SISTEM ANALISIS SENTIMEN TERHADAP PRODUK SUNSCREEN PADA MARKETPLACE SHOPEE MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)

LAPORAN TUGAS AKHIR

Laporan ini Disusun untuk Memenuhi Salah Satu Syarat Memperoleh Gelar Sarjana Strata 1 (S1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang

Disusun Oleh :

Nama : Thoriq Bahtiar

NIM : 32601900031

FAKULTAS TEKNOLOGI INDUSTRI

UNIVERSITAS ISLAM SULTAN AGUNG

SEMARANG

2024

SENTIMENT ANALYSIS SYSTEM FOR SUNSCREEN PRODUCTS ON THE SHOPEE MARKETPLACE USING SUPPORT VECTOR MACHINE (SVM)

FINAL ASSIGNMENT REPORT

This report is prepared to fulfill one of the requirements to obtain a Bachelor's Degree (S1) in the Informatics Engineering Study Program, Faculty of Industrial Technology, Sultan Agung Islamic University, Semarang.

Compiled By :

Name : Thoriq Bahtiar

NIM : 32601900031

FACULTY OF INDUSTRIAL TECHNOLOGY

SULTAN AGUNG ISLAMIC UNIVERSITY

SEMARANG

2024

LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul **"Sistem Analisis Sentimen Terhadap Produk *Sunscreen* Pada Marketplace Shopee Menggunakan *Support Vector Machine* (SVM)"** ini disusun oleh :

Nama : Thoriq Bahtiar

NIM : 32601900031

Program Studi : S1 Teknik Informatika

Telah disetujui oleh Dosen Pembimbing pada :

Hari : kamis

Tanggal : 5 September 2024

Mengesahkan,

Pembimbing I	Pembimbing II
Badie'ah S.T, M.Kom	Imam Much Ibnu Subroto S.T, M.Sc, Ph.d
NIK. 210615044	NIK. 210600017

Mengetahui,
Ketua Program Studi S1 Teknik Informatika
Fakultas Teknologi Industri
Universitas Islam Sultan Agung

Moch. Taufik S.T, M.IT
NIK. 210604034

LEMBAR PENGESAHAN PENGUJI

Laporan Tugas Akhir dengan judul **"Sistem Analisis Sentimen Terhadap Produk *Sunscreen* Pada Marketplace Shopee Menggunakan *Support Vector Machine* (SVM)"** ini telah dipertahankan di depan dosen penguji Tugas Akhir pada :

Hari : Senin

Tanggal : 2 September 2024

TIM PENGUJI

Penguji I	Penguji II
Bagus Satrio Waluyo Poetra, S.Kom, M.Cs	Andi Riansyah, S.T, M.Kom
NIK. 210616051	NIK. 210616053

SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama	: Thoriq Bahtiar
NIM	: 32601900031
Judul Tugas Akhir	: Sistem Analisis Sentimen Terhadap Produk Sunscreen Pada Marketplace Shopee Menggunakan Support Vector Machine (SVM)

Dengan ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, 20 Agustus 2024

Yang Menyatakan,

Thoriq Bahtiar

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama	: Thoriq Bahtiar
NIM	: 32601900031
Program Studi	: S1 Teknik Informatika
Fakultas	: Teknologi Industri

Dengan ini menyatakan Karya Ilmiah berupa Tugas Akhir dengan judul : **Sistem Analisis Sentimen Terhadap Produk *Sunscreen* Pada Marketplace Shopee Menggunakan *Support Vector Machine* (SVM)** menyetujui menjadi hak milik Universitas Islam Sultan Agung Semarang serta memberikan hak bebas royalti non-eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternetdan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan Agung Semarang.

Semarang, 20 Agustus 2024

Yang menyatakan,

Thoriq Bahtiar

KATA PENGANTAR

Dengan mengucap syukur alhamdulillah atas kehadirat Allah SWT yang telah memberikan rahmat dan karunianya kepada penulis, sehingga dapat menyelesaikan Laporan Tugas Akhir dengan judul "Sistem Analisis Sentimen Terhadap Produk Sunscreen Pada Marketplace Shopee Menggunakan Support Vector Machine (SVM)" ini untuk memenuhi salah satu syarat menyelesaikan studi serta dalam rangka memperoleh gelar sarjana (S-1) pada Program Studi Teknik Informatika Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang.

Tugas Akhir ini disusun dan dibuat dengan adanya bantuan dari berbagai pihak, materi maupun teknis, oleh karena itu saya selaku penulis mengucapkan terimakasih kepada :

1. Rektor UNISSULA Bapak Prof. Dr. H. Gunarto, S.H., M.H yang mengizinkan penulis menimba ilmu di kampus ini.
2. Dekan Fakultas Teknoloogi Industri Ibu Dr. Novi Marlyana, S.T., M.T.
3. Dosen pembimbing I penulis Badie'ah S.T, M.Kom yang telah meluangkan waktu dan memberi ilmu.
4. Dosen pembimbing II penulis Imam Much Ibnu Subroto S.T, M.Sc, Ph.D yang telah meluangkan waktu dan memberi ilmu.
5. Bapak Dahlan dan Ibu Jazimah sebagai orang tua penulis, serta Kakak Ulfa Lutfia dan juga kepada Dede Maskanah atas segala dukungan dan motivasi yang telah diberikan selama proses penyusunan skripsi ini.
6. Dan kepada semua pihak yang tidak dapat saya sebutkan satu persatu.

Dengan segala kerendahan hati, penulis menyadari masih terdapat banyak kekurangan dari segi kualitas atau kuantitas maupun dari ilmu pengetahuan dalam penyusunan laporan, sehingga penulis mengharapkan adanya saran dan kritikan yang bersifat membangun demi kesempurnaan laporan ini dan masa mendatang.

Semarang, 20 Agustus 2024

Thoriq Bahtiar

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

ABSTRAK

Pasar online seperti Shopee telah menjadi platform penting bagi bisnis untuk menjual produk dan layanan kepada khalayak luas, mendorong pertumbuhan platform marketplace. Kemunculan pasar digital mengubah cara konsumen berbelanja dengan menawarkan kenyamanan, variasi, dan harga yang kompetitif. Marketplace menyediakan ruang virtual di mana penjual dapat memamerkan produk mereka, menjangkau pelanggan potensial secara global, dan meningkatkan visibilitas merek. Penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen terhadap produk sunscreen di marketplace Shopee menggunakan metode Support Vector Machine (SVM). Analisis sentimen dilakukan untuk mengevaluasi tingkat kepuasan konsumen terhadap produk sunscreen yang dijual di Shopee. Melalui analisis ini, persepsi konsumen tentang produk tersebut dapat dievaluasi, sehingga dapat memberikan wawasan berharga untuk peningkatan produk dan pengambilan keputusan konsumen. Proses analisis sentimen mencakup tahapan pra-pemrosesan teks, termasuk lowercasing, tokenization, dan penghapusan stop words menggunakan Sastrawi. Berdasarkan hasil penelitian yang telah dilakukan, metode SVM dengan pembobotan TF-IDF menunjukkan performa yang memuaskan, dengan akurasi yang tinggi. Nilai precision tercatat pada angka 0,89, nilai recall sebesar 0,90, dan nilai f1-score mencapai 0,89, yang mengindikasikan kemampuan model dalam mengklasifikasikan sentimen konsumen secara efektif dan konsisten.

Kata kunci : analisis sentimen, marketplace Shopee, produk sunscreen, Support Vector Machine, pra-pemrosesan teks

ABSTRACT

Online marketplaces such as Shopee have become an important platform for businesses to sell products and services to a wider audience, driving the growth of marketplace platforms. The emergence of digital marketplaces has changed the way consumers shop by offering convenience, variety, and competitive prices. Marketplaces provide a virtual space where sellers can showcase their products, reach potential customers globally, and increase brand visibility. This study aims to develop a sentiment analysis system for sunscreen products on the Shopee marketplace using the Support Vector Machine (SVM) method. Sentiment analysis is carried out to evaluate the level of consumer satisfaction with sunscreen products sold on Shopee. Through this analysis, consumer perceptions of the product can be evaluated, so that it can provide valuable insights for product improvement and consumer decision making. The sentiment analysis process includes text pre-processing stages, including lowercasing, tokenization, and stop word removal using Sastrawi. Based on the results of the research that has been done, the SVM method with TF-IDF weighting shows satisfactory performance, with high accuracy. The precision value is recorded at 0.89, the rFecall value is 0.90, and the F1-score value reaches 0.89, which indicates the model's ability to classify consumer sentiment effectively and consistently.

Keywords : sentiment analysis, Shopee marketplace, sunscreen products, Support Vector Machine, text pre-processing

BAB I
PENDAHULUAN

1.1 Latar Belakang

Pasar *online* seperti Shopee telah menjadi *platform* penting bagi bisnis untuk menjual produk dan layanan kepada khalayak luas, mendorong pertumbuhan *platform marketplace*. Munculnya pasar digital telah mengubah cara konsumen berbelanja, menawarkan kenyamanan, variasi, dan harga yang kompetitif. *Marketplace* menyediakan ruang virtual di mana penjual dapat memamerkan produk mereka, menjangkau pelanggan potensial secara global, dan meningkatkan visibilitas merek. *Platform* seperti Shopee memfasilitasi transaksi, mengamankan proses pembayaran, dan menawarkan ulasan dan peringkat pelanggan untuk membangun kepercayaan dan kredibilitas. Keberhasilan pasar didorong oleh faktor-faktor seperti kemajuan teknologi, perubahan preferensi konsumen, dan kebutuhan bisnis untuk beradaptasi dengan lanskap digital. Dengan memanfaatkan data pasar dan analitik, penjual di pasar dapat memperoleh wawasan tentang perilaku konsumen, tren, dan preferensi untuk menyesuaikan penawaran mereka secara efektif (Rahmayanti, 2023).

Analisis sentimen pada *marketplace* Shopee merupakan umpan balik konsumen untuk mengevaluasi tingkat kepuasan dengan salah satu produk yang dijual di *platform*. Analisis ini bertujuan untuk memberikan wawasan tentang persepsi konsumen tentang produk tersebut, membantu menilai keamanan dan efektivitas produk. Dengan menganalisis sentimen konsumen, Shopee dapat menawarkan informasi berharga kepada konsumen untuk pengambilan keputusan saat membeli produk. Proses analisis sentimen melibatkan evaluasi tanggapan konsumen untuk mengidentifikasi umpan balik konsumen untuk peningkatan atau pengembangan produk, berdasarkan hasil analisis sentimen. Melalui analisis sentimen, perbedaan sentimen terhadap berbagai jenis produk di Shopee dapat diidentifikasi, membantu dalam memahami preferensi konsumen dan tingkat kepuasan. Analisis sentimen produk Shopee berkontribusi untuk meningkatkan kualitas produk, memungkinkan penjual dan produsen untuk merespons secara

efektif umpan balik konsumen dan rekomendasi untuk peningkatan produk (Valentini dkk., 2019).

Produk *sunscreen* yang digunakan dalam penelitian ini merupakan jenis *sunscreen* khusus yang hanya diformulasikan untuk memberikan perlindungan optimal terhadap kulit dari paparan sinar matahari yang berlebihan. Produk *sunscreen* yang diformulasikan dengan bahan-bahan aktif yang efektif dalam menyaring sinar ultraviolet (UV) dan mencegah kerusakan kulit yang disebabkan oleh paparan sinar UV yang berbahaya dan tidak mengandung bahan tambahan lain seperti pemutih atau pelembap yang mungkin ditemukan dalam *sunscreen* multifungsi, sehingga seluruh komposisinya dioptimalkan hanya untuk memberikan perlindungan maksimal terhadap sinar matahari.

Oleh karena itu, tujuan peneliti adalah untuk membuat sebuah sistem yang dapat membantu pelanggan membuat keputusan untuk membeli produk *sunscreen* perawatan badan di Shopee. Memungkinkan penjual dan produsen untuk segera menanggapi umpan balik pelanggan dan memberikan rekomendasi untuk produk yang lebih baik, meningkatkan kepercayaan dan kepuasan pelanggan, dan mendukung bisnis yang berada dalam *platform* Shopee.

1.2 Perumusan Masalah

Ulasan konsumen tidak hanya mencerminkan pengalaman pribadi pengguna dengan sebuah produk, tetapi juga memberikan informasi detail tentang efektivitas produk tersebut. Dengan menganalisis ulasan ini, calon pembeli bisa mendapatkan gambaran yang lebih jelas mengenai kelebihan dan kekurangan produk tersebut, sehingga bisa menjadi umpan balik yang berharga untuk peningkatan produk di masa mendatang. Kemudian bagaimana cara mengetahui perbedaan analisis sentimen yang terdapat pada beberapa merek pada *platform marketplace* Shopee tersebut.

1.3 Pembatasan Masalah

Dalam penelitian kali ini ada beberapa batasan masalah yang ditetapkan. Adapun batasan masalah dalam tugas akhir ini adalah :

1. Dataset yang digunakan dalam penelitian ini berasal dari ulasan atau komentar yang dikumpulkan dari berbagai produk *sunscreen* yang dijual di

toko resmi (*official store*) pada *marketplace* Shopee, yang mana toko-toko resmi ini adalah penjual yang telah diverifikasi oleh Shopee.

2. Jumlah data yang diambil untuk penelitian ini sebanyak 636 dataset, yang terdiri dari ulasan dan komentar konsumen yang dikumpulkan dari 6 toko resmi (*official store*) yang ada di platform *marketplace* Shopee. Setiap toko resmi tersebut *(official store)*, diambil sebanyak 106 dataset, sehingga total keseluruhan dataset yang digunakan mencapai 636, memastikan bahwa setiap toko berkontribusi secara merata terhadap jumlah data yang dianalisis dan memberikan sampel yang cukup besar untuk mendapatkan hasil yang akurat dan representatif mengenai sentimen konsumen terhadap produk *sunscreen* yang dijual di toko-toko tersebut.
3. Produk *sunscreen* yang diambil untuk penelitian ini merupakan kombinasi antara produk lokal dan produk impor, yang mencakup berbagai merek yang tersedia di pasar, baik yang diproduksi oleh perusahaan dalam negeri maupun oleh perusahaan luar negeri. Kombinasi ini bertujuan untuk memberikan gambaran yang lebih komprehensif dan mendalam mengenai preferensi dan kepuasan konsumen terhadap produk *sunscreen* di Shopee, memungkinkan analisis perbandingan antara produk yang diproduksi di dalam negeri dengan produk yang diimpor, serta untuk mengidentifikasi kelebihan dan kekurangan masing-masing jenis produk berdasarkan ulasan konsumen. Dengan demikian, penelitian ini dapat memberikan wawasan yang lebih luas dan menyeluruh tentang kualitas dan performa produk *sunscreen* di pasar, serta membantu calon pembeli untuk membuat keputusan yang lebih informasi dan tepat saat memilih produk yang sesuai dengan kebutuhan dan preferensi mereka.

1.4 Tujuan

Tujuan tugas akhir ini adalah untuk menganalisis sentimen konsumen dengan cara mengevaluasi tanggapan mereka terhadap beberapa produk *sunscreen* badan pada *marketplace* Shopee, baik dari segi kualitas, harga, maupun pengalaman penggunaan, sehingga dapat memberikan wawasan yang lebih mendalam mengenai preferensi, kepuasan, dan kekecewaan pelanggan. Dengan menggunakan teknik

pemrosesan bahasa alami dan algoritma pembelajaran mesin, analisis ini dapat mengidentifikasi pola-pola umum dalam tanggapan konsumen yang berguna bagi perusahaan dalam merancang strategi pemasaran dan perbaikan produk yang lebih efektif dan tepat sasaran.

1.5 Manfaat

Dalam upaya meningkatkan efektivitas perbaikan atau pengembangan produk, penulis melakukan analisis mendalam terhadap sentimen pelanggan, yang memungkinkan untuk mengidentifikasi area spesifik yang membutuhkan perhatian, sehingga perbaikan atau inovasi yang diimplementasikan dapat lebih tepat sasaran dan sesuai dengan kebutuhan serta harapan pasar.

Dengan menyediakan informasi yang komprehensif dan berharga, dapat membantu konsumen dalam proses pengambilan keputusan yang lebih baik dan lebih informatif, sehingga mereka dapat memilih produk atau layanan yang paling sesuai dengan kebutuhan dan preferensi mereka, serta merasa lebih percaya diri dengan pilihan yang dibuat. Dengan menggunakan sistem analisis data, dapat melihat dan memahami secara mendalam persepsi konsumen terhadap berbagai produk *sunscreen* yang dijual di *platform e-commerce* Shopee yang diberikan oleh pengguna, sehingga dapat memperoleh wawasan yang berharga mengenai kepuasan, kekhawatiran dan preferensi mereka terhadap produk-produk tersebut.

1.6 Sistematika Penulisan

Sistematika penulisan yang akan digunakan oleh penulis dalam pembuatan laporan tugas akhir adalah sebagai berikut :

Tabel 1. 1 Tabel Sistematika Penulisan

BAB I	:	PENDAHULUAN
		Pada BAB I berisi tentang pemaparan latar belakang pemilihan judul, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian dan sistematika penulisan.
BAB II	:	TINJAUAN PUSTAKA DAN DASAR TEORI
		Pada BAB II berisi tentang pemaparan penelitian-penelitian sebelumnya dan dasar teori yang berguna untuk membantu penulis dalam memahami tentang analisis sentimen.

BAB III	:	METODE PENELITIAN
		Pada BAB III berisi tentang pemaparan tahapan penelitian yang dimulai dari tahapan mendapatkan dataset hingga tahapan klasifikasi dan uji coba data.
BAB IV	:	HASIL DAN ANALISIS PENELITIAN
		Pada BAB IV berisi tentang pemaparan hasil penelitian yang dimulai dari hasil akhir sistem, klasifikasi data uji dan akurasi dari sistem.
BAB V	:	KESIMPULAN DAN SARAN
		Pada BAB V berisi tentang pemaparan kesimpulan yang menunjukkan apakah tujuan dari penelitian ini sudah tercapai dan saran agar penelitian dapat dikembangkan lagi oleh pihak yang ingin melakukan penelitian lebih lanjut.

BAB II
TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Analisis sentimen merupakan metode analisis data yang digunakan untuk menentukan opini, sentimen, atau emosi yang terkandung dalam teks. Penerapan analisis sentimen telah meluas ke berbagai bidang, termasuk pemasaran, layanan pelanggan, dan riset pasar. Dengan berkembangnya *e-commerce*, analisis sentimen memainkan peran penting dalam memahami persepsi konsumen terhadap produk yang dijual di *platform* seperti Shopee. Salah satu metode yang efektif untuk analisis sentimen adalah *Support Vector Machine* (SVM). Sistem analisis sentimen terhadap produk *sunscreen* di Shopee menggunakan *Support Vector Machine* (SVM) adalah pendekatan yang efektif untuk memahami opini konsumen. Dengan analisis sentimen, penjual dan produsen dapat memperoleh wawasan yang mendalam tentang ulasan konsumen dan melakukan perbaikan atau pengembangan produk yang lebih tepat sasaran. Implementasi SVM dalam analisis sentimen menawarkan akurasi yang tinggi dan mampu menangani data berdimensi tinggi, menjadikannya pilihan yang tepat untuk melakukan penelitian ini.

Dari hasil skenario pengujian yang dilakukan untuk analisis sentimen produk kecantikan dari ulasan *Female Daily*. Proses *preprocessing* yang meliputi penghapusan tanda baca, normalisasi, pembersihan data, dan stemming tanpa *stopword*, bersama dengan ekstraksi fitur menggunakan TF-IDF, membantu mempersiapkan data ulasan dengan baik untuk analisis sentimen. Penelitian ini menggunakan *Information Gain* tidak meningkatkan akurasi secara signifikan, penggunaannya bermanfaat dalam mengurangi kompleksitas komputasi dengan mengurangi jumlah fitur yang diproses. Menggunakan *Kernel Sigmoid* pada SVM menunjukkan performa yang lebih baik dibandingkan *kernel linear* dan RBF, dengan mencapai nilai akurasi tertinggi sebesar 85,98%. Hal ini menunjukkan bahwa pemilihan kernel dapat berpengaruh signifikan terhadap hasil akhir dalam analisis sentimen ini (Putri dkk., 2019).

Pertumbuhan *e-commerce* telah mengubah peran media sosial menjadi *platform social commerce*. Hal ini memungkinkan konsumen untuk tidak hanya berinteraksi sosial tetapi juga melakukan transaksi pembelian langsung melalui platform seperti *Female Daily*. Interaksi ini secara signifikan mempengaruhi bagaimana *Word of Mouth Marketing* (WOMM) beroperasi dari yang tradisional menjadi *modern*, di mana ulasan dan rekomendasi konsumen dapat dengan mudah diakses dan dipertimbangkan oleh pengguna lain. Dengan menggunakan metode analisis sentimen menggunakan *Support Vector Machine* (SVM) dan *Naïve Bayes Classifier* (NBC), evaluasi terhadap ulasan produk *The Tea Tree Skin Clearing Toner* dari *The Body Shop* dilakukan. Hasil menunjukkan bahwa SVM dengan *kernel linear* memberikan tingkat akurasi sebesar 86% dengan nilai *Area Under Curve* (AUC) sebesar 0,91, yang lebih tinggi dibandingkan dengan *metode Naïve Bayes* yang memberikan akurasi sebesar 83% dengan AUC 0,82. Oleh karena itu, SVM dengan *kernel linear* direkomendasikan untuk klasifikasi ulasan produk ini karena kinerja yang lebih baik dalam mengidentifikasi sentimen positif, negatif, dan netral (Nabila, 2022).

Pengujian yang dilakukan pada pengaruh proses pembobotan fitur dan seleksi fitur pada *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan *Chi Square* maupun *Term Frequency* (TF) dengan *Chi Square* yang kemudian masuk ke dalam tahapan klasifikasi menggunakan metode *Support Vector Machine* (SVM) pada 2 kernel yaitu *Kernel Linear* dan *Gaussian RBF* untuk mengeksekusi klasifikasi sentiment analysis pada review buku novel berbahasa Inggris. Diperoleh bahwa hasil perfomansi terbaik untuk klasifikasi sentiment analysis pada review buku novel berbahasa Inggris, yaitu pada penggunaan *Kernel Gaussian RBF* untuk setiap kedua pembobotan fitur dengan seluruh nilai persentase seleksi fitur yang digunakan dengan nilai performansi terbaik sebesar 74.2%. Selain itu, dari hasil penelitian pada kedua proses pembobotan fitur dalam satu seleksi fitur serta penggunaan *kernel Linear* pada klasifikasi *Support Vector Machine* (SVM), diperoleh bahwa hasil performansi yang diperoleh sebesar 70,7% jika menggunakan pembobotan fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) dan seleksi fitur *Chi Square*. Sedangkan, hasil performansi yang berhasil

diperoleh sebesar 71% jika menggunakan pembobotan *fitur Term Frequency* (TF) dan seleksi *fitur Chi Square*. Hal ini didapatkan, bahwa dengan penggunaan *kernel Gaussian RBF* untuk kedua proses pembobotan fitur dengan seluruh nilai persentase seleksi fitur yang digunakan, mampu meningkatkan nilai performansi lebih baik jika dibandingkan dengan penggunaan kernel Linear (Kencana C & Sibaroni Y, 2019).

Berdasarkan analisis mendalam yang telah dilakukan oleh (Auliya dkk., 2020), dapat disimpulkan bahwa *tweet* yang diperoleh dengan kata kunci Bukalapak, Shopee, dan Tokopedia cenderung lebih banyak mengarah ke sentimen positif daripada sentimen negatif. Selain itu, performa klasifikasi yang dihasilkan menunjukkan nilai G-mean dan AUC terbaik untuk data uji Bukalapak dengan masing-masing sebesar 0,85 dan 0,86 pada *fold* pertama. Sementara itu, untuk data uji Shopee, nilai G-mean dan AUC terbaik tercatat sebesar 0,76 dan 0,77 pada *fold* ketujuh. Adapun data uji Tokopedia menunjukkan performa yang optimal dengan nilai G-mean sebesar 0,82 dan AUC sebesar 0,83 pada *fold* keenam. Secara keseluruhan, hasil ini menunjukkan bahwa klasifikasi menggunakan *kernel RBF* lebih unggul dibandingkan dengan penggunaan *kernel linier* dalam proses analisis sentimen pada dataset yang digunakan. Hal ini mengindikasikan bahwa pendekatan yang lebih kompleks dan fleksibel dari kernel RBF mampu menangkap pola-pola data yang lebih baik, sehingga menghasilkan performa klasifikasi yang lebih akurat dan andal.

Berdasarkan hasil pengujian yang telah dilakukan (Valentini dkk., 2019), ditemukan bahwa skenario 3 dan skenario 8 memiliki performansi sistem yang lebih baik dalam mengklasifikasikan *tweet* dibandingkan dengan skenario lainnya. Pada skenario 3, yang menggunakan kombinasi TF-IDF dan *stemming*, sistem mencapai akurasi sebesar 81,58%. Sementara itu, pada skenario 8, yang menggunakan kombinasi *word count* dan *stemming*, akurasi yang dicapai adalah 77,56%. Meskipun demikian, perlu diingat bahwa nilai akurasi yang tinggi tidak selalu menjamin bahwa sistem yang dibangun adalah yang paling baik. Dalam penelitian ini, ditemukan bahwa masih terdapat beberapa kesalahan klasifikasi yang biasanya

disebabkan oleh adanya data bernegasi yang tidak ditangani dengan baik. Oleh karena itu, langkah selanjutnya dalam pengembangan sistem ini adalah fokus pada penanganan data bernegasi untuk meningkatkan akurasi dan keandalan sistem klasifikasi di masa depan.

Dari beberapa hasil penelitian diatas peneliti dapat mengambil kesimpulan untuk membuat sebuah sistem yang mengambil data dari sebuah *platform marketplace* yang nantinya diolah untuk dilakukan analisis sentimen berdasarkan komentar yang diberikan oleh pengguna. Dalam sistem yang akan dibuat ini menggunakan metode *Machine Learning* memiliki peran untuk mengolah data mentah yang didapatkan dari *marketplace* tersebut dan data mentah dari proses *Web Scraping* dengan melakukan beberapa pengolahan data kalimat seperti *case folding, stemming, filtering, tokenizing,* dan beberapa pengolahan lain yang mungkin dibutuhkan agar hasil pengujian memiliki akurasi yang tinggi. Untuk metode *Support Vector Machine* (SVM) sendiri memiliki peran untuk mengklasifikasikan sebuah kalimat yang didapat dari komentar *marketplace* tersebut.

2.2 Dasar Teori

2.2.1 Analisis Sentimen

Analisis sentimen adalah jenis studi komputasi yang didasarkan pada pengolahan bahasa alami dan komputasi linguistik yang melihat pendapat, penilaian, evaluasi, sikap, emosi, dan sentimen seseorang terhadap sesuatu seperti barang atau jasa, individu, organisasi, acara, topik dan elemen lainnya. Biasanya, data pengujian berupa ulasan produk online yang menunjukkan perasaan emosional, seperti sedih, senang atau marah, untuk menentukan apakah mendapatkan penilaian positif atau negatif (Auliya dkk., 2020).

Opinion mining atau analisis sentimen merupakan suatu bidang ilmu dari *data mining* yang berguna untuk menganalisis, mengolah, dan mengekstrak data tekstual pada entitas, seperti layanan, produk, individu, organisasi, peristiwa, atau masalah dan topik tertentu. Analisis ini berfungsi untuk mendapatkan sebuah informasi dari suatu himpunan data yang ada. Analisis sentimen adalah penelitian yang baru pada *Natural Language Processing* (NLP) dan bertujuan menemukan subjektivitas

dalam teks maupun mengekstraksi dan menjalankan klasifikasi sentimen pada opini.

Terdapat tiga teknik pada metode klasifikasi sentimen yakni *hybrid approach*, *lexicon based*, dan *machine learning*. Pada era ini, penelitian analisis sentimen dilakukan dengan machine learning karena dapat memprediksi polaritas sentimen (positif negatif, ataupun netral) berdasarkan data *training* pada data *testing*. Proses analisis sentimen sebagaimana diilustrasikan adalah teks tidak teratur, mencakup teks pada *review*, forum, *tweet*, dan *blog*. *Pre-processing* data mencakup proses *tokenisasi*, *stopword removal*, *stemming*, identifikasi sentimen, dan klasifikasi sentimen (Kevin dkk., 2020).

2.2.2 *Support Vector Machine* (SVM)

Support Vector Machine (SVM) adalah teknik pembelajaran dengan banyak kualitas yang diinginkan dan menjadikan algoritma SVM sangat popular. SVM mempunyai dasar teoritis yang kuat dan melakukan klasifikasi lebih akurat daripada kebanyakan algoritma lain di banyak aplikasi. Banyak penelitian telah melaporkan bahwa SVM merupakan metode yang paling akurat untuk klasifikasi teks. SVM juga banyak digunakan dalam klasifikasi sentimen (Idris dkk., 2023).

Support Vector Machine (SVM) adalah salah satu algoritma pembelajaran mesin yang paling populer digunakan dalam analisis sentimen. Menurut penelitian yang dipublikasikan pada tahun 2021 (Sugiarti & Iskandar, 2021), SVM bekerja dengan cara mencari *hyperplane* optimal yang memisahkan data ke dalam kelas-kelas yang berbeda. SVM efektif untuk analisis sentimen karena kemampuannya dalam menangani data dengan dimensi tinggi dan memberikan performa yang baik dalam klasifikasi. SVM digunakan untuk menganalisis sentimen komentar di Shopee, dan hasil menunjukkan bahwa SVM dengan *kernel* tertentu (misalnya, *linear* atau RBF) dapat memberikan akurasi yang tinggi dalam mengklasifikasikan komentar sebagai positif, negatif, atau netral.

SVM terbagi menjadi dua jenis yaitu SVM *data linier* dan SVM *data non-linier*. SVM *data linier* digunakan untuk data yang dapat dipisahkan secara *linier*, jika sebuah *dataset* dapat diklasifikasikan menjadi dua kelas dengan menggunakan

sebuah garis lurus tunggal disebut *linier* dan klasifikasinya disebut sebagai *linier* SVM *Classifier*. Sedangkan SVM *data non-linier* digunakan untuk data yang dapat dipisahkan secara *non-linier*, jika sebuah dataset tidak dapat diklasifikasi menggunakan garis lurus disebut data *non-linier* dan klasifikasinya disebut sebagai *Non-Linier SVM Classifier*.

2.2.3 *Marketplace* Shopee

Marketplace Shopee adalah *platform e-commerce* yang memungkinkan pengguna untuk membeli dan menjual berbagai produk secara online. Sebagai salah satu *marketplace* terbesar di Asia Tenggara, Shopee menawarkan fitur-fitur yang memudahkan transaksi *online*, seperti sistem pembayaran yang aman, layanan pengiriman yang cepat, dan fitur interaktif untuk berinteraksi dengan penjual dan pembeli. Shopee memiliki desain UI yang ramah pengguna dan UX yang responsif, memudahkan pengguna dalam mencari, memilih dan membeli produk serta mengomentari sebuah produk dengan leluasa. Shopee sendiri menawarkan berbagai metode pembayaran, termasuk transfer bank, kartu kredit dan *e-wallet*, yang memberikan fleksibilitas kepada pengguna. Dengan kerjasama bersama berbagai jasa logistik, Shopee memastikan pengiriman yang cepat dan tepat waktu. Sering adanya kampanye promosi seperti diskon besar, gratis ongkir dan *cashback* untuk menarik lebih banyak pengguna (Kusuma, 2023).

2.2.4 *Web Scraping*

Metode pengumpulan data yang dikenal sebagai *web scraping* memungkinkan Anda mendapatkan data dari situs web secara otomatis tanpa harus menyalin data secara manual. Tujuan *web scraping* adalah untuk menemukan informasi tertentu dan kemudian mengumpulkannya di internet yang baru. Teknik *web scraping* berfokus pada pengambilan dan ekstraksi data, yang memungkinkan pencarian lebih mudah. Aplikasi *web scraping* hanya berfokus pada teknik pengambilan dan ekstraksi data dengan berbagai ukuran data (Deviacita dkk., 2019).

2.2.5 *Text Preprocessing*

Metode data *preprocessing* merupakan langkah-langkah yang penting dalam analisis sentimen untuk mempersiapkan data teks agar sesuai untuk analisis. Berikut adalah langkah-langkah untuk data *preprocessing* :

a. *Case folding* dan *Cleansing*. Merupakan konversi seluruh huruf dalam teks menjadi huruf kecil atau huruf besar, tujuannya untuk memastikan konsistensi dalam representasi teks dan pembersihan teks dari kata yang tidak dibutuhkan untuk mengurangi noise, seperti *hastag (#), mention username (@),* angka dan url.

b. *Tokenisasi*. Proses ini adalah memecah teks yang panjang menjadi token atau memisahkan kata-kata menjadi unit yang lebih kecil, Sebagai contoh ketika ada kalimat "Saya pergi ke pasar" dari kalimat tersebut akan dipecah menjadi token atau kata sehingga akan menjadi "Saya", "Pergi", "Ke", "Pasar", biasanya acuan dari pemecahan token adalah spasi.

c. *Stemming*. Penghapusan awal atau akhir kata untuk mengubah kata ke bentuk dasarnya sesuai KBBI. Sebagai contoh Stemming antara lain kata "melihat", "memperlihatkan", "dilihat", "dilihatkan" akan ditransformasi menjadi kata "lihat".

d. Normalisasi. Normalisasi digunakan untuk mengambil kata penting dari hasil proses token. Bisa dengan menggunakan algoritma *stoplist* (membuang kata yang kurang penting). *Stoplist* atau *stopword* adalah kata – kata yang tidak deskriptif yang dapat dibuang dalam pendekatan *bag-of-word* (Darwis dkk., 2020).

2.2.6 TF IDF

Ekstraksi fitur TF-IDF *(Term Frequency-Inverse Document Frequency)* adalah proses mengubah teks menjadi representasi numerik yang mencerminkan pentingnya kata-kata dalam dokumen relatif terhadap korpus dokumen. Teknik ini sangat berguna dalam pemrosesan bahasa alami (NLP) dan pembelajaran mesin, khususnya untuk tugas-tugas seperti klasifikasi teks, penambangan teks, dan pencarian informasi. TF bertugas untuk mengukur frekuensi kemunculan suatu kata

dalam sebuah kalimat tertentu. Sedangkan IDF berguna untuk mengukur seberapa penting kata tersebut dalam sebuah dokumen.

Kata-kata yang sering muncul di banyak dokumen, seperti *stop words* yang meliputi kata-kata umum seperti "dan," "atau," "tetapi," dan "juga," mendapatkan bobot lebih rendah dalam perhitungan TF-IDF. Hal ini disebabkan karena kata-kata tersebut tidak memberikan informasi yang signifikan untuk membedakan satu dokumen dari yang lain. Sebaliknya, kata-kata yang spesifik dan jarang muncul di dokumen lain mendapatkan bobot lebih tinggi. Misalnya, kata-kata teknis atau istilah khusus yang hanya relevan dalam konteks tertentu, seperti *"neural networks"* dalam dokumen tentang pembelajaran mesin, akan memiliki bobot yang lebih tinggi. Dengan demikian, perhitungan TF-IDF efektif dalam menyoroti kata-kata kunci yang penting dan memastikan bahwa fitur-fitur yang paling informatif dan relevan diberikan perhatian lebih besar dalam model pembelajaran mesin. Hal ini membantu dalam meningkatkan akurasi dan efisiensi model dalam tugas-tugas seperti klasifikasi teks, penambangan teks, dan pencarian informasi (Septian dkk., 2019).

Pada tahap pembobotan dapat direpresentasikan kedalam bentuk vektor TF-IDF (*Term Frecuency – Inverse Document Frecuency*). TF-IDF ini dapat menghasilkan sebuah vektor dengan banyak *term* sehingga dapat dikenali tiap kata yang dihitung sebagai satu fitur. Proses TF ini akan menghitung jumlah kemunculan kata pada dataset, untuk menentukan bobot dari masing-masing *term*/kata dalam sebuah kalimat yang ada.

$$tf = 0.5 + 0.5\frac{tf}{\max(tf)} \qquad (1)$$

Keterangan :

t f = Banyaknya kata yang muncul pada sebuah kalimat

max(t f) = Panjang kata dari sebuah kalimat

Pada proses IDF merupakan jumlah kalimat yang berisikan *term* yang terdapat pada sebuah dataset menggunakan persamaan sebagai berikut :

$$idf = In\frac{N}{df} + 1 \quad (2)$$

Keterangan :

In : Logaritma natural

N : Jumlah semua kalimat

df : Jumlah kata pada kalimat

Proses pembobotan akan diklasifikasikan menjadi tiga kelas yaitu positif, negatif dan netral. Dengan ketentuannya apabila *score* kata >0 adalah positif, <0 adalah netral dan selain itu adalah netral (Darwis dkk., 2020).

2.2.7 *Confusion Matrix*

Proses ini adalah langkah krusial dalam pembangunan model pembelajaran mesin, termasuk model *Support Vector Machine* (SVM). Evaluasi hasil bertujuan untuk menilai kinerja model yang telah dilatih dengan menggunakan data pengujian *(test dataset)* yang tidak digunakan selama pelatihan. Beberapa metrik evaluasi umum yang digunakan untuk mengukur kinerja model sebagai berikut :

a. Akurasi adalah salah satu metrik evaluasi yang paling sederhana dan umum digunakan dalam pembelajaran mesin untuk mengukur kinerja model klasifikasi. Akurasi mengukur persentase prediksi yang benar dibandingkan dengan keseluruhan prediksi yang dibuat oleh model. Dengan kata lain, akurasi menunjukkan seberapa sering model membuat prediksi yang benar. Akurasi mudah dihitung dan diinterpretasikan, sehingga sering digunakan sebagai metrik dasar untuk mengevaluasi kinerja model. Menggunakan semua prediksi (baik yang benar maupun salah) untuk memberikan gambaran umum tentang kinerja model (Wolfgang, 2023). Perhitungan proporsi prediksi yang benar dari semua prediksi.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

b. *Precision* adalah salah satu metrik evaluasi yang penting dalam pembelajaran mesin, khususnya dalam tugas klasifikasi. *Precision* mengukur seberapa banyak prediksi positif yang benar dibandingkan dengan total prediksi positif

yang dibuat oleh model. *Precision* sangat berguna dalam konteks di mana biaya kesalahan positif palsu *(false positives)* tinggi. *Precision* sangat berguna ketika biaya kesalahan positif palsu tinggi, seperti dalam deteksi penipuan atau diagnosis penyakit. Memberikan penekanan pada kualitas prediksi positif, memastikan bahwa apa yang diprediksi positif benar-benar positif. Untuk mengukur akurasi dari prediksi positif.

$$Precision = \frac{TP}{TP + FP} \quad (4)$$

c. *Recall*, juga dikenal sebagai *True Positive Rate* (TPR) atau *Sensitivity*, adalah metrik evaluasi dalam *machine learning* yang mengukur seberapa baik model dalam mengenali semua *instance* positif yang ada dalam dataset. *Recall* sangat penting dalam situasi di mana menangkap semua *instance* positif lebih diutamakan daripada menghindari kesalahan positif palsu. *Recall* sangat berguna dalam situasi di mana menangkap semua instance positif sangat penting, seperti dalam diagnosis penyakit atau deteksi keamanan. Menekankan pada kemampuan model dalam mendeteksi semua *instance* positif. Berikut untuk mengukur seberapa baik model mendeteksi semua contoh positif yang sebenarnya.

$$Recall = \frac{TP}{TP + FN} \quad (5)$$

d. *F1 Score* juga merupakan salah satu metrik evaluasi yang sering digunakan dalam analisis klasifikasi untuk menilai kinerja model, terutama ketika data tidak seimbang. *F1 Score* menggabungkan dua metrik penting : *Precision* dan *Recall*, untuk memberikan gambaran yang lebih lengkap tentang seberapa baik model melakukan klasifikasi. *F1 Score* berkisar antara 0 hingga 1. Nilai 1 menunjukkan *precision* dan *recall* yang sempurna, sedangkan nilai 0 menunjukkan bahwa model tidak dapat mendeteksi kelas positif dengan baik. *F1 Score* memberikan gambaran tentang keseimbangan antara *Precision* dan *Recall*, sehingga berguna ketika distribusi kelas tidak seimbang dan ingin memastikan bahwa model tidak hanya mengutamakan salah satu aspek. *F1*

Score digunakan ketika jumlah contoh dari kelas positif sangat kecil dibandingkan kelas negatif, *F1 Score* dapat memberikan gambaran yang lebih baik tentang kinerja model dibandingkan hanya mengandalkan *Accuracy*.

$$F1\ Score = 2\ x\ \frac{Precision\ X\ Recall}{Precision + Recall} \quad (6)$$

Keterangan :

TP = *True Positif*. Adalah jumlah contoh di mana model benar-benar memprediksi kelas positif dan kelas sebenarnya juga positif. Ini berarti model berhasil mengidentifikasi contoh yang benar-benar positif.

TN = *True Negatif*. Adalah jumlah contoh di mana model benar-benar memprediksi kelas negatif dan kelas sebenarnya juga negatif. Ini berarti model berhasil mengidentifikasi contoh yang benar-benar negatif.

FP = *False Positive*. Adalah jumlah contoh di mana model memprediksi kelas positif padahal kelas sebenarnya adalah negatif. Ini berarti model salah dalam memprediksi contoh sebagai positif ketika seharusnya negatif.

FN = *False Negative*. Adalah jumlah contoh di mana model memprediksi kelas negatif padahal kelas sebenarnya adalah positif. Ini berarti model salah dalam memprediksi contoh sebagai negatif ketika seharusnya positif. (Wulan dkk., 2019).

BAB III
METODOLOGI PENELITIAN

3.1 Metode Penelitian

Dalam penelitian ini, metode atau algoritma yang digunakan adalah *Support Vector Machine* (SVM) dan *Term Frequency-Inverse Document Frequency* (TF-IDF). Kombinasi kedua metode ini mencakup SVM sebagai teknik klasifikasi dan TF-IDF sebagai metode penghitungan bobot kata untuk dataset. Dengan kombinasi ini, dihasilkan sebuah sistem analisis sentimen yang efisien. Adapun tahapan yang harus dilakukan dalam penelitian ini, antara lain :

3.1.1 *Web Scraping*

WebHarvy adalah alat *web scraping* visual yang dapat digunakan untuk mengumpulkan dataset dari berbagai situs *web* tanpa perlu menulis kode pemrograman. Berikut adalah langkah-langkah umum untuk mengumpulkan dataset menggunakan *WebHarvy* :

- Instalasi dan Pengaturan Awal. Unduh dan *instal WebHarvy* dari situs resminya. Buka aplikasi dan pilih opsi *"New Configuration"* untuk memulai konfigurasi pengumpulan data baru.
- Pemilihan URL. Masukkan URL situs *web* yang ingin dilakukan proses pengambilan data. *WebHarvy* akan memuat halaman tersebut di dalam *browser* internalnya.
- Pemilihan Data. Setelah halaman dimuat, Anda dapat mengklik elemen-elemen data yang ingin diambil (misalnya, judul artikel, tanggal, penulis, isi teks).
- *WebHarvy* akan otomatis mendeteksi pola dari elemen yang Anda pilih dan menyoroti elemen serupa di halaman.
- Konfigurasi *Paging*. Jika situs *web* memiliki beberapa halaman, Anda dapat mengatur *pagination* dengan mengklik tombol navigasi berikutnya di situs *web*. *WebHarvy* akan mengkonfigurasi *scraping data* dari semua halaman secara otomatis.

- Ekstraksi Data. Setelah semua elemen yang diinginkan dipilih, klik tombol *"Start"* untuk memulai proses *scraping*. Data yang diekstrak akan ditampilkan dalam format tabel di dalam *WebHarvy*.
- Ekspor Data. Data yang telah diekstrak dapat diekspor ke berbagai format seperti CSV, XML, JSON, atau Excel untuk analisis lebih lanjut.

3.1.2 *Text Preprocessing*

a. *Case Folding (Lowercasing)*. Proses normalisasi teks di mana semua karakter dalam *string* diubah menjadi huruf kecil (atau kadang-kadang huruf besar) untuk memastikan perbandingan dan pencarian tidak memperhatikan perbedaan huruf besar dan kecil. Ini berguna dalam konteks seperti pencarian teks, pemrosesan data dan pemrosesan bahasa alami.

Tabel 3. 1 Tabel Contoh *Lowercasing*

Sebelum	Packingan nya tebel bgt padahal cmn beli 1, pengemasan cepat, sesuai pesanan, exp masi 2025
Sesudah	packingan nya tebel bgt padahal cmn beli 1, pengemasan cepat, sesuai pesanan, exp masi 2025

b. Tokenisasi. Proses memecah teks menjadi unit-unit kecil yang disebut token. Token bisa berupa kata, frasa, atau bahkan karakter, tergantung pada tujuan analisis. Tokenisasi adalah langkah awal yang penting dalam banyak aplikasi pemrosesan bahasa alami (NLP) dan analisis teks, seperti pemrograman bahasa, pencarian informasi, dan analisis sentimen.

Tabel 3. 2 Tabel Contoh Tokenisasi

Sebelum	Nyoba produk baru, kemasan baru, praktis dan mungil. Pengiriman cepat. Semoga cocok di kulit aku.
Sesudah	'Nyoba', 'produk', 'baru', ',', 'kemasan', 'baru', ',', 'praktis', 'dan', 'mungil', '.', 'Pengiriman', 'cepat', '.', 'Semoga', 'cocok', 'di', 'kulit', 'aku', '.'

c. *Removing Punctuation*. Menghapus tanda baca adalah proses membersihkan teks dari karakter tanda baca seperti titik, koma, tanda tanya, dan lain-lain. Ini sering dilakukan setelah tokenisasi untuk memastikan bahwa analisis atau pemrosesan teks tidak terganggu oleh tanda baca.

Tabel 3. 3 Tabel Contoh *Removing Punctuation*

Sebelum	oke banget packaging super aman sukak! and over all very well done thank u yaa!
Sesudah	oke banget packaging super aman sukak and over all very well done thank u yaa

d. *Removing Numbers*. Proses menghilangkan angka dari teks atau data untuk tujuan tertentu, seperti pembersihan data, analisis teks, atau pemrosesan bahasa alami. Proses ini sering digunakan ketika angka tidak relevan untuk analisis atau dapat mengganggu hasil analisis.

Tabel 3. 4 Tabel Contoh *Removing Numbers*

Sebelum	Packingan nya tebel bgt padahal cmn beli 1, pengemasan cepat, sesuai pesanan, exp masi 2025
Sesudah	Packingan nya tebel bgt padahal cmn beli, pengemasan cepat, sesuai pesanan, exp masi

e. Penghapusan *Stop Words* Dengan Sastrawi. Menghapus *stop words* dalam bahasa Indonesia menggunakan pustaka *'sastrawi'*, dapat menggunakan modul *'sastrawi'* yang menyediakan daftar *stop words* dan fitur untuk menghapusnya dari teks. *'Sastrawi'* adalah pustaka pemrosesan bahasa alami (NLP) khusus untuk bahasa Indonesia yang sering digunakan dalam analisis teks.

Tabel 3. 5 Tabel Contoh Penghapusan *Stop Words*

Sebelum	Smp udh bner² abs pke nya.. Enk bgt d pke, hasilnya mate tpi gaa berat. Kyk pke bedak aja. Wangi nya jg enk kyk bedak bayi
Sesudah	Smp udh bner² abs pke Enk bgt pke hasilnya mate berat Kyk pke bedak Wangi enk kyk bedak bayi

f. *Stemming. Stemming* adalah proses mengurangi kata-kata ke bentuk dasarnya atau akar kata, sehingga variasi bentuk kata yang berbeda dapat diperlakukan sebagai satu bentuk dasar. Ini penting dalam pemrosesan bahasa alami untuk meningkatkan konsistensi dan efektivitas analisis teks.

Tabel 3. 6 Tabel Contoh *Stemming*

Sebelum	Bagus banget cocok untuk kulit remaja untuk kulit yang kusam harganya terjangkau
Sesudah	Bagus banget cocok untuk kulit remaja untuk kulit yang kusam harga jangkau

g. *Removing Non-Alphanumeric*. Menghapus karakter *non-alphanumeric* dalam bahasa Indonesia melibatkan proses menghilangkan karakter-karakter yang bukan huruf atau angka dari teks. Ini berguna untuk membersihkan teks dari tanda baca, simbol khusus, dan spasi ekstra yang mungkin tidak diperlukan dalam analisis atau pemrosesan data.

Tabel 3. 7 Tabel Contoh *Removing Non-alphanumeric*

Sebelum	PENGIRIMANNYA CEPET BANGET, GA SAMPAI 3 HARI UDAH NYAMPE😭🙏 INI 100% ORI YA GAISSS JADI JANGAN KHAWATIR, PACKINGNYA PAKE KARDUS DAN AMAN BGT, KURIRNYA RAMAH KARENA UDAH LANGGANAN, THANKS EMINAA🥰💔
Sesudah	PENGIRIMANNYA CEPET BANGET GA SAMPAI 3 HARI UDAH NYAMPE INI 100 ORI YA GAISSS JADI JANGAN KHAWATIR PACKINGNYA PAKE KARDUS DAN AMAN BGT KURIRNYA RAMAH KARENA UDAH LANGGANAN THANKS EMINAA

h. *Joining Tokens*. *Joining tokens* adalah proses menggabungkan token-token yang telah dipisahkan (misalnya, setelah teks di *stopword removing* atau *stemming*) menjadi satu *string* atau teks yang utuh. Proses ini sering dilakukan setelah tokenisasi untuk membentuk teks yang bersih dan siap untuk analisis atau pemrosesan lebih lanjut.

Tabel 3. 8 Tabel Contoh *Joining Tokens*

Sebelum	'Nyoba', 'produk', 'baru', ',', 'kemasan', 'baru', ',', 'praktis', 'dan', 'mungil', '.', 'Pengiriman', 'cepat', '.', 'Semoga', 'cocok', 'di', 'kulit', 'aku', '.'
Sesudah	Nyoba produk baru , kemasan baru , praktis dan mungil . Pengiriman cepat . Semoga cocok di kulit aku .

3.1.3 TF-IDF

TF-IDF *(Term Frequency-Inverse Document Frequency)* adalah teknik yang digunakan dalam pemrosesan bahasa alami dan *text mining* untuk mengevaluasi seberapa penting sebuah kata dalam sebuah kalimat relatif terhadap kumpulan kalimat. Berikut adalah penjelasan TF-IDF secara mendetail :

Term Frequency (TF) mengukur seberapa sering sebuah kata muncul dalam sebuah dokumen. Ini memberikan bobot yang lebih tinggi pada kata-kata yang sering muncul dalam dokumen tersebut.

Inverse Document Frequency (IDF) mengukur seberapa penting sebuah kata di seluruh dokumen dalam corpus. Kata-kata yang umum (misalnya, "dan", "ini", "yang") akan mendapatkan bobot yang lebih rendah, karena mereka muncul di banyak dokumen dan tidak memberikan banyak informasi spesifik.

3.1.4 *Support Vector Machine*

Pelatihan model *Support Vector Machine* (SVM) adalah proses kompleks yang melibatkan berbagai tahapan untuk melatih algoritma dengan tujuan menemukan *hyperplane* optimal yang memisahkan data ke dalam kelas yang berbeda dengan margin terbesar, yang dikenal sebagai margin maksimal. Langkah pertama dalam proses ini adalah mengumpulkan data yang relevan yang akan digunakan untuk pelatihan model. Data yang dikumpulkan harus mencakup fitur (*input*) yang merupakan atribut atau karakteristik dari data, serta label (*output*) yang menunjukkan kategori atau kelas yang menjadi target prediksi.

Setelah data dikumpulkan, langkah selanjutnya adalah melakukan pembersihan dan pra-pemrosesan data. Ini termasuk menghilangkan data yang tidak relevan yang mungkin tidak berguna untuk model, menangani nilai yang hilang untuk memastikan bahwa data yang digunakan lengkap dan konsisten, serta

melakukan normalisasi atau standarisasi fitur jika diperlukan untuk memastikan bahwa semua fitur berada dalam skala yang sama. Normalisasi atau standarisasi sangat penting, terutama jika fitur memiliki skala yang berbeda-beda, untuk menghindari dominasi fitur tertentu dalam proses pelatihan model.

Setelah tahap pra-pemrosesan selesai, data harus dibagi menjadi dataset pelatihan (*training dataset*) dan dataset pengujian (*test dataset*). Pembagian ini penting untuk mengevaluasi kinerja model. Dataset pelatihan digunakan untuk melatih model SVM, di mana algoritma belajar dari data untuk menemukan *hyperplane* yang memisahkan kelas-kelas dengan margin terbesar. Proses pelatihan ini melibatkan optimasi untuk memaksimalkan margin dan meminimalkan kesalahan klasifikasi pada data pelatihan.

Selama pelatihan, parameter model seperti parameter C (yang mengontrol *trade-off* antara kesalahan klasifikasi dan margin) dan parameter kernel (yang menentukan jenis kernel yang digunakan seperti *linear, polynomial* atau *radial basis function* (RBF)) dioptimalkan untuk meningkatkan kinerja model. Setelah model dilatih, dataset pengujian digunakan untuk mengevaluasi kinerja model yang telah dilatih. Penggunaan dataset pengujian ini memberikan indikasi seberapa baik model akan bekerja pada data yang belum pernah dilihat sebelumnya, membantu dalam mengidentifikasi potensi *overfitting* atau *underfitting*. Dengan demikian, melalui serangkaian langkah-langkah yang terstruktur ini, pelatihan model SVM memastikan bahwa model yang dihasilkan mampu mengklasifikasikan data dengan akurasi yang tinggi.

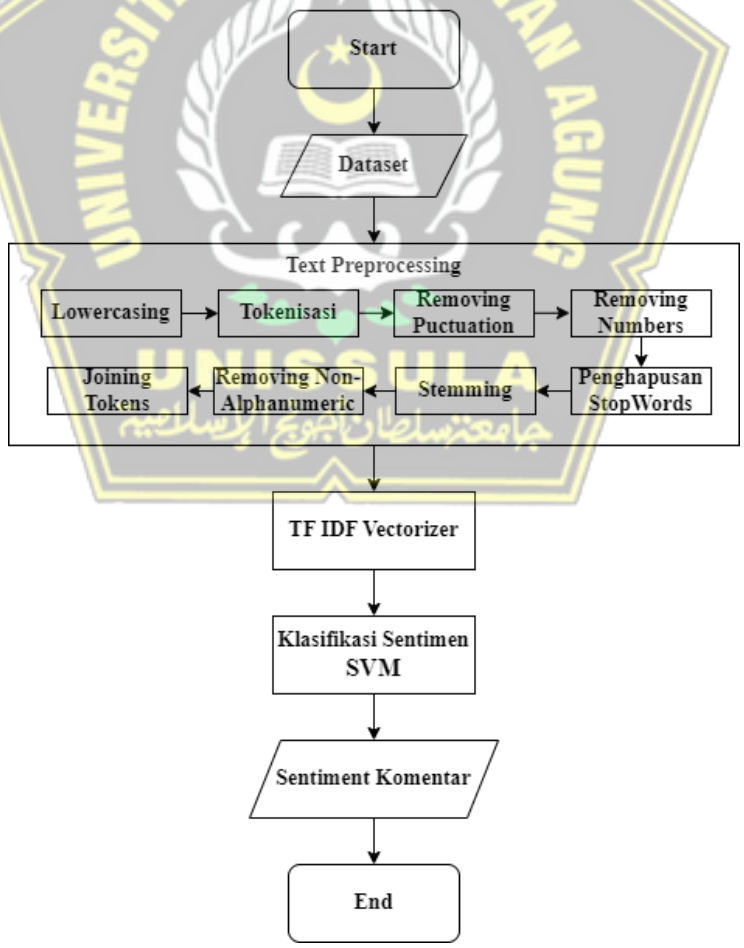
3.1.5 *Uppersampling*

Uppersampling diterapkan pada data train dan bukan pada data asli karena tujuan utama dari *uppersampling* adalah untuk memperbaiki ketidakseimbangan kelas selama pelatihan model. Sebelum *uppersampling*, data train menunjukkan ketidakseimbangan antara jumlah dataset positive yang sebanyak 6.416 dengan dataset negative yang sebanyak 3.436. Untuk mengatasi hal ini dan memastikan bahwa model tidak bias terhadap kelas yang dominan, dilakukan proses uppersampling pada data train, meningkatkan jumlah dataset negative hingga 6.416, setara dengan dataset positive. Dengan cara ini, data train yang digunakan

untuk melatih model menjadi lebih seimbang, memungkinkan model untuk belajar mengenali pola dari kedua kelas dengan lebih baik dan adil. Sementara itu, data test tetap dipertahankan dalam distribusi aslinya untuk memastikan bahwa evaluasi performa model mencerminkan kemampuannya dalam menghadapi data yang tidak seimbang seperti yang mungkin ditemui dalam aplikasi dunia nyata. Pendekatan ini membantu menghasilkan model yang lebih akurat dan dapat diandalkan, serta menghindari risiko *overfitting* yang bisa terjadi jika seluruh dataset diubah distribusinya.

3.1.6 Perancangan Arsiteksur Sistem

Dalam merancang arsitektur suatu sistem, diperlukan flowchart yang menunjukkan langkah-langkah bagaimana sistem berjalan, berikut ini merupakan *flowchart* rancangan model yang akan dibangun :



Gambar 3. 1 *Flowchart* Perancangan Arsitektur Sistem

Berikut adalah penjelasan alur *flowchart* yang akan dibangun :

a. *Start* : Langkah awal dari proses.
b. Dataset : Dataset merupakan hasil scraping komentar produk Shopee yang akan digunakan dalam analisis sentimen.
c. *Text Preprocessing* : Langkah-langkah pemrosesan teks menjadi lebih bersih, untuk menghasilkan sebuah sistem dengan akurasi yang tinggi.
d. *TF-IDF Vectorizer* : Mengubah teks yang sudah diproses menjadi vektor angka menggunakan metode TF-IDF *(Term Frequency-Inverse Document Frequency)* untuk mempresentasikan bobot pentingnya kata dalam suatu dokumen.
e. Klasifikasi Sentiment (SVM) : Menggunakan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen teks berdasarkan vektor yang sudah dihasilkan dari TF-IDF.
f. Sentimen Komentar : Hasil dari klasifikasi sentimen yang menunjukkan apakah sentimen komentar adalah positif atau negatif.
g. *End* : Langkah akhir dari proses.

BAB IV
HASIL DAN ANALISIS PENELITIAN

4.1 Cara Kerja

Pada proses penelitian ini yaitu dengan mengumpulkan ulasan dari Shopee, kemudian melakukan pra-pemrosesan untuk membersihkan dan menyiapkan data, melatih model SVM untuk mengklasifikasikan sentimen dan kemudian mengevaluasi serta mengimplementasikan model untuk menganalisis ulasan baru. Dalam proses pengumpulan data, sistem menggunakan teknik *web scraping* untuk mengambil ulasan pengguna dari halaman produk *sunscreen* di Shopee. Setelah data dikumpulkan, sistem melakukan pra-pemrosesan, yang mencakup penghapusan duplikasi, penghapusan *stop words*, tokenisasi, dan *stemming* atau *lemmatization* untuk memastikan bahwa data siap digunakan dalam tahap pelatihan model. Selanjutnya, fitur-fitur teks ulasan diubah menjadi vektor numerik menggunakan teknik TF-IDF. Data ini kemudian dibagi menjadi set pelatihan dan pengujian untuk melatih model SVM. Model SVM dilatih untuk mengklasifikasikan sentimen ulasan sebagai positif, negatif, atau netral. Setelah model dilatih, kinerjanya dievaluasi menggunakan metrik seperti akurasi, *precision, recall* dan *f1-Score* untuk memastikan model memiliki performa yang baik. Kemudian, model yang telah terlatih diimplementasikan untuk menganalisis sentimen ulasan baru, memberikan prediksi tentang apakah ulasan tersebut positif atau negatif.

4.2 Hasil *Web Scraping*

Proses pengambilan dataset dilakukan menggunakan *software WebHarvy*, sebuah *software* yang efisien untuk *web scraping*. Dalam penelitian ini, data yang diambil berjumlah total 636 dataset, yang masing-masing terdiri dari ulasan-ulasan produk *sunscreen*. Dataset ini diorganisasikan dengan rinci, di mana setiap merk mendapatkan 106 dataset. Merk-merk yang dianalisis meliputi Make Over, Wardah, Emina, Skintific, Garnier, dan CosRX. Pemilihan merk-merk ini

mencakup kombinasi dari merk lokal dan merk impor, memberikan representasi yang luas dan bervariasi dari produk-produk sunscreen yang tersedia di pasar. Make Over, Wardah, dan Emina merupakan merk lokal yang populer di Indonesia, dikenal dengan produk-produk kecantikan yang terjangkau dan berkualitas. Di sisi lain, Skintific, Garnier, dan CosRX adalah merk impor yang telah mendapatkan tempat di hati konsumen Indonesia melalui produk-produk perawatan kulit mereka yang efektif.

Dengan menggunakan *WebHarvy*, proses pengambilan data ulasan menjadi lebih terstruktur dan efisien, memungkinkan pengumpulan data yang akurat dan relevan untuk analisis sentimen. Data yang dikumpulkan ini kemudian akan digunakan dalam tahapan selanjutnya, termasuk pra-pemrosesan, pelatihan model, dan analisis sentimen, untuk memberikan wawasan berharga mengenai persepsi konsumen terhadap berbagai merk *sunscreen* di *marketplace* Shopee. Melalui analisis ini, diharapkan dapat diperoleh informasi yang berguna bagi produsen dan penjual untuk meningkatkan produk dan strategi pemasaran mereka.

Di bawah ini merupakan beberapa contoh hasil yang diperoleh dari proses *web scraping* yang telah berhasil dilaksanakan. Proses ini telah dilakukan dengan cermat untuk mengumpulkan data dari berbagai sumber secara otomatis, dan hasil-hasil berikut menunjukkan informasi yang berhasil diambil dan diproses selama kegiatan tersebut. Dengan menggunakan teknik *web scraping* yang efisien, kami dapat memperoleh data yang relevan dan terkini untuk analisis lebih lanjut.

Tabel 4. 1 Hasil *Web Scraping*

MakeOver	Lamaaaa bgt packing dan pengirimannya. Aman dan lengkap seluruh produk pesanannya, tanggal expired nya aman. Untuk bonusnya itu dipakenya gimana yah sama powdernya aja beda ukuran tempatnya hehe
Wardah	Pengiriman cepat, harga yang terjangkau dan produk sesuai dengan deskripsi 👍 Aku sih ngerasanya ini ringan bngt dipakai cepet nyerep gtu jdi gk bikin wajah jdi white cast luv deh ❤️

Emina	Pengiriman nya rapih dan respon sangat cepat. Rekomendasi deh beli di sini. Lain kali beli lagi. Semoga tetap menjadi toko yang amanah
Skintific	Dari kemaren pengen nyoba, akhirnya nyoba juga. Wangii yaa ternyata, cocok dibawa kemana aja nih simple tinggal semprot. Packingnya amann bgt tebel bublewarp nyaa 🤗🤗
Garnier	Profil Kecantikan: jenis kulit kering dan kusam Kemasan: keren Manfaat: belum di coba Semoga cocok yaa..
CosRX	Produknya blm saya coba ke wajah langsung, cuman pas saya coba ke tangan wangi terus lembut juga. Thank you!!! Pengirimannya sama pengemasannya cepet banget 🤩🤩👍👍

4.3 Hasil Implementasi Text Preprocessing

1. *Lowercasing*

Pada proses ini yaitu mengubah semua karakter dalam *string text* menjadi huruf kecil. Ini penting dalam pra-pemrosesan teks untuk memastikan bahwa perbedaan antara huruf besar dan kecil tidak mempengaruhi analisis. Berikut ini beberapa contoh implementasi *lowercasing* yang sudah berhasil dilakukan :

Sebelum :

Tabel 4. 2 Sebelum *Lowercasing*

Lamaaaa bgt packing dan pengirimannya. Aman dan lengkap seluruh produk pesanannya, tanggal expired nya aman. Untuk bonusnya itu dipakenya gimana yah sama powdernya aja beda ukuran tempatnya hehe
Tekstur: Cukup kental.. tdk bgtu cair Daya serap: Cepat meresap Efektivitas: Lumayan sih utk prlindungan outdoor....nyerap ke bedak aku krn akuh pake make over powersatay jg Sprtinya cocok pke make over.... Byakin potongan diskon hrga ya dan srg2 live ksih voucher dan koin😊
Baru pertama kali beli sunscreennya, biasanya sih cocok pakai produk make over. Semoga sunscreennya juga cocok. Expired masih lama (2025), pengemasannya selalu rapi pake bbw dan dus, makasih banyak seller..^^

Sesudah :

Tabel 4. 3 Hasil Sesudah Lowercasing

lamaaaa bgt packing dan pengirimannya. aman dan lengkap seluruh produk pesanannya, tanggal expired nya aman. untuk bonusnya itu dipakenya gimana yah sama powdernya aja beda ukuran tempatnya hehe
tekstur: cukup kental.. tdk bgtu cair daya serap: cepat meresap efektivitas: lumayan sih utk prlindungan outdoor....nyerap ke bedak aku krn akuh pake make over powersatay jg sprtinya cocok pke make over.... byakin potongan diskon hrga ya dan srg2 live ksih voucher dan koin😊
baru pertama kali beli sunscreennya, biasanya sih cocok pakai produk make over. semoga sunscreennya juga cocok. expired masih lama (2025), pengemasannya selalu rapi pake bbw dan dus, makasih banyak seller..^^

2. *Tokenization*

Proses ini melakukan tokenisasi pada teks *input* menggunakan fungsi word_tokenize. Tokenisasi adalah proses memecah teks menjadi unit-unit kecil yang disebut token, yang biasanya berupa kata atau tanda baca. Berikut ini beberapa contoh implementasi tokenisasi yang sudah berhasil dilakukan :

Sebelum :

Tabel 4. 4 Sebelum Tokenisasi

Make up no. 1 sih emang ini ga perlu ragu, harga atas dikit dg yg lain emang sesuai kinerjanya
Packing baik Cuma agak lama di pengemasan nya Kecepatan pengiriman baik Variasi sesuai dengan pesanan Terima kasih seller, shopee & kurir
Tekstur: lembut dan krim Daya serap: cepat menyerap Efektivitas: belum tau Barang sampai dengan selamat dan aman. Walau telat sehari.

Sesudah :

Tabel 4. 5 Hasil Sesudah Tokenisasi

'make', 'up', 'no.', '1', 'sih', 'emang', 'ini', 'ga', 'perlu', 'ragu,', 'harga', 'atas', 'dikit', 'dg', 'yg', 'lain', 'emang', 'sesuai', 'kinerjanya'

'packing', 'baik', 'cuma', 'agak', 'lama', 'di', 'pengemasan', 'nya', 'kecepatan', 'pengiriman', 'baik', 'variasi', 'sesuai', 'dengan', 'pesanan', 'terima', 'kasih', 'seller,', 'shopee', '&', 'kurir'
'tekstur:', 'lembut', 'dan', 'krim', 'daya', 'serap:', 'cepat', 'menyerap', 'efektivitas:', 'belum', 'tau', 'barang', 'sampai', 'dengan', 'selamat', 'dan', 'aman.', 'walau', 'telat', 'sehari.'

3. *Removing Punctuation*

Pada proses ini memfilter token untuk menghapus tanda baca. Langkah ini penting untuk memastikan bahwa tanda baca seperti koma, titik, tanda seru, dll., tidak mengganggu analisis teks. Fungsi ini untuk memeriksa setiap token dalam tokens dan hanya menyertakan token yang bukan tanda baca. Berikut ini beberapa contoh implementasi *removing punctuation* yang sudah berhasil dilakukan :

Sebelum :

Tabel 4. 6 Sebelum Removing Punctuation

'make', 'up', 'no.', '1', 'sih', 'emang', 'ini', 'ga', 'perlu', 'ragu,', 'harga', 'atas', 'dikit', 'dg', 'yg', 'lain', 'emang', 'sesuai', 'kinerjanya'
'packing', 'baik', 'cuma', 'agak', 'lama', 'di', 'pengemasan', 'nya', 'kecepatan', 'pengiriman', 'baik', 'variasi', 'sesuai', 'dengan', 'pesanan', 'terima', 'kasih', 'seller,', 'shopee', '&', 'kurir'
'tekstur:', 'lembut', 'dan', 'krim', 'daya', 'serap:', 'cepat', 'menyerap', 'efektivitas:', 'belum', 'tau', 'barang', 'sampai', 'dengan', 'selamat', 'dan', 'aman.', 'walau', 'telat', 'sehari.'

Sesudah :

Tabel 4. 7 Hasil Sesudah *Removing Punctuation*

'make', 'up', 'no', '1', 'sih', 'emang', 'ini', 'ga', 'perlu', 'ragu', 'harga', 'atas', 'dikit', 'dg', 'yg', 'lain', 'emang', 'sesuai', 'kinerjanya'
'packing', 'baik', 'cuma', 'agak', 'lama', 'di', 'pengemasan', 'nya', 'kecepatan', 'pengiriman', 'baik', 'variasi', 'sesuai', 'dengan', 'pesanan', 'terima', 'kasih', 'seller', 'shopee', 'kurir'

'tekstur', 'lembut', 'dan', 'krim', 'daya', 'serap', 'cepat', 'menyerap', 'efektivitas', 'belum', 'tau', 'barang', 'sampai', 'dengan', 'selamat', 'dan', 'aman', 'walau', 'telat', 'sehari'

4. *Removing Numbers*

Menghapus angka *(removing numbers)* dari teks adalah salah satu langkah dalam *preprocessing* data teks yang sering digunakan dalam *Natural Language Processing* (NLP). Tujuan utama dari langkah ini adalah untuk mengurangi *noise* dalam data dan fokus pada teks yang memiliki arti semantik lebih jelas. Berikut ini beberapa contoh implementasi removing numbers yang sudah berhasil dilakukan :

Sebelum :

Tabel 4. 8 Sebelum *Removing Numbers*

Perjuangan 9.9 bangun tengah malem buat dapetin voicher disc 50% dr Emina cm buat co sunscreen doang, anw thankyouu ya emina, packagingnya baru ya jd lucu, simple, thankyou thankyou
pesan hari jumat tgl 7 juli baru sampai tanggal 15 juli,huhu,lama sekali,udh chat admin nya katanya tunggu 3×24 jam,eh malah lebih,tapi gapapa kak,ttp suka bgt,ringan bgt dipake di wajah,langsung menyerap,trimakasi emina
Packingan nya tebel bgt padahal cmn beli 1, pengemasan cepat, sesuai pesanan, exp masi 2025

Sesudah :

Tabel 4. 9 Hasil Sesudah *Removing Numbers*

'perjuangan', 'bangun', 'tengah', 'malem', 'buat', 'dapetin', 'voicher', 'disc', 'dr', 'emina', 'cm', 'buat', 'co', 'sunscreen', 'doang', 'anw', 'thankyouu', 'ya', 'emina', 'packagingnya', 'baru', 'ya', 'jd', 'lucu', 'simple', 'thankyou', 'thankyou'
'pesan', 'hari', 'jumat', 'juli', 'baru', 'sampai', 'tanggal', 'juli', 'huhu', 'lama', 'sekali', 'udh', 'chat', 'admin', 'nya', 'katanya', 'tunggu', 'eh', 'malah', 'lebih', 'tapi', 'gapapa', 'kak', 'ttp', 'suka', 'bgt', 'ringan', 'bgt', 'dipake', 'di', 'wajah', 'langsung', 'menyerap', 'trimakasi', 'emina'
'packingan', 'nya', 'tebel', 'bgt', 'padahal', 'cmn', 'beli', 'pengemasan', 'cepat', 'sesuai', 'pesanan', 'exp', 'masi'

5. Menghapus *Stop Words*

Menggunakan *StopWordRemoverFactory* bertujuan untuk memudahkan pembuatan objek yang digunakan untuk menghapus *stop words* dari teks. Dengan objek ini, kita mendapatkan kemampuan untuk menghapus *stop words* dari teks secara efisien. Proses ini menghasilkan teks yang sudah bersih dari *stop words*, sehingga hanya kata-kata yang lebih bermakna yang tersisa untuk analisis lebih lanjut, seperti analisis sentimen atau pemodelan teks. Berikut ini beberapa contoh implementasi menghapus *stop words* yang sudah berhasil dilakukan :

Sebelum :

Tabel 4. 10 Sebelum Menghapus *Stop Words*

Perjuangan 9.9 bangun tengah malem buat dapetin voicher disc 50% dr Emina cm buat co sunscreen doang, anw thankyouu ya emina, packagingnya baru ya jd lucu, simple, thankyou thankyou
pesan hari jumat tgl 7 juli baru sampai tanggal 15 juli,huhu,lama sekali,udh chat admin nya katanya tunggu 3×24 jam,eh malah lebih,tapi gapapa kak,ttp suka bgt,ringan bgt dipake di wajah,langsung menyerap,trimakasi emina
Packingan nya tebel bgt padahal cmn beli 1, pengemasan cepat, sesuai pesanan, exp masi 2025

Sesudah :

Tabel 4. 11 Hasil Sesudah Menghapus *Stop Words*

'perjuangan', 'bangun', 'malem', 'dapetin', 'voicher', 'disc', 'dr', 'emina', 'cm', 'co', 'sunscreen', 'doang', 'anw', 'thankyouu', 'emina', 'packagingnya', 'baru', 'jd', 'lucu', 'simple', 'thankyou', 'thankyou'
'pesan', 'hari', 'jumat', 'juli', 'sampai', 'tanggal', 'juli', 'huhu', 'lama', 'sekali', 'udh', 'chat', 'admin', 'katanya', 'tunggu', 'eh', 'malah', 'lebih', 'tapi', 'gapapa', 'kak', 'ttp', 'suka', 'bgt', 'ringan', 'bgt', 'dipake', 'wajah', 'langsung', 'menyerap', 'trimakasi', 'emina'
'packingan', 'tebel', 'padahal', 'cmn', 'beli', 'pengemasan', 'cepat', 'sesuai', 'pesanan', 'exp', 'masi'

6. *Stemming*

Untuk memudahkan pembuatan objek *stemmer* yang akan digunakan dalam proses *stemming*, dapat menggunakan *StemmerFactory* untuk menghasilkan objek *stemmer* yang dapat mengubah kata-kata menjadi bentuk dasarnya. Dengan mengaplikasikan objek *stemmer* ini, kita dapat menyederhanakan kata-kata ke bentuk dasarnya, yang membantu dalam menyamakan kata-kata dengan variasi morfologis, sehingga analisis teks menjadi lebih konsisten dan akurat. Berikut ini beberapa contoh implementasi stemming yang sudah berhasil dilakukan :

Sebelum :

Tabel 4. 12 Sebelum *Stemming*

Perjuangan 9.9 bangun tengah malem buat dapetin voicher disc 50% dr Emina cm buat co sunscreen doang, anw thankyouu ya emina, packagingnya baru ya jd lucu, simple, thankyou thankyou
pesan hari jumat tgl 7 juli baru sampai tanggal 15 juli,huhu,lama sekali,udh chat admin nya katanya tunggu 3×24 jam,eh malah lebih,tapi gapapa kak,ttp suka bgt,ringan bgt dipake di wajah,langsung menyerap,trimakasi emina
Packingan nya tebel bgt padahal cmn beli 1, pengemasan cepat, sesuai pesanan, exp masi 2025

Sesudah :

Tabel 4. 13 Hasil Sesudah *Stemming*

'perjuang', 'bangun', 'tengah', 'malem', 'buat', 'dapet', 'voicher', 'disc', 'dr', 'emina', 'cm', 'buat', 'co', 'sunscreen', 'doang', 'anw', 'thankyouu', 'emina', 'packag', 'baru', 'jd', 'lucu', 'simpl', 'thankyou', 'thankyou'
'pesan', 'hari', 'jumat', 'juli', 'baru', 'sampai', 'tanggal', 'juli', 'huhu', 'lama', 'sekali', 'udh', 'chat', 'admin', 'kata', 'tunggu', 'eh', 'malah', 'lebih', 'tapi', 'gapapa', 'kak', 'ttp', 'suk', 'bgt', 'ringan', 'bgt', 'dipak', 'wajah', 'langsung', 'menyerap', 'trimakasi', 'emina'
'packag', 'tebel', 'bgt', 'padahal', 'cmn', 'beli', 'pengemas', 'cepat', 'sesuai', 'pesan', 'exp', 'masi'

7. *Removing Non-Alphanumeric*

Langkah ini dilakukan untuk menghapus token yang mengandung karakter *non-alfanumerik*, seperti tanda baca dan simbol, dari daftar tokens. Ini membantu memastikan bahwa hanya kata-kata yang valid dan relevan yang tersisa untuk analisis lebih lanjut, mengurangi kebisingan dalam data teks. Berikut ini beberapa contoh implementasi *removing non-alphanumeric* yang sudah berhasil dilakukan :

Sebelum :

Tabel 4. 14 Sebelum *Removing Non-Alphanumeric*

Bagus Banget , Padahal Cuma Pesan 2 Barang Kecil , Tapi Di Kasih Kardus + Bubble Wrap Tebal😉 Thankyou Seller 🙏😊 HARUS BANGET SIH DICONTOH SAMA TOKO LAIN 💢
Alhamdulilah barang diterima dengan sangat baik 😇 packing aman 👍 respon penjual sangat baik ❤️ diproses .. dikemas dan langsung dikirim dihari pemesanan 🤗
Parah emina sebagus itu udah lama pake ini🤩🤩🤩🖤🖤pengemasan aman pakai kardus terus buble n rapih jugaa😭😭😭🖤 Sebagus itu emang emina buat para remaja Pengiriman cepet juga😋😘 Thanks emina🤩🤩🖤🖤😘

Sesudah :

Tabel 4. 15 Hasil Sesudah *Removing Non-Alphanumeric*

'bagus', 'banget', 'padahal', 'cuma', 'pesan', 'barang', 'kecil', 'tapi', 'di', 'kasih', 'kardus', 'bubble', 'wrap', 'tebal', 'thankyou', 'seller', 'harus', 'banget', 'sih', 'dicontoh', 'sama', 'toko', 'lain'
'alhamdulilah', 'barang', 'diterima', 'dengan', 'sangat', 'baik', 'packing', 'aman', 'respon', 'penjual', 'sangat', 'baik', 'diproses', 'dikemas', 'dan', 'langsung', 'dikirim', 'dihari', 'pemesanan'
'parah', 'emina', 'sebagus', 'itu', 'udah', 'lama', 'pake', 'ini', 'pengemasan', 'aman', 'pakai', 'kardus', 'terus', 'buble', 'n', 'rapih', 'jugaa', 'sebagus', 'itu', 'emang', 'emina', 'buat', 'para', 'remaja', 'pengiriman', 'cepet', 'juga', 'thanks', 'emina'

8. *Joining Tokens*

Langkah ini dilakukan untuk menggabungkan kembali token-token yang telah diproses menjadi satu *string* yang siap untuk digunakan dalam analisis lebih lanjut atau disimpan dalam format yang mudah dibaca. Misalnya, setelah memproses teks dengan berbagai teknik pra-pemrosesan, hasil akhirnya biasanya diperlukan dalam bentuk string untuk analisis atau penyimpanan. Berikut ini beberapa contoh implementasi *joining tokens* yang sudah berhasil dilakukan :

Sebelum :

Tabel 4. 16 Sebelum *Joining Tokens*

'bagus', 'banget', 'padahal', 'cuma', 'pesan', 'barang', 'kecil', 'tapi', 'di', 'kasih', 'kardus', 'bubble', 'wrap', 'tebal', 'thankyou', 'seller', 'harus', 'banget', 'sih', 'dicontoh', 'sama', 'toko', 'lain'
'alhamdulilah', 'barang', 'diterima', 'dengan', 'sangat', 'baik', 'packing', 'aman', 'respon', 'penjual', 'sangat', 'baik', 'diproses', 'dikemas', 'dan', 'langsung', 'dikirim', 'dihari', 'pemesanan'
'parah', 'emina', 'sebagus', 'itu', 'udah', 'lama', 'pake', 'ini', 'pengemasan', 'aman', 'pakai', 'kardus', 'terus', 'buble', 'n', 'rapih', 'jugaa', 'sebagus', 'itu', 'emang', 'emina', 'buat', 'para', 'remaja', 'pengiriman', 'cepet', 'juga', 'thanks', 'emina'

Sesudah :

Tabel 4. 17 Hasil Sesudah *Joining Tokens*

bagus banget padahal cuma pesan barang kecil tapi kasih kardus bubble wrap tebal thankyou seller harus banget sih dicontoh sama toko lain
alhamdulilah barang diterima sangat baik packing aman respon penjual sangat baik diproses dikemas langsung kirim dihari pemesanan
parah emina sebagus udah lama pake ini pengemasan aman pakai kardus terus buble rapih jugaa sebagus emang emina buat para remaja pengiriman cepet thanks emina

4.4 TF-IDF

Pada proses perhitungan TF-IDF, yang merupakan teknik penting dalam pemrosesan teks dan analisis informasi. Teknik ini bertujuan untuk mengevaluasi seberapa penting sebuah kata dalam dokumen relatif terhadap kumpulan dokumen yang lebih besar, dengan mempertimbangkan frekuensi kemunculan kata dalam dokumen serta seberapa jarang kata tersebut muncul di seluruh dokumen. Kode program ini menjalankan berbagai langkah, mulai dari tokenisasi teks hingga perhitungan dan normalisasi nilai TF-IDF, untuk menghasilkan representasi yang berguna bagi analisis teks dan pencarian informasi. Kode program berikut ini digunakan untuk menghitung nilai *Term Frequency-Inverse Document Frequency* :

```
vectorizer = TfidfVectorizer(min_df = 5,
                             max_df = 0.8,
                             sublinear_tf = True,
                             use_idf = True)
```

`TfidfVectorizer` merupakan kelas dari pustaka scikit-learn yang mengubah teks menjadi matriks fitur TF-IDF *(Term Frequency-Inverse Document Frequency)*.

`min_df = 5` berfungsi untuk mengabaikan kata-kata yang muncul di kurang dari 5 dokumen. Ini membantu mengurangi dimensi matriks dengan menghapus kata-kata yang sangat jarang.

`max_df = 0.8` berfungsi untuk mengabaikan kata-kata yang muncul di lebih dari 80% dokumen. Ini membantu menghapus kata-kata yang terlalu umum dan tidak memberikan informasi yang relevan.

`sublinear_tf = True` yaitu menggunakan logaritma untuk menghitung frekuensi kata (TF) daripada frekuensi mentahnya. Ini membantu mengurangi dampak kata yang sangat sering muncul.

`use_idf = True` yaitu menggunakan IDF *(Inverse Document Frequency)* dalam perhitungan, yang memberikan bobot lebih pada kata-kata yang jarang muncul di seluruh korpus dokumen.

Berikut ini merupakan hasil perhitungan TF-IDF :

Train Data with TF-IDF:

	18th	2x	3x	50k	abad	abang	abc	abdi	abis	abu	...	zaadit	zadit	zaman	zon	zona	zonk	zoo	zupa	zuppa	text
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	makan ramai antre menu masakan masakan standar...
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	sagoo kichen kopi paris van java kafe nya hada...
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	malu in isu bohong melulu lu jokowi
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	hari sudah tiga kali nikmat lantai selalu dika...
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	walaupun tempat nya kecil sangat bantu rasa ma...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
12827	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	suka roti mereka saji roti iris sangat tipis p...
12828	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.294876	0.0	0.0	0.0	0.0	0.0	fadli zon amin rais gerindra orang ingin hancu...
12829	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	sering main sekitar dago beehive kafe eatery t...
12830	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	bohong banget tropicana slim bikin diabetes ni...
12831	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.000000	0.0	0.0	0.0	0.0	0.0	sama keluarga mes banyak menu sini total tagih...

12832 rows × 3718 columns

Gambar 4. 1 Hasil Perhitungan TF-IDF

4.5 *Support Vector Machine* (SVM)

Pada proses pelatihan model *Support Vector Machine* (SVM), yang merupakan metode pembelajaran mesin yang populer untuk klasifikasi dan regresi, kode program berikut ini digunakan untuk mengimplementasikan berbagai langkah kritis dalam pelatihan model. Proses ini melibatkan pengolahan data *input*, penentuan parameter model, pelatihan model menggunakan dataset yang telah disiapkan, serta evaluasi kinerja model berdasarkan metrik yang relevan. Kode ini mencakup tahapan seperti pemisahan data menjadi set pelatihan dan set pengujian, penerapan teknik normalisasi atau standardisasi data jika diperlukan, serta optimasi hyperparameter untuk mencapai performa terbaik dari model SVM yang dihasilkan. Berikut ini adalah kode programnya :

```
classifier_linear = svm.SVC(kernel='linear')
classifier_linear.fit(train_vectors, data_train['Target'])
prediction_linear = classifier_linear.predict(test_vectors)
```

Pada proses pelatihan model *Support Vector Machine* (SVM) dengan kernel linear, metode ini digunakan untuk memodelkan hubungan antara fitur teks yang telah diubah menjadi representasi TF-IDF dan label target dalam dataset. Kernel linear mengasumsikan bahwa data dapat dipisahkan dengan hyperplane linier, yang berarti model berusaha menemukan garis pemisah yang optimal untuk memisahkan kelas-kelas dalam ruang fitur. Selama tahap `fit`, model SVM dilatih menggunakan data pelatihan untuk belajar pola-pola yang ada dalam data, dengan tujuan

mengoptimalkan parameter model sehingga dapat memisahkan data dengan sebaik mungkin. Setelah model terlatih, tahap `predict` dilakukan untuk menerapkan model pada data uji, menghasilkan prediksi tentang label target, dan memungkinkan evaluasi kinerja model berdasarkan akurasi dan metrik evaluasi lainnya.

4.6 Hasil *Confusion Matrix*

Hasil evaluasi yang mencakup analisis kinerja model pada data test serta perbandingannya dengan kinerja pada data train (data pelatihan), memberikan wawasan mendalam mengenai efektivitas model dalam memprediksi label target. Proses ini melibatkan perhitungan berbagai metrik evaluasi seperti akurasi, *precision, recall* dan *f1-score* yang membantu dalam mengidentifikasi sejauh mana model dapat generalisasi pada data yang belum pernah dilihat sebelumnya, serta bagaimana model tersebut beradaptasi dengan data pelatihan. Dengan membandingkan hasil evaluasi antara data test dan data pelatihan, kita dapat menilai potensi overfitting atau underfitting, serta mengevaluasi apakah model memerlukan penyetelan lebih lanjut atau perubahan dalam strategi pelatihan untuk meningkatkan kinerja secara keseluruhan.

Tabel 4. 18 Hasil *Confusion Matrtix*

	Precision	Recall	F1-Score	Support
Poitive	0,89	0,90	0,89	735
Negative	0,90	0,89	0,89	735
Accuracy			0,89	1470
Macro Avg	0,89	0,89	0,89	1470
Weighted Avg	0,89	0,89	0,89	1470

Secara keseluruhan, hasil evaluasi model memiliki performa yang baik dengan accuracy sebesar 0.89, serta *precision, recall* dan *f1-score* yang relatif seimbang antara kelas positif dan negatif.

4.7 Hasil *Uppersampling*

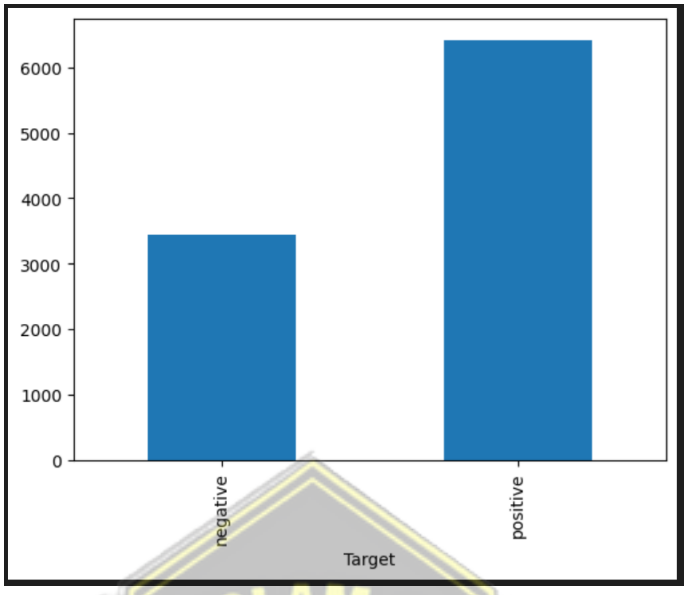
Uppersampling adalah teknik yang digunakan dalam pemrosesan data untuk menangani ketidakseimbangan kelas dalam sebuah dataset. Ketidakseimbangan

kelas terjadi ketika satu kelas memiliki jumlah instance yang jauh lebih banyak dibandingkan kelas lain, yang sering kali menyebabkan model belajar untuk lebih fokus pada kelas mayoritas dan mengabaikan kelas minoritas.

Pada analisis sentimen, kelas biasanya dibagi menjadi sentimen positif dan negatif. Jika data tidak seimbang (misalnya, lebih banyak data dengan sentimen positif dibandingkan sentimen negatif), model bisa jadi cenderung bias dan memiliki performa yang buruk dalam mendeteksi sentimen minoritas. Keuntungan dari uppersampling adalah memungkinkan model untuk belajar lebih baik dari instance kelas minoritas, meningkatkan performa prediksi untuk kelas tersebut. Namun, penting untuk berhati-hati karena *uppersampling* dapat menyebabkan *overfitting* jika tidak dilakukan dengan benar, terutama jika hanya menggunakan duplikasi sederhana.

Pada penelitian ini, peneliti melakukan *uppersampling* pada model yang sudah ada, yaitu pada data train, dengan cara menaikkan jumlah dataset negatif agar sesuai dengan jumlah dataset positif. Proses uppersampling ini dilakukan untuk mengatasi ketidakseimbangan kelas dalam dataset, yang sering kali menjadi tantangan dalam analisis data dan pembelajaran mesin. Dengan meningkatkan jumlah sampel dari kelas negatif hingga setara dengan kelas positif, dapat mengurangi bias dan meningkatkan akurasi model prediksi. Pendekatan ini juga membantu dalam memastikan bahwa model memiliki performa yang lebih stabil dan konsisten ketika diterapkan pada data baru yang belum pernah dilihat sebelumnya. Berikut ini merupakan hasil *uppersampling* pada data train :

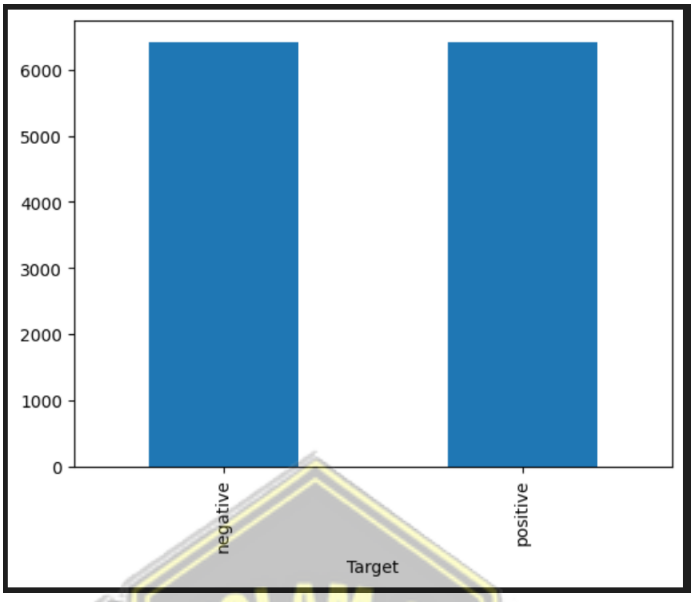
Data train sebelum di *uppersampling* :



Gambar 4. 2 Data Train Sebelum Di *Uppersampling*

Sebelum proses *uppersampling* dilakukan, jumlah dataset *positive* adalah sebanyak 6.416 dataset, sedangkan jumlah dataset *negative* hanya sebanyak 3.436 dataset. Ketidakseimbangan ini dapat menyebabkan model prediksi lebih cenderung mengidentifikasi sampel sebagai *positive*, mengabaikan atau mengurangi akurasi dalam mendeteksi sampel *negative*. Oleh karena itu, untuk mengatasi masalah ini, peneliti melakukan uppersampling pada dataset *negative* hingga mencapai jumlah yang setara dengan dataset *positive*.

Data train setelah proses *uppersampling* :

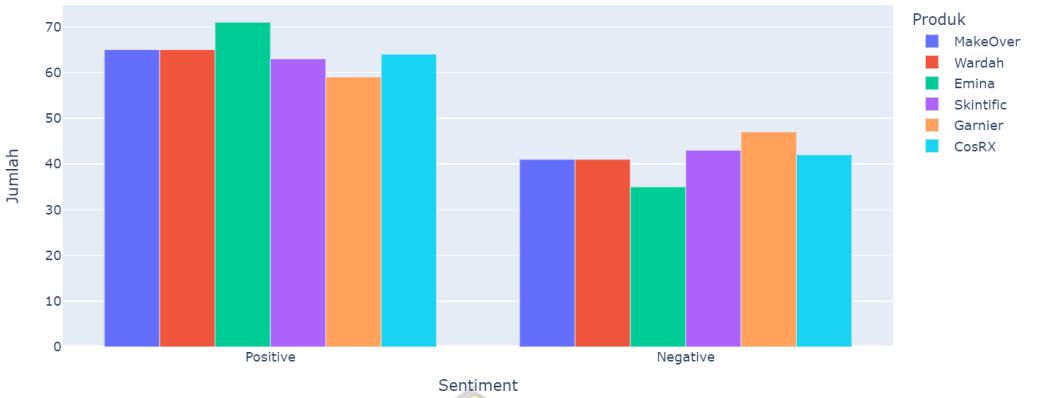


Gambar 4. 3 Data Train Setelah Di *Uppersampling*

Setelah proses *uppersampling* dilakukan, dataset *negative* mengalami kenaikan data sebesar 2.980 sampel, sehingga jumlah total dataset *negative* menjadi 6.416 sampel, setara dengan jumlah dataset *positive*. Dengan demikian, jumlah keseluruhan dataset *positive* dan *negative* pada data train mencapai 12.832 sampel. Proses *uppersampling* ini bertujuan untuk mengatasi ketidakseimbangan antara kelas *positive* dan *negative* yang sebelumnya ada dalam dataset, sehingga dapat meningkatkan akurasi dan performa model dalam memprediksi kedua kelas tersebut. Langkah ini sangat penting dalam memastikan bahwa model yang dilatih memiliki kemampuan yang lebih baik dalam mengenali pola dari kedua kelas secara adil dan akurat.

4.8 Hasil Penelitian

Berikut ini merupakan hasil analisis sentimen yang sudah dilakukan terhadap berbagai ulasan pengguna, di mana kami mengidentifikasi pola-pola emosional yang dominan, baik yang bersifat positif maupun negatif, untuk memberikan wawasan yang lebih mendalam mengenai persepsi umum terhadap produk atau layanan yang ditinjau.



Gambar 4. 4 Hasil Analisis Sentimen

Grafik batang yang ditampilkan menggambarkan jumlah sentimen positif dan negatif dari beberapa produk kecantikan, yaitu MakeOver, Wardah, Emina, Skintific, Garnier, dan CosRX. Berikut adalah penjelasan hasil dari grafik tersebut :

- MakeOver :
  Sentimen Positif : 65
  Sentimen Negatif : 41
- Wardah :
  Sentimen Positif : 65
  Sentimen Negatif : 41
- Emina :
  Sentimen Positif : 71
  Sentimen Negatif : 35
- Skintific :
  Sentimen Positif : 63
  Sentimen Negatif : 43
- Garnier :
  Sentimen Positif : 58
  Sentimen Negatif : 48
- CosRX
  Sentimen Positif : 64
  Sentimen Negatif : 42

Produk dengan Sentimen Positif Tertinggi : Emina memiliki jumlah sentimen positif tertinggi, mencapai sebanyak 71.

Produk dengan Sentimen Negatif Tertinggi : Garnier memiliki jumlah sentimen negatif tertinggi, mencapai sebanyak 48.

Sebagian besar produk memiliki distribusi sentimen yang relatif seimbang, dengan jumlah sentimen positif yang sedikit lebih tinggi daripada sentimen negatif. MakeOver, Wardah, dan CosRX memiliki jumlah sentimen positif dan negatif yang

hampir sama, menunjukkan bahwa ketiga produk ini diterima dengan baik oleh pengguna dengan jumlah ulasan positif yang signifikan dibandingkan dengan ulasan negatif.

Secara keseluruhan, grafik ini menunjukkan bahwa mayoritas produk memiliki ulasan positif yang lebih banyak daripada ulasan negatif, dengan Emina sebagai produk yang paling disukai berdasarkan jumlah sentimen positif yang diterima. Garnier, meskipun memiliki jumlah sentimen positif yang cukup tinggi, juga memiliki sentimen negatif yang cukup signifikan, menunjukkan adanya beberapa ketidakpuasan di antara pengguna.

BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berfokus pada analisis sentimen terhadap ulasan produk *sunscreen* di *platform marketplace* Shopee. Data ulasan dikumpulkan menggunakan teknik *web scraping* dan dianalisis menggunakan metode *Support Vector Machine* (SVM). Berikut adalah poin-poin utama dari hasil penelitian ini :

- Model SVM dilatih untuk mengklasifikasikan sentimen ulasan sebagai positif & negatif. Evaluasi model dilakukan menggunakan *confusion matrix* dengan nilai akurasi yang cukup bagus, nilai *precision* tercatat pada angka 0,89, nilai *recall* sebesar 0,90 dan nilai *f1-score* mencapai 0,89. Model ini menunjukkan performa yang baik dalam mengklasifikasikan sentimen ulasan.
- Dengan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa kombinasi algoritma *Support Vector Machine* (SVM) dan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) yang dijalankan menunjukkan performa yang cukup bagus dalam menjalankan analisis sentimen.
- Penelitian ini dapat memberikan wawasan berharga bagi produsen dan penjual dalam meningkatkan produk dan strategi pemasaran mereka berdasarkan analisis sentimen konsumen di Shopee. Analisis sentimen ini memungkinkan identifikasi area spesifik yang membutuhkan perhatian, sehingga perbaikan atau inovasi yang diimplementasikan dapat lebih tepat sasaran dan sesuai dengan kebutuhan serta harapan pasar.

5.2 Saran

Penelitian selanjutnya diharapkan dapat fokus pada penanganan data yang tidak sesuai, seperti kata-kata yang memiliki makna negatif yang mungkin mempengaruhi akurasi klasifikasi sentimen. Dengan peningkatan dalam pengolahan data tidak sesuai, akurasi sistem klasifikasi diharapkan dapat lebih baik

Dan disarankan untuk mencoba membandingkan berbagai metode dan model machine learning lainnya, seperti *Random Forest* atau *Neural Networks*, untuk melihat apakah ada peningkatan performa dalam klasifikasi sentimen. Eksperimen dengan berbagai teknik pemrosesan teks dan ekstraksi fitur juga dapat dilakukan untuk meningkatkan hasil klasifikasi.

DAFTAR PUSTAKA

Nabila, A. (2022). *Analisis Sentimen Ulasan Produk Toner Pada Beauty Brand "The Body Shop" Menggunakan Metode Naïve Bayes Classifier Dan Support Vector Machine: Studi Kasus Di Female Daily*. 1–87.

Auliya, A. D., Subanti, S., & Zukhronah, E. (2020). *Implementasi Text Mining Pada Analisis Sentimen Pengguna Twitter Terhadap Marketplace di Indonesia Menggunakan Algoritma Support Vector Machine*.

Darwis, D., Pratiwi, E. S., Ferico, A., & Pasaribu, O. (2020). Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. Dalam *Jurnal Ilmiah Edutic* (Vol. 7, Nomor 1).

Deviacita, D., Sasty, H., & Muhardi, H. (2019). *Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace*. *7*(4).

Sugiarti, D. I., & Iskandar, R. (2021). *Pengaruh Consumer Reviewterhadap Keputusan Pembeli Terhadap Toko Online Shopee*.

Wolfgang, E. (2023). *Machine Learning for Brain Disorders*. http://www.springer.com/series/7657

Putri, N. F., Al Faraby, S., & Dwifebri, M. (2019). *Analisis Sentimen pada Produk Kecantikan dari Ulasan Female Daily Menggunakan Information Gain dan SVM Classifier*.

Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). *Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Mengunakan Algoritma Support Vector Machine (SVM)*.

Septian, J. A., Fahrudin, T. M., & Nugroho, A. (2019). *Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF-IDF dan K-Nearest Neighbor*. https://t.co/9WloaWpfD5

Kencana, C. G., & Sibaroni, Y. (2019). *Klasifikasi Sentiment Analysis pada Review Buku Novel Berbahasa Inggris dengan Menggunakan Metode Support Vector Machine (SVM)*.

Kevin, V., Que, S., Iriani, A., & Purnomo, H. D. (2020). Analisis Sentimen Transportasi Online Menggunakan Support Vector Machine Berbasis Particle

Swarm Optimization (Online Transportation Sentiment Analysis Using Support Vector Machine Based on Particle Swarm Optimization). Dalam *Jurnal Nasional Teknik Elektro dan Teknologi Informasi* | (Vol. 9, Nomor 2). www.tripadvisor.com,

Kusuma, I. S. H. (2023). Pengaruh Online Customer Review terhadap Keputusan Pembelian pada Marketplace Shopee di Kalangan Mahasiswa Kota Bandung. *International Journal Administration Business and Organization*, *4*(2), 31–39. https://doi.org/10.61242/ijabo.23.266

Rahmayanti, N. P. (2023). Pengaruh Marketplace Dan Pembayaran Digital Terhadap Tingkat Penjualan Umkm Di Kota Banjarmasin. *Jurnal Komunikasi Bisnis dan Manajemen*, *10*(1).

Valentini, R., Siwabessy, P., Herdiani, A., & Romadhony, A. (2019). *Analisis Sentimen Masyarakat Terhadap Hasil Kerja Petahana Dalam Kaitan Dengan Pemilihan Presiden tahun 2019 Pada Sosial Media Twitter Menggunakan Support Vector Machine (SVM)*.

Wulan, S., Vitandy, U., Supianto, A. A., & Abdurrachman Bachtiar, F. (2019). *Analisis Sentimen Evaluasi Kinerja Dosen menggunakan Term Frequency-Inverse Document Frequency dan Naïve Bayes Classifier* (Vol. 3, Nomor 6). http://j-ptiik.ub.ac.id