

Optimasi *Support Vector Machine* Menggunakan *Particle Swarm Optimization* pada Analisis Sentimen Program Efisiensi Anggaran Pemerintah

¹Nurhayati, ²Lili Tanti, ³Budi Triandi

^{1,2,3}Ilmu Komputer, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, Medan, Indonesia

*Korespondensi: n.hayati2390@gmail.com

Submit : 20 Jan 2026 | Diterima : 02 Mar 2026 | Terbit : 10 Mar 2026

ABSTRACT

Public response to government policies, including budget efficiency programs, can be understood using a sentiment analysis approach. This study aims to compare the performance of Support Vector Machine (SVM) algorithms with Linear, Polynomial, and Radius Basis Functional (RBF) kernels, as well as evaluate the effect of Particle Swarm Optimization (PSO) on improving classification performance. The dataset was obtained from Twitter social media using a crawling technique, consisting of 4274 data points, and was processed through text cleaning stages such as cleaning, case folding, normalization, tokenization, stopword removal, stemming, labeling using a lexicon-based approach, and feature extraction using TF-IDF (Term Frequency-Inverse Document Frequency). The evaluation was conducted using accuracy, precision, recall, F1-score, and confusion matrix metrics. The results showed that the Linear SVM kernel successfully increased the accuracy to 0.7579 or 75.79%, while the RBF and Polynomial kernels did not provide significant improvements. In addition, the neutral class was the most difficult to classify. This study concludes that the combination of Linear SVM and PSO is the best model for analyzing sentiment on government budget efficiency policies, and emphasizes the importance of selecting the right kernel and optimization strategy in developing machine learning-based classification systems.

Keywords: *Sentiment Analysis, Budget Efficiency, Particle Swarm Optimization, Support Vector Machine, TF-IDF,*

ABSTRAK

Respon publik terkait kebijakan pemerintah, termasuk program efisien anggaran dapat dipahami menggunakan pendekatan analisis sentimen. Penelitian ini bertujuan membandingkan kinerja algoritma *Support Vector Machine* (SVM) dengan kernel *Radial Basis Functional* (RBF), kernel *Linear*, dan kernel *Polynomial*, serta mengevaluasi pengaruh *Particle Swarm Optimization* (PSO) terhadap peningkatan performa klasifikasi. Dataset sebanyak 4274 data diperoleh melalui teknik crawling dari media sosial X (Twitter) dan kemudian diproses melalui langkah-langkah pembersihan teks seperti *cleaning*, *case folding*, *normalization*, *tokenization*, *stopword removal*, *stemming*, *labeling* menggunakan *lexicon based* serta menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk ekstraksi fitur. Proses penilaian kinerja model dilaksanakan dengan memanfaatkan indikator *accuracy*, *precision*, *recall*, dan *F1-score*, serta didukung oleh analisis *confusion matrix*. Berdasarkan hasil pengujian yang diperoleh, penelitian ini menunjukkan bahwa SVM kernel *Linear* berhasil meningkatkan akurasi menjadi 0.7579 atau 75.79%, sedangkan pada kernel *RBF* dan *Polynomial* tidak memberikan peningkatan signifikan. Selain itu, kelas netral menjadi kelas yang paling sulit diklasifikasikan. Penelitian ini menyimpulkan bahwa kombinasi SVM *Linear* dan PSO merupakan model terbaik untuk analisis sentimen kebijakan efisiensi anggaran pemerintah, serta menegaskan pentingnya pemilihan kernel dan strategi optimasi yang tepat dalam pengembangan sistem klasifikasi berbasis *machine learning*.

Kata Kunci: Analisis Sentimen, Efisiensi Anggaran, Particle Swarm Optimization, Support Vector Machine, TF-IDF.

PENDAHULUAN

Perkembangan teknologi informasi dan media sosial telah mengubah cara masyarakat menyampaikan pendapat terhadap berbagai isu publik, termasuk kebijakan pemerintah. Media sosial khususnya Twitter menjadi sarana utama bagi masyarakat untuk mengekspresikan opini, kritik, maupun dukungan berupa tweet secara terbuka dan *real time*. Salah satu isu strategis yang banyak mendapat perhatian publik adalah program efisiensi anggaran pemerintah yang bertujuan untuk mengoptimalkan penggunaan anggaran negara. Namun, kebijakan ini menimbulkan beragam persepsi di masyarakat, baik berupa sentimen positif, negatif maupun netral.

Analisis sentimen yang sering disebut sebagai *opinion mining* telah menjadi bidang penelitian yang krusial dalam bidang *Natural Language Processing* (NLP), kecerdasan buatan dan analitik data. Pertumbuhan eksponensial konten yang dihasilkan pengguna di *platform* media sosial khususnya twitter telah menyebabkan ledakan data teks yang mengharuskan penggunaan metodologi canggih untuk mengekstrak opini, sentimen dan sikap dari data teks tersebut. Akan tetapi, hal ini sekaligus menghadirkan tantangan karena data yang dihasilkan bersifat tidak struktur, beragam dan mengandung banyak noise, seperti bahasa gaul, singkatan dan emotikon yang mempersulit analisis. Selain itu, pemahaman konteks dan nuansa emosi manusia mungkin tidak selalu dapat ditangkap dengan akurat yang menyebabkan interpretasi yang salah dalam klasifikasi sentimen (Hamza et al., 2023).

Pada dasarnya, analisis sentimen bertujuan untuk mengklasifikasikan emosi atau sikap yang terkandung dalam teks ke dalam kategori positif, negatif dan netral. Klasifikasi ini sangat penting dalam memahami perilaku konsumen, tren pasar, dan opini publik (Hossain & Rahman, 2022; Rajput, 2022; Sathya, 2023). Kerangka kerja yang digunakan dalam analisis sentimen sangat terkait dengan teori linguistik, paradigma *machine learning*, dan pendekatan statistik.

Dalam proses klasifikasi sentimen, pemilihan algoritma *machine learning* yang tepat sangat mempengaruhi kinerja sistem. *Support Vector Machine* (SVM) merupakan salah satu algoritma *supervised learning* bertujuan menemukan *hyperplane optimal* yang memisahkan data ke dalam kelas-kelas sentimen secara maksimal. SVM sangat efektif dalam menangani kasus di mana kelas-kelas tidak dapat dipisahkan secara linier, berkat trik kernel yang secara implisit memetakan fitur input ke ruang dimensi tinggi (Hameed et al., 2025). Berbagai penelitian menunjukkan bahwa kombinasi SVM dengan representasi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) mampu menghasilkan akurasi klasifikasi yang baik pada data teks media sosial (Nugroho & Maharani, 2024), (Ipadeola et al., 2025). Misalnya, penggunaan TF-IDF dalam analisis sentimen dan deteksi emosi telah menghasilkan akurasi melebihi 75% (Sutronggono & Imah, 2023) dengan penyempurnaan lebih lanjut yang meningkatkan metrik kinerja dengan akurasi 89% (Huang et al., 2022).

Meskipun demikian, menurut sejumlah penelitian menunjukkan bahwa kinerja SVM sangat dipengaruhi oleh pemilihan parameter yang tepat, dimana pengaturan parameter yang optimal mampu meningkatkan akurasi model secara signifikan (Manik et al., 2022; Wang et al., 2023). Parameter penalti (C) berperan dalam menyeimbangkan kesalahan pada data pelatihan dan kemampuan generalisasi model, sedangkan parameter kernel seperti *Radial Basis Function* (RBF) dan kernel *Polynomial*, menentukan fleksibilitas batas keputusan serta kemampuan SVM dalam memodelkan hubungan nonlinier pada data (Song et al., 2023; Wang et al., 2023). Oleh karena itu, diperlukan metode optimasi parameter untuk meningkatkan kinerja SVM. Salah satu metode yang banyak digunakan adalah *Particle Swarm Optimization* (PSO). PSO adalah algoritma optimasi metaheuristik yang terinspirasi oleh perilaku sosial yang diamati pada kawanan burung. Dikembangkan oleh Kennedy dan Eberhart pada tahun 1995, algoritma ini mensimulasikan pencarian kolaboratif untuk solusi optimal oleh sekelompok agen (partikel) dalam ruang pencarian multidimensi yang didefinisikan. Setiap partikel menyesuaikan posisinya berdasarkan pengalamannya sendiri dan pengalaman tetangganya menggunakan mekanisme pembaruan kecepatan yang menggabungkan pengetahuan individu dan kelompok. Mekanisme adaptif ini membuat PSO sangat cocok untuk menangani berbagai masalah optimasi diberbagai bidang, mulai dari desain teknik hingga aplikasi biomedis (Li-kun et al., 2023; Shami et al., 2022).

Beberapa studi telah menggunakan SVM untuk analisis sentimen tanpa optimasi, seperti penelitian yang dilakukan Iqrom Aidil Redho et al., (2025), menerapkan SVM untuk menganalisis sentimen pengguna pada ulasan *Google Play Store* untuk aplikasi Sirekap dengan menggunakan 15.000 data dengan kategori sentimen positif, negatif dan sangat negatif. Hasil menunjukkan bahwa akurasi yang diperoleh sebesar 88%, secara keseluruhan mayoritas ulasan bersifat sangat negatif dengan persentase sebesar 76.9%. Ulasan negatif menunjukkan

banyaknya keluhan pengguna terhadap bug, error, gagal login, dan lainnya yang membuat pengguna aplikasi Sirekap 2024 menjadi tidak nyaman.

Selain itu, penelitian Akhmad & Purwiyanto (2024), menggunakan SVM dengan Smote untuk menganalisis sentimen ulasan online tentang Bali Hai Cruises, menghasilkan tingkat akurasi sebesar 98.26%. Penelitian juga dilakukan oleh Pamungkas & Cahyono (2024), yang membandingkan SVM dengan KNN untuk menganalisis ulasan aplikasi ChatGPT menghasilkan SVM dengan akurasi sebesar 80% dan KNN sebesar 71%.

Pada Penelitian Supriyatna & Putri (2024), menggabungkan PSO dengan SVM untuk analisis ulasan game mencapai akurasi 97.61% dalam kondisi yang dioptimalkan dibandingkan dengan akurasi 97.28% tanpa optimasi. Hal ini menunjukkan bahwa optimasi dapat menghasilkan perbaikan marginal bahkan ketika hasil dasar sudah tinggi. Sementara itu, upaya optimasi lainnya dibahas dalam penelitian Ali & Hendrastuty (2024), yang membahas berbagai metode seperti Naive Bayes, Random Forest dan SVM, mencatat bahwa menggabungkan teknik optimasi dengan metode SVM dapat memberikan manfaat yang signifikan dengan akurasi sebesar 99%.

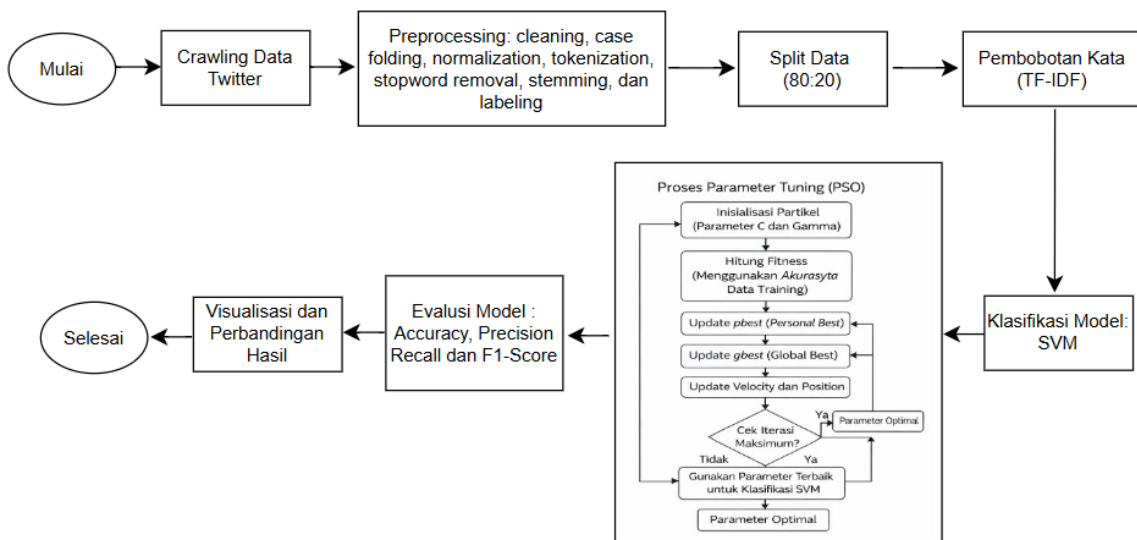
Dari berbagai penelitian tersebut menunjukkan perbedaan yang jelas antara analisis SVM yang dilakukan tanpa optimasi dan yang menerapkan optimasi. Studi tanpa optimasi melaporkan metrik kinerja yang solid dengan kisaran akurasi 80%-90%, menunjukkan bahwa tanpa optimasi SVM tetap menjadi metode yang tangguh. Namun disisi lain, SVM yang dikombinasikan dengan teknik optimasi cenderung melaporkan metrik akurasi yang sedikit lebih tinggi.

Berdasarkan uraian tersebut, penelitian ini menekankan pentingnya penerapan analisis sentimen berbasis *machine learning* untuk memahami opini publik di media sosial, khususnya dalam konteks evaluasi kebijakan atau program pemerintah. Dengan mengkombinasikan tahapan pra-pemrosesan teks, representasi fitur TF-IDF, algoritma SVM, serta optimasi parameter menggunakan PSO diharapkan dapat diperoleh model analisis sentimen yang memiliki kinerja optimal dan mampu memberikan gambaran yang lebih akurat mengenai persepsi masyarakat.

METODE PENELITIAN

2.1 Tahapan Penelitian

Metode penelitian ini disusun secara sistematis untuk membangun model klasifikasi sentimen terhadap program efisiensi anggaran pemerintah berdasarkan data teks dari media sosial twitter. Langkah-langkah penelitian ini dilakukan dengan beberapa tahapan, yaitu pengumpulan data, *preprocessing data*, *split data*, ekstraksi fitur menggunakan TF-IDF, proses klasifikasi menggunakan SVM, serta optimasi parameter model dengan PSO. Penelitian pada tahapan akhir dilakukan proses penilaian kinerja model menggunakan *confussion matrix*. Gambar 1 menyajikan alur penelitian secara umum sebagai panduan pelaksanaan penelitian.



Gambar 1. Tahapan Penelitian

2.2 Pengumpulan Data

Proses pengumpulan tweet (*crawling data*) dilakukan menggunakan *Twitter Application Programming Interface* (API). API ini memungkinkan pengambilan tweet publik secara *real-time* atau berdasarkan kata kunci, hastag, dan akun pengguna berdasarkan data historis. Kerangka kerja seperti *tweepy* dan *python library* sering digunakan untuk mengakses API twitter (Karuna et al., 2023; Vidyashree & Rajendra, 2023). Dataset yang diperoleh pada periode Januari s.d. Desember 2025 berjumlah 4274 data yang disimpan dalam format *.csv* (*comma separated values*) untuk dapat dianalisis lebih lanjut.

2.3 Preprocessing Data

Tahapan *preprocessing* data merupakan hal yang sangat penting karena mengubah teks mentah menjadi format yang lebih sesuai untuk analisis. Langkah-langkah awal yang dilakukan umumnya meningkatkan data, memastikan model dapat menginterpretasikan dan mempelajari pola dengan efektif. Misalnya, penelitian yang dilakukan bahwa *preprocessing* teks terstruktur, termasuk pembersihan data dan normalisasi meningkatkan kualitas input untuk pemodelan sentimen (Untoro & Farhan, 2025).

Tahap awal preprocessing data adalah *cleaning*, yaitu proses pembersihan teks yang mencakup penghapusan karakter yang tidak diinginkan, spasi kosong berlebihan, dan data yang tidak relevan yang dapat mempengaruhi hasil (Papia et al., 2024). Setelah melalui tahap *cleaning*, proses selanjutnya adalah *case folding*, yaitu mengubah semua teks menjadi huruf kecil secara seragam untuk menstandarkan data input dan menghindari ketidaksesuaian yang disebabkan oleh perbedaan huruf besar/kecil (Aufar et al., 2023). Selanjutnya ke tahapan *Normalization*, yaitu standarisasi data teks untuk mengurangi variasi dalam representasi data, tujuannya adalah memastikan bahwa berbagai bentuk kata diperlakukan sebagai entitas yang sama dalam analisis (Muhamad Agung Nulhakim et al., 2025). Setelah itu dilakukan proses *Tokenization*, yaitu memecah teks menjadi elemen-elemen individu atau token yang dapat berupa kata atau frasa (Syachrul et al., n.d.).

Tahapan selanjutnya adalah *Stopword Removal* yaitu, penyaringan kata-kata umum yang tidak menambah konten bermakna dalam analisis seperti, "yang", "di", "dan", "ke" dan sebagainya. Langkah ini dilakukan karena *stopword* dapat melemahkan skor sentimen sehingga analisis menjadi kurang efektif (Saputra et al., 2025). Setelah tahapan *Stopword Removal* langkah selanjutnya adalah *Stemming*, yaitu mengubah kata-kata menjadi bentuk dasar atau akar kata yang membantu mengurangi variasi bentuk kata dan meningkatkan efektivitas pencocokan terhadap kamus kata sentimen yang telah disiapkan sebelumnya (Minggow et al., 2025). Pada proses ini menggunakan fungsi *stemmer* pada *library sastrawi* untuk mengubah kata ke bentuk dasarnya. Tahapan terakhir dalam *preprocessing* adalah *labeling* yang menggunakan pendekatan *lexicon-based*, dimana setiap kata dalam teks diberi skor polaritas berdasarkan kamus sentimen yang telah ditentukan. Pendekatan ini tidak hanya memanfaatkan intensitas emosi tetapi juga memanfaatkan intuisi manusia dan pengetahuan sebelumnya untuk mengkategorikan sentimen secara akurat (Syachrul & Fatharani, 2025). Pada tahap ini, peneliti menggunakan kamus *lexicon* yang diperoleh dari repositori GitHub publik.

2.4 Split Data (Pembagian Data)

Pada tahap ini dataset dibagi ke dalam dua bagian, yaitu 80% sebagai data latih (*training*) dan 20% sebagai data uji (*testing*). Data pelatihan dimanfaatkan untuk melatih serta mengoptimasi kinerja algoritma klasifikasi agar mampu mempelajari pola dari data yang tersedia. Sementara data pengujian dilakukan untuk mengukur performa model menggunakan data yang tidak terlibat dalam proses pelatihan. Pembagian data ini bertujuan untuk memastikan evaluasi model dilakukan secara objektif serta meminimalkan resiko terjadinya *overfitting*.

2.5 Ekstraksi Fitur (TF-IDF)

Pada tahapan ini dilakukan proses ekstraksi fitur teks menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk merepresentasikan dokumen dalam bentuk numerik. TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen (TF) serta tingkat kemunculannya di seluruh kumpulan dokumen (IDF), sehingga kata-kata yang bersifat informatif memperoleh bobot lebih tinggi. Pendekatan ini memungkinkan pemilahan fitur yang relevan dengan menurunkan pengaruh kata

umum yang kurang bermakna. Nilai bobot TF-IDF berada pada rentang 0 hingga 1 yang merepresentasikan tingkat relevansi kata terhadap dokumen yang dianalisis (Addiga & Bagui, 2022). Nilai TF-IDF dihitung menggunakan persamaan rumus dengan langkah pertama yaitu perhitungan *Term Frequency* (TF) (1) :

$$tf(w, d) = \log(1 + f(w, d)) \quad (1)$$

Tahap selanjutnya adalah perhitungan *Inverse Dokument Frequency* (IDF) dengan persamaan (2):

$$idf(w) = \log\left(\frac{N}{df(w)}\right) \quad (2)$$

Tahap terakhir adalah perhitungan bobot TF-IDF dengan persamaan (3):

$$tf-idf(w, d, D) = tf(w, d) \times idf(w, D) \quad (3)$$

2.6 Klasifikasi

Model klasifikasi dibangun dengan menggunakan algoritma SVM dengan tujuan menganalisis pengaruh fungsi kernel terhadap kinerja klasifikasi. SVM dipilih karena kemampuannya dalam menangani data berdimensi tinggi serta efektivitasnya dalam membentuk batas keputusan optimal melalui pemaksimalan margin antar kelas. Pada penelitian ini digunakan tiga jenis kernel, yaitu kernel *Linear*, kernel *Polynomial* dan kernel *radial basis function* (RBF). Kernel *Linear* digunakan untuk memodelkan hubungan linier antar fitur, kernel *Polynomial* untuk menangkap pola nonlinier dengan kompleksitas tertentu, sedangkan RBF digunakan untuk memetakan data ke ruang berdimensi lebih tinggi sehingga mampu menangani pola data yang kompleks (Manik et al., 2022). SVM bekerja dengan membentuk sebuah *hyperplane* sebagai batas keputusan yang memisahkan data ke dalam kelas-kelas yang berbeda. Titik-titik data yang berada paling dekat dengan *hyperplane* yang dikenal sebagai *support vectors*, berperan penting dalam menentukan margin, yaitu jarak antara *hyperplane* dan titik terdekat dari setiap kelas (Nugroho & Maharani, 2024). Secara sistematis, *hyperplane* pada SVM direpresentasikan oleh fungsi keputusan dengan persamaan (4):

$$f(x) = wx + b \quad (4)$$

Dimana w merupakan vektor bobot, x adalah vektor fitur dari data pelatihan, dan b adalah nilai bias yang mengatur posisi batas keputusan.

2.7 Optimasi Parameter menggunakan PSO

Untuk mencari kombinasi parameter terbaik dan meningkatkan akurasi klasifikasi dilakukan optimasi parameter (*hyperparameter tuning*) menggunakan PSO. Karakteristik utamanya adalah bahwa semua partikel mengikuti pencarian kolaboratif di mana setiap partikel tertarik ke posisi global terbaik (G_{best}) dikawanan dan ke posisi terbaiknya sendiri (P_{best}). Pembaruan kecepatan dipengaruhi oleh bobot inersia yang merepresentasikan kontribusi kecepatan sebelumnya, serta dua parameter percepatan yang mengatur pengaruh komponen kognitif dan sosial terhadap pergerakan partikel. Proses iteratif ini terus berlangsung hingga kriteria penghentian terpenuhi, seperti batas iterasi maksimum atau tercapainya solusi optimal (Ballerini, 2023) yang dapat dilihat pada persamaan (5) dan persamaan (6) berikut:

$$V_i(t) = \omega V_i(t-1) + c_1 A(P_{best} - X_i(t-1)) + c_2 B(G_{best} - X_i(t-1)) \quad (5)$$

$$X_i(t) = V_i(t) + X_i(t-1) \quad (6)$$

2.8 Evaluasi Model

Setelah proses pelatihan dan pengujian selesai dilakukan, tahapan selanjutnya adalah evaluasi terhadap keberhasilan model analisis sentimen yang dibangun. Metrik evaluasi yang digunakan adalah *Multiclass Confusion Matrix* 3x3 yaitu positif, negatif dan netral. Dari hasil *confusion matrix*, maka dapat diidentifikasi dengan berbagai metrik evaluasi seperti *accuracy*, *precision*, *recall* dan *F1-Score*. *Accuracy* memberikan gambaran umum mengenai tingkat ketepatan klasifikasi, sedangkan *precision*, *recall* dan *f1-score* memberikan pemahaman yang

lebih rinci tentang kinerja model, terutama dalam kasus ketidakseimbangan kelas (Alsaedi, 2024).

HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data pada penelitian ini diperoleh dari komentar atau *tweet* pengguna twitter menggunakan teknik data *crawling* berbasis bahasa pemrograman *python* dengan memanfaatkan twitter API. Pengambilan data dilakukan secara bertahap berdasarkan kata kunci yang relevan seperti efisiensi anggaran, penghematan anggaran, pemotongan anggaran, pemangkasan anggaran dan beberapa kata kunci terkait lainnya. Data yang dikumpulkan kemudian diseleksi berjumlah 4274 data dan di simpan dalam format *.csv* guna memudahkan proses *preprocessing* dan tahap analisis lebih lanjut. Dataset akhir dari proses tersebut selanjutnya akan digunakan sebagai data utama dalam proses analisis sentimen, sebagaimana ditampilkan pada Gambar. 2

	conversation_id_str	created_at	favorite_count	full_text	id_str	image_url1
0	1942460910880407729	Tue Jul 08 05:49:04 +0000 2025	59	Efisiensi anggaran APBN 2025 lebih merupakan r...	1942460910880407729	https://pbs.twimg.com/media/GvUB4u7W4AAQURx.jpg
1	1942161695390196045	Mon Jul 07 10:00:06 +0000 2025	65	Outlook Pemerintah memprakirakan Belanja (Rp3...	1942161695390196045	https://pbs.twimg.com/media/GvPxoXBalAECEf4.png
2	1980701499715211590	Tue Oct 21 18:23:31 +0000 2025	1709	Mau diRestrukturisasi ampe 1000 tahun pun itu ...	1980701499715211590	https://pbs.twimg.com/media/G3zdgz7XcAADI3m.jpg
3	1964548223760289904	Sun Sep 07 04:36:10 +0000 2025	373	Sudah saatnya negara ini bubar. Sawah sepetak ...	1964548223760289904	https://pbs.twimg.com/media/G0N6M0nbUAAkSsL.jpg
4	1998193011197718995	Tue Dec 09 00:48:33 +0000 2025	2370	Saya menyimak dengan prihatin pernyataan dari ...	1998193011197718995	https://pbs.twimg.com/media/G7sB81WasAAkWBp.jpg

Gambar 2. Dataset Penelitian

3.2 Preprocessing Data

Tahap *preprocessing* dimulai dengan membaca dataset dalam format CSV untuk memastikan kelengkapan dan konsistensi data. Tahapannya yaitu *cleaning*, *case folding*, *normalization*, *tokenization*, *stopword removal*, *stemming* dan *labeling*.

3.2.1 Cleaning

Dataset yang memiliki *noise* akan dihapus. Data *noise* ini mencakup penghapusan karakter-karakter khusus seperti simbol, angka, tanda baca, URL dan tag HTML yang tidak memberikan makna sentimen, sebagaimana ditampilkan pada Tabel 1.

Tabel 1. Hasil *Cleaning*

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
Mahasiswa UI Kesulitan Akses Jurnal Setelah Kampus Terapkan Efisiensi Anggaran https://t.co/OtKSTVRkXp	Mahasiswa UI Kesulitan Akses Jurnal Setelah Kampus Terapkan Efisiensi Anggaran
Apakah penundaan pengangkatan ASN dan PPPK ada kaitannya dng efisiensi anggaran?bayangkan ada lebih dari 1 juta orang jadi terkatung-katung ada yg udah resign ngekos bareng dll https://t.co/gEhULoZyEf	Apakah penundaan pengangkatan ASN dan PPPK ada kaitannya dng efisiensi anggaran bayangkan ada lebih dari 1 juta orang jadi terkatung katung ada yg udah resign ngekos bareng dll
Presiden Prabowo Subianto menyatakan rasa bangga atas dedikasi para menteri di Kabinet Merah Putih yang tidak pernah	Presiden Prabowo Subianto menyatakan rasa bangga atas dedikasi para menteri di Kabinet Merah Putih yang tidak pernah

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
mengeluh di tengah kebijakan efisiensi anggaran. https://t.co/6w6pLJM7Tz	mengeluh di tengah kebijakan efisiensi anggaran

3.2.2 *Case Folding*

Pada tahap ini merupakan proses mengubah semua huruf kapital (*uppercase*) menjadi huruf kecil (*lowercase*), sebagaimana dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Case Folding*

Sebelum <i>Case_folding</i>	Sesudah <i>Case_folding</i>
Mahasiswa UI Kesulitan Akses Jurnal Setelah Kampus Terapkan Efisiensi Anggaran	mahasiswa ui kesulitan akses jurnal setelah kampus terapkan efisiensi anggaran
Apakah penundaan pengangkatan ASN dan PPPK ada kaitannya dng efisiensi anggaran bayangkan ada lebih dari 1 juta orang jadi terkatung katung ada yg udah resign ngekos bareng dll	apakah penundaan pengangkatan asn dan pppk ada kaitannya dng efisiensi anggaran bayangkan ada lebih dari 1 juta orang jadi terkatung katung ada yg udah resign ngekos bareng dll
Presiden Prabowo Subianto menyatakan rasa bangga atas dedikasi para menteri di Kabinet Merah Putih yang tidak pernah mengeluh di tengah kebijakan efisiensi anggaran	presiden prabowo subianto menyatakan rasa bangga atas dedikasi para menteri di kabinet merah putih yang tidak pernah mengeluh di tengah kebijakan efisiensi anggaran

3.2.3 *Normalization* (Normalisasi)

Tahap normalisasi merupakan merubah kata yang tidak baku, memperbaiki kata-kata yang salah ketik (*typo*), penyingkatan kata dan bahasa *slang* atau bahasa gaul menjadi kata baku atau kata yang seharusnya yang bersumber dari kamus dalam formal *excel* sebagaimana yang ditampilkan pada Tabel 3.

Tabel 3. Hasil Normalisasi

Sebelum Normalisasi	Sesudah Normalisasi
mahasiswa ui kesulitan akses jurnal setelah kampus terapkan efisiensi anggaran	mahasiswa ui kesulitan akses jurnal setelah kampus terapkan efisiensi anggaran
apakah penundaan pengangkatan asn dan pppk ada kaitannya dng efisiensi anggaran bayangkan ada lebih dari 1 juta orang jadi terkatung katung ada yg udah resign ngekos bareng	apakah penundaan pengangkatan asn dan pppk ada kaitannya dengan efisiensi anggaran bayangkan ada lebih dari 1 juta orang jadi terkatung katung ada yang sudah resign ngekos bareng
presiden prabowo subianto menyatakan rasa bangga atas dedikasi para menteri di kabinet merah putih yang tidak pernah mengeluh di tengah kebijakan efisiensi anggaran	presiden prabowo subianto menyatakan rasa bangga atas dedikasi para menteri di kabinet merah putih yang tidak pernah mengeluh di tengah kebijakan efisiensi anggaran

3.2.4 *Tokenization* (Tokenisasi)

Pada tahap tokenisasi, tweet-tweet akan dipisahkan menjadi potongan-potongan kata atau yang disebut sebagai *token* berdasarkan karakter spasi yang ada dalam tweet, dapat dilihat pada Tabel 4.

Tabel 4. Hasil Tokenisasi

Sebelum Tokenisasi	Sesudah Tokenisasi
mahasiswa ui kesulitan akses jurnal setelah kampus terapkan efisiensi anggaran	['mahasiswa', 'ui', 'kesulitan', 'akses', 'jurnal', 'setelah', 'kampus', 'terapkan', 'efisiensi', 'anggaran']
apakah penundaan pengangkatan asn dan pppk ada kaitannya dengan efisiensi anggaran bayangkan ada lebih dari 1 juta orang jadi terkatung katung ada yang sudah resign ngekos bareng dll	['apakah', 'penundaan', 'pengangkatan', 'asn', 'dan', 'pppk', 'ada', 'kaitannya', 'dengan', 'efisiensi', 'anggaran', 'bayangkan', 'ada', 'lebih', 'dari', '1', 'juta', 'orang', 'jadi', 'terkatung', 'katung', 'ada', 'yang', 'sudah', 'resign', 'ngekos', 'bareng', 'dll']
presiden prabowo subianto menyatakan rasa bangga atas dedikasi para menteri di kabinet merah putih yang tidak pernah mengeluh di tengah kebijakan efisiensi anggaran	['presiden', 'prabowo', 'subianto', 'menyatakan', 'rasa', 'bangga', 'atas', 'dedikasi', 'para', 'menteri', 'di', 'kabinet', 'merah', 'putih', 'yang', 'tidak', 'pernah', 'mengeluh', 'di', 'tengah', 'kebijakan', 'efisiensi', 'anggaran']

3.2.5 Stopword Removal

Pada proses ini dilakukan penghilangan kata-kata yang tidak penting dan kata-kata umum yang biasanya sering muncul dalam jumlah besar dan dianggap tidak memiliki makna sehingga kata tersebut akan dihilangkan, sebagaimana ditunjukkan pada Tabel 5.

Tabel 5. Hasil Stopword Removal

Sebelum Stopword	Sesudah Stopword
['mahasiswa', 'ui', 'kesulitan', 'akses', 'jurnal', 'setelah', 'kampus', 'terapkan', 'efisiensi', 'anggaran']	['mahasiswa', 'kesulitan', 'akses', 'jurnal', 'kampus', 'terapkan', 'efisiensi', 'anggaran']
['apakah', 'penundaan', 'pengangkatan', 'asn', 'dan', 'pppk', 'ada', 'kaitannya', 'dengan', 'efisiensi', 'anggaran', 'bayangkan', 'ada', 'lebih', 'dari', '1', 'juta', 'orang', 'jadi', 'terkatung', 'katung', 'ada', 'yang', 'sudah', 'resign', 'ngekos', 'bareng']	['penundaan', 'pengangkatan', 'asn', 'pppk', 'kaitannya', 'efisiensi', 'anggaran', 'bayangkan', 'juta', 'orang', 'terkatung', 'katung', 'resign', 'ngekos', 'bareng']
['presiden', 'prabowo', 'subianto', 'menyatakan', 'rasa', 'bangga', 'atas', 'dedikasi', 'para', 'menteri', 'di', 'kabinet', 'merah', 'putih', 'yang', 'tidak', 'pernah', 'mengeluh', 'di', 'tengah', 'kebijakan', 'efisiensi', 'anggaran']	['presiden', 'prabowo', 'subianto', 'bangga', 'dedikasi', 'menteri', 'kabinet', 'merah', 'putih', 'mengeluh', 'kebijakan', 'efisiensi', 'anggaran']

3.2.6 Stemming

Pada tahap *stemming* adalah proses mengubah kata-kata ke bentuk kata dasarnya sesuai dengan standar KBBI yang dapat dilihat pada Tabel 6.

Tabel 6. Hasil Stemming

Sebelum Stemming	Sesudah Stemming
mahasiswa ui kesulitan akses jurnal setelah kampus terapkan efisiensi anggaran	mahasiswa sulit akses jurnal kampus terap efisiensi anggar

Sebelum Stemming	Sesudah Stemming
apakah penundaan pengangkatan asn dan pppk ada kaitannya dengan efisiensi anggaran bayangkan ada lebih dari 1 juta orang jadi terkatung katung ada yang sudah resign ngekos bareng	tunda angkat asn pppk kait efisiensi anggaran bayang juta orang katung katung resign ngekos bareng
presiden prabowo subianto menyatakan rasa bangga atas dedikasi para menteri di kabinet merah putih yang tidak pernah mengeluh di tengah kebijakan efisiensi anggaran	presiden prabowo subianto bangga dedikasi menteri kabinet merah putih keluh bijak efisiensi anggaran

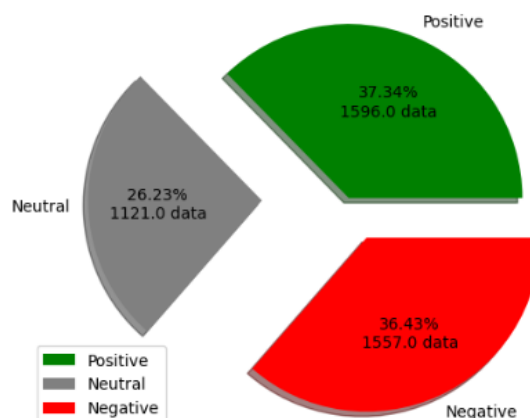
3.2.7 Labeling

Setelah seluruh data selesai dibersihkan proses selanjutnya adalah pelabelan. Pelabelan dilakukan secara otomatis menggunakan kamus kata Indonesia sentimen *lexicon* yang tiap katanya memiliki score polarity. Hasilnya dapat dilihat pada Tabel 7.

Tabel 7. Hasil Labeling

Text	Label	Polarity
mahasiswa sulit akses jurnal kampus terap efisiensi anggaran	negative	-3
tunda angkat asn pppk kait efisiensi anggaran bayang juta orang katung katung resign ngekos bareng dll	neutral	0
presiden prabowo subianto bangga dedikasi menteri kabinet merah putih keluh bijak efisiensi anggaran	positive	1

Mekanisme perhitungan skor dan penentuan label dilakukan setelah proses stemming. Setiap tweet yang telah melalui proses stemming kemudian dibandingkan dengan dua daftar kata tersebut. Proses analisis dilakukan melalui perhitungan skor polaritas, yaitu selisih antara jumlah kata positif dan negatif dalam satu tweet. Label sentimen ditentukan berdasarkan skor polaritas, yaitu skor >0 bernilai positif; skor <0 bernilai negatif; dan skor =0 bernilai netral. Berdasarkan proses pelabelan tersebut diperoleh 1596 data berlabel "Positif", 1121 data berlabel "Netral" dan 1557 data berlabel "Negatif", yang dapat dilihat pada Gambar 3.



Gambar 3. Diagram Hasil Pelabelan

3.3 Split Data (Pembagian Data)

Setelah dilakukan proses pelabelan, langkah selanjutnya adalah melakukan pembagian data dengan rasio 80:20, yaitu 3419 (80%) merupakan data latih dan 855 (20%) merupakan data uji, sebagaimana dapat Tabel 8.

Tabel 8. Split Data

Data Latih (Data Training)	3419 (80%)
Data Uji (Data Testing)	855 (20%)
Jumlah	4274

3.4 Ekstraksi Fitur (TF-IDF)

Pada tahap ini, data dikonversi ke dalam bentuk representasi numerik menggunakan metode TF-IDF. Hasil transformasi tersebut dapat dilihat pada Tabel 9.

Tabel 9. Hasil TF-IDF

Term	Indexing			TF		IDF (log D/df)	W (TF-IDF)		
	D1	D2	D3	df	D/df		D1	D2	D3
mahasiswa	1	0	0	1	3	0,4771	0,4771	0,0000	0,0000
sulit	1	0	0	1	3	0,4771	0,4771	0,0000	0,0000
akses	1	0	0	1	3	0,4771	0,4771	0,0000	0,0000
jurnal	1	0	0	1	3	0,4771	0,4771	0,0000	0,0000
kampus	1	0	0	1	3	0,4771	0,4771	0,0000	0,0000
terap	1	0	0	1	3	0,4771	0,4771	0,0000	0,0000
efisiensi	1	1	1	3	1	0,0000	0,0000	0,0000	0,0000
anggar	1	1	1	3	1	0,0000	0,0000	0,0000	0,0000
tunda	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
angkat	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
asn	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
pppk	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
kait	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
anggaran	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
bayangkan	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
juta	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
orang	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
terkatungtung	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
resign	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
ngekos	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
bareng	0	1	0	1	3	0,4771	0,0000	0,4771	0,0000
presiden	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
prabowo	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
subianto	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
bangga	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
dedikasi	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
menteri	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
kabinet	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
merah	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
putih	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
keluh	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771
bijak	0	0	1	1	3	0,4771	0,0000	0,0000	0,4771

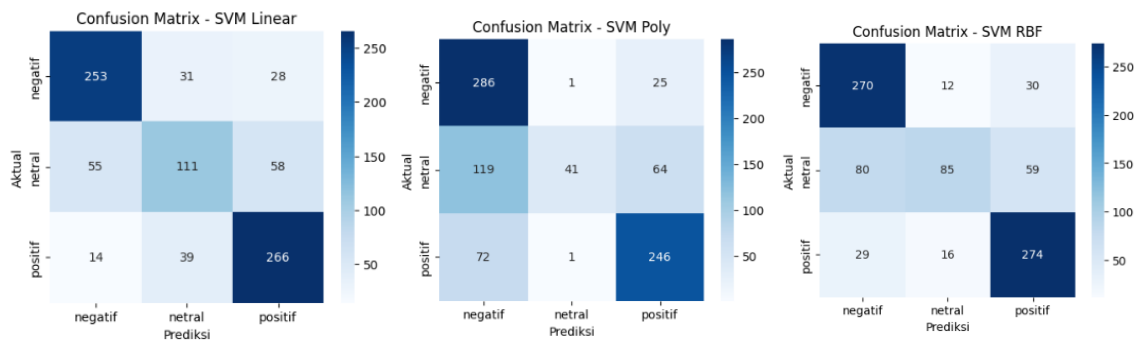
3.5 Klasifikasi SVM

Pada tahap awal, klasifikasi dilakukan menggunakan SVM tanpa optimasi parameter. Penelitian ini menggunakan metode SVM dengan tiga kernel yaitu *Linear*, *Polynomial* dan *RBF*. Parameter yang digunakan secara *default* untuk memperoleh *baseline* performa model. Ringkasan hasil pengukuran menggunakan metrik evaluasi yang ditampilkan pada Tabel 10.

Tabel 10. Hasil Metrik Evaluasi SVM *Baseline*

<i>Linear</i>		<i>Polynomial</i>		<i>RBF</i>	
Accuracy	0.7368	Accuracy	0.6702	Accuracy	0.7357
Precision	0.7293	Precision	0.7426	Precision	0.7387
Negatif	0.79	Negatif	0.60	Negatif	0.71
Netral	0.61	Netral	0.95	Netral	0.75
Positif	0.76	Positif	0.73	Positif	0.75
Recall	0.7368	Recall	0.6702	Recall	0.7357
Negatif	0.81	Negatif	0.92	Negatif	0.87
Netral	0.50	Netral	0.18	Netral	0.38
Positif	0.83	Positif	0.77	Positif	0.86
F1-Score	0.7307	F1-Score	0.6257	F1-Score	0.7171
Negatif	0.80	Negatif	0.72	Negatif	0.78
Netral	0.55	Netral	0.31	Netral	0.50
Positif	0.79	Positif	0.75	Positif	0.80

Berdasarkan hasil evaluasi Tabel 10 di atas, menunjukkan bahwa kernel *Linear* menghasilkan nilai *accuracy* tertinggi sebesar 0.7368 atau 73.68%, diikuti oleh kernel *RBF* sebesar 0.7357 atau 73.57%, sedangkan kernel *Polynomial* memperoleh nilai terendah sebesar 0.6702 atau 67.02%. Hasil tersebut mengindikasikan bahwa kernel *Linear* dan *RBF* lebih mampu merepresentasikan pola data dibandingkan kernel *Polynomial* pada konteks analisis sentimen yang digunakan. Rincian kinerja model dapat dilihat pada confusion matrix pada Gambar 4.



Gambar 4. Confusion Matrix SVM (Kernel *Linear*, *Polynomial* dan *RBF*)

Berdasarkan *confusion matrix* pada Gambar 4, kernel *Linear* dan *RBF* menunjukkan performa klasifikasi yang relatif lebih konsisten dibandingkan dengan kernel *Polynomial*. Kernel *Linear* mampu mengklasifikasikan data negatif dan positif dengan baik, namun masih mengalami kesalahan cukup signifikan pada kelas netral. Kernel *Polynomial* memperlihatkan kecenderungan bias terhadap kelas negatif, ditandai dengan tingginya jumlah prediksi benar pada kelas tersebut, tetapi sangat rendah dalam mengenali kelas netral. Sementara itu, kernel *RBF* menghasilkan distribusi prediksi yang lebih seimbang antar kelas, dengan tingkat klasifikasi yang baik pada kelas negatif dan positif serta performa yang lebih stabil pada kelas netral. Secara keseluruhan, hasil mengindikasikan bahwa kernel *Linear* dan *RBF* lebih representatif sebagai model *baseline* sebelum dilakukan optimasi menggunakan *Particle Swarm Optimization* (PSO).

3.6 Optimasi PSO

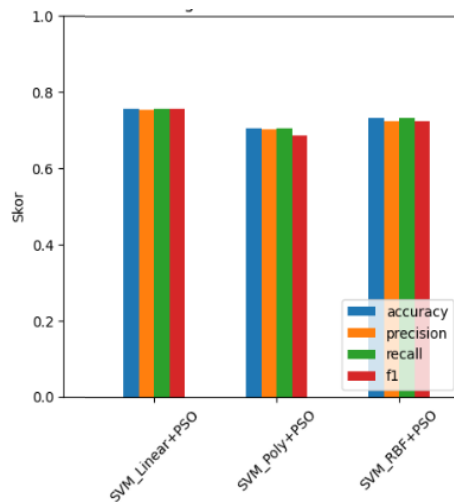
Pada klasifikasi SVM dengan optimasi PSO, penulis melakukan beberapa pengujian untuk menentukan parameter PSO yang terbaik pada penelitian ini. Pengujian yang dilakukan dapat dilihat pada Tabel 11.

Tabel 11. Pengujian Parameter

Pengujian	c1	c2	w	Iterasi	Partikel	Kernel	Akurasi
1	1.5	1.5	0.7	10	20	Linear	0.7578
						Poly	0.7064
						RBF	0.7321
2	1.5	1.5	0.7	20	20	Linear	0.7368
						Poly	0.7064
						RBF	0.7321
3	1.5	1.5	0.9	30	20	Linear	0.7368
						Poly	0.7040
						RBF	0.7321

Berdasarkan hasil pengujian parameter pada Tabel 11, dapat diketahui bahwa variasi konfigurasi PSO memberikan pengaruh yang berbeda terhadap performa SVM pada masing-masing kernel. Dengan mempertahankan nilai c1 dan c2 sebesar 1.5, eksperimen difokuskan pada perubahan inertia weight (w), jumlah iterasi dan jumlah partikel untuk mengevaluasi stabilitas akurasi model. Hasil pengujian menunjukkan, konfigurasi terbaik diperoleh pada pengujian pertama dengan nilai c1=1.5, c2=1.5, w=0.7, iterasi=10, dan jumlah partikel=20, dengan akurasi tertinggi sebesar 0.7578 atau 75.78% pada kernel Linear. Peningkatan jumlah iterasi dan partikel pada pengujian berikutnya tidak memberikan peningkatan performa yang signifikan, sehingga dapat disimpulkan bahwa konfigurasi tersebut telah mencapai kondisi konvergen. Oleh karena itu, kernel linear dengan parameter PSO pada pengujian pertama dipilih sebagai konfigurasi optimal untuk tahap evaluasi selanjutnya. Untuk hasil optimasi PSO dapat dilihat pada Gambar 5.

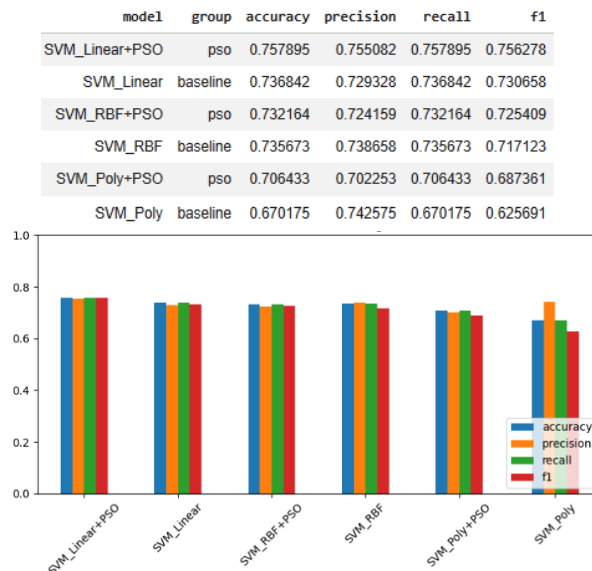
model	accuracy	precision	recall	f1
SVM_Linear+PSO	0.757895	0.755082	0.757895	0.756278
SVM_Poly+PSO	0.706433	0.702253	0.706433	0.687361
SVM_RBF+PSO	0.732164	0.724159	0.732164	0.725409



Gambar 5. Hasil Optimasi PSO

3.7 Evaluasi Model

Evaluasi dilakukan dengan membandingkan dua konfigurasi model, yaitu SVM tanpa optimasi PSO dan SVM yang diintegrasikan dengan PSO. Setiap model di analisis menggunakan metrik *accuracy*, *precision*, *recall*, serta *F1 score* untuk memperoleh gambaran menyeluruh mengenai performa klasifikasi yang disajikan pada Gambar 6.



Gambar 6. Hasil Perbandingan Evaluasi

Berdasarkan hasil yang tersaji pada Gambar 6 di atas, menunjukkan bahwa peningkatan performa paling signifikan terjadi pada kernel linear. SVM Linear + PSO mencapai akurasi sebesar 0.7579 atau 75.79%, lebih tinggi dibandingkan *baseline* sebesar 0.7368 atau (73.68%). Kenaikan kinerja tersebut juga konsisten pada nilai *precision*, *recall* dan *F1-score*. Sebaliknya, penerapan PSO pada kernel RBF tidak memberikan perbaikan kinerja, bahkan menunjukkan penurunan akurasi dari 0.7357 atau 73.57% menjadi 0.7322 atau 73.22%. pada kernel *Polynomial*, optimasi PSO hanya menghasilkan peningkatan yang sangat marginal pada akurasi, sementara *precision* mengalami penurunan.

KESIMPULAN

Penelitian ini mengkaji penerapan algoritma SVM dengan tiga kernel, yaitu *Linear*, *Polynomial* dan *RBF*, serta mengintegrasikannya dengan PSO untuk meningkatkan performa analisis sentimen terhadap kebijakan efisiensi anggaran pemerintah. Proses penilaian kinerja model dilakukan dengan memanfaatkan beberapa ukuran evaluasi, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, interpretasi hasil juga diperkuat melalui *confusion matrix* guna mengamati sebaran prediksi pada masing-masing kategori sentimen secara lebih rinci. Berdasarkan hasil pengujian yang telah dilakukan, model SVM dengan kernel linear menunjukkan performa yang relatif lebih konsisten dan stabil dibandingkan dengan penggunaan kernel *polynomial* maupun kernel *RBF*, baik pada kondisi *baseline* maupun setelah dilakukan optimasi. Integrasi PSO pada kernel *linear* berhasil meningkatkan akurasi model menjadi 0.7579 atau 75.79%, disertai peningkatan pada seluruh metrik evaluasi lainnya. Hal ini mengindikasikan bahwa PSO mampu menemukan konfigurasi parameter yang lebih optimal sehingga memperbaiki kemampuan klasifikasi model. Sebaliknya, penerapan PSO pada kernel *RBF* dan *Polynomial* tidak menghasilkan peningkatan performa yang signifikan, bahkan pada beberapa metrik mengalami penurunan. Temuan dalam studi ini mengindikasikan bahwa tingkat keberhasilan penerapan PSO dipengaruhi secara signifikan oleh jenis kernel yang digunakan serta karakteristik distribusi data yang dianalisis. Disamping itu, analisis terhadap *confusion matrix* menunjukkan bahwa kategori sentimen netral cenderung menjadi kelas yang paling sulit dikenali oleh seluruh model yang mengindikasikan adanya ambiguitas linguistik dalam teks sentimen. Berdasarkan keseluruhan hasil eksperimen, dapat disimpulkan bahwa kombinasi SVM (kernel linear) dengan optimasi PSO merupakan metode terbaik dalam penelitian ini. Penelitian

ini menegaskan pentingnya pemilihan kernel yang tepat serta penggunaan optimasi yang sesuai dalam meningkatkan performa sistem analisis sentimen berbasis *machine learning*.

REFERENSI

- Addiga, A., & Bagui, S. (2022). Sentiment Analysis on Twitter Data Using Term Frequency-Inverse Document Frequency. *Journal of Computer and Communications*, 10(08), 117–128. <https://doi.org/10.4236/jcc.2022.108008>
- Akhmad, E. P. A., & Purwiyanto, D. (2024). Sentiment Analysis of Bali Hai Cruises Reviews Using Support Vector Machine Algorithm. *JURNAL APLIKASI PELAYARAN DAN KEPSELABUHANAN*, 14(2), 149–157. <https://doi.org/10.30649/japk.v14i2.113>
- Ali, H., & Hendrastuty, N. (2024). COMPARISON OF NAÏVE BAYES CLASSIFIER, SUPPORT VECTOR MACHINE, RANDOM FOREST ALGORITHMS FOR PUBLIC SENTIMENT ANALYSIS OF KIP-K PROGRAM ON TWITTER. *Jurnal Teknik Informatika (Jutif)*, 5(6), 1701–1712. <https://doi.org/10.52436/1.jutif.2024.5.6.4030>
- Alsaedi, R. (2024). Sentiment Analysis of Arabic Tweets: Detecting Revilement. *Turkish Journal of Computer and Mathematics Education (Turcomat)*, 15(3), 312–323. <https://doi.org/10.61841/turcomat.v15i3.14726>
- Aufar, A. F., Mochamad Alfian Rosid, Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews. *JICTE (Journal of Information and Computer Technology Education)*, 7(2), 42–50. <https://doi.org/10.21070/jicte.v7i2.1650>
- Ballerini, L. (2023). *Particle Swarm Optimization in 3D Medical Image Registration: A Systematic Review*. <http://arxiv.org/abs/2302.11627>
- Hameed, S., Nauman, M., Akhtar, N., Fayyaz, M. A. B., & Nawaz, R. (2025). Explainable AI-driven depression detection from social media using natural language processing and black box machine learning models. *Frontiers in Artificial Intelligence*, 8. <https://doi.org/10.3389/frai.2025.1627078>
- Hamza, M. A., Alshahrani, H. J., Hassan, A. Q. A., Gaddah, A., Allheeb, N., Alsaif, S. A., Al-Onazi, B. B., & Mohsen, H. (2023). Computational Linguistics with Optimal Deep Belief Network Based Irony Detection in Social Media. *Computers, Materials and Continua*, 75(2), 4137–4154. <https://doi.org/10.32604/cmc.2023.035237>
- Hossain, M. S., & Rahman, M. F. (2022). Customer Sentiment Analysis and Prediction of Insurance Products' Reviews Using Machine Learning Approaches. *Fiib Business Review*, 12(4), 386–402. <https://doi.org/10.1177/23197145221115793>
- Huang, Y., Wang, X., Wang, R., & Min, J. (2022). Analysis and Recognition of Food Safety Problems in Online Ordering Based on Reviews Text Mining. In *Wireless Communications and Mobile Computing* (Vol. 2022). Hindawi Limited. <https://doi.org/10.1155/2022/4209732>
- Ipadeola, I. O., Ojo, J. A., & Adebayo, I. G. (2025). Sentiment Analysis of Movie Reviews using Word Embeddings and Machine Learning Techniques. *LAUTECH Journal of Engineering and Technology*, 19(3). <https://doi.org/10.36108/laujet/5202.91.0341>
- Iqrom Aidil Redho, Syahril Muhammad, Jakak Muhamad Pamuji, Irawan Indra, & Febyani Yanita. (2025). *Penerapan Metode Support Vector Machine Dalam Menganalisis Sentimen Pengguna Aplikasi Sirekap 2024 Di Google Playstore*.
- Karuna, G., Anvesh, P., Singh, C. S., Reddy, K. R., Shah, P. K., & Shankar, S. S. (2023). Feasible Sentiment Analysis of Real Time Twitter Data. *E3S Web of Conferences*, 430. <https://doi.org/10.1051/e3sconf/202343001045>
- Li-kun, D., Wang, J., Tang, P., & Lv, T. (2023). *An Improved PSO Algorithm Based on Stream Computing and Multi-Swarm*. 254. <https://doi.org/10.1117/12.3004262>
- Manik, A., Nababan, E. B., & Tulus, T. (2022). Improved Support Vector Machine Performance Using Particle Swarm Optimization in Credit Risk Classification. *Jurnal Teknik Informatika (Jutif)*, 3(6), 1739–1746. <https://doi.org/10.20884/1.jutif.2022.3.6.615>
- Minggow, L. F. S., Vitianingsih, A. V., Kacung, S., Maukar, A. L., & Rusdi, J. F. (2025). Sentiment Analysis on Ajaib App Using the SVM Method. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 14(4), 551–556. <https://doi.org/10.32736/sisfokom.v14i4.2402>
- Muhamad Agung Nulhakim, Yuliant Sibaroni, & Ku Muhammad Naim Ku Khalif. (2025). Geospatial Sentiment Analysis Using Twitter Data on Natural Disasters in Indonesia with Support Vector Machine (SVM) Algorithm. *International Journal on Information and*

- Communication Technology (IJoICT)*, 10(2), 242–258.
<https://doi.org/10.21108/ijoict.v10i2.1032>
- Nugroho, B. S., & Maharani, W. (2024). Support Vector Machine and Naïve Bayes for Personality Classification Based on Social Media Posting Patterns. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1717–1731. <https://doi.org/10.47065/bits.v6i3.6411>
- Pamungkas, A. S., & Cahyono, N. (2024). Analisis Sentimen Review ChatGPT di Play Store menggunakan Support Vector Machine dan K-Nearest Neighbor. *Edumatic: Jurnal Pendidikan Informatika*, 8(1), 1–10. <https://doi.org/10.29408/edumatic.v8i1.24114>
- Papia, S. K., Khan, M. A., Habib, T., Rahman, M., & Islam, M. N. (2024). DistilRoBiLSTMFuse: an efficient hybrid deep learning approach for sentiment analysis. *PeerJ Computer Science*, 10. <https://doi.org/10.7717/peerj-cs.2349>
- Rajput, L. (2022). Sentiment Analysis using Latent Dirichlet Allocation for Aspect Term Extraction. *Journal of Computers, Mechanical and Management*, 1(2), 30–35. <https://doi.org/10.57159/gadl.jcmm.1.2.22026>
- Saputra, S., Paradise, P., & Nugraha, N. A. S. (2025). Performance Evaluation of Naïve Bayes and SVM in Sentiment Analysis of Illegal Parking Attendants. *Edumatic: Jurnal Pendidikan Informatika*, 9(2), 532–541. <https://doi.org/10.29408/edumatic.v9i2.30714>
- Sathya, A. (2023). Evaluating Sentiment Classification to Specify Polarity by Lexicon-Based and Machine Learning Approaches for COVID-19 Twitter Data Sets. *Journal of Advanced Applied Scientific Research A. Sathya et al, JOAASR*, 5(4), 12–27. <http://www.joaasr.com>
- Shami, T. M., El-Saleh, A. A., Alswaitti, M., Al-Tashi, Q., Summakieh, M. A., & Mirjalili, S. (2022). Particle Swarm Optimization: A Comprehensive Survey. *IEEE Access*, 10, 10031–10061. <https://doi.org/10.1109/ACCESS.2022.3142859>
- Song, P., Chen, L., Cai, K., Xiong, Y., & Gong, T. (2023). Analog Circuit Diagnosis Based on Support Vector Machine With Parameter Optimization by Improved NKCGWO. *Analog Integrated Circuits and Signal Processing*, 119(3), 497–510. <https://doi.org/10.1007/s10470-023-02194-4>
- Supriyatna, B. L., & Putri, F. P. (2024). Optimized support vector machine for sentiment analysis of game reviews. *International Journal of Informatics and Communication Technology*, 13(3), 344–353. <https://doi.org/10.11591/ijict.v13i3.pp344-353>
- Sutranggono, A. N., & Imah, E. M. (2023). Tweets Emotions Analysis of Community Activities Restriction as COVID-19 Policy in Indonesia Using Support Vector Machine. In *CommIT Journal* (Vol. 17, Number 1).
- Syachrul, A.-M., Roba, R., Lailiyah, S., Yusnita, A., & Widya, S. (n.d.). Application of Naive Bayes Algorithm for Analysis of User Reviews on Mobile Legends Game: Bang Bang. *Journal of Information and Technology Accredited Sinta*, 4.
- Syahrul, E., & Fatharani, D. (2025). HYBRID SENTIMENT ANALYSIS OF MAXIM APP USERS USING SUPPORT VECTOR MACHINE AND LEXICON-BASED APPROACH. *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(3S1). <https://doi.org/10.23960/jitet.v13i3S1.8148>
- Untoro, M. C., & Farhan, M. (2025). Sentiment Analysis of Public Opinion on BAWASLU Using Random Forest and Particle Swarm Optimization. *Scientific Journal of Informatics*, 12(1), 171–182. <https://doi.org/10.15294/sji.v12i1.22234>
- Vidyashree, K. P., & Rajendra, A. B. (2023). An Improvised Sentiment Analysis Model on Twitter Data Using Stochastic Gradient Descent (SGD) Optimization Algorithm in Stochastic Gate Neural Network (SGNN). *SN Computer Science*, 4(2), 190. <https://doi.org/10.1007/s42979-022-01607-x>
- Wang, Y., Han, J., & Zhang, T. (2023). A Relief-PGS Algorithm for Feature Selection and Data Classification. *Intelligent Data Analysis*, 27(2), 399–415. <https://doi.org/10.3233/ida-216493>