

BAB II

TINJAUAN PUSTAKA

2.1. Studi Sebelumnya

Menurut penelitian terdahulu ada beberapa penelitian serupa, yang mana para peneliti tersebut melakukan analisis sentimen yang menggunakan aplikasi-aplikasi dengan metode yang sama. Berikut lima studi yang berkaitan:

Penelitian pertama ini dilakukan oleh Muhamad Fani Al-shufi, Adhitia Erfina. Peneliti melakukan Sentimen Analisis Mengenai Aplikasi Streaming Film Menggunakan Algoritma Support Vector Machine di Play Store. Kegiatan menonton film saat ini banyak dilakukan secara daring melalui berbagai aplikasi streaming, yang didukung oleh kemudahan akses internet serta variasi pilihan aplikasi yang tersedia. Ulasan dari pengguna menjadi aspek penting dalam menilai kualitas aplikasi tersebut, namun pengklasifikasian ulasan secara manual tidak memungkinkan karena jumlahnya sangat besar. Dalam penelitian ini, algoritma Support Vector Machine (SVM) digunakan untuk menganalisis 1000 ulasan pengguna dari lima aplikasi streaming yang ada di Playstore. Hasil analisis menunjukkan bahwa tingkat akurasi tertinggi dalam mengklasifikasikan ulasan positif dan negatif diperoleh pada aplikasi Iflix sebesar 92,67%, diikuti oleh Netflix 81,33%, Disney+ Hotstar 69,33%, WeTV 64,67%, dan Vidio 62,00% [4].

Penelitian kedua ini dilakukan oleh Vina Alviani, Syariful Alam, dan Imay Kurniawan (2023) menganalisis sentimen ulasan aplikasi WeTV di Twitter yang merupakan media sosial populer di Indonesia dengan tingkat penggunaan 59% pada

2020. Mereka menggunakan metode Support Vector Machine (SVM) yang dikenal memiliki akurasi lebih tinggi dibandingkan metode lain seperti Naive Bayes. Proses penelitian meliputi pengumpulan data, pembersihan, pelabelan, praproses teks, pembobotan kata dengan TF-IDF, klasifikasi menggunakan SVM, dan evaluasi dengan confusion matrix. Dari 4.024 data ulasan yang dianalisis, ditemukan 1.624 sentimen positif dan 2.400 sentimen negatif, dengan akurasi 89%, precision 87%, dan recall 83%. Sentimen negatif dominan disebabkan oleh banyaknya iklan, error aplikasi, dan fitur terjemahan yang belum optimal, menunjukkan kebutuhan perbaikan pada aplikasi WeTV[5]

Penelitian ketiga ini dilakukan oleh Stefan Novaldi Salombe, Syamsu Alam Valentino Aris (2025), menganalisis ulasan netizen pada Aplikasi Disney+ Hotstar di Aplikasi Google Play Store menggunakan Machine Learning. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna terhadap aplikasi Disney+ Hotstar di Google Play Store dengan pendekatan machine learning. Data ulasan diambil dari halaman aplikasi Disney+ Hotstar tersebut, kemudian melalui proses preprocessing seperti pembersihan data, tokenisasi, serta stemming. Data selanjutnya diklasifikasikan menggunakan dua algoritma machine learning, yaitu Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM). Hasil analisis menunjukkan SVM mencapai akurasi tertinggi dalam mengklasifikasikan sentimen ulasan. Analisis topik ulasan mengungkap bahwa sentimen positif sering membahas aspek layanan aplikasi, sentimen negatif didominasi keluhan masalah teknis, sementara sentimen netral berfokus pada deskripsi aplikasi tanpa penilaian emosional [6].

Penelitian keempat ini dilakukan Bima Santika dan Agung Kharisma Hidayah (2025), menganalisis Sentimen Ulasan Aplikasi Vidio pada Play Store Menggunakan Algoritma Support Vector Machine. Vidio: Sports, Movies, Series merupakan salah satu aplikasi layanan streaming yang banyak digunakan di Indonesia dan memperoleh penilaian sebesar 3,5 bintang di Google Play Store pada tahun 2024. Meskipun memiliki jumlah pengguna yang cukup besar, berbagai ulasan menunjukkan adanya ketidakpuasan pengguna terhadap sejumlah fitur serta kinerja

aplikasi. Oleh karena itu, penelitian ini bertujuan untuk menganalisis opini pengguna terhadap aplikasi Vidio berdasarkan ulasan yang terdapat di Google Play Store serta mengevaluasi tingkat akurasi algoritma Support Vector Machine (SVM) dalam mengklasifikasikan sentimen dengan pendekatan Knowledge Discovery in Databases (KDD). Pada tahap transformasi data, dataset dibagi menjadi data latih dan data uji dengan tiga skema pembagian, yaitu 90:10, 80:20, dan 70:30. Proses pembobotan kata dilakukan menggunakan teknik Term Frequency–Inverse Document Frequency (TF-IDF). Selanjutnya, performa algoritma SVM diuji dengan empat jenis kernel, yaitu Linear, Radial Basis Function (RBF), Polynomial, dan Sigmoid. Hasil pengujian menunjukkan bahwa pembagian data dengan rasio 70:30 menggunakan kernel Sigmoid memberikan kinerja terbaik, dengan nilai akurasi sebesar 83,33%, precision 86,13%, recall 83,33%, dan F1-score 76,25%. Model tersebut dinilai mampu menghasilkan keseimbangan yang optimal antara precision dan recall, meskipun masih terdapat sedikit penurunan pada nilai F1-score[7].

Penelitian kelima ini dilakukan Rezky Abdillah, Elin Haerani, dan Reski Mai Candra berjudul "Analisis Sentimen Ulasan Aplikasi WeTV untuk Peningkatan Layanan Menggunakan Metode Support Vector Machine" menjelaskan bahwa analisis sentimen merupakan metode efektif untuk mengidentifikasi opini pengguna aplikasi melalui ulasan teks. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) yang dipadukan dengan metode pembobotan Term Frequency-Inverse Document Frequency (TF-IDF) untuk mengklasifikasikan 12.000 ulasan pengguna WeTV di Google Play Store ke dalam tiga kategori sentimen: positif, netral, dan negatif. Proses pengolahan data meliputi cleaning, case folding, tokenizing, normalisasi, penghilangan stopword, dan stemming. Hasil pengujian menunjukkan bahwa SVM mampu mencapai akurasi 76%, precision 77%, recall 79%, dan f1-score 78%. Artikel ini juga mengacu pada penelitian terdahulu yang membuktikan bahwa SVM, baik dengan kernel linear maupun RBF, memiliki performa yang konsisten dan andal dalam analisis sentimen di berbagai konteks seperti opini masyarakat, layanan publik, dan media sosial [8].

Tabel 2.1 Penelitian Sebelumnya

No	Peneliti	Tahun	Tujuan	Metode	Hasil
1	Muhamad Fani Al-shufi, Adhitia Erfina.	2021	mengklasifikasikan ulasan menjadi positif atau negatif secara otomatis dan mengetahui tingkat akurasi klasifikasi pada masing-masing aplikasi streaming.	SVM	Mendapatkan ulasan positif,negatif dari setiap aplikasi streaming film
2	Vina Alviani, Syariful Alam, dan Imay Kurniawan	2023	Sentimen ulasan pengguna aplikasi WeTV di Twitter dengan menggunakan metode <i>Support Vector Machine</i> (SVM) agar dapat memahami opini positif dan negatif serta mengidentifikasi kendala utama yang harus diperbaiki pada aplikasi tersebut.	SVM	Jumlah sentimen negatif (2.400) lebih banyak daripada positif (1.624), terutama karena keluhan tentang iklan berlebihan, seringnya error, dan fitur terjemahan yang kurang baik, sehingga WeTV perlu melakukan perbaikan

3	Stefan Novaldi Salombe, Syamsu Alam,Valentino Aris	2025	Penelitian ini bertujuan menganalisis sentimen ulasan pengguna terhadap aplikasi Disney+ Hotstar di Google Play Store dengan pendekatan machine learning. Data ulasan diambil dari halaman aplikasi Disney+ Hotstar tersebut, kemudian melalui proses preprocessing seperti pembersihan data, tokenisasi, serta stemming. Data selanjutnya diklasifikasikan menggunakan dua algoritma <i>machine</i>	SVM & Naive Bayes	Hasil analisis menunjukkan SVM mencapai akurasi tertinggi dalam mengklasifikasikan sentimen ulasan. Analisis topik ulasan mengungkap bahwa sentimen positif sering membahas aspek layanan aplikasi, sentimen negatif didominasi keluhan masalah teknis, sementara sentimen netral berfokus pada deskripsi aplikasi tanpa penilaian emosional
---	--	------	--	-------------------	--

			<i>learning, yaitu Naive Bayes Classifier (NBC) dan Support Vector Machine (SVM)..</i>		
--	--	--	--	--	--

4	Bima Santika dan Agung Kharisma Hidayah	2025	penelitian ini bertujuan untuk menganalisis opini pengguna terhadap aplikasi Vidio berdasarkan ulasan yang terdapat di <i>Google Play Store</i> serta mengevaluasi tingkat akurasi algoritma Support Vector Machine (SVM)	SVM	Hasil pengujian menunjukkan bahwa pembagian data dengan rasio 70:30 menggunakan kernel Sigmoid memberikan kinerja terbaik, dengan nilai akurasi sebesar 83,33%, precision 86,13%, recall 83,33%, dan F1-score 76,25%. Model tersebut dinilai mampu menghasilkan keseimbangan yang optimal antara precision dan recall, meskipun masih terdapat sedikit
---	---	------	---	-----	--



					penurunan pada nilai F1-score
5	Rezky Abdillah, Elin Haerani, dan Reski Mai Candra	2023	Menganalisis pendapat pengguna aplikasi WeTV yang terdapat dalam ulasan di Google Play Store ke dalam tiga kelas sentimen: positif, netral, dan negatif.	SVM	Mendapatkan opini pengguna WeTV dalam ulasan untuk membantu pengembang memperbaiki dan meningkatkan layanan aplikasi

2.2. Dasar Teori

2.2.1. Analisis Sentimen

Analisis sentimen, yang juga dikenal sebagai *opinion mining*, adalah proses otomatis untuk memahami, mengekstrak, dan mengolah data teks guna memperoleh informasi mengenai sentimen yang terkandung dalam sebuah kalimat opini. Tujuan analisis sentimen adalah untuk mengidentifikasi pendapat atau kecenderungan opini seseorang terhadap suatu isu atau objek, apakah cenderung bersifat positif, negatif, atau netral. Proses ini melibatkan penelaahan pandangan atau opini individu terhadap topik tertentu dengan cara mengelompokkan sejumlah teks baik berupa dokumen, kalimat, maupun fitur ke dalam kategori sentimen seperti positif, negatif, atau netral.

Jadi Analisis sentimen sendiri merupakan suatu proses untuk mengidentifikasi, mengekstrak, dan mengolah informasi mengenai perasaan atau sikap seseorang yang berkaitan dengan opini tertentu, berdasarkan pendekatan komputasi terhadap opini, sentimen, maupun emosi seseorang terhadap suatu isu atau topik tertentu[9].

2.2.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) pertama kali diperkenalkan oleh Vapnik pada tahun 1992 sebagai salah satu metode machine learning yang mengadopsi prinsip *Structural Risk Minimization* (SRM). Metode *Support Vector Machine* (SVM) merupakan teknik pembelajaran mesin yang digunakan untuk mengelompokkan opini atau ulasan atau komentar ke dalam kategori sentimen positif, negatif, atau netral [10]. Dalam *Support Vector Machine* (SVM), fungsi utamanya adalah mencari dan membentuk batas pemisah terbaik (hyperplane) yang dapat memisahkan data ke dalam kelas-kelas tertentu. bekerja dengan mencari dan memisahkan data teks berdasarkan fitur yang sudah di input, misalnya menggunakan teknik TF-IDF, sehingga dapat membedakan kelas sentimen dengan maksimal antara kelas tersebut. Prinsip ini bertujuan untuk

memperoleh *hyperplane* optimal yang dapat memisahkan kelas berbeda dalam ruang input secara efektif

Selain itu, SVM memiliki keunggulan dalam mencegah *overfitting*, terutama ketika jumlah data antar kelas tidak seimbang [11]. SVM memungkinkan model tetap stabil meskipun terdapat perbedaan jumlah data antar kategori, seperti pada data sentimen yang jumlahnya lebih sedikit[12]. Berikut merupakan perbandingan Metode SVM dengan Metode serupa:

Aspek Perbandingan	SVM (Support Vector Machine)	Naive Bayes	Random Forest
Prinsip Kerja	Mencari <i>hyperplane</i> optimal dengan margin maksimum untuk memisahkan kelas	menentukan kelas suatu data berdasarkan peluang kemunculan kata-kata, dengan menganggap bahwa setiap kata tidak saling bergantung satu sama lain.	Menggabungkan banyak decision tree dan mengambil keputusan berdasarkan voting
Kinerja pada teks	Sangat baik pada data berdimensi tinggi (TF-IDF)	Cukup baik, namun menurun jika fitur saling berkorelasi	Kurang optimal pada data teks berdimensi sangat tinggi
Sensitivitas pada Noise	Umumnya tinggi dan stabil pada analisis sentimen	Sensitif terhadap kata ambigu dan noise	Lebih tahan terhadap noise dibanding Naive Bayes
Akurasi	Umumnya tinggi dan stabil pada analisis sentimen	Cepat tetapi akurasi relatif lebih rendah	Dapat tinggi, namun bergantung pada parameter dan ukuran data

Metode Support Vector Machine (SVM) dipilih dalam penelitian ini karena memiliki kemampuan yang sangat baik dalam menangani data teks berdimensi tinggi, seperti ulasan pengguna aplikasi streaming film yang direpresentasikan menggunakan

pembobotan TF-IDF[13]. Karakteristik data teks yang mengandung banyak kata dan fitur menyebabkan metode seperti Random Forest menjadi kurang efisien karena membutuhkan sumber daya komputasi yang besar dan waktu pelatihan yang lebih lama.

Dibandingkan dengan Naive Bayes, SVM tidak mengasumsikan bahwa setiap fitur bersifat independen, sehingga lebih sesuai untuk data ulasan pengguna yang memiliki keterkaitan antar kata. Selain itu, SVM dikenal mampu menghasilkan akurasi yang lebih tinggi dan stabil dalam klasifikasi sentimen, terutama pada data yang mengandung noise dan ketidakseimbangan kelas.

Dengan mempertimbangkan karakteristik data ulasan Genflix 2.0 yang bersifat tidak terstruktur, berdimensi tinggi, dan berpotensi mengandung noise, metode SVM dinilai lebih tepat dibandingkan Naive Bayes dan Random Forest untuk memperoleh hasil analisis sentimen yang lebih optimal dan reliabel.

2.2.3. Aplikasi Streaming Film

Aplikasi streaming film merupakan *platform digital* yang memberikan kemudahan bagi pengguna untuk menonton film dan program televisi secara online melalui internet tanpa perlu mengunduh file-film terlebih dahulu. Perkembangan teknologi yang pesat dalam beberapa dekade terakhir, khususnya di bidang hiburan digital dan jaringan internet, telah memberikan perubahan besar terhadap cara kita berkomunikasi dan memperoleh informasi [14]. Melalui aplikasi streaming film, pengguna dapat menonton berbagai film kapan pun dan di mana pun asalkan terhubung dengan internet. Aplikasi streaming film menggunakan teknologi pengiriman data secara langsung (real-time) untuk menyajikan konten berupa film yang bisa di tonton.

2.2.4. Aplikasi Genflix 2.0

Genflix 2.0 adalah layanan streaming *video on demand* (VOD) dan siaran langsung dari Indonesia yang merupakan versi pembaruan besar dari aplikasi Genflix sebelumnya. Aplikasi ini dirancang untuk memberikan pengalaman menonton yang lebih optimal dengan penambahan performa yang lebih baik, antarmuka pengguna yang lebih mudah digunakan, serta koleksi konten yang lebih lengkap. Dengan Genflix 2.0, pengguna dapat

menikmati berbagai konten mulai dari film Hollywood, serial TV, konten lokal, hingga siaran olahraga dengan harga yang terjangkau. Selain itu Aplikasi Genflix 2.0 memiliki konten eksklusif bernama ShortBite, yaitu fitur premium aplikasi yang menyajikan film atau drama yang berdurasi pendek. Selain itu juga, aplikasi ini dilengkapi dengan fitur unggulan seperti pemutar adaptif yang menyesuaikan kualitas video berdasarkan kecepatan internet pengguna, serta dukungan untuk berbagai jenis perangkat, menjadikan Genflix 2.0 salah satu platform hiburan digital yang kuat di pasar Indonesia.

2.2.5. Pre-processing

Preprocessing adalah proses awal dalam pengolahan data yang bertujuan untuk membersihkan, menyiapkan, dan mengubah data mentah agar menjadi lebih terstruktur dan mudah diproses. Proses ini membantu meningkatkan efektivitas algoritma pembelajaran mesin dan kualitas analisis data dengan memastikan data dalam format yang sesuai dan bebas dari kesalahan atau ketidak teraturan[15]

- **Text Cleaning** (Pembersihan Text) : Menghapus karakter-karakter yang tidak dibutuhkan seperti tanda baca, angka, tautan, tagar, sebutan pengguna, emoji, dan elemen lainnya.[16] Tujuan dari langkah ini adalah untuk menyingkirkan gangguan dan komponen yang tidak relevan dalam teks sehingga data menjadi lebih bersih dan siap untuk dianalisis.
- **Case Folding** : Proses ini melibatkan pengubahan seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) untuk menjaga konsistensi[17]. Misalnya, kata "Data *SCIENCE*" akan diubah menjadi "data *science*". Tujuan dari langkah ini adalah agar kata-kata yang sama tidak dianggap berbeda hanya karena perbedaan penggunaan huruf besar dan kecil.
- **Normalization** : Proses yang mengubah kata-kata yang tidak terstruktur atau singkat menjadi kata-kata yang sudah sesuai dengan standar Kamus Besar Bahasa Indonesia (KBBI)[18]. Proses ini meliputi identifikasi kata yang tidak baku lalu menggantinya dengan bentuk kata yang baku menurut KBBI.

- **Stop Word Removal:** Proses ini melibatkan penghilangan kata-kata umum yang sering muncul namun kurang memberikan informasi penting[19]. Seperti "di", "itu", dan sebagainya. Tujuannya adalah untuk meningkatkan ketepatan analisis dengan menyingkirkan kata-kata yang dianggap sebagai gangguan atau noise.
- **Tokenisasi:** Pemecahan teks menjadi bagian-bagian kecil yang disebut token[20]. Token tersebut dapat berupa kata, frasa, atau kalimat, sesuai dengan kebutuhan analisis. Tokenisasi memudahkan pengenalan elemen-elemen penting dalam teks yang nantinya dapat diproses lebih lanjut.
- **Stemming:** Tahap ini bertujuan mengonversi kata-kata yang memiliki imbuhan dan telah dipilih menjadi bentuk kata dasarnya dengan cara menghapus imbuhan yang menempel pada kata tersebut[21]. Dengan mengubah kata berimbuhan menjadi kata dasar, makna asli kata menjadi lebih mudah dipahami dan dianalisis dalam konteks bahasa.
- **Labeling:** Pelabelan dapat terdiri dari tiga kategori, yaitu positif, negatif, dan netral. Label-label ini digunakan dalam proses klasifikasi untuk mengelompokkan data yang telah melalui tahap pembobotan ke dalam masing-masing kelas tersebut.[22]

2.2.6.TF – IDF

TF-IDF adalah sebuah metode yang digunakan untuk menilai tingkat kepentingan suatu kata dalam sebuah dokumen dibandingkan dengan kumpulan dokumen lainnya. Konsep ini menggabungkan dua komponen utama, yaitu *Term Frequency* (TF) yang mengukur seberapa sering kata muncul dalam sebuah dokumen, serta *Inverse Document Frequency* (IDF) yang menilai seberapa jarang kata tersebut muncul di keseluruhan dokumen[23]. Metode TF-IDF sering digunakan dalam pengolahan teks, sebagai contoh dalam mesin pencari metode ini membantu menemukan kata kunci yang paling relevan dengan dokumen sehingga hasil pencarian menjadi lebih tepat[24]. Selain itu dengan adanya metode ini, data teks yang awalnya tidak

terstruktur dapat diolah menjadi informasi yang lebih bermakna dan bermanfaat untuk berbagai aplikasi digital[25].

2.2.7. Bahasa Pemrograman Python

Python adalah bahasa pemrograman tingkat tinggi yang mudah dipelajari, fleksibel, dan serbaguna. Bahasa ini pertama kali dikembangkan oleh Guido van Rossum pada tahun 1991 dengan tujuan agar kode dapat ditulis secara sederhana, mudah dibaca, dan efisien. Python dapat diterapkan dalam berbagai bidang, seperti pengembangan web menggunakan *framework Django* atau *Flask*, analisis data dan data science dengan *library Pandas* serta *NumPy*, hingga *machine learning* dan kecerdasan buatan menggunakan *TensorFlow* atau *Scikit-learn*[26]

