

ANALISIS SENTIMEN PENGGUNA APLIKASI iQIYI PADA GOOGLE PLAY BERBASIS SVM

Wenita Amelia ¹⁾

¹⁾ Teknik Informatika Institut Sains dan Bisnis Atma Luhur
email : 2311500005@mahasiswa.atmaluhur.ac.id¹⁾

Abstraksi

Perkembangan teknologi digital mendorong meningkatnya jumlah ulasan pengguna pada aplikasi streaming yang merepresentasikan pengalaman dan penilaian terhadap suatu layanan. Jumlah ulasan yang tinggi menimbulkan keterbatasan dalam proses evaluasi manual sehingga diperlukan pendekatan klasifikasi sentimen berbasis komputasi. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi iQIYI pada Google Play Store tahun 2025 menggunakan algoritma *Support Vector Machine* (SVM). Data yang digunakan sebanyak 4.929 ulasan pengguna yang diperoleh melalui proses *scraping* dan diproses melalui tahap *pre-processing* teks serta pembobotan menggunakan metode TF-IDF. Klasifikasi sentimen dilakukan menggunakan SVM dengan *linear kernel* dan *polynomial kernel* untuk membandingkan kinerja model. Hasil penelitian menunjukkan bahwa SVM dengan *linear kernel* memiliki performa klasifikasi yang lebih unggul dan stabil dibandingkan *polynomial kernel* dengan tingkat akurasi sebesar 88,92%. Temuan ini menunjukkan bahwa *linear kernel* efektif digunakan dalam klasifikasi sentimen ulasan pengguna aplikasi iQIYI.

Kata Kunci :

Analisis Sentimen, iQIYI, *Support Vector Machine* (SVM), *Linear Kernel*, Google Play Store

Abstract

The development of digital technology has driven an increasing number of user reviews on streaming apps, representing experiences and assessments of a service. The high number of reviews creates limitations in the manual evaluation process, necessitating a computational-based sentiment classification approach. This study aims to analyze the sentiment of iQIYI app user reviews on the Google Play Store in 2025 using the Support Vector Machine (SVM) algorithm. The data used were 4,929 user reviews obtained through scraping and processed through a text pre-processing stage and weighting using the TF-IDF method. Sentiment classification was performed using SVM with a linear kernel and a polynomial kernel to compare model performance. The results showed that SVM with a linear kernel had superior and more stable classification performance than the polynomial kernel with an accuracy rate of 88.92%. These findings indicate that the linear kernel is effective in classifying sentiment of iQIYI app user reviews.

Keywords :

Sentiment Analysis, iQIYI, *Support Vector Machine* (SVM), *Linear Kernel*, Google Play Store.

Pendahuluan

Transformasi digital secara signifikan telah menggeser paradigma konsumsi media masyarakat melalui layanan *Over-the-Top* (OTT) yang menawarkan fleksibilitas akses konten multimedia secara luas [1]. Salah satu platform yang mendominasi pasar Indonesia saat ini adalah iQIYI, yang dikenal luas melalui distribusi konten drama dan film Asia. Sebagai aplikasi yang dapat diunduh secara gratis melalui Google Play Store, iQIYI menjadi salah satu platform *streaming* yang paling populer dan mudah diakses oleh masyarakat untuk menikmati berbagai konten hiburan. Dengan basis pengguna yang besar iQIYI memperoleh ulasan masif yang merepresentasikan pengalaman pengguna secara *real-time*. Namun, proses evaluasi opini secara manual untuk meningkatkan kualitas layanan menjadi sulit karena banyaknya ulasan tekstual yang tidak terorganisir [2].

Aplikasi iQIYI menghadapi tantangan besar dalam mempertahankan loyalitas pengguna karena banyaknya ulasan negatif tentang kendala teknis dan kualitas layanan. Menurut penelitian sebelumnya, keluhan pengguna aplikasi serupa seringkali berfokus pada masalah sistem langganan premium dan masalah antarmuka yang mengganggu kemudahan akses [3]. Mengingat bahwa narasi ulasan tekstual mengandung lebih banyak informasi daripada hanya rating bintang untuk mengukur kepuasan pelanggan, opini-opini ini dapat merusak reputasi platform [4]. Maka otomatisasi yang dapat memahami pengguna diperlukan untuk membantu pengelola aplikasi membuat kebijakan yang tepat [5].

Upaya penyelesaian masalah klasifikasi tersebut sebelumnya telah banyak dilakukan menggunakan algoritma seperti *Random Forest* yang berfokus pada pengolahan tekstual [6] dan *K-Nearest Neighbor* (KNN) yang berbasis pada kedekatan fitur data, untuk menyelesaikan masalah klasifikasi sentimen [2].

Namun demikian, algoritma *Support Vector Machine* (SVM) dipilih dalam penelitian ini karena keunggulannya dalam menangani data teks yang sangat besar melalui pencarian *hyperplane* yang ideal untuk memisahkan kelas sentimen secara konsisten [7][8]. Metode ini digunakan pertama kali dengan mengekstraksi data menggunakan *scraping web*. Setelah itu, proses *pre-processing* untuk mengubah teks mentah menjadi data terstruktur. Metode pembobotan *Term Frequency-Inverse Document Frequency* (TF-IDF) dapat mengekstraksi fitur kata yang paling representatif dalam membedakan persepsi positif dan negatif untuk meningkatkan akurasi klasifikasi [9][10].

Tujuan utama dari penelitian ini adalah untuk membangun model klasifikasi sentimen yang mampu membedakan persepsi positif dan negatif pengguna iQIYI secara akurat. Hasil dari penelitian ini akan divalidasi melalui instrumen *confusion matrix* yang mencakup metrik *accuracy*, *precision*, *recall*, dan *f1-score* untuk menjamin keandalan model yang dibangun [2][9]. Melalui pencapaian nilai metrik yang optimal, penelitian ini diharapkan dapat memberikan kontribusi ilmiah serta menjadi landasan bagi pengelola aplikasi iQIYI dalam meningkatkan kualitas layanan pelanggan berdasarkan data ulasan terbaru [5].

Kebaruan (*novelty*) dalam penelitian ini adalah menganalisis secara komparatif penggunaan fungsi *linear kernel* dan *polynomial kernel* pada algoritma SVM. Pendekatan ini diterapkan untuk mengeksplorasi serta menentukan model klasifikasi yang paling stabil dan adaptif terhadap karakteristik data ulasan pada objek aplikasi iQIYI. Berbeda dengan penelitian Fathon (2025), yang menggunakan *Random Forest* [6], dan Pratama dan Rosyida (2023), yang menggunakan KNN [2], penelitian ini mengeksplorasi efektivitas fungsi kernel yang berbeda untuk menangani dinamika data ulasan tahun 2025. Integrasi tahapan *pre-processing* yang komprehensif serta pengujian stabilitas kernel SVM ini diharapkan mampu menghasilkan performa klasifikasi yang lebih presisi dan kompetitif dibandingkan studi-studi sebelumnya [7][8].

Tinjauan Pustaka

Penelitian terdahulu

Penelitian mengenai analisis sentimen pada platform streaming dan aplikasi digital telah banyak dilakukan dengan berbagai algoritma. Penelitian sebelumnya, [6] menganalisis ulasan pengguna aplikasi iQIYI, yang menemukan akurasi sebesar 83% dari algoritma *Random Forest*. Meskipun cukup representatif untuk mengklasifikasikan sentimen secara keseluruhan, penelitian tersebut memiliki kelemahan signifikan pada nilai presisi sentimen negatif, yang hanya mencapai 62%. Akibatnya, model dianggap tidak sensitif untuk mengidentifikasi keluhan teknis khusus pengguna. Selain itu, [2] juga melakukan klasifikasi pada objek yang sama, tetapi terbatas hanya menggunakan algoritma *K-Nearest Neighbor* (KNN).

Metode ini dinilai kurang efektif karena efisiensi komputasi dan ketidakmampuan KNN untuk menangani dimensi fitur teks yang kompleks secara optimal dibandingkan dengan pendekatan berbasis *hyperplane*.

Studi lainnya, [11] menggunakan algoritma *Support Vector Machine* (SVM) pada ulasan WeTV dengan nilai hasil yang lebih baik daripada *Naive Bayes*. Meskipun SVM terbukti unggul dalam domain video *Over-the-Top* (OTT), penelitian tersebut memiliki kelemahan metodologis karena penelitian tersebut belum mengkaji teknik penanganan ketidakseimbangan data (*imbalanced data*), yang sering menyebabkan bias pada model ulasan aplikasi populer. [12] menunjukkan akurasi 93,16% pada data ulasan aplikasi BYOND by BSI yang menunjukkan keunggulan SVM. Namun, studi tersebut memiliki kelemahan dalam generalisasi model karena dataset yang digunakan didominasi dalam konteks perbankan formal. Sehingga kinerjanya belum diuji pada dinamika bahasa ulasan aplikasi hiburan seperti iQIYI, yang penuh dengan istilah non-formal. Penelitian oleh [13] mencapai akurasi 91,9% untuk penggunaan algoritma *Naive Bayes* pada ulasan aplikasi Agoda di Google Play Store. Namun masih memiliki kesulitan untuk memahami susunan kata yang lebih kompleks dalam kalimat ulasan. Selain itu, [9] mengimplementasikan SVM dengan *linear kernel* tunggal pada ulasan Google Play Store dengan akurasi 91%. Kelemahan utama penelitian tersebut terletak pada penggunaan fungsi kernel tunggal yang kurang fleksibel dalam menangani karakteristik data ulasan yang memiliki fitur non-linear.

Untuk meningkatkan efektivitas klasifikasi pada penelitian sebelumnya, penelitian ini melakukan analisis komparatif penggunaan *linear kernel* dan *polynomial kernel* pada ulasan iQIYI tahun 2025. Melalui eksplorasi optimasi fungsi kernel ini, diharapkan tercipta model klasifikasi yang lebih adaptif terhadap kompleksitas data non-linear serta mampu memberikan hasil yang lebih akurat dan kredibel.

Landasan Teori

iQIYI

iQIYI merupakan sebuah platform hiburan yang memungkinkan pengguna untuk mengakses berbagai konten video seperti film dan drama melalui perangkat mobile, dan ulasan yang diberikan pengguna di platform distribusi merupakan data penting untuk mengukur tingkat kepuasan pelanggan terhadap konten dan layanan teknis [6].

Analisis Sentimen

Analisis sentimen atau *opinion mining* adalah bidang studi yang menganalisis opini, evaluasi, dan emosi individu terhadap suatu entitas. Dalam konteks aplikasi mobile, analisis ini bertujuan untuk mengklasifikasikan polaritas teks ulasan ke dalam kategori positif atau negatif guna memahami persepsi publik terhadap layanan aplikasi secara otomatis [7].

Google Play Store

Google Play Store adalah ekosistem distribusi digital yang dimiliki oleh Google yang berfungsi sebagai penyedia utama berbagai jenis konten multimedia untuk perangkat berbasis Android, termasuk aplikasi, game, dan literatur digital. Selain memenuhi kebutuhan pengunduhan dan pembaruan perangkat lunak dalam berbagai kategori, platform ini juga berfungsi sebagai media interaksi dua arah melalui fitur ulasan dan rating [14]. Untuk menilai kualitas dan kepuasan pengguna terhadap layanan digital, kumpulan ulasan merupakan alat penting.

Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin berbasis *supervised learning* yang bekerja dengan membangun sebuah *hyperplane* untuk memisahkan kelas data secara optimal [12]. Secara mekanis, algoritma ini berfokus pada identifikasi bidang pemisah yang memiliki *margin* atau ruang kosong maksimal antara *hyperplane* dengan titik data terdekat dari setiap kategori, yang secara teknis didefinisikan sebagai *support vectors* [8]. Dalam konteks pemrosesan bahasa alami, SVM memiliki keunggulan dalam menangani dimensi fitur tekstual yang besar serta memiliki resistensi yang tinggi terhadap fenomena *overfitting*. Karakteristik tersebut memungkinkan SVM untuk mempertahankan stabilitas dan akurasi yang konsisten dalam mengklasifikasikan polaritas sentimen pada ulasan pengguna di Google Play Store [11][9].

Text Mining

Text mining adalah proses pengambilan informasi penting dari data teks yang tidak terstruktur. Untuk menemukan pola dalam kumpulan data teks sebelum diklasifikasikan, metode ini terdiri berbagai tahapan *pre-processing*, seperti *cleaning*, *case folding*, *word normalization*, *tokenizing*, *stopword removal* dan *stemming* [15][13]. Tujuannya untuk mengurangi *noise* pada data teks, yang akan meningkatkan akurasi analisis frekuensi dan kinerja model klasifikasi.

Fungsi Kernel (Linear dan Polynomial)

Salah satu elemen fundamental dalam algoritma *Support Vector Machine* (SVM) adalah fungsi kernel, yang berfungsi untuk mentransformasikan data dari ruang masukan (*input space*) ke dalam ruang fitur berdimensi lebih tinggi guna mengatasi batasan linearitas pada data asli [12]. Secara teknis, fungsi ini memungkinkan algoritma untuk menentukan *hyperplane* optimal tanpa harus melakukan perhitungan koordinat data pada dimensi tinggi secara eksplisit, sebuah mekanisme efisiensi komputasi yang dikenal sebagai *kernel trick*.

1. Linear Kernel

Merupakan fungsi kernel yang memetakan data secara langsung tanpa melakukan transformasi non-linear yang rumit. Karena pemisahan antar kelas seringkali sudah cukup optimal melalui garis pemisah linear di ruang fitur, kernel ini

sangat efektif digunakan pada dataset ulasan Google Play Store yang memiliki ukuran fitur teks yang sangat besar (ribuan kata) [16].

2. Polynomial Kernel

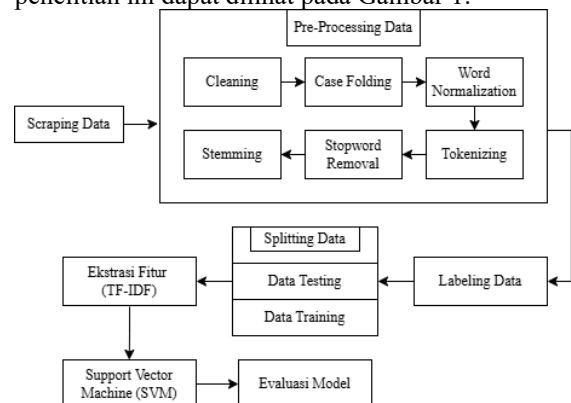
Polynomial kernel adalah fungsi kernel non-linear yang mengonversi data ke dalam ruang fitur dengan dimensi yang lebih tinggi, agar dapat mengatasi pola sentimen yang rumit dan tumpang tindih. Dengan memanfaatkan kernel ini, batas keputusan yang dibentuk menjadi jauh lebih fleksibel melalui pengaturan parameter derajat (*degree*), menjadikannya lebih responsif dalam mengklasifikasikan fitur-fitur data ulasan aplikasi yang selalu berubah. [17][18].

Confusion Matrix

Confusion matrix merupakan alat evaluasi performa model klasifikasi yang menyajikan tabel perbandingan antara hasil prediksi dengan data aktual. instrumen ini menghasilkan metrik berupa *accuracy*, *precision*, *recall*, dan *f1-score* yang memberikan validasi objektif mengenai stabilitas model, terutama dalam menghadapi ketidakseimbangan jumlah data sentimen [19].

Metode Penelitian

Diagram alur penelitian yang digunakan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Diagram Alur Penelitian

Scraping Data

Pada tahap awal penelitian ini, data ulasan pengguna untuk aplikasi iQIYI dikumpulkan melalui Google Play Store. Dengan menggunakan bahasa pemrograman Python pada platform Google Colab, proses pengambilan data dilakukan secara otomatis menggunakan teknik *web scraping*. Peneliti menarik data secara *real-time*, selanjutnya, data mentah yang telah dikumpul diekspor dan disimpan dalam bentuk dataset.csv untuk diolah pada tahap penelitian selanjutnya.

Pre-Processing Data

Pada tahap selanjutnya, dataset.csv diproses melalui tahap *pre-processing* untuk diubah menjadi data yang bersih dan konsisten. Proses ini dimulai dengan *Cleaning* untuk menghapus *noise* seperti angka, simbol, tanda baca, dan URL. Kemudian, *case folding*

mengubah teks menjadi huruf *lowercase* secara keseluruhan. selanjutnya, *word normalization* digunakan untuk mengubah kata tidak baku menjadi kata standar, kemudian dilakukan *tokenizing* untuk memotong kalimat menjadi potongan kata tunggal dan proses berikutnya, menggunakan *stopword removal* untuk mengeliminasi kata-kata umum yang tidak bermakna lalu *stemming* untuk mengonversi kata berimbuhan kembali ke bentuk dasarnya.

Labeling Data

Labeling data dilakukan dengan mengkategorikan ulasan ke dalam sentimen positif, negatif dan netral. Ulasan dengan skala tinggi dikategorikan sebagai sentimen positif, sedangkan ulasan dengan skala rendah dikategorikan sebagai sentimen negatif. Hasil pelabelan ini sangat penting karena berfungsi sebagai data dasar (*ground truth*) yang digunakan pada tahap penelitian berikutnya untuk melatih dan menguji akurasi model klasifikasi.

Splitting Data

Setelah tahap pelabelan selesai, dataset yang memiliki label sentimen dibagi menjadi dua bagian utama, yaitu *data training* dan *data testing*. dalam penelitian ini, pembagian data menggunakan rasio sebesar 80% untuk *data training* yang digunakan untuk melatih model dalam mengenali pola kata dalam kategori sentimen positif dan negatif dan 20% untuk *data testing* yang digunakan untuk menguji ketepatan model dalam mengklasifikasikan ulasan baru. untuk memastikan bahwa model yang dihasilkan objektif dan menghindari *overfitting*, proses pembagian ini dilakukan secara acak. Sebelum mencapai tahap klasifikasi dengan algoritma *Support Vector Machine* (SVM), dua komponen data ini akan diproses melalui tahap ekstraksi fitur menggunakan *TF-IDF* untuk mengubah teks menjadi bentuk numerik.

Ekstraksi Fitur (TF-IDF)

Setelah pembagian data selesai, langkah berikutnya mengonversi teks menjadi numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Tahap ini menghitung bobot setiap kata berdasarkan frekuensi kemunculannya (*Term Frequency*) dan tingkat kelangkaannya dalam seluruh dataset (*Inverse Document Frequency*). Tujuan proses ini untuk memberikan bobot atau nilai tinggi pada kata-kata unik yang informatif serta mengurangi dampak kata-kata umum yang tidak memiliki nilai diskriminatif. Hasilnya adalah matriks fitur, yang digunakan oleh model SVM untuk mengonstruksi *hyperplane* optimal dengan membedakan kategori sentimen. Hasil ekstraksi fitur ini juga disajikan dalam bentuk WordCloud untuk menunjukkan distribusi kata paling sering muncul dalam ulasan iQIYI. Ukuran kata yang lebih besar mencerminkan frekuensi kemunculan paling tinggi yang merupakan ciri utama dalam membedakan sentimen positif dan negatif.

Support Vector Machine (SVM)

Implementasi SVM dalam penelitian ini dilakukan dengan membandingkan *linear kernel* dan *polynomial kernel* untuk menemukan *hyperplane* paling optimal pada data ulasan iQIYI. Pada *linear kernel* digunakan parameter *default* berupa nilai C sebesar 1.0. Sedangkan pada *polynomial kernel* diterapkan parameter *degree* sebesar 3, nilai C sebesar 1.0, serta *gamma* menggunakan *scale* untuk membangun model klasifikasi sentimen. Hasil klasifikasi dari kedua kernel tersebut dipetakan menggunakan instrumen *confusion matrix*. Tabel kontingensi ini berfungsi untuk membandingkan hasil prediksi model terhadap label aktual (*ground truth*) pada data uji.

Evaluasi Model

Tahap akhir dari metodologi penelitian ini adalah evaluasi model, yang bertujuan untuk mengukur sejauh mana keandalan dan keakuratan algoritma SVM dalam mengklasifikasikan sentimen ulasan iQIYI. Proses evaluasi ini dilakukan dengan menggunakan *classification report*, sehingga performa model dapat diukur secara komprehensif menggunakan empat metrik utama:

Accuracy

Accuracy menggambarkan persentase total ulasan yang diklasifikasikan dengan benar (baik positif maupun negatif) oleh model dibandingkan dengan keseluruhan data uji. Rumus *accuracy* menggunakan Persamaan (1).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision

Precision menunjukkan seberapa akurat model dalam memprediksi kelas positif dari seluruh data yang diprediksi sebagai positif. Rumus *precision* menggunakan Persamaan (2).

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall

Recall mengukur keberhasilan model dalam menemukan kembali informasi atau mengenali pola pada kelas positif dari keseluruhan data yang benar-benar positif. Rumus *recall* menggunakan Persamaan (3).

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

F1-Score

F1-Score merupakan rata-rata harmonik antara *precision* dan *recall*. Metrik ini memberikan gambaran performa model yang lebih seimbang, terutama untuk memastikan bahwa model tetap memiliki kinerja yang stabil meskipun terdapat potensi ketidakseimbangan jumlah data pada kelas sentimen tertentu. Rumus *F1-Score* menggunakan Persamaan (4).

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Hasil dari perhitungan metrik-metrik tersebut akan memberikan validasi ilmiah mengenai efektivitas penggunaan algoritma SVM dalam menganalisis sentimen pengguna aplikasi iQIYI.

Hasil dan Pembahasan

Scraping Data

Tahap awal penelitian ini adalah mengumpulkan ulasan pengguna aplikasi iQIYI pada Google Play Store menggunakan library *Google Play Scraper* dengan bahasa pemrograman Python. Data dikumpul secara otomatis melalui Google Colab untuk periode Januari hingga Desember 2025 dengan total hasil 4.929 ulasan. Dataset ini mencakup lima atribut utama, yaitu *Review ID*, *Username*, *Date*, *Score*, dan *Content*. Seluruh data kemudian disimpan dalam format .CSV untuk diolah lebih lanjut pada tahap pre-processing. Sampel dataset mentah hasil *scraping data* dapat dilihat pada Tabel 1.

Tabel 1. Hasil *Scraping Data*

reviewid	username	date	score	content
48cb2f8f-de8c-4e7c-9a02-61a838c8fc50	Pengguna Google	2025-12-27	5	aplikasi ini bagus tapi terlalu banyak iklan
cea0e42c-13c6-410d-9559-4c04b28094f5	Pengguna Google	2025-12-27	3	Apk di android oke, tolong dong apk yang di smart tv subtitle nya digedein atau ada settingan buat memperbesar subtitle nya, kecil banget kek semut padahal film nya udah oke banget.
75b4b279-28c4-4063-a329-95ede995cd26	Pengguna Google	2025-12-27	2	hadehh awal doang lancar makin kesini sering eror, udah nyedot pulsa buat langganan otomatis tp tetep nggak bisa nonton, eh nyoba perpanjang langganan gagal mulu

Pre-Processing

Tahapan *pre-processing* yang dilakukan untuk mengolah data ulasan iQIYI sebagai berikut:

Cleaning

Untuk menjaga kualitas dataset, proses *cleaning* dilakukan dengan mengeliminasi seluruh komponen yang tidak relevan atau dianggap sebagai *noise*. Tahap ini mencakup penghapusan tanda baca (seperti titik, koma, dan tanda seru), angka, simbol-simbol khusus, tautan URL, *username* (@), dan emoji yang tidak relevan untuk analisis sentimen karena data ulasan aplikasi iQIYI masih berupa teks mentah. Proses ini dilakukan untuk memastikan bahwa algoritma SVM hanya memproses teks ulasan murni,

sehingga model klasifikasi dapat mengidentifikasi pola sentimen dengan optimal tanpa gangguan karakter non-alfabet yang dapat memengaruhi akurasi hasil akhir. Hasil *cleaning* dapat dilihat pada Tabel 2.

Tabel 2. Hasil *Cleaning*

Sebelum	Sesudah
aplikasi ini bagus tapi terlalu banyak iklan	aplikasi ini bagus tapi terlalu banyak iklan
Apk di android oke, tolong dong apk yang di smart tv subtitle nya digedein atau ada settingan buat memperbesar subtitle nya, kecil banget kek semut padahal film nya udah oke banget	Apk di android oke tolong dong apk yang di smart tv subtitle nya digedein atau ada settingan buat memperbesar subtitle nya kecil banget kek semut padahal film nya udah oke banget
hadehh awal doang lancar makin kesini sering eror, udah nyedot pulsa buat langganan otomatis tp tetep nggak bisa nonton, eh nyoba perpanjang langganan gagal mulu	hadehh awal doang lancar makin kesini sering eror udah nyedot pulsa buat langganan otomatis tp tetep nggak bisa nonton eh nyoba perpanjang langganan gagal mulu

Case Folding

Pada tahap *case folding*, semua huruf dalam data ulasan aplikasi iQIYI diubah menjadi huruf kecil (*lowercase*). Proses ini dilakukan untuk menghindari duplikasi kata karena perbedaan penggunaan huruf kapital, seperti kata "Apk" dan "apk", yang dianggap sebagai dua entitas berbeda oleh sistem. Dengan menyeragamkan teks menjadi huruf kecil, konsistensi data tetap terjaga. Sehingga memudahkan algoritma SVM untuk menghitung bobot kata secara akurat tanpa adanya redundansi fitur. Hasil penerapan *case folding* dapat dilihat pada Tabel 3.

Tabel 3. Hasil *Case Folding*

Sebelum	Sesudah
aplikasi ini bagus tapi terlalu banyak iklan	aplikasi ini bagus tapi terlalu banyak iklan
Apk di android oke, tolong dong apk yang di smart tv subtitle nya digedein atau ada settingan buat memperbesar subtitle nya kecil banget kek semut padahal film nya udah oke banget	apk di android oke tolong dong apk yang di smart tv subtitle nya digedein atau ada settingan buat memperbesar subtitle nya kecil banget kek semut padahal film nya udah oke banget
hadehh awal doang lancar makin kesini sering eror udah nyedot pulsa buat langganan otomatis tp tetep nggak bisa nonton eh nyoba perpanjang langganan gagal mulu	hadehh awal doang lancar makin kesini sering eror udah nyedot pulsa buat langganan otomatis tp tetep nggak bisa nonton eh nyoba perpanjang langganan gagal mulu

Word Normalization

Untuk memperbaiki penulisan kata-kata yang tidak baku, singkatan, dan kesalahan penulisan (*typo*) yang ditemukan dalam ulasan pengguna iQIYI, dilakukan proses normalisasi kata. Tahap ini dilakukan karena

data yang diperoleh dari komentar media sosial biasanya berbahasa informal, sehingga penting untuk mengubah kata-kata tersebut ke dalam bentuk baku dalam kamus bahasa Indonesia. Dengan normalisasi, variasi kata yang bermakna sama dapat diseragamkan seperti kata “tapiii” menjadi “tapi”, meningkatkan kualitas dan kejelasan informasi data dataset. Hasil penggunaan *word normalization* dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Word Normalization*

Sebelum	Sesudah
aplikasi ini bagus tapiiii trlalu bnyakkk iklaannn	aplikasi ini bagus tapi terlalu banyak iklan
apk di android oke tolong dong apk yang di smart tv subtitle nya digedein atau ada settingan buat memperbesar subtitle nya kecil banget kek semut padahal film nya udah oke banget	aplikasi di android bagus tolong dong aplikasi yang di smart tv subtitle nya diperbesar atau ada settingan buat memperbesar subtitle nya kecil banget seperti semut padahal film nya sudah bagus banget
hadehh awal doang lancar makin kesini sering eror udah nyedot pulsa buat langganan otomatis tp tetep nggak bisa nonton eh nyoba perpanjang langganan gagal mulu	mengeluh awal saja lancar semakin ke sini sering error sudah boros pulsa buat langganan otomatis tapi tetap tidak bisa nonton eh mencoba perpanjang langganan gagal selalu

Tokenizing

Setelah teks melalui proses normalisasi, dilakukan tahap tokenizing untuk membagi kalimat ulasan menjadi potongan kata-kata tunggal atau token. Proses ini memisahkan setiap kata berdasarkan spasi dan menghilangkan karakter yang tersisa sehingga setiap kata dapat berfungsi sebagai fitur sendiri. Tahap ini sangat penting untuk membantu sistem mengidentifikasi pola sentimen dalam ulasan iQIYI lebih akurat dan mengidentifikasi frekuensi kemunculan setiap kata secara khusus. Hasil proses *tokenizing* dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Tokenizing*

Sebelum	Sesudah
aplikasi ini bagus tapi terlalu banyak iklan	['aplikasi', 'ini', 'bagus', 'tapi', 'terlalu', 'banyak', 'iklan']
aplikasi di android bagus tolong dong aplikasi yang di smart tv subtitle nya diperbesar atau ada settingan buat memperbesar subtitle nya kecil banget seperti semut padahal film nya sudah bagus banget	['aplikasi', 'di', 'android', 'bagus', 'tolong', 'dong', 'aplikasi', 'yang', 'di', 'smart', 'tv', 'subtitle', 'nya', 'diperbesar', 'atau', 'ada', 'settingan', 'buat', 'memperbesar', 'subtitle', 'nya', 'kecil', 'banget', 'seperti', 'semut', 'padahal', 'film', 'nya', 'sudah', 'bagus', 'banget']
mengeluh awal saja lancar semakin ke sini sering error sudah boros pulsa buat langganan otomatis tapi tetap tidak	['mengeluh', 'awal', 'saja', 'lancar', 'semakin', 'ke', 'sini', 'sering', 'error', 'sudah', 'boros', 'pulsa', 'buat', 'langganan', 'otomatis', 'tapi', 'tetap', 'tidak', 'bisa', 'nonton', 'eh', 'mencoba', 'perpanjang', 'langganan', 'gagal', 'selalu']

bisa nonton eh mencoba perpanjang langganan gagal selalu	['otomatis', 'tapi', 'tetap', 'tidak', 'bisa', 'nonton', 'eh', 'mencoba', 'perpanjang', 'langganan', 'gagal', 'selalu']
--	---

Stopword Removal

Selanjutnya *stopword removal*, proses menghilangkan kata-kata umum yang sering terjadi tetapi tidak memiliki nilai informasi yang signifikan untuk menentukan suatu sentimen. Dalam ulasan aplikasi iQIYI, kata seperti “ini”, “di”, “dong”, “yang”, “dan”, “nya”, “atau”, “saja”, hingga “eh” dihilangkan karena dianggap sebagai kata tugas yang tidak mewakili pendapat. Dengan menghilangkan kata-kata tersebut, dataset menjadi lebih sederhana, hanya memiliki kata-kata inti yang menunjukkan keluhan atau perasaan pengguna terhadap aplikasi iQIYI. Hasil tahap *stopword removal* dapat dilihat pada Tabel 6.

Tabel 6. Hasil *Stopword Removal*

Sebelum	Sesudah
['aplikasi', 'ini', 'bagus', 'tapi', 'terlalu', 'banyak', 'iklan']	['aplikasi', 'bagus', 'terlalu', 'banyak', 'iklan']
['aplikasi', 'di', 'android', 'bagus', 'tolong', 'dong', 'aplikasi', 'yang', 'di', 'smart', 'tv', 'subtitle', 'nya', 'diperbesar', 'atau', 'ada', 'settingan', 'buat', 'memperbesar', 'subtitle', 'nya', 'kecil', 'banget', 'seperti', 'semut', 'padahal', 'film', 'nya', 'sudah', 'bagus', 'banget']	['aplikasi', 'android', 'bagus', 'aplikasi', 'subtitle', 'diperbesar', 'settingan', 'memperbesar', 'subtitle', 'kecil', 'semut', 'bagus']
['mengeluh', 'awal', 'saja', 'lancar', 'semakin', 'ke', 'sini', 'sering', 'error', 'sudah', 'boros', 'pulsa', 'buat', 'langganan', 'otomatis', 'tapi', 'tetap', 'tidak', 'bisa', 'nonton', 'eh', 'mencoba', 'perpanjang', 'langganan', 'gagal', 'selalu']	['mengeluh', 'lancar', 'sering', 'error', 'boros', 'pulsa', 'langganan', 'otomatis', 'perpanjang', 'langganan', 'gagal', 'selalu']

Stemming

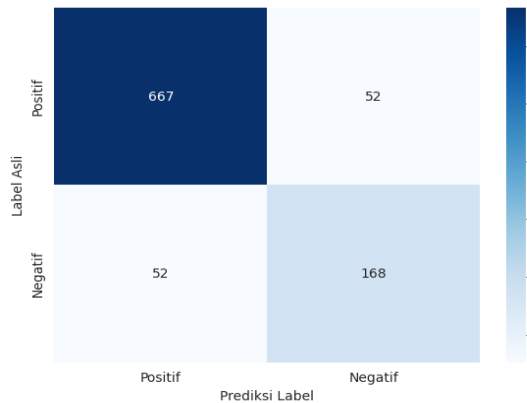
Tahap terakhir pada *pre-processing*, yaitu *stemming* proses mengubah kata berimbuhan menjadi kata dasar aslinya. Dalam penelitian ini, *library Sastrawi* digunakan untuk melakukan proses *stemming* dengan tujuan untuk mengurangi variasi kata dengan akar yang sama misalnya, kata-kata “diperbesar” dan “memperbesar” diubah menjadi kata dasar “besar” dan kata “langganan” diubah menjadi “langgan”. Dengan menyatukan berbagai bentuk kata ke dalam satu akar kata tunggal, dataset menjadi lebih konsisten dan efisien. Hasil tahap *stemming* dapat dilihat pada Tabel 7.

Tabel 7. Hasil *Stemming*

Sebelum	Sesudah
['aplikasi', 'bagus', 'terlalu', 'banyak', 'iklan']	Aplikasi bagus terlalu banyak iklan
['aplikasi', 'android', 'bagus', 'aplikasi']	Aplikasi android bagus aplikasi subtitle besar set

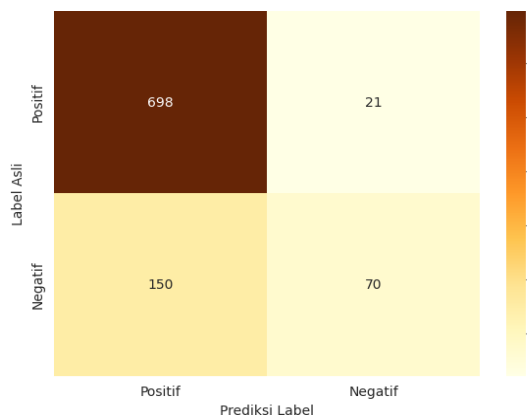
Support Vector Machine (SVM)

Penelitian ini bertujuan untuk memodelkan pola sentimen ulasan iQIYI ke dalam kategori positif atau negatif dengan menggunakan algoritma *Support Vector Machine* (SVM). Pengujian dilakukan dengan membandingkan dua fungsi, *linear kernel* dan *polynomial kernel* untuk mengetahui efektivitas tiap kernel dalam mengidentifikasi ruang fitur teks berdimensi tinggi yang dinilai menggunakan *confusion matrix* dan dapat dilihat pada Gambar 6 dan Gambar 7.



Gambar 6. *Confusion Matrix Linear Kernel*

Pengujian *linear kernel* pada Gambar 6 menunjukkan kemampuan klasifikasi yang stabil dengan presisi tinggi pada kedua kelas sentimen. Model ini menghasilkan 667 ulasan positif (*True Positive*) dan 168 ulasan negatif (*True Negative*) yang terklasifikasi akurat. Rendahnya jumlah kesalahan klasifikasi, yaitu masing-masing 52 ulasan pada sentimen negatif yang terprediksi sebagai positif (*False Positive*) dan sentimen positif yang terprediksi sebagai negatif (*False Negative*) menunjukkan bahwa dataset ulasan iQIYI bersifat terpisah secara linear. Hal ini menyatakan perbedaan leksikal antara sentimen positif dan negatif cukup berbeda sehingga pembentukan *hyperplane* linear sangat efektif untuk memisahkan data secara optimal dan efisien.



Gambar 7 *Confusion Matrix Polynomial Kernel*

Pada Gambar 7, hasil klasifikasi pengujian dengan *polynomial kernel* menunjukkan bahwa 698 ulasan sentimen positif (*True Positive*), dan 70 ulasan sentimen negatif (*True Negative*). Namun, terdapat tingkat kesalahan yang cukup tinggi dimana 150 sentimen negatif diprediksi positif (*False Positive*)

dan 21 sentimen positif diprediksi negatif (*False Negative*). Dengan pola ini, model lebih cenderung fokus pada pola sentimen positif dan mengalami kesulitan untuk membedakan karakteristik khusus dari pola sentimen negatif karena pemodelan fitur dengan fungsi polinomial membuat batas pemisahan antar kelas menjadi kurang jelas.

Dalam mengklasifikasikan sentimen ulasan iQIYI, *linear kernel* menunjukkan hasil klasifikasi yang lebih seimbang dan stabil daripada *Polynomial Kernel* berdasarkan hasil *confusion matrix*.

Evaluasi Model

Evaluasi model dilakukan untuk membandingkan kinerja *linear kernel* dan *polynomial kernel* pada algoritma SVM. Hasil perbandingan dapat dilihat pada Tabel 8.

Tabel 8. Perbandingan SVM *Linear Kernel* dan *Polynomial Kernel*

Kernel	Kelas	Precision	Recall	F1-Score	Accuracy
Linear	Positif	0.76	0.76	0.76	88.92 %
	Negatif	0.93	0.93	0.93	
Polynomial	Positif	0.77	0.32	0.45	81.79 %
	Negatif	0.82	0.97	0.89	

Dalam penelitian ini, peneliti menemukan perbedaan antara penggunaan *linear kernel* dan *polynomial kernel* terutama pada data ulasan aplikasi iQIYI. Berdasarkan hasil evaluasi menunjukkan bahwa *linear kernel* menghasilkan kinerja yang lebih optimal dibandingkan *polynomial kernel*. *Linear kernel* mencapai tingkat akurasi sebesar 88,92%, sedangkan *polynomial kernel* memperoleh akurasi sebesar 81,79%. Pada kategori sentimen positif, *linear kernel* menunjukkan performa yang stabil dengan nilai *precision*, *recall*, dan *f1-score* yang konsisten sebesar 0,76. sementara itu, *polynomial kernel* memiliki nilai *precision* yang relatif tinggi pada kelas positif sebesar 0,77, namun diikuti oleh nilai *recall* yang rendah sebesar 0,32 dan *f1-score* sebesar 0,45, yang menunjukkan keterbatasan model dalam mengenali seluruh ulasan positif. pada kategori sentimen negatif, *linear kernel* memperlihatkan kinerja yang sangat baik dengan nilai *precision*, *recall*, dan *f1-score* masing-masing sebesar 0,93. *polynomial kernel* juga menunjukkan performa yang cukup tinggi pada kelas negatif dengan nilai *precision* sebesar 0,82, *recall* sebesar 0,97, dan *f1-score* sebesar 0,89. namun, ketidakseimbangan performa antar kelas pada *polynomial kernel* mengindikasikan adanya kecenderungan bias dalam proses klasifikasi, sehingga *linear kernel* dinilai lebih unggul karena mampu menjaga konsistensi dan akurasi prediksi pada seluruh kelas ulasan aplikasi iQIYI.

Kesimpulan dan Saran

Hasil penelitian dapat disimpulkan algoritma *Support Vector Machine* (SVM) dapat memprediksi dan mengklasifikasikan sentimen pengguna aplikasi

iQIYI pada Google Play tahun 2025 dengan baik. *Linear Kernel* menunjukkan performa hasil yang lebih unggul dan stabil pada model SVM dibandingkan *Polynomial Kernel*, dengan tingkat akurasi 88,92%. Meskipun perbedaan kinerja pada kedua kernel, *Linear Kernel* memiliki keunggulan dalam menghasilkan kinerja klasifikasi yang stabil dalam mengenali sentimen positif dan negatif pada ulasan pengguna aplikasi iQIYI berbasis fitur teks. Salah satu kelemahan penelitian ini adalah dominasi ulasan positif dan penghapusan kelas netral yang dapat menyebabkan distribusi kelas sentimen yang tidak seimbang dan berpotensi memengaruhi hasil klasifikasi. Secara keseluruhan, tujuan penelitian untuk mengonstruksi model klasifikasi sentimen yang akurat telah dicapai melalui penggunaan SVM *Linear Kernel*. SVM *Linear Kernel* terbukti mampu menghasilkan kinerja konsisten dalam membedakan persepsi pengguna iQIYI.

Untuk mengurangi dominasi kelas sentimen tertentu, penelitian selanjutnya disarankan untuk menggunakan metode penanganan data tidak seimbang (*Imbalanced Data*). Selain itu, penelitian tentang metode klasifikasi lain dan penggunaan jumlah data ulasan yang lebih besar dapat dilakukan untuk meningkatkan akurasi dan kemampuan model untuk menangani kompleksitas suatu bahasa ulasan pengguna.

Daftar Pustaka

- [1] M. Y. Siregar, A. Davy Wiranata, and R. A. Saputra, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Streaming Vidio Menggunakan Metode Naïve Bayes," *Media Online*, vol. 4, no. 5, pp. 2419–2429, 2024, doi: 10.30865/klik.v4i5.1787.
- [2] S. Rosyida and A. Pratama, "Classification of Positive and Negative Sentiments Using the K-Nearest Neighbor Algorithm on iQIYI Application," *Sinkron*, vol. 8, no. 2, pp. 631–636, 2023, doi: 10.33395/sinkron.v8i2.12204.
- [3] U. Kulsum, M. Jajuli, and N. Sulistiyowati, "Analisis Sentimen Aplikasi WETV di Google Play Store Menggunakan Algoritma Support Vector Machine," *J. Appl. Informatics Comput.*, vol. 6, no. 2, pp. 205–212, 2022, doi: 10.30871/jaic.v6i2.4802.
- [4] V. Fitriyana, L. Hakim, D. C. R. Novitasari, and A. H. Asyhar, "Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine," *J. Buana Inform.*, vol. 14, no. 01, pp. 40–49, 2023, doi: 10.24002/jbi.v14i01.6909.
- [5] M. Yusuf Rismanda Gaja, I. Maulana, and O. Komarudin, "Analisis Sentimen Opini Pengguna Aplikasi Vidio Pada Ulasan Playstore Menggunakan Algoritma Naive Bayes," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 4, pp. 2767–2774, 2024, doi: 10.36040/jati.v7i4.7197.
- [6] U. Fathon, M. Arifin, and A. Setiawan, "Analisis Sentimen Ulasan Pengguna Aplikasi Crunchyroll, iQIYI, Wibuku di Google Play Store Menggunakan Metode Random Forest," *STORAGE J. Ilm. Tek. dan Ilmu Komput.*, vol. 4, no. 3, pp. 265–276, 2025, doi: 10.55123/storage.v4i3.6027.
- [7] N. Fauzan and P. W. Setyaningsih, "Sentiment Analysis of BPD DIY Mobile Banking Application Using SVM and KNN Methods," *J. Inovtek Polbeng*, vol. 10, no. 2, pp. 934–944, 2025, [Online]. Available: <https://jurnal.polbeng.ac.id/index.php/ISI/article/view/594>
- [8] S. Butsianto and A. M. Rifa'i, "Analisis Sentimen Ulasan Aplikasi Jamsostek dengan SVM, Random Forest, dan Logistic Regression," *J. Inform. Ekon. Bisnis*, vol. 7, pp. 700–706, 2025, doi: 10.37034/inf.v7i3.1266.
- [9] R. A. Afif, A. Supriyanto, R. F. Damaryanti, and W. Prasetya Adi, "Analisis Sentimen Aplikasi Adiraku di Google Play Store Menggunakan Metode Support Vectore Machine," *J. Fasilkom*, vol. 15, no. 1, pp. 163–171, 2025, doi: 10.37859/jf.v15i1.8510.
- [10] I. Surya Kumala Idris, Y. Aril Mustofa, and I. Abraham Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, pp. 32–35, 2023, [Online]. Available: <https://ejournal.ung.ac.id/index.php/jjee/article/view/16830>
- [11] P. P. Allorerung and R. Rismayani, "Analisis Sentimen Pada Ulasan Aplikasi WeTV di Google Play Store Menggunakan Algoritma NBC dan SVM Sentiment Analysis on WeTV App Reviews on Google Play Store Using NBC and SVM Algorithms," *Sist. J. Sist. Inf.*, vol. 12, pp. 404–414, 2023, [Online]. Available: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/viewFile/2518/555>
- [12] F. N. Firzatullah and N. Nuroji, "Analisis Sentimen Pengguna Aplikasi Byond BSI Pada Google Play Store Menggunakan Algoritma SVM Dan Random Forest," *METIK J.*, vol. 9, no. 2, pp. 265–276, 2025, doi: 10.47002/metik.v9i2.1089.
- [13] R. S. Nuraini, D. Krisbiantoro, M. Imron, and P. Subarkah, "ANALISIS SENTIMEN PENGGUNA APLIKASI AGODA DI GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAIVE BAYES," vol. 7, no. 1, 2025, [Online]. Available: <https://jurnal.amikom.ac.id/index.php/joism/article/view/2066>
- [14] P. Apria Ananda Anam, D. Abdul Fatah, and M. Ali Syakur, "Analisis Sentimen Review Aplikasi Zalora Di Google Playstore Menggunakan Metode Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 2, pp. 2443–2450, 2025, doi: 10.36040/jati.v9i2.13083.
- [15] A. S. A. Rafsanjani, D. L. Fithri, and S. Supriyono, "Analisis Sentimen Ulasan Pengguna Aplikasi KitaLulus pada Google Play Store dengan menggunakan algoritma Support Vector Machine (SVM)," *Sistemasi*, vol. 14, no. 5, p. 2519, 2025, doi: 10.32520/stmsi.v14i5.5519.
- [16] U. Hendriyanto, M. D., Ridha, A. A. Enri, "ANALISIS SENTIMEN ULASAN APLIKASI MOLA PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE," *J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 120–128, 2022, doi: 10.36987/informatika.v11i2.5860.
- [17] T. Tinaliah and T. Elizabeth, "Analisis Sentimen

- Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 9, no. 4, pp. 3436–3442, 2022, doi: 10.35957/jatisi.v9i4.3586.
- [18] P. Fremmuzar and A. Baita, “Uji Kernel SVM dalam Analisis Sentimen Terhadap Layanan Telkomsel di Media Sosial Twitter,” *Komputika J. Sist. Komput.*, vol. 12, no. 2, pp. 57–66, 2023, doi: 10.34010/komputika.v12i2.9460.
- [19] S. Dermawan and A. T. Ayunda, “Sentiment Analysis of Coretax on Social Media X Using Naive Bayes , SVM , and LSTM for Service Improvement,” vol. 9, no. 6, 2025, [Online]. Available: <https://jurnal.polibatam.ac.id/index.php/JAIC/article/view/11063>