

Analisis Sentimen Ulasan Aplikasi TikTok di Google Play Store Menggunakan Algoritma Support Vector Machine

Rahman Qolbi¹⁾, Abertun Sagit Sahay²⁾, Septian Geges³⁾

¹⁾²⁾³⁾ Teknik Informatika, Fakultas Teknik, Universitas Palangka Raya
Jl. Yos Sudarso, Kec. Jekan Raya, Kota Palangka Raya, Kalimantan Tengah 74874

¹⁾ rahmanqolbi@mhs.eng.upr.ac.id

²⁾ abertun@it.upr.ac.id

³⁾ septian.geges@it.upr.ac.id

Abstrak

TikTok adalah sebuah aplikasi jejaring sosial dan platform *video music* di mana pengguna bisa membuat, mengedit, dan berbagi klip video pendek lengkap dengan *filter* dan disertai musik sebagai pendukung. Di *Google Play Store* aplikasi *TikTok* telah diunduh sebanyak 500 juta lebih pengguna per Februari 2024. Dengan banyaknya pengguna *TikTok* terdapat ulasan yang beragam, baik yang positif maupun negatif. Dari ulasan tersebut dapat diketahui bagaimana *sentimen* pengguna terhadap aplikasi *TikTok*. *Analisis sentimen* menjadi penting karena memungkinkan pihak *TikTok* untuk memahami opini dan persepsi pengguna terhadap produk atau layanan mereka. *Analisis sentimen* dilakukan menggunakan algoritma *Support Vector Machine* karena mempunyai kemampuan yang lebih baik dibandingkan algoritma lainnya. Data diolah terlebih dahulu pada tahap *preprocessing* kemudian diberi label dengan pendekatan *lexicon based*. *Data splitting* yang dipilih untuk *modeling* adalah 90% data latih dan 10% data uji. Model yang telah dilatih menghasilkan performa yang baik dengan *akurasi* sebesar 93.78%, *presision* positif sebesar 95.00% dan negatif sebesar 90.08%, *recall* positif sebesar 96.67% dan negatif sebesar 85.58%, nilai *f1-score* positif sebesar 95.83% dan negatif sebesar 87.77%. Hasil dari penelitian ini kemudian diimplementasikan dalam bentuk *interface web* menggunakan *library streamlit*.

Kata kunci: Analisis Sentimen, *Support Vector Machine*, TikTok, Algoritma

Abstract

TikTok is a social networking app and video music platform where users can create, edit, and share short video clips complete with filters and accompanying music. In the *Google Play Store*, the *TikTok* application has been downloaded by more than 500 million users as of February 2024. With so many *TikTok* users there are various reviews, both positive and negative. From these reviews it can be seen how user sentiment towards the *TikTok* application. Sentiment analysis is important because it allows *TikTok* to understand user opinions and perceptions of their products or services. Sentiment analysis is performed using the *Support Vector Machine* algorithm because it has better capabilities than other algorithms. The data was first processed in the preprocessing stage and then labeled with a lexicon-based approach. The data splitting chosen for modeling is 90% training data and 10% test data. The model that has been trained produces good performance with an accuracy of 93.78%, positive precision of 95.00% and negative of 90.08%, positive recall of 96.67% and negative of 85.58%, positive *f1-score* value of 95.83% and negative of 87.77%. The results of this study were then implemented in the form of a web interface using the *streamlit* library.

Keywords: Sentiment Analysis, *Support Vector Machine*, TikTok, Algorithm

1. PENDAHULUAN

Teknologi seiring waktu terus berkembang, salah satunya dalam bidang komunikasi. Komunikasi saat ini sudah sangat berkembang melalui media digital. Hal tersebut tak lepas dari akses internet yang mudah diakses, khususnya melalui *smartphone*. Saat ini *smartphone* menjadi salah satu perangkat yang banyak digunakan karena mudah digunakan di mana saja. Aplikasi yang tersedia pada *smartphone* untuk mengakses media digital pun beragam, salah satunya adalah TikTok.

TikTok adalah sebuah aplikasi jejaring sosial dan platform video musik dimana pengguna bisa membuat, mengedit, dan berbagi klip video pendek lengkap dengan filter dan disertai musik sebagai pendukung [1]. Aplikasi TikTok sendiri tersedia pada *Google Play Store* untuk pengguna *smartphone* dengan sistem operasi Android. *Google Play Store* sendiri merupakan sebuah toko daring di Android yang menyediakan aplikasi, *game*, buku, dan konten lainnya. Di *Google Play Store* aplikasi TikTok telah diunduh sebanyak 500 juta lebih pengguna per Februari 2024. Dengan banyaknya pengguna TikTok terdapat ulasan yang beragam, baik yang positif maupun negatif. Dari ulasan tersebut dapat diketahui bagaimana sentimen pengguna terhadap aplikasi TikTok.

Analisis sentimen menjadi penting karena memungkinkan pihak TikTok untuk memahami opini dan persepsi pengguna terhadap produk atau layanan mereka. Sentimen positif dapat menunjukkan kepuasan pengguna dan keberhasilan fitur tertentu, sedangkan sentimen negatif dapat mengidentifikasi area yang memerlukan perbaikan. Dengan melakukan analisis sentimen, pihak TikTok dapat mengambil langkah-langkah proaktif untuk meningkatkan pengalaman pengguna, mengatasi masalah yang sering dihadapi, dan merancang strategi pemasaran yang lebih efektif.

Untuk memulai analisis sentimen banyak hal yang perlu dilakukan, salah satunya adalah memilih metode klasifikasi yang akan dipakai [2]. Metode yang digunakan adalah algoritma *Support Vector Machine* (SVM). SVM adalah metode *learning machine* yang bertujuan untuk menemukan *hyperplane* terbaik yang memisahkan dua buah kelas pada ruang *input* (*input space*). Algoritma klasifikasi SVM menggunakan data *training* untuk membentuk model klasifikasi, model yang terbentuk digunakan sebagai prediksi kelas data baru yang belum pernah ada sebelumnya yang disebut dengan *testing data* [3].

Berdasarkan pembahasan diatas, maka melatarbelakangi peneliti untuk melakukan penelitian menggunakan algoritma *Support Vector Machine* yang berjudul “Analisis Sentimen Ulasan Aplikasi TikTok Di *Google Play Store* Menggunakan Algoritma *Support Vector Machine*”.

2. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Yang pertama adalah penelitian [4] dengan judul “Perbandingan Algoritma *Support Vector Machine* dan *Random Forest* untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022”. Pada penelitian ini dilakukan perbandingan algoritma untuk analisis sentimen terhadap kebijakan pemerintah terkait kenaikan harga BBM di Indonesia. Dari hasil penelitian tersebut didapatkan algoritma *Support Vector Machine* mendapatkan akurasi sebesar 77% dan algoritma *Random Forest* mendapatkan akurasi sebesar 76%.

Yang kedua adalah penelitian [5] dengan judul “Analisis Sentimen Menggunakan SVM Dan KNN Pada Review Drama Korea Di Mydramalist”. Pada penelitian ini dilakukan analisis sentimen review atau ulasan drama korea di situs web MyDramaList. Dari hasil penelitian tersebut didapatkan algoritma *Support Vector Machine* mendapatkan akurasi sebesar 97% dan algoritma K-Nearest Neighbors mendapatkan akurasi sebesar 95%.

Yang ketiga adalah penelitian [6] dengan judul “Analisis Sentimen Top 10 Traveler Ranked Hotel di Kota Makassar Menggunakan Algoritma *Decision Tree* dan *Support Vector Machine*”. Pada penelitian ini dilakukan analisis sentimen top 10 ranked hotel di Kota Makassar. Pada penelitian ini diterapkan *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi masalah keseimbangan data. Dari hasil penelitian tersebut didapatkan algoritma *Decision Tree* tanpa SMOTE mendapatkan akurasi sebesar 91.59% dan algoritma *Support Vector Machine*

tanpa SMOTE mendapatkan akurasi sebesar 92.08% dengan data splitting 70:30. Lalu algoritma Decision Tree dengan SMOTE mendapatkan akurasi sebesar 93.60% dan algoritma Support Vector Machine tanpa SMOTE mendapatkan akurasi sebesar 98.98% dengan data splitting 70:30.

Yang keempat adalah penelitian [7] dengan judul “Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD)”. Pada penelitian ini dilakukan analisis sentimen mengenai tanggapan masyarakat pada aplikasi Tiktok. Algoritma yang digunakan Naïve Bayes. Data yang digunakan diperoleh dari Google Play Store mulai dari rating 1 sampai dengan rating 5 dan disimpan pada file CSV. Dari hasil penelitian tersebut didapatkan menggunakan 5-Fold Cross Validation dengan 100% Term algoritma Naïve Bayes dengan seleksi fitur Categorical Proportional Difference mendapatkan nilai Accuracy sebesar 0,729947, nilai Precision sebesar 0,746854, nilai Recall sebesar 0,926118, dan nilai f-measure 0,824511.

2.2 Analisis Sentimen

Analisis sentimen merupakan sebuah metode yang digunakan untuk mengekstrak data opini, memahami serta mengolah tekstual data secara otomatis untuk melihat sentimen yang terkandung dalam sebuah opini [8].

2.3 TikTok

TikTok adalah sebuah aplikasi jejaring sosial dan platform video music dimana pengguna bisa membuat, mengedit, dan berbagi klip video pendek lengkap dengan filter dan disertai musik sebagai pendukung. [1].

2.4 Support Vector Machine

Support Vector Machine (SVM) adalah suatu teknik yang relatif baru untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi [9]. Konsep klasifikasi dengan SVM adalah mencari *hyperlane* terbaik yang memiliki fungsi sebagai pemisah dari dua kelas data [10]. *Support Vector Machine* hanya menggunakan beberapa titik data terpilih yang berkontribusi (*Support Vector*) untuk membentuk model yang akan digunakan dalam proses klasifikasi. Dalam memperoleh *hyperplane* pada SVM, dapat menggunakan persamaan berikut [11] :

$$(w \cdot x_i) + b = 0 \tag{1}$$

Di dalam data x_i , yang termasuk pada kelas -1 dapat dirumuskan seperti pada persamaan berikut:

$$(w \cdot x_i + b) \leq 1, y_i = -1 \tag{2}$$

Sedangkan data x_i yang termasuk pada kelas +1 dapat dirumuskan seperti pada persamaan berikut:

$$(w \cdot x_i + b) \geq 1, y_i = 1 \tag{3}$$

Keterangan:

w = parameter *hyperplane* yang dicari (garis yang tegak lurus antara garis *hyperplane* dan titik *support vector*)

x_i = titik data masukan *Support Vector Machine*

y_i = kelas data

b = nilai bias

2.5 K-fold Cross Validation

K-fold Cross validation adalah suatu metode validasi yang pada prinsipnya bahwa proses validasi akan dilakukan secara berulang sebanyak nilai K berdasarkan catatan data yang

dimasukkan. Dari data yang diinput tersebut akan diacak sehingga diperoleh data *testing* dan *data training* dimana pada kedua data tersebut memiliki jumlah data yang sama [12].

2.6 Confusion Matrix

Confusion matrix merupakan sebuah tabel yang memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem (prediksi) dengan hasil klasifikasi yang sebenarnya [13]. Dari confusion matrix dapat diketahui nilai *accuracy*, *precision*, *recall* dan *f-score* yang dapat dilihat pada persamaan berikut:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$f1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

Keterangan:

TP = nilai True Positive

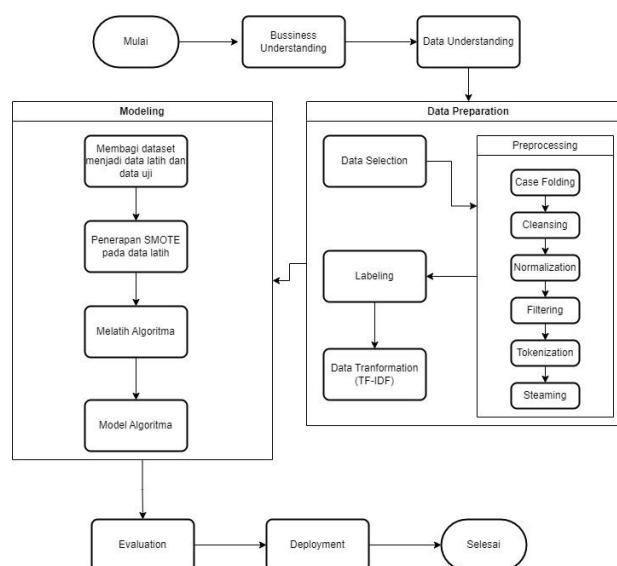
FP = nilai False Positive

TN = nilai True Negative

FN = nilai False Negative

3. METODE PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode *Cross-Industry Standard Procedure for Data Mining* (CRISP-DM) dengan tahapan *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation* dan *deployment*. Tahapan *business understanding* dilakukan untuk memahami objek penelitian. Tahapan *data understanding* dilakukan dengan melakukan *pengambilan dataset* dengan *scraping data* ulasan aplikasi TikTok yang diperoleh dari *Google Play Store* menggunakan menggunakan *library* Python, yaitu *google_play_scrapper*. Tahapan *data preparation* terdiri dari tiga hal yang dilakukan untuk mempersiapkan data yang masih mentah, yaitu *data Selection*, *text preprocessing*, *labeling* dan *data transformation*. Tahapan *modeling* merupakan penerapan dari algoritma yang digunakan untuk mengklasifikasikan sentimen ulasan aplikasi TikTok. Algoritma yang digunakan dalam penelitian ini adalah algoritma *Support Vector Machine*. Tahapan *evaluation* dilakukan untuk menguji model yang telah dibuat. Tahapan *deployment* dilakukan dengan memvisualisasi dan memaparkan hasil penelitian melalui grafik agar dapat lebih dimengerti.



Gambar 1. Tahapan Penelitian

4. PEMBAHASAN

4.1 Business Understanding

Pemahaman bisnis dilakukan dengan mengidentifikasi masalah, yaitu bagaimana cara analisis sentimen ulasan aplikasi TikTok di *Google Play Store*. Dengan analisis sentimen dapat diketahui bagaimana aplikasinya menurut pengguna sehingga dapat meningkatkan kualitas aplikasinya. Penelitian ini bertujuan untuk mengetahui gambaran umum mengenai sentimen pengguna aplikasi TikTok dan menerapkan algoritma *Support Vector Machine* untuk analisis sentimen.

4.2 Data Understanding

Data yang digunakan adalah ulasan dari pengguna aplikasi Tiktok pada *Google Play Store*. Data diperoleh dengan teknik *scraping* menggunakan *library Python* yaitu *google-play-scrapers* yang dilakukan pada Google Colab. Ulasan yang diambil adalah ulasan terbaru (*newest*) pada bulan Januari 2025. Ulasan yang diambil hanya mencakup ulasan berbahasa Indonesia dan dari pengguna yang berasal dari Indonesia. Data yang didapat terdiri dari 51.317 ulasan dengan atribut-atribut seperti *reviewId*, *userName*, *userImage*, *content*, *score*, *thumbsUpCount*, *reviewCreatedVersion*, *at*, *replyContent*, *repliedAt*, dan *appVersion*.

Tabel 1. Hasil data scraping

No	Baris ini	Regular	score	at
1	Pengguna Google	Mantap	5	31/01/2025 23:59
2	Pengguna Google	Saya ingin bilang kenapa setiap hari postingan dan pencarian nya aneh aneh semua	2	31/01/2025 23:58
4	Pengguna Google	Ini sangat luar biasa sangat bagus	5	31/01/2025 23:56
5	Pengguna Google	Sangat cocok	5	31/01/2025 23:52

4.3 Data Preparation

4.3.1 Data Selection

Data yang telah diperoleh sebelumnya kemudian akan diseleksi sesuai kebutuhan. Data yang digunakan adalah bagian ulasan pengguna yang terletak pada atribut *content*.

4.3.2 Text Preprocessing

Case Folding

Case Folding dilakukan dengan fungsi *string lower* yang berfungsi untuk membuat huruf alfabet menjadi huruf kecil (lowercase).

Tabel 2. Penerapan *case folding*

No	Sebelum	Sesudah
1	Aplikasinya jelek karena ada bug-nya, tolong diperbaiki bug-nya. Saya jadinya harus hapus install tiktok terus, selagi bug-nya belum diperbaiki, saya akan tetap kasih bintang 1!!!!	aplikasinya jelek karena ada bug-nya, tolong diperbaiki bug-nya. saya jadinya harus hapus install tiktok terus, selagi bug-nya belum diperbaiki, saya akan tetap kasih bintang 1!!!!
2	Terimakasih, lebih byak teman dan seru	terimakasih, lebih byak teman dan seru

Cleansing

Cleansing dilakukan dengan *library re* dari Python yang berfungsi untuk menerapkan *regular expression* atau RegEx pada ulasan. *Dataset* yang sudah di *case folding* akan diterapkan fungsi

clean_text untuk membersihkan ulasan dari elemen yang tidak diperlukan seperti *non-ascii character, special character, simbol, angka, duplicate character, dan double space.*

Tabel 3. Penerapan *cleansing*

No	Sebelum	Sesudah
1	aplikasinya jelek karena ada bug-nya, tolong diperbaiki bug-nya. saya jadinya harus hapus install tiktok terus, selagi bug-nya belum diperbaiki, saya akan tetap kasih bintang 1!!!!	aplikasinya jelek karena ada bug nya tolong diperbaiki bug nya saya jadinya harus hapus instal tiktok terus selagi bug nya belum diperbaiki saya akan tetap kasih bintang
2	terimakasih, lebih banyak teman dan seru	terimakasih lebih banyak teman dan seru

Tokenizing

Tokenizing dilakukan dengan fungsi *word_tokenize* dari *library Natural Language Toolkit* atau *nlk*. Kata-kata pada ulasan akan dipecah menjadi *item* atau token dalam bentuk *List*.

Tabel 4. Penerapan *tokenizing*

No	Sebelum	Sesudah
1	aplikasinya jelek karena ada bug nya tolong diperbaiki bug nya saya jadinya harus hapus instal tiktok terus selagi bug nya belum diperbaiki saya akan tetap kasih bintang	['aplikasinya', 'jelek', 'karena', 'ada', 'bug', 'nya', 'tolong', 'diperbaiki', 'bug', 'nya', 'saya', 'jadinya', 'harus', 'hapus', 'instal', 'tiktok', 'terus', 'selagi', 'bug', 'nya', 'belum', 'diperbaiki', 'saya', 'akan', 'tetap', 'kasih', 'bintang']
2	terimakasih lebih banyak teman dan seru	['terimakasih', 'lebih', 'banyak', 'teman', 'dan', 'seru']

Normalization

Normalization dilakukan dengan mengubah ulasan yang memiliki kata tidak baku menjadi baku menggunakan *dataset kamus-alay* dari penelitian [14].

Tabel 5. Penerapan *tokenizing*

No	Sebelum	Sesudah
1	['aplikasinya', 'jelek', 'karena', 'ada', 'bug', 'nya', 'tolong', 'diperbaiki', 'bug', 'nya', 'saya', 'jadinya', 'harus', 'hapus', 'instal', 'tiktok', 'terus', 'selagi', 'bug', 'nya', 'belum', 'diperbaiki', 'saya', 'akan', 'tetap', 'kasih', 'bintang']	['aplikasinya', 'jelek', 'karena', 'ada', 'bug', 'nya', 'tolong', 'diperbaiki', 'bug', 'nya', 'saya', 'jadinya', 'harus', 'hapus', 'instal', 'tiktok', 'terus', 'selagi', 'bug', 'nya', 'belum', 'diperbaiki', 'saya', 'akan', 'tetap', 'kasih', 'bintang']
2	['terimakasih', 'lebih', 'banyak', 'teman', 'dan', 'seru']	['terimakasih', 'lebih', 'banyak', 'teman', 'dan', 'seru']

Filtering

Filtering merupakan proses menghilangkan kata pada ulasan yang termasuk pada *stopword list*. *Stopword* merupakan kata-kata yang tidak berpengaruh pada ulasan. *Stopword list* yang digunakan berasal dari *library nltk* dan ditambahkan beberapa kata yang banyak muncul pada *dataset* namun tidak memiliki pengaruh terhadap sentimen.

Tabel 6. Penerapan *tokenizing*

No	Sebelum	Sesudah
1	['aplikasinya', 'jelek', 'karena', 'ada', 'bug', 'nya', 'tolong', 'diperbaiki', 'bug', 'nya', 'saya', 'jadinya', 'harus', 'hapus', 'instal', 'tiktok', 'terus', 'selagi', 'bug', 'nya', 'belum', 'diperbaiki', 'saya', 'akan', 'tetap', 'kasih', 'bintang']	['aplikasinya', 'jelek', 'bug', 'tolong', 'diperbaiki', 'bug', 'hapus', 'instal', 'tiktok', 'selagi', 'bug', 'belum', 'diperbaiki', 'kasih', 'bintang']
2	['terimakasih', 'lebih', 'banyak', 'teman', 'dan', 'seru']	['terimakasih', 'lebih', 'teman', 'seru']

Stemming

Stemming menggunakan *library* Sastrawi yang akan mengubah kata turunan menjadi kata dasar dengan menghilangkan sufiks, prefiks dan konfiks yang terdapat pada kata.

Tabel 7. Penerapan *tokenizing*

No	Sebelum	Sesudah
1	['aplikasinya', 'jelek', 'bug', 'tolong', 'diperbaiki', 'bug', 'hapus', 'instal', 'tiktok', 'selagi', 'bug', 'belum', 'diperbaiki', 'kasih', 'bintang']	['aplikasi', 'jelek', 'bug', 'tolong', 'baik', 'bug', 'hapus', 'instal', 'tiktok', 'selagi', 'bug', 'belum', 'baik', 'kasih', 'bintang']
2	['terimakasih', 'lebih', 'teman', 'seru']	['terimakasih', 'lebih', 'teman', 'seru']

4.3.3 Labeling

Labeling dilakukan secara otomatis dengan menggunakan gabungan leksikon berbahasa Indonesia yaitu *InSet (Indonesia Sentiment Lexicon)* dari penelitian [15] dan *sentiword_id* dari penelitian [16]. Ulasan yang memiliki bobot diatas 0 akan dilabeli positif dan ulasan yang memiliki bobot dibawah 0 akan dilabeli negatif. Didapatkan dari 42279 ulasan, yang diberi label bersentimen positif sebanyak 30527 ulasan (72,2%) dan yang bersentimen negatif sebanyak 11752 ulasan (27,8%).

Tabel 8. Hasil *labeling*

No	Text Preprocessing	Label
1	mantap	Positif
2	bilang postingan cari aneh aneh	Negatif
3	update ngelag kena imbas of buka hanya logo tiktok muncul tiktok tidak jalan keluar uninstal download login	Negatif
4	sangat sangat bagus	Positif
5	sangat cocok	Positif

4.3.4 Data Transformation

Ulasan yang telah diterapkan *text Preprocessing* akan diubah menjadi bentuk lain agar dapat digunakan oleh algoritma *Support Vector Machine*. Pada penelitian ini menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah representasi teks menjadi vektor-vektor fitur yang memuat informasi tentang pentingnya setiap kata dalam setiap dokumen.

4.4 Modeling

Pada tahap ini, berfokus terhadap penerapan dari algoritma yang digunakan untuk mengklasifikasikan sentimen ulasan aplikasi TikTok. Proses *modeling* menggunakan *library scikit-learn*. *Dataset* yang telah diolah pada tahap *data preparation* dibagi menjadi dua, yaitu data latih (90%) dan data uji (10%). Setelah melakukan *data splitting* kemudian mengoptimalkan model yang akan dibuat dengan menggunakan *GridSearchCV*. *GridSearchCV* adalah bagian dari *library scikit-learn* yang berfungsi untuk mencari parameter optimal untuk model algoritma berdasarkan parameter yang disediakan. Dari *GridSearchCV* dengan *3 fold cross validation* didapatkan dari 85 kombinasi parameter yang dicoba yang merupakan parameter terbaik untuk model algoritma SVM yang akan dilatih adalah nilai $C=10$, $\gamma=0.1$ dan *kernel*-nya *rbf* seperti pada Tabel 9.

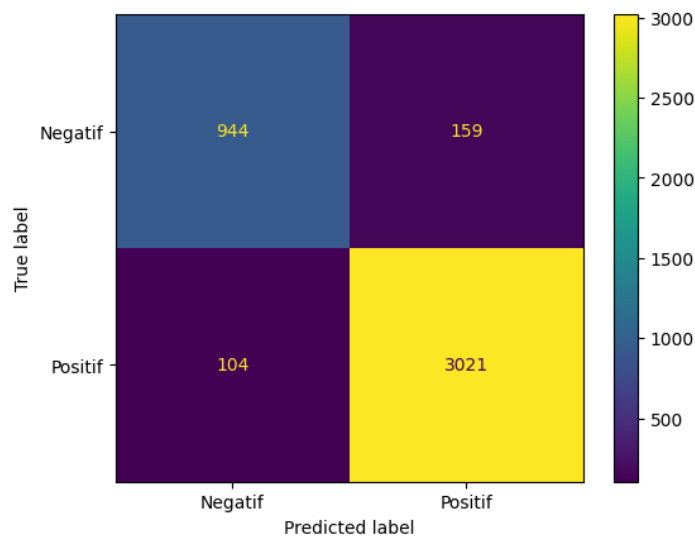
Tabel 9. Hasil *grid search*

Pengujian Ke	C	Kernel	Gamma	Degree	Accuracy
17	10	rbf	0.1	-	0.928951
22	100	rbf	0.01	-	0.928504
21	100	rbf	0.1	-	0.927426
2	1	linear	-	-	0.926454
3	10	linear	-	-	0.926217

Setelah mendapatkan parameter optimal kemudian pada data latih akan diterapkan SMOTE untuk menyeimbangkan data. Dari hasil penerapan SMOTE untuk data latih yang awalnya data positif berjumlah 27452 ulasan dan data negatif berjumlah 10599 menjadi sama-sama berjumlah 27452. Setelah mendapatkan parameter optimal dan menerapkan SMOTE lalu melatih algoritma menggunakan data latih dengan fungsi fit. Kemudian model algoritma yang telah dilatih melakukan prediksi pada data uji untuk mengklasifikasikan data uji menggunakan fungsi *predict*.

4.5 Evaluation

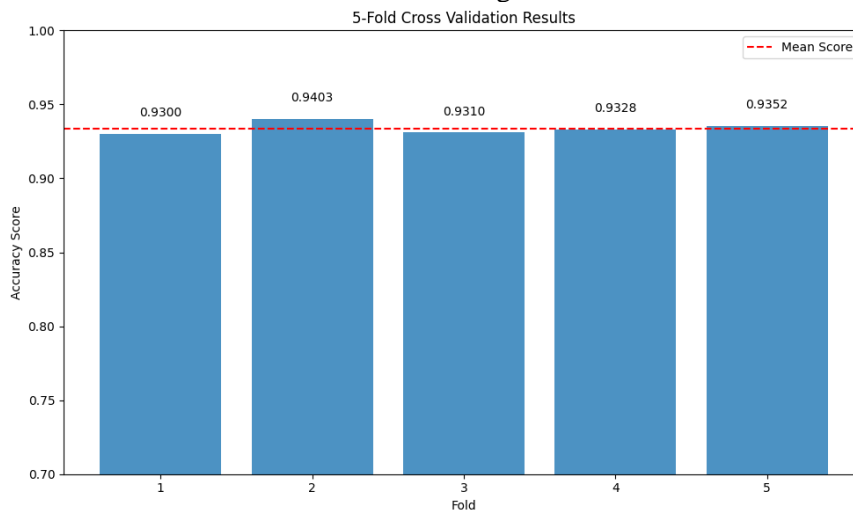
Pada tahap ini berfokus terhadap proses pengujian atau evaluasi dari model yang telah dibuat. Pengujian terhadap model dilakukan dengan menggunakan *Confusion Matrix*. *Confusion Matrix* menampilkan hasil klasifikasi dengan model algoritma dan hasil klasifikasi yang sebenarnya.



Gambar 2. Hasil *confusion matrix*

Dari hasil klasifikasi diatas dapat diketahui nilai *accuracy*, *presicion*, *Recall* dan *f1-score*. *Accuracy* yang didapat sebesar 93,78%. Nilai tersebut berarti bahwa model dapat mengklasifikasikan data secara benar dari keseluruhan data uji dengan baik. *Precision* yang didapat untuk ulasan positif sebesar 95,00% dan untuk ulasan negatif sebesar 90,08%. Nilai tersebut berarti model memiliki tingkat ketepatan yang baik dalam mengklasifikasikan ulasan pada masing-masing kelas. *Recall* yang didapat untuk ulasan positif sebesar 96,67% dan untuk ulasan negatif sebesar 85,58%. Nilai tersebut berarti model dapat secara efektif mendeteksi sebagian besar ulasan dengan benar pada masing-masing kelas. *F1-score* yang didapat untuk ulasan positif sebesar 95,83% dan ulasan negatif sebesar 87,77%. Nilai tersebut berarti model memiliki kinerja yang seimbang dan optimal dalam mengklasifikasikan ulasan, baik ulasan yang positif maupun ulasan yang negatif.

Untuk menguji apakah model yang telah dilatih memiliki performa yang baik ketika mengklasifikasikan data baru maka dilakukan *k-fold cross validation*. Metode *k-fold cross validation* pada penelitian ini menggunakan nilai $k=5$ seperti pada penelitian [7] dimana data latih akan diuji sebanyak 5 kali dengan 4 bagian sebagai data latih dengan SMOTE dan 1 bagian sebagai data validasi tanpa SMOTE secara bergantian. Dari hasil pengujian didapatkan rata-rata akurasi pengujianya cukup stabil sebesar 93.39%. Setiap pengujian menghasilkan akurasi yang tidak berbeda jauh satu sama lain seperti yang terlihat pada **Error! Reference source not found.** yang berarti model tidak memiliki masalah dalam mengklasifikasikan data baru.

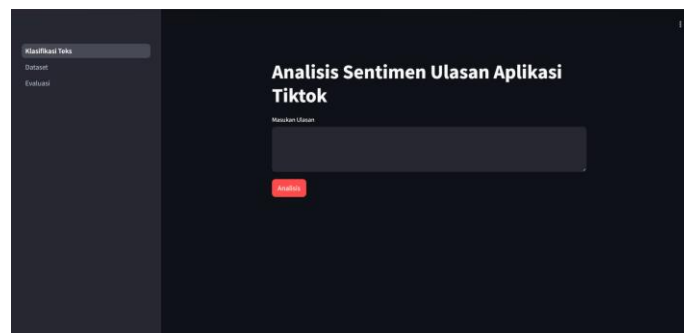


Gambar 3. Hasil 5-fold cross validation

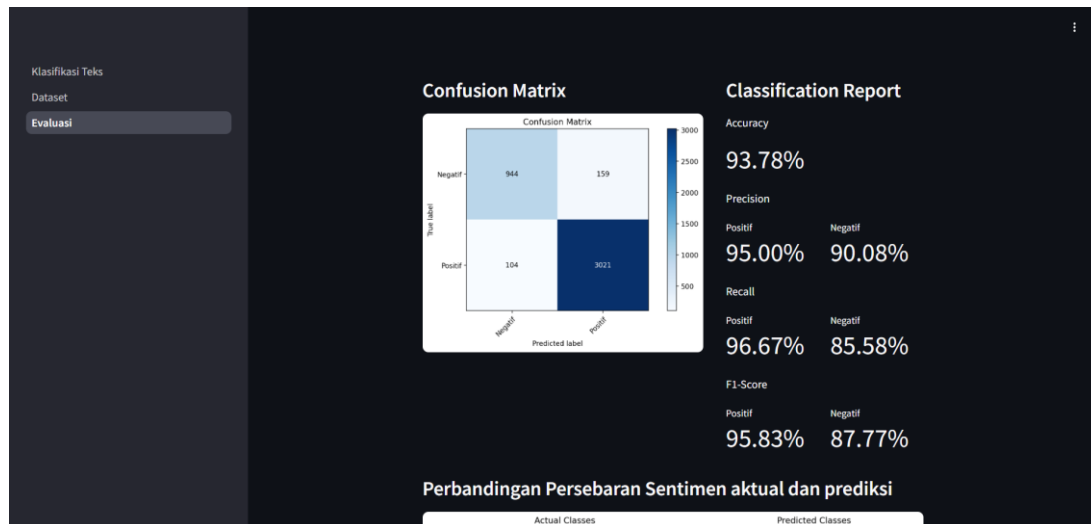
4.6 Deployment

4.6.1 Halaman Klasifikasi

Pada halaman klasifikasi terdapat *text input* untuk memasukkan ulasan yang ingin diklasifikasikan. Dengan menekan tombol "analisis" akan mengeluarkan hasil klasifikasi dari ulasannya.



Gambar 4. Halaman klasifikasi teks



Gambar 10. Halaman evaluasi

5. KESIMPULAN

Penelitian ini menggunakan data dari ulasan aplikasi TikTok pada Google Play Store. Data yang digunakan merupakan ulasan terbaru pada bulan Januari tahun 2025 dan didapatkan data sebanyak 51066 ulasan. Data diolah terlebih dahulu pada tahap preprocessing sehingga tersisa 42279 ulasan. Data yang telah diolah lalu diberi label sesuai sentimennya yaitu positif dan negatif. Pelabelan dilakukan menggunakan pendekatan lexicon based. Dari hasil pelabelan didapatkan ulasan positif sebanyak 30527 ulasan dan ulasan negatif sebanyak 11752 ulasan. Analisis sentimen dilakukan menggunakan algoritma Support Vector Machine, dengan pembagian data 90% sebagai data latih dan 10% sebagai data uji. Evaluasi terhadap model yang telah dilatih menghasilkan accuracy sebesar 93.78%, precision untuk sentimen positif sebesar 95.00% dan sentimen negatif sebesar 90.08%, recall untuk sentimen positif sebesar 96.67% dan sentimen negatif sebesar 85.58%, dan f1-score untuk sentimen positif sebesar 95.83% dan sentimen negatif sebesar 87.77%. Hasil ini menunjukkan bahwa algoritma Support Vector Machine dapat menganalisis sentimen ulasan aplikasi TikTok secara efektif. Hasil dari penelitian ini kemudian diimplementasikan dalam bentuk interface web yang interaktif menggunakan library streamlit.

DAFTAR PUSTAKA

- [1] B. Winarso, "Apa Itu TikTok dan Apa Saja Fitur-fiturnya?," DailySocial.id. Diakses: 26 Februari 2024. [Daring]. Tersedia pada: <https://dailysocial.id/post/apa-itu-tik-tok>
- [2] S. Fanissa, M. A. Fauzi, dan S. Adinugroho, "Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking," 2018. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [3] E. Suryati, A. Ari Aldino, N. Penulis Korespondensi, dan E. Suryati Submitted, "Analisis Sentimen Transportasi Online Menggunakan Ekstraksi Fitur Model Word2vec Text Embedding Dan Algoritma Support Vector Machine (SVM)," vol. 4, no. 1, hlm. 96–106, 2023, doi: 10.33365/jtsi.v4i1.2445.
- [4] M. Samantri dan Afiyati, "Perbandingan Algoritma Support Vector Machine dan Random Forest untuk Analisis Sentimen Terhadap Kebijakan Pemerintah Indonesia Terkait Kenaikan Harga BBM Tahun 2022," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, hlm. 2024, 2024, doi: 10.35870/jti.
- [5] N. Raisa, N. Riza, dan W. I. Rahayu, "Analisis Sentimen Menggunakan SVM Dan KNN Pada Review Drama Korea Di Mydramalist," 2023.
- [6] Y. A. Singgalen, "Analisis Sentimen Top 10 Traveler Ranked Hotel di Kota Makassar Menggunakan Algoritma Decision Tree dan Support Vector Machine," *KLIK: Kajian*

-
- Ilmiah Informatika dan Komputer*, vol. 4, no. 1, hlm. 323–332, 2023, doi: 10.30865/klik.v4i1.1153.
- [7] J. Alfiah Zulqornain, Indriati, dan P. Pandu Adikara, “Analisis Sentimen Tanggapan Masyarakat Aplikasi Tiktok Menggunakan Metode Naïve Bayes dan Categorical Proportional Difference (CPD),” *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 5, no. 7, hlm. 2886–2890, 2021, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [8] F. V. Sari dan A. Wibowo, “Analisis Sentimen Pelanggan Toko Online JD.ID Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi,” *Jurnal SIMETRIS*, vol. 10, no. 2, 2019.
- [9] R. Mahendrajaya, G. A. Buntoro, dan M. B. Setyawan, “Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based Dan Support Vector Machine,” *Komputek*, vol. 3, no. 2, hlm. 52–63, 2019, doi: 10.24269/jkt.v3i2.270.
- [10] S. Y. Pangestu, Y. Astuti, dan L. D. Farida, “Algoritma Support Vector Machine Untuk Klasifikasi Sikap Politik Terhadap Partai Politik Indonesia,” *Jurnal Mantik Penusa*, vol. 3, no. 1, hlm. 236–241, 2019, [Daring]. Tersedia pada: <https://t.co/eF>
- [11] H. C. Husada dan A. S. Paramita, “Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM),” *Teknika*, vol. 10, no. 1, hlm. 18–26, Feb 2021, doi: 10.34148/teknika.v10i1.311.
- [12] S. Lende *dkk.*, “Analisis Sentimen Siswa Terhadap Pelajaran Informatika di SMPK St. Yohanes Kalembu Lona Dengan Metode Naive Bayes Classifier,” *Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI)*, vol. 6, hlm. 217–225, 2023.
- [13] M. I. Fikri, T. S. Sabrila, Y. Azhar, dan U. M. Malang, “Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter,” *SMATIKA Jurnal: STIKI Informatika Jurnal*, vol. 10, no. 02, hlm. 71–76, 2020.
- [14] N. Aliyah Salsabila, Y. Ardhito Winatmoko, A. Akbar Septiandri, dan A. Jamal, “Colloquial Indonesian Lexicon,” dalam *International Conference on Asian Language Processing (IALP)*, 2018, hlm. 226–229. doi: 10.1109/IALP.2018.8629151.
- [15] F. Koto dan G. Y. Rahmaningtyas, “Inset lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs,” dalam *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017*, Institute of Electrical and Electronics Engineers Inc., Jul 2017, hlm. 391–394. doi: 10.1109/IALP.2017.8300625.
- [16] D. H. Wahid dan S. N. Azhari, “Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity,” *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 10, no. 2, hlm. 207–218, 2016.
-