebijakan Kendaraan Listrik Menggunakan Algoritma Naïve Bayes Dan Support Vector Machine," 2025.

[13] R. Rahman Salam, M. Fajri Jamil, and Y. Ibrahim, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of Cash Direct Assistance Distribution for Fuel Oil Using Support Vector Machine Analisis Sentimen Terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine," vol. 3, pp. 27–35, 2023.

[14] G. Lifiano Jamot Munthe and M. Yasir Alghifari, "Analisis Sentimen Publik Terhadap Kebijakan Pemerintah Terbaru Tentang Pendistribusian Gas Elpigi Subsidi Pada Masyarakat Menggunakan Metode Algoritma SVM," 2025.

[15] A. Fauziah Nur and Y. Salim, "Analisis Sentimen Pengguna X Terhadap Perkembangan Artificial Intelligence (AI) Menggunakan Algoritma Machine Learning," *Literatur Informatika & Komputer*, vol. 1, no. 4, pp. 347–357, 2024, doi: 10.33096/linier.v1i4.2534.