OPTIMALISASI ANALISIS SENTIMEN DAN KLASIFIKASI FILM VINA: SEBELUM 7 HARI MENGGUNAKAN *SUPPORT VECTOR MACHINE* DAN *PARTICLE SWARM OPTIMIZATION*

Luthfi Arie Zulfikri[1], Rizqullah Abiyyu Hade[2]

[1,2]Informatika, Fakultas Teknik, Universitas Jenderal Soedirman, Indonesia
NIM : [1]H1D022061, [2]H1D022091
Email: [1]luthfi.zulfikri@mhs.unsoed.ac.id, [2]rizqullah.hade@mhs.unsoed.ac.id

(Artikel dikirimkan tanggal : dd mmm yyyy)

Abstrak

Perkembangan media sosial memungkinkan pengguna untuk mengekspresikan pendapat dan sentimen, terutama melalui platform seperti X (sebelumnya dikenal sebagai Twitter). Salah satu topik yang menarik perhatian adalah opini publik Indonesia terhadap film Vina: Sebelum 7 Hari. Penelitian ini bertujuan untuk menganalisis sentimen terkait film tersebut menggunakan metode *support vector machine* (SVM). SVM dikenal efektif dalam klasifikasi sentimen, namun performanya sangat dipengaruhi oleh pemilihan fungsi *kernel*, parameter γ, dan parameter regularisasi C. Untuk mengoptimalkan parameter-parameter tersebut, penelitian ini menerapkan *particle swarm optimization* (PSO), yang bertugas mencari kombinasi nilai terbaik dari parameter γ dan C. Hasil analisis menunjukkan bahwa implementasi SVM standar mampu mencapai akurasi klasifikasi sebesar 78,55%. Namun, setelah optimasi parameter menggunakan algoritma PSO, akurasi meningkat secara signifikan menjadi 83,04%. Temuan ini membuktikan bahwa pendekatan SVM yang dimodifikasi dengan PSO lebih efektif dalam membedakan sentimen positif dan negatif pada data opini publik Indonesia terhadap film Vina: Sebelum 7 Hari dari platform X.

Kata kunci: *Analisis Sentimen, Klasifikasi, Particle Swarm Optimization, Support Vector Machine, Twitter*

OPTIMIZING SENTIMENT ANALYSIS AND CLASSIFICATION OF THE FILM VINA: SEBELUM 7 HARI USING SUPPORT VECTOR MACHINE AND PARTICLE SWARM OPTIMIZATION

Abstract

The development of social media allows users to express opinions and sentiments, especially through platforms such as X (formerly known as Twitter). One topic that has attracted attention is the Indonesian public's opinion on the film Vina: Sebelum 7 Hari. This research aims to analyze sentiment related to the film using the support vector machine (SVM) method. SVM is known to be effective in sentiment classification, but its performance is greatly influenced by the selection of kernel function, parameter γ, and regularization parameter C. To optimize these parameters, this study applies particle swarm optimization (PSO), which is tasked with finding the best value combination of parameters γ and C. The analysis showed that the standard SVM implementation was able to achieve a classification accuracy of 78.55%. However, after parameter optimization using the PSO algorithm, the accuracy increased significantly to 83.04%. This finding proves that the SVM approach modified with PSO is more effective in distinguishing positive and negative sentiments on Indonesian public opinion data towards the film Vina: Sebelum 7 Hari from platform X.

Keywords: *Classification, Particle Swarm Optimization, Sentiment Analysis, Support Vector Machine, Twitter*

1. PENDAHULUAN

Internet telah menjadi bagian penting dalam kehidupan sehari-hari, yang menghubungkan banyak orang melalui media sosial untuk berbagi emosi dan pendapat di berbagai platform. Emosi ini biasanya berupa sentimen atau evaluasi terhadap produk dan layanan tertentu [1]. Dalam konteks film, ulasan yang diberikan oleh penonton dapat memberikan gambaran mendalam tentang kualitas sebuah film sekaligus membantu orang lain menentukan apakah film tersebut layak untuk ditonton [2].

X, yang sebelumnya bernama Twitter merupakan salah satu platform media sosial yang

populer dan banyak digunakan untuk berbagi opini, pemikiran, serta pandangan secara bebas di ruang publik [3]. Topik yang dibahas di X sangat beragam, mulai dari isu politik, kesadaran sosial, konflik perang, ulasan film, sistem pendidikan, hingga tanggapan terhadap produk baru maupun yang sudah ada. Selain itu, data dari X yang tersedia secara publik menjadi sumber informasi berharga bagi para peneliti untuk menganalisis opini pengguna [4].

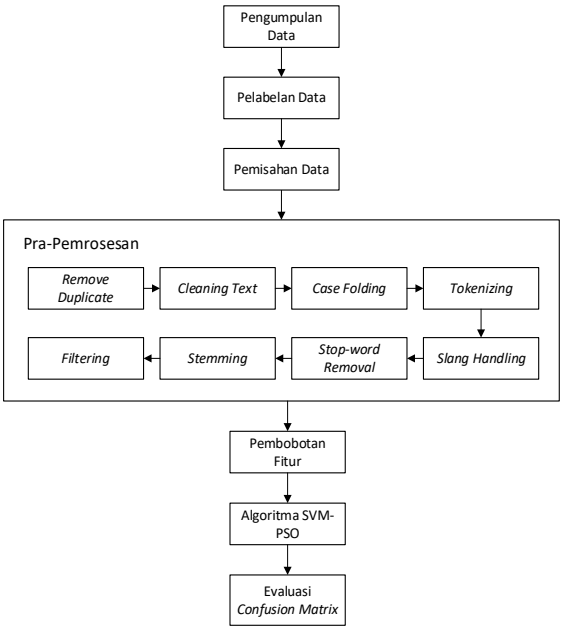
Analisis sentimen merupakan teknik yang digunakan untuk mengolah teks dan mengidentifikasi isi opini yang terkandung dari teks tersebut [5]. Tujuan utama dari analisis sentimen adalah menentukan apakah suatu teks menyampaikan opini positif, negatif, atau netral [6]. Umumnya, proses klasifikasi dilakukan dengan menggunakan metode pembelajaran mesin yang diawasi seperti pengklasifikasi *support vector machine* (SVM), *naïve bayes classifier, dan k-nearest neighbor* (KNN). *Support vector machine* (SVM) mampu memberikan akurasi yang lebih tinggi dibandingkan metode lainnya [7], [8], [9], [10].

Support vector machine (SVM) adalah model klasifikasi yang menggunakan pendekatan pembelajaran terawasi [11]. Kinerja SVM sangat bergantung pada fungsi *kernel*, parameter gama *(*γ), dan parameter regularisasi C, sehingga pemilihan nilai γ dan C yang tepat menjadi sangat penting untuk meningkatkan akurasi klasifikasi. Untuk mengoptimalkan parameter tersebut, metode *particle swarm optimization* (PSO) telah diusulkan sebagai solusi untuk menentukan batasan optimal dalam SVM [11], [12].

SVM yang dimodifikasi dengan PSO pada tahap pembelajaran dirancang untuk meminimalkan kesalahan pelatihan dengan mengacu pada prinsip *structural risk minimization* (SRM). Ketika terjadi peningkatan kesalahan pelatihan, PSO digunakan untuk mengontrol dan mencari parameter dengan error terkecil. Dengan demikian, diperlukan proses untuk menemukan parameter optimal. Setelah parameter optimal ditemukan, parameter ini diterapkan pada model SVM yang dilatih ulang [12]. Penelitian ini memanfaatkan analisis sentimen untuk mengevaluasi opini pengguna tentang film Vina: Sebelum 7 Hari berdasarkan ulasan di platform X, dengan algoritma SVM yang dioptimalkan menggunakan PSO.

2. METODE

Penelitian ini berfokus pada opini publik pengguna X terhadap film Vina: Sebelum 7 Hari. Data yang digunakan berupa twit berbahasa Indonesia. Langkah-langkah yang dilakukan dalam penelitian ini dijelaskan secara rinci pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Pada tahap ini, data twit dari X akan dikumpulkan menggunakan API X yang diakses melalui bahasa pemrograman Python. Data yang dikumpulkan berupa twit berbahasa Indonesia yang menyebutkan film Vina: Sebelum 7 Hari, dengan pencarian dilakukan menggunakan berbagai kata kunci terkait.

2.2. Pelabelan Data

Setelah data berhasil dikumpulkan, langkah selanjutnya adalah mempersiapkannya untuk tahap pelatihan. Data ini disebut sebagai data pelatihan dan perlu diberi label terlebih dahulu. Pada tahap ini, pelabelan dilakukan secara manual berdasarkan kriteria yang telah ditetapkan, dengan setiap data dalam *training set* diklasifikasikan ke dalam dua kategori, yaitu positif atau negatif.

2.3. Pemisahan Data

Pemisahan data adalah proses membagi data menjadi dua atau lebih *subset*. Umumnya, data dipisahkan menjadi dua bagian, satu bagian digunakan untuk mengevaluasi atau menguji dan bagian lainnya digunakan untuk melatih model [13]. Dalam penelitian ini, rasio pembagian data sebesar 80:20 dipilih karena secara umum telah terbukti menghasilkan performa yang optimal di berbagai kasus [14].

2.4. Pra-Pemrosesan Data

Model bahasa X memiliki karakteristik unik yang membedakannya dari model lainnya. Penyesuaian pada ruang fitur dilakukan secara sengaja dengan memanfaatkan ciri khas tersebut. Tahapan pra-pemrosesan data dirancang untuk

memanfaatkan sifat unik model bahasa X, sehingga meningkatkan kinerja pada tahap-tahap pengembangan selanjutnya [15]. Langkah-langkah pra-pemrosesan yang diterapkan meliputi:

2.4.1. *Remove Duplicate*

Penghapusan duplikat dilakukan untuk mengeliminasi teks yang identik. Langkah ini bertujuan menghindari perlambatan proses analisis model yang disebabkan oleh data redundan [16].

2.4.2. *Cleaning Text*

Pada tahap ini menghapus elemen-elemen yang tidak relevan, seperti karakter khusus, simbol, nama pengguna (*@username*), URL, dan tanda baca yang tidak diperlukan [17].

2.4.3. *Case Folding*

Pada tahap ini, semua huruf dalam teks diubah menjadi huruf kecil untuk menyamakan format penulisan [18].

2.4.4. *Tokenizing*

Proses *tokenizing* memecah teks menjadi unit-unit kecil yang disebut token. Token ini dapat berupa simbol, kata, frasa, atau kata kunci tertentu [19].

2.4.5. *Slang Handling*

Proses ini dilakukan dengan mengganti kata-kata informal atau tidak baku dengan bentuk kata yang lebih resmi dan sesuai kaidah bahasa [20].

2.4.6. *Stop-word Removal*

Tahap ini menghilangkan kata-kata yang dianggap tidak memiliki makna penting, seperti "yang," "dengan," atau "maka" [21].

2.4.7. *Stemming*

Stemming mengubah kata-kata turunan menjadi bentuk dasarnya. *Library* Sastrawi digunakan dalam proses ini karena mampu meminimalkan kesalahan *stemming*, baik *over-stemming* maupun *under-stemming*, serta menawarkan efisiensi waktu pemrosesan yang lebih baik [22], [23].

2.4.8. *Filtering*

Filtering dilakukan menghapus kata-kata berdasarkan panjang karakter tertentu. Dalam hal ini, hanya kata dengan panjang 4–25 karakter yang dipertahankan [16].

2.5. Pembobotan Fitur

TF-IDF adalah metode yang sering digunakan untuk memberikan bobot pada kata dalam teks. Metode ini bekerja dengan menghitung seberapa sering sebuah kata muncul dalam satu dokumen (TF atau *Term Frequency*) dan membandingkannya dengan seberapa jarang kata tersebut muncul di seluruh dokumen lainnya (IDF atau Inverse *Document Frequency*). Dengan kombinasi TF dan IDF, TF-IDF membantu menentukan seberapa penting sebuah kata dalam sebuah dokumen tertentu [24]. Rumus TF-IDF ditunjukkan pada Persamaan 1.

$$TF - IDF = TF \times IDF \quad (1)$$

2.6. Pemodelan Klasifikasi

2.6.1. *Support Vector Machine* (SVM)

Support vector machine (SVM) adalah model klasifikasi yang dirancang menggunakan mekanisme pembelajaran terawasi [11]. SVM lebih berfokus pada minimalisasi risiko struktural (*Structural Risk Minimization*) dibandingkan pendekatan minimalisasi kesalahan empiris yang biasanya digunakan oleh *Neural Networks*. Model ini mampu menangani tugas klasifikasi baik untuk data linier maupun *nonlinier*. Kemampuan ini dicapai dengan memproyeksikan data pelatihan ke dalam ruang berdimensi lebih tinggi menggunakan teknik yang disebut *kernel trick*, yang memungkinkan pemetaan secara *nonlinier* [25].

Dalam ruang dimensi yang lebih tinggi tersebut, SVM bertugas menemukan *hyperplane* (bidang pemisah) yang paling optimal untuk memisahkan berbagai kategori data pelatihan. Proses pembelajaran SVM bertujuan menentukan *hyperplane* linier yang optimal di ruang yang telah ditransformasi tersebut [25]. Dengan menerapkan minimalisasi risiko struktural, SVM berhasil mengatasi masalah yang sering muncul pada pendekatan lain [11]. Fungsi keluaran SVM dirumuskan pada Persamaan 2.

$$f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^{'})K(x, z) + b_i \quad (2)$$

2.6.2. *Particle Swarm Optimization* (PSO)

Particle swarm optimization (PSO) adalah algoritma stokastik yang bersifat iteratif dan tidak deterministik. PSO menggunakan pendekatan kecerdasan kawanan, di mana sekelompok individu yang disebut partikel bekerja sama untuk mengeksplorasi solusi dalam ruang pencarian masalah tertentu. Ketika partikel-partikel ini bergerak bersama, mereka membentuk kawanan yang saling berbagi informasi [26].

Dalam PSO, setiap partikel memperbaiki posisinya dengan belajar dari pengalaman terbaiknya sendiri (personal *best* atau pbest) serta solusi terbaik secara global (*global best* atau gbest). Posisi gbest mencerminkan solusi yang paling optimal di antara seluruh partikel dalam populasi [26]. Evaluasi kinerja setiap partikel dilakukan berdasarkan nilai

fungsi objektif untuk menentukan tingkat kecocokan solusi [27].

Proses pencarian solusi dilakukan secara iteratif, di mana partikel memperbarui posisi mereka di ruang pencarian dengan kecepatan tertentu. Kecepatan dan posisi setiap partikel dihitung berulang kali menggunakan rumus tertentu hingga solusi optimal ditemukan [27]. Iterasi kecepatan dan posisi partikel ditunjukkan masing-masing pada Persamaan 3 dan 4.

$$v_{id}^{t+1} = \omega \cdot v_{id}^{t} + c_1 + r_1 \cdot (p_{id} - x_{id}^{t}) + c_2 + r_2 \cdot (p_{gd} - x_{id}^{t}), \quad (3)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (4)$$

2.7. Evaluasi

Evaluasi teknik klasifikasi dokumen dapat dilakukan dengan mengukur akurasi berdasarkan statistik seperti *true positives* (TP), *true negatives* (TN), *false positives* (FP), dan *false negatives* (FN). Akurasi ini dihitung menggunakan metrik yang memanfaatkan *confussion matrix* sebagai alat analisis utama [28]. Persamaan akurasi ditunjukkan pada Persamaan 5.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

3. HASIL

3.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berisi sentimen terhadap film Vina: Sebelum 7 Hari, yang dikumpulkan dari X (Twitter), salah satu platform media sosial terpopuler di Indonesia [28]. Data diperoleh dengan menggunakan kata kunci "film vina" dalam rentang waktu antara Mei hingga Juni 2024. Dari proses ini, berhasil dikumpulkan sebanyak 2012 data twit yang relevan.

Pengumpulan data dilakukan menggunakan Google Collaboratory dengan memanfaatkan operasi *crawling*. Google Collaboratory dihubungkan ke X melalui token API yang diperoleh dari hasil inspeksi pada platform X menggunakan Google Chrome. Setelah data terkumpul, semua ulasan disimpan dalam fail CSV untuk keperluan analisis lebih lanjut.

3.2. Pelabelan Data

Sebelum melakukan pra-pemrosesan, data terlebih dahulu diberi label untuk menentukan polaritas dari setiap objek data. Proses pelabelan ini melibatkan pemilihan satu kolom yang relevan dan menghilangkan empat belas kolom lainnya. Hasil pelabelan manual menghasilkan total 1265 twit dengan sentimen negatif, 747 twit dengan sentimen positif, sehingga terdapat 2012 data yang siap dianalisis. Tabel 1 menunjukkan contoh hasil dari proses pelabelan data.

Tabel 1. Hasil Pelabelan Data

No	Twit	Polaritas
1	film vina bagus bgtttt plis	Positif
2	Besok di ajakin nonton film Vina. Katanya sih seru ngikut aja lah mumpung libur.	Positif
3	film Vina adalah film terbodoh yg pernah dibuat karena bisa membuka kembali ingatan keluarga korban tentang kejadian tersebut.	Negatif
4	@moviemenfes Gausah nonton vina. Yg ngide itu di film in jelek otaknya	Negatif
...	...	...
2012	Vina adalah film sampah yang dibuat dengan ketiadaan nurani produser dan sutradaranya yang ingin menularkan ketiadaan nurani itu ke penonton haus sensasi pliss jangan tonton biar gak ketularan bobrok hati biar itu duo begundal stop bertengik-tengik diri.	Negatif

3.3. Pemisahan Data

Dalam penelitian ini, *dataset* dibagi menjadi data latih dan data uji dengan rasio 80:20, 80% dari *dataset* digunakan untuk pelatihan dan 20% untuk pengujian. Pembagian data dilakukan secara acak dengan menggunakan parameter *random_state* bernilai 42 untuk memastikan konsistensi hasil. Berdasarkan hasil pembagian, jumlah data latih adalah 1609, sedangkan jumlah data uji adalah 403.

3.4. Pra-Pemrosesan Data

Pra-pemrosesan adalah tahap penting untuk membersihkan data dari kata-kata atau twit yang tidak relevan serta kata-kata yang tidak memiliki makna signifikan. Proses ini dilakukan sesuai dengan konten data yang diperoleh dari pengambilan atau *crawling* data X. Tahapan pra-pemrosesan terdiri dari beberapa langkah yang dilakukan secara berurutan, yaitu: 1) *Remove Duplicate*, 2) *Cleaning Text*, 3) *Case Folding*, 4) *Tokenizing*, 5) *Slang Handling*, 6) *Stop-word Removal*, 7) *Stemming*, dan 8) *Filtering*.

Contoh twit mentah yang dapat ditemukan adalah "Film Vina: Sebelum 7 Hari Cetak Rekor di Hari Pertama Tayang https://t.co/g3y5Y5xQHj." Twit ini kemudian diproses melalui beberapa langkah hingga menjadi teks bersih yang ditunjukkan pada Tabel 2.

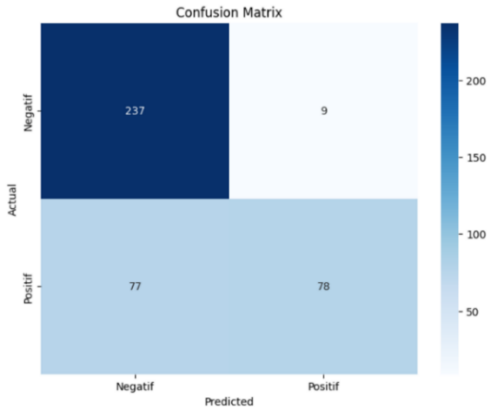
Tabel 2. Proses Pra-Pemrosesan Data

No	Tahap	Sebelum	Sesudah
1	*Cleaning Text*	Film Vina: Sebelum 7 Hari Cetak Rekor di Hari Pertama Tayang https://t.co/g3y5Y5xQHj	Film Vina Sebelum Hari Cetak Rekor di Hari Pertama Tayang
2	*Case Folding*	Film Vina Sebelum Hari Cetak Rekor di Hari Pertama Tayang	film vina sebelum hari cetak rekor di hari pertama tayang
3	*Tokenizing*	film vina sebelum hari cetak rekor di hari pertama tayang	['film', 'vina', 'sebelum', 'hari', 'cetak', 'rekor', 'di', 'hari', 'pertama', 'tayang']
4	*Slang Handling*	['film', 'vina', 'sebelum', 'hari', 'cetak', 'rekor', 'di', 'hari', 'pertama', 'tayang']	['film', 'vina', 'sebelum', 'hari', 'cetak', 'rekor', 'di', 'hari', 'pertama', 'tayang']
5	*Stop-word Removal*	['film', 'vina', 'sebelum', 'hari', 'cetak', 'rekor', 'di', 'hari', 'pertama', 'tayang']	['film', 'vina', 'sebelum', 'cetak', 'rekor', 'tayang']
6	*Stemming*	['film', 'vina', 'sebelum', 'cetak', 'rekor', 'tayang']	['film', 'vina', 'belum', 'cetak', 'rekor', 'tayang']
7	*Filtering*	['film', 'vina', 'belum', 'cetak', 'rekor', 'tayang']	['film', 'vina', 'belum', 'cetak', 'rekor', 'tayang']

3.5. Pemodelan Klasifikasi

3.5.1. Implementasi *Support Vector Machine* (SVM)

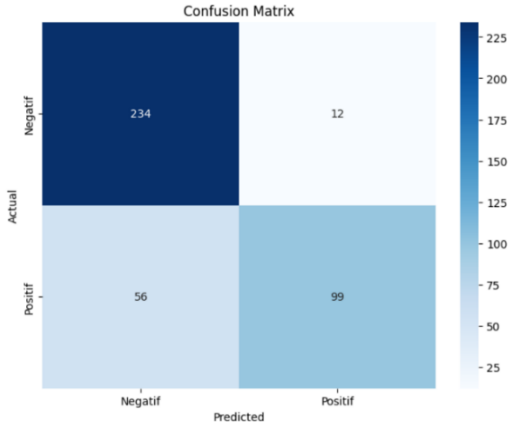
Berdasarkan hasil eksperimen yang dilakukan menggunakan data twit di platform X terkait film Vina: Sebelum 7 Hari, algoritma Support Vector Machine (SVM) diuji pada 2012 data ulasan pengguna. Pengujian ini menghasilkan nilai akurasi yang dihitung menggunakan *confusion matrix*, seperti yang ditampilkan pada Gambar 2.



Gambar 2. *Confusion Matrix* SVM

3.5.2. Implementasi PSO pada *Support Vector Machine* (SVM)

Setelah mengukur akurasi awal dari model s*upport vector machine* (SVM), langkah selanjutnya adalah menerapkan algoritma *particle swarm optimization* (PSO) untuk meningkatkan performa model. Algoritma PSO digunakan untuk menyesuaikan parameter C dan gama (γ) pada SVM, sekaligus melakukan seleksi atribut dan fitur guna mencapai akurasi yang lebih optimal. Dalam eksperimen ini, PSO diatur dengan parameter berupa ukuran populasi sebanyak 20, jumlah generasi maksimum 50, batas nilai C antara 0,1 hingga 100, serta batas nilai γ antara 0,0001 hingga 1. Data yang digunakan berasal dari ulasan pengguna mengenai film Vina: Sebelum 7 Hari sebanyak 2012 data. Hasil eksperimen menunjukkan bahwa penerapan PSO secara signifikan meningkatkan akurasi model SVM, yang terlihat pada *confusion matrix* yang ditampilkan dalam Gambar 3.



Gambar 3. *Confusion Matrix* SVM dengan PSO

Berdasarkan hasil klasifikasi, sebanyak 234 data berhasil diklasifikasikan sebagai Positif (*True Positive*), sementara 12 data keliru diklasifikasikan sebagai Positif (*False Positive*). Untuk data yang diklasifikasikan sebagai Negatif, terdapat 99 data yang benar (*True Negative*) dan 56 data yang keliru diklasifikasikan sebagai Negatif (*False Negative*). Secara keseluruhan, tingkat akurasi model SVM yang dimodifikasi dengan PSO mencapai 83,04%.

3.6. Perbandingan SVM dan SVM-PSO

Berdasarkan hasil eksperimen menggunakan data twit berbahasa di platform X terkait film Vina: Sebelum 7 Hari, algoritma *support vector machine* (SVM) dan SVM yang dioptimasi dengan *particle swarm optimization* (PSO) diuji pada *dataset* yang telah difilter, terdiri dari 2012 data twit. Nilai akurasi, *recall*, dan *precision* dari masing-masing algoritma ditampilkan dalam Tabel 3 dan Tabel 4.

Tabel 3. Performa Algoritma SVM

Class	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	75.48%	96.34%	84.64%	246
Positif	89.66%	50.32%	64.46%	155
Akurasi			78.55%	401

Tabel 4. Performa Algoritma SVM dengan PSO

Class	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	80.69%	95.12%	87.31%	246
Positif	89.19%	63.87%	74.44%	155
Akurasi			83.04%	401

Hasil perbandingan menunjukkan bahwa algoritma SVM yang dimodifikasi dengan PSO menghasilkan akurasi lebih tinggi dibandingkan SVM standar, yaitu 83.04% dibandingkan 78.55%. Selain itu, pada kelas negatif, algoritma SVM yang dioptimasi dengan PSO menunjukkan peningkatan nilai *precision* dari 75.48% menjadi 80.69% dan *recall* dari 96.34% menjadi 95.12%. Sedangkan pada kelas positif, terdapat peningkatan *precision* dari 89.66% menjadi 89.19% dan recall dari 50.32% menjadi 63.87%.

Secara keseluruhan, algoritma SVM yang dioptimasi dengan PSO menunjukkan peningkatan kinerja yang signifikan, terutama dalam meningkatkan nilai *recall* pada kelas positif, yang menjadi salah satu kelemahan SVM standar. Dengan demikian, kombinasi SVM dan PSO terbukti lebih unggul dan efektif untuk menyelesaikan permasalahan analisis sentimen pada twit terkait film Vina: Sebelum 7 Hari di platform X.

4. KESIMPULAN

Hasil penelitian ini menunjukkan bahwa pengoptimalan algoritma *Support Vector Machine* (SVM) menggunakan *Particle Swarm Optimization* (PSO) berhasil meningkatkan akurasi menjadi 83,04%. Sebagai perbandingan, penerapan algoritma SVM tanpa optimasi hanya mencapai akurasi 78,55%. Temuan ini mengindikasikan bahwa kombinasi SVM dan PSO memberikan kinerja klasifikasi yang lebih baik dibandingkan SVM saja. Algoritma tersebut terbukti efektif untuk mengklasifikasikan twit dengan sentimen 'Positif' atau 'Negatif' terkait film Vina: Sebelum 7 Hari di platform media sosial X.

DAFTAR PUSTAKA

[1] P. P. Rokade, P. V. R. D. P. Rao, dan A. K. Devarakonda, "Forecasting movie rating using k-nearest neighbor based collaborative filtering," *International Journal of Electrical and Computer Engineering*, vol. 12, no. 6, hlm. 6506–6512, Des 2022, doi: 10.11591/ijece.v12i6.pp6506-6512.

[2] Z. Shaukat, A. A. Zulfiqar, C. Xiao, M. Azeem, dan T. Mahmood, "Sentiment analysis on IMDB using lexicon and neural networks," *SN Appl Sci*, vol. 2, no. 2, hlm. 1–10, Feb 2020, doi: 10.1007/S42452-019-1926-X/TABLES/1.

[3] Y. M. Tun dan M. Khaing, "A large-scale sentiment analysis using political tweets," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 6, hlm. 6913, Des 2023, doi: 10.11591/ijece.v13i6.pp6913-6925.

[4] V. Ramanathan, T. Meyyappan, dan S. M. Thamarai, "Sentiment Analysis: An Approach for Analysing Tamil Movie Reviews Using Tamil Tweets," dalam *Recent Advances in Mathematical Research and Computer Science Vol. 3*, Book Publisher International (a part of SCIENCEDOMAIN International), 2021, hlm. 28–39. doi: 10.9734/bpi/ramrcs/v3/4845F.

[5] P. Mehta dan S. Pandya, "A review on sentiment analysis methodologies, practices and applications," *International Journal of Scientific and Technology Research*, vol. 9, no. 2, hlm. 601–609, 2020.

[6] K. Ullah, A. Rashad, M. Khan, Y. Ghadi, H. Aljuaid, dan Z. Nawaz, "A Deep Neural Network-Based Approach for Sentiment Analysis of Movie Reviews," *Complexity*, vol. 2022, no. 1, hlm. 5217491, Jan 2022, doi: 10.1155/2022/5217491.

[7] M. D. H. Jasy, S. Al Hasan, M. I. K. Sagor, A. Noman, dan J. M. Ji, "A Performance Evaluation of Sentiment Classification Applying SVM, KNN, and Naive Bayes," dalam *2021 International Conference on Computing, Networking, Telecommunications & Engineering Sciences Applications (CoNTESA)*, IEEE, Des 2021, hlm. 56–60. doi: 10.1109/CoNTESA52813.2021.9657115.

[8] A. Salma dan W. Silfianti, "Sentiment Analysis of User Reviews on Covid-19 Information Applications using Naive Bayes Classifier, Support Vector Machine, and K-Nearest Neighbor," *International Research Journal of Advanced Engineering and Science*, vol. 6, no. 4, hlm. 158–162, 2021.

[9] J. W. Iskandar dan Y. Nataliani, "Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 6, hlm. 1120–1126, Des 2021, doi: 10.29207/resti.v5i6.3588.

[10] F. S. Pamungkas dan I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNN untuk Studi Tanggapan Masyarakat Indonesia Terhadap Pandemi Covid-19 pada Media Sosial Twitter," dalam *PRISMA, Prosiding Seminar Nasional Matematika*, 2021, hlm. 628–634.

[11] D. J. Kalita, V. P. Singh, dan V. Kumar, "SVM Hyper-Parameters Optimization using Multi-PSO for Intrusion Detection," dalam *Social Networking and Computational Intelligence: Proceedings of SCI-2018*, 2020, hlm. 227–241. doi: 10.1007/978-981-15-2071-6_19.

[12] S. Samantaray, A. Sahoo, dan A. Agnihotri, "Prediction of Flood Discharge Using Hybrid PSO-SVM Algorithm in Barak River

Basin," *MethodsX*, vol. 10, hlm. 102060, 2023, doi: 10.1016/j.mex.2023.102060.

[13] A. Erfina dan M. R. N. Ramdani Alamsyah, "Implementation of Naive Bayes classification algorithm for Twitter user sentiment analysis on ChatGPT using Python programming language," *Data and Metadata*, vol. 2, hlm. 45, Mei 2023, doi: 10.56294/dm202345.

[14] H. Bichri, A. Chergui, dan M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," *International Journal of Advanced Computer Science and Applications*, vol. 15, no. 2, hlm. 331–339, 2024, doi: 10.14569/IJACSA.2024.0150235.

[15] V. Gupta dan Dr. P. Rattan, "Improving Twitter Sentiment Analysis Efficiency with SVM-PSO Classification and EFWS Heuristic," *Procedia Comput Sci*, vol. 230, hlm. 698–715, 2023, doi: 10.1016/j.procs.2023.12.125.

[16] A. Mustopa, Hermanto, Anna, E. B. Pratama, A. Hendini, dan D. Risdiansyah, "Analysis of User Reviews for the PeduliLindungi Application on Google Play Using the Support Vector Machine and Naive Bayes Algorithm Based on Particle Swarm Optimization," dalam *2020 Fifth International Conference on Informatics and Computing (ICIC)*, IEEE, Nov 2020, hlm. 1–7. doi: 10.1109/ICIC50835.2020.9288655.

[17] A. Winanto dan C. Budihartanti, "Comparison of the Accuracy of Sentiment Analysis on the Twitter of the DKI Jakarta Provincial Government during the COVID-19 Vaccine Time," *Journal of Computer Science and Engineering (JCSE)*, vol. 3, no. 1, hlm. 14–27, Feb 2022, doi: 10.36596/jcse.v3i1.249.

[18] R. Novendri, A. S. Callista, D. N. Pratama, dan C. E. Puspita, "Sentiment Analysis of YouTube Movie Trailer Comments Using Naïve Bayes," *Bulletin of Computer Science and Electrical Engineering*, vol. 1, no. 1, hlm. 26–32, Jun 2020, doi: 10.25008/bcsee.v1i1.5.

[19] E. O. Omuya, G. Okeyo, dan M. Kimwele, "Sentiment analysis on social media tweets using dimensionality reduction and natural language processing," *Engineering Reports*, vol. 5, no. 3, Mar 2023, doi: 10.1002/eng2.12579.

[20] M. A. Ayu dan A. H. Muhendra, "Preprocessing of Slang Words for Sentiment Analysis on Public Perceptions in Twitter," dalam *Advances in Sentiment Analysis*, J. Li, Ed., Rijeka: IntechOpen, 2023, 2. doi: 10.5772/intechopen.113725.

[21] A. R. Nur Hasifah, M. Nur Amirah, T. O. Nur Saidatul Sa'adiah, M. Muhammad Firdaus, dan M. Muhammad Firdaus, "Amazon product sentiment analysis using RapidMiner," *Applied Mathematics and Computational Intelligence*, vol. 11, no. 1, hlm. 336–349, 2022, Diakses: 8 Juni 2024. [Daring]. Tersedia pada: http://dspace.unimap.edu.my:80/handle/123456789/77661

[22] D. D. Saputra *dkk.*, "Optimization Sentiments of Analysis from Tweets in myXLCare using Naïve Bayes Algorithm and Synthetic Minority Over Sampling Technique Method," *J Phys Conf Ser*, vol. 1471, no. 1, hlm. 012014, Feb 2020, doi: 10.1088/1742-6596/1471/1/012014.

[23] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, dan H. A. Gozali, "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi," *IOP Conf Ser Mater Sci Eng*, vol. 874, no. 1, hlm. 012017, Jun 2020, doi: 10.1088/1757-899X/874/1/012017.

[24] Fatihah Rahmadayana dan Yuliant Sibaroni, "Sentiment Analysis of Work from Home Activity using SVM with Randomized Search Optimization," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 5, hlm. 936–942, Okt 2021, doi: 10.29207/resti.v5i5.3457.

[25] N. S. Bajaj, A. D. Patange, R. Jegadeeshwaran, S. S. Pardeshi, K. A. Kulkarni, dan R. S. Ghatpande, "Application of metaheuristic optimization based support vector machine for milling cutter health monitoring," *Intelligent Systems with Applications*, vol. 18, hlm. 200196, Mei 2023, doi: 10.1016/j.iswa.2023.200196.

[26] M. Jain, V. Saihjpal, N. Singh, dan S. B. Singh, "An Overview of Variants and Advancements of PSO Algorithm," *Applied Sciences*, vol. 12, no. 17, hlm. 8392, Agu 2022, doi: 10.3390/app12178392.

[27] S. Kumar, M. Badruddin Khan, M. Hoque Abul Hasanat, A. Khader Jilani Saudagar, A. AlTameem, dan M. AlKhathami, "Sigmoidal Particle Swarm Optimization for Twitter Sentiment Analysis," *Computers, Materials & Continua*, vol. 74, no. 1, hlm. 897–914, 2023, doi: 10.32604/cmc.2023.031867.

[28] M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, dan R. Budiarto, "Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking," *IEEE Access*, vol. 8, hlm. 90847–90861, 2020, doi: 10.1109/ACCESS.2020.2994222.