PENERAPAN SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN TWITTER TERHADAP EFISIENSI ANGGARAN

Rizky Ananda Hafika, Stefen Agus Waruwu ,Yuda Advis Ambrosius Sitohang, Muhammad Yazid Noor, Muhammad Haikal Al Majid, Arnita, Fanny Ramadhani
Ilmu Komputer, Universitas Negeri Medan
Jalan W. Iskandar Psr V Medan Esatate Kab. Deli Serdang , Indonesia
rizkyanandahafika@gmail.com

ABSTRAK

Media sosial, khususnya Twitter, menjadi wadah utama bagi masyarakat untuk mengekspresikan opini terhadap kebijakan pemerintah, termasuk efisiensi anggaran. Analisis sentimen berbasis machine learning diperlukan untuk memahami respons publik secara otomatis. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat di Twitter terhadap kebijakan efisiensi anggaran menggunakan metode Support Vector Machine (SVM). Data yang digunakan telah melalui tahap preprocessing, labeling manual, dan untuk memastikan kualitas analisis. Sentimen yang diklasifikasikan terdiri dari tiga kategori: negatif, netral, dan positif. Hasil penelitian menunjukkan bahwa mayoritas sentimen masyarakat terhadap kebijakan ini bersifat negatif, menandakan ketidakpuasan atau ketidaksetujuan. Dalam perbandingan performa model SVM, Kernel Linear menunjukkan akurasi terbaik sebesar 81.46%, sedikit lebih tinggi dibandingkan Kernel RBF yang memperoleh 81.17%. Kernel Linear terbukti lebih stabil dalam mengklasifikasikan sentimen negatif dan netral, sementara Kernel RBF lebih unggul dalam menangani pola data yang kompleks tetapi kurang stabil dalam mengklasifikasikan sentimen positif. Oleh karena itu, model SVM dengan Kernel Linear dipilih sebagai model terbaik dalam penelitian ini. Penelitian ini memberikan wawasan mengenai persepsi publik terhadap kebijakan pemerintah serta menunjukkan efektivitas SVM dalam analisis sentimen media sosial.

Kata Kunci: *Analisis Sentimen, Efisiensi Anggaran, Media Sosial, Support Vector Machine (SVM), Twitter.*

1. PENDAHULUAN

Dalam era digital, media sosial menjadi salah satu wadah utama bagi masyarakat untuk mengekspresikan opini terhadap berbagai isu, termasuk kebijakan pemerintah. Twitter, sebagai salah satu platform media sosial terbesar, menyajikan data yang kaya akan sentimen publik terkait kebijakan yang diterapkan oleh pemerintah. Namun, tingginya volume data yang dihasilkan setiap hari menyebabkan sulitnya mengekstrak opini secara manual, sehingga diperlukan metode yang dapat mengolah dan menganalisis sentimen masyarakat secara sistematis.

Salah satu kebijakan yang kerap menimbulkan perdebatan adalah kebijakan efisiensi anggaran. Kebijakan ini bertujuan untuk mengoptimalkan penggunaan anggaran negara, tetapi di sisi lain juga dapat memicu kekhawatiran terkait potensi penurunan kualitas layanan publik. Perbedaan pandangan ini terlihat jelas di media sosial, di mana masyarakat bebas mengungkapkan opini mereka, baik yang bersifat mendukung maupun menentang. Sayangnya, tanpa adanya analisis sistematis, opini-opini ini sulit untuk disimpulkan secara objektif. Oleh karena itu, diperlukan pendekatan berbasis machine learning untuk mengolah data opini masyarakat dengan lebih akurat dan efisien.

Salah satu metode yang dapat digunakan adalah analisis sentimen berbasis machine learning, yang memungkinkan klasifikasi opini publik secara otomatis. Support Vector Machine (SVM) merupakan salah satu algoritma yang sering digunakan dalam analisis sentimen karena kemampuannya menangani data teks berdimensi tinggi serta memberikan akurasi tinggi dalam klasifikasi sentimen. Beberapa penelitian sebelumnya telah membuktikan efektivitas SVM dalam menganalisis sentimen pengguna di berbagai platform digital.

Misalnya, Penelitian yang dilakukan oleh [1] berjudul 'Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine' bertujuan untuk menganalisis review pengguna terhadap aplikasi LinkedIn menggunakan metode SVM. Hasil penelitian menunjukkan bahwa model SVM mencapai akurasi 85,54%, dengan kata 'ovo' sebagai ulasan positif yang paling sering muncul, sedangkan kata 'driver' paling sering muncul dalam ulasan negatif. Penelitian lain yang dilakukan oleh [2] membandingkan performa metode SVM dan KNN dalam melakukan klasifikasi ulasan aplikasi Gojek di Google Play Store. Hasil penelitian menunjukkan bahwa metode KNN dengan nilai K=22 memperoleh akurasi 82,14%, presisi 82,28%, dan recall 95,43%. Sementara itu, metode SVM dengan kernel linear dan parameter C=1 memiliki performa lebih baik, dengan akurasi 87,98%, presisi 88,55%, dan recall 95,43%. Dengan demikian, Penelitian sebelumnya menunjukkan bahwa SVM unggul dalam analisis sentimen, terutama dibandingkan dengan metode lain seperti KNN. Oleh karena itu, penelitian ini akan menguji efektivitas SVM dalam menganalisis sentimen terhadap kebijakan efisiensi anggaran di Twitter.

Berdasarkan penelitian sebelumnya yang menunjukkan keunggulan SVM, penelitian ini

bertujuan untuk menganalisis sentimen masyarakat di Twitter terhadap kebijakan efisiensi anggaran setelah melalui proses preprocessing, labeling manual dan ekstrasi fitur. Selain itu, penelitian ini akan membandingkan performa model SVM dengan kernel linear dan RBF untuk menentukan model yang lebih optimal serta mengetahui tingkat akurasi terbaik.

Diharapkan hasil penelitian ini dapat memberikan wawasan lebih dalam mengenai persepsi masyarakat terhadap kebijakan efisiensi anggaran serta menjadi referensi bagi pemerintah dalam memahami respons masyarakat. Selain itu, penelitian ini juga berkontribusi dalam pengembangan model klasifikasi sentimen yang lebih akurat dalam domain kebijakan publik.

2. TINJAUAN PUSTAKA

2.1. Penelitian terdahulu

Penelitian mengenai analisis sentimen menggunakan Support Vector Machine (SVM) telah banyak dilakukan dalam berbagai konteks, termasuk kebijakan publik dan media sosial. Berikut adalah beberapa penelitian terdahulu yang mendukung penelitian ini:

Penelitian oleh [3] berjudul *"Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi"* menganalisis opini masyarakat terhadap akun Twitter resmi @dkijakarta selama pandemi COVID-19. Penelitian ini menggunakan TF-IDF Vectorizer untuk pembobotan kata dan membandingkan beberapa algoritma klasifikasi, yaitu Random Forest (75,81%), Naïve Bayes (75,22%), dan Support Vector Machine (77,58%). Hasil analisis menunjukkan mayoritas tweet bersentimen netral (83,6%), dengan negatif (8,8%) dan positif (7,6%). Studi ini menunjukkan bahwa SVM memiliki akurasi tertinggi, sehingga lebih optimal dalam klasifikasi sentimen.

Penelitian oleh [4] dengan judul *"Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine"* menunjukkan bahwa algoritma SVM dapat digunakan untuk mengklasifikasikan ulasan pengguna Google Meet ke dalam sentimen positif, negatif, dan netral. Dari 10.000 data ulasan, sentimen positif mendominasi dengan 4.468 ulasan yang banyak mengandung kata "bagus", diikuti oleh 4.056 ulasan netral dan 1.476 ulasan negatif dengan kata dominan "aplikasi". Dengan pembagian data training 90% dan data testing 10%, model SVM yang digunakan mencapai akurasi 94%, precision 98%, recall 95%, dan f1-score 96%. Temuan ini menunjukkan bahwa SVM mampu mengklasifikasikan sentimen ulasan dengan tingkat akurasi tinggi, yang relevan dalam analisis kepuasan pengguna terhadap aplikasi berbasis ulasan teks.

Selanjutnya Penelitian oleh [5] berjudul "Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM" menganalisis opini masyarakat terhadap bakal calon presiden 2024 di Twitter. Dengan pengguna media sosial yang mencapai 170 juta pada 2021, Twitter menjadi platform utama untuk menyampaikan opini politik. Penelitian ini menggunakan Support Vector Machine (SVM) untuk menganalisis sentimen tweet terkait tiga bakal calon presiden: aniesbaswedan, ganjarpranowo, dan prabowosubianto. Data yang digunakan berjumlah 1719 tweet dengan distribusi 597 untuk aniesbaswedan, 627 untuk ganjarpranowo, dan 495 untuk prabowosubianto.

Hasil analisis menunjukkan metode SVM mencapai akurasi 78,3%. Akurasi ini dipengaruhi oleh jumlah dan komposisi data positif serta negatif. Studi ini memberikan wawasan tentang sentimen publik terkait bakal calon presiden 2024 dan dapat menjadi referensi dalam memahami preferensi politik menjelang pemilu. Penelitian oleh [6] berjudul "Support Vector Machine Berbasis Particle Swarm Optimization Pada Analisis Sentimen Anggota KPPS" membahas opini masyarakat terhadap anggota KPPS dalam Pemilu 2024 yang menjadi trending di media sosial X. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) berbasis Particle Swarm Optimization (PSO) untuk analisis sentimen. Data yang diambil dari media sosial X berjumlah 702 opini, dan setelah proses preprocessing, tersisa 688 data bersih. Kata yang paling sering muncul dalam opini adalah "meninggal". Hasil analisis menunjukkan bahwa metode SVM berbasis PSO mencapai akurasi 70%. Dari analisis sentimen, 99% opini masyarakat bersifat positif, sementara 1% bersifat negatif.

2.2. Analisis Sentimen

Analisis sentimen merupakan metode untuk mengukur opini seseorang mengenai tingkat persetujuan mereka terhadap suatu hal, produk, layanan, atau peristiwa tertentu [7].

Metode ini bertujuan untuk memahami dan mengekstrak data sentimen yang umumnya dikategorikan berdasarkan polaritasnya, yaitu positif atau negatif. Proses ini dilakukan dengan memanfaatkan teknik pemrosesan bahasa alami (NLP) [8].

Sentimen umumnya dikategorikan menjadi positif, negatif, dan netral. Dalam penelitian ini, analisis sentimen diterapkan untuk mengetahui bagaimana reaksi masyarakat terhadap kebijakan Efisiensi Anggaran.

2.3. Twitter Sebagai Sumber Data

Twitter, dengan 24 juta pengguna aktif, merupakan salah satu platform media sosial dengan jumlah pengguna terbanyak di Indonesia. Informasi yang dibagikan di Twitter mencerminkan tanggapan pengguna terhadap berbagai objek [7].

Data dari Twitter bersifat real-time, menjadikannya sumber data yang relevan dalam menganalisis tren opini publik terhadap kebijakan pemerintah.

2.4. Crawling

Proses awal dalam pengumpulan data disebut crawling, yang dilakukan dengan memanfaatkan fasilitas Twitterscraper [9].

Crawling data adalah proses otomatis untuk mengumpulkan data dari sumber tertentu, dalam hal ini Twitter. Crawling dilakukan dengan bantuan Pustaka Tweepy. Proses ini memungkinkan penelitian mendapatkan opini masyarakat secara langsung dan real-time terkait kebijakan yang sedang dikaji.

2.5. Preprocesing

Preprocessing bertujuan untuk menormalkan istilah yang terdapat dalam suatu kalimat. Proses ini dilakukan agar diperoleh data latih yang berkualitas serta fitur yang sesuai dengan kebutuhan. Dengan demikian, preprocessing dapat menyederhanakan pemrosesan data secara lebih efektif [10].

Proses ini bertujuan untuk menghapus kata-kata yang tidak relevan, sehingga menghasilkan teks yang lebih sederhana dan fokus pada kata-kata yang mengandung sentiment [9].

2.6. Cleansing

Cleaning adalah proses menghapus semua karakter non-alfabet dari tweet. Langkah ini bertujuan untuk menghilangkan simbol atau karakter yang tidak diinginkan dan tidak relevan dalam analisis sentimen, sehingga data menjadi lebih bersih dan terstruktur [11].

Proses ini mencakup penghapusan karakter khusus, tanda baca, URL, mention (@username), dan angka yang tidak berpengaruh pada analisis sentimen. Hashtag (#kata) dihapus tandanya, tetapi kata tetap dipertahankan. Selain itu, spasi ganda serta spasi di awal dan akhir teks juga diperbaiki. Dengan cleaning, data menjadi lebih bersih dan siap untuk tahap preprocessing berikutnya.

2.7. Case Folding

Pada tahap case folding, teks dalam kumpulan kalimat akan diubah ke dalam bentuk standar, yaitu semua huruf kecil (lowercase). Langkah ini bertujuan untuk menyamakan format teks sehingga memudahkan proses analisis data [12].

Proses ini dilakukan agar kata dengan bentuk huruf berbeda, seperti "Efisiensi" dan "efisiensi," dikenali sebagai entitas yang sama. Dengan demikian, case folding membantu mengurangi variasi kata yang tidak perlu dan meningkatkan akurasi dalam pemrosesan data teks.

2.8. Tokenisasi

Proses tokenizing membagi kalimat dalam tweet menjadi bagian-bagian yang lebih kecil, yang disebut token. Token ini berupa kata-kata yang berdiri sendiri, dipisahkan berdasarkan spasi atau tanda baca, sehingga memudahkan analisis lebih lanjut [12].

2.9. Normalisasi

Proses ini bertujuan untuk memastikan bahwa kata-kata yang diperpanjang atau disingkat dikonversi menjadi bentuk yang sesuai dengan definisi dalam Kamus Besar Bahasa Indonesia (KBBI). Mengubah kata slang menjadi kata baku dikenal sebagai "konversi slang", yang membantu meningkatkan keakuratan dalam analisis teks [13].

2.10. Stopword removal

Proses ini dilakukan untuk menghapus kata-kata yang tidak memiliki makna signifikan dalam analisis data tweet. Kata-kata tersebut disebut stopwords, seperti "dari", "ke", "yang", "di", "dan", serta kata umum lainnya yang tidak berkontribusi pada pemahaman sentimen [14].

Kata-kata ini umumnya sering muncul tetapi tidak memberikan informasi penting dalam pemrosesan teks. Dengan menghapus stopword, data yang diolah menjadi lebih bersih dan fokus pada kata-kata yang lebih bermakna untuk analisis sentimen atau klasifikasi teks.

2.11. Stemmer

Pada tahap stemming, kata-kata dalam dokumen dikonversi ke bentuk dasarnya dengan menerapkan aturan tertentu. Proses ini bertujuan untuk menyederhanakan kata dengan menghilangkan imbuhan, sehingga analisis teks menjadi lebih efektif [15].

Dalam penelitian ini, proses stemming dilakukan menggunakan Sastrawi, sebuah pustaka NLP (Natural Language Processing) yang khusus untuk bahasa Indonesia. Sastrawi bekerja dengan menghapus imbuhan (awalan, sisipan, akhiran) dari sebuah kata sehingga diperoleh bentuk dasar atau kata dasar. Misalnya, kata "mengerjakan" akan diubah menjadi "kerja", dan "diberikan" menjadi "beri".

2.12. Labeling

Labeling adalah proses pemberian kategori atau kelas pada data untuk digunakan dalam model pembelajaran mesin. Dalam analisis sentimen, setiap tweet diklasifikasikan ke dalam label tertentu, seperti positif, negatif, atau netral. Pada penelitian ini, labeling dilakukan dalam dua tahap:

a. Labeling otomatis, menggunakan pemrograman untuk mengklasifikasikan sentimen berdasarkan kata-kata tertentu atau model yang telah dilatih sebelumnya.
b. Labeling manual, dilakukan dengan pengecekan ulang untuk memastikan akurasi dan memperbaiki kesalahan klasifikasi yang terjadi pada tahap otomatis.

Labeling ini sangat penting karena menentukan kualitas data latih yang akan digunakan dalam membangun model klasifikasi sentimen menggunakan Support Vector Machine (SVM).

2.13. TF-IDF

Dalam penelitian ini, metode TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk menentukan bobot setiap kata dalam tweet yang telah melewati tahap preprocessing. TF-IDF merupakan salah satu teknik pembobotan kata dalam proses ekstraksi fitur, yang menghitung frekuensi kemunculan kata dalam sebuah dokumen relatif terhadap keseluruhan kumpulan dokumen, sehingga kata yang lebih relevan memiliki bobot lebih tinggi dalam analisis data [13].

Pada penelitian ini, TF-IDF digunakan untuk mengekstraksi fitur dari tweet yang telah dikumpulkan, sehingga setiap tweet direpresentasikan dalam bentuk vektor numerik berdasarkan bobot kata-kata yang ada di dalamnya. Dengan pendekatan ini, model Support Vector Machine (SVM) dapat mengolah data teks secara lebih optimal dalam mengklasifikasikan sentimen masyarakat terhadap kebijakan efisiensi anggaran.

2.14. Support Vector Machine

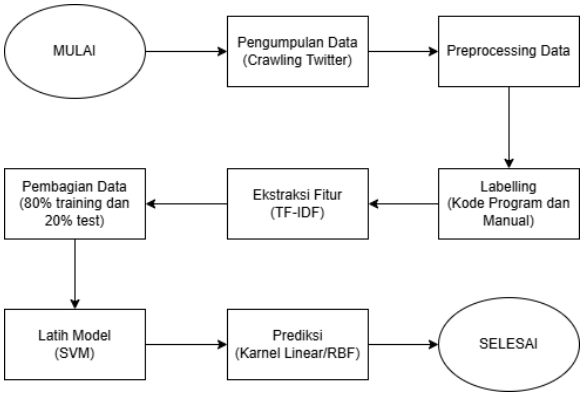
SVM (Support Vector Machine) adalah salah satu metode dalam supervised learning yang digunakan untuk proses klasifikasi. Dalam model klasifikasi, SVM memiliki dasar matematis yang kuat dan konsep yang lebih jelas. Hyperplane merupakan fungsi yang digunakan untuk memisahkan kelas dalam data, di mana SVM berupaya memaksimalkan jarak antar kelas guna menemukan hyperplane optimal untuk meningkatkan akurasi klasifikasi [16].

Algoritma SVM dikenal sebagai salah satu metode yang efektif dalam menghasilkan solusi klasifikasi yang optimal [17].

Dalam penelitian ini, SVM digunakan untuk analisis sentimen dengan dua jenis kernel, yaitu linear dan Radial Basis Function (RBF). Kernel linear digunakan karena cocok untuk data yang dapat dipisahkan secara linier, sedangkan kernel RBF digunakan untuk menangani data dengan pola yang lebih kompleks dan tidak terpisahkan secara linier. Penggunaan kedua kernel ini memungkinkan perbandingan performa untuk menentukan model yang paling optimal dalam mengklasifikasikan sentimen terhadap kebijakan efisiensi anggaran. Algoritma SVM dipilih karena kemampuannya dalam menangani data berdimensi tinggi serta keandalannya dalam klasifikasi teks, termasuk analisis sentimen di Twitter.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif menganalisis sentimen masyarakat terhadap kebijakan Efisiensi Anggaran di Twitter menggunakan algoritma Support Vector Machine (SVM). Beikut alur analisisnya.



Gambar 1. Alur penelitian

3.1. Sumber data

Data yang digunakan dalam penelitian ini diperoleh melalui teknik web crawling menggunakan Tweet-Harvest. Proses ini dilakukan untuk mengumpulkan tweet yang mengandung kata kunci “efisiensi anggaran”. Data yang berhasil dikumpulkan terdiri dari 3.503 tweet yang diperoleh pada tahun 2025. Data mentah ini berisi 15 atribut, yaitu: conversation_id_str, created_at, favorite_count, full_text, id_str, image_url, in_reply_to_screen_name, lang, location, quote_count, reply_count, retweet_count, tweet_url, user_id_str, dan username.

Namun, dalam penelitian ini, hanya kolom "full_text" yang digunakan sebagai data utama untuk analisis sentimen. Kolom ini berisi teks lengkap dari setiap tweet yang akan diproses lebih lanjut menggunakan metode preprocessing. Sebelum dilakukan preprocessing, atribut lain yang tidak relevan dengan analisis sentimen akan dihapus untuk mengurangi kompleksitas data dan meningkatkan efisiensi pemrosesan.

3.2. Preprocesing

Data yang telah dikumpulkan memerlukan preprocessing sebelum dapat digunakan untuk analisis sentimen. Tahapan preprocessing dalam penelitian ini meliputi:

a. Cleaning: Menghapus karakter khusus, URL, tanda baca, angka, dan emoji agar data lebih bersih.
b. Case Folding: Mengubah seluruh teks menjadi huruf kecil agar konsistensi data tetap terjaga.
c. Tokenisasi: Memecah teks menjadi kata-kata individu agar dapat dianalisis lebih lanjut.
d. Normalisasi:
e. Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna penting dalam analisis sentimen, menggunakan daftar stopword dari pustaka Sastrawi.
f. Stemming: Mengubah kata-kata ke bentuk dasarnya agar memiliki makna yang seragam dalam analisis teks.

3.3. Labeling Data

Setelah preprocessing, data diberikan label secara manual untuk menentukan apakah sentimen dalam tweet bersifat positif, negatif, atau netral. Labeling dilakukan dengan membaca isi tweet dan mengklasifikasikannya sesuai dengan konteks opini yang disampaikan.

3.4. Pembobotan Fitur

Dalam penelitian ini, digunakan metode TF-IDF (Term Frequency-Inverse Document Frequency) untuk memberikan bobot pada setiap kata dalam tweet. TF-IDF membantu mengidentifikasi kata-kata yang paling berpengaruh dalam menentukan sentimen dengan mengukur seberapa sering suatu kata muncul dalam dokumen dan membandingkannya dengan keseluruhan data.

3.5. Pembagian Data

Dataset yang telah diberi label kemudian dibagi menjadi dua bagian:

a. 80% data training, digunakan untuk melatih model SVM.
b. 20% data testing, digunakan untuk menguji performa model.

Pembagian ini dilakukan secara stratified split untuk memastikan distribusi sentimen dalam data training dan testing tetap seimbang. Pemilihan rasio 80/20 didasarkan pada keseimbangan antara jumlah data latih yang cukup untuk mengenali pola dengan baik dan jumlah data uji yang representatif untuk evaluasi performa model. Rasio ini juga membantu menghindari overfitting maupun underfitting.

3.6. Penerapan Model Support Vector Machine (SVM)

Model Support Vector Machine (SVM) digunakan untuk mengklasifikasikan sentimen tweet. Dalam penelitian ini, dua jenis kernel diterapkan, yaitu:

a. Kernel Linear: Digunakan untuk memisahkan data yang dapat dipisahkan secara linier.
b. Kernel RBF (Radial Basis Function): Digunakan untuk menangani data dengan pola non-linear.

Model dilatih menggunakan data training dan diuji dengan data testing untuk melihat performanya.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Preprocessing Data

Pada tahap ini, data yang diperoleh melalui proses crawling dari Twitter mengalami beberapa tahap preprocessing agar lebih bersih dan siap untuk dianalisis. Langkah-langkah yang dilakukan meliputi : cleansing, case folding, tokenisasi, normalisasi, stopword removal, stemming.

```
                                                full_text
0  ini kl ga ndang dibenerin defisit makin meleba...
1  Nih kelakuan institusi tukang gebukmu yg angga...
2                        @linuzgacor dampak efisiensi
3  eh masa kampusku kena efisiensi tiap jumat jd ...
4  @RetroKamen Efisiensi tai kucing lah. Sama ser...

                                                clean_text
0  ini kl ga ndang dibenerin defisit makin meleba...
1  Nih kelakuan institusi tukang gebukmu yg angga...
2                                  dampak efisiensi
3  eh masa kampusku kena efisiensi tiap jumat jd ...
4  Efisiensi tai kucing lah Sama serdos dosen jg ...
```

Gambar 2. Preprocesing cleansing

```
                                                clean_text
0  ini kl ga ndang dibenerin defisit makin meleba...
1  Nih kelakuan institusi tukang gebukmu yg angga...
2                                  dampak efisiensi
3  eh masa kampusku kena efisiensi tiap jumat jd ...
4  Efisiensi tai kucing lah Sama serdos dosen jg ...

                                               case_folding
0  ini kl ga ndang dibenerin defisit makin meleba...
1  nih kelakuan institusi tukang gebukmu yg angga...
2                                  dampak efisiensi
3  eh masa kampusku kena efisiensi tiap jumat jd ...
4  efisiensi tai kucing lah sama serdos dosen jg ...
```

Gambar 3. Preprocesing cleansing – case folding

```
                                               case_folding
0  ini kl ga ndang dibenerin defisit makin meleba...
1  nih kelakuan institusi tukang gebukmu yg angga...
2                                  dampak efisiensi
3  eh masa kampusku kena efisiensi tiap jumat jd ...
4  efisiensi tai kucing lah sama serdos dosen jg ...

                                                tokenized
0  [ini, kl, ga, ndang, dibenerin, defisit, makin...
1  [nih, kelakuan, institusi, tukang, gebukmu, yg...
2                               [dampak, efisiensi]
3  [eh, masa, kampusku, kena, efisiensi, tiap, ju...
4  [efisiensi, tai, kucing, lah, sama, serdos, do...
```

Gambar 4. Preprocesing case folding - tokenisasi

```
                                                tokenized
0  [ini, kl, ga, ndang, dibenerin, defisit, makin...
1  [nih, kelakuan, institusi, tukang, gebukmu, yg...
2                               [dampak, efisiensi]
3  [eh, masa, kampusku, kena, efisiensi, tiap, ju...
4  [efisiensi, tai, kucing, lah, sama, serdos, do...

                                               normalized
0  [ini, kalau, tidak, ndang, dibenerin, defisit,...
1  [nih, kelakuan, institusi, tukang, gebukmu, ya...
2                               [dampak, efisiensi]
3  [eh, masa, kampusku, kena, efisiensi, tiap, ju...
4  [efisiensi, tai, kucing, lah, sama, serdos, do...
```

Gambar 5. Preprocesing tokenisasi - normalisasi

```
                                                normalized
0  [ini, kalau, tidak, ndang, dibenerin, defisit,...
1  [nih, kelakuan, institusi, tukang, gebukmu, ya...
2                                  [dampak, efisiensi]
3  [eh, masa, kampusku, kena, efisiensi, tiap, ju...
4  [efisiensi, tai, kucing, lah, sama, serdos, do...

                                              stopword_removed
0  [ndang, dibenerin, defisit, melebar, efeknya, ...
1  [nih, kelakuan, institusi, tukang, gebukmu, an...
2                                  [dampak, efisiensi]
3  [eh, kampusku, kena, efisiensi, jumat, jd, dar...
4  [efisiensi, tai, kucing, serdos, dosen, jg, la...
```

Gambar 6. Preprocesing normalisasi – stopword removal

```
                                              stopword_removed
0  [ndang, dibenerin, defisit, melebar, efeknya, ...
1  [nih, kelakuan, institusi, tukang, gebukmu, an...
2                                  [dampak, efisiensi]
3  [eh, kampusku, kena, efisiensi, jumat, jd, dar...
4  [efisiensi, tai, kucing, serdos, dosen, jg, la...

                                                  stemmed
0  [ndang, dibenerin, defisit, lebar, efek, namba...
1  [nih, laku, institusi, tukang, gebuk, anggar, ...
2                                  [dampak, efisiensi]
3  [eh, kampus, kena, efisiensi, jumat, jd, darin...
4  [efisiensi, tai, kucing, serdos, dosen, jg, la...
```

Gambar 7. Preprocesing stopword removal - stemming

Hasil preprocessing ini menunjukkan transformasi data mentah menjadi lebih terstruktur dan siap untuk analisis sentimen.

4.2. Distribusi Sentimen

Setelah melalui tahap preprocessing, data tweet diklasifikasikan ke dalam tiga kategori sentimen: positif, netral, dan negatif. Klasifikasi dilakukan dengan dua metode, yaitu labeling otomatis dan labeling manual, yang hasilnya dibandingkan untuk mengidentifikasi perbedaan dalam distribusi sentimen.

4.3. Labeling Otomatis

Proses labeling otomatis dilakukan dengan pendekatan berbasis kata kunci (lexicon-based), di mana daftar kata positif dan negatif telah ditentukan sebelumnya untuk mengkategorikan sentimen dalam tweet. Setiap tweet yang mengandung kata dari daftar positive words akan diberi label positif (1), sementara yang mengandung kata dari daftar negative words diberi label negatif (-1). Jika suatu tweet tidak mengandung kedua jenis kata tersebut, maka dikategorikan sebagai netral (0).

Proses ini diawali dengan membaca dataset yang telah melalui tahap preprocessing, termasuk stemming untuk menyederhanakan kata menjadi bentuk dasarnya. Selanjutnya, setiap tweet dianalisis untuk mendeteksi keberadaan kata kunci yang sesuai dengan daftar kata positif dan negatif.

a. Jika sebuah tweet mengandung setidaknya satu kata dari daftar positif, maka langsung diberi label positif.
b. Jika mengandung setidaknya satu kata dari daftar negatif, maka diberi label negatif.
c. Jika tidak ditemukan kata kunci dari kedua daftar, tweet dikategorikan sebagai netral.

Setelah proses labeling selesai, hasilnya disimpan dalam file baru yang akan digunakan untuk tahap analisis lebih lanjut.

```
Jumlah Tweet Berdasarkan Sentimen:
Positif (1)   : 205
Netral (0)    : 2002
Negatif (-1)  : 1296
```

Gambar 8. Hasil labeling otomatis

4.4. Labeling Manual

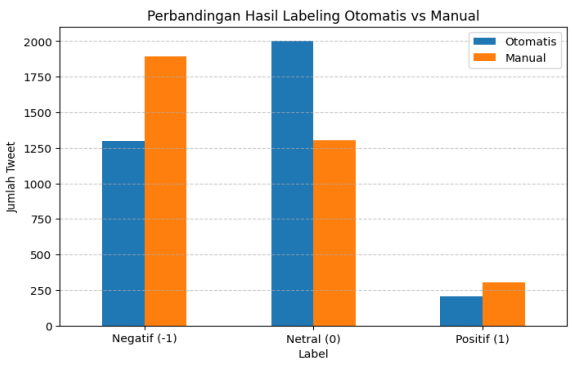
Labeling manual dilakukan untuk meningkatkan akurasi dengan mempertimbangkan konteks kalimat secara lebih mendalam, yang sering kali tidak dapat diidentifikasi hanya dengan pendekatan berbasis kata kunci.

```
Jumlah Tweet Berdasarkan Sentimen (Setelah Perbaikan Manual):
Positif (1)   : 303
Netral (0)    : 1306
Negatif (-1)  : 1894
```

Gambar 9. Hasil labeling manual

4.5. Perbandingan Hasil Labeling Otomatis dan Manual

Dari hasil di atas, terlihat bahwa labeling otomatis cenderung mengklasifikasikan lebih banyak tweet sebagai netral, sedangkan labeling manual menunjukkan jumlah yang lebih tinggi untuk sentimen negatif dan positif. Hal ini menunjukkan bahwa model berbasis kata kunci memiliki keterbatasan dalam memahami konteks dan ironi dalam teks Twitter.



Gambar 10. Hasil labeling otomatis dan manual

4.6. Evaluasi Performa Model SVM

Dalam penelitian ini, model Support Vector Machine (SVM) digunakan untuk klasifikasi sentimen pada data Twitter. Evaluasi performa dilakukan dengan membandingkan dua jenis kernel, yaitu Linear dan Radial Basis Function (RBF), menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score.

```
Best parameter: {'C': 1}
Accuracy (Linear Kernel with Best C): 0.8146

Classification Report (Linear Kernel - Best Model):
              precision    recall  f1-score   support

          -1       0.88      0.83      0.85       400
           0       0.72      0.84      0.78       255
           1       0.89      0.54      0.68        46

    accuracy                           0.81       701
   macro avg       0.83      0.74      0.77       701
weighted avg       0.82      0.81      0.81       701
```

Gambar 11. Model karnel linear

```
Best Parameters (RBF Kernel): {'C': 1, 'gamma': 1}
Accuracy (RBF Kernel): 0.8117

Classification Report (RBF Kernel):
              precision    recall  f1-score   support

          -1       0.85      0.86      0.86       400
           0       0.74      0.78      0.76       255
           1       0.88      0.50      0.64        46

    accuracy                           0.81       701
   macro avg       0.83      0.72      0.75       701
weighted avg       0.81      0.81      0.81       701
```

Gambar 12. Model karnel RBF

Berdasarkan hasil evaluasi, model SVM dengan kernel Linear lebih unggul dibandingkan kernel RBF. Kernel Linear memiliki akurasi lebih tinggi (81.46% vs 81.17%) serta performa lebih baik dalam recall dan F1-score makro. Model dengan kernel Linear mencapai akurasi 81.46% dengan parameter terbaik C = 1, sedangkan model dengan kernel RBF memiliki akurasi 81.17% dengan parameter C = 1, gamma = 1. Pada kernel Linear, precision tertinggi diperoleh untuk sentimen positif (0.89), namun recall lebih rendah (0.54), yang menyebabkan F1-score lebih kecil dibandingkan sentimen lainnya. Sebaliknya, kernel RBF sedikit lebih baik dalam mendeteksi sentimen negatif dengan F1-score lebih tinggi dibandingkan kernel Linear.

Dari hasil evaluasi kedua model, akurasi terbaik diperoleh menggunakan Kernel Linear dengan parameter C = 1, yang mencapai 81.46%. Kernel Linear lebih stabil dalam mengklasifikasikan berbagai sentimen dibandingkan Kernel RBF. Beberapa kelebihan dari Kernel Linear adalah memberikan akurasi lebih tinggi dan performa yang lebih seimbang dalam klasifikasi sentimen, lebih sederhana serta cepat dibandingkan Kernel RBF pada data berdimensi tinggi, dan lebih efektif dalam menangani klasifikasi sentimen dengan fitur yang terdistribusi linier. Namun, Kernel Linear juga memiliki kekurangan, seperti kurang efektif untuk data dengan pola non-linear serta F1-score yang lebih rendah untuk sentimen positif dibandingkan sentimen lainnya.

Sementara itu, Kernel RBF memiliki keunggulan dalam mendeteksi sentimen negatif dengan F1-score lebih tinggi serta mampu menangani pola data yang lebih kompleks dibandingkan Kernel Linear. Namun, kelemahan dari Kernel RBF adalah akurasinya yang lebih rendah dibandingkan Kernel Linear serta kecenderungan mengalami overfitting jika parameter tidak dioptimalkan dengan baik. Dari hasil evaluasi ini, SVM dengan Kernel Linear merupakan model terbaik untuk analisis sentimen dalam penelitian ini. Model ini memiliki keseimbangan yang baik antara akurasi, precision, recall, dan F1-score sehingga dapat digunakan secara efektif dalam klasifikasi sentimen Twitter.

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa sentimen masyarakat di Twitter terhadap kebijakan efisiensi anggaran terbagi ke dalam tiga kategori, yaitu negatif, netral, dan positif. Dari hasil analisis, mayoritas sentimen yang ditemukan bersifat negatif, menunjukkan adanya ketidakpuasan atau ketidaksetujuan dari masyarakat terhadap kebijakan tersebut. Sentimen netral juga cukup dominan, sementara sentimen positif memiliki proporsi yang lebih kecil. Dalam perbandingan performa model Support Vector Machine (SVM) dengan kernel yang berbeda, diperoleh bahwa Kernel Linear memiliki akurasi terbaik sebesar 81.46%, sedikit lebih tinggi dibandingkan dengan Kernel RBF yang mencapai 81.17%. Model dengan Kernel Linear lebih stabil dalam mengklasifikasikan sentimen negatif dan netral dibandingkan Kernel RBF, yang cenderung lebih baik dalam menangani pola data yang lebih kompleks tetapi memiliki performa yang kurang stabil dalam klasifikasi sentimen positif. Oleh karena itu, SVM dengan Kernel Linear dipilih sebagai model terbaik untuk analisis sentimen dalam penelitian ini.

Penelitian selanjutnya disarankan untuk mengeksplorasi model algoritma lain, seperti Random Forest, Naïve Bayes, atau deep learning, guna membandingkan performanya dengan SVM yang telah digunakan dalam penelitian ini. Selain itu, peningkatan jumlah data dapat dilakukan agar model lebih mampu melakukan generalisasi dan menangkap pola sentimen dengan lebih akurat serta andal, terutama dalam menghadapi variasi bahasa dan konteks yang lebih luas.

DAFTAR PUSTAKA

[1] M. M. Maarif dan N. Setiyawati, “Analisis Sentimen Review Aplikasi LinkedIn di Google Play Store Menggunakan Support Vector Machine,” *Progresif J. Ilm. Komput.*, vol. 20, no. 1, hal. 454, 2024, doi: 10.35889/progresif.v20i1.1614.

[2] M. N. Muttaqin dan I. Kharisudin, “Analisis Sentimen Pada Ulasan Aplikasi Gojek Menggunakan Metode Support Vector Machine dan K Nearest Neighbor,” *UNNES J. Math.*, vol. 10, no. 2, hal. 22–27, 2021, [Daring]. Tersedia pada: http://journal.unnes.ac.id/sju/index.php/ujm

[3] R. D. Himawan dan E. Eliyani, “Perbandingan Akurasi Analisis Sentimen Tweet terhadap

Pemerintah Provinsi DKI Jakarta di Masa Pandemi," *J. Edukasi dan Penelit. Inform.*, vol. 7, no. 1, hal. 58, 2021, doi: 10.26418/jp.v7i1.41728.

[4] D. Angraina dan A. Putri, "Analisis Sentimen Pengguna Aplikasi Google Meet Menggunakan Algoritma Support Vector Machine," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, hal. 472–478, 2022, doi: 10.37859/coscitech.v3i3.4260.

[5] A. Handayani dan I. Zufria, "Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM," *J. Inf. Syst. Res.*, vol. 5, no. 1, hal. 53–63, 2023, doi: 10.47065/josh.v5i1.4379.

[6] F. A. Artanto, "Support Vector Machine Berbasis Particle Swarm Optimization Pada Analisis Sentimen Anggota KPPS," vol. 14, no. 1, hal. 75–79, 2024.

[7] R. Ramlan, N. Satyahadewi, dan W. Andani, "Analisis Sentimen Pengguna Twitter Menggunakan Support Vector Machine Pada Kasus Kenaikan Harga BBM," *Jambura J. Math.*, vol. 5, no. 2, hal. 431–445, 2023, doi: 10.34312/jjom.v5i2.20860.

[8] D. Puspitasari dan T. Sutabri, "Analisis Sentimen Berdasarkan pada Twitter (X) terhadap Layanan Indihome Menggunakan Algoritma Support Vector Machine (SVM)," vol. 3, no. 2, hal. 58–71, 2024, doi: 10.55123/jumintal.v3i2.4449.

[9] D. Darwis, E. S. Pratiwi, dan A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, hal. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.

[10] O. I. Gifari, M. Adha, F. Freddy, dan F. F. S. Durrand, "Film Review Sentiment Analysis Using TF-IDF and Support Vector Machine," *J. Inf. Technol.*, vol. 2, no. 1, hal. 36–40, 2022.

[11] H. Syah dan A. Witanti, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 Pada Media Sosial Twitter Menggunakan Algoritma Support Vector Machine (Svm)," *J. Sist. Inf. dan Inform.*, vol. 5, no. 1, hal. 59–67, 2022, doi: 10.47080/simika.v5i1.1411.

[12] Ali Nur Ikhsan, Pungkas Subarkah, Alifah Dafa Iftinani, dan A. N. Fadilah, "Analisis sentimen aplikasi tiktok menggunakan algoritma naïve bayes dan support vector machine," *TEKNOSAINS J. Sains, Teknol. dan Inform.*, vol. 10, no. 2, hal. 176–184, 2023, doi: 10.37373/tekno.v10i2.419.

[13] H. C. Husada dan A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, hal. 18–26, 2021, doi: 10.34148/teknika.v10i1.311.

[14] T. Safitri, Y. Umaidah, dan I. Maulana, "Analisis Sentimen Pengguna Twitter Terhadap Grup Musik BTS Menggunakan Algoritma Support Vector Machine," *J. Appl. Informatics Comput.*, vol. 7, no. 1, hal. 28–35, 2023, doi: 10.30871/jaic.v7i1.5039.

[15] A. Muhammadin dan I. A. Sobari, "Analisis Sentimen Pada Ulasan Aplikasi Kredivo Dengan Algoritma Svm Dan Nbc," *Reputasi J. Rekayasa Perangkat Lunak*, vol. 2, no. 2, hal. 85–91, 2021, doi: 10.31294/reputasi.v2i2.785.

[16] R. W. Pratiwi, S. F. H, D. Dairoh, D. I. Af'idah, Q. R. A, dan A. G. F, "Analisis Sentimen Pada Review Skincare Female Daily Menggunakan Metode Support Vector Machine (SVM)," *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 1, hal. 40–46, 2021, doi: 10.20895/inista.v4i1.387.

[17] T. M. Permata Aulia, N. Arifin, dan R. Mayasari, "Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19," *SINTECH (Science Inf. Technol. J.*, vol. 4, no. 2, hal. 139–145, 2021, doi: 10.31598/sintechjournal.v4i2.762.