

Analisis Sentimen Pengguna *Twitter* Menggunakan *Support Vector Machine* Pada Kasus Kenaikan Harga BBM

Rahadi Ramlan¹, Neva Satyahadewi^{1,*}, Wirda Andani¹

¹Program Studi Statistika, Fakultas MIPA, Universitas Tanjungpura, Pontianak 78124, Indonesia

*Corresponding author. Email: neva.satya@math.untan.ac.id

ABSTRAK

Twitter merupakan salah satu media sosial dengan pengguna aktif terbanyak yaitu sebesar 24 juta pengguna aktif. Informasi yang dipublikasikan pada *twitter* mengandung komentar dari pengguna terhadap suatu objek. Analisis sentimen digunakan untuk menentukan data tersebut termasuk komentar negatif atau komentar positif karena komentar yang diambil di *twitter* yaitu data tekstual. Metode dipakai dalam analisis sentimen ini yaitu *Support Vector Machine* (SVM) tentang komentar masyarakat terhadap kenaikan harga BBM pada *twitter*. Data komentar yang digunakan sebanyak 258 data pada tanggal 4 September 2022 karena pada tanggal tersebut tepat sehari setelah kenaikan harga BBM. Pertama dilakukan *preprocessing* untuk menghilangkan kata maupun informasi yang tidak diperlukan. Kemudian data dibagi menjadi data *training* sebesar 80% dan data *testing* sebesar 20%. Didapatkan tingkat akurasi sebesar 82,69%, *sensitivity* sebesar 100%, dan *specificity* sebesar 79,07%. Kemudian dari hasil *testing* 52 data didapat hasil 43 komentar negatif dan 9 komentar positif sehingga dapat disimpulkan masyarakat lebih banyak yang tidak setuju dengan adanya kenaikan harga BBM.

Kata Kunci:

Komentar; *Twitter*; *Preprocessing*; Sentimen; *Support Vector Machine*; Harga BBM

ABSTRACT

Twitter is one of the social media with the most active users, which is 24 million active users. Information published on *twitter* contains comments from users on an object. Sentiment analysis is used to determine whether the data includes negative comments or positive comments because the comments taken on *twitter* are textual data. The method used in this sentiment analysis is *Support Vector Machine* (SVM) about public comments on fuel price increases on *twitter*. The comment data used was 258 data on September 4, 2022 because on that date it was exactly the day after the fuel price increase. First, *preprocessing* is done to remove unnecessary words or information. Then the data is divided into training data by 80% and testing data by 20%. The accuracy rate is 82.69%, *sensitivity* is 100%, and *specificity* is 79.07%. Then from the results of testing 52 data obtained the results of 43 negative comments and 9 positive comments so that it can be concluded that more people disagree with the increase in fuel prices.

Keywords:

Comment; Twitter; Preprocessing; Sentiment; Support Vector Machine; Fuel Price

Style Sitasi:

R. Ramlan, N. Satyahadewi, and W. Andani, "Analisis Sentimen Pengguna *Twitter* Menggunakan *Support Vector Machine* Pada Kasus Kenaikan Harga BBM", *Jambura J. Math.*, vol. 5, No. 2, pp. 431–445, 2023, doi: <https://doi.org/10.34312/jjom.v5i2.20860>

1. Pendahuluan

Bahan Bakar Minyak (BBM) merupakan komoditas yang memegang peranan penting dalam kegiatan ekonomi suatu negara. Kenaikan harga BBM tidak hanya berdampak pada masyarakat kecil tetapi juga berdampak pada perusahaan khususnya [1]. Pada tanggal 3 September 2022, tepat jam 14.00 harga BBM dinaikkan. Perubahan harga BBM yang paling disorot yaitu kenaikan harga Pertalite dari Rp7.650/liter menjadi Rp10.000/liter [2]. Kenaikan harga BBM menyebabkan ketidakpuasan terutama pada kelompok menengah ke bawah. Dampak dari kenaikan harga BBM yang dirasakan masyarakat termasuk meningkatnya biaya rumah tangga, transportasi, dan bahan pokok. Dari segi sosial, kenaikan harga BBM juga dapat berdampak pada kemiskinan yang semakin tinggi [3]. Kenaikan harga BBM selalu mendapat tanggapan pro dan kontra dari masyarakat. Tanggapan tersebut lebih banyak dilakukan pada media sosial, salah satu media yang digunakan adalah *Twitter*.

Dilaporkan terdapat 556 juta pengguna *twitter* di seluruh dunia pada Januari 2023 dan Indonesia menempati peringkat kelima dengan negara pengguna *Twitter* terbanyak [4]. *Twitter* merupakan salah satu media sosial yang dapat digunakan menyampaikan komentar atau pandangan mereka terhadap suatu objek, seperti produk atau layanan yang terkait dengan kejadian yang terjadi di sekitar masyarakat [5]. Pada *twitter* memiliki banyak data komentar, jadi menyalin data secara manual akan membutuhkan waktu yang lama. Salah satu teknik yang digunakan untuk mengambil data atau informasi berupa data tekstual dari *website* adalah penggunaan *web scraping*. *Web scraping* akan melakukan ekstraksi data tertentu saja dari *website* yang dituju. Data hasil *web scraping* tersebut akan digunakan untuk analisis. Analisis hasil dari *web scraping* biasa digunakan dalam klasifikasi. Klasifikasi merupakan proses mendapatkan sebuah model atau fungsi yang menjelaskan dan membedakan data menjadi kelompok-kelompok tertentu. Dalam klasifikasi salah satu analisis yang umum dipakai dalam analisis data teks yaitu *sentiment analysis* [6].

Analisis sentimen adalah pengukuran pendapat orang tentang tingkat kesepakatan mengenai topik tertentu, produk, atau layanan, atau bahkan suatu kejadian. Dua pendekatan telah digunakan untuk mempelajari analisis sentimen: *natural language processing*, and *machine learning algorithms*. Dalam pengklasifikasian *machine learning* pada *sentiment analysis* ada beberapa metode yang bisa dipakai, yaitu *Naïve Bayes Classifier* (NBC), *Maximum Entropy*, dan *Support Vector Machine* (SVM) [7]. Metode SVM dianggap sebagai metode klasifikasi teks terbaik. SVM merupakan pengklasifikasi linear yang memiliki kemampuan untuk memisahkan kelas-kelas secara linear dengan margin yang besar, sehingga menjadi salah satu pengklasifikasi paling kuat yang mampu menangani vektor fitur berdimensi tak terhingga [8].

Penelitian analisis sentimen pengguna *twitter* juga sebelumnya pernah dilakukan oleh Pravina, *et al.* [9] yang melakukan penelitian tentang opini masyarakat tentang

maskapai penerbangan pada *twitter* menggunakan metode SVM. Dalam penelitian ini terdapat perbedaan dengan penelitian yang peneliti lakukan yaitu terdapat penggunaan *translation data* teks dan *Lexicon Based Features* dalam penelitian yang dilakukan ini.

Kemudian Ali, et al. [10] melakukan penelitian analisis sentimen untuk dataset IMDB yang terdiri dari 50 ulasan film menggunakan 3 *deep learning* modes yaitu MLP (*Multilayer Perceptron*), CNN (*Convolutional Neural Network*), dan LSTM (*Long short-term memory*) dan gabungan models CNN dan LSTM, didapat hasil akhir model CNN_LSTM menggungguli teknik *deep learning* yang lainnya.

Penelitian tentang sentimen masyarakat terhadap prakerja di *twitter* menggunakan metode *support vector machine* pernah dilakukan oleh [11]. Hasil yang didapatkan dari penggunaan 2 kernel yaitu linear dan RBF yaitu nilai akurasi kernel linear 98.67%, *precision* 98%, *recall* 99%, dan *F1-Score* 98%, sedangkan pada nilai akurasi kernel RBF 98.34%, *precision* 97%, *recall* 98%, *F1-Score* 98%, dapat disimpulkan bahwa sentiment masyarakat dari pengguna *twitter* terhadap program kartu prakerja dimasa pandemi lebih condong ke netral sebesar 98,34%.

Penelitian terkait selanjutnya dilakukan oleh Nahar, et al. [12] yang membahas sistem identifikasi membaca al-qur'an dengan menggunakan metode *artificial neural network* (ANN) dan *support vector machine* (SVM). Hasil akurasi SVM yaitu 96,59% dan jauh lebih baik daripada akurasi ANN yaitu sebesar 86,17%.

Penelitian tentang metode SVM juga pernah dilakukan oleh Saranya [13] yang membahas tentang klasifikasi tumor yaitu ganas dan jinak. Penelitian tersebut menggunakan hasil CT scans, X-Ray and MRI. Hasil MRI yang sering digunakan karena paling dapat diandalkan dan aman. Tumor yang dilakukan penelitian yaitu tumor otak karena struktur otak yang rumit otak, penggunaan klinis dari beberapa yang umum digunakan biopsi terbatas untuk tumor otak.

Pada penelitian ini, digunakan metode SVM karena dari beberapa referensi menunjukkan nilai akurasi metode SVM cukup baik sehingga cocok dipakai dalam analisis sentimen. Penelitian ini menggunakan data hasil *tweet* di *twitter* dengan kata kunci BBM Naik. Alasan memilih topik kenaikan harga BBM karena pada saat itu sedang ramai diperbincangkan kenaikan harga BBM yang cukup tinggi sehingga diambil topik sentimen terhadap kenaikan harga BBM. Data yang digunakan adalah data latih dan data uji. Tujuan utama dari penelitian ini untuk melihat sentimen pengguna *twitter* terhadap kenaikan harga BBM berdasarkan metode *support vector machine*.

Penelitian ini penting dilakukan karena hasil penelitian ini bisa membantu pemerintah dan pemangku kepentingan lainnya dalam mengambil keputusan terkait kebijakan harga BBM. Mereka dapat menggunakan informasi ini untuk memahami bagaimana masyarakat merasa tentang kenaikan harga BBM, dan dapat membantu meningkatkan layanan publik dan memberikan respon yang lebih baik terhadap kebutuhan dan aspirasi masyarakat. Hasil penelitian ini juga dapat dijadikan referensi pengembangan metode analisis sentimen lebih lanjut, khususnya dalam konteks media sosial seperti *Twitter*. Penggunaan metode *Support Vector Machine* dalam penelitian ini dapat memberikan kontribusi untuk pengembangan teknik dan algoritma analisis sentimen yang lain yang lebih efisien dan akurat. Penelitian ini juga dapat membantu memprediksi potensi peristiwa sosial, protes, atau aksi massa sebagai respons terhadap kenaikan harga BBM.

2. Metode

2.1. Data Penelitian

Data yang dipakai dalam penelitian ini yaitu data sekunder dengan data yang diperoleh dari media sosial *twitter* pada tanggal 4 September 2022. Pada penelitian ini digunakan data *tweet* yang ada di *twitter* dengan pengambilan data menggunakan teknik *web scraping* dengan bantuan *jupyter notebook*. Data *tweet* yang digunakan yaitu data *tweet* dengan kata kunci "BBM Naik" pada tanggal 4 September 2022 dengan *tweet* yang digunakan merupakan *tweet* yang berbahasa Indonesia. Jumlah data yang didapat merupakan sampel data dari hasil *web scraping* dari *twitter* berjumlah 1001 data dan jumlah data setelah dilakukan *data selection* berjumlah 258 data. Berikut contoh sampel data *twitter* pada Tabel 1 yang diambil menggunakan teknik *web scraping*.

Tabel 1. Sampel data *twitter*

No	Tweet Hasil <i>Web Scraping</i>
1	BBM naik tinggi, rokok tak terbeli, orang pintar narik subsidi, lambe kami sepo lagi
2	@permaisuri.id Kuncinya di ""meski harga global menggila"" ... Karena harga global tidak menggila maka harga bbm perlu naik
3	Bbm naik tuh tujuannya untuk ngurangin orang orang yang suka naik motor tanpa tujuan

Web scraping merujuk pada teknik pengambilan dokumen semi-terstruktur dari internet yang biasanya berupa halaman web yang dibuat menggunakan bahasa markup seperti HTML atau XHTML. Tujuan utama dari *web scraping* ini adalah untuk mengumpulkan informasi dari halaman web tersebut, baik secara keseluruhan maupun sebagian untuk dipakai dalam konteks lainnya [14][14].

2.2. Prosedur Penelitian

2.2.1. Data Selection

Data selection dilakukan dengan tujuan untuk memilih data mana saja yang termasuk ke dalam data komentar. Data yang tidak termasuk ke dalam komentar contohnya *tweet* yang mengenai berita tentang kenaikan BBM dengan menyebarkan *link* di *twitter*, dan komentar BBM naik dikaitkan dengan klasemen salah satu klub bola juga naik, dll. Setelah dilakukan seleksi manual, jumlah data yang diperoleh yaitu sebanyak 258 data. Data komentar kemudian dilakukan pembagian kelas yaitu komentar positif dan negatif yang juga dilakukan secara manual. Komentar positif yaitu pendapat dari masyarakat di *twitter* yang mendukung atau setuju terhadap kenaikan BBM, begitupula sebaliknya.

2.2.2. Text Preprocessing

Preprocessing data komentar *twitter* perlu dilakukan sebelum proses klasifikasi agar dimensi dari model ruang vektor menjadi lebih kecil. Dengan mengurangi dimensi model ruang vektor, proses klasifikasi akan menjadi lebih cepat. Tujuan dari *preprocessing* ini adalah untuk menghilangkan kata-kata yang tidak cocok untuk penelitian, menyamakan bentuk kata-kata atau huruf (*lowercase*), dan mengurangi jumlah kata. Tahap *preprocessing* dalam penelitian ini terdiri dari beberapa proses, yaitu *cleaning*, *case folding*, *tokenization*, *stopwords removal*, dan *stemming* [15]:

1. *Cleaning* yaitu data dilakukan pembersihan dengan tujuan untuk menghapus tanda

- baca, *mention*, *hashtag*, *link*, dan karakter lainnya pada dokumen.
2. *Case folding* yaitu untuk menyamakan karakter dari setiap dokumen setiap huruf dalam kalimat atau kata dari setiap dokumen dikonversikan menjadi huruf kecil (*lowercase*).
 3. *Tokenizing* yaitu digunakan untuk menghapus karakter-karakter tertentu seperti tanda baca. Karakter spasi berfungsi sebagai pemisah untuk membagi kalimat menjadi sekelompok kata.
 4. *Stopwords removal* adalah proses yang dilakukan untuk menghilangkan kata-kata yang tidak memiliki makna. Tahap ini akan menggunakan kamus *stopwords*. Kata-kata yang terdapat dalam daftar *stopwords* akan dihapus.
 5. *Stemming* yaitu proses yang dilakukan untuk mengubah kata-kata yang tidak baku menjadi baku dengan menghilangkan imbuhan seperti awalan dan akhiran.

2.2.3. Pembobotan Kata TF-IDF (*Term Frequency-Inverse Document Frequency*)

Metode TF-IDF merupakan suatu teknik untuk menghitung nilai bobot dari setiap kata yang paling umum digunakan dalam proses pengambilan informasi. Metode ini terkenal karena efisien, mudah digunakan, serta memberikan hasil yang akurat. Prinsip dasar dari metode TF-IDF adalah memberikan nilai bobot pada hubungan antara kata tertentu (*term*) dengan sebuah dokumen. TF-IDF digunakan sebagai indikator statistik untuk mengevaluasi tingkat pentingnya sebuah kata dalam sebuah dokumen atau dalam sekumpulan kata. Dalam konteks dokumen tunggal, setiap kalimat dianggap sebagai dokumen yang terpisah. Frekuensi kemunculan kata dalam dokumen tersebut mencerminkan tingkat kepentingan kata tersebut dalam dokumen tersebut. Bobot kata akan lebih tinggi jika kata tersebut seiring muncul dalam dokumen tersebut, dan lebih rendah jika kata tersebut muncul dalam beberapa dokumen yang berbeda [16].

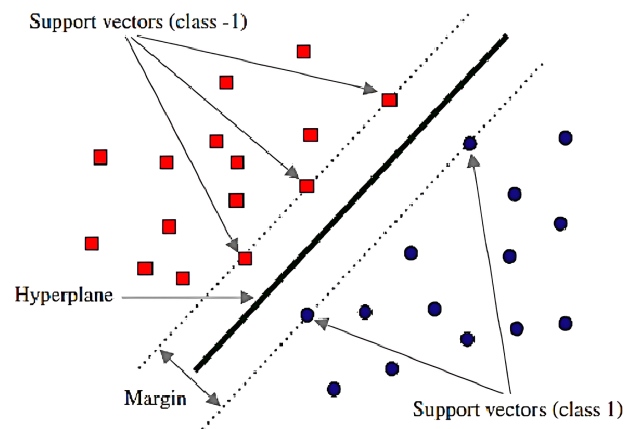
Metode TF-IDF digunakan secara luas sebagai metode perbandingan terhadap metode pembobotan kata baru. Metode ini menghitung bobot kata dalam sebuah dokumen dilakukan dengan mengalikan nilai *Term Frequency* dengan *Inversi Document Frequency* [17]. Rumus untuk menghitung TF-IDF [18] diberikan pada persamaan (1):

$$W_{dt} = tf_{dt} \times idf_{dt} = tf_{dt} \times \log \frac{D}{df} \quad (1)$$

2.3. Konsep Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode klasifikasi yang diperkenalkan oleh Boser, Guyon, dan Vapnik pada tahun 1992. Metode SVM banyak digunakan karena akurasi yang tinggi dan kemampuannya dalam mengatasi data berdimensi tinggi [19]. SVM adalah suatu metode *machine learning* yang bertujuan untuk mencari jarak terbaik yang dapat memisahkan dua kelas, yang juga dikenal sebagai *hyperplane*. Metode SVM bekerja dengan memisahkan anggota dari dua kelas, yaitu kelas 1 dan -1. Pada Gambar 1, titik warna merah mewakili pola dengan kelas -1, sementara warna biru digunakan untuk menggambarkan pola pada kelas 1. Untuk memecahkan masalah ini, kita dapat menentukan garis pembatas atau *hyperplane* sebagai pemisah antara kedua kelompok kelas [17].

Konsep SVM dapat dijelaskan secara sederhana sebagai usaha mencari garis pemisah terbaik antar dua kelas (*hyperplane*) dengan memaksimalkan jarak antar kelas. Dengan



Gambar 1. *Hyperplane SVM*

cara ini, SVM dapat menjamin kemampuan generalisasi yang tinggi untuk data-data yang akan datang [20].

Menurut [21] *hyperplane* pemisah dapat ditulis pada Persamaan (2).

$$w \cdot x_i + b = 0 \quad (2)$$

Pertidaksamaan berikut dapat digunakan untuk merumuskan nilai x yang termasuk dalam kelas -1 (sampel negatif):

$$w \cdot x_i + b \leq -1 \quad (3)$$

Nilai x yang termasuk kelas 1 (sampel positif) dapat dinyatakan sebagai nilai yang memenuhi pertidaksamaan berikut:

$$w \cdot x_i + b \geq 1 \quad (4)$$

dengan w = bobot *vector*, x_i = data ke- i , dan b = bias.

Adapun beberapa langkah metode SVM adalah sebagai berikut:

1. Melakukan inisialisasi terhadap parameter yang diaplikasikan, yaitu λ , α_x , γ , C , dan ϵ . Misalkan $\lambda = 0,5$ (*Lambda*), $\alpha_x = 0,5$ (*alpha tetapan*), $\gamma = 0,5$ (*learning rate*), $C = 1$ (variabel *slack*), dan $\epsilon = 0,001$ (*epsilon*).
2. Menghitung Matriks Hessian (D_{ij}) dengan Persamaan (5):

$$D_{ij} = y_i y_j K(x_i \cdot x_j) + \lambda^2 \quad (5)$$

dengan y adalah kelas data ke- i dan K adalah fungsi kernel.

3. Perhitungan nilai *error* (E_i) dengan menggunakan Persamaan (6):

$$E_i = \sum_{j=1}^n \alpha_x \cdot D_{ij} \quad (6)$$

4. Perhitungan nilai *delta alpha* (δa_i) dengan menggunakan Persamaan (7):

$$\delta\alpha_i = \min\{\max[\gamma(1 - E_i), -\alpha_x], C - \alpha_x\}. \tag{7}$$

5. Perhitungan nilai *alpha* baru (α_i) dengan menggunakan Persamaan (8):

$$\alpha_i = \alpha_x + \delta\alpha_i \tag{8}$$

6. Menghitung nilai w^+ (bobot nilai *alpha* terbesar pada kelas positif) dan w^- (bobot nilai *alpha* terbesar pada kelas negatif) menggunakan persamaan (9) dan (10):

$$w^+ = \sum_{i=1}^n \alpha_i y_i (K(x_i, x^+)) \tag{9}$$

$$w^- = \sum_{i=1}^n \alpha_i y_i (K(x_i, x^-)). \tag{10}$$

7. Perhitungan nilai bias (b) dengan menggunakan Persamaan (11):

$$b = -\frac{1}{2} (w^+ + w^-). \tag{11}$$

8. Pengujian terhadap data uji kemudian perhitungan keputusan $h(x)$ menggunakan persamaan (12):

$$h(x) = \begin{cases} +1, & \text{if } w \cdot x + b \geq 0 \\ -1, & \text{if } w \cdot x + b < 0 \end{cases} \tag{12}$$

Apabila hasil perhitungan keputusan memiliki nilai 0 atau lebih besar, maka tanda $sign h(x)$ adalah +1, menunjukkan kelas positif. Sebaliknya, jika hasil perhitungan keputusan memiliki nilai kurang dari 0, tanda $sign h(x)$ adalah -1. Perhitungan $sign h(x)$ mengikuti persamaan (13):

$$h(x) = w \cdot x + b = \sum_{i=1}^n \alpha_i \cdot y_i \cdot K(x_i, x_j) + b \tag{13}$$

Setelah melalui tahap metode SVM, selanjutnya dilakukan evaluasi hasil klasifikasi.

2.4. Evaluasi Hasil Klasifikasi

Confusion Matrix adalah cara yang dengan efektif menunjukkan kinerja pengklasifikasi dan memiliki keuntungan dalam interpretasi hasil yang mudah. *Confusion Matrix* dapat digunakan untuk mengevaluasi kinerja model atau algoritma apapun.

Tabel 2. *Confusion matrix*

		Kelas Hasil Prediksi	
		Positif	Negatif
Kelas Asli	Positif	<i>True Positive (TP)</i>	<i>False Positif (FP)</i>
	Negatif	<i>False Negative (FN)</i>	<i>True Negative (TN)</i>

Seperti yang ditunjukkan pada Tabel 2, baris-baris dalam *confusion matrix* yang mewakili nilai kelas asli dan kolom-kolom mewakili nilai kelas prediksi. *Confusion matrix* memberikan informasi komparatif antara hasil klasifikasi sistem dan hasil

sebenarnya [22].

Terdapat empat istilah yang mencerminkan hasil dari proses klasifikasi dalam *Confusion matrix*. Istilah-istilah tersebut adalah: *True Positive* (data positif yang diprediksi dengan tepat), *True Negative* (data negatif yang diprediksi dengan tepat), *False Positive* (data positif yang diprediksi secara salah), dan *False Negative* (data negatif yang diprediksi secara salah) [16]. Setelah data uji dimasukkan ke dalam *confusion matrix* nilai-nilai tersebut kemudian digunakan untuk menghitung sensitivitas, spesifisitas, dan akurasi. Berdasarkan [12], sensitivitas digunakan untuk membandingkan jumlah data yang diprediksi dengan benar (*True Positive*) dalam klasifikasi terhadap jumlah data yang masuk ke dalam kelas positif.

$$Sensitivity = \frac{TP}{TP + FN} \tag{14}$$

Specificity adalah jumlah perbandingan data yang salah (*True Negative*) terhadap jumlah data yang masuk kelas negatif. Untuk mengitung nilai-nilai tersebut digunakan persamaan di bawah ini.

$$Specificity = \frac{TN}{TN + FP} \tag{15}$$

Accuracy adalah persentase dari data yang terklasifikasi dengan benar dari semua baris dalam set data, baik yang terklasifikasi dengan benar maupun yang terklasifikasi secara salah, berikut persamaannya.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{16}$$

3. Hasil dan Pembahasan

3.1. Data Selection

Data selection pada penelitian ini dilakukan untuk memilih data mana saja yang termasuk ke dalam data komentar dan bukan komentar terkait kenaikan harga BBM. Setelah dilakukan seleksi, diperoleh 258 data komentar dan bukan komentar. Beberapa contoh hasil *data selection* disajikan pada Tabel 3.

Tabel 3. Contoh *Data Selection*

No	Teks	Keterangan
1	malas kalau bbm naik harga ikut naik, ancam demo, orang tinggal ibu kota sumpah lelah banget	Komentar
2	BBM naik gara gara MU beli antony, Kalah arsenal cuma nutupin bbm naik	Bukan Komentar

3.2. Text Preprocessing

Preprocessing data komentar *twitter* perlu dilakukan sebelum proses klasifikasi, untuk menghilangkan kata-kata yang tidak cocok untuk penelitian, menyamakan bentuk kata-kata atau huruf (*lowercase*), dan mengurangi jumlah kata. Dengan melakukan *preprocessing data*, dapat dipastikan bahwa data yang digunakan untuk pelatihan adalah data yang lebih baik, lebih bersih, dan lebih cocok untuk digunakan pada analisis SVM.

Beberapa sampel hasil proses *text preprocessing* disajikan pada Tabel 4.

Tabel 4. Contoh hasil *text preprocessing*

No	Text Awal	Text Hasil Preprocessing
1	BBM naik tinggi, rokok tak terbeli. orang pintar narik subsidi, lambe kami sepo lagi.	bbm naik tinggi rokok beli orang pintar tarik subsidi
2	@permaisuri_id Kuncinya di ""meski harga global menggila""... Karena harga global tidak menggila maka harga bbm perlu naik	kunci meski harga global gila harga bbm perlu naik
3	Bbm naik tuh tujuannya untuk ngurangin orang yang suka naik motor tanpa tujuan	bbm naik itu tujuan kurang orang orang suka naik motor tujuan

Setelah dilakukan *text preprocessing*, selanjutnya perlu dilakukan pembobotan kata.

3.3. Pembobotan Kata TF-IDF (Term Frequency-Inverse Document Frequency)

Pembobotan kata dengan TF-IDF dilakukan untuk melihat bobot kata yang paling umum digunakan pada proses pengambilan informasi. Proses pembobotan dengan TF-IDF dilakukan dengan memberikan bobot hubungan antar kata tertentu yang berkaitan dengan kenaikan harga BBM. Hal ini dilakukan untuk melihat data statistik penting tidaknya sebuah kata yang berkaitan dengan kenaikan harga BBM.

Sampel data *tweet* yang digunakan untuk perhitungan TF-IDF dengan persamaan (1) adalah 3 sampel data *tweet* hasil dari *preprocessing* pada Table 4. Hasil perhitungan keseluruhan TF-IDF selanjutnya disajikan pada Tabel 5.

Tabel 5. Perhitungan TF-IDF

Kata	TF			DF	IDF	TF-IDF		
	D1	D2	D3			D1	D2	D3
bbm	1	1	1	3	0	0	0	0
naik	1	1	2	3	0	0	0	0
tinggi	1	0	0	1	0,477	0,477	0	0
rokok	1	0	0	1	0,477	0,477	0	0
beli	1	0	0	1	0,477	0,477	0	0
orang	1	0	2	2	0,176	0,176	0	0,352
pintar	1	0	0	1	0,477	0,477	0	0
tarik	1	0	0	1	0,477	0,477	0	0
subsidi	1	0	0	1	0,477	0,477	0	0
kunci	0	1	0	1	0,477	0	0,477	0
meski	0	1	0	1	0,477	0	0,477	0
harga	0	3	0	1	0,477	0	1,431	0
global	0	2	0	1	0,477	0	0,954	0
gila	0	2	0	1	0,477	0	0,954	0
perlu	0	1	0	1	0,477	0	0,477	0
tujuan	0	0	2	1	0,477	0	0	0,954
kurang	0	0	1	1	0,477	0	0	0,477
suka	0	0	1	1	0,477	0	0	0,477
motor	0	0	1	1	0,477	0	0	0,477
							y	
							-1	1
								1

Tabel 5 menunjukkan variabel D yang merupakan data penelitian yang diolah, seperti

D1 yaitu data ke-1, dan seterusnya. Variabel y menunjukkan label sentimen dari ketiga sampel tweet yang digunakan. Dari Tabel 5 data kedua nilai TF-IDF yang paling tinggi yaitu kata “harga” yang bernilai 1,431 yang berarti kata “harga” ini penting dan khas pada data kedua karena memiliki nilai TF-IDF yang tinggi dari kata-kata lainnya, demikian juga dengan data-data selanjutnya.

3.4. Klasifikasi Support Vector Machine (SVM)

Proses klasifikasi dimulai dengan memisahkan data uji dan data latih. Dalam penelitian ini, persentase data latih dan data uji digunakan untuk membagi data. Data latih dan data uji yang digunakan mencakup data pembobotan TF-IDF dan label kelas, dengan perbandingan 20% untuk data uji dan 80% untuk data latih. Pemilihan data uji dan data latih didasarkan pada urutan komentar. Dari total 258 data komentar, 206 data komentar pertama digunakan sebagai data latih, sementara 52 data sisanya digunakan sebagai data uji. Dalam penelitian ini, proses klasifikasi menggunakan fungsi kernel. Fungsi kernel ini digunakan untuk menghitung setiap data terhadap data itu sendiri dan data lainnya. Rincian fungsi kernel dapat ditemukan dalam Tabel 6.

Tabel 6. Fungsi Kernel

	D1	D2	D3
D1	$K(D1,D1)$	$K(D1,D2)$	$K(D1,D3)$
D2	$K(D2,D1)$	$K(D2,D2)$	$K(D2,D3)$
D3	$K(D3,D1)$	$K(D3,D2)$	$K(D3,D3)$

Nilai fungsi kernel keseluruhan didapatkan matriks 3×3 yang hasilnya diuraikan dalam Tabel 7.

Tabel 7. Hasil Perhitungan Nilai Fungsi Kernel

	D1	D2	D3
D1	1,396	0	0,062
D2	0	4,553	0
D3	0,062	0	1,718

Tabel 7 memberikan nilai kesamaan (*similarity*) antara pasangan data dalam ruang fitur berdasarkan *dot product*. Semakin besar nilai kernel linear, semakin mirip vektor-vektor tersebut. Misalnya, nilai yang lebih tinggi antara D1 dan D3 menunjukkan bahwa kedua vektor tersebut lebih mirip atau serupa dalam ruang fitur. Nilai yang lebih rendah, seperti antara D2 dan D3, menunjukkan bahwa kedua vektor tersebut kurang mirip dalam ruang fitur.

Selanjutnya adalah melakukan perhitungan matriks Hessian dengan menggunakan Persamaan (5) dilakukan untuk setiap data yang hasilnya dapat dilihat pada Tabel 8.

Tabel 8. Hasil Perhitungan Nilai Matriks Hessian

	D1	D2	D3
D1	1,647	0,250	0,188
D2	0,250	4,803	0,250
D3	0,188	0,250	1,968

Dalam SVM merujuk pada elemen-elemen matriks jarak antara pasangan data. Jika nilai

matriks hessian bernilai positif menunjukkan bahwa pasangan data memiliki kesamaan atau korelasi positif dalam ruang, begitupula sebaliknya. Berdasarkan Tabel 8 nilai matriks hessian semua bernilai positif berarti semua pasangan data memiliki kesamaan atau korelasi positif dalam ruang.

Tahap selanjutnya adalah menghitung nilai *error* menggunakan Persamaan (6) yang dilakukan untuk setiap data yang hasilnya dapat dilihat pada Tabel 9.

Tabel 9. Hasil Perhitungan Nilai *Error*

Dokumen	<i>Error</i>
D1	1,042
D2	2,652
D3	1,203

Tabel 9 menunjukan nilai *error* dari data, nilai *error* dari data pertama cukup baik karena nilainya lebih rendah daripada nilai *error* data lainnya.

Setelah mendapat nilai *error*, kemudian hitung nilai *delta alpha* menggunakan Persamaan (7). Perhitungan nilai *delta alpha* dilakukan untuk memperbarui bobot (*alpha*) dari model SVM ketika data baru ditambahkan ke dalam dataset pelatihan. Dalam pendekatan *sequential training*, tidak perlu melatih kembali seluruh model dari awal setiap kali ada data baru. Sebaliknya harus memanfaatkan model yang sudah ada dan melakukan penyesuaian pada bobotnya untuk mencerminkan informasi dari data baru. Hasil perhitungan data lainnya dapat dilihat pada Tabel 10.

Tabel 10. Hasil Perhitungan Nilai *Delta Alpha*

Dokumen	<i>Delta alpha</i>
D1	-0,021
D2	-0,5
D3	-0,101

Berdasarkan pada Tabel 10, nilai *delta alpha* semua data bernilai negatif yang artinya data tersebut harus dikurangi dengan nilai awalnya yaitu 0.5, yang prosesnya ada pada perhitungan nilai *alpha* baru menggunakan Persamaan (8). Perhitungan dilakukan untuk setiap data yang hasilnya dapat dilihat pada Tabel 11.

Tabel 11. Hasil Perhitungan Nilai *Alpha* Baru

Dokumen	<i>Alpha</i> baru
D1	0,479
D2	0
D3	0,399

Tabel 11 menunjukkan hasil pengurangan nilai *alpha* awal dengan nilai *delta alpha* yang bertujuan agar semua nilai *alpha* bernilai positif. Tujuan utama SVM adalah menemukan *hyperplane* dengan margin maksimal antara dua kelas sehingga w^+ dan w^- dihitung untuk menentukan bagaimana vektor pendukung berkontribusi terhadap pembentukan *hyperplane*.

Proses selanjutnya menghitung nilai w^+ dan w^- . Bobot nilai *alpha* terbesar pada kelas positif pada data ke-3 dan bobot nilai *alpha* terbesar kelas negatif yaitu pada data ke-1.

$$w * x^+ = (0,479 \times (-1) \times 0,062) + (0 \times (1) \times 0) + (0,399 \times 1 \times 1,718) = 0,655$$

$$w * x^- = (0,479 \times (-1) \times 1,399) + (0 \times (-1) \times 0) + (0,399 \times 1 \times 0,062) = -0,644$$

Didapat hasil $w^+ = 0.655$ yang berarti vektor pendukung dari kelas positif (1) memberikan kontribusi positif dalam menentukan posisi *hyperplane* yang memisahkan kedua kelas. Semakin besar nilai w^+ semakin besar kontribusi vektor pendukung dari kelas positif terhadap pembentukan *hyperplane*, begitu juga sebaliknya untuk nilai w^- .

Selanjutnya, dihitung nilai bias (b) dengan memakai Persamaan (11), menghasilkan

$$b = -\frac{1}{2} (0,655 + (-0,644)) = -0,006.$$

Nilai bias yang didapat bernilai negatif yang berarti mendorong *hyperplane* ke arah kelas negatif, sehingga lebih cenderung mengklasifikasikan data masuk kedalam kelas negatif.

Setelah nilai α , w dan b sudah didapat, kemudian bisa dilakukan pengujian pada sampel data uji berikut ini:

D_x = apa cuma saya setuju bbm naik tinggi.

Kalimat dari data uji jika dipecah mengandung kata/*term* 'apa' 'cuma' 'saya' 'setuju' 'bbm' 'naik' 'tinggi' dan dibuat TF-IDF untuk proses klasifikasi yang dapat dilihat pada Tabel 12. Variabel D_x menunjukkan pada data ke- x dan variabel y menunjukkan sentimen yang akan dicari dari data uji.

Tabel 12. TF-IDF Data uji

D_x	TF-IDF	D_x	TF-IDF
bbm	0	global	0
naik	0	gila	0
tinggi	0,301	perlu	0
rokok	0	tujuan	0
beli	0	kurang	0
orang	0	suka	0
pintar	0	motor	0
tarik	0	apa	0,602
subsidi	0	cuma	0,602
kunci	0	saya	0,602
meski	0	setuju	0,602
harga	0		
		y	
		?	

Pertama dalam menguji adalah dengan mengalikan *dot product* data uji dengan semua data latih memakai fungsi kernel dengan perhitungan.

$$K(x, y) = xy$$

dimana x = data latih dan y = data uji. Hasil perhitungan *dot product* antara data uji

dengan data latih pada Tabel 13.

Tabel 13. Hasil perhitungan *dot product* data uji dan data latih

Dokumen	<i>Dot Product</i>
D1	0,144
D2	0
D3	0

Tabel 13 menunjukkan hasil perhitungan *dot product* data uji dan data latih. Hasil perhitungan memberikan nilai data pertama sebesar 0,144 dan data lainnya bernilai 0. Selanjutnya perhitungan fungsi keputusan memakai Persamaan (12) menghasilkan nilai $h(x)$:

$$\begin{aligned} h(x) &= ((0,479 \times (-1) \times 0,144) + (-0,006)) + ((0 \times (1) \times 0) + (-0,006)) + \\ &\quad ((0,399 \times 1 \times 0) + (-0,006)) \\ &= (-0,086) = \text{sign}(-0,086) = -1 \end{aligned}$$

Berdasarkan dari perhitungan fungsi keputusan terhadap data uji diperoleh nilai $h(x)$ *sign* data uji tersebut diklasifikasi kedalam kelas negatif. Artinya berdasarkan pelatihan data latih dan dilakukan pengujian data baru yang belum pernah dilihat oleh model didapat hasil klasifikasi data baru tersebut masuk kedalam klasifikasi negatif karena hasil akhir bernilai -1.

3.5. Evaluasi Hasil Klasifikasi

Hasil analisa menggunakan metode SVM dengan penggunaan data sebesar 258 data yang terbagi atas 206 data *training* dan 52 data *testing*. Tabel 14 menjelaskan tentang *Confusion Matrix* hasil klasifikasi sentimen.

Tabel 14. Hasil *Confusion Matrix*

		Kelas Hasil Prediksi	
		Positif	Negatif
Kelas Asli	Positif	9	9
	Negatif	0	34

Berdasarkan nilai *confusion matrix* pada Tabel 14, dapat dihitung nilai *sensitivity*, *specificity*, dan akurasi dengan menggunakan Persamaan (14), Persamaan (15), dan Persamaan (16), yaitu

$$\text{Sensitivity} = \frac{TP}{TP + FN} = \frac{9}{9 + 0} = 1 = 100\%$$

$$\text{Specificity} = \frac{TN}{TN + FP} = \frac{34}{34 + 9} = 0,790 = 79,07\%$$

$$\text{Akurasi} = \frac{TP + TN}{TP + FP + FN + TN} = \frac{9 + 34}{9 + 9 + 0 + 34} = 0,826 = 82,69\%$$

Berdasarkan perhitungan memberikan nilai *sensitivity* sebesar 100% yang artinya

perbandingan jumlah data *true positive* (9 data) dalam pengklasifikasian terhadap jumlah data yang masuk kelas positif (9 data), *specificity* sebesar 79.07% yang artinya jumlah perbandingan data *True Negative* (34 data) terhadap jumlah data yang masuk kelas negatif (43 data) dan nilai akurasi sebesar 82,69% yang artinya prediksi yang benar dari total prediksi dari model.

4. Kesimpulan

Berdasarkan hasil analisis sentimen pengguna *twitter* menggunakan metode SVM diperoleh kesimpulan bahwa metode SVM dapat melakukan pengklasifikasian respon publik tentang isu kenaikan BBM mengacu hasil evaluasi dengan nilai akurasi sebesar 82,69%, *sensitivity* sebesar 100%, dan *specificity* sebesar 79,07%.

Berdasarkan hasil nilai *confusion matrix*, sentimen terhadap kenaikan BBM menggunakan data uji, dengan menggunakan metode SVM mendapatkan hasil positif sebanyak 9 dari 52 data uji (17,31%) dan hasil nilai negatif sebanyak 43 data dari 52 data uji (82,69%). Berdasarkan hasil tersebut dapat disimpulkan bahwa sebagian besar pengguna BBM tidak setuju dengan kenaikan harga BBM.

Referensi

- [1] N. M. Ardani, "Studi komparatif reaksi pasar sebelum dan sesudah kenaikan harga bbm atas saham Iq-45 di bei (even study pada peristiwa kenaikan bbm tanggal 5 januari 2017)," *Jurnal Pendidikan Ekonomi Undiksha*, vol. 10, no. 2, pp. 616–626, 2019, doi: 10.23887/jjpe.v10i2.20142.
- [2] J. C. Sihombing, "Kenaikan harga bbm : Jahat atau sepakat..???" 2022, [Online] Available: <https://www.djkn.kemenkeu.go.id/kpknl-sidempuan/baca-artikel/15373/Kenaikan-Harga-BBM-Jahat-atau-Sepakat.html><https://www.djkn.kemenkeu.go.id/kpknl-sidempuan/baca-artikel/15373/Kenaikan-Harga-BBM-Jahat-atau-Sepakat.html>.
- [3] B. C. Sandy, D. Manongga, and A. Iriani, "Analisis sentimen terhadap kenaikan harga bahan bakar minyak (bbm) pada media online," in *Seminar Nasional & Konferensi Ilmiah Sistem Informasi, Informatika & Komunikasi*, 2015, pp. 24–30, [Online] Available: <https://publikasi.uyelindo.ac.id/index.php/semmau/article/view/56/35>.
- [4] C. M. Annur, "Pengguna twitter di indonesia capai 24 juta hingga awal 2023, peringkat berapa di dunia?" *Databoks*, 2023, [Online] Available: <https://databoks.katadata.co.id/datapublish/2023/02/27/pengguna-twitter-di-indonesia-capai-24-juta-hingga-awal-2023-peringkat-berapa-di-dunia>.
- [5] R. Tinages, A. Triayudi, and I. D. Sholihati, "Analisis sentimen terhadap layanan indihome berdasarkan twitter dengan metode klasifikasi support vector machine (SVM)," *Jurnal Media Informatika Budidarma*, vol. 4, no. 3, p. 650, jul 2020, doi: 10.30865/mib.v4i3.2181.
- [6] D. D. A. Yani, H. S. Pratiwi, and H. Muhandi, "Implementasi web scraping untuk pengambilan data pada situs marketplace," *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, vol. 7, no. 4, p. 257, oct 2019, doi: 10.26418/justin.v7i4.30930.
- [7] R. S. Karan, K. K. Shirsat, P. L. Kasar, and R. Chaudhary, "Sentiment analysis on twitter data: A new approach," in *2018 International Conference on Current Trends towards Converging Technologies (ICCTCT)*. IEEE, mar 2018, doi: 10.1109/icctct.2018.8551169.
- [8] Y. A. Amrani, M. Lazaar, and K. E. E. Kadiri, "Random forest and support vector machine based hybrid approach to sentiment analysis," *Procedia Computer Science*, vol. 127, pp. 511–520, 2018, doi: 10.1016/j.procs.2018.01.150.
- [9] A. M. Pravina, I. Cholisoddin, and P. P. Adikara, "Analisis sentimen tentang opini maskapai penerbangan pada dokumen twitter menggunakan algoritme support vector machine (svm)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 3, no. 3, pp. 2789–2797, 2019, [Online] Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/4793/2232>.
- [10] N. M. Ali, M. M. A. E. Hamid, and A. Youssif, "Sentiment analysis for movies reviews dataset

- using deep learning models," *International Journal of Data Mining & Knowledge Management Process (IJDKP)*, vol. 09, no. 03, pp. 19–27, may 2019, doi: 10.5121/ijdkp.2019.9302.
- [11] S. Styawati, N. Hendrastuty, and A. R. Isnain, "Analisis sentimen masyarakat terhadap program kartu prakerja pada twitter dengan metode support vector machine," *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 6, no. 3, pp. 150–155, 2021, [Online] Available: <http://ejournal.poltektegal.ac.id/index.php/informatika/article/view/2870>.
- [12] K. M. O. Nahar, M. Al-Shannaq, A. Manasrah, R. Alshorman, and I. Alazzam, "A holy quran reader/reciter identification system using support vector machine," *International Journal of Machine Learning and Computing*, vol. 9, no. 4, pp. 458–464, aug 2019, doi: 10.18178/ijmlc.2019.9.4.826.
- [13] J. Saranya and R. R. M.E, "Classification of benign and malignant tumour using svm," *International Journal of Recent Trends in Engineering & Research (IJRTER)*, vol. 05, no. Special Issue 07, pp. 105–110, mar 2019, doi: 10.23883/ijrter.conf.20190304.017.hrer2.
- [14] A. Priyanto and M. R. Maárif, "Implementasi web scrapping dan text mining untuk akuisisi dan kategorisasi informasi dari internet (studi kasus: Tutorial hidroponik)," *Indonesian Journal of Information Systems*, vol. 1, no. 1, pp. 25–33, aug 2018, doi: 10.24002/ijis.v1i1.1664.
- [15] W. Bourequat and H. Mourad, "Sentiment analysis approach for analyzing iPhone release using support vector machine," *International Journal of Advances in Data and Information Systems*, vol. 2, no. 1, pp. 36–44, apr 2021, doi: 10.25008/ijadis.v2i1.1216.
- [16] S. Menaka and N. Radha, "Text classification using keyword extraction technique," *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 12, pp. 734–740, 2013.
- [17] D. R. Manalu, M. C. L. Tobing, and M. Yohanna, "Analisis sentimen twitter terhadap wacana penundaan pemilu dengan metode support vector machine," *METHOMIKA Jurnal Manajemen Informatika dan Komputerisasi Akuntansi*, vol. 6, no. 6, pp. 149–156, oct 2022, doi: 10.46880/jmika.vol6no2.pp149-156.
- [18] O. Somantri, S. Wiyono, and D. Dairoh, "Metode k-means untuk optimasi klasifikasi tema tugas akhir mahasiswa menggunakan support vector machine (SVM)," *Scientific Journal of Informatics*, vol. 3, no. 1, pp. 34–45, jun 2016, doi: 10.15294/sji.v3i1.5845.
- [19] M. Belouch, S. E. Hadaj, and M. Idhammad, "Performance evaluation of intrusion detection based on machine learning using apache spark," *Procedia Computer Science*, vol. 127, pp. 1–6, 2018, doi: 10.1016/j.procs.2018.01.091.
- [20] D. T. Wisudawati, T. W. Utami, and P. R. Arum, "Analisis sentimen terhadap dampak covid-19 pada performa tokopedia menggunakan support vector machine," in *Seminar Nasional Variansi*, 2020.
- [21] A. P. Putra, N. N. Debatara, and D. Kusnandar, "Tingkat akurasi klasifikasi jarak kelahiran di kampung keluarga berencana (kb) dengan metode support vector machine (svm)," *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 9, no. 3, 2020, [Online] Available: <https://jurnal.untan.ac.id/index.php/jbmstr/article/view/40891>.
- [22] H. Yun, "Prediction model of algal blooms using logistic regression and confusion matrix," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 3, p. 2407, jun 2021, doi: 10.11591/ijece.v11i3.pp2407-2413.



This article is an open-access article distributed under the terms and conditions of the [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). Editorial of JJoM: Department of Mathematics, Universitas Negeri Gorontalo, Jln. Prof. Dr. Ing. B.J. Habibie, Moutong, Tilongkabila, Kabupaten Bone Bolango, Provinsi Gorontalo 96554, Indonesia.