

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Dalam beberapa tahun terakhir, kemajuan teknologi mobile di Indonesia telah mendorong transformasi digital yang signifikan, khususnya dalam peningkatan penggunaan aplikasi pengeditan video di kalangan masyarakat umum. Fenomena ini didukung oleh penetrasi perangkat seluler yang masif, di mana tercatat bahwa 80% populasi Indonesia telah memiliki *smartphone* pada tahun 2024. Tingginya aksesibilitas terhadap teknologi ini telah mengubah perilaku konsumsi dan produksi konten digital, mendorong masyarakat untuk beralih menggunakan platform penyuntingan canggih namun praktis seperti CapCut guna menciptakan konten kreatif yang dapat didistribusikan secara instan ke berbagai jejaring sosial.[1]

Merespons tren tersebut, perusahaan ByteDance mengembangkan capcut sebagai aplikasi *all-in-one* yang dirancang secara khusus untuk memfasilitasi kebutuhan penyuntingan video pada perangkat seluler. Kehadirannya di Google Play Store telah menarik perhatian global, terbukti dengan jutaan unduhan dan interaksi pengguna. Ulasan yang masuk memberikan gambaran ganda mengenai performa aplikasi; di satu sisi, pengguna mengapresiasi antarmuka yang intuitif dan fitur penyuntingan yang *user-friendly*, namun di sisi lain, tidak sedikit pengguna yang menyuarakan keluhan terkait kendala teknis krusial, seperti terjadinya *lag* saat proses ekspor video hingga keberadaan iklan yang dinilai mengganggu pengalaman pengguna[2].

Popularitas aplikasi ini terlihat dari data statistik terbaru di Google Play Store yang diakses pada 31 Januari 2026, di mana CapCut telah mendapatkan jumlah unduhan dengan lebih dari 1 miliar unduhan secara global. Selain itu, tingginya tingkat keterlibatan pengguna tercermin dari volume respons yang mencapai 12,4 juta ulasan. Angka interaksi yang masif ini menegaskan bahwa metode analisis manual tidak lagi memadai, sehingga penerapan analisis sentimen

otomatis menjadi kebutuhan mendesak untuk memahami dinamika kepuasan pengguna secara efisien dan akurat.[3]

Besarnya volume ulasan yang tersedia pada platform Google Play Store menjadikan platform ini sebagai sumber data primer yang representatif untuk melihat sentimen pengguna secara luas. Mengingat keterbatasan dalam pengambilan data secara manual pada skala jutaan ulasan, penelitian ini memfokuskan pengambilan data dengan menggunakan teknik scraping dengan bantuan pustaka `google-play-scraper`. Pendekatan komputasi ini dipilih karena keunggulannya dalam mengakuisisi data bervolume besar secara efisien serta kemampuannya mempertahankan integritas metadata penting, seperti tanggal ulasan dan rating bintang. Penggunaan `google-play-scraper` sebagai instrumen akuisisi data ini sejalan dengan temuan penelitian[4], yang membuktikan bahwa library ini efektif digunakan untuk scraping data ulasan pada Google Play Store untuk keperluan analisis sentimen yang akurat.

Kendati demikian, data ulasan yang diperoleh dari platform Google Play Store memiliki karakteristik data yang tidak terstruktur. Data ini seringkali mengandung *noise* berupa penggunaan emoji, kesalahan tipografi, serta ragam bahasa informal atau *slang* yang mempersulit proses ekstraksi informasi. Kondisi ini menjadi hambatan bagi pengembang aplikasi untuk mendapatkan wawasan yang dapat ditindaklanjuti (*actionable insights*) secara langsung. Oleh karena itu, analisis sentimen hadir sebagai solusi sistematis untuk mengolah ketidakteraturan tersebut dan mengkategorikan opini pengguna ke dalam kelas sentimen positif, negatif, atau netral melalui pendekatan yang terstruktur dan terukur[5].

Dalam kerangka ilmu komputer, analisis sentimen sebagai bagian dari Pemrosesan Bahasa Alami (Natural Language Processing/NLP) dimanfaatkan sebagai sumber data utama yang bernilai tinggi untuk menjalankan penilaian kepuasan pengguna secara komprehensif. Penelitian empiris menunjukkan urgensi analisis ini, di mana terungkap bahwa hingga 40% dari total sentimen negatif seringkali berkaitan erat dengan masalah performa teknis aplikasi yang krusial. Temuan ini mengharuskan adanya tahapan pra-pemrosesan data yang mendalam

dan cermat, meliputi pembersihan data, stemming, hingga pembobotan kata menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) agar fitur teks dapat dikenali dengan baik oleh mesin[6].

Selain tantangan pada struktur teks, masalah ketidakseimbangan kelas data (imbalanced data) juga sering ditemukan dalam dataset ulasan aplikasi populer, di mana jumlah ulasan positif cenderung mendominasi secara signifikan dibandingkan ulasan negatif. Ketimpangan ini dapat menyebabkan model klasifikasi menjadi bias. Untuk mengatasi hal tersebut, penerapan teknik *resampling* seperti SMOTE (Synthetic Minority Over-sampling Technique) menjadi langkah strategis. Teknik ini terbukti secara empiris mampu menyeimbangkan distribusi kelas data dan meningkatkan akurasi model prediksi sebesar 5-10% dalam berbagai skenario pengujian[7].

Beberapa penelitian telah banyak menerapkan penggunaan algoritma Support Vector Machine (SVM) dalam bidang Analisis sentimen. Salah satunya penelitian[8] Penelitian ini memilih algoritma Support Vector Machine (SVM) karena keunggulannya dalam menangani data pada teks yang berdimensi tinggi. Penggunaan SVM sebagai metode pembelajaran mesin yang sangat baik untuk tugas klasifikasi teks sentimen, memiliki kemampuan unik untuk membentuk hyperplane yang ideal yang memaksimalkan margin pemisah antar kelas di ruang fitur yang besar dan kompleks.

Selanjutnya pada studi komparatif yang dilakukan menunjukkan bahwa SVM bekerja dengan cara mencari hyperplane ideal yang memaksimalkan margin pemisah antar kelas di ruang fitur yang kompleks sehingga algoritma SVM lebih baik dibandingkan dengan algoritma Naive Bayes pada dataset dengan fitur bahasa sehari-hari. Skor metrik F1 berhasil mencapai tingkat 88% setelah integrasi fitur yang tepat dan optimal[9]. Di sisi lain efektivitas SVM juga telah terbukti secara empiris pada berbagai objek penelitian serupa. Pada analisis sentimen ulasan aplikasi Grab dengan penggunaan 1000 data ulasan pengguna aplikasi Grab dilakukan analisis sentimen menggunakan SVM berhasil mendeteksi keluhan

pengguna terkait layanan seperti keterlambatan respons driver dengan tingkat akurasi 85% [10].

Lebih lanjut, konfigurasi kernel pada SVM juga memegang peranan penting dalam performa model. Hasil pada penelitian analisis sentimen terhadap aplikasi e-hadrah menunjukkan bahwa SVM dengan kernel linear lebih unggul daripada kernel RBF. SVM dengan kernel linear menghasilkan akurasi 95,46%, ketepatan 81,83%, recall 55,61%, dan *f-measure* 62,82% dengan menggunakan 976 data. Kernel RBF di sisi lain menghasilkan akurasi 94,15% ketepatan 58,03%, recall 40,33%, dan *f-measure* 43,24%[4]. Untuk analisis sentimen ulasan aplikasi jansostek mobile yang telah dilakukan dengan menggunakan Support Vector Machine menunjukkan bahwa pengklasifikasian metode SVM menghasilkan nilai akurasi terbaik dengan 96% akurasi dan 92% ketepatan[11].

Selanjutnya penelitian analisis sentimen terhadap ulasan aplikasi CapCut dengan menggunakan algoritma Support Vector Machine (SVM) yang dilakukan proses preprocessing dan metode vektorisasi TF-IDF digunakan untuk mengubah data menjadi representasi numerik, model SVM digunakan untuk mengkategorikan sentimen ke dalam tiga kategori positif, negatif, dan netral. Model SVM memiliki akurasi terbaik sebesar 90,12%[12]. Dengan menggunakan SVM analisis aplikasi M-Paspor menunjukkan beberapa masalah penting terkait aksesibilitas. Penemuan ini sangat relevan untuk meningkatkan inklusivitas pengelolaan fitur premium di aplikasi[13].

Berdasarkan paparan masalah dan tinjauan literatur tersebut, penelitian ini bermaksud mengembangkan sistem analisis sentimen pada ulasan aplikasi CapCut di Google Play Store menggunakan algoritma Support Vector Machine (SVM) yang dioptimalkan dengan TF-IDF dan penanganan ketidakseimbangan data. Berbeda dengan penelitian sebelumnya, penelitian ini akan memfokuskan klasifikasi pada dua kelas utama (positif dan negatif) dengan mengeliminasi ulasan netral (rating 3). Pendekatan *binary classification* ini dipilih untuk mempertegas polaritas opini sehingga menghasilkan wawasan yang lebih tajam. Hasil penelitian ini diharapkan dapat memberikan wawasan mendalam dan kontribusi nyata bagi

pengembang aplikasi dalam melakukan optimalisasi fitur sekaligus memperkaya literatur analisis sentimen berbahasa Indonesia pada ranah aplikasi penyuntingan video.

### **1.2 Rumusan Masalah**

- 1) Bagaimana cara pemberian label secara otomatis pada ulasan pengguna pada aplikasi CapCut?
- 2) Bagaimana penerapan pra-pemrosesan teks menggunakan library Sastrawi dan pembobotan TF-IDF dengan metode SMOTE untuk penyeimbangan data, serta dampaknya terhadap hasil klasifikasi ulasan aplikasi CapCut di Google Play Store menggunakan algoritma Support Vector Machine (SVM) setelah dilakukan filtering bintang 3?

### **1.3 Batasan Masalah**

- 1) Data diperoleh pada ulasan diambil hanya dari Google Play sebanyak 3.000 data ulasan terbaru menggunakan web crawling Python.
- 2) Hanya ulasan berbahasa Indonesia yang diproses.
- 3) Ulasan dengan bintang 3 difilter dan tidak digunakan dalam dataset.
- 4) Pelabelan sentimen otomatis, dilakukan dengan kriteria untuk Bintang 1 dan 2 negatif, sementara bintang 4 dan 5 positif. Dalam pelabelan ini tidak memiliki kelas netral.
- 5) Klasifikasi hanya menggunakan SVM dan tidak ada model lainnya.
- 6) Pengukuran kinerja menggunakan akurasi, presision, recall, F1 score.

### **1.4 Tujuan Penelitian**

- 1) Untuk pemberian label secara otomatis pada ulasan pengguna pada aplikasi CapCut?
- 2) Untuk melakukan pra-pemrosesan teks dengan library sastrawi pembobotan TF-IDF dengan metode SMOTE untuk penyeimbangan data dan hasil klasifikasi pada ulasan terbaru aplikasi CapCut di Google Play Store menggunakan algoritma Support Vector Machine setelah dilakukan filtering bintang 3.

## 1.5 Manfaat Penelitian

### 1) Manfaat Teoritis

Studi ini meningkatkan pengetahuan tentang pemrosesan bahasa alami (NLP) pemrosesan teks berbahasa Indonesia khususnya tentang penggunaan algoritma SVM dengan teknik pelabelan otomatis berbasis peringkat atribut bintang (skor) dan eliminasi kelas netral. Hasil penelitian ini memberikan referensi yang berharga untuk kemajuan penelitian analisis sentimen pada ulasan aplikasi mobile di Indonesia.

### 2) Manfaat Praktis

Perbaikan kualitas aplikasi yang lebih sesuai dengan kebutuhan pengguna aplikasi CapCut yang akan membantu pengembang membuat pembaruan yang cepat dan tepat sasaran untuk meningkatkan pengalaman pengguna yang lebih baik dan nyaman.

### 3) Manfaat Bagi Peneliti

Hasil dari penelitian ini dapat digunakan sebagai alat referensi dan bahan pengembangan oleh peneliti lain yang akan melakukan studi lanjutan pada bidang analisis sentimen baik untuk aplikasi lain maupun penyempurnaan teknik pra-pemrosesan teks berbahasa Indonesia.

## 1.6 Sistematika Penulisan

### 1.6.1 BAB I PENDAHULUAN

Pada bab ini berisi latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

### 1.6.2 BAB II TINJAUAN PUSTAKA

Bab ini menguraikan teori-teori penunjang yang dikutip dari jurnal ilmiah yang relevan dengan topik penelitian. Teori yang dibahas meliputi konsep analisis Sentimen, tahapan Preprocessing, teknik pembobotan kata TF-IDF, metode Support Vector Machine (SVM), teknik SMOTE untuk penanganan ketidakseimbangan data, serta metode evaluasi Confusion Matrix. Bab ini juga memuat tinjauan studi terdahulu.

### **1.6.3 BAB III METODE PENELITIAN**

Bab ini berisi tentang alur dan tahapan prosedur yang digunakan untuk menyelesaikan masalah penelitian dijelaskan. Seperti metode pengumpulan data ulasan dari Google Play Store, perancangan sistem, langkah pra-pemrosesan data (preprocessing), dan penerapan algoritma SMOTE dan SVM akan dibahas.

### **1.6.4 BAB IV HASIL DAN PEMBAHASAN**

Bab ini memaparkan hasil implementasi dari sistem yang telah dibangun serta analisis terhadap hasil pengujian. Hal yang dibahas meliputi deskripsi data hasil scraping, hasil tahapan preprocessing, visualisasi data sebelum dan sesudah penerapan SMOTE, serta analisis performa klasifikasi SVM berdasarkan nilai akurasi, presisi, recall, dan f1-score.

### **1.6.5 BAB V PENUTUP**

Pada bab ini merupakan bagian akhir yang berisi kesimpulan dan saran berdasarkan hasil penelitian dan pembahasan yang telah dilakukan untuk menjawab rumusan masalah.