



Implementation of *Support Vector Machine* for Sentiment Analysis on Blibli App Reviews on *Play Store*

Bima Santika*¹, Agung Kharisma Hidayah²

¹Universitas Muhammadiyah Bengkulu, Kota Bengkulu, Indonesia

*Corresponding Author: santikabima83@gmail.com

Abstract - Blibli is one of the leading e-commerce platforms in Indonesia with a 4.7-star rating on the Google Play Store. Although it has many positive reviews, automatic sentiment analysis is needed to understand user perceptions objectively and efficiently. This study aims to classify the sentiment of user reviews of the Blibli application using the Support Vector Machine (SVM) algorithm. A total of 2,000 recent reviews were collected through web scraping and underwent preprocessing processes including case folding, cleaning, normalization, filtering, stemming, and tokenization. To address class imbalance, the SMOTE method was applied so that positive and negative sentiment data became balanced, with 1,413 reviews each. The data were then divided into three training and testing ratio scenarios: 90:10, 80:20, and 70:30. Text transformation was carried out using the TF-IDF method. Experiments were conducted by comparing four SVM kernels: Linear, RBF, Polynomial, and Sigmoid. The best results were obtained in the 90:10 scenario using the Linear kernel, achieving an accuracy of 96.82%, precision of 96.93%, recall of 96.82%, and F1-score of 96.82%. These findings indicate that SVM, particularly with the Linear kernel, is highly effective and balanced in classifying user review sentiments. The results of this study are expected to contribute to application developers in improving service quality based on user feedback obtained automatically and in a structured manner.

Keywords— Sentiment analysis, Blibli, Play Store, SVM, SMOTE

I. PENDAHULUAN

Sekarang ini, perkembangan teknologi berlangsung sangat pesat dan telah menjadi bagian penting dalam menunjang berbagai kebutuhan manusia. Berkat kemajuan teknologi, berbagai aktivitas kini dapat dilakukan dengan lebih efisien dan praktis [1]. Salah satu contohnya adalah kegiatan jual beli yang telah bertransformasi ke dalam bentuk digital, menggantikan metode konvensional. Proses ini dikenal dengan istilah *e-commerce* [2]. Menurut laporan terbaru, pasar *e-commerce* Indonesia diperkirakan akan mencapai US\$44,7 miliar pada tahun 2024, dengan proyeksi pertumbuhan tahunan sebesar 10,65%. Selain itu, sekitar 6 dari 10 konsumen Indonesia kini berpartisipasi dalam belanja *online* melalui *platform live shopping*, yang mencerminkan perubahan besar dalam cara berbelanja [3].

Google Play Store merupakan sebuah layanan distribusi digital yang menyajikan berbagai konten multimedia, seperti aplikasi, permainan, film, buku digital, serta musik, yang

dikategorikan dalam beragam jenis untuk memenuhi kebutuhan pengguna [4]. Salah satu aplikasi yang dapat ditemukan di *platform* ini adalah Blibli, yang merupakan salah satu *e-commerce* terdepan di Indonesia yang menyediakan beragam produk, mulai dari kebutuhan rumah tangga hingga tiket perjalanan. Beberapa fitur andalannya mencakup layanan pengiriman cepat, sistem pembayaran yang *fleksibel*, serta promo menarik. Aplikasi ini berhasil meraih *rating* 4,7 bintang di *Play Store* pada tahun 2024, dengan jumlah unduhan yang telah melampaui angka sepuluh juta serta mendapatkan respons positif dari pengguna [5].

Google Play menyediakan fitur ulasan pengguna yang memungkinkan siapa saja untuk mengetahui pengalaman orang lain dalam menggunakan sebuah aplikasi. Fitur ini sering dijadikan referensi praktis dalam mengakses informasi terkait produk atau layanan tertentu [6]. Hal ini menjadi penting karena banyak ulasan yang tidak secara langsung menunjukkan apakah pengguna puas atau tidak. Beberapa komentar terlihat positif, tapi sebenarnya menyindir atau mengandung sarkasme. Ada juga ulasan yang terlalu singkat atau tidak jelas, sehingga sulit untuk dipahami maksud sebenarnya. Selain itu, jumlah ulasan positif dan negatif sering kali tidak seimbang. Karena itu, diperlukan otomatis agar opini pengguna bisa diklasifikasikan dengan lebih tepat dan objektif. Meski ulasan-ulasan tersebut sangat membantu, membaca dan mengevaluasi ribuan komentar secara manual bukanlah hal yang efisien serta berisiko menimbulkan bias. Karena itu, diperlukan metode analisis sentimen otomatis untuk mengelompokkan opini pengguna ke dalam sentimen positif maupun negatif, sehingga dapat mempermudah pengembang dan pengguna dalam

DOI: <https://doi.org/10.24036/voteteknika.v13i2.133666>

Received : 2025-05-12
Revised : 2025-06-10
Accepted : 2025-06-26
Published : 2025-06-29



For all articles published in VOTETEKNIKA
<https://ejournal.unp.ac.id/index.php/voteknika>, © copyright is retained by the authors. This is an open-access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

menilai kualitas aplikasi yang tersedia[7]. Namun, analisis sentimen pada teks berbahasa Indonesia, terutama di bidang *e-commerce*, sering menghadapi kesulitan seperti penggunaan bahasa sehari-hari yang tidak baku, singkatan, dan ungkapan yang sulit dipahami oleh komputer, sehingga membuat proses analisis menjadi lebih rumit[8].

Support Vector Machine (SVM) dikenal sebagai salah satu algoritma yang umum diterapkan karena efektivitasnya dalam menangani permasalahan klasifikasi dan regresi secara akurat. Saat digunakan pada data latih yang terbatas. Metode ini diperkenalkan oleh Vapnik dan bekerja optimal pada data yang bersifat *Linear*, sementara untuk data *non-Linear*, *SVM* memanfaatkan fungsi kernel agar data dapat dipisahkan secara efisien[9]. Keunggulan lain dari *SVM* adalah kemampuannya dalam melakukan generalisasi terhadap data baru, menjaga tingkat kesalahan tetap rendah, serta tetap andal meskipun jumlah atau dimensi fitur sangat besar[10]. Berkat karakteristik tersebut, *SVM* sering dianggap sebagai salah satu metode klasifikasi paling andal, dengan kinerja yang melampaui banyak pendekatan lainnya [11].

Adapun penelitian sebelumnya menyimpulkan bahwa *SVM* memiliki akurasi tertinggi (81%) dalam analisis sentimen terhadap Tiktokshop di media sosial X dibanding tiga algoritma lainnya. Namun, penelitian ini masih terbatas karena hanya menilai performa berdasarkan akurasi tanpa mempertimbangkan metrik lain seperti *precision*, *recall*, dan *F1-score*. Keunikannya terletak pada penggunaan data terkini dari media sosial X serta fokus pada topik Tiktokshop yang masih jarang diteliti, dengan pendekatan komparatif antar algoritma[12].

Temuan lain, Penelitian ini menyimpulkan bahwa *SVM* memiliki performa sedikit lebih baik daripada *Naive Bayes* dalam analisis sentimen Tiktokshop, dengan *F1-score* tertinggi 99%. Namun, penelitian ini belum mengevaluasi pengaruh teknik praproses atau pemilihan fitur terhadap hasil model. Keunikannya terletak pada penggunaan data terkini dari media sosial serta fokus pada topik Tiktokshop yang masih jarang diteliti dengan pendekatan komparatif berbasis metrik evaluasi lengkap[13].

Penelitian berbeda juga menyimpulkan bahwa *SVM* memiliki performa terbaik dalam analisis sentimen ulasan aplikasi Loket X dengan akurasi 85% dan *F1-score* 71%, sementara *Naive Bayes* menunjukkan kelemahan pada *F1-score* meski akurasinya sama. Keterbatasan penelitian ini adalah penggunaan 500 data dan klasifikasi manual yang berpotensi subjektif. Keunikannya terletak pada data dari dua platform aplikasi dan fokus pada aplikasi lokal yang jarang diteliti, dengan evaluasi menggunakan berbagai metrik performa[14].

Selain itu, Penelitian ini menganalisis sentimen pengguna media sosial X terhadap serangan ransomware data KIP-K dengan membandingkan *Naive Bayes* dan *SVM*. Dari 2.648 data mentah yang diproses menjadi 1.738 data bersih, *SVM* unggul dengan akurasi 88% dan *F1-score* tinggi dibanding *Naive Bayes* yang akurasinya 70%. Namun, penelitian ini terbatas pada satu platform dan belum mengeksplorasi pengaruh praproses atau parameter model secara mendalam. Keunikannya adalah fokus pada isu keamanan data publik dan penggunaan data *real-time* dari media sosial untuk menggambarkan respons masyarakat[15].

Konsistensi performa *SVM* juga terlihat pada penelitian lain, Penelitian ini mengimplementasikan *SVM* dan *Naive Bayes* untuk klasifikasi sentimen 3.000 ulasan aplikasi OVO dengan pelabelan *lexicon-based VADER*, menemukan dominasi sentimen negatif 66%. *SVM* unggul dengan akurasi 89% dan *F1-score* 88%, dibanding *Naive Bayes* 86% dan 87%. Keterbatasan meliputi kurangnya eksplorasi teknik praproses kompleks dan variasi parameter, serta ketergantungan pada pelabelan *lexicon-based* yang kurang akurat untuk bahasa Indonesia. Keunikan terletak pada penggunaan data terkini aplikasi lokal dan kombinasi metode *machine learning* dengan *lexicon-based*[16].

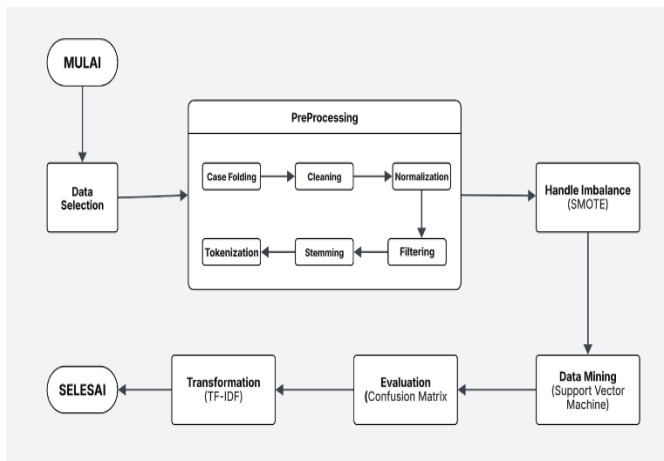
Berdasarkan kinerja dari berbagai penelitian sebelumnya, *SVM* menunjukkan performa terbaik dalam melakukan klasifikasi data. Keunggulan dibuktikan melalui skor akurasi dan *F1-score* memiliki performa yang lebih unggul ketika dibandingkan dengan pendekatan lain, salah satunya *Naive Bayes*, *Random Forest*, dan *KNN*.

Meskipun demikian, sebagian besar penelitian terdahulu cenderung hanya menguji kinerja *SVM* dengan satu jenis kernel saja, seperti *Linear* atau *RBF*, tanpa membandingkan pengaruh kombinasi kernel yang berbeda terhadap hasil klasifikasi. Selain itu, pengaruh variasi rasio pembagian data latih dan data uji terhadap performa model juga jarang dianalisis secara mendalam. Hal ini dapat menjadi keterbatasan dalam menghasilkan model yang optimal dan aplikatif untuk berbagai skenario nyata.

Tujuan dari penelitian ini adalah untuk menyelidiki sentimen yang terdapat dalam ulasan aplikasi Bilibili di *Google Play Store* menggunakan algoritma *Support Vector Machine (SVM)*. Pemilihan algoritma ini didasarkan pada berbagai penelitian sebelumnya yang menunjukkan performanya yang konsisten dan unggul dalam tugas klasifikasi, termasuk dalam konteks analisis sentimen. Dalam konteks ini, ulasan pengguna sering kali mengandung variasi bahasa, ekspresi, dan makna yang beragam, sehingga menimbulkan tantangan tersendiri dalam proses klasifikasi. Oleh karena itu, penelitian ini berupaya mengevaluasi seberapa efektif *SVM* dalam mengatasi tantangan tersebut dan mengklasifikasikan opini pengguna ke dalam sentimen positif atau negatif. Dengan mempertimbangkan kemampuan *SVM* dalam menangani data berdimensi tinggi dan melakukan generalisasi yang baik, penelitian ini menduga bahwa *SVM* memiliki potensi untuk memberikan hasil klasifikasi yang akurat dan efisien meskipun dihadapkan pada kompleksitas teks ulasan pengguna.

II. METODE PENELITIAN

Studi ini menggunakan metode klasifikasi *Support Vector Machine (SVM)* dalam proses *KDD (Knowledge Discovery in Database)* untuk menganalisis data. *KDD* sendiri adalah tahapan dalam data mining yang mencakup analisis, pembuatan model, dan menemukan pola tersembunyi dalam data. Gambar 1 memperlihatkan proses penelitian berdasarkan metode *KDD*.



Gambar 1. Proses Penelitian

A. Data Selection

Pengambilan data dilakukan secara otomatis menggunakan metode *web scraping* dengan bantuan *library google-play-scraper* untuk mengekstrak ulasan langsung dari platform *Google Play*, di mana ulasan yang diambil merupakan ulasan terbaru dengan menggunakan parameter *Sort.NEWEST* agar mencerminkan pengalaman pengguna terkini..

B. Preprocessing

Proses *preprocessing* data dilakukan untuk mengonversi data yang masih belum terorganisir menjadi format yang lebih rapi dan dapat digunakan dalam analisis. Tahapan dalam proses ini mencakup:

1. *Case folding*
2. *Cleaning*
3. *Normalization*
4. *Filtering*
5. *Stemming*
6. *Tokenization*

C. Handle imbalance

Handle imbalance merujuk pada upaya menangani ketidakseimbangan jumlah data antar kelas dalam suatu masalah klasifikasi pada pembelajaran mesin[17]. Ketidakseimbangan kelas dalam analisis sentimen dapat menyebabkan model bias, karena kelas mayoritas mendominasi dan model kurang sensitif terhadap kelas minoritas. Hal ini mengurangi akurasi prediksi dan efektivitas analisis. Oleh karena itu, penanganan ketidakseimbangan kelas penting agar model dapat mengenali semua kelas dengan baik[18]. *SMOTE* (*Synthetic Minority Over-sampling Technique*) adalah metode yang digunakan untuk mengatasi ketidakseimbangan kelas. Teknik ini menambah jumlah data pada kelas minoritas dengan membuat data sintetis baru yang berasal dari kombinasi atau interpolasi data yang sudah ada. Dengan cara ini, *SMOTE* membantu menjaga distribusi kelas mayoritas tetap sama sambil membuat dataset menjadi lebih seimbang[19]. Oleh karena itu, penerapan *SMOTE* dapat membantu model lebih baik dalam mengenali kelas minoritas.

D. Transformation

Tahap transformasi bertujuan untuk mempersiapkan data dalam format yang dapat diproses lebih jauh untuk keperluan analisis data mining. Dalam langkah ini, data akan dibagi menjadi data pelatihan dan data pengujian menggunakan tiga

perbandingan rasio: 90:10, 80:20, dan 70:30. Setelah itu, metode *TF-IDF* digunakan untuk melakukan pembobotan kata sehingga data teks dapat diubah menjadi bentuk vektor berbobot.

E. Data Mining

Dalam penelitian ini, pendekatan *Support Vector Machine* (*SVM*) menggunakan empat tipe kernel yang berbeda, yaitu *Linear*, *Radial Basis Function* (*RBF*), *Sigmoid*, dan *Polynomial*. Setiap kernel memiliki rumus perhitungan sebagai berikut:

1. kernel *Linear* : $k(x_i, x) = x_i \cdot x$
2. kernel *RBF* : $k(x_i, x) = \exp\left(-\frac{\|x_i - x\|^2}{2\sigma^2}\right)$
3. kernel *Polynomial* : $k(x_i, x) = (x_i \cdot x)^d$
4. kernel *Sigmoid* : $k(x_i, x) = \tanh(\sigma(x_i \cdot x) + c)$

F. Evaluation

Tahap evaluasi dilakukan dengan menganalisis kinerja berbagai jenis kernel pada algoritma *SVM*. Evaluasi menggunakan metrik seperti akurasi, presisi, *recall*, dan *F1-score*, yang dihitung berdasarkan *confusion matrix* untuk menilai efektivitas dan konsistensi masing-masing kernel dalam mengklasifikasikan sentimen.

G. Visualisasi Word Cloud

Word Cloud merupakan tampilan grafis yang menggambarkan kumpulan kata-kata dari sebuah teks, di mana kata yang sering muncul akan memiliki ukuran font lebih besar untuk menonjolkan frekuensinya. Tujuan dari penggunaan *Word Cloud* adalah untuk mempermudah dalam menemukan kata-kata utama atau tema yang paling sering muncul dalam sebuah data teks secara cepat dan intuitif.

III. HASIL DAN PEMBAHASAN

A. Data Selection

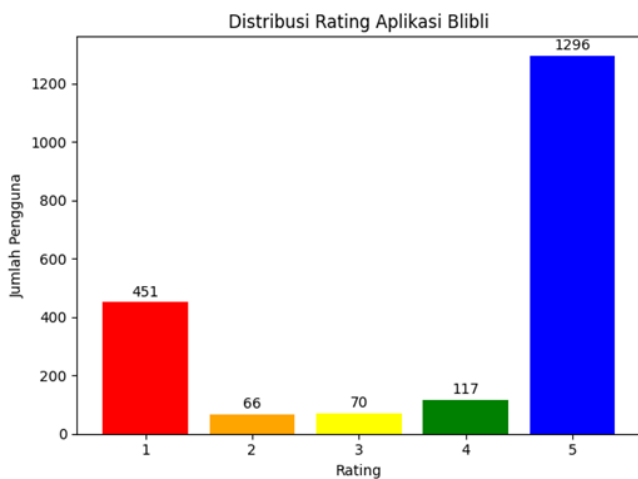
Pada fase Proses pengambilan data dari ulasan aplikasi *Blibli*, langkah ini dilaksanakan menggunakan metode *web scraping*, dengan bantuan pustaka *Google-play-scraper*, *Library* ini menyediakan antarmuka *API* yang mempermudah penarikan data secara langsung dari *Google Play Store* dan dapat digunakan melalui bahasa pemrograman *Python*. Dalam proses ini, peneliti berhasil mengumpulkan sebanyak 2000 ulasan terbaru, yang diambil dalam rentang waktu mulai dari 08 November 2024 hingga 27 April 2025.

userName	rating	tanggal	ulasan
Nikmatul Qodri	5	27/04/2025 04:38	mantap betul
Muhammad Jaka Alf	2	27/04/2025 03:15	banyak toko akun shareing ga jelas trs kalau order udah bayar
Samuel Wijaya	1	27/04/2025 02:29	tidak bisa COD ...barang harga Rp 10000 saja harus transfer...
Aby Adawiyah	5	26/04/2025 23:45	aplikasi nya bagus. tolong makin ditingkatkan ya untuk pilihan
TV MEDIA	3	26/04/2025 22:40	Jujur ya. Murah beli di toko terdekat daripada di Blibli, udah la
BuHar [] Budi Hardi	5	26/04/2025 19:31	Aplikasi Olshop yg Bagus, klw bisa perbanyak promo Gratis On
Afif Hidayat	4	26/04/2025 14:26	baru pertama pake agak bingung si tapi aman
SUHU JETSUN ARAH#	5	26/04/2025 13:30	goodddd
Miftakhul Hasan	1	26/04/2025 11:39	aplikasi lemot. tampilan ribet,harga lbh mahal,dan proses refu
Yupi Yuhana	5	26/04/2025 11:00	Hay Blibli,, aku selalu puas belanja di market mu, selain kemas
Ahmad Murtadla	2	26/04/2025 09:13	penarikan saldo refund ke rekening sangat lama, tidak seperti.
Yani Komariah	5	26/04/2025 07:26	GOOD
Faruq Abiyu	5	26/04/2025 06:01	aplikasinya sangat membantu mudah simpel bagus nih aplikasi
M Rizki as syarif	1	26/04/2025 05:15	Mohon maaf ini kalau mau buat suara notifikasi jangan gede ge
Zen Fahri	5	26/04/2025 03:13	semua jadi praktis trm kasih
Siska Mariska	5	25/04/2025 10:05	apk mudah digunakan

Gambar 2. Hasil Scrapping Data Ulasan Aplikasi Blibli

Pada informasi terkait ulasan aplikasi *Blibli*, terdapat empat atribut, yaitu "*userName*," "*rating*," "*tanggal*," dan "*ulasan*." Untuk keperluan pelabelan dalam penelitian ini, peneliti

memanfaatkan Pemberian label dari *review* aplikasi di *Google Play* dilakukan berdasarkan *rating* yang disampaikan oleh pengguna. Ulasan dengan nilai 1 hingga 2 diasosiasikan sebagai opini bernuansa negatif, nilai 3 diklasifikasikan sebagai sentimen netral, sedangkan nilai 4 dan 5 mencerminkan sentimen positif dari pengguna.



Gambar 3. Jumlah Ulasan Tiap Rating

Gambar 3 menunjukkan bahwa ulasan terbanyak diberikan pada *rating* 5, dengan total sebanyak 1.296 ulasan. Diikuti oleh *rating* 1 yang memperoleh 451 ulasan. Sebaliknya, *rating* 2 menjadi yang paling sedikit memperoleh ulasan, yakni hanya 66 ulasan.

Berdasarkan hasil klasifikasi, sebanyak 1.413 ulasan tergolong dalam sentimen positif, yang diperoleh dari *rating* 4 dan 5. Sentimen negatif terdiri dari 517 ulasan, berasal dari *rating* 1 dan 2. Sementara itu, 70 ulasan dengan *rating* 3 diklasifikasikan sebagai sentimen netral, mengingat *rating* tersebut cenderung merepresentasikan penilaian yang moderat atau ambivalen. Dengan demikian, proporsi ulasan bernuansa positif masih mendominasi dibandingkan dengan kategori netral maupun negatif.

B. Preprocessing

1. Case folding

Proses ini melibatkan transformasi seluruh tulisan ke dalam bentuk lowercase untuk menyeragamkan format teks. Hal ini dilakukan agar kata yang sebenarnya sama tidak dianggap berbeda hanya karena perbedaan huruf besar atau kecil.

Tabel 1. Case folding

NO	Input	Output
1	tidak bisa COD ...barang harga Rp 10000 saja harus transfer...	tidak bisa cod ...barang harga rp 10000 saja harus transfer...
2	Aplikasi OlShop yg Bagus, klw bisa perbanyak promo Gratis OngKir nya.	aplikasi olshop yg bagus, klw bisa perbanyak promo gratis ongkir nya.

2. Cleaning

Tahapan ini bertujuan untuk menghapus komponen-komponen yang tidak memiliki keterkaitan dengan analisis, memastikan bahwa hanya karakter alfabet yang diperlukan yang tersisa dalam data ulasan. Langkah ini mencakup pembersihan data dengan menghapus karakter-karakter khusus, angka, nilai

kosong, serta simbol-simbol seperti tanda seru, persen, tanda tambah, tanda kurung, URL, emoji, dan elemen lain.

Tabel 2. Cleaning

NO	Input	Output
1	kerreeen.... ternyata blibli mantap banget....ðŸ˜ˆ'☐	kerreeen ternyata blibli mantap banget
2	baru pake semoga nyamanðŸ˜ˆ%ðŸ˜ˆ'☐	baru pake semoga nyaman

3. Normalization

Pada tahap ini, tujuan utamanya adalah untuk memperbaiki kesalahan ejaan dan mengubah kata-kata yang tidak tepat menjadi bentuk yang benar, sehingga kata-kata dengan arti yang sama bisa disamakan. Selain itu, singkatan-singkatan yang sering digunakan, seperti "bgt" menjadi "banget" dan "tdk" menjadi "tidak", juga akan diubah menjadi bentuk lengkapnya.

Tabel 3. Normalization

NO	Input	Output
1	aplikasi olshop yg bagus klw perbanyak promo gratis ongkir nya	aplikasi olshop yang bagus kalau perbanyak promo gratis ongkir nya
2	apk yg bermanfaat byk promonya	aplikasi yang bermanfaat banyak promonya

4. Filtering

Filtering atau *Stopwords* Dimaksudkan untuk menghilangkan kata-kata yang tidak memberikan kontribusi penting dalam analisis, seperti 'dan', 'atau', 'yang' dan kata lainnya yang cuma jadi penghubung atau pelengkap saja.

Tabel 4. Filtering

NO	Input	Output
1	tokoh akun shareing tidak terus order sudah bayar tokoh nya tidak bener blibli susah kalau batalin pesanan beda sama tokoh orange	tokoh akun shareing order bayar tokoh nya bener blibli susah batalin pesanan beda tokoh orange
2	tokoh yang proses barangnya cancel otomatis merugikan	tokoh proses barangnya cancel otomatis merugikan

5. Stemming

Pada tahap *Stemming*, kata-kata yang memiliki imbuhan seperti awalan atau akhiran akan disederhanakan menjadi bentuk dasarnya. Misalnya, kata "berlari" akan diubah menjadi "lari". Proses ini bertujuan untuk membuat kata-kata lebih simpel dan sesuai dengan aturan bahasa. Biasanya Tahap ini dilakukan dengan menggunakan perangkat lunak Sastrawi, yang dirancang khusus untuk memproses bahasa Indonesia.

Tabel 5. Stemming

NO	Input	Output
1	mantap belanja motor blibli sungguh mudah terjamin	mantap belanja motor blibli sungguh mudah jamin
2	aplikasi bagus mencari barang	aplikasi bagus cari barang

6. Tokenization

Tahap terakhir dalam proses ini adalah tokenisasi, yaitu proses memisahkan teks menjadi unit-unit kata berdasarkan pemisah seperti spasi atau tanda baca. Langkah ini bertujuan untuk mempermudah analisis teks, karena setiap kata diperlakukan sebagai entitas yang berdiri sendiri sehingga dapat dianalisis secara lebih terstruktur oleh sistem.

Tabel 6. *Tokenization*

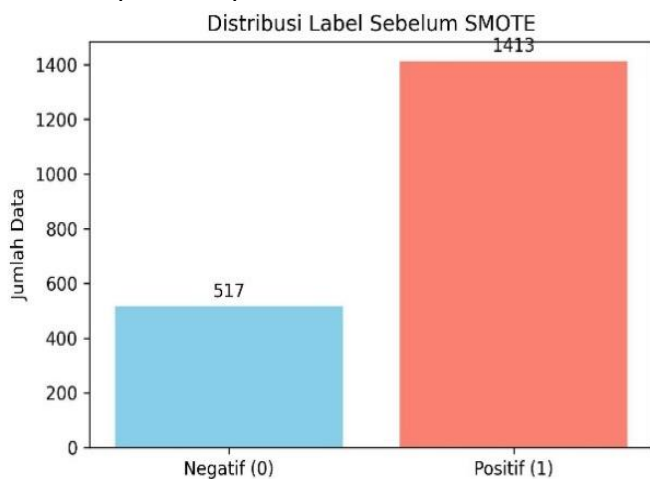
NO	Input	Output
1	aplikasi nya bagus tolong tingkat ya pilih gratis vouchernya terima kasih	['aplikasi', 'nya', 'bagus', 'tolong', 'tingkat', 'ya', 'pilih', 'gratis', 'vouchernya', 'terima', 'kasih']
2	tari saldo refund rekening aplikasi belanja langsung masuk	['tari', 'saldo', 'refund', 'rekening', 'aplikasi', 'belanja', 'langsung', 'masuk']

C. Handle imbalance

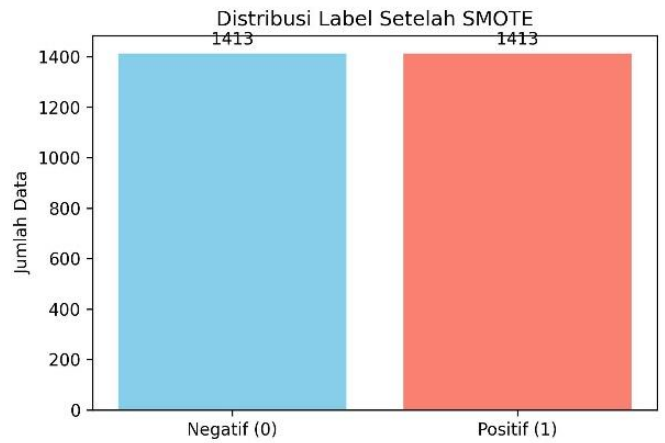
Seperti telah dijelaskan pada bagian hasil, subbab Data Selection, data awal dalam penelitian ini mencakup tiga kategori sentimen, yaitu positif, negatif, dan netral (*rating* 3). Namun, karena fokus penelitian ini adalah pada klasifikasi biner, ulasan dengan *rating* netral dihapus. Penghilangan kelas netral dilakukan untuk menghindari ambiguitas dalam proses pelabelan dan agar model dapat lebih fokus dalam membedakan antara sentimen positif dan negatif secara jelas.

Setelah penghapusan data netral, jumlah data antar kelas masih menunjukkan ketidakseimbangan, dengan dominasi kelas tertentu yang dapat menyebabkan bias dalam pelatihan model. Untuk mengatasi hal tersebut, diterapkan metode *Synthetic Minority Over-sampling Technique (SMOTE)*. *SMOTE* menghasilkan data sintesis pada kelas minoritas sehingga jumlah data pada kedua kelas menjadi seimbang.

Hasil dari penerapan *SMOTE* menunjukkan bahwa jumlah data sentimen positif dan negatif masing-masing menjadi 1.413 ulasan, sehingga total dataset berjumlah 2.826 ulasan. Perbandingan distribusi data sebelum dan sesudah penerapan *SMOTE* dapat dilihat pada Gambar 4 dan Gambar 5.



Gambar 4. Distribusi jumlah klasifikasi sentimen sebelum penerapan *SMOTE*



Gambar 5. Distribusi jumlah klasifikasi sentimen setelah penerapan *SMOTE*

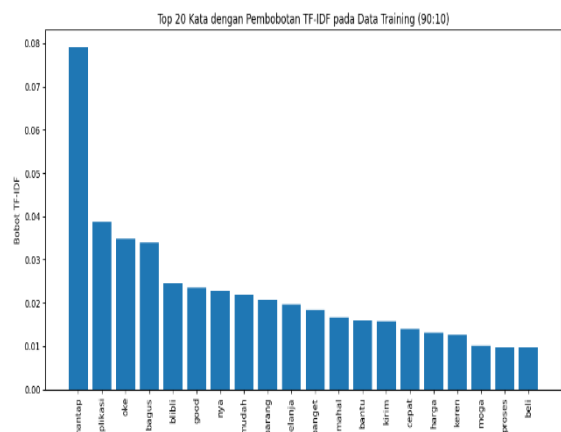
D. Transformation

Pada tahap transformasi, dataset dibagi menjadi data yang digunakan untuk pelatihan dan pengujian dibagi dalam tiga konfigurasi berbeda. Konsistensi pembagian dijaga dengan menggunakan fungsi dan parameter *random_state=42*, sehingga hasil eksperimen tetap stabil. Rincian dari tiap skenario disajikan pada Tabel 7 sebagai dasar evaluasi performa model terhadap variasi data.

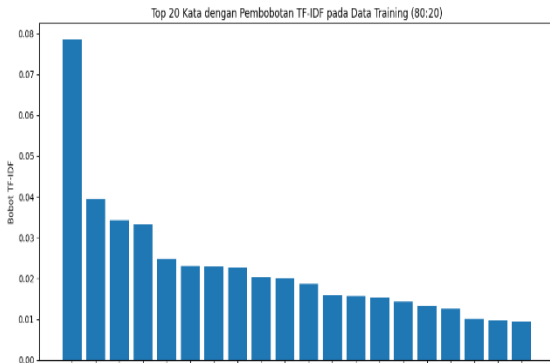
Tabel 7. Pembagian Dataset

Rasio Pembagian		Jumlah Data	Jumlah Data Uji
Data	Data Uji	Latih	
90%	10%	2543	283
80%	20%	2260	566
70%	30%	1978	848

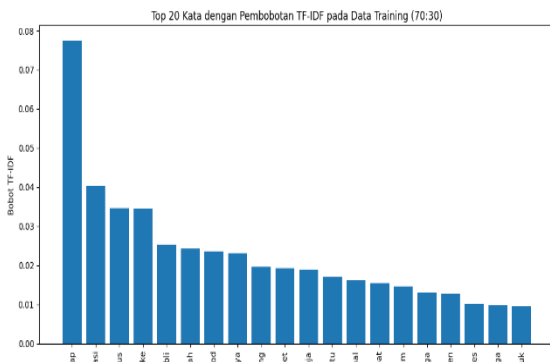
Setelah Proses pemisahan data dilakukan ke dalam dua bagian, yaitu data latih dan data uji. Selanjutnya adalah melakukan transformasi menggunakan metode *TF-IDF* untuk mengubah kata menjadi nilai numerik berdasarkan frekuensi kemunculannya. Hasil pembobotan ini digunakan untuk analisis lebih lanjut, dan visualisasi menampilkan 20 kata dengan bobot tertinggi sesuai kriteria tertentu.



Gambar 6. Contoh penerapan metode *TF-IDF* (90:10)



Gambar 7. Contoh penerapan metode *TF-IDF* (80:20)



Gambar 8. Contoh penerapan metode *TF-IDF* (70:30)

E. Data Mining

Pada tahap klasifikasi sentimen atau data mining ini, data dibagi ke dalam tiga rasio yang bervariasi, yaitu 90:10, 80:20, dan 70:30. Algoritma *Support Vector Machine (SVM)* diterapkan dengan empat tipe kernel yang berbeda, yakni *Linear*, *Radial Basis Function (RBF)*, *Polynomial*, dan *Sigmoid*. untuk mengevaluasi performa model. Hasil pengujian untuk masing-masing kombinasi kernel dan rasio pembagian data disajikan dalam Tabel 8.

Tabel 8. Hasil klasifikasi *SVM*

Skenario	Kernel	Accuracy
90:10	<i>Linear</i>	96,82%
	<i>RBF</i>	95,76%
	<i>poly</i>	85,87%
	<i>Sigmoid</i>	92,23%
80:20	<i>Linear</i>	95,05%
	<i>RBF</i>	93,82%
	<i>poly</i>	84,10%
	<i>Sigmoid</i>	93,64%
70:30	<i>Linear</i>	94,93%
	<i>RBF</i>	92,81%
	<i>poly</i>	84,32%
	<i>Sigmoid</i>	93,16%

F. Evaluation

Tahap evaluasi merupakan bagian akhir dari penelitian ini, di mana dilakukan penilaian terhadap performa model yang dikembangkan menggunakan algoritma *Support Vector Machine (SVM)*. Penilaian ini mengandalkan *confusion matrix* sebagai pendekatan evaluasi dalam menghitung efektivitas model. Pada proses ini, performa model dibandingkan

berdasarkan empat jenis kernel *SVM* yang diterapkan dalam beberapa skenario data mining yang berbeda.

Skenario	Kernel	Akurasi	Precision	Recall	F1-Score
90:10:00	<i>Linear</i>	96,82%	96,93%	96,82%	96,82%
90:10:00	<i>Rbf</i>	95,76%	96,09%	95,76%	95,75%
90:10:00	<i>Poly</i>	85,87%	88,97%	85,87%	85,57%
90:10:00	<i>Sigmoid</i>	92,23%	93,08%	92,23%	92,19%
80:20:00	<i>Linear</i>	95,05%	95,33%	95,05%	95,05%
80:20:00	<i>Rbf</i>	93,82%	94,50%	93,82%	93,79%
80:20:00	<i>Poly</i>	84,10%	87,94%	84,10%	83,69%
80:20:00	<i>Sigmoid</i>	93,64%	94,14%	93,64%	93,62%
70:30:00	<i>Linear</i>	94,93%	95,17%	94,93%	94,92%
70:30:00	<i>Rbf</i>	92,81%	93,65%	92,81%	92,77%
70:30:00	<i>Poly</i>	84,32%	87,94%	84,32%	83,93%
70:30:00	<i>Sigmoid</i>	93,16%	93,67%	93,16%	93,14%

Gambar 9. Perbandingan Hasil Evaluation

Gambar 9 menunjukkan bahwa pada skenario 90:10:00, kernel *Linear* memberikan hasil terbaik dengan akurasi 96,82%, *precision* 96,93%, *recall* 96,82%, dan *F1-score* 96,82%. Hasil tersebut mengindikasikan bahwa model mampu melakukan klasifikasi sentimen positif dan negatif secara akurat dan seimbang.

G. Visualisasi Word Cloud



Gambar 10. wordcloud positif



Gambar 11. Wordcloud Negatif

Berdasarkan gambar 10 dan 11, penggunaan analisis visual melalui wordcloud berhasil mengungkapkan kata-kata yang dominan digunakan oleh pengguna dalam *review* aplikasi di Play Store. Pada *review* dengan sentimen positif, istilah seperti “mantap”, “mudah”, dan “bibli” sering terlihat, menggambarkan pengalaman pengguna yang menyenangkan dengan aplikasi tersebut. Sementara itu, ulasan negatif didominasi oleh kata-kata seperti “mahal”, “aplikasi”, “harga”, dan “kirim”, yang menunjukkan aspek-aspek yang perlu perbaikan oleh pengembang aplikasi.

Studi ini menunjukkan bahwa pendekatan *Support Vector Machine (SVM)*, khususnya memanfaatkan kernel *Linear*,

mampu mengklasifikasikan tanggapan pengguna secara efektif. Analisis yang dilakukan menghasilkan pemahaman mendalam mengenai bagaimana pengguna memandang aplikasi tersebut, yang bisa dimanfaatkan sebagai bahan pertimbangan penting untuk membantu pengembang dalam memperbaiki kualitas dan fungsionalitas aplikasi ke depannya.

IV. KESIMPULAN

Penelitian ini membuktikan bahwa algoritma *Support Vector Machine (SVM)*, khususnya dengan kernel *Linear*, mampu mengklasifikasikan sentimen ulasan pengguna aplikasi Blibli secara akurat dan seimbang. Eksperimen dilakukan dengan membandingkan empat jenis kernel *SVM (Linear, RBF, Polynomial, dan Sigmoid)* serta tiga skenario rasio data latih dan uji (90:10, 80:20, dan 70:30). Setelah dilakukan proses *preprocessing*, pembobotan *TF-IDF*, dan penanganan ketidakseimbangan data menggunakan *SMOTE*, model *SVM* menunjukkan performa terbaik pada skenario 90:10 dengan akurasi 96,82%, *precision* 96,93%, *recall* 96,82%, dan *F1-score* 96,82%. Hasil ini mengindikasikan bahwa *SVM* sangat efektif untuk diterapkan dalam analisis sentimen, dan dapat menjadi alat bantu penting dalam evaluasi kualitas aplikasi berdasarkan ulasan pengguna.

DAFTAR PUSTAKA

- [1] A. A. Muhammad, Ermatita, and D. S. Prasvita, "ANALISIS SENTIMEN PENGGUNA APLIKASI DANA BERDASARKAN ULASAN PADA *GOOGLE PLAY* MENGGUNAKAN METODE SUPPORT VECTOR MACHINE Prodi S1 Informatika / Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta," *Semin. Nas. Mhs. Ilmu Komput. dan Apl.*, pp. 194–204, 2022.
- [2] M. Saifurridho, M. Martanto, and U. Hayati, "Analisis Algoritma K-Nearest Neighbor terhadap Sentimen Pengguna Aplikasi Shopee," *J. Inform. Terpadu*, vol. 10, no. 1, pp. 21–26, 2024, doi: 10.54914/jit.v10i1.1054.
- [3] I. H. Kusuma and N. Cahyono, "Analisis Sentimen Masyarakat Terhadap Penggunaan *E-commerce* Menggunakan Algoritma K-Nearest Neighbor," *J. Inform. J. Pengemb. IT*, vol. 8, no. 3, pp. 302–307, 2023, doi: 10.30591/jpit.v8i3.5734.
- [4] B. Girsang *et al.*, "Analisis Sentimen Aplikasi Lion Parcel Menggunakan Lexicon Based dan Support Vector Machine P - ISSN : 2302-3295," vol. 12, no. 3, 2024.
- [5] A. Bagas Pranata, A. R. Abdillah, and F. Irwiensyah, "KLIK: Kajian Ilmiah Informatika dan Komputer Analisis Sentimen Ulasan Pengguna Aplikasi Netflix Pada *Google Play* Menggunakan Algoritma Naïve Bayes," *Media Online*, vol. 4, no. 6, pp. 3091–3098, 2024, doi: 10.30865/klik.v4i6.1964.
- [6] P. Store, "Algoritma *Naïve Bayes* untuk Analisis Sentiment Review Blibli . com di *Google Play Store Naïve Bayes* Algorithm for Sentiment Analysis of Blibli . com Review on Google," vol. 13, pp. 831–840, 2024.
- [7] M. D. Hendriyanto, A. A. Ridha, and U. Enri, "Analisis Sentimen Ulasan Aplikasi Mola Pada *Google Play Store* Menggunakan Algoritma Support Vector Machine," *INTECOMS J. Inf. Technol. Comput. Sci.*, vol. 5, no. 1, pp. 1–7, 2022, doi: 10.31539/intecom.v5i1.3708.
- [8] M. Xanderina *et al.*, "Analisis Sentimen Ulasan *E-commerce* Shopee Pada *Google Play Store* Menggunakan Machine Learning," *J. Eng. Sustain. Technol.*, vol. 10, no. 2, pp. 990–998, 2024, [Online]. Available: <https://ejournal.unma.ac.id/index.php/j-ensitec/article/view/9071>
- [9] T. M. Permata Aulia, N. Arifin, and R. Mayasari, "Perbandingan Kernel *Support Vector Machine (SVM)* Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19," *SINTECH (Science Inf. Technol. J.)*, vol. 4, no. 2, pp. 139–145, 2021, doi: 10.31598/sintechjournal.v4i2.762.
- [10] N. Pratiwi and Y. Setyawan, "Analisis Akurasi Dari Perbedaan Fungsi Kernel Dan Cost Pada Support Vector Machine Studi Kasus Klasifikasi Curah Hujan Di Jakarta," *J. Fundam. Math. Appl.*, vol. 4, no. 2, pp. 203–212, 2021, doi: 10.14710/jfma.v4i2.11691.
- [11] F. Bei and S. Sudin, "Analisis Sentimen Aplikasi Tiket Online Di Play Store Menggunakan Metode *Support Vector Machine (SVM)*," *Sismatik*, vol. 01, no. 01, pp. 91–97, 2021.
- [12] A. Rohanah, D. L. Rianti, B. N. Sari, T. Informatika, U. S. Karawang, and K. Pelanggan, "Perbandingan *Naïve Bayes* dan Support Vector Machine," vol. 6, no. 1, 2021.
- [13] J. T. Informatika, "J-TIFA," vol. 2617, no. 1, pp. 18–28, 2025.
- [14] M. Husni, A dan Randi, "Jurnal Inovasi Global," *J. Inov. Glob.*, vol. 2, no. 3, pp. 543–551, 2024.
- [15] N. S. Ramadan and D. Darwis, "PERBANDINGAN METODE *NAÏVE BAYES* DAN *SVM* UNTUK SENTIMEN ANALISIS MASYARAKAT TERHADAP SERANGAN RANSOMWARE PADA DATA KIP-K," vol. 8, no. 1, pp. 12–23, 2025.
- [16] A. Lowell, A. Lowell, K. Candra, and E. Indra, "Perbandingan Metode Support Vector Machine (*SVM*) Dan *Naïve Bayes* Pada Analisis Sentimen Ulasan Aplikasi OVO JURNAL MEDIA INFORMATIKA [JUMIN]," vol. 6, no. 2, pp. 896–905, 2025.
- [17] L. Qadrini, H. Hikmah, and M. Megasari, "Oversampling, Undersampling, *SMOTE SVM* dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejava Timur Tahun 2017," *J. Comput. Syst. Informatics*, vol. 3, no. 4, pp. 386–391, 2022, doi: 10.47065/josyc.v3i4.2154.
- [18] F. Dwi Astuti and F. Nova Lenti, "Implementasi *SMOTE* untuk mengatasi Imbalance Class pada Klasifikasi Car Evolution menggunakan K-NN," *J. JUPITER*, vol. 13, no. 1, pp. 89–98, 2021.
- [19] N. S. Ramadhanti, W. A. Kusuma, and A. Annisa, "Optimasi Data Tidak Seimbang pada Interaksi Drug Target dengan Sampling dan Ensemble Support Vector Machine," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, p. 1221, 2020, doi: 10.25126/jtiik.2020762857.