Sistemasi: Jurnal Sistem Informasi
Volume 15, Nomor 2, 2026: 658-668
ISSN:2302-8149
e-ISSN:2540-9719

Analisis Sentimen Ulasan Aplikasi *CapCut* pada *Google Play Store* menggunakan *Support Vector Machine* dengan Teknik *SMOTE*

Sentiment Analysis of CapCut Application Reviews using Support Vector Machine with the SMOTE Technique

[1]Faridah Ayu Shefia*, [2]Pratomo Setiaji, [3]Wiwit Agus Triyanto
[1,2,3]Program Studi Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus
[1,2,3]Jl. Lingkar Utara, Gondangmanis, Kecamatan Bae, Kota Kudus, Jawa Tengah 59327, Indonesia
e-mail: 202153124@std.umk.iac.id, pratomo.setiaji@umk.ac.id, at.wiwit@umk.ac.id

(***received*:** 16 December 2025, ***revised*:** 5 January 2026, ***accepted*:** 6 January 2026)

Abstrak

Popularitas konten video pendek di berbagai *platform* media sosial mendorong meningkatnya penggunaan aplikasi pengedit video lintas perangkat, baik melalui smartphone, desktop, maupun layanan berbasis web. *CapCut* merupakan salah satu aplikasi yang banyak dimanfaatkan untuk menghasilkan konten kreatif, dan ulasan pengguna pada *Google Play Store* menjadi indikator penting dalam menilai kualitas pengalaman pengguna. Namun, data ulasan umumnya tidak seimbang, dengan dominasi sentimen positif dan jumlah sentimen netral yang jauh lebih sedikit, sehingga menjadi tantangan dalam proses klasifikasi. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna *CapCut* menggunakan *Support Vector Machine (SVM)* dengan penerapan *Synthetic Minority Over-sampling Technique (SMOTE)* untuk menangani ketidak seimbangan data. Data dikumpulkan melalui pemindaian ulasan pada *Google Play Store* dan menghasilkan 4.381 data setelah tahap pembersihan. Ulasan kemudian melalui prapemrosesan teks, pembobotan *TF-IDF*, dan pelatihan model. Hasil pengujian menunjukkan *SVM* mampu mencapai akurasi sebesar 73.54% dengan nilai *F1-Score* berbobot 0.736. Temuan ini menunjukkan *SMOTE* berperan dalam meningkatkan performa model pada kelas minoritas. Penelitian ini menggambarkan persepsi pengguna terhadap *CapCut* serta menunjukkan potensi *SVM* sebagai metode klasifikasi sentimen pada ulasan aplikasi berbasis teks.

Kata kunci : analisis sentimen, *capcut, google play store*, SMOTE, *support vector machine*

Abstract

The growing popularity of short-form video content across various social media platforms has increased the use of cross-device video editing applications, accessible through smartphones, desktops, and web-based services. CapCut is one of the most widely used applications for creating creative content, and user reviews on the Google Play Store serve as an important indicator for evaluating user experience quality. However, review datasets are often imbalanced, with positive sentiment dominating and neutral sentiment appearing in much smaller proportions, which poses challenges for sentiment classification. This study aims to analyze user sentiment toward CapCut reviews using Support Vector Machine (SVM) and applying the Synthetic Minority Over-sampling Technique (SMOTE) to address data imbalance. The data were collected by scraping reviews from the Google Play Store, resulting in 4,381 cleaned review entries after the data cleaning stage. The reviews then underwent text preprocessing, TF-IDF feature weighting, and model training. The experimental results show that the SVM model achieved an accuracy of 73.54% with a weighted F1-score of 0.736. These findings indicate that SMOTE contributes to improving model performance on minority classes. Overall, this study provides insights into user perceptions of CapCut and highlights the potential of SVM as an effective sentiment classification method for text-based application reviews.

Keywords: *sentiment analysis, capcut, google play store, SMOTE, support vector machine*

 ISSN:2302-8149
e-ISSN:2540-9719

1 Pendahuluan

Perkembangan konten video pendek telah mengubah cara pengguna internet dalam mengonsumsi informasi dan hiburan. Platform seperti *TikTok, Instagram Reels*, dan *YouTube Shorts* memicu tingginya kebutuhan terhadap aplikasi pengedit video yang ringkas, mudah digunakan, serta tersedia lintas perangkat. CapCut menjadi salah satu aplikasi yang banyak diadopsi oleh pembuat konten karena menawarkan fitur pengeditan lengkap tanpa memerlukan perangkat profesional. CapCut tidak hanya terlihat dari tingginya jumlah pengguna di platform media sosial, tetapi juga dari pemanfaatannya di lingkungan pendidikan sebagai media pembelajaran [1]. Popularitas aplikasi ini tercermin dari tingginya jumlah unduhan dan banyaknya ulasan yang diberikan pengguna pada *Google Play Store*, yang dapat menjadi indikator kualitas pengalaman penggunaan aplikasi [2].

Analisis sentimen pada ulasan aplikasi merupakan pendekatan yang digunakan untuk memahami persepsi pengguna secara otomatis melalui pemanfaatan algoritma machine learning [2], [3]. Karakteristik ulasan pada platform digital cenderung tidak terstruktur, melibatkan penggunaan bahasa informal, singkatan, campuran bahasa, serta emoji, sehingga menyulitkan model dalam mengenali sentimen secara akurat [4]. Selain itu, distribusi sentimen dalam ulasan aplikasi sering kali tidak seimbang, dengan dominasi sentimen positif dan jumlah ulasan netral atau negatif yang lebih sedikit. Kondisi ini dapat menyebabkan model klasifikasi bias terhadap kelas mayoritas dan menghasilkan performa yang rendah pada kelas minoritas [5]. Penanganan ketidakseimbangan kelas menjadi faktor penting dalam meningkatkan kualitas model klasifikasi sentimen berbasis machine learning [6],

Berbagai metode klasifikasi telah digunakan dalam penelitian analisis sentimen, seperti *Naive Bayes, K-Nearest Neighbor, Random Forest,* dan *Support Vector Machine (SVM)* [7]. Sejumlah studi menunjukkan bahwa *SVM* memiliki kinerja yang stabil pada data teks berbasis fitur TF–IDF [7], [8], [9]. Namun, performa model dapat menurun ketika jumlah data minoritas sangat sedikit, sehingga diperlukan teknik penyeimbangan data [10], [11]. Untuk mengatasi hal tersebut, *Synthetic Minority Over-sampling Technique (SMOTE)* digunakan untuk menyeimbangkan data dengan membangkitkan sampel sintetik pada kelas minoritas [8], [10]. Studi terdahulu melaporkan bahwa penerapan *SMOTE* dapat membantu meningkatkan performa model, terutama pada metrik yang sensitif terhadap kelas minoritas seperti recall dan F1-Score [12].

Penelitian terkait ulasan aplikasi CapCut masih terbatas dan sebagian besar menggunakan algoritma lain seperti *Naive Bayes* tanpa penanganan ketidakseimbangan data atau hanya menggunakan skema klasifikasi yang lebih sederhana [2], [13]. Hal ini menunjukkan adanya celah penelitian untuk mengevaluasi kinerja SVM yang dikombinasikan dengan SMOTE pada data ulasan CapCut. Kebaruan penelitian ini terletak pada evaluasi klasifikasi sentimen ulasan *CapCut* tiga kelas (positif, netral, negatif) menggunakan SVM kernel *linear* pada fitur TF–IDF n-gram (1,2) dengan pembatasan 5.000 fitur, serta penerapan SMOTE hanya pada data latih untuk mengatasi ketidakseimbangan kelas [10], [11], [12]. Selain melaporkan akurasi dan F1-Score berbobot, penelitian ini juga menekankan analisis kesalahan prediksi khususnya pada kelas netral serta membandingkan kinerja SVM dengan model pembanding untuk menunjukkan karakteristik performa pada data teks yang bersifat sparse [6] [7]. Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis sentimen ulasan pengguna *CapCut* pada *Google Play Store* menggunakan algoritma SVM dengan penerapan SMOTE untuk mengatasi ketidakseimbangan data, serta mengevaluasi klasifikasi tiga kelas sentimen: positif, netral, dan negatif [10].

2 Tinjauan Literatur

Penelitian mengenai analisis sentimen pada ulasan aplikasi digital menunjukkan bahwa pendekatan machine learning efektif digunakan untuk mengidentifikasi persepsi pengguna secara otomatis melalui teks ulasan yang tidak terstruktur [3], [6]. Ulasan pada *Google Play Store* umumnya bersifat informal, menggunakan singkatan dan campuran bahasa, serta dapat memuat emoji yang menambah kompleksitas pemrosesan bahasa [4]. Selain tantangan linguistik, beberapa studi juga menemukan bahwa ulasan aplikasi sering membentuk distribusi kelas yang tidak seimbang, sehingga model cenderung lebih kuat pada kelas mayoritas dan kurang baik dalam memprediksi kelas minoritas [6], [8], [14]. Kondisi tersebut dapat membuat nilai akurasi terlihat cukup tinggi, tetapi performa pada metrik seperti recall atau *F1-Score* untuk kelas minoritas rendah [12].

 ISSN:2302-8149
e-ISSN:2540-9719

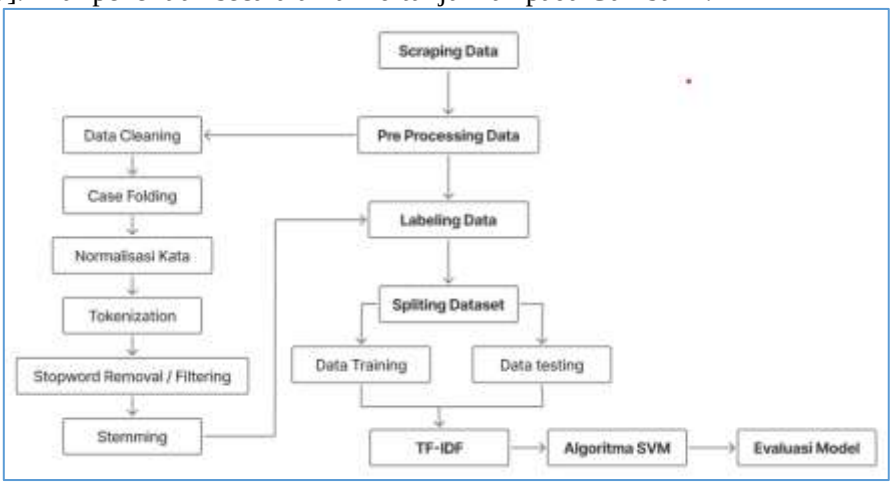
Sejumlah metode telah digunakan dalam penelitian analisis sentimen, antara lain *Naive Bayes*, *K-Nearest Neighbor*, *Random Forest*, dan *Support Vector Machine (SVM)* [3], [7], [9]. Hasil studi terdahulu menunjukkan bahwa pemilihan metode dipengaruhi oleh karakteristik dataset dan teknik representasi teks yang digunakan [6]. *Naive Bayes* sering diterapkan karena sederhana tetapi memiliki keterbatasan ketika data tidak seimbang [3], [8]. Di sisi lain, SVM dinilai stabil untuk menangani data teks yang berdimensi tinggi terutama dengan pembobotan fitur TF-IDF dan menunjukkan performa kompetitif dibandingkan algoritma lain pada berbagai kasus analisis sentimen [7], [15].

Ketidakseimbangan data merupakan permasalahan umum dalam analisis sentimen karena dapat menurunkan kinerja model pada kelas minoritas [8], [16]. Salah satu teknik yang banyak digunakan untuk mengatasinya adalah *Synthetic Minority Over-sampling Technique (SMOTE)*, yang bekerja dengan membangkitkan data sintetis sehingga distribusi kelas menjadi lebih seimbang [10], [17], [18]. Beberapa penelitian menunjukkan bahwa penerapan SMOTE mampu meningkatkan nilai F1-Score serta membantu model mengenali pola sentimen yang sebelumnya kurang terwakili [8]. Kombinasi SMOTE dengan *Support Vector Machine (SVM)* juga terbukti efektif dalam meningkatkan stabilitas dan performa klasifikasi, khususnya pada data teks berdimensi tinggi berbasis TF–IDF [12].

Penelitian analisis sentimen pada aplikasi CapCut sendiri masih terbatas. Studi yang ada umumnya menggunakan *Naive Bayes* pada ulasan CapCut tanpa menerapkan teknik penanganan ketidakseimbangan data dan sering menggunakan skema klasifikasi yang lebih sederhana [13]. Di sisi lain, kombinasi SVM dan SMOTE telah banyak diterapkan pada konteks aplikasi lain namun belum memberikan gambaran performa pada ulasan *CapCut* untuk tiga kelas sentimen. Oleh karena itu, penelitian ini mengembangkan hipotesis bahwa penerapan SMOTE pada ulasan CapCut yang tidak seimbang dapat meningkatkan performa SVM, khususnya pada kelas minoritas, sehingga hasil klasifikasi menjadi lebih proporsional [2], [11], [13].

3 Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif melalui proses klasifikasi sentimen berbasis machine learning pada ulasan pengguna aplikasi *CapCut di Google Play Store*. Tahapan penelitian terdiri dari pengumpulan data, pra-pemrosesan teks, pelabelan sentimen, pembobotan fitur menggunakan TF–IDF, penyeimbangan data menggunakan SMOTE, serta pelatihan dan evaluasi model SVM [19]. Alur penelitian secara umum ditunjukkan pada Gambar 1.



Gambar 1 Flowchart analisis sentimen

3.1 Pengumpulan Data

Data dikumpulkan menggunakan teknik web scraping pada platform *Google Play Store.* dengan menggunakan bahasa pemrograman Python. Pengambilan data difokuskan pada ulasan pengguna aplikasi *CapCut* yang berisi informasi berupa teks ulasan, rating, serta tanggal publikasi. Proses scraping dilakukan secara otomatis untuk mengumpulkan data dalam jumlah besar [2]. Dataset awal yang diperoleh masih mengandung berbagai kekurangan, seperti data duplikat, ulasan kosong, serta komentar yang tidak relevan dengan tujuan analisis sentimen. Oleh karena itu, dilakukan tahap pembersihan data (*data cleaning*) untuk menghilangkan elemen-elemen tersebut [6].

 ISSN:2302-8149
e-ISSN:2540-9719

Setelah melalui proses pembersihan, diperoleh sebanyak 4.381 ulasan yang dinilai layak digunakan sebagai dataset penelitian. Selanjutnya, data tersebut digunakan pada tahapan pra-pemrosesan, pelabelan sentimen, pembobotan fitur, serta pelatihan model klasifikasi [2] [3].

3.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk menyiapkan data teks agar dapat diproses oleh model. Proses ini terdiri dari beberapa tahap umum dalam analisis sentimen, yaitu:

- ***Cleaning***: menghapus karakter non-alfabet, hyperlink, angka, tanda baca, emoji, serta teks berulang yang tidak relevan.
- ***Case folding***: mengubah seluruh teks menjadi huruf kecil, sehingga kata dengan bentuk berbeda tetapi makna sama diperlakukan secara konsisten.
- ***Normalisasi***: memperbaiki ejaan tidak baku dan bentuk singkatan ke bentuk standar berdasarkan kamus normalisasi.
- ***Tokenization***: memecah kalimat menjadi token kata untuk memudahkan proses analisis.
- ***Stopword removal***: menghapus kata-kata umum yang tidak memberikan makna sentimen seperti “dan”, “yang”, atau “pada”.
- ***Stemming***: mengembalikan kata berimbuhan ke bentuk dasar untuk mengurangi variasi kata dengan makna sama.

Hasil dari tahap pra-pemrosesan menghasilkan data teks yang lebih bersih dan siap digunakan pada tahap ekstraksi fitur [6].

3.3 Pelabelan Data

Pelabelan data dilakukan menggunakan pendekatan *rating-based labeling*, yaitu mengonversi nilai rating pengguna menjadi kelas sentimen. Kategori pelabelan ditentukan sebagai berikut:

- Rating 4-5 : Sentimen Positif
- Rating 3 : Sentimen Netral
- Rating 1-2 : Sentimen Negatif

Pendekatan ini dipilih karena bersifat objektif dan sesuai untuk analisis ulasan aplikasi di Google Play Store, serta mampu mengurangi subjektivitas dalam proses pelabelan manual.

3.4 Visualisasi Word Cloud

Sebelum dilakukan pembobotan fitur menggunakan TF–IDF, penelitian ini menerapkan visualisasi word cloud untuk memperoleh gambaran awal mengenai kata-kata yang paling sering muncul pada ulasan pengguna. *Word cloud* digunakan sebagai analisis eksploratif guna mengidentifikasi kecenderungan istilah yang merepresentasikan sentimen positif, netral, maupun negatif.

Visualisasi ini membantu memastikan bahwa hasil pra-pemrosesan telah berjalan dengan baik serta memberikan pemahaman awal terhadap karakteristik data sebelum dilakukan ekstraksi fitur numerik [6]. Dengan demikian, word cloud berperan sebagai tahap pendukung yang memperkuat analisis sebelum proses pembobotan TF–IDF dan pelatihan model klasifikasi dilakukan.

3.5 Pembobotan Fitur Menggunakan TF–IDF

Pembobotan fitur dilakukan menggunakan metode *Term Frequency–Inverse Document Frequency* (TF–IDF) untuk merepresentasikan data teks dalam bentuk numerik. Metode ini digunakan karena mampu menggambarkan tingkat kepentingan suatu kata dalam dokumen relatif terhadap keseluruhan korpus, serta efektif dalam pengolahan teks berdimensi tinggi serta sering digunakan dalam klasifikasi sentimen [7].

Nilai Term Frequency (TF) digunakan untuk menunjukkan seberapa sering suatu kata muncul dalam sebuah dokumen. Perhitungan TF ditunjukkan pada Persamaan (1).

- Term Frequency (TF):

$$TF(t,d) = \frac{f_{t,d}}{\Sigma_k f_{k,d}} \quad (1)$$

Selanjutnya, pada *Inverse Document Frequency* (IDF) digunakan untuk mengurangi bobot kata yang sering muncul di banyak dokumen dan meningkatkan bobot kata yang lebih spesifik. Perhitungan IDF ditunjukkan pada Persamaan (2).

- Inverse Document Frequency (IDF):

$$IDF(t) = log\,\frac{N}{n_t} \quad (2)$$

 ISSN:2302-8149
e-ISSN:2540-9719

Nilai TF dan IDF kemudian dikalikan untuk memperoleh bobot akhir TF–IDF, seperti ditunjukkan pada Persamaan (3).

- TF–IDF:

$$TF - IDF(t,d) = TF(t,d)\ x\ IDF(t) \quad (3)$$

Keterangan: $f_{t,d}$ adalah frekuensi kata t dalam dokumen d , N adalah jumlah seluruh dokumen, dan n_t adalah jumlah dokumen yang mengandung kata t.

Hasil pembobotan TF–IDF ini digunakan sebagai representasi fitur teks yang selanjutnya diproses pada tahap pelatihan model klasifikasi.

3.6 Spliting Dataset

Dataset dibagi menjadi data pelatihan dan data pengujian dengan rasio 80:20. Proporsi ini banyak digunakan dalam penelitian analisis sentimen untuk menjaga keseimbangan antara kemampuan model mempelajari pola dari dataset dan kemampuan model diuji pada data baru. Pembagian dilakukan secara acak untuk menghindari bias distribusi kelas.

3.7 Penyeimbangan Data Menggunakan SMOTE

Distribusi sentimen pada ulasan menunjukkan ketidakseimbangan kelas, di mana kelas negatif memiliki jumlah data yang lebih dominan dibandingkan kelas positif dan netral. Ketidakseimbangan ini berpotensi menyebabkan model klasifikasi cenderung bias terhadap kelas mayoritas, sehingga menurunkan performa prediksi pada kelas dengan jumlah data lebih sedikit. Untuk mengatasi permasalahan tersebut, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE). SMOTE bekerja dengan menghasilkan sampel sintetis pada kelas minoritas berdasarkan kedekatan fitur antar data, sehingga distribusi kelas menjadi lebih seimbang [10].

Pada penelitian ini, SMOTE diterapkan setelah proses pembagian data (train-test split) dan hanya pada data latih, guna menghindari terjadinya data leakage. Parameter yang digunakan adalah nilai default dengan k_neighbors = 5, yang berarti setiap sampel minoritas dibangkitkan berdasarkan lima tetangga terdekatnya.

Penerapan SMOTE bertujuan untuk meningkatkan kemampuan model dalam mengenali pola pada kelas minoritas serta meningkatkan performa klasifikasi, khususnya pada metrik recall dan F1-Score, sebagaimana ditunjukkan dalam penelitian sebelumnya [20].

3.8 Pelatihan dan Evaluasi Model

Model klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM) dengan kernel linear [21]. Pemilihan SVM didasarkan pada kemampuannya dalam menangani data berdimensi tinggi serta kinerjanya yang stabil pada representasi fitur TF–IDF. Model dilatih menggunakan data latih yang telah melalui proses SMOTE. Evaluasi dilakukan menggunakan data uji yang tidak terlibat dalam proses pelatihan. Metrik evaluasi yang digunakan meliputi: akurasi, precision, recall, F1-Score, confusion matrix [7].

Formulasi metrik evaluasi adalah sebagai berikut:

- Akurasi

  Akurasi mengukur proporsi prediksi yang benar terhadap seluruh data pengujian. Perhitungan nilai akurasi ditunjukkan pada Persamaan (4) sebagai berikut:

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

- Precision

  Precision mengukur proporsi prediksi positif yang benar terhadap seluruh prediksi positif. Rumus precision ditunjukkan pada Persamaan (5).

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

- Recall

  Recall menunjukkan kemampuan model dalam menemukan seluruh data yang benar-benar positif. Perhitungan recall ditunjukkan pada Persamaan (6).

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

 ISSN:2302-8149
e-ISSN:2540-9719

- F1-Score
  F1-Score merupakan nilai rata-rata harmonik antara precision dan recall. Rumus F1-Score ditunjukkan pada Persamaan (7).

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

Evaluasi ini digunakan untuk menilai performa model secara menyeluruh, terutama pada kondisi data yang tidak seimbang

3.9 Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma klasifikasi berbasis margin yang bertujuan mencari hyperplane optimal untuk memisahkan data antar kelas [7]. Pada penelitian ini digunakan kernel linear, karena sesuai dengan karakteristik data TF–IDF yang bersifat *sparse* dan berdimensi tinggi.

Model SVM dikonfigurasi menggunakan parameter default dari *scikit-learn*, dengan nilai C = 1.0. Untuk klasifikasi multi-kelas, digunakan strategi *One-vs-One* (OvO) yang merupakan mekanisme bawaan dari implementasi SVC. Strategi ini membangun model klasifikasi untuk setiap pasangan kelas sehingga mampu menangani klasifikasi tiga kelas secara efektif [11].

Kombinasi TF–IDF, SMOTE, dan SVM diharapkan mampu menghasilkan model klasifikasi yang stabil serta memiliki performa yang baik dalam mengidentifikasi sentimen pengguna aplikasi CapCut.

4 Hasil dan Pembahasan

Setelah seluruh tahapan penelitian dilakukan, mulai dari proses scraping, pra-pemrosesan data, pelabelan sentimen, pembobotan fitur menggunakan TF–IDF, penyeimbangan data dengan SMOTE, pelatihan model SVM, hingga evaluasi performa, bagian ini menyajikan hasil analisis yang diperoleh dari eksperimen. Penjabaran hasil disusun berdasarkan urutan proses penelitian agar mudah dipahami.

4.1 Scraping Data

Tahap awal dalam penelitian ini adalah pengambilan data ulasan pengguna aplikasi CapCut dari Google Play Store menggunakan teknik *web scraping*. Proses ini menghasilkan sebanyak 5.000 data ulasan mentah yang terdiri dari teks ulasan, rating, dan informasi pendukung lainnya.
Selanjutnya dilakukan proses penyaringan data untuk memastikan kualitas dataset. Tahapan ini meliputi penghapusan data duplikat, penghapusan ulasan kosong, serta penyaringan komentar yang tidak relevan dengan tujuan analisis sentimen. Setelah proses pembersihan dilakukan, jumlah data yang digunakan dalam penelitian ini menjadi 4.381 ulasan.

Data hasil scraping ini kemudian digunakan sebagai dasar dalam tahapan berikutnya, yaitu pra-pemrosesan teks, pelabelan sentimen, pembobotan fitur menggunakan TF–IDF, serta proses klasifikasi menggunakan metode *Support Vector Machine* (SVM). Visualisasi hasil pengambilan data ditampilkan pada Gambar 2.

	Review ID	Username	Rating	Review Text	Date
0	d907ce2a-5e8d-4b92-aa38-47b910e8318d	Pengguna Google	5	kok nya aku gak bisa	2025-12-09 07:07:18
1	879230ad-9333-4c09-9474-d5cc587bd904	Pengguna Google	5	suka banget si tapi seringkali iklan jadi kada...	2025-12-09 07:01:28
2	89f5b36f-8385-46d9-a414-7bce09a3ef18	Pengguna Google	5	Bagus	2025-12-09 07:01:05
3	365d0148-2e92-4493-b30a-9c01e008c474	Pengguna Google	5	bagu bgt	2025-12-09 06:59:34
4	67ed6e26-e874-49b8-8cad-ffc39e7bc4a0	Pengguna Google	2	ga suka Capcut sekarang bnyak iklannya	2025-12-09 06:59:24
5	d3c062b1-7ea5-45d5-a3a5-6d3391dd8cf7	Pengguna Google	5	bagus	2025-12-09 06:55:34
6	8b503738-514d-4f01-8d20-e7e64c526135	Pengguna Google	1	capcut sangat sangat tidak saya sukaaaaaa doni...	2025-12-09 06:52:03
7	f4c195b4-a08b-4111-8f97-be31c2ce4646	Pengguna Google	4	masalahnya nggak ada pfitur diperlambat	2025-12-09 06:46:03
8	55d7e0d5-c078-41dc-a005-acf97708f4c7	Pengguna Google	3	bagus tapi selalu ngebug pas mau masuk	2025-12-09 06:46:00
9	208b2585-45ca-48ef-a2cb-4149ba3061a8	Pengguna Google	5	aplikasi ini sangat luar biasa bagi saya tugas...	2025-12-09 06:45:45

Gambar 2 Hasil scraping ulasan pengguna capcut

4.2 Pra-Pemrosesan Data

Tahap pra-pemrosesan dilakukan untuk menyiapkan data agar dapat diolah secara optimal oleh model machine learning. Proses yang diterapkan meliputi:

 ISSN:2302-8149
e-ISSN:2540-9719

- ***Cleaning***
  Menghapus angka, emoji, hyperlink, tanda baca, dan karakter non-alfabet.
- ***Case Folding***
  Mengubah seluruh teks menjadi huruf kecil.
- **Normalisasi**
  Mengonversi kata tidak baku menjadi bentuk baku.
- ***Tokenization***
  Memecah kalimat menjadi kata-kata.
- ***Stopword Removal***
  Menghapus kata umum yang tidak memiliki makna sentimen.
- ***Stemming***
  Mengembalikan kata berimbuhan ke bentuk dasarnya.

Setelah tahap pra-pemrosesan, seluruh teks menjadi lebih bersih dan siap digunakan untuk representasi fitur. Contoh hasil pra-pemrosesan ditunjukkan pada Gambar 3.

```
<class 'pandas.core.frame.DataFrame'>
Index: 4381 entries, 0 to 4445
Data columns (total 10 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   Date              4381 non-null   object
 1   Username          4381 non-null   object
 2   Rating            4381 non-null   int64
 3   Review Text       4381 non-null   object
 4   cleaning          4381 non-null   object
 5   case_folding      4381 non-null   object
 6   normalisasi       4381 non-null   object
 7   tokenize          4381 non-null   object
 8   stopword removal  4381 non-null   object
 9   stemming_data     4381 non-null   object
dtypes: int64(1), object(9)
memory usage: 376.5+ KB
```

Gambar 3 Hasil pra-pemrosesan data

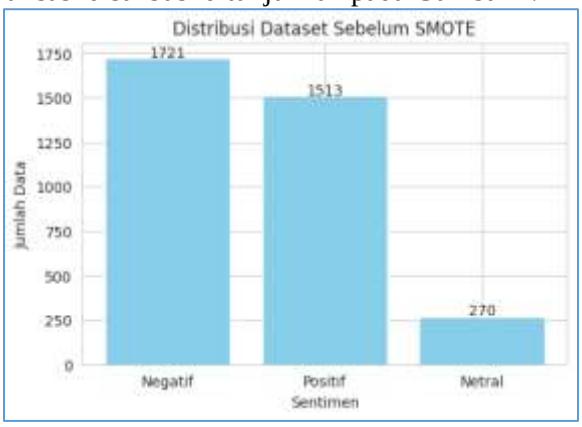
4.3 Pelabelan Sentimen

Pelabelan dilakukan menggunakan pendekatan berbasis rating. Aturan pelabelan sebagai berikut:

Distribusi data sebelum SMOTE adalah:

- Negatif: 1.721 ulasan
- Positif: 1.513 ulasan
- Netral: 270 ulasan

Hasil ini menunjukkan adanya ketidakseimbangan kelas, terutama pada kelas netral yang memiliki jumlah paling sedikit. Visualisasi distribusi ditunjukkan pada Gambar 4.



Gambar 4 Distribusi sentimen sebelum SMOTE

4.4 Visualisasi Word Cloud

Word cloud digunakan untuk melihat kata-kata yang paling sering muncul dalam ulasan pengguna.

 ISSN:2302-8149
e-ISSN:2540-9719

- Sentimen Positif sering menampilkan kata seperti *"bagus", "fitur", "mudah"*.
- Sentimen Negatif didominasi kata *"iklan", "error", "berat"*.
- Sentimen Netral memperlihatkan variasi kata yang lebih datar.

Wordcloud ditunjukkan pada Gambar 5.

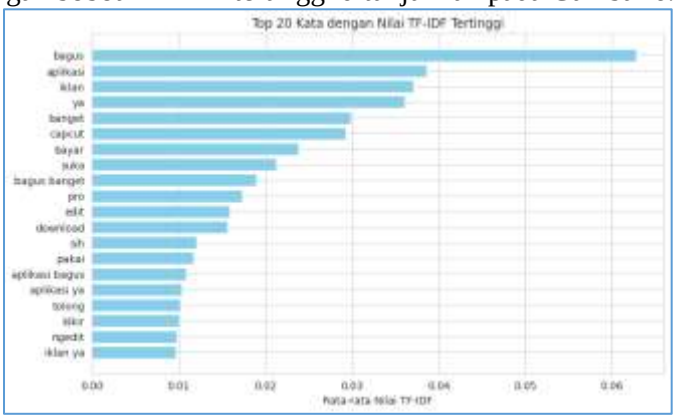
Gambar 5 Word cloud untuk sentimen positif, negatif, dan netral

4.5 Pembobotan Fitur TF–IDF

Representasi fitur dilakukan menggunakan TF–IDF dengan konfigurasi:

- max_features = 5.000
- ngram_range = (1,2)

TF–IDF menghasilkan fitur yang cukup kaya untuk menangkap konteks kata dalam ulasan. Visualisasi 20 kata dengan bobot TF–IDF tertinggi ditunjukkan pada Gambar 6.



Gambar 6 Kata dengan bobot TF–IDF tertinggi

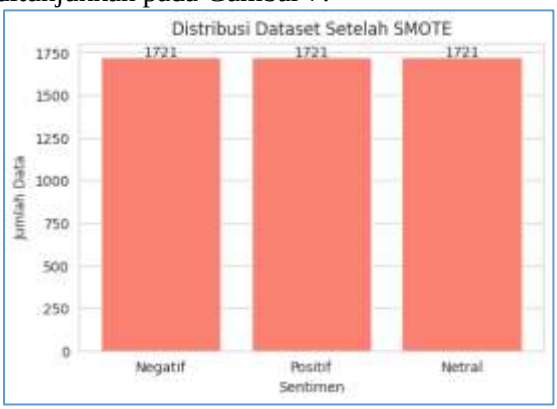
4.6 Split Dataset

Dataset dibagi menjadi dua bagian, yaitu 80% data latih dan 20% data uji. Pembagian dilakukan secara acak untuk memastikan representasi kelas tetap proporsional dan menghindari bias distribusi data. Proporsi ini umum digunakan dalam penelitian analisis sentimen karena memberikan keseimbangan antara kemampuan model mempelajari pola data dan kemampuan generalisasi pada data baru.

4.7 Penyeimbangan Data Menggunakan SMOTE

Sebelum SMOTE, dataset sangat tidak seimbang terutama pada kelas netral. Setelah SMOTE diterapkan, jumlah data pada masing-masing kelas menjadi seimbang sehingga model dapat belajar pola sentimen secara lebih optimal.

Distribusi setelah SMOTE ditunjukkan pada Gambar 7.



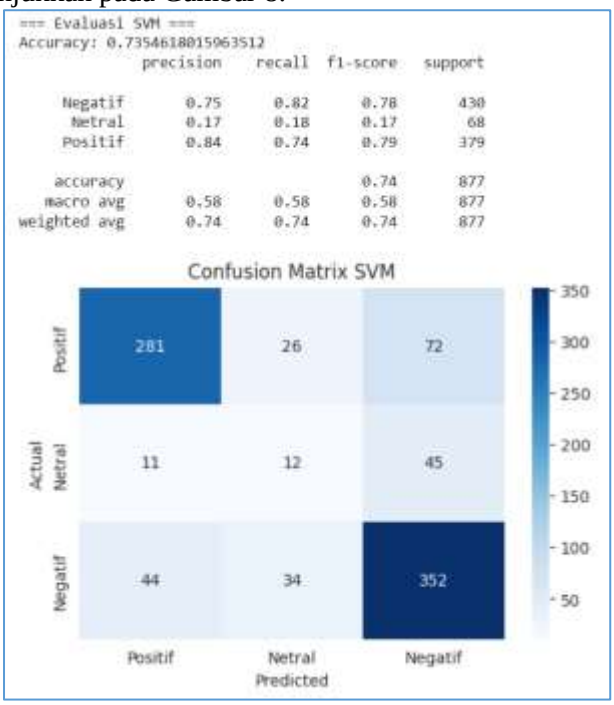
 ISSN:2302-8149
e-ISSN:2540-9719

Gambar 7 Distribusi sentimen setelah SMOTE

4.8 Evaluasi Model SVM

Model utama yang digunakan dalam penelitian ini adalah Support Vector Machine (SVM) dengan kernel linear. Evaluasi dilakukan pada data uji setelah melalui proses prapemrosesan, pembobotan fitur menggunakan TF–IDF, serta penyeimbangan data dengan SMOTE. Berdasarkan hasil pengujian, model SVM berhasil mencapai akurasi sebesar 73,54% dan F1-Score berbobot sebesar 0,736.

Performa model menunjukkan bahwa kelas positif dan negatif dapat dikenali dengan cukup baik, sedangkan kelas netral memiliki tingkat kesalahan prediksi yang lebih tinggi, yang kemungkinan disebabkan oleh pola bahasa yang lebih ambigu dan sulit dibedakan dari dua kelas lainnya. Visualisasi hasil evaluasi model ditunjukkan pada Gambar 8.



Gambar 8 Confusion matrix model SVM

Model cenderung mampu mengenali sentimen positif serta negatif, namun masih mengalami kesulitan dalam membedakan sentimen netral karena ekspresi yang sering kali ambigu dan mirip dengan kategori lainnya.

4.9 Perbandingan Model

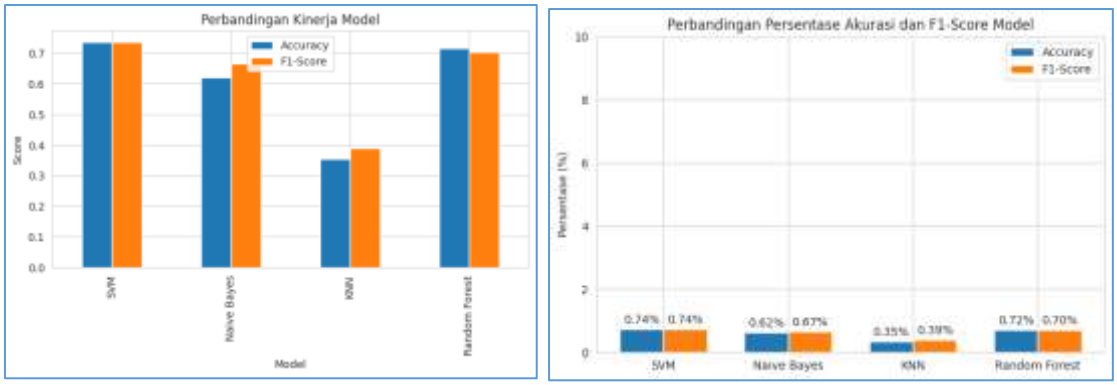
Untuk melihat posisi performa SVM, penelitian ini juga melakukan pengujian terhadap tiga model lain, yaitu Naive Bayes, Random Forest, dan KNN. Hasil evaluasi ditampilkan pada Tabel 1.

Tabel 1 Perbandingan performa model

Model	Akurasi	F1-Score
1. SVM	0.73693	0.73693
2. Random Forest	0.71607	0.70373
3. Naive Bayes	0.62029	0.66548
4. KNN	0.35347	0.39017

Dari tabel di atas, dapat disimpulkan bahwa SVM merupakan model paling unggul, memiliki akurasi dan F1-score tertinggi dibandingkan tiga model lainnya. KNN menunjukkan performa paling rendah, sedangkan Random Forest memberikan hasil kompetitif namun tetap berada di bawah SVM yang ditunjukkan pada Gambar 9.

 ISSN:2302-8149
e-ISSN:2540-9719



Gambar 9 Perbandingan kinerja model dan persentase

5 Kesimpulan

Analisis sentimen terhadap ulasan pengguna aplikasi *CapCut* di *Google Play Store* telah berhasil dilakukan menggunakan metode *Support Vector Machine (SVM)* dengan penerapan *Synthetic Minority Over-sampling Technique (SMOTE).* Proses analisis mencakup tahapan pengumpulan data, pra-pemrosesan teks, representasi fitur menggunakan *TF–IDF*, penyeimbangan data, serta evaluasi model klasifikasi. Hasil pengujian menunjukkan bahwa *model SVM* mencapai akurasi sebesar 73,54% dengan nilai *F1-Score* berbobot sebesar 0,736, di mana penerapan *SMOTE* terbukti membantu meningkatkan keseimbangan distribusi kelas sehingga model mampu mengenali sentimen positif dan negatif dengan cukup baik. Namun demikian, performa pada kelas netral masih relatif lebih rendah, yang diduga disebabkan oleh karakteristik bahasa ulasan yang ambigu dan sulit dibedakan secara tegas. Secara keseluruhan, pendekatan *SVM* dengan *SMOTE* dapat digunakan sebagai metode yang efektif untuk klasifikasi sentimen ulasan aplikasi berbasis teks, serta memberikan gambaran umum mengenai persepsi pengguna terhadap aplikasi *CapCut.* Untuk pengembangan selanjutnya, disarankan penggunaan representasi fitur yang lebih kontekstual atau pendekatan berbasis deep learning guna meningkatkan performa klasifikasi, khususnya pada kelas netral.

Referensi

[1] L. Deng and M. Syafwin, "*Using the Capcut Application as a Learning Media*," 2022.

[2] F. N. Hasan, "Analisis Sentimen Pengguna Aplikasi *CapCut* pada Ulasan di *Play Store* menggunakan Metode *Naïve Bayes*," *Media Online)*, Vol. 4, No. 4, 2024, DOI: 10.30865/klik.v4i4.1555.

[3] M. R. Hanafi and R. K. R, "Analisis Sentimen pada Ulasan Aplikasi Sirekap di *Google Play* menggunakan *Algoritma Naive Bayes*," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, Vol. 4, No. 4, pp. 1578–1586, Oct. 2024, DOI: 10.57152/malcom.v4i4.1693.

[4] S. F. Huwaida, R. Kusumawati, and B. Isnaini, "Analisis Sentimen Komentar YouTube terhadap Pemindahan Ibu Kota Negara menggunakan Metode *Naïve Bayes*," *Jambura Journal of Informatics*, Vol. 6, No. 1, pp. 26–39, Apr. 2024, DOI: 10.37905/jji.v6i1.24718.

[5] A. F. Anjani, D. Anggraeni, and I. M. Tirta, "Implementasi *Random Forest* menggunakan *SMOTE* untuk Analisis Sentimen Ulasan Aplikasi *Sister for Students* UNEJ," *Jurnal Nasional Teknologi dan Sistem Informasi*, Vol. 9, No. 2, pp. 163–172, Sep. 2023, DOI: 10.25077/teknosi.v9i2.2023.163-172.

[6] N. S. Sediatmoko, Y. Nataliani, and I. Suryady, "*Sentiment Analysis of Customer Review using Classification Algorithms and SMOTE for Handling Imbalanced Class*," 2024.

[7] T. A. Khan *et al.*, "*Sentiment Analysis using Support Vector Machine and Random Forest*," *Journal of Informatics and Web Engineering*, Vol. 3, No. 1, pp. 2821–370, 2024, DOI: 10.33093/jiwe.2023.3.1.5.

[8] I. Purnamasari and D. D. Saputra, "Analisis Sentimen dengan Metode *Naïve Bayes*, *SMOTE* dan *Adaboost* pada *Twitter* Bank BTN," *Jurnal Teknologi Informasi dan Komunikasi)*, Vol. 7, No. 2, 2023, DOI: 10.35870/jti.

 ISSN:2302-8149
e-ISSN:2540-9719

[9] P. S. Silitonga, A. R. Tatipang, E. S. -, and A. P. S. -, "Analisis Sentimen Ulasan Pengguna Aplikasi Tiket.Com dengan *K-Nearest Neighbor*," *Jurnal Informatika dan Teknik Elektro Terapan*, Vol. 13, No. 3, Jul. 2025, DOI: 10.23960/jitet.v13i3.7012.

[10] A. Mulianti, Y. Chrisnanto, and H. Ashaury, "Optimalisasi Klasifikasi *Support Vector Machine* dengan *SMOTE*: Studi Kasus Ulasan Pengguna Aplikasi Alfagift," *Jurnal Pekommas*, Vol. 9, No. 2, pp. 249–258, Dec. 2024, DOI: 10.56873/jpkm.v9i2.5583.

[11] D. F. Nawulansih, N. C. Santi, and I. A. Sa'ida, "Analisis Sentimen Ulasan Aplikasi DANA di *Google Play Store*: Penerapan *Support Vector Machine* dan *Synthetic Minority Over-sampling Technique*," *Jurnal Pendidikan dan Teknologi Indonesia*, Vol. 5, No. 9, pp. 2660–2671, Sep. 2025, DOI: 10.52436/1.jpti.1053.

[12] R. Obiedat *et al.*, "*Sentiment Analysis of Customers' Reviews using a Hybrid Evolutionary SVM-based Approach in an Imbalanced Data Distribution*," *IEEE Access*, Vol. 10, pp. 22260–22273, 2022, DOI: 10.1109/ACCESS.2022.3149482.

[13] O. Alvianto and W. Prihartono, "*Journal of Artificial Intelligence and Engineering Applications Implementation of Naive Bayes in Sentiment Analysis of CapCut App Reviews on the Play Store*," 2025. [Online]. Available: https://ioinformatic.org/

[14] L. Ashbaugh and Y. Zhang, "*A Comparative Study of Sentiment Analysis on Customer Reviews using Machine Learning and Deep Learning*," *Computers*, Vol. 13, No. 12, Dec. 2024, DOI: 10.3390/computers13120340.

[15] F. Azhari and R. K. R, "*Sentiment Analysis Towards Full Movie Dirty Vote 2024 in X using Support Vector Machine Method*," *Journal La Multiapp*, Vol. 5, No. 4, pp. 377–387, Aug. 2024, DOI: 10.37899/journallamultiapp.v5i4.1459.

[16] A. N. Ma'aly, D. Pramesti, A. D. Fathurahman, and H. Fakhrurroja, "*Exploring Sentiment Analysis for the Indonesian Presidential Election Through Online Reviews using Multi-Label Classification with a Deep Learning Algorithm*," *Information (Switzerland)*, Vol. 15, No. 11, Nov. 2024, DOI: 10.3390/info15110705.

[17] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiasari, "Analisis Sentimen pada Rating Aplikasi *Shopee* menggunakan Metode *Decision Tree* berbasis *SMOTE*," *AITI: Jurnal Teknologi Informasi*, Vol. 18, No. Agustus, pp. 173–184, 2021.

[18] F. Dwianasari *et al.*, "Analisis Sentimen Masyarakat terhadap Aktivitas Pertambangan di Raja Ampat menggunakan *Support Vector Machine* dan *Naïve Bayes* dengan Teknik *SMOTE*," pp. 234–244, 2025, DOI: 10.61132/keat.v2i2.1208.

[19] K. Pramayasa, I. Md, D. Maysanjaya, G. Ayu, and A. Diatri Indradewi, "Analisis Sentimen Program MBKM pada Media Sosial *Twitter* menggunakan *KNN* dan *SMOTE*", [Online]. Available: https://doi.org/10.31598

[20] A. H. Luthfi, A. Faqih, and G. Dwilestari, "*Journal of Artificial Intelligence and Engineering Applications Enhancing Model Accuracy in Sentiment Analysis of the by.U Application using Naïve Bayes and SMOTE Techniques*," 2025. [Online]. Available: https://ioinformatic.org/

[21] M. Saputra, M. Iqbal, and M. Teknologi Informasi, "*Analysis of SVM Algorithm and SMOTE Technique on Public Sentiment with Google Maps Reviews Towards the Quality of Education at LP3I Across Indonesia*," 2025.