ANALISIS SENTIMEN PUBLIK TERHADAP KEBIJAKAN PEMERINTAH TERBARU TENTANG PENDISTRIBUSIAN GAS ELPIGI SUBSIDI PADA MASYARAKAT MENGGUNAKAN METODE ALGORITMA SVM

Fathoni, Gabriel Lifiano Jamot Munthe, M. Yasir Alghifari
Sistem Informasi, Universitas Sriwijaya
Jalan Palembang-Prabumulih km 32 Palembang, Indonesia
gabriellifiano@gmail.com

ABSTRAK

Kebijakan pemerintah terbaru mengenai pendistribusian gas elpiji subsidi hanya melalui pangkalan resmi Pertamina menuai berbagai respons dari masyarakat, khususnya di media sosial X (Twitter). Permasalahan muncul karena perubahan sistem ini dinilai menyulitkan masyarakat tertentu, serta menimbulkan keresahan terkait aksesibilitas dan kelangkaan pasokan. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kebijakan tersebut dengan memanfaatkan algoritma *Support Vector Machine* (SVM) sebagai metode klasifikasi. Data dikumpulkan melalui proses *crawling* menggunakan sejumlah kata kunci relevan selama periode 1 Februari hingga 28 Februari 2025. Setelah itu data diproses melalui tahapan *preprocessing*, normalisasi, ekstraksi fitur menggunakan metode N-gram, serta pelabelan sentimen secara otomatis dengan bantuan pustaka *TextBlob*. Berdasarkan hasil analisis terhadap 1.091 data, ditemukan bahwa sentimen publik didominasi oleh sentimen positif sebesar 47,3%, diikuti oleh sentimen negatif sebesar 27,0%, dan sentimen netral sebesar 25,7%. Evaluasi performa model menunjukkan tingkat akurasi sebesar 68,35%, dengan kinerja terbaik pada klasifikasi sentimen positif. Hasil penelitian ini menunjukkan bahwa masyarakat cenderung menerima kebijakan tersebut secara positif, meskipun masih terdapat kekhawatiran yang perlu diantisipasi melalui strategi komunikasi publik yang lebih informatif, terbuka, dan berkelanjutan.

Kata kunci : *Analisis Sentimen, Gas Elpiji Subsidi, Media Sosial, X, Support Vector Machine.*

1. PENDAHULUAN

Perkembangan kebijakan energi di Indonesia selalu menarik perhatian publik, terutama kebijakan pendistribusian gas elpigi yang bersubsidi. Pada tanggal 1 Februari 2025, pemerintah mengimplemntasikan kebijakan baru mengenai transformasi kegiatan pendistribusian gas elpigi dimana penjualan gas elpiji subsidi melalui pengecer dihentikan dan hanya disalurkan melalui pangkalan resmi dari PT Pertamina[1]. Kebijakan ini diterapkan guna memastikan bahwa pasokan gas yang terbatas dapat tepat sasaran serta harga jualnya sesuai dengan ketentuan yang berlaku.

Perubahan sistem distribusi ini tidak lepas dari berbagai permasalahan yang terjadi selama pelaksanaan konversi penggunaan elpiji. Pada sistem distribusi yang sebelumnya memungkinkan penjualan gas elpigi di pasar bebas menyebabkan ketidaksesuaian sasaran, di mana masyarakat kelas menengah ke atas turut memperoleh subsidi yang seharusnya diperuntukkan bagi masyarakat berpenghasilan rendah. Akibatnya, kelangkaan elpiji terjadi, sehingga menimbulkan dampak negatif bagi warga yang benar-benar membutuhkan[2].

Kebijakan pendistribusian elpiji sendiri merupakan respon pemerintah terhadap permasalahan distribusi subsidi yang telah terjadi sejak awal pelaksanaannya. Landasan hukum yang mendasari kebijakan ini, antara lain Undang-Undang No. 22/2001 tentang Minyak dan Gas Bumi, Peraturan Presiden No. 5/2006, Perpres No. 104/2007, serta Peraturan Menteri ESDM No. 26/2009 [3], menunjukkan bahwa kebijakan energi di Indonesia selalu terikat dengan aspek regulasi yang ketat. Namun, implementasi di lapangan masih menemui kendala terutama dalam hal distribusi yang tidak merata serta dampak sosial ekonomi yang timbul akibat kelangkaan elpiji.

Selain dari aspek teknis pendistribusian, dinamika kebijakan ini juga turut memicu respon dan opini dari masyarakat yang terekspresikan melalui berbagai media sosial dan forum diskusi online. Kondisi ini menimbulkan peluang bagi peneliti untuk melakukan analisis sentimen sebagai upaya mengukur persepsi publik terhadap kebijakan pemerintah. Dengan mengukur sentimen positif dan negatif, penelitian ini dapat memberikan gambaran yang lebih komprehensif mengenai dampak kebijakan tersebut serta mengidentifikasi isu-isu yang masih perlu penanganan.

Dalam era digital saat ini, analisis sentimen dapat memanfaatkan teknik pengolahan bahasa alami (*Natural Language Processing*/NLP) dan algoritma *machine learning* untuk mengklasifikasikan opini dari data teks. Salah satu metode yang terbukti efektif dalam melakukan klasifikasi teks adalah *Support Vector Machine* (SVM)[4]. *Support Vector Machine* (SVM) adalah suatu metode klasifikasi yang menentukan hyperplane optimal yaitu bidang pemisah dengan margin maksimum yang digunakan untuk memisahkan data ke dalam dua atau lebih kelas. Keunggulan SVM terletak pada pemanfaatan support vector dalam perhitungan jarak antarkelas sehingga proses komputasi tetap berjalan efisien. Dalam berbagai penelitian analisis sentimen terkini, metode

ini terbukti mampu menangani data berdimensi tinggi serta menghasilkan model klasifikasi dengan tingkat akurasi yang baik [5]. Penelitian-penelitian terdahulu telah menunjukkan bahwa metode SVM efektif dalam mengidentifikasi sentimen dari berbagai sumber data, mulai dari *review* produk hingga opini publik terkait kebijakan. *Support Vector Machine* (SVM) dapat menjadi pilihan algoritma klasifikasi yang terbukti unggul dalam menangani data opini publik, khususnya dalam konteks analisis sentimen berbasis teks. Penelitian yang dilakukan oleh Putri dan Saragih menunjukkan hasil bahwa SVM menghasilkan tingkat akurasi yang lebih tinggi dibandingkan algoritma lain seperti Naive Bayes ketika diterapkan pada analisis konflik sipil-militer dalam isu pengungsi Rohingya [6]. Algoritma ini mampu membedakan opini positif dan negatif dengan presisi tinggi sehingga menjadikannya metode yang efektif untuk menangkap nuansa emosi dalam teks. Hal serupa diungkapkan oleh Agarkar et al., dimana SVM secara konsisten menunjukkan performa tinggi dalam klasifikasi nada (*tone*) dan opini pada platform Reddit. Hal ini menegaskan peran SVM sebagai pilihan utama dalam klasifikasi teks berbasis opini masyarakat [7]. Implementasi metode ini dalam penelitian mengenai kebijakan pendistribusian elpiji diharapkan dapat memberikan insight yang mendalam mengenai persepsi masyarakat terhadap kebijakan subsidi yang sedang berlangsung.

Dengan menggabungkan analisis sentimen berbasis SVM dan tinjauan terhadap kebijakan pendistribusian elpiji penelitian ini bertujuan untuk memberikan gambaran menyeluruh mengenai respons publik terhadap kebijakan tersebut. Hasil analisis diharapkan tidak hanya bermanfaat bagi akademisi dalam pengembangan sistem informasi dan teknik analisis data, tetapi juga dapat menjadi masukan bagi pihak-pihak terkait dalam penyempurnaan kebijakan pendistribusian subsidi elpiji di masa depan.

2. TINJAUAN PUSTAKA

Dalam beberapa tahun sebelumnya, analisis sentimen publik telah banyak diterapkan untuk mengevaluasi respons masyarakat terhadap berbagai kebijakan pemerintah. Salah satunya adalah penelitian yang dilakukan oleh Isnain dkk.[8] yaitu menganalisis sentimen publik terhadap kebijakan *lockdown* pemerintah Jakarta dengan menggunakan algoritma *Support Vector Machine* (SVM) yang didukung oleh ekstraksi fitur TF-IDF. Penelitian ini dilakukan dengan cara mengumpulkan data tweet melalui teknik *crawling* dan menerapkan tahap *preprocessing* seperti *case folding*, *cleansing*, *stopword removal*, *stemming*, dan *tokenization* untuk mempersiapkan data yang akan dianalisis. Hasil dari penelitian ini menunjukkan distribusi sentimen dengan 68,75% tweet positif dan 31,25% tweet negatif, serta nilai akurasi sebesar 74%, *precision* 75%, *recall* 92%, dan F1-*score* 83%. Penelitian ini menyoroti keunggulan SVM yang efektif dan cepat dalam mengklasifikasikan data teks, meskipun terdapat kekurangan berupa kebutuhan tuning parameter untuk mengoptimalkan performa model, terutama ketika data tidak seimbang. Berdasarkan hasil dari temuan tersebut memberikan dasar kuat bahwa SVM dapat menjadi metode yang potensial untuk diterapkan pada analisis sentimen publik dalam konteks kebijakan pemerintah.

Penelitian lain mengenai analisis sentimen juga dilakukan oleh Rahmaddeni, dkk, pada tahun 2022 [9]. Penelitian ini mengangkat tema perbandingan antara SVM dan model hybrid SVM-XGBoost (XGBSVM) untuk menganalisis opini publik mengenai vaksinasi COVID-19 di Twitter. Dalam penelitian tersebut, data tweet dibagi dengan rasio 90:10 antara data pelatihan dan pengujian, di mana SVM murni menghasilkan akurasi sebesar 83% dan XGBSVM hanya mencapai 79%. Hasil tersebut menunjukkan bahwa meskipun metode hybrid dapat memberikan pendekatan yang lebih kompleks namun SVM tradisional tetap unggul dalam hal akurasi untuk analisis sentimen. Peneliti juga menyoroti bahwa pemilihan fitur dan teknik pembobotan seperti TF-IDF sangat berpengaruh terhadap performa klasifikasi. Studi ini mengindikasikan pentingnya pemilihan algoritma yang tepat dan pengoptimalan parameter dalam meningkatkan kinerja model analisis sentimen.

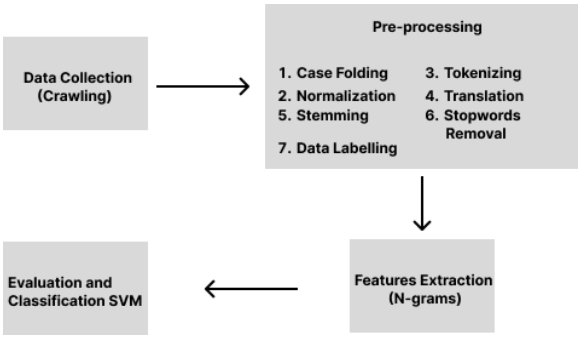
Dalam penelitian yang dilakukan oleh Yunita dan Kamayani yaitu melakukan perbandingan antara algoritma SVM dan Naïve Bayes untuk menganalisis sentimen terhadap kebijakan penghapusan kewajiban skripsi[10]. Hasil penelitian menunjukkan bahwa SVM memiliki performa yang lebih baik dengan akurasi mencapai 80%, *recall* 83%, *precision* 76%, dan F1-*score* 79%, dibandingkan dengan Naïve Bayes. Meskipun dataset yang digunakan relatif terbatas, penelitian ini berhasil mengungkapkan bahwa SVM lebih efektif dalam mengklasifikasikan data sentiment pada konteks kebijakan pendidikan. Kelebihan penelitian ini terletak pada penggunaan metrik evaluasi yang komprehensif, sementara kekurangannya adalah tidak adanya eksplorasi mendalam terhadap faktor-faktor eksternal yang mungkin mempengaruhi opini publik. Temuan dari studi ini memberikan wawasan bahwa SVM dapat dijadikan pilihan untuk mengembangkan model analisis sentimen yang lebih baik, khususnya dalam konteks kebijakan pendistribusian gas elpigi dimana parameter dan fitur tambahan perlu dieksplorasi guna mengakomodasi keragaman opini masyarakat.

Berdasarkan tinjauan literatur dari penelitian sebelumnya, peneliti mengusulkan metode SVM untuk diterapkan pada tweets topik kebijakan pendisitribusian gas elpigi terbaru untuk tujuan klasifikasi kelas sentimen pada media sosial X. *Support Vector Machine* (SVM) merupakan metode terbaik diantara beberapa metode lainnya karena mampu memetakan data ke dalam ruang berdimensi tinggi menggunakan teknik kernel trick, yang memungkinkan pemisahan kelas secara optimal dan akurasi yang lebih tinggi dalam pemrosesan data opini

public sehingga tingkat akurasi yang dihasilkan lebih baik [11].Teknis klasifikasi dilakukan dengan cara mengklasifikasikan menjadi 3 kelas yakni positif, netral dan negatif. Dengan penerapan SVM ini diharapkan mendapatkan hasil akurasi yang baik dan memberikan gambaran yang lebih komprehensif mengenai persepsi masyarakat serta menjadi acuan bagi pemerintah dalam merancang strategi distribusi energi yang lebih adaptif dan responsif.

3. METODE PENELITIAN

Tahap penelitian merupakan suatu rangkaian langkah sistematis yang dilakukan mulai dari identifikasi masalah hingga mendapatkan kesimpulan hasil penelitian. Alur penelitian terbagi atas 4 tahap yaitu pengambilan data dari media sosial X, *pre-processing data*, *features extraction* (TF-IDF), serta evaluasi dan klasifikasi algoritma SVM seperti yang terlihat pada Gambar 1.



Gambar 1. Alur Penelitian

3.1. *Data collection (Crawling)*

Data collecting atau *crawling* merupakan tahap awal penting dalam analisis sentimen yang bertujuan untuk mengumpulkan data mentah dari media sosial melalui API, web scraping, atau crawler khusus [12].

3.2. Tahap *preprocessing*

Preprocessing data adalah tahapan yang dilakukan untuk melakukan sebuah proses awal dalam pengolahan data. Pada tahap ini data yang telah dikumpulkan akan diolah untuk menghindarkan dari data yang mengganggu (noise) atau data yang tidak konsisten [13].

3.3. *Case folding*

Case folding merupakan tahapan proses normalisasi teks dengan mengubah semua huruf menjadi huruf kecil agar seragam dan mengurangi variasi kata[14]. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter(pembatas).

3.4. *Normalization*

Proses normalisasi adalah tahap yang dilakukan untuk membantu pengolahan dan analisis data untuk algoritma pengajaran mesin. Ini juga meningkatkan kualitas data dan mempermudah kompatibilitas antara aplikasi, sistem, dan tipe data[15].

3.5. *Stopwords removal*

Stopwords removal adalah proses menghapus kata-kata yang umum dan tidak memberikan banyak informasi penting dalam analisis teks. Kata-kata tersebut disebut "stop words". Tujuan dari stopword removal adalah untuk meningkatkan efisiensi dan akurasi analisis teks dengan menghilangkan kata-kata yang sering muncul namun memiliki sedikit kontribusi dalam pemahaman konten atau makna teks [16].

3.6. *Tokenizing*

Tokenizing adalah proses pemisahan teks menjadi unit-unit yang lebih kecil, seperti kata atau frasa, agar dapat diolah lebih lanjut dalam analisis teks [17].

3.7. *Stemming*

Stemming adalah proses pemetaan dan penguraian berbagai bentuk kata menjadi bentuk dasarnya . Proses pemetaan dan penguraian digunakan untuk menemukan kata dasar dari sebuah kata yang mengalami imbuhan dengan cara menghilangkan atau menghapus imbuhan-imbuhan tersebut [18].

3.8. *Translation*

Translation adalah proses menerjemahkan teks dari satu bahasa ke bahasa lain guna meningkatkan keterbacaan, konsistensi, atau kompatibilitas dalam pemrosesan bahasa alami [19].

3.9. *Data labeling*

Data labeling adalah suatu proses memberi sebuah kategori pada data teks berdasarkan sentimen yang terkandung di dalamnya [20].

3.10. *Features extraction* (n-grams)

Ekstraksi fitur dalam analisis sentimen adalah proses konversi teks mentah menjadi representasi numerik yang dapat diolah oleh algoritma pembelajaran mesin, seperti *Support Vector Machine* (SVM). Proses ini bertujuan untuk menangkap aspek penting dari data teks yang mencerminkan sentimen pengguna, baik positif, negatif, maupun netral.

Model N-Gram adalah teknik statistik yang digunakan dalam pengolahan bahasa alami untuk memperkirakan kemungkinan munculnya suatu kata berdasarkan urutan kata-kata sebelumnya dalam teks atau kalimat. Dalam model ini, teks atau kalimat dibagi menjadi sejumlah unit n-gram, yang masing-masing terdiri dari n kata yang berurutan. Misalnya, jika n = 3, maka unit n-gram tersebut akan terdiri dari tiga kata yang berdekatan dalam teks atau kalimat. Selanjutnya, model ini menghitung frekuensi kemunculan n-gram dalam teks atau korpus yang lebih besar dan menggunakan informasi tersebut untuk memprediksi kata yang mungkin muncul selanjutnya. Model ini berguna untuk berbagai aplikasi seperti analisis sentimen, klasifikasi teks, dan pengenalan suara [21].

3.11. Evaluasi dan klasifikasi SVM

Evaluasi *Support Vector Machine* (SVM) adalah langkah penting dalam mengukur kinerja model SVM dalam tugas klasifikasi atau regresi. Proses evaluasi ini umumnya melibatkan penggunaan metrik seperti akurasi, *precision*, *recall*, dan F1-*score*. Selain itu, teknik validasi seperti *cross-validation* juga diterapkan untuk memastikan bahwa model mampu melakukan generalisasi dengan baik terhadap data baru [22]

Klasifikasi SVM merupakan salah satu metode dalam pembelajaran mesin yang berfungsi untuk memisahkan data ke dalam dua kelas atau lebih berdasarkan *hyperplane* optimal. Apabila data tidak dapat dipecah secara linear, SVM memanfaatkan fungsi kernel untuk mentransformasi data ke dalam dimensi yang lebih tinggi, sehingga memungkinkan pemisahan yang lebih efektif.

4. HASIL DAN PEMBAHASAN

Pada bagian ini menjelaskan secara rinci setiap tahapan yang ditempuh dalam metode penelitian. Proses penelitian dilakukan dengan menggunakan bahasa pemrograman *Python* dan diimplementasikan pada platform *Google Colab*. Seluruh rangkaian kegiatan, mulai dari pengumpulan data, tahap *preprocessing* data, pembangunan model, hingga evaluasi kinerja model, dilaksanakan secara sistematis dengan memanfaatkan berbagai *library* yang tersedia.

4.1. *Data collection (Crawling)*

Penelitian ini menggunakan satu dataset yang diperoleh dari tiga kata kunci pencarian berbeda yang berasal dari media sosial X. Setelah data dikumpulkan berdasarkan ketiga kata kunci tersebut, seluruh file digabungkan menjadi satu dataset utama yang siap untuk dapat diolah. Proses pencarian data dilakukan dalam rentang waktu satu bulan yaitu antara tanggal 1 Februari 2025 hingga 28 April 2025. Untuk melakukan *crawling data*, digunakan *library* Tweet *Harvest* dengan menjalankan perintah kode program tertentu guna mengumpulkan data dari media sosial X. Menjalankan *command* "*search_keyword* = 'aturan baru elpiji lang:id'" untuk mengambil data pencarian pertama, lalu "*search_keyword* = 'gas elpiji subsidi lang:id'" untuk pencarian kedua, untuk pencarian ketiga, "*search_keyword* = 'kelangkaan elpiji subsidi lang:id'" untuk pencarian keempat. Hasil pengumpulan data setelah menjadi satu sebanyak 1091 data.

4.2. *Preprocessing data*

Tahap *preprocessing* merupakan langkah awal yang dilakukan untuk membersihkan dan menyiapkan data teks mentah sebelum dianalisis lebih lanjut. Tahap ini penting untuk meningkatkan akurasi model analisis sentiment dengan mengurangi kebisingan dalam data dan memastikan bahwa fitur yang digunakan dalam pemodelan benar-benar relevan terhadap makna sentimen dalam teks.

4.3. *Case folding*

Pada tahapan ini, data dibersihkan dengan cara memodifikasi, mengubah, atau menghapus data-data yang dianggap tidak perlu. Tahap ini menghilangkan angka, karakter spesial, tanda baca, link, hashtag, emoji, dan lain-lain.

Tabel 1. Case Folding

Sebelum	Sesudah
2025 Indonesia ga baik aja kasus pagar laut gas Elpiji 3 kg dana danantara korupsi Pertamina Kapan negara mau maju	indonesia ga baik aja kasus pagar laut gas elpiji kg dana danantara korupsi pertamina kapan negara mau maju

4.4. *Tokenizing*

Tokenizing dalam penelitian ini adalah tahapan dalam memecah string pada suatu teks yang telah melewati tahap data cleansing menjadi pecahan kata-kata yang disebut dengan token.

Tabel 2. Tokenizing

Sebelum	Sesudah
indonesia ga baik aja kasus pagar laut gas elpiji kg dana danantara korupsi pertamina kapan negara mau maju	['indonesia', 'ga', 'baik', 'aja', 'kasus', 'pagar', 'laut', 'gas', 'elpiji', 'kg', 'dana', 'danantara', 'korupsi', 'pertamina', 'kapan', 'negara', 'mau', 'maju']

4.5. *Normalization*

Normalisasi merupakan tahapan berikutnya untuk mengubah kata tidak baku atau *slang words* menjadi kata baku. Pada tahap ini, terlebih dahulu dibuat secara manual sebuah file Bernama "kamusnormalisasi.csv" yang berisi daftar bahasa tidak baku dan bahasa bakunya. Pada kode program normalisasi ini, *row* pertama merupakan bahasa tidak baku, *row* kedua merupakan bahasa baku. Pada bagian *row* pertama menggantikan semua kata di *row* kedua.

Tabel 3. Normalization

Sebelum	Sesudah
'Indonesia', 'ga', 'baik', 'aja', 'kasus', 'pagar', 'laut', 'gas', 'elpiji', 'kg', 'dana', 'danantara', 'korupsi', 'pertamina', 'kapan', 'negara', 'mau', 'maju'	['indonesia', 'tidak', 'baik', 'saja', 'kasus', 'pagar', 'laut', 'gas', 'elpiji', 'kg', 'dana', 'danantara', 'korupsi', 'pertamina', 'kapan', 'negara', 'mau', 'maju']

4.6. *Stopwords removal*

Pada tahapan ini terjadi proses selanjutnya yaitu menghapus kata-kata umum seperti "dan", "yang", "di", "ke", dan lain-lain yang biasanya tidak terlalu penting dalam analisis teks karena dianggap tidak memiliki makna penting dalam analisis.

Tabel 4. Stopwords removal

Sebelum	Sesudah
['indonesia', 'tidak', 'baik', 'saja', 'kasus', 'pagar', 'laut', 'gas', 'elpiji', 'kg', 'dana', 'danantara', 'korupsi', 'pertamina', 'kapan', 'negara', 'mau', 'maju']	['indonesia', 'baik', 'kasus', 'pagar', 'laut', 'gas', 'elpiji', 'kg', 'dana', 'danantara', 'korupsi', 'pertamina', 'kapan', 'negara', 'mau', 'maju']

4.7. *Stemming*

Pada tahapan ini data akan mencari *root* (dasar) kata dari tiap kata hasil filtering dengan menghapus kata imbuhan di depan maupun imbuhan di belakang kata.

Tabel 5. Stemming

Sebelum	Sesudah
['kamis', 'kepala', 'kejaksaan', 'negeri', 'aceh', 'besar', 'diwakili', 'kepala', 'seksi', 'intelijen', 'filman', 'ramadhan', 'sh', 'mh']	['kamis', 'kepala', 'jaksa', 'negeri', 'aceh', 'besar', 'wakil', 'kepala', 'seksi', 'intelijen', 'film', 'ramadhan', 'sh', 'mh']

4.8. *Translation*

Pada tahapan ini data akan diterjemahkan kedalam bahasa inggris agar dapat dipolarisasi label. Namun sebelum itu hasil stemming tadi digabungkan terlebih dahulu menjadi satu kalimat utuh..

Tabel 6. Translation

Sebelum	Sesudah
['kamis', 'kepala', 'jaksa', 'negeri', 'aceh', 'besar', 'wakil', 'kepala', 'seksi', 'intelijen', 'film', 'ramadhan', 'sh', 'mh', 'hadir', 'tinjau', 'pasar', 'jelang', 'ramadhan', 'awas', 'gas', 'elpiji', 'kg', 'kejaksaanri', 'kejatiaceh', 'kejatiaceh', 'kejariacehbesar', 'abes', 'akuntabel']	*Thursday Head of Aceh Besar Attorney General Deputy Head of the Intelligence Section of the film Ramadhan SH MH was present to review the market ahead of the Ramadhan Awas Gas LPG KG Prosecutor's Office of Kejatiaceh Kejariaceh Besar ABES Accountable*

4.9. *Data labeling*

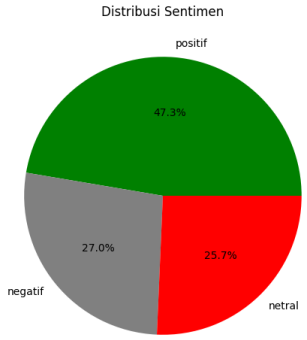
Tahap *Data labelling* dilakukan dengan menggunakan *library python TextBlob* dengan melihat polaritas yang dimiliki oleh teks tweet yang telah diterjemahkan ke dalam bahasa inggris. *TextBlob* saat ini hanya menyediakan layanan *labeling* untuk data berbahasa inggris sehingga dataset harus diterjemahkan ke dalam Bahasa inggris terlebih dahulu.

Berdasarkan hasil analisis sentimen terhadap kebijakan pemerintah terbaru mengenai pendistribusian gas elpiji subsidi, diketahui bahwa mayoritas tanggapan dari masyarakat tergolong positif, dengan jumlah mencapai 518 data atau 47,3% dari keseluruhan tanggapan. Hal ini mencerminkan adanya apresiasi dan penerimaan masyarakat terhadap kebijakan yang diberlakukan.

Tabel 7. Data labeling

Teks Bahasa Inggris	Sentimen
Meaning of LPG GAS Tube Kg Indonesian green color green color LPG gas cylinders are characteristic	Negatif
Bukudjati Prophetfa Emilydiparis now Prabowo's pattern is a superhero of a tax case that seems not to ride LPG gas, the case of added.	Netral
Thursday Head of Aceh Besar Attorney General Deputy Head of the Intelligence Section of the film Ramadhan SH MH was present to review the market ahead of the Ramadhan Awas Gas LPG KG KG Prosecutor's Office of Kejatiaceh Kejariaceh Besar ABES Accountable	Positif

Sementara itu, terdapat 291 tanggapan atau 27,0% yang menunjukkan sentimen negatif, yang menandakan adanya kritik atau ketidakpuasan dari sebagian masyarakat terhadap kebijakan tersebut. Adapun 281 tanggapan lainnya atau 25,7% tergolong netral, hal ini menunjukkan bahwa sebagian masyarakat memberikan respon yang tidak secara eksplisit menunjukkan dukungan maupun penolakan. Temuan ini menunjukkan bahwa secara umum persepsi publik terhadap kebijakan pendistribusian gas elpiji subsidi cenderung berada pada spektrum positif, meskipun tetap terdapat suara-suara tau opini yag bersifat netral maupun kritis.

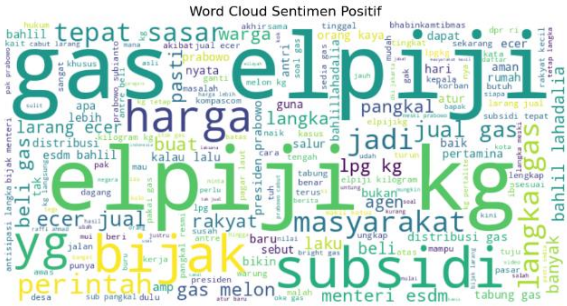


Gambar 2. *Pie chart* hasil labeling

4.10. *Features extraxtion (N-grams)*

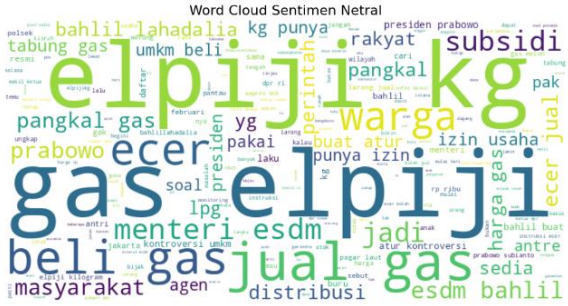
Tahap ekstraksi fitur dalam penelitian ini menggunakan metode N-gram untuk mengubah teks menjadi representasi numerik. Metode ini berfokus pada mengidentifikasi kata-kata penting berdasarkan frekuensi kemunculannya dalam dokumen dan seberapa jarang kata tersebut ditemukan di dokumen lain. Proses ini membantu model pembelajaran mesin mengenali fitur teks yang relevan dengan lebih efektif. Untuk mempermudah pemahaman, hasil dari metode N-gram akan divisualisasikan dalam bentuk word cloud. Kata-kata dengan skor tertinggi akan ditampilkan dengan ukuran lebih besar, sehingga

memudahkan identifikasi kata kunci penting secara visual.



Gambar 3. Word cloud sentiment positif

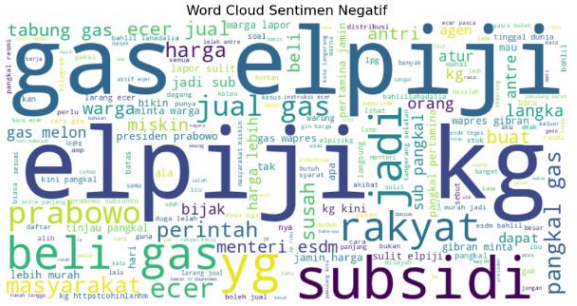
Pada Gambar 3 menunjukkan *word cloud* sentimen positif yang didominasi oleh kata-kata seperti "gas", "elpiji", "subsidi", "kg", dan "bijak". Kemunculan kata-kata tersebut mencerminkan respons positif masyarakat terhadap kebijakan distribusi gas elpiji bersubsidi, khususnya dalam hal keterjangkauan dan ketepatan sasaran. Istilah seperti "harga", "tepat", "masyarakat", dan "ecer" menunjukkan bahwa distribusi dinilai efektif, membantu masyarakat kecil, serta mudah diakses. Secara keseluruhan, *word cloud* ini mengindikasikan bahwa kebijakan tersebut diterima dengan baik dan dianggap memberikan manfaat nyata bagi masyarakat.



Gambar 4. Word cloud sentiment netral

Pada gambar 4 menunjukkan *word cloud* sentimen netral yang didominasi oleh kata-kata seperti "elpiji", "gas", "kg", "ecer", dan "beli". Kata-kata tersebut menunjukkan bahwa perbincangan publik bersifat informatif dan deskriptif tanpa menunjukkan kecenderungan sentimen tertentu. Istilah seperti "subsidi", "izin", "usaha", "distribusi", dan "menteri esdm" mengindikasikan bahwa pembahasan seputar kebijakan dan regulasi masih bersifat netral dan berfokus pada penyampaian fakta atau informasi. Secara keseluruhan, *word cloud* ini mencerminkan persepsi masyarakat yang cenderung menjelaskan kondisi atau situasi tanpa ekspresi penilaian emosional yang kuat.

Pada gambar 5 menampilkan *word cloud* sentimen negatif yang didominasi oleh kata-kata seperti "elpiji", "gas", "subsidi", "kg", dan "ecer". Kemunculan kata-kata seperti "langka", "sulit", "antri", "miskin", dan "lapor" memiliki arti adanya keluhan atau ketidakpuasan publik terhadap ketersediaan dan distribusi gas elpiji subsidi. Adapun istilah seperti "harga", "turun", "beli", dan "permintaan" mencerminkan keresahan terhadap fluktuasi harga serta keterbatasan akses bagi masyarakat. Selain itu, penyebutan tokoh atau institusi seperti "prabowo", "menteri esdm", dan "pertamina" menunjukkan bahwa masyarakat turut menyoroti tanggung jawab pemerintah dalam mengatasi permasalahan ini. Secara keseluruhan, *word cloud* ini mencerminkan persepsi negatif terkait kesulitan akses dan distribusi gas subsidi yang dirasakan masyarakat.



Gambar 5. Word cloud sentiment negatif

4.11. *Evaluasi dan klasifikasi SVM*

Setelah proses fitur ekstraksi, dataset dibagi untuk keperluan pelatihan dan pengujian model. Pembagian dilakukan dengan rasio 80% untuk *Train* dan 20% untuk *Test*, di mana 80% data digunakan untuk melatih model, sementara 20% sisanya digunakan untuk menguji kinerjanya. Rasio ini merupakan pembagian yang sering digunakan karena memberikan model data yang cukup untuk belajar sekaligus menyediakan data yang memadai untuk evaluasi performa.

Tabel 8. Hasil evaluasi model

	Precision	***Recall***	***F1-Score***	***Support***
Negatif	0.70	0.54	0.61	59
Netral	0.64	0.64	0.64	56
Positif	0.70	0.79	0.74	103
Accuracy			0.68	218
Macro avg	0.68	0.66	0.66	218
Weighted avg	0.68	0.68	0.68	218
Accuracy : 68.35 %				

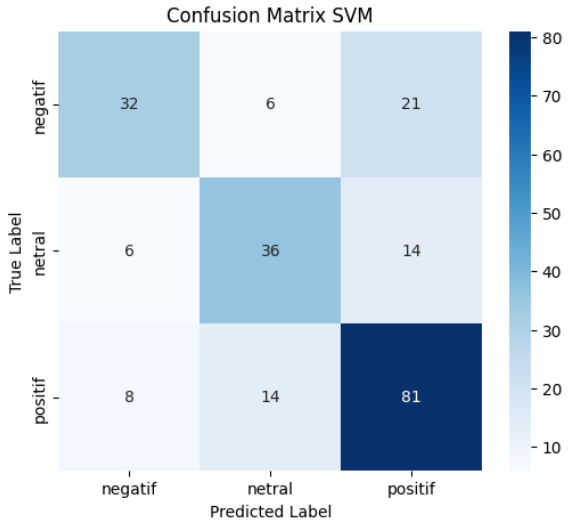
Berdasarkan hasil evaluasi model seperti yang ditampilkan pada Tabel 10, performa model dalam mengklasifikasikan sentimen tergolong cukup baik, dengan akurasi keseluruhan sebesar 68,35%. Model menunjukkan kinerja paling baik pada kelas sentimen positif, dengan *precision* sebesar 0.70, *recall* 0.79, dan f1-*score* sebesar 0.74 dari total 103 data. Hal ini mengindikasikan bahwa model mampu mengenali dan mengklasifikasikan sentimen positif dengan cukup akurat dan konsisten.

Untuk kelas sentimen netral, model mencatat *precision*, *recall*, dan f1-*score* yang seimbang yaitu

masing-masing sebesar 0.64. Hal ini menunjukkan bahwa model cukup stabil dalam mengenali sentimen netral, meskipun masih terdapat ruang untuk peningkatan. Sementara itu, kelas sentimen negatif menunjukkan performa yang paling rendah, khususnya pada *recall* sebesar 0.54, yang berarti model belum mampu mengenali seluruh data negatif secara optimal. Meskipun *precision* pada kelas ini mencapai 0.70, f1-*score* nya tercatat sebesar 0.61 dari total 59 data.

Dari sisi evaluasi rata-rata, nilai *macro average*, *precision*, *recall*, dan f1-*score* masing-masing adalah 0.68, 0.66, dan 0.66, yang menunjukkan performa model secara merata untuk ketiga kelas. Sedangkan untuk nilai *weighted average* yang mempertimbangkan distribusi jumlah data per kelas, masing-masing berada di angka 0.68 untuk *precision*, *recall*, dan f1-*score*.

Hasil ini menunjukkan bahwa secara keseluruhan model memiliki performa yang layak, terutama pada kelas sentimen positif, namun masih menghadapi tantangan dalam meningkatkan akurasi klasifikasi terhadap sentimen negatif yang dapat disebabkan oleh jumlah data yang relatif lebih sedikit dibandingkan kelas lainnya.



Gambar 6. Confusion matrix SVM

Pada gambar 6 menunjukkan *confusion matrix* hasil klasifikasi sentimen publik menggunakan algoritma *Support Vector Machine* (SVM). Model ini dikembangkan untuk mengelompokkan data sentimen dari media sosial X (Twitter) ke dalam tiga kategori yaitu negatif, netral, dan positif. Berdasarkan *confusion matrix*, terlihat bahwa algoritma SVM mampu mengklasifikasikan data sentimen dengan cukup baik, khususnya pada kategori sentimen positif. Tercatat sebanyak 81 data positif berhasil diklasifikasikan dengan benar, meskipun terdapat 14 *instance* positif yang diklasifikasikan sebagai netral, dan 8 lainnya diklasifikasikan sebagai negatif. Hal ini menunjukkan bahwa model cukup andal dalam mengidentifikasi sentimen positif, namun masih mengalami kesulitan membedakan opini yang cenderung positif dengan yang bersifat netral.

Pada kategori sentimen netral, terdapat 36 prediksi yang benar, namun masih terdapat 14 instance yang salah diklasifikasikan sebagai positif dan 6 sebagai negatif. Hal ini mencerminkan bahwa ekspresi netral dalam opini publik cenderung memiliki kemiripan struktur bahasa dengan kategori lain, sehingga menyebabkan ambiguitas dalam prediksi model. Sementara itu, untuk sentimen negatif terdapat sebanyak 32 *instance* yang telah berhasil diklasifikasikan dengan tepat, tetapi terdapat 21 *instance* yang diklasifikasikan sebagai positif, serta 6 *instance* lainnya sebagai netral. Tingginya kesalahan klasifikasi pada kategori ini mengindikasikan bahwa model masih kesulitan mengenali ekspresi negatif terutama yang disampaikan secara halus atau menggunakan bahasa informal khas media sosial.

Secara keseluruhan, hasil ini mengindikasikan bahwa sentimen publik terhadap kebijakan pendistribusian gas elpiji subsidi didominasi oleh sentimen netral. Dominasi ini menunjukkan bahwa sebagian besar masyarakat masih bersikap menunggu dan belum membentuk opini yang kuat, baik mendukung maupun menolak. Sikap netral ini kemungkinan besar disebabkan oleh minimnya informasi yang diterima publik terkait mekanisme distribusi, dampak kebijakan di tingkat akar rumput, serta belum terlihat realisasi langsung di lapangan.

Akan tetapi, muncul juga sentimen positif yang menandakan adanya dukungan terhadap kebijakan tersebut terutama dari pihak-pihak yang menilai kebijakan ini sebagai langkah reformasi distribusi subsidi yang lebih tepat sasaran. Namun di sisi lain, sentimen negatif tetap hadir sebagai refleksi kekhawatiran masyarakat terhadap potensi ketidaktepatan sasaran penerima, ancaman kelangkaan, serta transparansi dalam pelaksanaan kebijakan.

Berdasarkan temuan ini, pemerintah disarankan untuk memperkuat strategi komunikasi publik yang lebih inklusif dan transparan. Informasi terkait kriteria penerima, mekanisme penyaluran, serta bentuk pengawasan harus disampaikan secara terbuka melalui media sosial, disertai infografis edukatif dan ruang diskusi publik yang aktif dan efektif. Hal ini penting guna membangun kepercayaan publik serta meredam kekhawatiran terhadap pelaksanaan kebijakan.

Adapun keterbatasan dalam penelitian ini mencakup potensi bias data dari media sosial X (Twitter) yang tidak sepenuhnya merepresentasikan keseluruhan populasi dan ketergantungan pada pelabelan otomatis menggunakan pustaka *TextBlob* yang masih memiliki keterbatasan dalam menangkap konteks lokal dan bahasa informal. Untuk penelitian selanjutnya disarankan untuk mengeksplorasi platform media sosial lain guna memperoleh perspektif yang lebih luas dan beragam. Selain itu, penggunaan model pembelajaran mendalam berbasis bahasa Indonesia seperti IndoBERT atau IndoLEM dapat ditinjau untuk

meningkatkan akurasi dalam memahami konteks lokal.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, algoritma *Support Vector Machine* (SVM) yang diterapkan dalam analisis sentimen publik terhadap kebijakan terbaru pendistribusian gas elpiji subsidi menunjukkan performa yang cukup baik dengan akurasi sebesar 68,35%. Dari total 1091 data yang dikumpulkan dari media sosial X (Twitter), diketahui bahwa mayoritas sentimen publik cenderung positif (518 data atau 47,3%), diikuti oleh sentimen negatif (291 data atau 27,0%) dan netral (281 data atau 25,7%). Hasil ini menunjukkan bahwa kebijakan tersebut mendapatkan penerimaan yang cukup baik dari masyarakat, meskipun masih terdapat kritik dan keraguan terutama terkait keterbatasan akses, potensi kelangkaan, dan efektivitas distribusi gas elpiji di lapangan. Evaluasi model juga menunjukkan bahwa SVM paling andal dalam mengklasifikasikan sentimen positif namun masih menghadapi tantangan dalam mengidentifikasi sentimen negatif secara akurat akibat jumlah data yang relatif lebih sedikit. Untuk itu, disarankan agar pemerintah memperkuat strategi komunikasi publik secara terbuka dan edukatif melalui media sosial termasuk transparansi mekanisme distribusi, kriteria penerima subsidi serta pengawasan. Selain itu, untuk penelitian selanjutnya disarankan untuk mengeksplorasi platform media sosial lain seperti Instagram, TikTok, atau YouTube untuk memperluas cakupan data serta mempertimbangkan penggunaan model pembelajaran mendalam berbasis bahasa Indonesia seperti IndoBERT atau IndoLEM guna meningkatkan akurasi klasifikasi dan pemahaman konteks lokal. Perbaikan pelabelan data melalui pendekatan manual atau crowdsourcing juga dapat menjadi opsi dalam meningkatkan kualitas dataset dan keakuratan analisis secara keseluruhan.

DAFTAR PUSTAKA

[1] berkas.dpr.go.id (2025, 4 Februari). MEWUJUDKAN PENYALURAN LPG BERSUBSIDI 3 KILOGRAM YANG TEPAT SASARAN. Diakses pada 10 Februari 2025, dari https://berkas.dpr.go.id/pusaka/files/info_singkat/Info%20Singkat-XVII-3-I-P3DI-Februari-2025-245.pdf

[2] UGM.ac.id. (2025, 3 Februari). Kebijakan Pelarangan Menjual LPG 3Kg di Tingkat Pengecer Dinilai Menyusahkan Rakyat Kecil. Diakses pada 10 Februari 2025, dari https://ugm.ac.id/id/berita/kebijakan-pelarangan-menjual-lpg-3kg-di-tingkat-pengecer-dinilai-menyusahkan-rakyat-kecil/.

[3] Maulana, I. S. (2024). PENEGAKAN HUKUM UNDANG–UNDANG NOMOR 22 TAHUN 2001 TENTANG MINYAK DAN GAS BUMI TERHADAP PENJUALAN BAHAN BAKAR MINYAK OLEH PERTAMINI DI KOTA SAMARINDA. LEGALITAS: Jurnal Ilmiah Ilmu Hukum, 7(2), 51-62.

[4] Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma Svm.

[5] Fajria, A. M., Faqih, A., & Dwilestari, G. (2025). The Impact of Principal Component Analysis Dimensionality Reduction on Sentiment Classification Performance Using Support Vector Machine. Journal of Artificial Intelligence and Engineering Applications (JAIEA), 4(2), 764-770.

[6] Putri, N. S., Saragih, H., & Heikhmakhtiar, A. K. (2024). Comparison of Naïve Bayes Classifier and Support Vector Machine for sentiment analysis on civil military relations conflict among Rohingya refugees as recommendation for defense policy making. Journal of Intelligent Decision Support System (IDSS), 7(3), 227-235.

[7] Agarkar, A. A., Betgeri, S., Nimbalkar, A., Nikam, V., Rane, C., & Motade, P. (2025, February). Sentiment and Tone Analysis of Reddit Posts Using Python. In 2025 4th International Conference on Sentiment Analysis and Deep Learning (ICSADL) (pp. 100-104).

[8] Isnain, A. R., Sakti, A. I., Alita, D., & Marga, N. S. (2021). Sentimen Analisis Publik Terhadap Kebijakan Lockdown Pemerintah Jakarta Menggunakan Algoritma SVM. JDMSI, 2(1), 31–37.

[9] Rahmaddeni, R., Anam, M. K., Irawan, Y., Susanti, S., & Jamaris, M. (2022). Comparison of support vector machine and XGBSVM in analyzing public opinion on Covid-19 vaccination. ILKOM Jurnal Ilmiah, 14(1), 32-38.

[10] Yunita, R., & Kamayani, M. (2023). Perbandingan algoritma SVM dan Naïve Bayes pada analisis sentimen penghapusan kewajiban skripsi. The Indonesian Journal of Computer Science, 12(5).

[11] Ramadhan, S., Siswanto, T., & Sari, S. (2025). Sentiment Analysis And Topic Modelling Of Candidate News In The 2024 General Election On Twitter Social Media Using Latent Dirichlet Allocation (LDA) Method. Intelmatics, 5(1), 33-41.

[12] Abidin, A. Z., Furqon, M. A., & Bukhori, S. (2025). Combining Rule Based, Lexicon Based and Support Vector Machine for Improve Accuracy in Sentiment Analysis of ChatGPT Usage. Journal of Research in Artificial Intelligence for Systems and Applications, 1(1), 20-28.

[13] Alghifari, F., & Juardi, D. (2021). Penerapan Data Mining Pada Penjualan Makanan dan Minuman Menggunakan Metode Algoritma Naïve Bayes: Studi Kasus: Makan Barbeque Sepuasnya. Jurnal Ilmiah Informatika, 9(02), 75-81.

[14] Juanita, S., Daeli, A. H., Syafrullah, M.,

Anggraeni, W., & Purnomo, M. H. (2025). A machine learning approach for text pattern diagnosis in mental health consultations. Decision Analytics Journal, 100572.

[15] Azmi, A. F., & Voutama, A. (2024). Prediksi Churn Nasabah Bank Menggunakan Klasifikasi Random Forest Dan Decision Tree Dengan Evaluasi Confusion Matrix. Komputa: Jurnal Ilmiah Komputer Dan Informatika, 13(1), 111-119.

[16] Ikram, H., Hasibuan, M. A., Rianda, H., Fahriza, A., Suryadi, S., & Suhendra, R. (2024). IMPLEMENTASI METODE KLASIFIKASI NAÏVE BAYES PADA ANALISIS SENTIMEN ULASAN APLIKASI MOBILE LEGENDS DI GOOGLE PLAY STORE. Jurnal Jambo Digitech, 1(1).

[17] Tarigan, V. T., & Yusupa, A. (2024). Perbandingan Algoritma Maching Learning dalam Analisis Sentimen Mobil Listrik di Indonesia pada Media Sosial Twitter/X. Jurnal Informatika Polinema, 10(4), 479-490.

[18] Albab, M. U., & Fawaiq, M. N. (2023). Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic. Jurnal Transformatika, 20(2), 1-12.

[19] Alam, M. S., & Suryani, A. A. (2021). Minang and Indonesian Pharase-Based Statistical Machine Translation. Journal of Informatics and Telecommunication Engineering, 5(1), 216-224.

[20] Subagyo, A. K., Syahputra, H. S. H., Khoiri, Z. F., & Sukmadiningtyas, S. (2024). ANALISIS SENTIMEN MASYARAKAT TERHADAP PELANTIKAN KABINET MERAH PUTIH PADA MEDIA SOSIAL. In Proceedings of the National Conference on Electrical Engineering, Informatics, Industrial Technology, and Creative Media (Vol. 4, No. 1, pp. 1080-1089).

[21] Kusuma, A. T. A. (2024). Perbandingan Metode Peter Norvig, Lstm, dan N-Gram untuk Spell Correction Bahasa Indonesia (Doctoral dissertation, Universitas Islam Indonesia).

[22] Nyakundi, G. N., Ndiritu, J., Ivivi, J. M., & Kamanu, T. (2024). Class Prediction of High-Dimensional Data with Class Imbalance: Breast Cancer Gene Expression Data.