ANALISIS SENTIMEN PUBLIC TWITTER TERHADAP KEBIJAKAN PEMERINTAH MENGGUNAKAN METODE SVM (STUDI KASUS : PROGRAM EFISIENSI ANGGARAN)

Fadia Rizka Mumtaz, M. Jodi Pratama, Muhammad Azmi Zaky,
Iqbal Akbar Kurniawan, Eka Saputra, Ken Ditha Tania, Winda Kurnia Sari
Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang - Prabumulih No.KM. 32, Indralaya Indah,
Kec. Indralaya, Kabupaten Ogan Ilir, Sumatera Selatan, Indonesia
09031282227049@student.unsri.ac.id

ABSTRAK

Kebijakan efisiensi anggaran pemerintah, sebagaimana diamanatkan dalam Instruksi Presiden Nomor 1 Tahun 2025, memicu beragam reaksi publik, terutama di media sosial. Pemotongan anggaran sebesar Rp306,69 triliun menimbulkan kekhawatiran terkait dampaknya terhadap layanan publik dan kesejahteraan masyarakat. Oleh karena itu, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kebijakan efisiensi anggaran melalui media sosial Twitter dengan menggunakan metode Support Vector Machine (SVM). Data dikumpulkan melalui teknik web scraping pada periode tertentu, kemudian diproses melalui tahapan preprocessing seperti cleansing, tokenisasi, dan stemming. Model SVM diterapkan untuk mengklasifikasikan opini publik menjadi kategori positif, netral, dan negatif. Hasil penelitian menunjukkan bahwa model yang digunakan memiliki akurasi sebesar 96,37%, dengan nilai recall tertinggi pada kelas netral sebesar 1,00, yang mengindikasikan bahwa mayoritas respons publik bersifat netral. Namun, model masih menunjukkan variasi dalam klasifikasi opini positif dan negatif dengan nilai recall masing-masing sebesar 0,99 dan 0,90. Studi ini memberikan wawasan mengenai pola sentimen publik terhadap kebijakan efisiensi anggaran serta dapat menjadi bahan pertimbangan bagi pemerintah dalam mengambil kebijakan yang lebih responsif terhadap opini masyarakat.

Kata kunci : *Analisis Sentimen; Support Vector Machine; Efisiensi Anggaran; Media Sosial; Kebijakan Publik*

1. PENDAHULUAN

Kebijakan efisiensi anggaran pendapatan dan belanja negara merupakan isu yang kerap menjadi perbincangan hangat di kalangan masyarakat Indonesia. Pada awal tahun 2025, pemerintah Indonesia menerbitkan Instruksi Presiden Nomor 1 Tahun 2025 yang mengamanatkan efisiensi belanja dalam pelaksanaan APBN dan APBD guna mengalokasikan anggaran yang lebih optimal untuk program prioritas nasional. Berdasarkan instruksi tersebut, pemerintah menargetkan pemotongan belanja sebesar Rp306,69 triliun yang terdiri dari pengurangan anggaran kementerian/lembaga sebesar Rp256,1 triliun dan pemangkasan transfer ke daerah sebesar Rp50,59 triliun [1].

Kebijakan ini selaras dengan postur APBN 2025 yang memiliki total belanja negara sebesar Rp3.621,3 triliun, dengan komposisi belanja pemerintah pusat Rp2.701,4 triliun dan transfer ke daerah Rp919,9 triliun [2].

Terkait adanya rencana efisiensi anggaran menjadikan masyarakat menyampaikan berbagai tanggapan, baik tanggapan positif maupun negatif. Dengan perkembangan teknologi saat ini, informasi tersebar dengan cepat melalui sosial media, yang menghasilkan berbagai tanggapan, baik positif maupun negatif. Opini yang diberikan oleh masyarakat perlu dikaji dengan melibatkan pemrosesan teks karena bentuk data opini publik yang belum terstruktur. Hal tersebut dilakukan agar menjadi bahan pertimbangan bagi pemerintah dalam menghadapi permasalahan ini dengan mempertimbangkan kondisi maupun opini publik [3].

Jumlah pengguna internet di Indonesia diperkirakan mencapai 221.563.479 jiwa pada 2024, dari total penduduk 278.696.200 jiwa pada 2023. Survei Penetrasi Internet Indonesia 2024 menunjukkan Tingkat penetrasi internet mencapai 79,5%, dengan peningkatan 1,4% dibandingkan periode sebelumnya[4].

Dalam hal ini, studi terdahulu menunjukkan bahwa penerapan *Knowledge Management System* (KMS) yang dikombinasikan dengan teknik *Knowledge Data Discovery* (KDD) dapat membantu dalam mengelola dan mengekstraksi wawasan berharga dari kumpulan data dalam skala besar. Pada penelitian terdahulu yang menganalisis ulasan aplikasi dengan membandingkan model *Word Embedding* dan CNN-LSTM untuk analisis sentimen. Penelitian ini mengindikasikan bahwa penggunaan teknik *machine learning* yang lebih mutakhir dapat meningkatkan akurasi dalam menginterpretasikan opini publik dari data yang tidak terstruktur [5].

Penelitian terkini telah mengkaji penggunaan metode *Support Vector Machine* (SVM) untuk analisis sentimen kebijakan publik, contohnya penelitian oleh Yasir dkk [6] yang menunjukkan akurasi masing-masing mencapai 82% dan 63% dalam mengklasifikasikan opini masyarakat.

Di sisi lain, studi oleh Chairunnisa dkk [7] menerapkan pendekatan SVM dalam konteks analisis sentimen terhadap kebijakan vaksinasi COVID-19 serta pembatasan sosial, menghasilkan tingkat presisi dan recall hingga 90%. Selain itu, penelitian lain yang dilakukan oleh Dendy & Sugihartono [8] mengungkap bahwa optimalisasi parameter SVM melalui teknik *grid search* mampu meningkatkan akurasi analisis sentimen hingga mencapai 91%.

Berdasarkan penelitian sebelumnya, peneliti bertujuan melakukan penelitian dengan tujuan melakukan analisis sentimen opini publik mengenai kebijakan efisiensi anggaran di Indonesia menggunakan Algoritma *Support Vector Machine* (SVM). Algoritma SVM merupakan metode yang memiliki performa dan tingkat akurasi yang paling tinggi dibandingkan dengan beberapa metode klasifikasi *machine learning* lainnya karena mampu mengenali pola dan menganalisis data dengan dimensi tinggi sehingga tingkat akurasi yang dihasilkan menjadi lebih baik. [9]

Penelitian ini menganalisis sentimen pengguna aplikasi X terhadap kebijakan efisiensi anggaran pemerintah menggunakan metode *Support Vector Machine* (SVM). Sistematikanya mencakup pendahuluan yang membahas latar belakang kebijakan dan dampaknya, metode penelitian yang meliputi pengumpulan serta pra-pemrosesan data, implementasi SVM, serta evaluasi performa model. Hasil dan pembahasan memaparkan tingkat akurasi, presisi, dan recall dari model yang digunakan, serta tren sentimen yang muncul.

Kajian terdahulu menunjukkan bahwa SVM telah digunakan secara luas dalam analisis opini publik, seperti studi Pradhana dkk [10] yang mengkaji sentimen terhadap kebijakan PPKM dan penelitian Hasibuan & Suhardi [11] terkait kebijakan vaksin COVID-19, keduanya menunjukkan efektivitas SVM dalam klasifikasi sentimen. Berbeda dari penelitian sebelumnya, studi ini fokus pada analisis sentimen setelah kebijakan efisiensi anggaran diterapkan, sehingga memberikan gambaran real-time mengenai reaksi masyarakat di media sosial. Dengan memahami pola sentimen publik, hasil penelitian ini diharapkan dapat menjadi masukan bagi pemerintah dalam mengevaluasi efektivitas kebijakan serta menyusun strategi yang lebih responsif terhadap opini masyarakat [12].

2. TINJAUAN PUSTAKA

Penelitian oleh Hasibuan & Serdano menunjukkan bahwa dalam analisis sentimen kebijakan pembelajaran tatap muka, SVM mencapai akurasi 88,09%, sedangkan *Naïve Bayes* hanya memperoleh akurasi 75,92%, dengan selisih 12,17% yang membuktikan keunggulan SVM dalam menangani data dengan pola yang lebih kompleks [13].

Hasil ini mengindikasikan bahwa SVM mampu membangun model yang lebih baik dalam mengklasifikasikan opini publik, terutama pada kebijakan yang menimbulkan pro dan kontra di masyarakat. Selain itu, penelitian oleh Nada mengenai analisis sentimen terhadap BPJS di Twitter menemukan bahwa SVM dengan kernel RBF mencapai akurasi 97,1%, sedangkan *Naïve Bayes* hanya memperoleh 86,7%, menunjukkan perbedaan sebesar 10,4% yang mengonfirmasi bahwa SVM dapat lebih unggul dalam menangani data dalam jumlah besar dan berdimensi tinggi [14].

Keunggulan ini menjadi faktor penting dalam analisis sentimen, mengingat data yang diambil dari media sosial seringkali memiliki karakteristik teks yang panjang, tidak terstruktur, dan mengandung banyak variasi dalam penulisan.

Selain penelitian sebelumnya, studi yang dilakukan oleh Widyanto tentang analisis sentimen terhadap RUU Kesehatan di Twitter juga menunjukkan hasil yang serupa, di mana SVM dengan kernel linear memiliki tingkat akurasi 87%, lebih tinggi dibandingkan dengan Multinomial *Naïve Bayes* yang hanya mencapai 82%, dengan selisih 5% [15].

Hasil ini memperlihatkan bahwa meskipun *Naïve Bayes* tetap menjadi metode yang cukup akurat dalam klasifikasi teks, SVM tetap lebih unggul dalam mengidentifikasi sentimen dengan perbedaan yang lebih jelas antara opini positif, negatif, dan netral. Penelitian oleh Ikhsan yang berfokus pada analisis sentimen kebijakan kenaikan PPN di Indonesia juga mendukung keunggulan SVM, di mana SVM mencapai akurasi 98% sebelum *balancing*, sedangkan *Naïve Bayes* hanya memperoleh akurasi 97% sebelum *balancing* dan menurun menjadi 90% setelah balancing, menunjukkan bahwa SVM memiliki keunggulan 1% lebih tinggi sebelum *balancing* dan 8% lebih baik setelah *balancing* [3].

Ini menunjukkan bahwa SVM lebih stabil dalam menghadapi perubahan distribusi data, karena pada kasus tertentu, metode probabilistik seperti *Naïve Bayes* bisa mengalami penurunan akurasi ketika distribusi data berubah secara signifikan.

Penelitian terbaru oleh Salam yang membahas analisis sentimen terhadap Bantuan Langsung Tunai (BLT) BBM juga menunjukkan bahwa SVM memiliki akurasi 85,98%, dengan nilai precision 82,25% dan recall 66,35%, sementara metode lain cenderung menghasilkan performa yang lebih rendah dalam presisi dan *recall* [16].

Perbedaan akurasi ini menunjukkan bahwa SVM mampu menghasilkan keputusan klasifikasi yang lebih presisi dibandingkan metode lainnya, terutama dalam kasus analisis opini publik terhadap kebijakan pemerintah yang sering kali mengandung sentimen yang bercampur. Penelitian oleh Idris dkk. (2020) menggunakan metode Support Vector Machine (SVM) dalam klasifikasi sentimen terhadap ulasan aplikasi Google Play Store. Hasil penelitian menunjukkan bahwa SVM mampu menghasilkan akurasi yang tinggi dalam membedakan sentimen positif dan negatif setelah melalui proses praproses

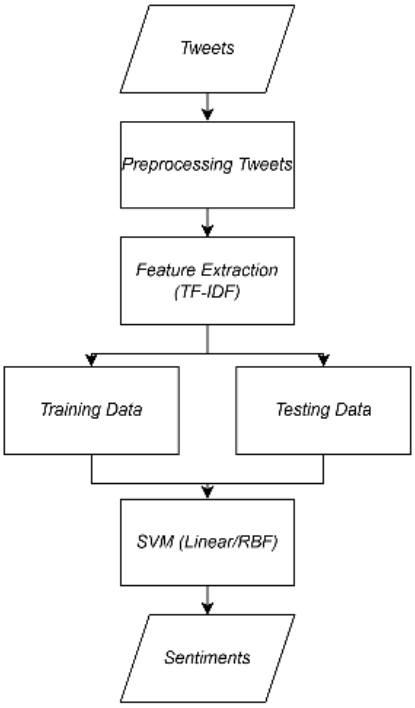
data dan ekstraksi fitur menggunakan TF-IDF [17]. Sementara itu, penelitian oleh Anna dan Zulfikar (2022) menerapkan metode SVM untuk analisis sentimen terhadap tweet mengenai vaksin COVID-19. Mereka menggunakan pendekatan supervised learning dengan proses pelabelan manual, serta menunjukkan bahwa SVM memberikan hasil akurasi yang cukup kompetitif dibandingkan metode lain seperti Naïve Bayes dan K-NN [18].

Berdasarkan berbagai penelitian yang telah dibahas, metode SVM dipilih dalam penelitian ini karena kemampuannya dalam menangani data berdimensi tinggi, kestabilannya dalam berbagai distribusi data, serta akurasinya yang lebih tinggi dibandingkan metode lainnya. Dengan demikian, penggunaan SVM dalam analisis sentimen terhadap kebijakan efisiensi anggaran pemerintah diharapkan dapat memberikan hasil yang lebih valid dan dapat dijadikan sebagai dasar dalam perumusan kebijakan publik yang lebih responsif terhadap opini masyarakat.

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin yang digunakan untuk klasifikasi dan regresi yang bekerja dengan mencari hyperplane terbaik yang memisahkan data ke dalam dua atau lebih kelas dengan margin maksimal [19].

Prinsip utama dari SVM adalah menggunakan fungsi kernel untuk mentransformasikan data ke dalam dimensi yang lebih tinggi, memungkinkan pemisahan yang lebih optimal [6].

Adapun alur dari proses kerja SVM dapat dilihat pada Gambar 1.



Gambar 1. Proses Kerja SVM

Proses kerja SVM dalam analisis sentimen dimulai dengan pengumpulan data dari media sosial seperti Twitter yang dilakukan melalui metode crawling untuk mendapatkan data mentah yang akan digunakan dalam penelitian [20].

Data yang telah dikumpulkan selanjutnya melalui proses preprocessing yang meliputi pembersihan teks dengan menghapus stopwords, melakukan stemming, lemmatization, serta tokenisasi agar data lebih terstruktur dan siap untuk dianalisis [21].

Setelah tahap preprocessing, data dikonversi menjadi representasi numerik menggunakan metode seperti TF-IDF atau Word Embeddings yang membantu dalam meningkatkan performa model SVM dengan memberikan bobot yang lebih relevan terhadap kata-kata yang sering muncul dalam data [22].

Tahapan berikutnya adalah membagi data menjadi training dan testing set untuk memastikan model SVM mendapatkan generalisasi yang baik saat diterapkan pada data baru [23].

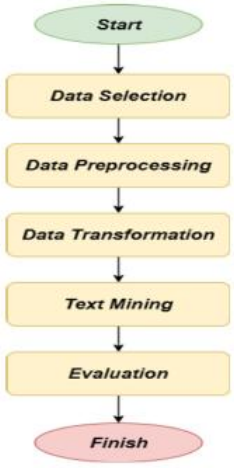
Model kemudian dilatih menggunakan kernel yang sesuai seperti Linear, RBF, atau Polynomial yang dipilih berdasarkan karakteristik data yang dianalisis [24]. Proses evaluasi dilakukan dengan menguji model menggunakan metrik akurasi, precision, recall, dan F1-score untuk mengukur seberapa baik model SVM dalam mengklasifikasikan sentimen secara tepat [25].

Setelah mendapatkan model terbaik, model diterapkan pada data baru untuk melakukan analisis sentimen secara otomatis yang kemudian hasilnya dapat digunakan dalam mendukung pengambilan keputusan kebijakan publik [26].

3. METODE PENELITIAN

Metode penelitian adalah sebuah prosedur sistematis yang disusun untuk melaksanakan penelitian guna mencapai tujuan yang diinginkan. Dalam penelitian ini menggunakan metode *Knowledge Discovery Database* (KDD). *Knowledge Discovery Database* (KDD) didefinisikan sebagai proses mengekstraksi informasi bermanfaat dari sekumpulan data [5].

Adapun alur rancangan penelitian yang dilakukan dapat dilihat pada Gambar 2.



Gambar 2. Alur penelitian

3.1. Data Selection

Pada tahap ini dilakukan proses pengumpulan data dengan teknik *crawling*. *Crawling* merupakan teknik mengumpulkan data pada sebuah website dengan memasukkan *Uniform Resource Locator* (URL). URL ini menjadi acuan untuk mencari semua *hyperlink* yang ada pada website. Kemudian dilakukan *indexing* untuk mencari kata dalam dokumen pada setiap link yang ada. Data yang diambil merupakan *tweets* yang terdapat dalam media sosial twitter dengan pemanfaatan media API twitter. Akses tweets membutuhkan hak akses berupa *consumer key*, *consumer secret*, *acces token* dan *access token secret*[27].

Kata kunci yang menjadi keyword adalah Efisiensi Anggaran. *Crawling* data diambil dari data dengan *tweets* tanggal 1 Desember 2024 sampai dengan 22 Februari 2025 dengan total perolehan tweets sejumlah 1.006 tweets.

3.2. Data Preprocessing

Setelah mendapatkan 1006 data *tweets* dengan kata kunci Efisiensi Anggaran. Data akan masuk ke tahap selanjutnya yaitu *preprocessing* data. *Preprocessing* data dilakukan dalam enam tahapan yaitu :

a. *Cleansing* : untuk mengurangi noise pada data tweet yang sudah dikumpulkan, seperti simbol-simbol, link URL, hashtag "#", dan at "@".
b. *Case Folding* : mengubah huruf kapital (uppercase) menjadi huruf kecil (lowercase), yang bertujuan untuk menghindari redudansi data.
c. *Tokenization* : yaitu melakukan pemotongan kalimat pada teks yang dipisahkan berdasarkan tiap kata yang menyusunnya. Kata-kata tersebut dipisahkan dengan spasi dan tanda petik tunggal.
d. *Stopword Removal* : pada tahap ini dilakukan penghapusan kata-kata yang dianggap tidak diperlukan pada data tweet, seperti kata penghubung "dan", "yang", "ke", dan lain sebagainya.
e. *Normalization* : merupakan proses untuk mengubah kata tidak baku menjadi kata yang baku pada data tweet.
f. *Stemming* : pada tahap ini dilakukan perbaikan kata-kata pada data tweet yang memiliki kata imbuhan menjadi kata dasar.

3.3. Labelling

Merupakan proses yang dilakukan untuk menentukan kelas dari setiap ulasan yang didapat, apakah termasuk ulasan positif atau negatif.

3.4. Data Transformation

Pada tahapan ini mengubah data label yang berupa positif, netral, dan negatif ke dalam data numeric. Positif diasumsikan sebagai 2, negatif diasumsikan sebagai 1, dan netral diasumsikan sebagai 0.

3.5. Text Mining

Pada tahap text mining dilakukan proses pengolahan data dengan mengklasifikasi data tweet menggunakan algoritma *Support Vector Machine* (SVM) dengan tujuan untuk mengolah data mentah menjadi informasi yang berguna.

3.6. Evaluation

Pada tahapan terakhir dilakukan evaluasi untuk mengukur validitas dari model yang terbentuk dari algoritma yang telah digunakan. Pengukuran performa algoritma dilakukan berdasarkan 3 faktor yaitu, *Accuracy*, *Precision*, dan *F1-Score* [28].

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini adalah proses analisis sentimen pengguna aplikasi "X" terhadap kebijakan efisiensi anggaran menggunakan algoritma *Support Vector Machine*, dan diklasifikasikan menjadi tiga kategori yaitu sentimen positif, sentimen netral, dan sentimen negatif dengan menggunakan bahasa pemrograman python.

4.1. D*ata Selection*

Data selection dilakukan dengan cara melakukan *crawling data* pada aplikasi "X", dengan mengumpulkan cuitan yang mengandung kata "kebijakan efisiensi anggaran". Data yang berhasil dikumpulkan sebanyak 1006 data. Tabel 1 dibawah menunjukkan dataset awal dari proses pengambilan data dengan teknik *crawling*.

Tabel 1. Dataset awal hasil *crawling data*

No	*Date*	*User*	*Tweet*
1	2025-02-20 08:21:14	Blackzodiac81	Si tolol loe memahami konteks *hingga ada kebijakan selanjutnya aja gak bisa. tolol nih gw jelasin gara gara efisiensi sudah dpt di pastikan 2025 ini gak ada anggaran dana utk sekolah spesialis. kecuali ada kebijakan selanjutnya yang blm bisa dipastikan kapan ???? ngerti ga
2	2025-02-19 11:26:01	Alamsyahsaragih	Ada-ada saja... Efisiensi itu (1) output tercapai dengan biaya lebih sedikit (2) tanpa menambah biaya output melebihi yang direncanakan. Bgmn publik bisa tahu bhw target 2025 akan tercapai meski anggaran dikurangi? Sdh ada transisi 6 bln lho?
3	2025-02-19 07:19:30	Baron224455	Efisiensi anggaran tahun 2025 untuk optimalkan program berdampak langsung pada rakyat
4	2025-02-18 13:29:47	Nonox2314	Inpres no. 1/2025 adalah langkah baik pemerintah utk melakukan efisiensi dalam tataran makro dgn memangkas anggaran pada kegiatan rutin puluhan

No	*Date*	*User*	*Tweet*
			tahun. Bila Pimpinan K/L dan Pemda bisa mengikuti kebijakan tsb pada masing2 sektor tentu ke depan efisiensi benar2 efektif.
5	2025-02-17 04:49:4	samudrafakta77	Untuk tahun anggaran 2025 Sri Mulyani menjelaskan ada 1.040.192 mahasiswa yang akan menerima beasiswa KIP. Anggaran beasiswa tersebut mencapai Rp14 7 triliun. Sri menekankan alokasi tersebut tidak mengalami efisiensi.

4.2. *Data Preprocessing*

Pada tahap ini data-data yang telah dikumpulkan akan melalui *preprocessing*. Tujuan dari data *preprocessing* adalah untuk membersihkan, mengorganisir, dan mempersiapkan data sehingga dapat diolah dengan baik. *Preprocessing* dilakukan melalui Google Collab dengan menggunakan bahasa Phyton.

4.2.1. *Cleansing*

Sebelum dilakukan cleansing terdapat banyak *noise* pada data, dan setelah proses *cleansing* data menjadi lebih terstruktur dan bersih, sehingga akan mudah untuk diproses pada tahap selanjutnya.

Tabel 2. *Cleansing*

Sebelum *Cleansing*	**Setelah *Cleansing***
Si tolol loe memahami konteks *hingga ada kebijakan selanjutnya aja gak bisa. tolol nih gw jelasin gara gara efisiensi sudah dpt di pastikan 2025 ini gak ada anggaran dana utk sekolah spesialis. kecuali ada kebijakan selanjutnya yang blm bisa dipastikan kapan ???? ngerti ga	Si tolol loe memahami konteks hingga ada kebijakan selanjutnya aja gak bisa tolol nih gw jelasin gara gara efisiensi sudah dpt di pastikan 2025 ini gak ada anggaran dana utk sekolah spesialis kecuali ada kebijakan selanjutnya yang blm bisa dipastikan kapan ngerti ga

4.2.2. *Case Folding*

Hasil dari proses *case folding* huruf pada setiap kata pada dataset yang sebelumnya huruf *uppercase* diubah menjadi huruf *lowercase* semua.

Tabel 3. *Case Folding*

Sebelum *Case Folding*	**Setelah *Case Folding***
Si tolol loe memahami konteks hingga ada kebijakan selanjutnya aja gak bisa tolol nih gw jelasin gara gara efisiensi sudah dpt di pastikan 2025 ini gak ada anggaran dana utk sekolah spesialis kecuali ada kebijakan selanjutnya yang blm bisa dipastikan kapan ngerti ga	si tolol loe memahami konteks hingga ada kebijakan selanjutnya aja gak bisa tolol nih gw jelasin gara gara efisiensi sudah dpt di pastikan 2025 ini gak ada anggaran dana utk sekolah spesialis kecuali ada kebijakan selanjutnya yang blm bisa dipastikan kapan ngerti ga

4.2.3. *Tokenization*

Pada tahap ini, teks dipisahkan menjadi kata kata individu berdasarkan spasi.

Tabel 4. *Tokenization*

Sebelum *Tokenization*	**Setelah *Tokenization***
si tolol loe memahami konteks hingga ada kebijakan selanjutnya aja gak bisa tolol nih gw jelasin gara gara efisiensi sudah dpt di pastikan 2025 ini gak ada anggaran dana utk sekolah spesialis kecuali ada kebijakan selanjutnya yang blm bisa dipastikan kapan ngerti ga	['si', 'tolol', 'loe', 'memahami', 'konteks', 'hingga', 'ada', 'kebijakan', 'selanjutnya', 'aja', 'gak', 'bisa', 'tolol', 'nih', 'gw', 'jelasin', 'gara', 'gara', 'efisiensi', 'sudah', 'dpt', 'di', 'pastikan', '2025', 'ini', 'gak', 'ada', 'anggaran', 'dana', 'utk', 'sekolah', 'spesialis', 'kecuali', 'ada', 'kebijakan', 'selanjutnya', 'yang', 'blm', 'bisa', 'dipastikan', 'kapan', 'ngerti', 'ga']

4.2.4. *Stopword Removal*

Pada tahap ini dilakukan penghapusan kata-kata yang umum digunakan tetapi tidak terlalu penting dalam analisis.

Tabel 5. *Stopword Removal*

Sebelum *Stopword Removal*	**Setelah *Stopword Removal***
['si', 'tolol', 'loe', 'memahami', 'konteks', 'hingga', 'ada', 'kebijakan', 'selanjutnya', 'aja', 'gak', 'bisa', 'tolol', 'nih', 'gw', 'jelasin', 'gara', 'gara', 'efisiensi', 'sudah', 'dpt', 'di', 'pastikan', '2025', 'ini', 'gak', 'ada', 'anggaran', 'dana', 'utk', 'sekolah', 'spesialis', 'kecuali', 'ada', 'kebijakan', 'selanjutnya', 'yang', 'blm', 'bisa', 'dipastikan', 'kapan', 'ngerti', 'ga']	['tolol', 'loe', 'memahami', 'konteks', 'kebijakan', 'selanjutnya', 'gak', 'bisa', 'tolol', 'nih', 'gw', 'jelasin', 'gara', 'gara', 'efisiensi', 'sudah', 'dpt', 'pastikan', '2025', 'gak', 'anggaran', 'dana', 'utk', 'sekolah', 'spesialis', 'kecuali', 'kebijakan', 'selanjutnya', 'blm', 'bisa', 'dipastikan', 'kapan', 'ngerti', 'ga']

4.2.5. *Normalization*

Tahap ini dilakukan dengan mengganti kata tidak baku ke kata baku.

Tabel 6. *Normalization*

Sebelum *Normalization*	**Setelah *Normalization***
['tolol', 'loe', 'memahami', 'konteks', 'kebijakan', 'selanjutnya', 'gak', 'bisa', 'tolol', 'nih', 'gw', 'jelasin', 'gara', 'gara', 'efisiensi', 'sudah', 'dpt', 'pastikan', '2025', 'gak', 'anggaran', 'dana', 'utk', 'sekolah', 'spesialis', 'kecuali', 'kebijakan', 'selanjutnya', 'blm', 'bisa', 'dipastikan', 'kapan', 'ngerti', 'ga']	['tolol', 'kamu', 'memahami', 'konteks', 'kebijakan', 'selanjutnya', 'tidak', 'bisa', 'tolol', 'nih', 'saya', 'jelasin', 'gara', 'gara', 'efisiensi', 'sudah', 'dapat', 'pastikan', '2025', 'tidak', 'anggaran', 'dana', 'untuk', 'sekolah', 'spesialis', 'kecuali', 'kebijakan', 'selanjutnya', 'belum', 'bisa', 'dipastikan', 'kapan', 'ngerti', 'tidak']

4.2.6. *Stemming*

Terakhir, kata-kata yang memiliki imbuhan akan dikembalikan ke dalam bentuk dasarnya.

Tabel 7. *Stemming*

Sebelum *Stemming*	Sesudah *Stemming*
['tolol', 'kamu', 'memahami', 'konteks', 'kebijakan', 'selanjutnya', 'tidak', 'bisa', 'tolol', 'nih', 'saya', 'jelasin', 'gara', 'gara', 'efisiensi', 'sudah', 'dapat', 'pastikan', '2025', 'tidak', 'anggaran', 'dana', 'untuk', 'sekolah', 'spesialis', 'kecuali', 'kebijakan', 'selanjutnya', 'belum', 'bisa', 'dipastikan', 'kapan', 'ngerti', 'tidak']	['tolol', 'kamu', 'paham', 'konteks', 'kebijakan', 'lanjut', 'tidak', 'bisa', 'tolol', 'nih', 'saya', 'jelas', 'gara', 'gara', 'efisiensi', 'sudah', 'dapat', 'pasti', '2025', 'tidak', 'anggaran', 'dana', 'untuk', 'sekolah', 'spesialis', 'kecuali', 'kebijakan', 'lanjut', 'belum', 'bisa', 'pasti', 'kapan', 'ngerti', 'tidak']

4.3. *Data Transformation*

Pada tahap ini, data yang telah melewati proses *preprocessing* akan melalui *data transformation*. Tujuan dari *data transformation* adalah mengubah format, struktur, atau nilai data agar lebih sesuai untuk analisis atau pemrosesan lebih lanjut. Transformasi data dilakukan untuk meningkatkan kualitas data, menyamakan skala, serta memastikan kompatibilitas dengan metode atau model yang digunakan.

4.3.1. *Labelling*

Proses pelabelan dilakukan secara manual dengan mempertimbangkan keberadaan kata-kata yang memiliki makna sentimen tertentu. Tabel 8 di bawah menunjukkan daftar kata yang dikategorikan sebagai *positive words* dan *negative words*. Jika suatu ulasan mengandung lebih banyak kata dari daftar *positive words*, maka ulasan tersebut diberi label positif. Sebaliknya, jika ulasan lebih banyak mengandung kata dari daftar *negative words*, maka diberi label negatif. Ulasan yang tidak mengandung kata dari kedua kategori tersebut atau memiliki keseimbangan antara kata positif dan negatif akan diberi label netral.

Tabel 8. *Labelling*

Positive word	Negative Word
['bagus', 'baik', 'puas', 'senang', 'mantap', 'suka', 'hebat','bijak', 'dapat', 'kerja', 'sesuai', 'langkah', 'lengkap', 'optimal']	['buruk', 'kecewa', 'jelek', 'parah', 'benci', 'payah', 'mengecewakan',"kurang", "dampak", "potong", "kena"]

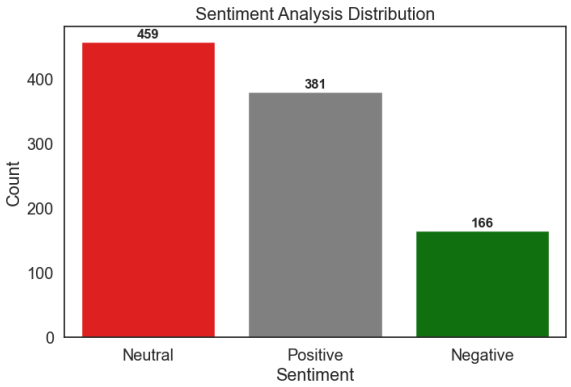
4.3.2. *Label Encoding*

Setelah dilakukan proses *labelling*, masing-masing kelas yang telah diberikan label akan diubah menjadi angka agar selanjutnya bisa diproses oleh *machine learning.* Tabel 9 dibawah ini menunjukkan hasil *label encoding* dari kelas-kelas yang ada.

Tabel 9. *Label Encoding*

Label	*Label Encoding*
Positive	2
Negative	1
Neutral	0

Proses *label encoding* menghasilkan jumlah sentimen netral sebanyak 459 data, sentimen positif sebanyak 381 data dan sentimen negatif sebanyak 166 data. Grafik visualisasi distribusi sentimen dari hasil proses *label encoding* ditunjukkan oleh gambar 3.



Gambar 3. Grafik distribusi sentimen

4.4. *Text Mining*

Pada tahap *text mining*, proses klasifikasi dilakukan dengan menerapkan algoritma SVM terhadap data teks yang telah melalui tahap *preprocessing*. Data yang digunakan dalam klasifikasi terdiri dari fitur yang diekstraksi menggunakan TF-IDF serta fitur tambahan berupa jumlah kata positif dan negatif dalam teks. Sementara itu, label ulasan yang sebelumnya berbentuk kategori (positif, negatif, dan netral) telah dikonversi ke dalam format numerik untuk dapat digunakan dalam model SVM.

Selanjutnya, dilakukan optimasi hyperparameter menggunakan GridSearchCV untuk menentukan kombinasi parameter terbaik, seperti kernel, nilai C, dan gamma. Model yang telah dilatih dievaluasi menggunakan metrik akurasi, *precision*, *recall*, *F1-score*, dan confusion matrix.

```
              precision    recall  f1-score   support

           0       0.98      1.00      0.99        83
           1       0.94      0.99      0.96       105
           2       0.99      0.90      0.94        88

    accuracy                           0.96       276
   macro avg       0.97      0.96      0.96       276
weighted avg       0.96      0.96      0.96       276

Akurasi Model SVM: 0.9637681159420289
```

Gambar 4. Hasil program klasifikasi SVM

Gambar diatas merupakan hasil skor akurasi, laporan klasifikasi dan tabel confusion matrix. Adapun tingkat akurasi model SVM yang diterapkan mencapai 96,37%.

4.5. *Evaluation*

Evaluasi dilakukan untuk mengukur performa model *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen publik terhadap kebijakan efisiensi anggaran. Beberapa aspek yang dievaluasi meliputi akurasi, *precision*, *recall*, *F1-score*, serta rekomendasi perbaikan berdasarkan analisis hasil model. Model SVM yang digunakan dalam penelitian ini memiliki akurasi sebesar 96,37%, dengan nilai *recall* tertinggi pada kelas netral sebesar 1,00, yang mengindikasikan bahwa model dapat mengenali semua data dengan sentimen netral secara sempurna. Sementara itu, nilai *recall* untuk kelas negatif dan positif masing-masing sebesar 0,99 dan 0,90, menunjukkan bahwa model mampu mengklasifikasikan sentimen publik dengan tingkat keakuratan yang sangat baik.

Precision, *recall*, dan *F1-score* dihitung untuk mengevaluasi keseimbangan antara prediksi positif dan negatif yang benar. *Precision* mengukur tingkat ketepatan prediksi positif yang benar dibandingkan dengan seluruh prediksi positif yang dibuat oleh model:

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

Sedangkan *recall* menunjukkan kemampuan model dalam mendeteksi seluruh data positif yang sebenarnya ada:

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

F1-score digunakan sebagai metrik keseluruhan untuk menyeimbangkan *precision* dan *recall*, dengan rumus:

$$F1\text{-}score = 2 \times \frac{Precision\ x\ Recall}{Precision + Recall} \quad (3)$$

Hasil evaluasi menunjukkan bahwa model memiliki nilai F1-score rata-rata sebesar 0,96, yang mencerminkan keseimbangan antara precision dan recall dalam klasifikasi sentimen. Tingginya nilai metrik evaluasi ini mengindikasikan bahwa model SVM yang digunakan memiliki performa yang optimal dalam mengklasifikasikan sentimen publik.

5. KESIMPULAN DAN SARAN

Kesimpulan dari penelitian "Analisis Sentimen Publik Twitter terhadap Kebijakan Pemerintah, Studi Kasus: Program Efisiensi Anggaran" menunjukkan bahwa model *Support Vector Machine* (SVM) memiliki akurasi 96.37%, dengan *recall* tertinggi 1.00 pada kelas netral, serta *recall* 0.90 pada kelas positif serta *recall* 0.99 pada kelas negatif. Hasil ini mengindikasikan bahwa model dapat mengenali sentimen netral dengan sempurna, tetapi masih perlu peningkatan dalam mendeteksi sentimen positif dan negatif. Untuk perbaikan, perlu dilakukan evaluasi terhadap kemungkinan *overfitting* pada kelas netral, *tuning hyperparameter* seperti nilai C dan pemilihan kernel, serta penambahan data pelatihan agar model lebih seimbang. Selain itu, penggunaan teknik *feature extraction* yang lebih kompleks seperti *word embeddings* dapat meningkatkan pemahaman model terhadap konteks sentimen.

DAFTAR PUSTAKA

[1] Kementerian Sekretariat Negara, "Instruksi Presiden Nomor 1 Tahun 2025 Tentang Efisiensi Anggaran," 2025.

[2] Informasi APBN 2025, "Akselerasi Pertumbuhan Ekonomi yang Inklusif dan Berkelanjutan," 2025.

[3] A. N. Ikhsan, P. Subarkah, A. D. Iftinani, and A. N. Fadilah, "Analisis Sentimen Kenaikan PPN Menggunakan Algoritma Naïve Bayes dan Support Vector Machine," *Infotekmesin*, vol. 16, no. 1, pp. 181–189, 2025.

[4] S. M. Prasetiyo, R. Gustiawan, and F. R. Albani, "Analisis Pertumbuhan Pengguna Internet di Indonesia," *Buletin Ilmiah Ilmu Komputer dan Multimedia*, vol. 2, no. 1, 2024.

[5] V. Novalia, K. D. Tania, A. Meiriza, and A. Wedhasmara, "Knowledge discovery of application review using word embedding's comparison with CNN–LSTM model on sentiment analysis," in *Proc. Int. Conf. Computer, Control, Informatics and Its Applications (IC3INA)*, Jakarta, Indonesia, 2023, pp. 29–34.

[6] M. Yasir, M. Haque, R. Suraji, and I. Sastrodiharjo, "Analisis Sentimen Terhadap Kontroversi Fatwa MUI Nomor 83 Tahun 2023 Tentang Pemboikotan Produk yang Terafiliasi Israel," *Jurnal Ekonomi Manajemen Sistem Informasi*, vol. 5, pp. 409–422, 2024.

[7] Q. A. Chairunnisa, Y. Herdiyeni, M. K. D. Hardhienata, and J. Adisantoso, "Analisis Sentimen Pengguna Twitter Terhadap Program Vaksinasi Covid-19 di Indonesia Menggunakan Algoritme Support Vector Machine," *Jurnal Ilmu Komputer dan Agri-Informatika*, vol. 9, no. 1, pp. 79–89, 2022.

[8] Dendy, A., & Sugihartono, T., "Perbandingan Teknik Optimasi Grid Search dan Randomized Search dalam Meningkatkan Akurasi Metode Klasifikasi SVM," *SKANIKA*, vol. 8, no. 1, pp. 13–22, 2025.

[9] B. Ermawan and N. Cahyono, "Optimasi Metode Klasifikasi Menggunakan FastText dan Grid Search," *JIKO*, vol. 9, p. 226, 2025.

[10] R. M. Pradhana, "Analisis Sentimen Publik terhadap Kebijakan Pemberlakuan Pembatasan Kegiatan Masyarakat Skala Mikro Menggunakan Algoritma Support Vector Machine Studi Kasus Twitter," Doctoral dissertation, Universitas Dinamika, 2021.

[11] M. S. Hasibuan and T. Suhardi, "Analysis of Covid-19 Vaccine Policy Sentiment Using SVM

and C4.5," *Jurnal Teknik Elektro dan Komputer TRIAC*, vol. 9, no. 2, pp. 67–69, 2022.

[12] H. Setiawan, E. Utami, and S. Sudarmawan, "Analisis Sentimen Twitter Kuliah Online Pasca Covid-19," *Jurnal Komtika*, vol. 5, no. 1, pp. 43–51, 2021.

[13] M. S. Hasibuan and A. Serdano, "Analisis Sentimen Kebijakan Pembelajaran Tatap Muka Menggunakan Support Vector Machine dan Naïve Bayes," *Jurnal Riset Sains dan Teknologi*, vol. 6, no. 2, pp. 199–204, 2022.

[14] D. D. Nada, "Perbandingan Analisis Sentimen Mengenai BPJS pada Media Sosial Twitter Menggunakan Naïve Bayes Classifier dan Support Vector Machine," Undergraduate thesis, Institut Teknologi Sepuluh Nopember (ITS), 2022.

[15] T. Widyanto, I. Ristiana, and A. Wibowo, "Komparasi Naïve Bayes dan SVM dalam Analisis Sentimen RUU Kesehatan di Twitter," *SINTECH Journal*, vol. 6, no. 3, 2023.

[16] R. R. Salam, M. F. Jamil, Y. Ibrahim, S. Rahmaddeni, and Herianto, "Analisis Sentimen terhadap Bantuan Langsung Tunai (BLT) Bahan Bakar Minyak (BBM) Menggunakan Support Vector Machine," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 3, no. 1, pp. 27–35, 2023.

[17] A. R. Idris, D. Yudhistira, and I. G. A. D. Cahyani, "Analisis Sentimen Masyarakat Terhadap Program Makan Gratis Menggunakan Algoritma K-Nearest Neighbor (K-NN)," *eProceedings of Applied Science*, vol. 9, no. 1, pp. 347–354, Apr. 2023.

[18] A. U. Rahmawati, A. D. A. Pangestu, N. R. Putri, M. R. Hidayat, and D. Alfina, "Sentiment Analysis of Twitter Users Regarding the Free Lunch Program Using IndoBERT," *Procedia Computer Science*, vol. 229, pp. 1903–1912, 2023.

[19] H. Akbar, F. Ramdhan, and S. Wijaya, "The Impact of Feature Engineering on Sentiment Analysis Performance Using SVM," *Journal of Machine Learning Applications*, vol. 14, no. 1, pp. 54–72, 2024.

[20] M. Rahman and R. Fadhil, "Support Vector Machine for Text Classification: A Comprehensive Review," *Journal of Artificial Intelligence Research*, vol. 45, pp. 112–130, 2021.

[21] D. Suryana and N. Lestari, "Text Preprocessing Techniques for Sentiment Analysis in Social Media Data," *Journal of Computer Science and Information Systems*, vol. 12, no. 4, pp. 221–238, 2023.

[22] P. Wijaya and T. Prasetyo, "TF-IDF vs. Word Embeddings: Which One is Better for Sentiment Analysis?," in *Proc. Int. Conf. on Data Science*, pp. 33–45, 2022.

[23] R. Hakim and S. Rahman, "Training and Testing Strategies in Machine Learning-Based Sentiment Analysis," *Int. J. Artif. Intell. Data Min*., vol. 11, no. 2, pp. 87–102, 2024.

[24] B. Santoso, A. Nugraha, and M. Setyawan, "Comparison of Kernel Types in SVM for Sentiment Analysis," *J. Intell. Syst. Appl*., vol. 9, no. 4, pp. 211–226, 2025.

[25] M. Firdaus and A. Ramadhan, "Evaluating Sentiment Classification Models: Metrics and Best Practices," *J. Data Sci. Mach. Learn*., vol. 13, no. 1, pp. 45–61, 2023.

[26] T. Nugroho, F. Mahendra, and D. Satria, "Application of Sentiment Analysis for Public Policy Decision Making," *Int. J. Gov. Technol. Stud*., vol. 8, no. 3, pp. 190–204, 2025.

[27] M. T. S. Jaye, I. M. A. D. Suarjaya, and N. K. D. Rusjayanthi, “Rancang Bangun Aplikasi Get Data di Media Sosial Twitter dan YouTube Berbasis Desktop,” *JITTER: J. Ilm. Teknol. Komput*., vol. 3, no. 2, pp. 1165–1175, 2022.

[28] D. Putra and A. Wibowo, "Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, vol. 2, pp. 84–92, 2020.