

Ekstraksi Pengetahuan dari Ulasan Aplikasi CapCut Menggunakan Metode *Aspect-Based Sentiment Analysis* dan Klasifikasi

Ishlah Putri Ariyani¹, Ken Ditha Tania^{2*}, Ari Wedhasmara³, Allsela Meiriza⁴

^{1,2,3,4} Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya
Jalan Palembang – Prabumulih Raya KM 32-UNSRI Indralaya, Sumatera Selatan
Email: ¹ishlahariyani@gmail.com, ²kenya.tania@gmail.com, ³a_wedhasmara@unsri.ac.id,
⁴allsela@unsri.ac.id

Abstract. *Knowledge Extraction from CapCut App Reviews Using Aspect-Based Sentiment Analysis and Classification.* Indonesia is experiencing rapid technological development, especially in the use of the internet and editing platforms like CapCut. These platforms enable video editing on various devices; however, user satisfaction is not always guaranteed due to individual differences in experience. This research aims to identify user sentiment towards the CapCut application based on aspects, using an Aspect-Based Sentiment Analysis (ABSA) approach supported by Machine Learning algorithms for the aspect-based sentiment classification task. The algorithm used in the classification process is Support Vector Machine. The data used are reviews of the CapCut application from the Google Play Store, with a total of 22,668 data points. The results show that the Support Vector Machine (SVM) algorithm performs well in each aspect, with accuracy values of 0.88 for the feature aspect and 0.87 for the user experience aspect. The results of knowledge extraction are obtained in the form of XML, which contains user sentiment information on two main aspects: features and user experience.

Keywords: *ABSA, CapCut, knowledge extraction, classification*

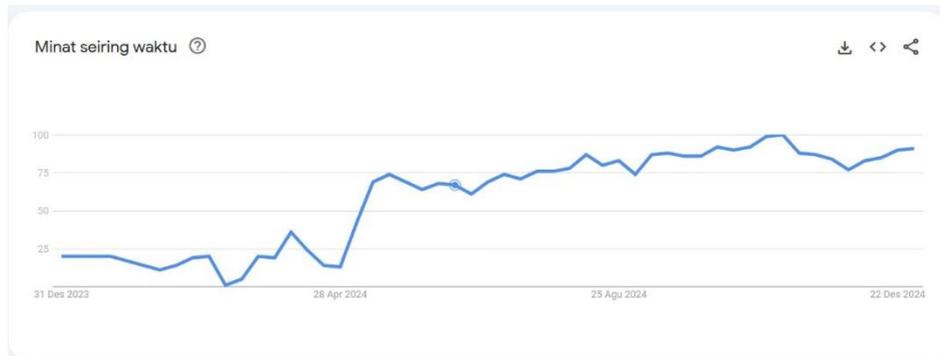
Abstrak. *Indonesia mengalami perkembangan teknologi yang pesat, khususnya dalam penggunaan internet dan platform editing seperti CapCut. Platform ini memungkinkan pengeditan video di berbagai perangkat, namun kepuasan pengguna tidak selalu terjamin karena perbedaan pengalaman individu. Penelitian ini bertujuan untuk mengidentifikasi sentimen pengguna terhadap aplikasi CapCut berdasarkan aspek. Dengan menggunakan pendekatan Aspect-Based Sentiment Analysis (ABSA) yang didukung oleh algoritma Machine Learning untuk tugas klasifikasi sentimen berdasarkan aspek. Algoritma yang digunakan dalam proses klasifikasi adalah Support Vector Machine. Data yang digunakan adalah ulasan aplikasi CapCut dari Google Play Store sebanyak 22.668 data. Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki performa yang baik untuk masing-masing aspek dengan nilai akurasi untuk aspek fitur 0,88 dan aspek user experience 0,87. Hasil ekstraksi pengetahuan yang diperoleh berupa XML yang memuat informasi sentimen pengguna terhadap dua aspek utama, yaitu fitur dan user experience.*

Kata Kunci: *ABSA, CapCut, ekstraksi pengetahuan, klasifikasi*

1. Pendahuluan

Indonesia mengalami perkembangan teknologi yang signifikan, terutama terhadap penggunaan internet yang meluas, terlebih dalam bidang teknologi kreatif seperti penggunaan *platform editing*. *Platform editing* menjadi salah satu faktor penting untuk memenuhi kebutuhan kreatif para pengguna daring. *Platform editing* merupakan aplikasi atau perangkat lunak yang menyediakan *tools* dan fitur untuk mengedit konten digital seperti gambar, teks, audio maupun video. *Platform editing* yang semakin populer saat ini adalah CapCut. CapCut menempati urutan kedua dalam kategori "*Photo & Video*" dan masuk 12 besar aplikasi "*Top Free*" di Google Play Store, dengan lebih dari 1 miliar unduhan yang mencerminkan tingginya adopsi pengguna.

CapCut yang dikembangkan oleh *ByteDance* telah menarik perhatian khalayak ramai berkat fitur-fitur inovatifnya.



Gambar 1. Statistik Pengguna Aplikasi CapCut Tahun 2024

CapCut adalah sebuah platform kreatif yang komprehensif, yang memungkinkan pengguna mengedit video lewat *browser*, iOS, Android, dan perangkat terhubung lainnya. CapCut menyediakan serangkaian fitur menarik yang dapat dimanfaatkan untuk mengedit video, seperti menambahkan musik, teks, stiker, serta efek dan filter untuk memperindah tampilan, menghapus latar belakang, menggunakan *green screen*, mengubah ukuran dan durasi video, serta *text-to-speech* atau sebaliknya yang mendukung beragam bahasa [1]. Melonjaknya popularitas penggunaan aplikasi CapCut mempengaruhi preferensi pengguna terhadap perubahan dan peningkatan fitur yang ada. Aplikasi CapCut pun tidak mampu menjamin kepuasan penggunaannya, sebab setiap pengguna memiliki sudut pandangan serta pengalaman yang berbeda-beda.

Selain CapCut, terdapat pula platform penyuntingan video lainnya seperti VN, yang dikenal dengan fitur-fitur canggih seperti pengeditan *multi-layer*, penyesuaian efek, serta dukungan terhadap berbagai format video. Namun, perbedaan utama antara CapCut dan VN terletak pada karakteristik ulasan penggunaannya. Ulasan terhadap CapCut cenderung lebih ekspresif dan subjektif, dengan penekanan pada kenyamanan, inovasi fitur, serta pengalaman personal dalam memanfaatkan fitur kreatif yang memudahkan proses pembuatan konten, khususnya untuk media sosial sebaliknya, ulasan pengguna terhadap aplikasi VN lebih banyak menyoroti aspek teknis, seperti performa aplikasi yang berat, kendala saat *rendering*, dan tuntutan spesifikasi sistem operasi yang tinggi [2]. Menurut penelitian [3], aplikasi CapCut dinilai lebih unggul dibandingkan VN dalam beberapa aspek. Tampilan antarmuka CapCut yang intuitif dan dirancang untuk kemudahan penggunaan membuatnya lebih disukai oleh sebagian besar pengguna. Pengguna mengapresiasi CapCut karena kemudahan navigasi dan pengoperasiannya yang sederhana, yang tidak selalu ditemukan pada VN, yang memiliki kurva pembelajaran sedikit lebih tinggi. Fakta ini menegaskan bahwa preferensi pengguna tidak hanya dipengaruhi oleh kecanggihan fitur teknis, tetapi juga oleh kenyamanan dan pengalaman personal dalam menggunakan aplikasi tersebut.

Untuk memahami lebih dalam mengenai bagaimana pengguna merespons dan menilai fitur maupun pengalaman mereka dalam menggunakan aplikasi seperti CapCut, diperlukan pendekatan yang mampu menganalisis opini secara sistematis. Salah satu pendekatan tersebut adalah analisis sentimen, yakni metode untuk mengeksplorasi opini atau teks yang diperoleh dari berbagai platform media sosial, menggunakan pembelajaran mesin untuk analisisnya [4]. Analisis sentimen cocok untuk mengamati bagaimana masyarakat, terutama pengguna aplikasi CapCut, merespons kebijakan tersebut [5].

Kualitas layanan aplikasi dapat dievaluasi dengan memanfaatkan data mining melalui analisis sentimen terhadap ulasan yang diterima secara massal di berbagai platform *online*, khususnya di Play Store. Analisis ini diperlukan untuk memperoleh informasi berharga terkait persepsi dan tingkat kepuasan pengguna terhadap aplikasi CapCut [5]. Meski Play Store memberikan gambaran awal melalui rating dan ulasan, informasi tersebut sering kali tidak

mencakup konteks rinci mengenai aspek-aspek penting yang relevan untuk mengevaluasi pengalaman pengguna. Untuk mengatasi kekurangan ini, analisis sentimen berbasis aspek dapat digunakan sebagai solusi untuk memahami opini dan persepsi pengguna terhadap aplikasi tersebut [6]. Penting untuk menganalisis persepsi dan kepuasan pengguna terhadap CapCut guna mengidentifikasi aspek yang menjadi keunggulan utama serta kelemahan yang perlu diperbaiki. Analisis ini diharapkan bisa memberikan penjelasan yang lebih jelas kepada para pengembang untuk memperbaiki mutu aplikasi dan menjaga kesetiaan pengguna di tengah persaingan yang semakin sengit.

Penelitian ini bertujuan untuk mengevaluasi tanggapan pengguna mengenai aplikasi CapCut berdasarkan aspek dengan memperhatikan aspek menggunakan algoritma *Support Vector Machine* (SVM). Proses ini menghasilkan ekstraksi pengetahuan berupa analisis sentimen yang memetakan opini pengguna terhadap aspek fitur dan *user experience*. Proses ini menghasilkan ekstraksi pengetahuan dalam bentuk analisis sentimen berbasis aspek yang memetakan opini pengguna terhadap masing-masing aspek. Namun, ekstraksi pengetahuan dalam penelitian ini tidak terbatas pada pelaporan hasil klasifikasi model. Penelitian ini mengonversi hasil klasifikasi sentimen menjadi informasi terstruktur dalam format XML, yang menyajikan metadata sentimen berdasarkan aspek. Format ini memudahkan pengembang dan pemangku kepentingan untuk mengakses, meninjau ulang, dan memanfaatkan hasil analisis secara lebih sistematis. Dengan demikian, hasil analisis tidak hanya berupa angka atau kategori sentimen, tetapi juga menjadi bentuk pengetahuan baru yang dapat digunakan untuk mendukung pengambilan keputusan strategis serta peningkatan kualitas layanan aplikasi. Hal ini membedakan ekstraksi pengetahuan dalam penelitian ini dari sekadar dokumentasi hasil pengujian model.

Proses ekstraksi pengetahuan memainkan peran kunci dalam mengelola informasi besar yang tidak terstruktur, yang perlu diubah menjadi pengetahuan yang berguna dan dapat diproses secara komputer [7]. Dalam konteks ini, format XML dipilih karena sifatnya yang terstruktur, fleksibel, dan mudah dipahami, baik oleh sistem maupun oleh manusia. Representasi ini diharapkan dapat memberikan nilai tambah bagi pengembang maupun perusahaan dalam merumuskan kebijakan dan strategi pengembangan aplikasi CapCut ke depannya dan informasi ini sangat berguna bagi perusahaan dalam pengambilan keputusan strategis di masa depan [8].

Sampai saat ini, banyak penelitian terkait analisis sentimen diantaranya oleh Hasan [1], Analisis dalam studi ini memanfaatkan algoritma *Naïve Bayes* yang menghasilkan matriks kebingungan, tingkat akurasi mencapai 84,09%, presisi berada di angka 91,91%, dan tingkat recall sebesar 73,53%. Penelitian terkait analisis sentimen aplikasi CapCut oleh Zai dan Isnain [9], analisis sentimen aplikasi CapCut dilakukan dengan memanfaatkan dua algoritma, yaitu *Naïve Bayes* dan *Support Vector Machine* (SVM), menggunakan teknik *Synthetic Minority Oversampling Technique* (SMOTE). Hasil optimasi menunjukkan bahwa model SVM mengalami peningkatan kinerja setelah penerapan SMOTE. Sebaliknya, model *Naïve Bayes* menunjukkan penurunan dalam nilai presisi.

Penelitian berikutnya oleh Pandunata, Nurdiansyah dan Alfina [10], penelitian ini menggunakan metode *Aspect-Based Sentiment Analysis* (ABSA) dan menggunakan metode SMOTE. Hasil studi ini menunjukkan bahwa dengan memanfaatkan SMOTE, algoritma SVM menjadi yang paling efektif dalam setiap aspek yang telah ditentukan. Berdasarkan penelitian oleh Ramadina dan Tania [11] dinyatakan bahwa pengguna memiliki tingkat kepuasan yang cukup terhadap perilaku pengemudi. Hasil dari penerapan algoritma pada analisis sentimen berbasis aspek untuk ulasan aplikasi Gojek tidak menunjukkan hasil yang memuaskan, di mana pengambilan pengetahuan mengenai pengguna aplikasi Gojek untuk faktor harga cenderung menghasilkan lebih banyak sentimen negatif dibandingkan yang positif. Ini dapat diartikan bahwa pengguna merasa bahwa tarif untuk layanan Gojek cukup mahal. Penelitian ini memberikan wawasan mengenai efektivitas algoritma SVM dalam memahami persepsi dan kepuasan pengguna dengan memanfaatkan ABSA pada tanggapan aplikasi CapCut.

2. Tinjauan Pustaka

2.1. Ekstraksi Pengetahuan

Ekstraksi pengetahuan adalah proses mendapatkan informasi dan relasinya, menggeneralisasi informasi tersebut, dan menyimpannya dalam format terstruktur seperti XML atau basis pengetahuan. Dengan menyimpan hasil ekstraksi dalam format XML, informasi tersebut dapat diorganisir secara sistematis sehingga memudahkan pengembang atau pihak tertentu dalam mengakses data dan memahami pola serta tren yang muncul. XML di pilih karena kemampuannya dalam mendukung struktur data yang kompleks dan standarisasi yang di perlukan untuk pertukaran data antar sistem. Pengetahuan yang di ekstraksi harus dalam format yang dapat dipahami oleh mesin dan memungkinkan mesin untuk melakukan inferensi. Ekstraksi Pengetahuan dapat memanfaatkan teknik ekstraksi informasi yang bertujuan untuk mengekstraksi informasi eksplisit dengan kategori tertentu dari kumpulan dokumen [7], [12].

2.2. Aplikasi CapCut

Aplikasi CapCut adalah aplikasi pengeditan video yang dirancang untuk pengguna android ataupun iOS yang dikembangkan oleh *ByteDance*, perusahaan yang juga mengembangkan TikTok. CapCut menawarkan beragam fungsi menarik yang bisa dimanfaatkan untuk mengolah video demi berinovasi [1]. Aplikasi CapCut menyediakan sejumlah fitur, seperti pemangkasan video, penambahan elemen visual, perpindahan, tulisan, dan iringan musik, yang ditujukan untuk mendukung pengguna dalam menciptakan konten yang menarik perhatian [9].

2.3. Text Mining

Metode untuk memperoleh data dan pola yang tidak terduga, sebelumnya tidak teridentifikasi, dan memiliki nilai potensial dari sekumpulan besar data yang tidak terorganisir, misalnya, teks yang ditulis dengan bahasa alami disebut *text mining*. Di mana, dinyatakan bahwa *sentiment classification* dan *opinion classification* merupakan dua pendekatan yang sering digunakan dalam *text mining* untuk menentukan sejauh mana suatu teks mendukung atau menentang suatu subjek.

2.4. Aspect-Based Sentiment Analysis (ABSA)

Pemanfaatan pendekatan analisis sentimen memungkinkan pemahaman yang lebih mudah terhadap bagaimana persepsi atau respons masyarakat terhadap suatu produk atau layanan berdasarkan ulasan maupun komentar yang mereka berikan. Menurut Alqaryouti, Siyam, Monem dan Shaalan [13], *Aspect-Based Sentiment Analysis* (ABSA) adalah proses di mana sentimen terhadap berbagai aspek dari suatu entitas yang diekstraksi dan diklasifikasikan sehingga dapat memberikan wawasan yang lebih mendalam mengenai opini pengguna. Dengan menggunakan ABSA dapat memberikan wawasan yang lebih mendalam mengenai hal yang disukai atau sebaliknya oleh pengguna mengenai masing-masing aspek.

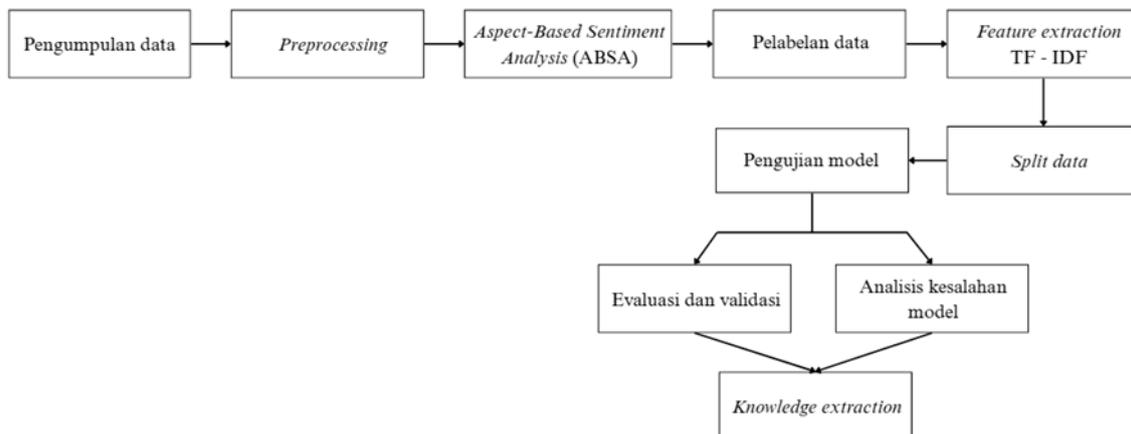
Untuk klasifikasi sentimen berbasis aspek, penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) yang dikenal efektif dalam tugas pengklasifikasian teks dengan performa yang baik dalam menangani data ulasan. SVM bekerja dengan mencari *hyperplane* terbaik yang memisahkan kelas sentimen secara optimal dalam ruang fitur, sehingga mampu menghasilkan model yang akurat dan stabil dalam memprediksi sentimen berdasarkan fitur teks ulasan. Dengan menggabungkan ABSA dan SVM, penelitian ini berhasil mengekstraksi pengetahuan yang lebih detail tentang opini pengguna terhadap aplikasi CapCut, sehingga hasilnya dapat digunakan untuk memberikan masukan strategis dalam pengembangan produk dan peningkatan kualitas layanan.

2.5 Machine Learning

Machine learning adalah sebuah metode dalam kecerdasan buatan yang dirancang untuk meniru cara manusia berpikir dan bertindak dalam memecahkan masalah [14]. Dalam konteks analisis sentimen, *supervised learning* banyak digunakan karena data ulasan biasanya sudah memiliki label sentimen (positif, negatif, netral) yang dapat dijadikan acuan pelatihan. Algoritma populer yang sering dipakai antara lain *Support Vector Machine* (SVM), *Naïve Bayes*, *Decision Tree*, dan *Random Forest*.

3. Metodologi Penelitian

Penelitian ini dilakukan dengan menggunakan Google Colab untuk membantu dalam proses pengolahan data dan klasifikasi. Pada Gambar 2 menunjukkan tahapan penelitian yang telah dikembangkan berdasarkan penelitian sebelumnya.



Gambar 2. Tahapan Penelitian

3.1. Pengumpulan Data

Tahapan ini dimulai dengan pengumpulan informasi menggunakan Google Colab bahasa pemrograman python, melalui metode *scraping* data. Informasi yang diperoleh berasal dari penilaian pengguna aplikasi CapCut di situs web Google Play Store. Pengumpulan data dilakukan pada tanggal 01 juni 2024 sampai dengan 30 juni 2024 yang diperoleh sebanyak 21.668 data.

3.2. Preprocessing

Tahapan *preprocessing* digunakan untuk menyiapkan data yang sudah dibersihkan dari data yang belum tersistematisasi. Tahapan *preprocessing* data sebagai berikut.

3.2.1. Cleaning

Cleaning merupakan proses pembersihan data seperti yang akan digunakan, memperbaiki kesalahan pengetikan ulasan dengan *slang*, menghapus tanda baca dan *emoji*, menghapus data yang terduplikat serta menghapus *missing values*. Tabel 1 menunjukkan hasil proses *cleaning* tahapan ini.

Tabel 1. Hasil Cleaning

Komponen	Sebelum	Sesudah
Slang	bgt	Banget
Slang	klo	kalau
Slang	pakek	pake
Emoji	👉	(dihapus)
Tanda Baca	, +	(dihapus)

3.2.2. Case Folding

Case folding merupakan suatu tahapan dalam mengonversi semua huruf besar menjadi huruf kecil dengan cara yang sistematis, untuk memastikan konsistensi dalam pengolahan teks. Dalam penelitian ini, kata-kata yang memiliki huruf besar akan diubah menjadi huruf kecil. Berikut pada Tabel 2 menunjukkan hasil dari proses tahapan *case folding* ini.

Tabel 2. Hasil Case Folding

Sebelum	Sesudah
Suka banget kalo pake capcut tanpa iklan seperti dulu meng ekspor video aja tanpa iklandata sekarang mah harus pake iklan datanya mohon ganti capcut yang dulu aja biar mudah meng ekspor	suka banget kalo pake capcut tanpa iklan seperti dulu meng ekspor video aja tanpa iklandata sekarang mah harus pake iklan datanya mohon ganti capcut yang dulu aja biar mudah meng ekspor

3.2.3. Tokenize

Tokenisasi dalam pra-pemrosesan data teks melibatkan pemecahan *string* yang lebih panjang menjadi token yang lebih kecil. Tabel 3 menunjukkan hasil dari proses *tokenize*.

Tabel 3. Hasil Tokenize

Sebelum	Sesudah
suka banget kalo pake capcut tanpa iklan seperti dulu meng ekspor video aja tanpa ikldata sekarang mah harus pake iklan datanya mohon ganti capcut yang dulu aja biar mudah meng ekspor	['suka', 'banget', 'kalo', 'pake', 'capcut', 'tanpa', 'iklan', 'seperti', 'dulu', 'meng', 'ekspor', 'video', 'aja', 'tanpa', 'ikldata', 'sekarang', 'mah', 'harus', 'pake', 'iklan', 'datanya', 'mohon', 'ganti', 'capcut', 'yang', 'dulu', 'aja', 'biar', 'mudah', 'meng', 'ekspor']

3.2.4. Filtering

Tahap berikutnya adalah proses penyaringan, di mana peneliti melakukan penghapusan kata-kata yang tidak signifikan. Proses *stopword removal* ini menggunakan *stopwords* dari NLTK. Tabel 4 menunjukkan hasil *stopword removal*.

Tabel 4. Hasil Stopword Removal

Sebelum	Sesudah
['suka', 'banget', 'kalo', 'pake', 'capcut', 'tanpa', 'iklan', 'seperti', 'dulu', 'meng', 'ekspor', 'video', 'aja', 'tanpa', 'ikldata', 'sekarang', 'mah', 'harus', 'pake', 'iklan', 'datanya', 'mohon', 'ganti', 'capcut', 'yang', 'dulu', 'aja', 'biar', 'mudah', 'meng', 'ekspor']	['suka', 'banget', 'kalo', 'pake', 'capcut', 'iklan', 'meng', 'ekspor', 'video', 'aja', 'ikldata', 'mah', 'pake', 'iklan', 'datanya', 'mohon', 'ganti', 'capcut', 'aja', 'biar', 'mudah', 'meng', 'ekspor']

3.2.5. Stemming

Stemming adalah proses mengurangi kata-kata yang mengalami infleksi atau derivasi ke bentuk dasar atau akar katanya. Proses ini juga dapat meningkatkan akurasi model klasifikasi dengan menghilangkan kata-kata yang berlebihan. Dalam proses *stemming* ini juga menggunakan *Library* sastra *stemming* bahasa Indonesia. Tabel 5 menunjukkan *output* proses *stemming*.

Tabel 5. Hasil Stemming

Sebelum	Sesudah
['suka', 'banget', 'kalo', 'pakek', 'capcut', 'iklan', 'meng', 'ekspor', 'video', 'aja', 'ikldata', 'mah', 'pakek', 'iklan', 'datanya', 'mohon', 'ganti', 'capcut', 'aja', 'biar', 'mudah', 'meng', 'ekspor']	suka banget kalo pake capcut iklan meng ekspor video aja ikldata mah pake iklan data mohon ganti capcut aja biar mudah meng ekspor

3.3. Aspect-Based Sentiment Analysis (ABSA)

Tahapan prapemrosesan diikuti dengan identifikasi aspek sentimen dalam ulasan pengguna karena studi ini menggunakan pendekatan *Aspect-Based Sentiment Analysis* (ABSA). Dua aspek yang di analisis yaitu, fitur dan *user experience* yang mengandung kata-kata positif dan negatif. Tabel 6 menunjukkan aspek-aspek ABSA.

Tabel 6. Aspek-Aspek ABSA

No.	Aspek	Tipe	Kata
1.	Fitur	Positif	Bagus, mantap, mudah, keren, lancar, recommended, baik, good, menarik, seru, best, lengkap, sip.
		Negatif	Jelek, kurang, bug, loading, ribet, macet, lag, error, lama, lemot, gajelas, bad, ngelek, buruk.
2.	User Experience	Positif	Suka, puas, terbaik, bantu, enak, guna, ok, lima, senang, manfaat, gampang, cocok, lumayan, apik, wow.
		Negatif	Iklan, Kesel, Ganggu, kebanyakan, bayar, susah, ekspor, pro, update, kecewa, gabisa

3.4. Pelabelan Data

Aspek ditentukan menggunakan metode *Aspect-Based Sentiment Analysis* (ABSA), setiap ulasan diberi label sentimen positif atau negatif berdasarkan konteks ulasan terhadap aspek fitur dan *user experience*. Untuk merepresentasikan hal ini, dilakukan transformasi data menggunakan metode *explode*, yaitu memecah data berdasarkan aspek. Dengan metode ini, satu komentar dapat menghasilkan lebih dari satu baris data, sesuai jumlah aspek yang dikenali. Proses ini memastikan setiap baris hanya memuat satu aspek dan satu label sentimen, sesuai pendekatan ABSA yang benar.

Tabel 7 menyajikan tiga contoh data yang telah melalui tahap pelabelan menggunakan pendekatan *Aspect-Based Sentiment Analysis* (ABSA). Proses pelabelan menghasilkan sebanyak 14.469 data, yang merupakan hasil ekstraksi aspek dan sentimen dari setiap ulasan yang telah dianalisis. Penentuan sentimen dilakukan berdasarkan konteks kemunculan kata positif atau negatif terhadap masing-masing aspek. Melalui proses ini, satu komentar dapat menghasilkan lebih dari satu pasangan (komentar, aspek, sentimen) dengan label sentimen berbeda. Hal ini memastikan pendekatan yang dilakukan merupakan *Aspect-Based Sentiment Analysis* (ABSA), bukan hanya analisis sentimen umum.

Tabel 7. Hasil Pelabelan Data

Review	Aspek Positif	Aspek Negatif	Total Positif	Total Negatif	% Positif	% Negatif	Sentimen
fitur editnya lancar saja	Fitur	-	1	0	100.0	0	Positif
lumayan bagus sih kadang lag video	User Experience, Fitur	Fitur	2	1	66.67	33.33	Positif
error lag sinyal bagus ekspor lambat sinyal bagus update bug ny	Fitur	User Experience, Fitur	1	5	16.67	83.33	Negatif

3.5. Feature Extraction

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan teknik merepresentasikan teks menjadi fitur numerik berdasarkan frekuensi kata (*Term Frequency*, *TF*) dan seberapa unik kata tersebut dalam seluruh dokumen (*Inverse Document Frequency*, *IDF*). Metode ini memberikan nilai pada setiap istilah dalam korpus untuk meningkatkan efektivitas model pembelajaran [15]. Penelitian ini menggunakan 'TfidfVectorizer()' untuk mengubah teks menjadi fitur berbasis TF-IDF.

3.6. Split Data

Data yang telah direpresentasikan melalui TF-IDF kemudian dibagi menjadi dua kelompok, yakni kelompok pelatihan dan kelompok pengujian [13]. Karena pendekatan yang digunakan adalah *Aspect-Based Sentiment Analysis* (ABSA), maka data dibagi secara terpisah berdasarkan aspek. Setiap model klasifikasi dibangun dan diuji untuk masing-masing aspek, yaitu *Fitur* dan *User Experience*. Hal ini memastikan bahwa analisis sentimen dilakukan secara spesifik terhadap setiap aspek, bukan sebagai sentimen umum terhadap keseluruhan komentar. Peneliti membagi data 80% untuk data pelatihan 20% data pengujian. Tabel 8 merupakan pembagian data yang digunakan untuk setiap aspek.

Tabel 7. Split Data

Aspek	Data Pelatihan	Data Pengujian
Fitur	6516	1629
User Experience	5059	1265

3.7. Model Testing

Tahapan pengujian dilaksanakan dengan mengimplementasikan algoritma klasifikasi, yakni *Support Vector Machine* (SVM) yang dinilai menggunakan metrik akurasi, presisi, *recall*, dan AUC-ROC untuk mengamati kinerjanya klasifikasi berdasarkan sentimen yang telah dilabeli. Selain itu, dilakukan pula analisis kesalahan klasifikasi menggunakan *confusion matrix* guna mengetahui jumlah prediksi yang benar dan salah pada tiap kategori sentimen.

3.8. Knowledge Extraction

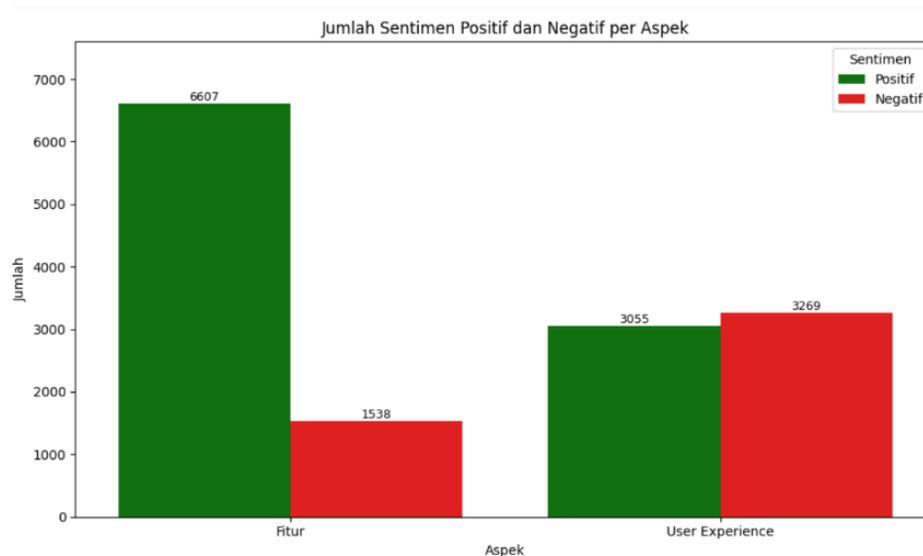
Hasil akhir dari analisis kemudian diekstraksi sebagai pengetahuan yang dimanfaatkan dalam pengambilan keputusan atau pengembangan sistem. Informasi yang dihasilkan dari analisis tersebut disusun secara sistematis dan direpresentasikan dalam format XML.

4. Hasil dan Diskusi

4.1 Analisis Sentimen

Analisis sentimen yang dilakukan menggunakan metode klasifikasi dengan memanfaatkan *Aspect-Based Sentiment Analysis* (ABSA) berhasil mengklasifikasikan data ke dalam dua aspek yaitu fitur dan *user experience*. Proses ini memanfaatkan pendekatan berbasis kamus kata kunci yang telah ditentukan sebelumnya untuk masing-masing aspek dan polaritas sentimen (positif atau negatif). Dalam proses pelabelan, jika sebuah ulasan mengandung lebih banyak kata kunci positif daripada negatif, maka sistem menetapkan label “Positif” sebagai sentimen dominan, dan sebaliknya. Namun, untuk memastikan pendekatan yang digunakan benar-benar berbasis aspek (ABSA), penelitian ini tidak langsung menggunakan label sentimen secara keseluruhan untuk satu komentar. Sebaliknya, setiap komentar dianalisis dan diekstraksi aspek-aspeknya secara terpisah, baik untuk aspek fitur maupun *user experience*.

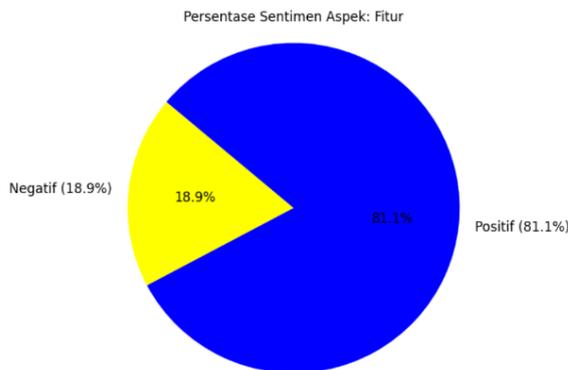
Melalui proses ini, satu komentar dapat memiliki lebih dari satu label sentimen, bergantung pada konteks dan aspek yang dibahas. Misalnya, sebuah ulasan dapat menyatakan kepuasan terhadap fitur tetapi juga menyampaikan kekecewaan terhadap pengalaman pengguna. Untuk itu, ulasan tersebut akan diekspansi (*explode*) menjadi dua entri: satu entri untuk aspek Fitur dan satu entri untuk aspek User Experience, masing-masing dengan label sentimen yang sesuai. Dengan demikian, pendekatan yang digunakan tetap mengikuti kaidah analisis sentimen berbasis aspek (ABSA), bukan sekadar analisis sentimen umum.



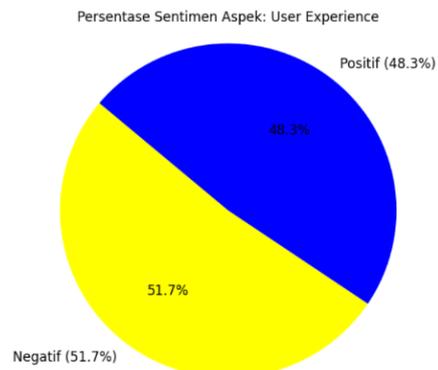
Gambar 3. Sentimen Berdasarkan Aspek

Gambar 3 memvisualisasikan hasil sentimen berdasarkan dua aspek dari proses analisis sentimen yang sudah dilakukan. Pada hasil tersebut didapatkan sentimen positif dan negatif pada

masing-masing aspek. Aspek fitur didapatkan 6.607 data dengan sentimen positif dan 1.528 data dengan sentimen negatif, sedangkan aspek *user experience* didapatkan 3.055 data dengan sentimen positif dan 3269 data dengan sentimen negatif. Adapun pada Gambar 3 dan Gambar 4 merupakan visualisasi dari proposi sentimen yang digunakan untuk mencari aspek dominan berdasarkan presentase positif dan negatif pada masing masing aspek. Berdasarkan hasil dari Gambar 4 dan Gambar 5 dapat disimpulkan bahwa fitur aplikasi CapCut menunjukkan dominasi sentimen positif dengan persentase sebesar 81.1%, yang mengindikasikan bahwa mayoritas pengguna merasa puas terhadap fitur-fitur yang ditawarkan oleh aplikasi.



Gambar 4. Pie Chart Sentimen Aspek Fitur



Gambar 5. Pie Chart Sentimen Aspek User Experience

4.2 Evaluasi Model

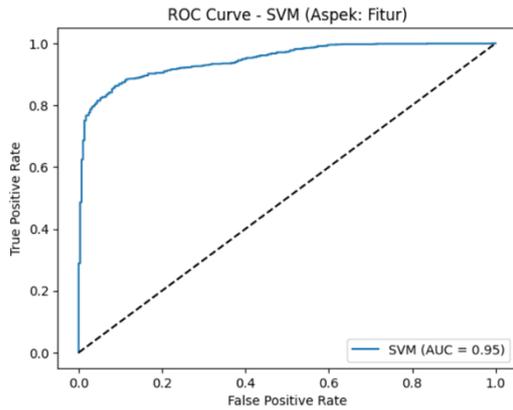
Evaluasi model dilakukan menggunakan algoritma Support Vector Machine (SVM) sebagai salah satu metode klasifikasi yang efektif dalam menangani data teks, terutama dalam tugas klasifikasi biner seperti analisis sentimen. Pendekatan yang digunakan tetap mengacu pada kerangka kerja *Aspect-Based Sentiment Analysis* (ABSA), sehingga evaluasi model dilakukan secara terpisah berdasarkan masing-masing aspek, yaitu Fitur dan *User Experience*. Penilaian performa model dievaluasi menggunakan beberapa metrik penting dari *classification report*, yakni akurasi, presisi, *recall*, *F1-score*, dan AUC-ROC.

Tabel 8. Hasil *Classification Model*

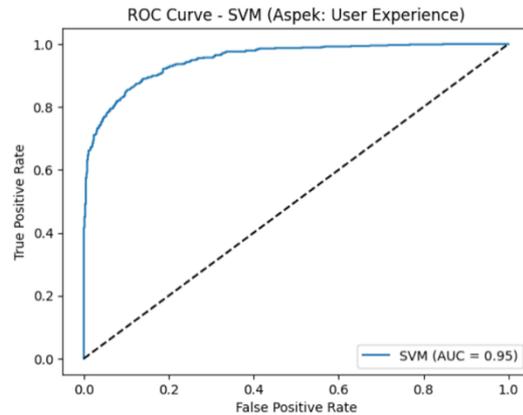
Aspek	Akurasi	Presisi	Recall	F1-Score	AUC-ROC	Cross Validation Mean Accuracy
Fitur	0,887	0,89	0,89	0,89	0,95	0,884
User Experience	0,875	0,88	0,88	0,88	0,95	0,885

Berdasarkan Tabel 8, hasil evaluasi model SVM menunjukkan performa yang baik pada aspek fitur dan *user experience* dengan akurasi sekitar 87,5% – 88,7%. Nilai presisi, *recall*, dan *F1-score* yang seimbang mengindikasikan kemampuan model dalam mengklasifikasikan sentimen positif dan negatif secara tepat. AUC-ROC sebesar 0,95 memperlihatkan kemampuan model membedakan kelas sentimen dengan baik. Rata-rata akurasi *cross-validation* di atas 88% menunjukkan model cukup stabil dan tidak *overfit*. Secara keseluruhan, metode SVM dengan pendekatan ABSA efektif untuk analisis sentimen berdasarkan aspek. Untuk memperkuat evaluasi performa model, dilakukan visualisasi kurva *Receiver Operating Characteristic* (ROC) pada masing-masing aspek, yaitu fitur dan *user experience*.

Gambar 6 dan Gambar 7 memperlihatkan kurva ROC untuk aspek fitur dan *user experience* secara berurutan. Terlihat bahwa kedua aspek memiliki nilai AUC yang tinggi, mendekati 0,95 yang mengindikasikan model memiliki performa klasifikasi yang sangat baik untuk kedua aspek tersebut. Visualisasi ini mendukung metrik evaluasi numerik yang telah disajikan sebelumnya dan memperkuat kesimpulan bahwa metode SVM dengan pendekatan ABSA efektif dalam analisis sentimen berbasis aspek.



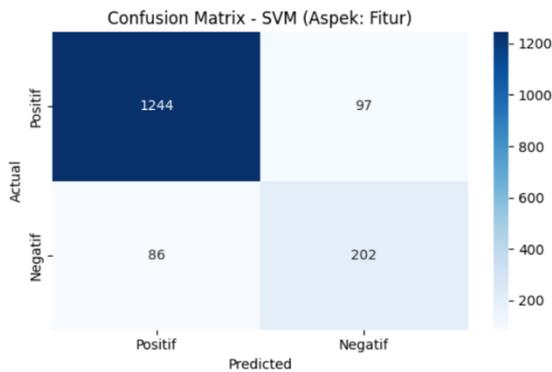
Gambar 6. AUC-ROC Aspek Fitur



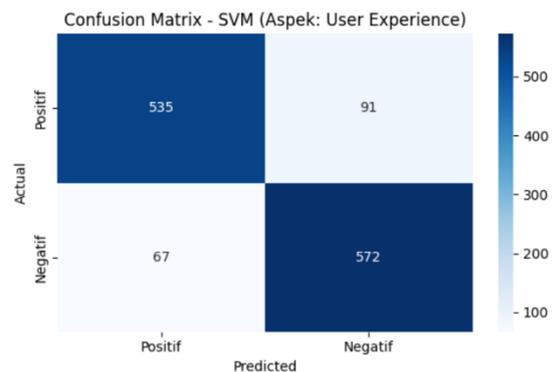
Gambar 7. AUC-ROC Aspek User Experience

4.3. Analisis Kesalahan Model

Untuk mengetahui jenis kesalahan yang terjadi dalam proses klasifikasi, dilakukan analisis terhadap *confusion matrix* dari masing-masing algoritma. Analisis ini bertujuan untuk mengidentifikasi jumlah kesalahan prediksi yang dihasilkan oleh model, khususnya pada nilai *False Positive* (FP) dan *False Negative* (FN). Penelitian ini menganalisis kesalahan model menggunakan *confusion matrix* [16]. Berikut ditampilkan *confusion matrix* berdasarkan setiap aspek pada Gambar 7 dan Gambar 8.



Gambar 8. Confusion Matrix Aspek Fitur



Gambar 9. Confusion Matrix Aspek User Experience

Berdasarkan Gambar 8 yang menunjukkan hasil *confusion matrix* berdasarkan aspek fitur disimpulkan bahwa model menunjukkan kinerja yang baik dengan jumlah prediksi benar untuk data positif adalah 1.244 dan prediksi benar untuk data negatif sebanyak 202 yang berarti banyak data positif yang berhasil teridentifikasi dengan benar. Gambar 9 menunjukkan hasil *confusion matrix* berdasarkan aspek *user experience*. Model berhasil mengidentifikasi data positif dengan benar, tetapi juga memiliki 91 *False Positives* yang menunjukkan beberapa data negatif salah di klasifikasikan sebagai positif.

4.4 Ekstraksi Pengetahuan

Ekstraksi pengetahuan dalam penelitian ini merupakan proses transformasi hasil klasifikasi sentimen ke dalam format terstruktur XML. Tujuannya bukan hanya menyimpan hasil evaluasi model, tetapi juga menghasilkan informasi baru berupa metadata yang bermanfaat dan siap digunakan dalam pengambilan keputusan. Representasi dalam format XML bertujuan untuk memudahkan pengembang aplikasi dan pemangku kepentingan dalam mengakses, menganalisis ulang, dan memanfaatkan hasil analisis. Dengan pendekatan ini, hasil ABSA tidak hanya berupa klasifikasi, tetapi juga menjadi pengetahuan terstruktur yang dapat diolah lebih lanjut untuk peningkatan kualitas layanan aplikasi.

```

<document>
<title> Ekstraksi Pengetahuan dari Ulasan Aplikasi CapCut Menggunakan Metode Analisis Sentimen Berbasis Aspek dan Metode Klasifikasi</title>
<content>
<aspect name="Fitur">
<positif>6607</positif>
<negatif>1538</negatif>
<kata_positif> Bagus, mantap, mudah, keren, lancar, recommended, baik, good, menarik, seru, best, lengkap, sip. </kata_positif>
<kata_negatif> jelek, kurang, bug, loading, ribet, macet, lag, error, lama, lemot, gajelas, bad, ngelek, burik. </kata_negatif>
<data_kata_positif>
mantap aplikasi capcut
bagus banget jj epep bjirrrr kyk ailight motion
fitur editnya lancar saja
fitur mudah untuk dipakai content creator mula
gatau poko emang bagus aja fitur nya
</data_kata_positif>
<data_kata_negatif>
jelek text to audio error kandung katakata aneh kena sensor alias langgar atur aplikasi sebal
bug aja nih template teks fungsi ngetik trus enter kali auto ilang gotul nget sih asli padahl teks editing sederhana
loading template nya macet 30 habis
</data_kata_negatif>
<evaluasi_model>
<akurasi>0.8876611418047882</akurasi>
<presisi>0.89</presisi>
<recall>0.89</recall>
<f1-score>0.89</f1-score>
<AUC-ROC>0.95</AUC-ROC>
<Cross_validation_mean_accuracy>0.884</Cross_validation_mean_accuracy>
<Confusion_Matrix>
<True_Positives>1244</True_Positives>
<False_Positives>97</False_Positives>
<False_Negatives>86</False_Negatives>
<True_Negatives>202</True_Negatives>
</Confusion_Matrix>
</evaluasi_model>
</aspect>
<aspect name="User Experience">
<positif>3055</positif>
<negatif>3269</negatif>
<kata_positif> Suka, puas, terbaik, bantu, enak, guna, ok, lima, senang, manfaat, gampang, cocok, lumayan, apik, wow </kata_positif>
<kata_negatif> Iklan, Kesel, Ganggu, kebanyakan, bayar, susah, ekspor, pro, update, kecewa, gabisa </kata_negatif>
<data_kata_positif>
puas pake aplikasinya
aplikasinya bantu banget
senang menggunakan capcut editor
</data_kata_positif>
<data_kata_negatif>
kadang ada iklan nya berisik ganggu
diupdate susah load ga masuk terang hubung internet pdhl kenceng
sistem capcut gimana kreator ngeluh sistem bayar template
</data_kata_negatif>
<evaluasi_model>
<Akurasi>0.8759988142292491</Akurasi>
<presisi>0.88</presisi>
<recall>0.88</recall>
<f1-score>0.88</f1-score>
<AUC-ROC>0.95</AUC-ROC>
<Cross_validation_mean_accuracy>0.885</Cross_validation_mean_accuracy>
<Confusion_Matrix>
<True_Positives>535</True_Positives>
<False_Positives>91</False_Positives>
<False_Negatives>67</False_Negatives>
<True_Negatives>572</True_Negatives>
</Confusion_Matrix>
</evaluasi_model>
</aspect>
</content>
</document>

```

Gambar 9. Ekstraksi Pengetahuan

Berdasarkan Gambar 9 didapatkan bahwa hasil ekstraksi menunjukkan bahwa aspek *fitur* mendapat dominasi sentimen positif sebanyak 6.607 ulasan dibandingkan dengan 1.538 ulasan negatif. Sebaliknya, pada aspek *user experience* terjadi dominasi sentimen negatif, yaitu sebanyak 3.269 ulasan dibandingkan 3.055 ulasan positif. Pengguna sering mengeluhkan iklan yang mengganggu, akses fitur berbayar, serta gangguan saat proses ekspor video. Model SVM yang digunakan untuk klasifikasi sentimen menunjukkan performa yang baik. Untuk aspek *fitur*, model menghasilkan akurasi sebesar 88,76% dengan skor AUC-ROC 0,95, menandakan model mampu membedakan kelas dengan baik. Demikian pula, aspek *user experience* menghasilkan akurasi 87,51% dan AUC-ROC yang sama tinggi.

Hasil evaluasi ini memperkuat bahwa model mampu menangkap karakteristik sentimen dari masing-masing aspek secara efektif. Dari hasil tersebut, dapat disimpulkan bahwa fitur CapCut secara umum mendapatkan respon positif dari pengguna, namun terdapat ruang untuk perbaikan pada aspek pengalaman pengguna, khususnya dalam hal iklan dan kenyamanan penggunaan secara keseluruhan. Pengetahuan yang diekstraksi ini penting sebagai dasar pengambilan keputusan strategis bagi pengembang aplikasi dalam meningkatkan kualitas layanan dan kepuasan pengguna.

5. Kesimpulan dan Saran

Penelitian ini menganalisis sentimen ulasan aplikasi CapCut dari *PlayStore* menggunakan metode *Aspect-Based Sentiment Analysis* (ABSA) dengan menggunakan algoritma seperti *Support Vector Machine* (SVM) dengan menunjukkan hasil performa algoritma seperti nilai akurasi, precision, recall, AUC, *ROC-Curve*, *F1-score* antar algoritma yang diujicobakan, kemudian menghasilkan ekstraksi pengetahuan berupa XML.

Penelitian ini menggunakan dataset hasil *crawling* sebanyak 21.668 data ulasan yang dianalisis berdasarkan dua aspek utama, yaitu aspek fitur dan *user experience*. Pada aspek fitur, sentimen positif mencapai 81,1% dan sentimen negatif sebesar 18,9%, sedangkan pada aspek *user experience*, sentimen positif hanya 48,3% dengan sentimen negatif mencapai 51,7%. Dari hasil uji coba terhadap model, menunjukkan bahwa algoritma *Support Vector Machine* (SVM) untuk setiap aspek menunjukkan performa yang baik. Untuk aspek fitur, model menghasilkan akurasi sebesar 88,76% dan aspek *user experience* menghasilkan akurasi 87,51%.

Ekstraksi pengetahuan disajikan dalam bentuk XML di mana pada penggunaan aplikasi CapCut untuk aspek fitur memiliki persentase sentimen positif lebih tinggi dibandingkan sentimen negatif. Hal ini menunjukkan bahwa untuk penggunaan fitur pada aplikasi cenderung lebih baik dan efektif bagi pengguna dan berdasarkan aspek *user experience*, pengguna merasa kurang puas dengan aplikasi CapCut. Pengetahuan yang berhasil diekstraksi ini memberikan wawasan penting bagi pengembang aplikasi. Dengan mempertahankan dan terus menyempurnakan fitur yang telah diapresiasi, serta mengurangi gangguan dari sisi pengalaman pengguna—khususnya terkait iklan dan aksesibilitas—pengembang dapat meningkatkan kepuasan dan loyalitas pengguna secara keseluruhan. Oleh karena itu, hasil ini dapat menjadi dasar dalam pengambilan keputusan strategis, baik untuk pengembangan produk maupun peningkatan kualitas layanan aplikasi CapCut ke depannya. Untuk pengembangan dari hasil penelitian ini, disarankan untuk menambahkan lebih banyak data akan meningkatkan keterwakilan dan generalisasi model serta dapat mengembangkan analisis sentimen yang mampu mengidentifikasi dan dapat menggunakan algoritma lainnya seperti *decision tree* atau *algoritma klasifikasi lainnya*. Hal ini akan memberikan wawasan yang lebih detail dan kaya, terutama dalam memahami persepsi pengguna terhadap berbagai fitur aplikasi.

Referensi

- [1] F. Noor 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.
- [2] R. Syahmewah, "PENGARUH PENGGUNAAN TEMPLATE PADA APLIKASI CAPCUT YANG MEMUDAHKAN MAHASISWA UNTUK MENGEDIT VIDIO SEBAGAI MEDIA PEMBELAJARAN," 2023.
- [3] M. Ardiansyah, M. M, and Riswanto, "Analisis Komparasi Ketertarikan Masyarakat Kota Batam Dalam Penggunaan Video Editor Capcut dan VN," *Jurnal Informasi dan Teknologi*, vol. 5, no. 3, pp. 91–102, Nov. 2023, doi: 10.60083/jidt.v5i3.398.
- [4] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc Netw Anal Min*, vol. 11, no. 1, p. 81, 2021.
- [5] O. Czeranowska *et al.*, "Migrants vs. stayers in the pandemic—A sentiment analysis of Twitter content," *Telematics and Informatics Reports*, vol. 10, p. 100059, 2023.
- [6] S. Samsir, K. Kusmanto, A. H. Dalimunthe, R. Aditiya, and R. Watrianthos, "Implementation naïve bayes classification for sentiment analysis on internet movie database," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 1, pp. 1–6, 2022.
- [7] Y. Chasseray, A.-M. Barthe-Delanoë, S. Négny, and J.-M. Le Lann, "Knowledge extraction from textual data and performance evaluation in an unsupervised context," *Inf Sci (N Y)*, vol. 629, pp. 324–343, 2023.

- [8] I. Tahyudin, A. R. Hananto, S. A. Rahayu, R. M. Anjani, and A. Nurhopipah, "Sentiment Analysis Model Development on E-Money Service Complaints.," *TEM Journal*, vol. 12, no. 4, 2023.
- [9] C. Zai and A. Rahman Isnain, "Komparasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) pada Analisis Sentimen Capcut," vol. 9, no. 1, p. 2024.
- [10] P. Pandunata, Y. Nurdiansyah, and F. D. Alfina, "Aspect-Based Sentiment Analysis of Avatar 2 Movie Reviews on IMDb Using Support Vector Machine," in *E3S Web of Conferences*, EDP Sciences, Nov. 2023. doi: 10.1051/e3sconf/202344802041.
- [11] A. Ramadina and K. D. Tania, "Knowledge Extraction of Gojek Application Review Using Aspect-based Sentiment Analysis," *The Indonesian Journal of Computer Science*, vol. 13, no. 3, 2024.
- [12] T. S. Sonnenschein, G. A. de Wit, N. R. den Braver, R. C. H. Vermeulen, and S. Scheider, "Validating and constructing behavioral models for simulation and projection using automated knowledge extraction," *Inf Sci (N Y)*, vol. 662, p. 120232, 2024.
- [13] O. Alqaryouti, N. Siyam, A. Abdel Monem, and K. Shaalan, "Aspect-based sentiment analysis using smart government review data," *Applied Computing and Informatics*, vol. 20, no. 1–2, pp. 142–161, Jan. 2024, doi: 10.1016/j.aci.2019.11.003.
- [14] I. Akbar, F. Supriadi, and D. I. Junaedi, "PEMANFAATAN MACHINE LEARNING DI BIDANG KESEHATAN," 2025.
- [15] M. Mujahid *et al.*, "Sentiment analysis and topic modeling on tweets about online education during COVID-19," *Applied Sciences*, vol. 11, no. 18, p. 8438, 2021.
- [16] N. B. Z. Adli, M. Ahmad, N. A. Ghani, S. D. Ravana, and A. A. Norman, "An Ensemble Classification of Mental Health in Malaysia related to the Covid-19 Pandemic using Social Media Sentiment Analysis," *KSII Transactions on Internet and Information Systems (TIIS)*, vol. 18, no. 2, pp. 370–396, 2024.