



Comparative Evaluation of SVM Kernels for Sentiment Classification in Fuel Price Increase Analysis

Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM

Salsabila Rabbani¹, Dea Safitri², Nadila Rahmadhani³,
Al Amin Fadillah Sani⁴, M. Khairul Anam^{5*}

^{1,2,4,5}Teknik Informatika, STMIK Amik Riau, Indonesia

³Sistem Informasi, STMIK Amik Riau, Indonesia

E-Mail: ¹201003182104@sar.ac.id, ²2010031802001@sar.ac.id, ³2210031806049@sar.ac.id,
⁴2110031802011@sar.ac.id, ⁵khairulanam@sar.ac.id

Received Aug 10th 2023; Revised Sept 12th 2023; Accepted Oct 15th 2023
Corresponding Author: M. Khairul Anam

Abstract

The policy of changing the price of fuel oil (BBM) by the government in September 2022 caused controversy among social media users including Twitter. To understand how the change in fuel price increase affects people's perceptions and emotions on Twitter, this study conducted sentiment analysis using the Support Vector Machine (SVM) algorithm with three different kernel types, namely linear, RBF (Radial Basis Function), and polynomial. This research aims to classify tweets as positive, negative, or neutral, and compare the performance of the three SVM kernels. This research also tries to overcome class imbalance by applying the SMOTE (Synthetic Minority Over-sampling Technique) oversampling technique to the dataset. The results showed that Twitter users predominantly gave negative reactions to the fuel price increase. In applying the SVM algorithm, the RBF kernel produces the best performance of 87% using TF-IDF word weighting. In addition, the use of TF-IDF word weighting has the best accuracy results compared to the BoW word weighting model. The application of SMOTE oversampling technique in the polynomial kernel of TF-IDF word weighting on 70:30 and 80:20 data division successfully improved the algorithm performance by 2%. The results of this study provide an in-depth insight into the public's views on fuel price policy and enable public decision-makers and industry to design more responsive and pro-people policies.

Keyword: Classification, Fuel Oil, Sentiment Analysis, SMOTE, Support Vector Machine

Abstrak

Kebijakan perubahan harga Bahan Bakar Minyak (BBM) oleh pemerintah pada September 2022 lalu menimbulkan kontroversi pengguna sosial media termasuk *Twitter*. Untuk memahami bagaimana perubahan kenaikan harga BBM apakah mempengaruhi persepsi dan emosi masyarakat di *Twitter* maka dalam penelitian ini dilakukan analisis sentimen menggunakan algoritma *Support Vector Machine* (SVM) dengan tiga jenis kernel berbeda, yaitu linier, RBF (*Radial Basis Function*), dan polinomial. Penelitian ini bertujuan untuk mengklasifikasikan *tweet-tweet* sebagai positif, negatif, atau netral, serta membandingkan kinerja ketiga kernel SVM tersebut. Penelitian ini juga mencoba mengatasi ketidakseimbangan kelas dengan menerapkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) *oversampling* pada dataset. Hasil penelitian menunjukkan bahwa pengguna *Twitter* dominan memberikan reaksi negatif terhadap kenaikan harga BBM. Dalam mengaplikasikan algoritma SVM, kernel RBF menghasilkan kinerja terbaik yaitu sebesar 87% menggunakan pembobotan kata TF-IDF. Selain itu, penggunaan pembobotan kata TF-IDF memiliki hasil akurasi terbaik dibandingkan dengan model pembobotan kata BoW. Penerapan teknik SMOTE *oversampling* dalam kernel *polynomial* pembobotan kata TF-IDF pada pembagian data 70:30 dan 80:20 berhasil meningkatkan kinerja algoritma sebesar 2%. Hasil penelitian ini memberikan wawasan mendalam tentang pandangan masyarakat terhadap kebijakan harga BBM dan memungkinkan pengambil keputusan publik serta industri untuk merancang kebijakan yang lebih responsif dan berpihak kepada kepentingan rakyat.

Kata Kunci: Analisis Sentimen, BBM, Klasifikasi, SMOTE, Support Vector Machine

1. PENDAHULUAN

Kenaikan harga Bahan Bakar Minyak (BBM) telah menjadi isu global yang kompleks dengan dampak yang signifikan pada berbagai aspek ekonomi, sosial, dan lingkungan. Bahan bakar minyak, terutama bahan bakar fosil seperti bensin dan solar, memainkan peran sentral dalam perekonomian modern dengan menjadi sumber energi dominan dalam sektor transportasi, industri, dan kebutuhan rumah tangga. Pada awal bulan September 2022 pemerintah menetapkan kebijakan kenaikan harga BBM di Indonesia. Kenaikan harga BBM telah memicu perhatian luas dari berbagai pihak, termasuk pemerintah, konsumen, industri, dan peneliti. Kelabilan harga bahan bakar minyak dapat mempengaruhi biaya aktivitas masyarakat dalam kegiatan sehari-hari [5]. Topik kenaikan harga bahan bakar minyak ini menjadi polemik di seluruh media sosial, salah satunya media sosial Twitter. Pengguna Twitter dapat mengeluarkan pendapat mereka secara pribadi, seperti masukan, kritikan ataupun pujian.

Semakin banyaknya pengguna media sosial yang memanfaatkan Twitter untuk mengungkapkan pendapatnya, maka diperlukan tindakan berupa analisis sentimen untuk memecahkan masalah berdasarkan opini publik. Berbagai metode dapat digunakan untuk menganalisis tanggapan masyarakat khususnya di media sosial, salah satunya adalah sentimen analisis. Analisis sentimen adalah disiplin ilmu yang memeriksa dan mengevaluasi pendapat atau pandangan, ekstraksi opini, penambangan sentimen, analisis pengaruh dan emosi terhadap suatu peristiwa atau topik [3]. Selain menentukan sentimen analisis, penelitian ini juga menggunakan *text mining* untuk mempermudah proses klasifikasi yang membutuhkan bantuan algoritma.

Penelitian ini menerapkan metode Support Vector Machine (SVM) sebagai algoritma untuk melakukan klasifikasi, SVM akan memprediksi label ke daerah kelas mana yang merupakan tempat dari data tersebut [4]. Model *Support Vector Machine* (SVM) membangun *hyperplane* (bidang pemisah) yang memiliki margin maksimum antara dua kelas, sehingga dapat memaksimalkan pemisahan antara kelas positif dan kelas negatif. Penelitian [18] RBF-SVM digunakan untuk memprediksi keadaan cacat, dan memiliki akurasi prediksi yang lebih baik dibandingkan dengan tiga metode yang umum digunakan yaitu SVM linier, ANN, dan pohon keputusan. Penelitian tersebut mendapatkan hasil kinerja RBF-SVM sebesar 92,58%.

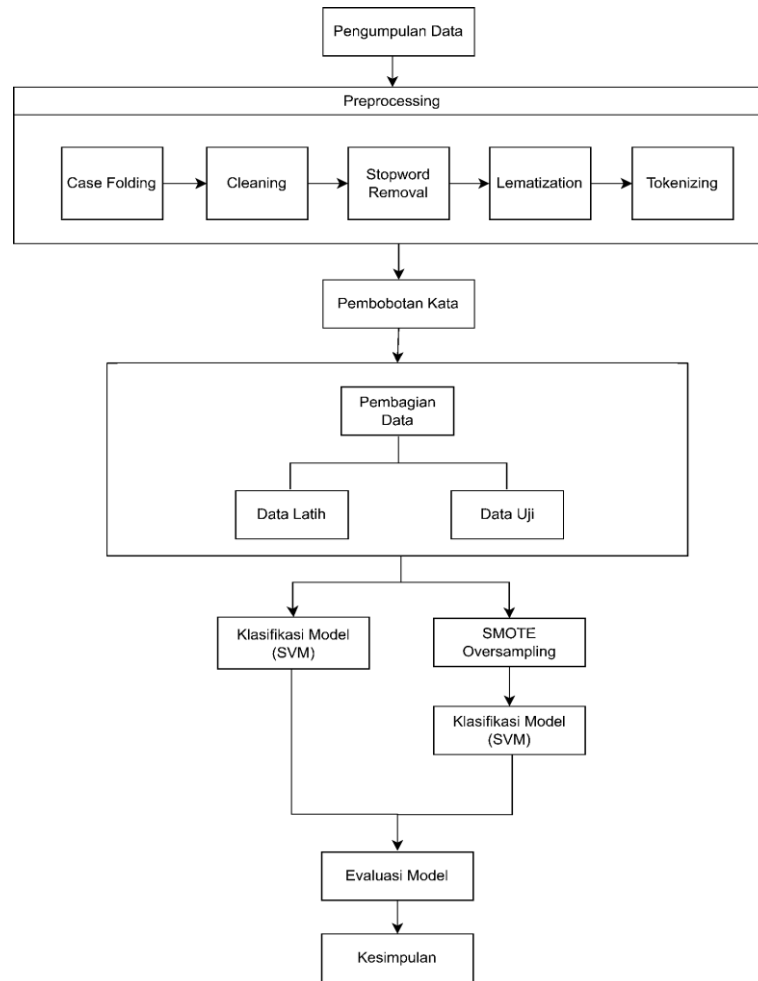
Dalam penelitian [1] yang membahas kepuasan pelanggan terhadap layanan Traveloka menunjukkan bahwa dengan menggunakan model SVM memperoleh akurasi tertinggi dari model *logistic regression* dan *naive bayes*. Dengan menggunakan pembobotan kata TF-IDF, SVM mendapatkan akurasi model sebesar 84,58%. Penelitian lain dari [2] membuktikan bahwa model SVM dengan pembagian data 70:30 menghasilkan akurasi model sebesar 85,98%. Hal ini menunjukkan bahwa model SVM adalah pemodelan yang baik untuk klasifikasi teks.

Penelitian ini juga juga mencoba mengatasi ketidakseimbangan label dengan menerapkan teknik SMOTE (*Synthetic Minority Over-sampling Technique*) *oversampling* pada dataset. SMOTE menghasilkan sampel sintetik dari kelas minoritas dengan melakukan *oversampling* pada setiap titik data dengan mempertimbangkan kombinasi linier dari tetangga kelas minoritas yang ada [6]. Dalam penelitian [17] menggunakan metode SMOTE *oversampling* dengan membandingkan beberapa pengklasifikasi pembelajaran mesin seperti, *Regresi Logistik*, *Random Forest*, dan *XGBoost*. Hasilnya model *Random Forest* mencapai kinerja yang lebih baik di antara algoritma tersebut dengan akurasi 99,07%, presisi dan perolehan 99,0%. Penelitian lain yang membahas mengenai penggunaan SMOTE *oversampling* [22] menggunakan model prediktif yang berbeda berhasil menaikkan kinerja model termasuk model SVM dengan hasil akurasi sebesar 98,9%. Penambahan SMOTE *oversampling* telah terbukti memahami dampak fitur terhadap pengembangan model.

Berdasarkan beberapa penelitian diatas, maka peneliti tertarik untuk melakukan penelitian dengan menggunakan perbandingan penggunaan SMOTE *oversampling* dalam kernel SVM untuk menentukan kinerja terbaik dalam mengolah dataset sentimen analisis kenaikan harga BBM. Oleh karena itu, melalui penelitian ini diharapkan dapat memberikan informasi bagaimana tanggapan pengguna Twitter tentang kebijakan pemerintah terhadap kenaikan harga bahan bakar minyak. Dengan memahami bagaimana persepsi dan sikap masyarakat berkembang terhadap kenaikan harga bahan bakar minyak memungkinkan pengambil keputusan publik serta industri untuk merancang kebijakan yang lebih responsif dan berpihak kepada kepentingan rakyat.

2. METODOLOGI PENELITIAN

Penelitian sentimen analisis terkait kenaikan harga BBM ini terdiri dari proses yang terencana dan sistematis untuk memberikan solusi terbaik bagi masalah. Data yang diambil dari *opini* masyarakat pada media sosial *Twitter*, penting untuk memastikan bahwa dataset relevan dengan permasalahan yang ada, dan berisi data yang cukup untuk melatih serta mengevaluasi model. Pada Gambar 1, terdapat urutan langkah-langkah yaitu mengumpulkan data, kemudian tahap pra-pemrosesan data, selanjutnya tahap *splitting* data, dengan membagi data latih dan data uji, selanjutnya pada data latih, tahap klasifikasi model serta menerapkan teknik SMOTE *oversampling*, setelah itu dilakukan tahap evaluasi model, dan diakhiri dengan kesimpulan dari hasil perbandingan metode *Support Vector Machine*.



Gambar 1. Metode Penelitian

2.1 Pengumpulan Data

Penelitian ini menggunakan dataset dari informasi *tweet* yang diambil dari platform *Twitter* melalui *crawling* data menggunakan API *Twitter* yang diperoleh dari pembuatan akun developer yang akan memberikan API Key dan Access Token unik. Jumlah data yang digunakan sebanyak 5.000 *tweet*. *Keyword* yang dipakai untuk mengambil data *tweet* adalah frasa yang mengandung frasa "Bahan Bakar Minyak". Pelabelan setiap data dilakukan secara manual dan divalidasi oleh seorang psikologi, ada 3 label yang diberikan, yaitu kelas negatif, positif dan netral. Beberapa pustaka yang digunakan dalam pengujian menggunakan bahasa pemrograman Python melalui IDE Jupyter, yang meliputi *pandas*, *numpy*, *nltk*, *matplotlib*, dan *wordcloud*.

2.2 Preprocessing (Pemrosesan Awal)

Dataset yang terkumpul masih dalam bentuk *unstructured* data, sehingga perlu dilakukan *preprocessing* agar data menjadi lebih terstruktur. Proses *preprocessing* yang dilakukan dalam penelitian ini melibatkan beberapa tahapan, yaitu perubahan huruf menjadi huruf kecil (*case folding*), pembersihan data (*cleaning*), penyaringan kata (*filtering*) atau *stopword removal*, pengembalian kata dasar (*lemmatization*), dan pemisahan kata-kata menjadi token (*tokenizing*),

1. Tahap *Case folding* melibatkan proses di mana huruf-huruf kapital diubah menjadi huruf kecil secara keseluruhan [7]. Penerapan *case folding* juga memiliki manfaat dalam menangani ketidakkonsistenan penulisan dalam data, seperti penggunaan yang tidak sesuai standar antara huruf besar dan kecil atau masalah ejaan yang mungkin terjadi. [2].
2. Tahap *Cleaning* yaitu tahap menghilangkan kata yang tidak diperlukan [7]. Misalnya, angka (1), pemisah kata yaitu koma (,), titik (.), tanda seru (!), dan tanda baca lainnya. Tokenisasi adalah memisahkan sepotong teks menjadi kata-kata atau frase [8].
3. Selanjutnya tahap *filtering*, dimana dilakukan seleksi kata-kata yang dianggap penting dari hasil tokenisasi [9]. Kata-kata yang tidak memberikan informasi seperti yang, dan, di, atau, dengan, dll., akan dihapus.

4. Kemudian lematisasi akan diterapkan pada ulasan yang telah di *filtering*. Lematisasi mengacu pada transformasi kata-kata ke dalam bentuk akarnya untuk meningkatkan kinerja model [10]. Contoh: *menolong, ditolong, penolong* akan menjadi *tolong*.
5. Token berupa kata unik yang akan menjadi identifikasi untuk pengelompokan sentimen. Sebelum melakukan tokenisasi, kata-kata dalam *tweet* dipisahkan menggunakan karakter spasi. Data *tweet* diubah menjadi huruf kecil sebelum diproses. Tautan, nama pengguna, dan simbol dihapus, dan kesalahan ejaan atau penulisan diperbaiki agar teks menjadi lebih akurat.

2.3 Pembobotan Kata

Pembobotan kata adalah suatu proses yang mengubah kata menjadi representasi numerik (vektor kata) [11]. Penelitian ini menggunakan dua metode pembobotan kata untuk mengetahui metode mana yang menghasilkan nilai frekuensi kata yang baik. Pembobotan kata yang dilakukan yaitu metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dan *Bag of Word* (BoW).

2.3.1 Term Frequency-Inverse Document Frequency (TF-IDF)

Dalam dokumen, seringkali digunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) pada penyematan kata. Dalam TF-IDF, dilakukan konversi data teks menjadi bentuk vektor agar dapat mempertimbangkan urutan kata yang tepat. Setiap kata dalam korpus berkorelasi dengan angka oleh TF-IDF yang menunjukkan seberapa signifikan setiap kata untuk korpus [14]. Setelah kata-kata diubah menjadi angka, nilai numerik TF-IDF diumpankan ke pengklasifikasi pembelajaran yang diawasi [15] dalam konteks yang dapat ditafsirkan oleh metode pembelajaran mesin. Untuk melakukan perhitungan TF-IDF menggunakan *library Python Sklearn*, kita bisa menggunakan *TfidfVectorizer*. Berikut rumus pembobotan kata.

$$W_{ij} = tf_{ij} \times \log \left(\frac{n}{df} \right) \quad (1)$$

Dimana W_{ij} adalah bobot yang menentukan pentingnya kata atau *term*, tf_{ij} adalah frekuensi *term* yaitu berapa kali *term* muncul, dan df adalah frekuensi dokumen total dari *term* dalam seluruh koleksi dokumen.

2.3.2 Bag Of Word (BoW)

Teknik *Bag of Word* (BoW) digunakan untuk mengekstraksi fitur dari data teks [14]. Haluan ini mudah diimplementasikan dan dipahami selain merupakan metode paling sederhana untuk mengekstrak fitur dari data teks. Teknik ini sangat cocok dan berguna untuk pemodelan bahasa dan klasifikasi teks. *Library "CountVectorizer"* digunakan untuk mengimplementasikan BoW. *CountVectorizer* menghitung kemunculan kata-kata dan membuat matriks database cadangan kata-kata [15]. BoW merupakan kumpulan kata atau fitur, dimana setiap fitur dikategorikan sebagai label yang menandakan kemunculan fitur yang dikategorikan tersebut. Berikut representasi dari BoW.

$$BoW(d) = [tf(t_1, d), tf(t_2, d), \dots, tf(t_n, d)] \quad (2)$$

Dimana, $BoW(d)$ adalah representasi vektor BoW untuk dokumen 'd', dan t_1, t_2, \dots, t_n adalah kata-kata unik dalam kamus.

2.4 Pembagian Data

Dalam memisahkan data, dilakukan pembagian ke dalam dua kelompok, yakni data latih dan data uji. Ada tiga percobaan pemisahan data yang dilakukan, yaitu:

1. Pada percobaan pertama, data latih terdiri dari 70% dan data uji 30%.
2. Pada percobaan kedua, data latih terdiri dari 80% dan data uji 20%.
3. Sementara pada percobaan ketiga, data latih terdiri dari 90% dan data uji 10%.

2.4.1 Data Latih (*Data Training*)

Penggunaan data latih (*Training*) untuk melatih sistem penelitian ini. Data pelatihan digunakan untuk mengajar metode klasifikasi SVM sehingga dapat memahami cara mengategorikan komentar menjadi kategori negatif, netral, atau positif.

2.4.2 Data Uji (*Data Testing*)

Setelah model klasifikasi dilatih, selanjutnya adalah mengevaluasi kinerja metode tersebut menggunakan data pengujian. Pengujian data bertujuan untuk menguji metode klasifikasi SVM dengan memasukkan data baru, lalu metode tersebut akan melakukan klasifikasi yang tepat terhadap data baru tersebut, mengategorikannya sebagai negatif, netral, atau positif..

2.5 SMOTE Oversampling

Penelitian ini menggunakan SMOTE *oversampling* untuk mengatasi ketidakseimbangan dalam penambahan data. Dalam pembelajaran ketidakseimbangan kita memiliki satu atau lebih kelas dengan data yang sangat sedikit, dan beberapa lainnya dengan jumlah data yang cukup. Pendekatan tingkat algoritma menyesuaikan spesifik algoritma klasifikasi untuk memperhitungkan masalah ketidakseimbangan [19]. Dataset yang tidak seimbang mengandung rasio kelas target yang tidak seimbang dan dapat menimbulkan masalah tugas klasifikasi karena model bisa terlalu cocok pada kelas mayoritas [21]. Untuk mengatasi masalah ini, dilakukan teknik pengambilan sampel ulang data, yaitu SMOTE *oversampling*. SMOTE adalah metode *oversampling* paling populer yang diusulkan untuk meningkatkan *oversampling* acak. Ada beberapa variasi SMOTE yang bertujuan untuk memerangi kelemahan algoritma asli [20]. Dalam penelitian ini, kami menyajikan analisis karakteristik distribusi sampel sintesis yang dihasilkan oleh SMOTE. Ini membantu dalam mengevaluasi kualitas data, dalam arti seberapa baik data yang dihasilkan meniru distribusi dasar yang sebenarnya.

2.6 Klasifikasi Model

Penelitian ini mencoba untuk menerapkan empat metode klasifikasi yang berbeda untuk menganalisis sentimen pada data *tweet*. Dalam penelitian ini, algoritma klasifikasi yang diterapkan adalah *Support Vector Machine* (SVM) yang bertujuan untuk membandingkan dan mengevaluasi kinerja masing-masing metode dalam mengklasifikasikan sentimen *tweet*.

Metode SVM (*Support Vector Machine*) mengandalkan vektor pendukung untuk memisahkan kelas dari data yang memiliki ciri-ciri yang berbeda-beda [12]. Konsep dasar dari algoritma SVM ini yaitu mencari *hyperplane* yang optimal. *Hyperplane* merupakan fungsi yang berfungsi sebagai pemisah antar data [13]. SVM mencoba untuk menemukan *hyperplane* yang paling baik memisahkan data *tweet* dengan sentimen negatif, netral, dan positif. SVM terdapat dua jenis *hyperplane* yang penting: *hyperplane* linear dan *hyperplane* non-linear. Jika data dapat dipisahkan dengan sempurna oleh sebuah *hyperplane* linear, SVM disebut sebagai SVM linier. Namun, jika data tidak dapat dipisahkan dengan sempurna secara linear, SVM menggunakan teknik transformasi kernel untuk mengubah data ke dalam dimensi fitur yang lebih tinggi, dimana *hyperplane* linier dapat dibuat. Dengan menggunakan berbagai jenis fungsi kernel, seperti kernel linier, kernel polinomial atau kernel Gauss, SVM dapat mengatasi data yang memiliki karakteristik non-linier dengan efisien. Fungsi kernel adalah fungsi yang mengubah data ke dimensi yang lebih tinggi dengan tujuan meningkatkan struktur data sehingga mempermudah proses pemisahan [16]. Dalam hal ini, rumus umum untuk SVM linear dapat dituliskan sebagai berikut:

$$f(x) = \text{sign}(w \cdot x + b) \quad (3)$$

Di mana, $f(x)$ adalah fungsi prediksi, w adalah vektor normal *hyperplane*, x adalah vektor fitur input, dan b adalah bias atau *intercept*.

Pada penelitian ini, ada 3 kernel yang digunakan untuk mengatasi masalah klasifikasi, yaitu linear, polinomial dan *Radial Basis Function* (RBF). Berikut adalah penjelasan tentang ketiga fungsi kernel tersebut beserta rumusnya:

1. Kernel Linear

Kernel linear mengukur produk titik (dot product) dari dua vektor input dalam ruang asli tanpa melakukan transformasi ke ruang fitur yang lebih tinggi.

$$K(x, y) = (x \cdot y) \quad (4)$$

2. Kernel Polinomial

Polinomial mengukur hubungan polinomial antara dua vektor input dalam ruang asli.

$$K(x, y) = (x \cdot y + c)^d \quad (5)$$

3. Kernel Radial Basis Function (RBF)

Kernel RBF menggunakan fungsi *Gaussian* (dikenal sebagai fungsi basis radial) untuk mengukur kesamaan antara dua vektor input dalam ruang fitur.

$$K(x, x') = \exp(-\gamma(x, x')^2) \quad (6)$$

Dimana C adalah parameter Cost, γ adalah parameter gamma, c adalah *coefficient* dan d adalah pangkat atau derajat polinomial. Parameter gamma mengindikasikan kemiringan, yang merujuk pada perubahan output sehubungan dengan perubahan input. Sementara itu, parameter *coefficient* mencerminkan titik

potong (intercept) dari model, sedangkan parameter pangkat atau derajat menggambarkan tingkat polinomial dari fungsi tersebut.

2.7 Evaluasi Model

Diperlukan pengujian model untuk mengevaluasi kinerja dari metode *Support Vector Machine (SVM)*. Dalam penelitian dievaluasi menggunakan nilai akurasi. Nilai akurasi adalah matrik evaluasi yang digunakan untuk mengukur seberapa baik model klasifikasi atau prediksi dapat memberikan hasil yang benar atau sesuai dengan data yang ada. Akurasi menggambarkan persentase keberhasilan model dalam memprediksi kelas atau label yang tepat.

Mengukur sejauh mana model dapat mengklasifikasikan data dengan benar secara keseluruhan. Dalam kata lain, akurasi menggambarkan persentase keberhasilan model dalam memprediksi kelas yang tepat.

$$\text{Akurasi} = \frac{(\text{Jumlah Prediksi Benar})}{(\text{Total Jumlah Data})} \quad (7)$$

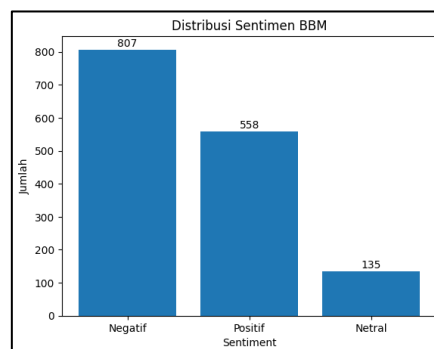
3. HASIL DAN PEMBAHASAN

Terdapat sebanyak 5000 data mentah yang dilabelkan menjadi 3 kategori yaitu, positif, negatif, dan netral. Data mentah akan melalui *preprocessing text* terlebih dahulu agar data yang digunakan lebih akurat nantinya. Langkah dalam *preprocessing data* yang dilakukan yaitu meliputi penghapusan redundansi data, link, Username, karakter, mengubah ke huruf kecil, melakukan *stopword*, *stemming*, dan tokenisasi.

Tabel 1. Hasil preprocessing text

Tweets	Cleaning	Stopword Removal	Lemmatization	Tokenizing
partai demokrat instruksikan kadernya turun ke...	partai demokrat instruksikan kadernya turun ke...	partai demokrat instruksikan kadernya turun jl...	partai demokrat instruksi kader turun jln aksi...	[partai, demokrat, instruksi, kader, turun, jl...
kasat lantas polres prabumulih akp muthemainah... terima audiensi driver ojol pemkab garut tola...	kasat lantas polres prabumulih akp muthemainah... terima audiensi driver ojol pemkab garut tolak...	kasat lantas polres prabumulih akp muthemainah... terima audiensi driver ojol pemkab garut tolak...	kasat lantas polres prabumulih akp muthemainah... terima audiensi driver ojol pemkab garut tolak...	[kasat, lantas, polres, prabumulih, akp, muthe... [terima, audiensi, driver, ojol, pemkab, garut...

Setelah dilakukannya *preprocessing text* ternyata terdapat 1500 data bersih yang akan digunakan untuk diproses selanjutnya. Sebelum itu, data yang sudah dibersihkan akan dihitung label sentimennya untuk mengetahui seberapa banyak sentimen publik terhadap kenaikan harga BBM.



Gambar 2. Hasil sentimen kenaikan harga BBM

Gambar diatas menunjukkan bahwa ternyata pengguna Twitter dominan memberikan reaksi negatif terhadap kenaikan harga BBM. Dalam penelitian ini selanjutnya akan menampilkan *wordcloud* untuk label dengan sentimen negatif.

Visualisasi pada gambar 3 menunjukkan kata-kata yang sering muncul dalam sentimen negatif. Proses selanjutnya yaitu, mengimplementasikan data kedalam algoritma SVM dengan 3 kernelnya yaitu, *RBF*, *Linear*, dan *Polynomial* menggunakan *splitting data* 70:30, 80:30, dan 90:10. Penelitian ini akan dilakukan menggunakan 2 metode pembobotan kata yaitu TF-IDF dan BoW.

- [3] Liu, B. (2022). *Sentiment analysis and opinion mining*. Springer Nature.
- [4] Fikri, M.I., Sabrila, T.S., Azhar, Y., Malang, U.M., 2020. Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter. *SMATIKA Jurnal* 10.
- [5] Kurniasih, U., & Suseno, A. T. (2022). Analisis Sentimen Terhadap Bantuan Subsidi Upah (BSU) pada Kenaikan Harga Bahan Bakar Minyak (BBM). *Jurnal Media Informatika Budidarma*, 6(4), 2335-2340.
- [6] Turlapati, V. P. K., & Prusty, M. R. (2020). Outlier-SMOTE: A refined oversampling technique for improved detection of COVID-19. *Intelligence-based medicine*, 3, 100023.
- [7] Sidik, F., Suhada, I., Anwar, A. H., & Hasan, F. N. (2022). Analisis Sentimen Terhadap Pembelajaran Daring Dengan Algoritma Naive Bayes Classifier. *Jurnal Linguistik Komputasional*, 5(1), 34-43.
- [8] A. S. Alammary, "Arabic Questions Classification Using Modified TF-IDF," in *IEEE Access*, vol. 9, pp. 95109-95122, 2021, doi: 10.1109/ACCESS.2021.3094115.
- [9] Khaira, U., Johanda, R., Utomo, P. E. P., & Suratno, T. (2020). Sentiment analysis of cyberbullying on twitter using SentiStrength. *Indones. J. Artif. Intell. Data Min*, 3(1), 21.
- [10] Al-Saqqa, S., Awajan, A., & Ghoul, S. (2019, October). Stemming effects on sentiment analysis using large arabic multi-domain resources. In *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS)* (pp. 211-216). IEEE.
- [11] Rahmadden, R., & Akbar, F. Comparison of Naïve Bayes Algorithm, Support Vector Machine and Decision Tree in Analyzing Public Opinion on COVID-19 Vaccination in Indonesia. *Indonesian Journal of Artificial Intelligence and Data Mining*, 6(1), 8-17.
- [12] Pratama, M. O., Satyawan, W., Jannati, R., Pamungkas, B., Syahputra, M. E., & Neforawati, I. (2019, April). The sentiment analysis of Indonesia commuter line using machine learning based on twitter data. In *Journal of Physics: Conference Series* (Vol. 1193, No. 1, p. 012029). IOP Publishing.
- [13] Rahman, O. H., Abdillah, G., & Komarudin, A. (2021). Klasifikasi Ujara Kebencian pada Media Sosial Twitter Menggunakan Support Vector Machine. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 5(1), 17-23.
- [14] Rupapara, V., Rustam, F., Shahzad, H. F., Mehmood, A., Ashraf, I., & Choi, G. S. (2021). Impact of SMOTE on imbalanced text features for toxic comments classification using RVVC model. *IEEE Access*, 9, 78621-78634.
- [15] Eshan, S. C., & Hasan, M. S. (2017, December). An application of machine learning to detect abusive bengali text. In *2017 20th International Conference of Computer and Information Technology (ICCI)* (pp. 1-6). IEEE.
- [16] Feta, N. R., & Ginanjar, A. R. (2019). Komparasi Fungsi Kernel Metode Support Vector Machine Untuk Pemodelan Klasifikasi Terhadap Penyakit Tanaman Kedelai. *BRITech, Jurnal Ilmiah Ilmu Komputer, Sains dan Teknologi Terapan*, 1(1), 33-39.
- [17] Ghosh, M. (2021, July). An Enhanced Stroke Prediction Scheme Using SMOTE and Machine Learning Techniques. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.
- [18] Wang, X., Yan, Z., Zeng, Y., Liu, X., Peng, X., & Yuan, H. (2021). Research on correlation factor analysis and prediction method of overhead transmission line defect state based on association rule mining and RBF-SVM. *Energy Reports*, 7, 359-368.
- [19] Elreedy, D., & Atiya, A. F. (2019). A comprehensive analysis of synthetic minority oversampling technique (SMOTE) for handling class imbalance. *Information Sciences*, 505, 32-64.
- [20] Douzas, G., Bacao, F., & Last, F. (2018). Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. *Information Sciences*, 465, 1-20.
- [21] Omar, B., Rustam, F., Mehmood, A., & Choi, G. S. (2021). Minimizing the overlapping degree to improve class-imbalanced learning under sparse feature selection: application to fraud detection. *IEEE Access*, 9, 28101-28110.
- [22] Bujang, S. D. A., Selamat, A., Ibrahim, R., Krejcar, O., Herrera-Viedma, E., Fujita, H., & Ghani, N. A. M. (2021). Multiclass prediction model for student grade prediction using machine learning. *IEEE Access*, 9, 95608-95621.