

**PENGUNAAN METODE SUPPORT VECTOR
MACHINE UNTUK ANALISIS SENTIMEN PADA
ULASAN PENGGUNA APLIKASI DANA PADA
GOOGLE PLAY STORE**

**OLEH
ANDRES FERNANDA RAHIM
T3119002**

SKRIPSI

Untuk memenuhi salah satu syarat ujian
guna memperoleh gelar Sarjana



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2024**

PERSETUJUAN SKRIPSI

PENGGUNAAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI DANA PADA GOOGLE PLAY STORE

Oleh
ANDRES FERNANDA RAHIM
T3119002

SKRIPSI

Untuk memenuhi salah satu syarat ujian
Guna memperoleh gelar Sarjana
Program Studi Teknik Informatika
Ini telah disetujui oleh tim pembimbing

Gorontalo, September 2024

Pembimbing 1



Yasin Aril Mustofa, M.Kom
NIDN.0926088503

Pembimbing 2



Abd Rahmat Karim Haba M.Kom
NIDN.0923118703

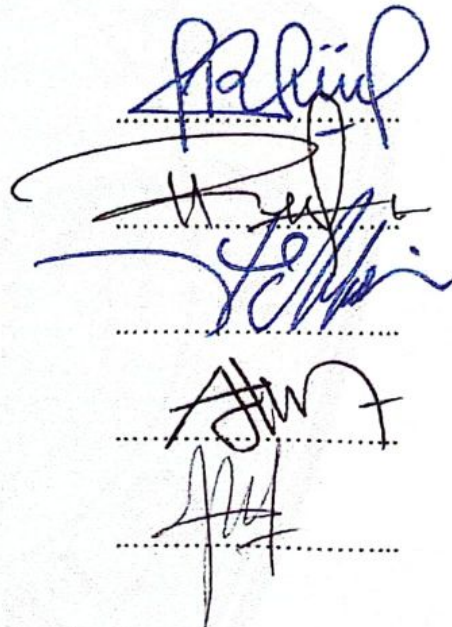
PENGESAHAN SKRIPSI

PENGUNAAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI DANA PADA GOOGLE PLAY STORE

Oleh
ANDRES FERNANDA RAHIM
T3119002

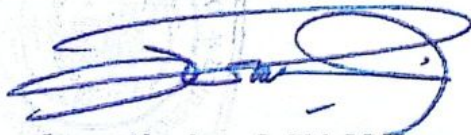
Diperikasa Oleh Panitia Ujian Strata Satu (S1)
Universitas Ichhan Gorontalo

1. Ketua Penguji
Rezqiwati Ishak, M.Kom
2. Anggota
Rofiq Harun, M. Kom
3. Anggota
Serwin, M. Kom
4. Anggota
Yasin Aril Mustofa, M. Kom
5. Anggota
Abd. Rahmat Karim Haba, M. Kom



Mengetahui

Dekan Fakultas Ilmu Komputer



Irvan Abraham Salihi, M.Kom
NIDN : 0918077302

Ketua Program Studi



Sudirman S. Panna, M.Kom
NIDN : 0924038025

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa:

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah di ajukan untuk mendapatkan gelar akademik (Sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan dari Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis di cantumkan sebagai acuan/sitasi dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Kota,

Yang Membuat Pernyataan


Andres Fernanda Rahim

ABSTRACT

ANDRES FERNANDA RAHIM. T3119002. USING THE SUPPORT VECTOR MACHINE METHOD FOR SENTIMENT ANALYSIS ON USER REVIEWS OF THE DANA APPLICATION ON GOOGLE PLAY STORE

This research aims to develop a sentiment classification system for users of the DANA application on the Google Play Store using the Support Vector Machine (SVM) method. It also seeks to apply the SVM method to enhance the accuracy and effectiveness of the classification. The research employs a quantitative approach with a descriptive analysis. Data collection runs by scraping user reviews from the Google Play Store, resulting in a dataset of 5,999 reviews. The data experience several pre-processing stages, including cleaning, normalization, tokenization, stop word removal, and stemming. After pre-processing, it applies TF-IDF weighting, followed by classification using the SVM algorithm. To address class imbalance, the data balancing application uses a Random Over Sampler. The classification results indicate that the model successfully categorized the reviews into three groups: positive (56.3%), negative (26.9%), and neutral (16.8%). Following data balancing, the model obtains an accuracy of 91%, a significant improvement from the previous accuracy of 79.71%. Additionally, the model's implementation based on Streamlit facilitates efficient and automatic sentiment analysis of the DANA application.

Keywords: Support Vector Machine, sentiment analysis, user reviews, DANA application, Google Play Store

ABSTRAK

ANDRES FERNANDA RAHIM. T3119002. PENGGUNAAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI DANA PADA GOOGLE PLAY STORE.

Penelitian ini bertujuan untuk mengembangkan sistem klasifikasi sentimen pengguna aplikasi DANA di Google Play Store menggunakan metode *Support Vector Machine* (SVM), dan 2) menerapkan metode SVM guna meningkatkan akurasi dan efektivitas klasifikasi. Penelitian ini menggunakan pendekatan kuantitatif dengan penyajian secara deskriptif. Teknik pengumpulan data dilakukan dengan scraping ulasan pengguna dari Google Play Store. Data yang diperoleh sebanyak 5.999 ulasan kemudian dianalisis melalui tahapan *preprocessing* (*cleaning*, normalisasi, tokenisasi, *stopword removal*, dan *stemming*), pembobotan TF-IDF, serta klasifikasi menggunakan algoritma SVM. Untuk mengatasi ketidakseimbangan kelas, dilakukan *balancing* data menggunakan *Random Over Sampler*. Hasil klasifikasi menunjukkan bahwa model berhasil mengelompokkan ulasan ke dalam tiga kategori: positif (56,3%), negatif (26,9%), dan netral (16,8%). Evaluasi performa model setelah *balancing* data menunjukkan akurasi sebesar 91% yang sebelumnya hanya mencapai 79,71%. Implementasi model dengan berbasis *streamlit* juga dapat membantu proses analisis sentimen secara otomatis dan efisien terhadap aplikasi DANA.

Kata kunci: *Support Vector Machine*, analisis sentimen, ulasan pengguna, aplikasi Dana, *Google Play Store*

KATA PENGANTAR

Alhamdulillah, penulis dapat menyelesaikan usulan penelitian dengan judul “Penggunaan Metode *Support Vector Machine* Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Dana Pada *Google Play Store*.”, untuk memenuhi salah satu syarat Ujian Akhir Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.

Penulis Menyadari sepenuhnya skripsi ini tidak mungkin terwujud tanpa bantuan dan dorongan dari berbagai pihak, baik bantuan moril maupun materil. Untuk itu, dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan banyak terima kasih dan penghargaan yang setinggi-tingginya kepada:

1. Ibu Dr. Hj. Juriko Abdussamad, M. Si, selaku ketua yayasan pengembangan ilmu pengetahuan dan teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar Latjokke, M. Si, Selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan Abraham Salihi, S. Kom. M.Kom, Selaku dekan fakultas ilmu komputer universitas ichsan gorontalo;
4. Bapak Sudirman Melangi, M.Kom, Selaku pembantu Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
5. Ibu Irma Surya Kumala M.Kom, Selaku Pembantu Dekan II Bidang Administrasi Umum dan keuangan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo;
6. Bapak Yasin Aril Mustofa, M. Kom, Selaku Pembimbing I; yang telah memberikan arahan, bimbingan, dan motivasi selama penyusunan skripsi ini.
7. Bapak Abd Rahmat Karim Haba, M. Kom, Selaku Pembimbing II; yang dengan sabar membimbing dan memberikan masukan yang konstruktif hingga skripsi ini dapat diselesaikan.
8. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah mendidik dan mengajarkan berbagai disiplin ilmu kepada penulis;

9. Kedua Orang tua saya yang tercinta, atas segala kasih sayang, jerih payah dan doa restunya dalam membesarkan dan mendidik penulis;
10. Rekan-rekan seperjuangan yang telah banyak memberikan bantuan dan dukungan moril yang sangat besar pada penulis;
11. Kepada semua pihak yang ikut membantu dalam penyelesaian skripsi ini yang tak sempat penulis sebutkan satu-persatu

Semoga Allah, SWT melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis menyadari sepenuhnya bahwa apa yang telah dicapai ini masih jauh dari kesempurnaan dan masih banyak terdapat kekurangan. Oleh karena itu, penulis sangat mengharapkan adanya kritik dan saran yang konstruktif. Akhirnya penulis berharap semoga hasil yang telah dicapai ini bermanfaat bagi kita semua, Aamiin.

Gorontalo, 10 Agustus 2024

Penulis

DAFTAR ISI

PERSETUJUAN SKRIPSI.....	ii
PENGESAHAN SKRIPSI.....	iii
PERNYATAAN SKRIPSI.....	iv
ABSTRACT	v
ABSTRAK.....	vi
KATA PENGANTAR.....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xii
DAFTAR TABEL	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	5
1.3 Rumusan Masalah.....	5
1.4 Tujuan Penelitian.....	5
1.5 Manfaat Penelitian	5
BAB II LANDASAN TEORI.....	6
2.1 Tinjauan Studi	6
2.2 Tinjauan Pustaka	7
2.2.2 Pengertian analisis Sentimen	7
2.2.3 SVM (Support Vector Machine).....	7
2.2.4 Dompot digital.....	7
2.2.5 Dana.....	8
2.2.6 Manfaat Dana.....	8
2.2.7 Klasifikasi	8
2.2.8 Model Klasifikasi	9
2.2.9 Data Mining	9
2.2.10 Algoritma Support Vector Machine	10
2.2.11 penerapan SVM Dalam Kasus Sebelumnya	11
2.3 Kerangka Pikir.....	19

BAB III METODE PENELITIAN	20
3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian	20
3.2 Studi Literatur	20
3.3 Pengumpulan Data.....	21
3.4 Tahap Praproses Data	21
3.5 Uji Coba dan Proses Klasifikasi.....	22
3.6 Tahap pelatihan dengan metode SVM (Support Vector Machine)	22
3.7 Evaluasi performa klasifikasi	22
3.8 Pembahasan hasil percobaan.....	23
3.9 Evaluasi.....	23
BAB IV HASIL PENELITIAN	24
4.1 Pengumpulan Data.....	24
4.2 Hasil Pemodelan	28
4.2.1 Hasil Preprocessing	29
4.2.2 Hasil Pembobotan TF-IDF.....	45
4.2.3 Klasifikasi <i>Support Vector Machine</i> (SVM).....	60
4.2.4 Penanganan <i>Imbalanced Data</i> dengan <i>Random Over Sampler</i>	62
4.3 Uji Model <i>Confusion Matriks</i>	63
4.4 Visualisasi Hasil Penelitian.....	65
4.4.1 Distribusi Prediksi Model	65
4.4.2 WordCloud.....	66
BAB V PEMBAHASAN	67
5.1 Pembahasan Model.....	67
5.1.1 Preprocessing.....	67
5.1.2 Pembobotan Kata.....	71
5.1.3 Klasifikasi	73
5.1.4 Evaluasi.....	74
5.2 Implementasi Model	76
BAB VI PENUTUP	78
6.1 Kesimpulan.....	78
6.2 Saran	78

DAFTAR PUSTAKA.....	79
LAMPIRAN	81

DAFTAR GAMBAR

Gambar 2. 1 Model Klasifikasi	9
Gambar 2. 2 Preprocessing.....	12
Gambar 2. 3 Klasifikasi SVM (support vector Machine)	16
Gambar 2. 4 Grafik presentase	17
Gambar 2. 5 Hasil Akurasi, Recall, Precession, dan F-1 Score	18
Gambar 3. 1 Tahapan Preprocessing	22
Gambar 4. 1 Hasil Pengumpulan data.....	27
Gambar 4. 2 Pemodelan SVM.....	28
Gambar 4. 3 Tahapan Preprocessing	29
Gambar 4. 4 Source Code Klasifikasi SVC	61
Gambar 4. 5 hasil klasifikasi sebelum balancing data	62
Gambar 4. 6 Source code Random Over Sampler	63
Gambar 4. 7 Confusion Matriks	64
Gambar 4. 8 Hasil Klasifikasi setelah balancing data.....	65
Gambar 4. 9 Distribusi Prediksi label	65
Gambar 4. 10 Wordcloud	66
Gambar 5. 1 Tahap Cleantext.....	67
Gambar 5. 2 Case Folding.....	68
Gambar 5. 3 Normalisasi Kata	68
Gambar 5. 4 Tokenization	69
Gambar 5. 5 Stopword removal.....	69
Gambar 5. 6 Stemming	70
Gambar 5. 7 Pembobotan TF	71
Gambar 5. 8 Hasil TF-IDF	72
Gambar 5. 9 Sample Prediksi	73
Gambar 5. 10 Confusion Matriks (sebelum balancing)	74
Gambar 5. 11 Laporan Klasifikasi (sebelum balancing).....	74
Gambar 5. 12 Confusion Matrix (sesudah balancing)	75

Gambar 5. 13 Laporan Klasifikasi (sesudah balancing)	75
Gambar 5. 14 Hasil Implementasi Model 1	76
Gambar 5. 15 Hasil Implementasi Model 2	77
Gambar 5. 16 Hasil Implementasi Model 3	77

DAFTAR TABEL

Tabel 1. 1 Contoh Data Ulasan Aplikasi Dana	3
Tabel 2. 1Tabel Tinjauan Studi	6
Tabel 2. 2 Crawling Data	11
Tabel 2. 3Proses Casefolding dan Cleansing	13
Tabel 2. 4 Proses Tokenisasi	13
Tabel 2. 5Proses Normalisasi	13
Tabel 2. 6Proses Stemming	13
Tabel 2. 7Stopwords Removal	13
Tabel 2. 8Menghitung TF dan IDF	15
Tabel 2. 9Pengujian dan evaluasi	17
Tabel 2. 10Hasil Confussion Matrix	18
Tabel 4. 1 Hasil Pengumpulan Data	24
Tabel 4. 2 Hasil Text Cleaning	30
Tabel 4. 3 Hasil Case folding	32
Tabel 4. 4 Hasil Normalisasi Kata	35
Tabel 4. 5 Hasil Tokenizing	38
Tabel 4. 6 Hasil Stopword Removal	41
Tabel 4. 7 Hasil Stemming	44
Tabel 4. 8 Sample Term dan Perhitungan TF	46
Tabel 4. 9 Sampel Perhitungan TF Normalisasi	51
Tabel 4. 10 Sampel Perhitungan DF dan IDF	54
Tabel 4. 11 Sampel Perhtungan TF-IDF	57
Tabel 4. 12 Akurasi SVM berdasarkan C value	61

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, masyarakat perlu bijak dalam memanfaatkan kemudahan dan efektivitas berinteraksi dengan sesama. Banyaknya inovasi digital di berbagai sektor menunjukkan partisipasi aktif masyarakat dalam mengikuti perkembangan zaman yang semakin modern. Pertumbuhan bisnis digital juga berperan dalam mendorong kemunculan perusahaan startup yang bergerak di sektor keuangan digital. Salah satu produk digital yang populer adalah uang elektronik, atau lebih dikenal dengan sebutan E-money[1]. Kehadiran *E-money* ini memungkinkan masyarakat untuk melakukan transaksi keuangan tanpa perlu menggunakan uang tunai.

Dana merupakan sebuah aplikasi dompet digital yang berasal dari Indonesia dan dapat diunduh melalui Google Play Store. Aplikasi ini dirancang untuk memfasilitasi berbagai jenis pembayaran secara digital, baik secara online maupun offline, tanpa memerlukan uang tunai (*cashless*) atau kartu fisik (*cardless*). Dengan mengintegrasikan teknologi pembayaran digital, Dana memungkinkan pengguna untuk melakukan transaksi dengan lebih mudah dan praktis, mengurangi ketergantungan pada uang tunai maupun kartu kredit atau debit.[1]

Perkembangan teknologi informasi dan komputer telah mengubah berbagai aspek kehidupan masyarakat Indonesia, termasuk dalam pengelolaan keuangan. Dompet digital, seperti DANA, telah menjadi bagian penting dalam transaksi sehari-hari karena kenyamanannya melalui aplikasi mobile. Sejak diluncurkan pada November 2018, DANA cepat menjadi populer dan menarik 20 juta pengguna hanya dalam beberapa bulan, dengan rata-rata 1,5 juta transaksi harian. Pada tahun 2023, pengguna DANA di Indonesia tercatat mencapai 170 juta orang.[2] Penelitian yang akan dilakukan adalah mengklasifikasi sentimen masyarakat terhadap aplikasi DANA berdasarkan ulasan pengguna yang terdapat pada google play, ulasan ini perlu diklasifikasi menggunakan machine learning karena jumlahnya banyak, jika diklasifikasikan secara manual maka itu akan sulit dan memakan waktu yang lama. Oleh karena itu peneliti menggunakan machine

learning untuk penelitian ini dan metode yang dipakai adalah Ensemble Learning untuk penelitian ini dan Metode yang di pakai adalah Metode Ensemble Learning karena kombinasi dari beberapa model biasanya lebih kuat daripada model individu. Ini disebabkan oleh fakta bahwa kesalahan yang dibuat oleh model-model individu cenderung berbeda, sehingga kombinasi dari hasil model-model ini dapat mengurangi total kesalahan prediksi [3].

Sebagaimana dalam penelitian terdahulu oleh D. Sondakh, dkk, pada tahun 2023 dengan berjudul Sistem Analisis Sentimen Ulasan Aplikasi Belanja Online Menggunakan Metode Ensemble Learning Penelitian ini mengembangkan sistem berbasis web untuk menganalisis sentimen komentar dan ulasan pengalaman pelanggan aplikasi belanja online menggunakan teknik ensemble learning. Model prediksi yang dihasilkan menggunakan voting mayoritas dari tiga algoritma: k-NN, SVM, dan Random Forest, dengan indikator akurasi 81.8%, precision 83%, recall 82%, dan F1-score 82%. Namun, ketiga algoritma tersebut menghasilkan file model classifier yang cukup besar. Untuk memaksimalkan akurasi hasil prediksi, perlu dipertimbangkan beberapa teknik yang menghasilkan file model klasifikasi yang lebih kecil dan melakukan tuning untuk menentukan parameter yang optimal untuk setiap model.[3]

Ada apa dengan dana dan kenapa harus aplikasi dana? Karena terdapat beberapa masalah aplikasi DANA yaitu merchant yang masih terbatas. Bila dibandingkan dengan kompetitor lainnya, seperti OVO, Go-pay dan LinkAja. Merchant yang telah bekerja sama dengan DANA jauh lebih sedikit dan juga belum menjangkau pedagang kecil, seperti pedagang kaki lima. Pasalnya aplikasi OVO dan Go-pay telah menjangkau ke sektor-sektor yang lebih kecil dalam sejauh ini.[4] Aplikasi Dana diluncurkan pada tahun 2018 memungkinkan pengguna melakukan transaksi keuangan nontunai mulai dari pembayaran kebutuhan sehari-hari hingga top up saldo investasi. Berdasarkan survei YouGov, Dana adalah dompet digital tercepat dalam peningkatan pengguna dengan pertumbuhan tercepat sekitar 4% di masa COVID-19 (Liputan 6, 2021). Aplikasi Dana memiliki banyak pengguna, sehingga sering kali terdapat ulasan positif, negatif dan netral yang tidak relevan dengan rating yang diberikan di Google Play Store.[1]

Beberapa masalah umum dalam penggunaan aplikasi digital meliputi transaksi yang gagal, proses verifikasi akun yang lama, serta kendala saat top-up saldo. Selain itu, keamanan dan risiko penipuan menjadi perhatian penting bagi pengguna. Masalah pada cashback dan promosi, aplikasi yang sering crash, layanan pelanggan yang tidak responsif, dan sinkronisasi data yang tidak lancar juga sering dikeluhkan. [5] berikut ini adalah contoh ulasan di aplikasi DANA.

Tabel 1. 1 Contoh Data Ulasan Aplikasi Dana

Ulasan/Komentar	Keterangan
"Transaksi saya gagal padahal saldo sudah mencukupi. Uang sudah terpotong tapi belum masuk ke penerima. Layanan pelanggan juga lambat merespons, dan tidak ada solusi yang jelas. Ini sangat mengecewakan!"	Negatif
"Transaksi sempat gagal, tapi saya senang karena pihak DANA cepat menangani masalahnya. Uang langsung dikembalikan ke saldo dalam waktu singkat. Semoga ke depan semakin lancar!"	Positif
"Saya mengalami transaksi gagal saat melakukan pembayaran, tapi uangnya dikembalikan setelah beberapa jam. Meski tidak terlalu lama, akan lebih baik kalau tidak ada masalah seperti ini dari awal."	Netral
"Dih Najis kalau mau bayar kasih notif apa kek ini main bayar sendiri hilang duit gua goblok tiba" aja berlangganan apple music goblok bangek dah."	Negatif

"Sangat membantu cukup praktis dan banyak promo"	Positif
"Verifikasi akun memakan waktu lebih lama dari yang diharapkan, sekitar 3 hari. Agak merepotkan, tapi setelah selesai semua berjalan dengan lancar. Semoga prosesnya bisa lebih cepat di masa mendatang."	Netral
"Ga ngerti lagi banyak errornya masa di buka ga bias?"	Negatif
"Semua Proses sangat smooth untuk pengguna baru."	Positif
"Top-up saldo di DANA umumnya berjalan lancar, tapi ada kalanya saldo tidak langsung masuk. Biasanya hanya menunggu beberapa menit, tapi pernah juga sampai lebih dari setengah jam. Tidak terlalu buruk, tapi berharap bisa lebih konsisten ke depannya."	Netral

Berdasarkan uraian diatas, maka dalam penelitian ini akan menerapkan metode Ensemble Learning untuk mengkalsifikasi sentimen sentimen pengguna terhadap aplikasi DANA dompet digital berdasarkan ulasan yang ada pada googleplay, analisis sentimen dilakukan untuk mengetahui besarnya presentase tanggapan pengguna terhadap layanan aplikasi DANA dengan mengelompokan ulasan pengguna dalam kategori positif, negatif, dan netral.

Bersadarkan berbagai pemaparan di atas, maka peneliti tertarik untuk melakukan penelitian dengan judul: **"Penggunaan Teknik Ensemble Learning Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Dana Pada Google Play Store.**

1.2 Identifikasi Masalah

1. Peningkatan ulasan pengguna terhadap aplikasi DANA di Google Play sulit diolah secara manual untuk menilai kualitas layanan.
2. Beberapa ulasan kurang spesifik atau ambigu, sehingga sulit untuk menentukan apakah ulasan itu positif, negative dan netral.

1.3 Rumusan Masalah

1. Bagaimana cara otomatis menilai kualitas layanan aplikasi DANA dari ulasan pengguna di Google Play?
2. Bagaimana mengatasi ulasan yang tidak jelas atau membingungkan untuk menentukan apakah ulasannya positif, negatif, atau netral?

1.4 Tujuan Penelitian

1. Mengembangkan sistem klasifikasi sentimen yang efisien dan akurat untuk ulasan pengguna aplikasi DANA di Google Play menggunakan metode SVM (Support Vector Machine) .
2. Menerapkan metode SVM (Support Vector Machine). untuk meningkatkan akurasi dan efektivitas klasifikasi sentimen dibandingkan dengan model individu.

1.5 Manfaat Penelitian

1. Dapat membantu mahasiswa sebagai bahan referensi dalam melaksanakan penelitian yang sama.
2. Dapat menambah wawasan mahasiswa dalam memanfaatkan teknologi untuk melakukan inovasi.

BAB II

LANDASAN TEORI

2.1 Tinjauan Studi

Tugas akhir ini memiliki konteks dasar metode klasifikasi sentimen Dana Dan Zoom. Dalam pengerjaan tugas akhir ini, akan digunakan beberapa penelitian sebelumnya sebagai pedoman dan referensi pengerjaan. Pada tabel di bawah ini akan dijelaskan deskripsi, metode, data yang digunakan dan hasilnya.

Tabel 2. 1Tabel Tinjauan Studi

NO	PENELITI	JUDUL	TAHUN	METODE	HASIL
1	Fanka Angelina Larasati	Analisis Sentimen Ulasan Aplikasi Dana	2022	Random Forest	accuracy 84%.
2.	Nuraeni Herlinawati	Analisis sentimen pengguna aplikasi dana berdasarkan ulasan pada google play	2020	Support vector machine	accuracy 81,22%
3.	Debby E. Sondak	Sistem Analisis Sentimen Ulasan Aplikasi	2023	Ensemble Learning	81.8% precision 83%, recall 82%, F1-score 82%.

		Belanja Online			
--	--	-------------------	--	--	--

2.2 Tinjauan Pustaka

2.2.2 Pengertian analisis Sentimen

Analisis sentimen adalah bidang penelitian yang berkembang pesat dalam pemrosesan bahasa lami (NLP) dan klasifikasi teks. Teknik ini telah menjadi bagian penting dari berbagai aplikasi termasuk politik, bisnis, periklanan dan pemasaran[5]

Berdasarkan hasil analisis sentimen tweet dengan metode *text mining* didapatkan sentimen positif lebih dominan yaitu sebanyak 22.246 (51,2%) tweet, diikuti oleh sentimen negatif sebanyak 12.015 (27,7%) tweet dan sentimen netral dengan jumlah 9.174 (21,1%) dari total 43.435 tweet. Nilai akurasi dengan algoritma *Naïve Bayes Classifier* didapatkan hasil sebesar 94% dan berada dalam kategori “*Excellent Classification*” yang artinya algoritma *Naive Bayes Classifier* berhasil memprediksi kategori sentimen dalam penelitian ini dengan baik.[6]

2.2.3 SVM (Support Vector Machine)

Support Vector Machine (SVM) adalah Teknik pembelajaran dengan banyak kualitas yang diinginkan dan menjadikan algoritma SVM sangat populer. 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[8]

2.2.4 Dompot digital

Dompot digital adalah layanan yang bersifat elektronik dan berfungsi untuk menyimpan data serta instrumen dari pembayaran. Dengan dompot digital pengguna dapat menyimpan uangnya untuk melakukan transaksi keuangan.[9]

Pada dasarnya, dompot digital berfungsi layaknya dompot fisik yang dapat digunakan untuk melakukan pembayaran yang aman. Jika menggunakan dompot digital, seseorang dapat melakukan pembayaran secara cashless ketika hendak membayar makanan, tiket pesawat, tiket bioskop, hingga belanja online.[9] berikut ini salah satu dompot digital.

2.2.5 Dana

DANA merupakan dompet digital serta tempat pembayaran digital secara terbuka yang umum digunakan oleh aplikasi-aplikasi yang sudah terdaftar, maupun gerai-gerai online yang konvensional. Platform dompet digital ini mengembangkan infrastruktur pembayaran digital dengan menyelidiki praktik terbaik yang sudah digunakan dan disediakan oleh jutaan pengguna dan sudah terbukti kelegalan dan keamanannya secara umum (Gideon, 2018). Dengan menggunakan aplikasi DANA, pengguna dapat melakukan bermacam transaksi *cashless* pada gerai – gerai online seperti membeli pulsa, membayar tagihan (internet, listrik, air, hingga BPJS), membayar cicilan, dan seterusnya yang berbasis online untuk mengetahui tingkat kepuasan konsumen secara keseluruhan dengan tingkat kepentingan dari layanan. [10]

2.2.6 Manfaat Dana

Aplikasi DANA menawarkan kemudahan yang luas dalam berbagai aspek keuangan pengguna. Dengan DANA, Anda dapat membayar berbagai tagihan bulanan seperti listrik, air, telepon, TV kabel, BPJS, dan internet secara online dengan mudah. Transfer uang antar pengguna DANA atau ke rekening bank lain juga dapat dilakukan dengan cepat. DANA memfasilitasi pembayaran untuk belanja online di berbagai e-commerce dan di toko fisik yang menerima metode pembayaran ini, cukup dengan memindai kode QR atau menggunakan fitur pembayaran digital lainnya. Pengguna dapat membeli pulsa dan paket data langsung dari aplikasi tanpa perlu mencari kios. Menambah saldo e-wallet DANA juga praktis, dengan opsi seperti transfer bank, kartu kredit, atau melalui gerai mitra. Beberapa versi DANA bahkan menawarkan fitur investasi seperti pembelian reksa dana dan SBN. Keamanan transaksi menjadi prioritas dengan penggunaan teknologi enkripsi canggih dan fitur perlindungan untuk menjaga data dan informasi finansial pengguna [11].

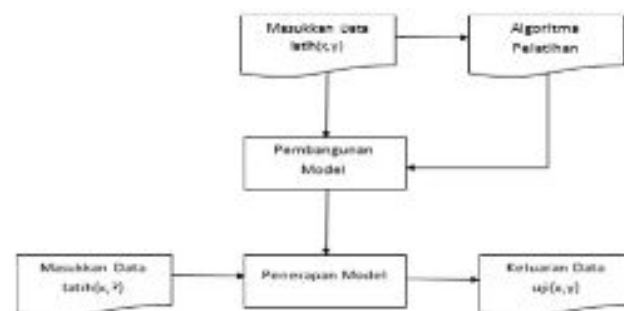
2.2.7 Klasifikasi

Klasifikasi adalah proses memprediksi kelas atau kategori dari suatu entitas berdasarkan fitur atau atributnya dan merupakan bagian dari supervised learning, yaitu model machine learning yang diajarkan menggunakan dataset berlabel. Dalam

pelatihan, dataset diberikan label kelasnya, sehingga model dapat belajar dari data tersebut. Tujuan utama dari klasifikasi adalah membangun model yang mampu mengidentifikasi kelas dari data baru yang belum pernah dilihat sebelumnya, dengan mengacu pada data yang telah digunakan untuk melatih model tersebut.[12]

2.2.8 Model Klasifikasi

Model dalam klasifikasi mempunyai arti yang sama dengan kotak hitam, dimana ada suatu model yang menerima masukan kemudian mampu melakukan pemikiran terhadap masukan tersebut dan memberikan jawaban atau sebuah hasil. Kerangka kerja klasifikasi bisa kita lihat pada gambar 2.1 di bawah ini



Gambar 2.1 Model Klasifikasi

Disediakan sejumlah data latih(x, y) untuk digunakan sebagai data pembangunan model. Model tersebut kemudian dipakai untuk memprediksi kelas dari data uji (x, y) sehingga diketahui kelas y yang sesungguhnya [14]. Model yang sudah dibangun pada saat pelatihan kemudian dapat digunakan memprediksi label kelas baru yang belum diketahui. Dalam pembangunan model selama proses pelatihan tersebut diperlukan suatu algoritma untuk membangunnya, yang disebut algoritma pelatihan. Dalam penelitian ini akan menggunakan algoritma SVM (Support Vector Machine)[13]

2.2.9 Data Mining

Data Mining merupakan proses menggali pengetahuan dari sebuah dokumen dari waktu tertentu menggunakan beberapa teknik Text Mining, seperti information extraction, dan natural language processing. Text Mining bertujuan untuk mencari kata-kata yang dapat menjadi kata kunci didalam suatu dokumen sehingga dapat dilakukan analisis hubungan antar dokumen[14]. Proses dalam Text

Mining yaitu: Text, tahap awal yaitu data akan ditampung dalam sebuah database yang besar.

1. Text Preprocessing, untuk melakukan pembersihan teks sehingga menjadi text yang bersih dari beberapa karakter yang tidak diinginkan, tahapan ini seperti text clean up, Case Folding, Tokenization, dan part of speech
2. Text Transformation, merubah teks dan membentuk teks menjadi sebuah atribut yang mengacu pada proses yang digunakan biasanya dilakukan merubah menjadi matriks.
3. Feature Selection, melakukan pengurangan kata yang dianggap tidak relevan. Tahapan yang digunakan adalah Stop Word Removal dan Stemming.

Proses penemuan pengetahuan ditunjukkan dalam gambar di bawah ini sebagai urutan berulang dari langkah-langkah berikut:

2.2.10 Algoritma Support Vector Machine

Support Vector Machine adalah metode klasifikasi jenis (Supervised) karena ketika proses pelatihan memerlukan goal atau target pembelajaran tertentu. SVM merupakan algoritma yang bekerja menggunakan pemetaan nonlinier untuk mengubah data pelatihan menjadi dimensi yang tinggi. Pada dimensi baru akan mencari hyperplane untuk memisahkan secara linier dan dengan pemetaan nonlinier yang tepat kedimensi yang lebih tinggi lagi. SVM muncul pertama kali pada tahun 1992 oleh Vladimir Vapnik Bersama dengan Bernhard Boser dan Isabell Guyon. Pada dasarnya SVM difungsikan untuk mengklasifikasikan data dengan 2 kelas dan linier separable. Namun masalah yang muncul kebanyakan non-linear. untuk mengatasi persoalan tersebut maka digunakan cara kernel trick pada data training yang bertujuan untuk mengelompokkan setiap data pada ruang input menjadi ruang vektor baru dengan dimensi yang lebih tinggi. SVM hanya menggunakan beberapa titik data terpilih yang berkontribusi untuk membentuk model yang akan dipergunakan dalam proses klasifikasi[14]. Sebuah hyperplane dapat ditulis sebagai persamaan SVM, ilustrasi metode SVM ditunjukkan pada persamaan, Adapun fungsi kernel yang biasanya digunakan dalam SVM yaitu :

1. Kernal Lianer : $X^T X$
2. Kernal polynominal : $(X^T X_i + 1)^p$
3. Kernal RBF : $\exp\left(-\frac{1}{2\sigma^2} |x - x_i|^2\right)$
4. Kernal Sigmoid : $\tanh\left(\sigma(x \cdot y) + x\right)$

2.2.11 penerapan SVM Dalam Kasus Sebelumnya

Berikut contoh penerapan algoritma SVM pada penelitian oleh Darwis,pratiwidan pasaribu yang berjudul PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA Penelitian ini bertujuan untuk memberikan rekomendasi tentang opini masyarakat tentang kinerja KPK berdasarkan data yang diambil dari twitter kemudian diklasifikasikan menggunakan algoritma SVM dengan label klasifikasi yaitu positif, negatif dan netral.berikut data yang di peroleh dari hasil penelitian.[15]

1. Crawling Data

Proses crawling menghasilkan 2000 data komentar tweet twitter KPK RI yang diambil dengan rentang waktu tahun 2018-2019 dengan memanfaatkan fasilitas twitterscraper dengan menginstal pip dan tools python 3.7.6 cara ini dapat memudahkan proses crawling, karena dalam proses crawling tidak ada limitasi tweet dan hasilnya dapat langsung disimpan dalam bentuk file .json atau .csv lalu disimpan dalam bentu file excel. Contoh dokumen hasil crawling dapat dilihat pada Tabel Berikut.

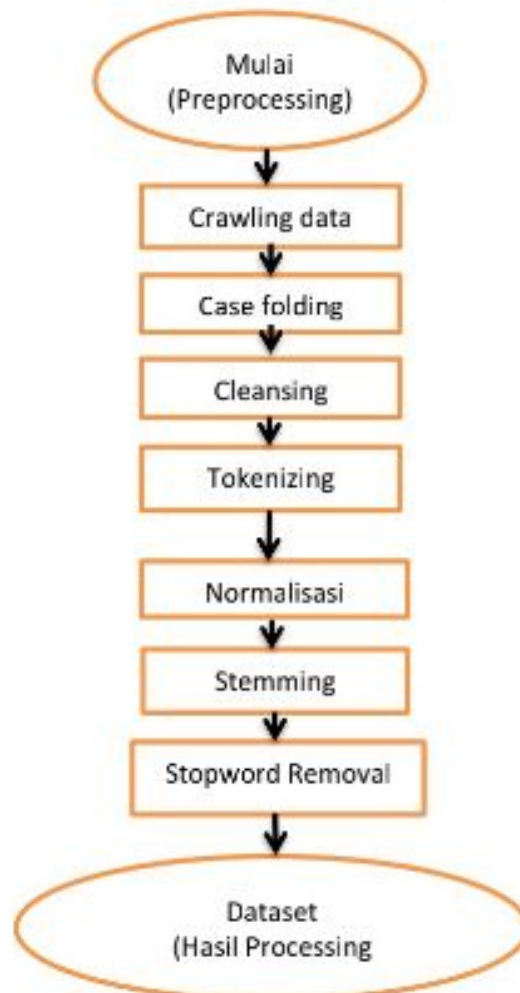
Tabel 2. 2 Crawling Data

Dokumen	Tweet
Dokumen 1	Lah haru tau.. kalo korupsi 2triliun trs lapor lumayan dong 2M balik lagi. Gitu ?

2. Preprocessing

Preprocessing merupakan tahap awal pada proses pengolahan data teks hasil ekstraksi sebelum diolah lebih lanjut hingga sampai ke proses klasifikasi. Tahapan preprocessing disajikan pada Gambar 2.2

Gambar 2. 2Preprocessing



Gambar 2 merupakan tahapan preprocessing yang terdiri dari :

1. Case folding dan Cleansing

Proses ini merupakan proses Poses menyeragamkan huruf- huruf kapital menjadi huruf kecil. dan pembersihan dokumen dari kata- kata yang tidak dibutuhkan untuk mengurangi noise. Kata yang dihilangkan merupakan suatu karakter HTML, kata kunci, emoticon, hastag (#), mention username (@username), angka dan url. Dapat dilihat pada contoh yang di ambil dari dokumen pertama pada Tabel 2.

Tabel 2. 3Proses Casefolding dan Cleansing

lah baru tau kalo korupsi t trs lapor lumayan dong m balik lagi gitu
--

2. ProsesTokenisasi

Merupakan proses menyeleksi, memecah dan pemotongan kata pada dokumen menjadi term-term berdasarkan spasi. Contoh hasil tokenisasi dapat dilihat pada Tabel

Tabel 2. 4TaProses Tokenisasi

['lah','baru','tau','kalo','korupsi','t','trs','lapor','lumayan','dong','m','balik','lagi','gitu']
--

3. Proses Normalisasi

Tahap ini mengubah kalimat tidak baku/slang menjadi kalimat baku yang sesuai dengan KBBI seperti contoh yang disajikan pada Tabel 4.

Tabel 2. 5Proses Normalisasi

['lah','baru','tahu','kalau','korupsi','triliun','terus','lapor','lumayan','dong','miliar','balik','lagi','begitu']

4. Proses Stemming

Merupakan proses untuk menghilangkan awalan atau akhiran kata yang terdapat kata sambung, kata depan, kata ganti, menjadi kata dasar yang sesuai dengan KBBI. Hasil dari proses stemming disajikan pada Tabel 5.

Tabel 2. 6Proses Stemming

['lah','baru','tahu','kalau','korupsi','triliun','terus','lapor','lumayan','dong','miliar','balik','lagi','begitu']

5. Proses *Stopwords Removal*

Merupakan proses penyaringan kata pada dokumen yang tidak ada hubungannya dalam analisis sentimen seperti yang disajikan pada Table 6.

Tabel 2. 7Stopwords Removal

['korupsi','triliun','lapor','lumayan','miliar']
--

3. Pembobotan

Pada tahapan ini kata tersebut direpresentasikan kedalam bentuk vektor dan TF-IDF. Penggunaan metode TF-IDF dalam proses pembobotan dapat menghasilkan sebuah vektor dengan banyak term sehingga dapat dikenali tiap kata yang dihitung sebagai satu fitur.

1. Term Frequency (TF)

Proses ini akan menghitung jumlah kemunculan kata pada dataset. Untuk menentukan bobot dari masing-masing term/kata dalam sebuah dokumen yang ada pada dataset menggunakan augmented TF berdasarkan persamaan.

$$tf = 0,5 + 0,5 \frac{t f}{\max(t f)}$$

Keterangan:

tf = banyaknya kata yang muncul pada sebuah dokumen
 $\max(t f)$ = panjang kata dari sebuah dokumen itu sendiri.

Sebagai contoh, misalkan Term Frequency (TF) diambil dari dokumen ke-1 sebagai berikut :

['korupsi', 'triliun', 'lapor', 'lumayan', 'miliar'].

Jika dokumen tersebut dimasukkan ke dalam persamaan maka:

Korupsi = $0,5 + 0,5 \left(\frac{1}{5}\right) = 0,2$ Karena setiap kata hanya muncul 1 kali pada dokumen maka hasilnya akan sama, kemudian bobot TF pada dokumen diurutkan sesuai dengan abjad sehingga hasilnya sebagai berikut: ['korupsi' = 0,2, 'lapor' = 0,2, 'lumayan' = 0,2, 'miliar' = 0,2, 'triliun' = 0,2]

2. Inverse Document Frequency (IDF)

Proses ini merupakan jumlah dokumen yang berisikan term yang dicari dalam dokumen dataset menggunakan persamaan.

$$idf = \ln \frac{N}{df} + 1$$

$$\text{miliar} = \ln \left(\frac{1890}{25}\right) + 1$$

$$= 5.3254562831854875$$

Kemudian bobot IDF pada dokumen di urutkan sesuai dengan abjad.

Sehingga hasilnya sebagai berikut:

['korupsi'=3.3132234911991016, 'lapor'=4.369944838158051, 'lumayan'
=6.752572638825633, 'miliar'=5.3254562831854875, 'triliun'=5.01797158
3437527]

3. TF IDF

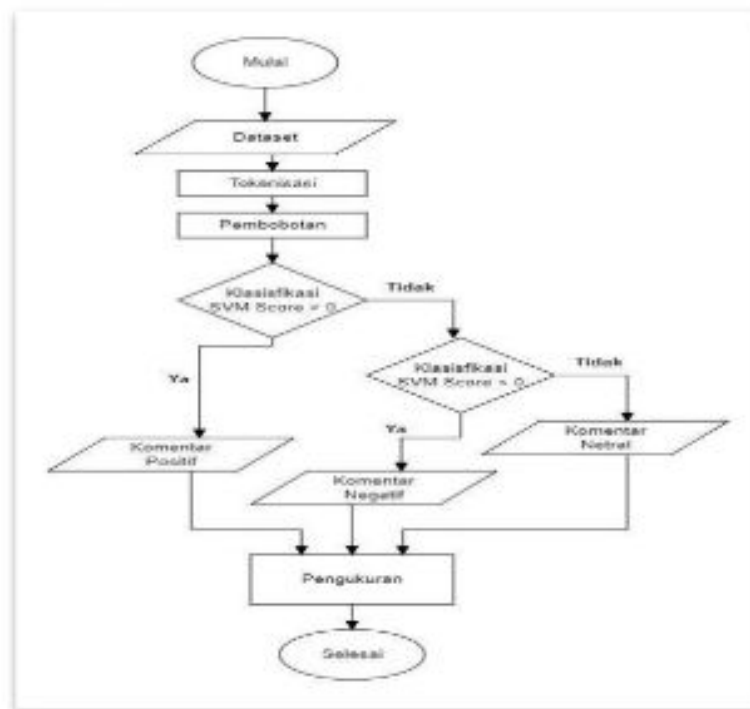
Cara menghitungnya adalah dengan mengalikan TF dan IDF dengan hasil yang disajikan pada Tabel 8.

Tabel 2. 8Menghitung TF dan IDF

Dokumen	TF*IDF	Label
['korupsi', 'triliun', 'lapor', 'lumayan', 'miliar']	[0.6626446982398204, 0.8739889676316102, 1.3505145277651267, 1.0650912566370976, 1.0035943166875054]	Negatif
['mahfud', 'lapor', 'lapor', 'harap', 'peduli', 'berantas', 'kosupsi', 'ikhlas', 'lapor', 'koruptor', 'ikhlas', 'peduli', 'nkri']	[0.4127906367465956, 0.4393168279998056, 1.1454953568285506, 0.25486334547685396, 0.34625237232454914, 1.008448808805704, 0.5727476784142753, 0.51942866 45250488, 1.2078746042298067]	Negatif
['skenario']	[8.544332108053688]	Netral

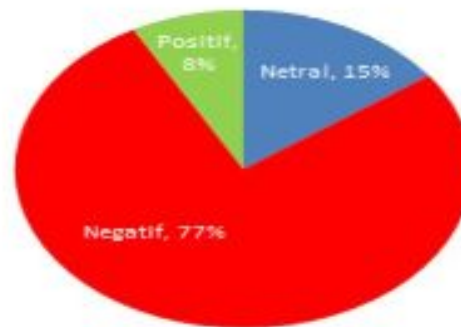
4. Klasifikasi SVM

Penelitian ini menggunakan scikit-learn machine learning pada python versi 3.7.6 dengan Support Vector machine. Alur dari proses klasifikasi disajikan pada Gambar 3.



Gambar 2. 3 Klasifikasi SVM (support vector Machine)

Gambar 2.3 menjelaskan pengolahan data yang telah melalui proses pembobotan dan kemudian akan diklasifikasikan menjadi tiga kelas, yaitu kelas positif, kelas negatif, dan kelas netral. Dengan ketentuan apabila bobot/ score kata >0 adalah positif, <0 adalah negatif dan selain itu adalah netral. Dari hasil pengklasifikasian mesin dengan algoritma SVM didapatkan presentase kelas yaitu kelas positif, kelas negatif, dan kelas netral yang disajikan pada Gambar2. 4.



Gambar 2. 4 Grafik presentase

Gambar 2.4 menunjukkan hasil presentase dari klasifikasi, dapat dijelaskan bahwa hasil sentimen lebih condong kepada sentimen negatif, yaitu sebesar 77%, untuk netral sebesar 15% dan positif hanya 8%. Data hasil klasifikasi ini diperoleh dengan membagi data sebagai informasi pelatihan serta pengujian, dengan ketentuan jumlah data pelatihan sebanyak 80% dan data pengujian sebanyak 20%.

5. Pengujian dan Evaluasi

Pengujian dilakukan menggunakan confusion matrix melalui library SVM, dari model training dan testing yang telah diklasifikasikan sebelumnya. Hingga menghasilkan matriks dengan ukuran 3x3 sebagai representatif kelas aktual dan kelas prediksi. Hasil dari model training dengan menggunakan data baru yang belum dilakukan training sebelumnya. Tabel confusion matrix disajikan Tabel 9.

Tabel 2. 9 Pengujian dan evaluasi

Data		Data Prediksi	
Aktual	Negatif	Netral	Positif
Negatif	TNg	NgN	FN
Netral	NNg	TN	NP
Positif	FP	PN	TP

Berdasarkan hasil pengujian yang dilakukan, penelitian ini menghasilkan confusion matrix sebagai berikut

Tabel 2. 10 Hasil Confussion Matrix

Data	Data Prediksi		
Aktual	Negatif	Netral	Positif
Negatif	252	31	4
Netral	10	53	0
Positif	19	3	6

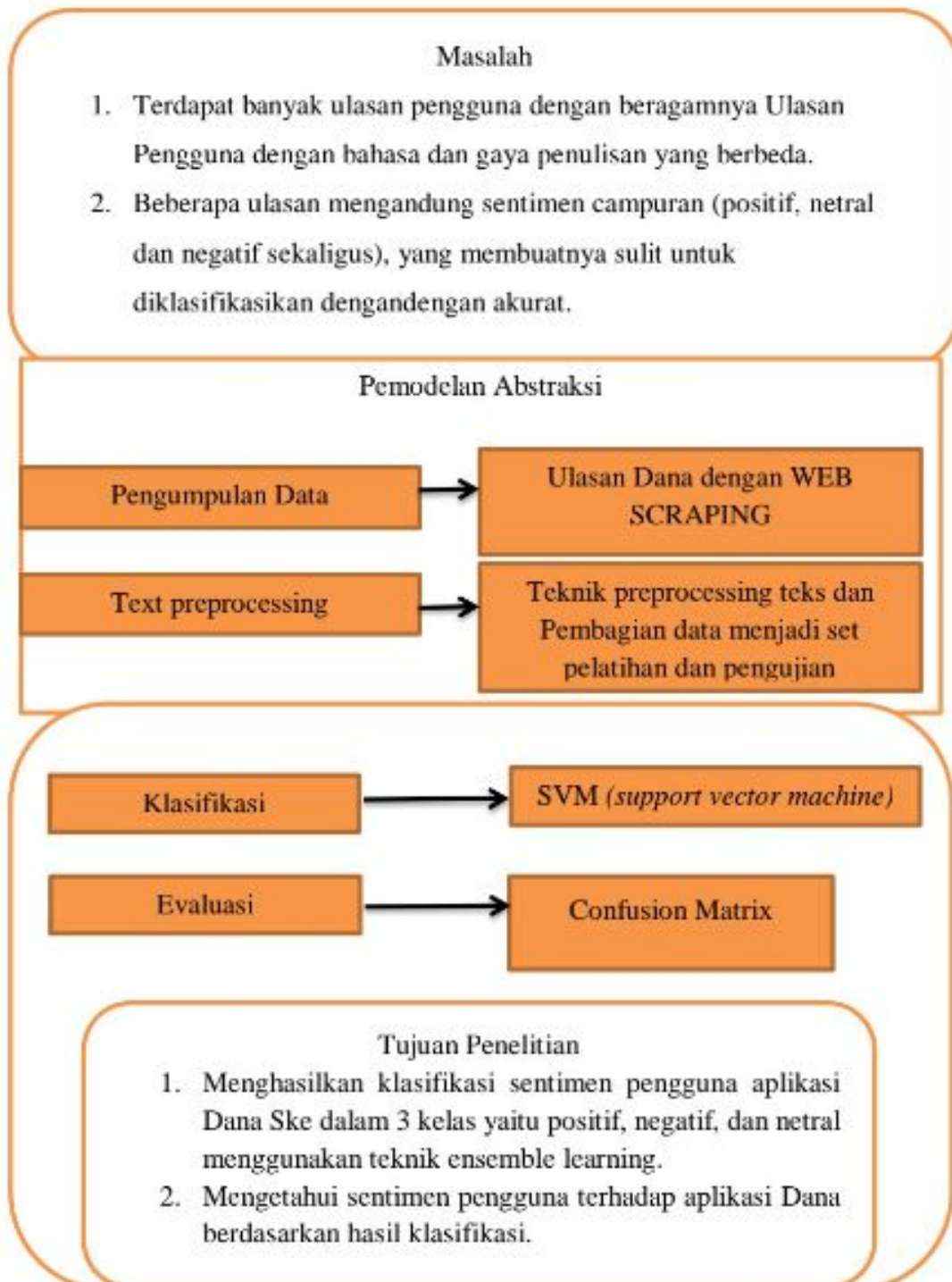
Pengujian ini menghasilkan akurasi, recall, percision serta F1-Score/F-Measure pada consol spyder yang disajikan pada Gambar 5.

	precision	recall	f1-score	support
Negatif	0.90	0.88	0.89	287
Netral	0.61	0.84	0.71	63
Positif	0.60	0.21	0.32	28
accuracy			0.82	378
macro avg	0.70	0.64	0.64	378
weighted avg	0.83	0.82	0.81	378

Gambar 2. 5 Hasil Akurasi, Recall, Precession, dan F-1 Score

Gambar 2.5 menunjukkan bahwa hasil pengujian untuk metode klasifikasi Support Vector Machine mendapatkan akurasi sebesar 82%, sebesar 90%, serta recall sebesar 88% dan f1-score sebesar 89%.

2.3 Kerangka Pikir



Gambar 2. 6 Kerangka Pemikiran

BAB III METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian

Dipandang dari tingkat penerapannya, penelitian ini merupakan penelitian terapan, apabila dipandang dari jenis informasi yang diolah, maka penelitian ini merupakan penelitian kuantitatif. Dalam penelitian ini metode yang digunakan adalah metode Ensemble Learning. untuk klasifikasi kelasnya ada 3 yaitu positif, negative dan netral. Dengan demikian penelitian ini adalah penelitian yang deskriptif, dengan tujuan untuk membuat gambaran secara sistematis, faktual, dan akurat sesuai fakta dan sifat dalam suatu objek penelitian tertentu.

Berdasarkan latar belakang dan kerangka pemikiran yang telah di jelaskan pada bab sebelumnya maka yang menjadi objek dalam penelitian ini adalah "Penggunaan Metode SVM (Support Vector Machine) Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi (Dana) di Google Play Store Penelitian ini dimulai dari bulan Agustus 2024 yang berlokasi di Di Kota Gorontalo

3.2 Studi Literatur

Studi literatur yang dilakukan dalam pembuatan tugas akhir ini adalah pembelajaran literatur terkait dengan konsep serta metode yang digunakan untuk menyelesaikan permasalahan dalam tugas akhir ini, seperti pembelajaran mengenai pola data, metode algoritma yang mendukung, dan sebagainya. Sumbertudi literatur yang digunakan dalam tugas akhir ini adalah:

1. E- book

Ebook digunakan untuk memperdalam teori penulis dalam mengerjakan tugas akhir. E-book yang digunakan meliputi teori klasifikasi dengan SVM (Support vector machine) dan teknik dalam melakukan data mining.

2. Jurnal Ilmiah

Ebook digunakan untuk memperdalam teori penulis dalam mengerjakan tugas akhir. E-book yang digunakan meliputi teori klasifikasi dengan SVM (Support Vector Machine) dan teknik dalam melakukan data mining.

3. Artikel

Artikel ini digunakan untuk menemukan fakta-fakta yang ada saat ini untuk mendukung latar belakang permasalahan. Artikel yang digunakan dalam tugas akhir ini adalah artikel berita dan yang terkait dengan kinerja Aplikasi Dana.

3.3 Pengumpulan Data

Pengumpulan data yang di gunakan dalam penelitin ini dibagi menjadi dua jenis pengumpulan data yaitu data primer dan data sekunder.

1. Data Primer

Data primer merupakan data yang dikumpulkan langsung oleh peneliti di Ulasan Aplikasi DANA berupa komentar atau Rating dari beberapa orang mengenai kinerja atau Pelayanan dari aplikasi dana

2. Data Sekunder

Data sekunder merupakan data yang di peroleh dari buku, jurnal, catatan dan informasi lain yang terkait dengan penelitian.

3.4 Tahap Praproses Data

Pada tahapan ini data mentah yang diperoleh dilakukan pemrosesan untuk mengubahnya menjadi data yang siap digunakan. Adapun proses-proses yang dilakukan dalam tahapan ini adalah adalah membersihkan dokumen dari kata yang tidak diperlukan untuk mengurangi noise, menyeragamkan bentuk huruf serta penghapusan angka dan tanda baca, mengubah dari kalimat menjadi kata-kata, melakukan proses pemilahan kata yang merupakan kata penghubung dan mengambil kata dasar dari sebuah kata yang memiliki imbuhan. Untuk tahap praproses data ini, semua langkah-langkahnya dilakukan secara manual. Berikut adalah alur praproses data yang akan dilakukan dalam penelitian ini:



Gambar 3. 1 Tahapan Preprocessing

3.5 Uji Coba dan Proses Klasifikasi

Pada tahapan ini dilakukan uji Metode SVM (Support Vector Machine) dan juga evaluasi model atau evaluasi performa dari algoritma yang digunakan dalam penelitian ini. Dalam tahapan uji coba digunakan tools R untuk mengerjakan semua prosesnya.

3.6 Tahap pelatihan dengan metode SVM (Support Vector Machine)

Pada tahap ini akan dilakukan tahap latih terhadap opini masyarakat yang sudah diketahui klasifikasinya, sehingga didapatkan model probabilitas Metode SVM (Support Vector Machine) yang nantinya akan digunakan pada tahap uji.

3.7 Evaluasi performa klasifikasi

Pada tahap ini akan dilakukan perhitungan evaluasi performa algoritma yang telah dilakukan pada saat tahap uji yaitu dengan menghitung akurasi, precision dan recallnya. Tahapan ini bertujuan untuk mengetahui keakurasian model klasifikasi yang diklasifikasikan menjadi 3 kategori yang telah diberikan di awal. Jika pada hasil evaluasi terdapat kekurangan maka dilakukan proses perbaikan pada pembuatan model klasifikasi dan dilakukan evaluasi lagi. Jika hasil sudah baik langkah selanjutnya adalah melakukan analisis terhadap hasil yang diperoleh.

3.8 Pembahasan hasil percobaan

Pada tahap ini akan dilakukan pembahasan dari semua hasil percobaan model yang telah dilakukan pada proses klasifikasi dengan menggunakan SVM (*Support Vector Machine*). Adapun dalam pengerjaan model klasifikasi tidak hanya menggunakan SVM.

3.9 Evaluasi

Pada proses evaluasi bertujuan untuk mengetahui hasil kinerja dari metode yang digunakan. Evaluasi dilakukan pada seluruh data output yang dihasilkan akan dipetakan kedalam Confusion Matrix untuk dinilai keakurasiannya.

BAB IV HASIL PENELITIAN

4.1 Pengumpulan Data

Data dalam penelitian ini diperoleh dari ulasan pengguna aplikasi Dana di *Google Play Store* menggunakan *library Python google-play-scraper*. Proses pengumpulan data menghasilkan total 5.999 ulasan yang telah melalui proses seleksi dan pembersihan. Pelabelan sentimen dilakukan menggunakan kamus sentimen Indonesia, di mana setiap kata dalam ulasan dibandingkan dengan daftar kata positif dan negatif. Ulasan dengan skor kata positif lebih banyak diberi label sentimen positif, sebaliknya ulasan dengan skor kata negatif lebih banyak diberi label sentimen negatif, dan jika jumlahnya seimbang diberi label netral.

Hasil pelabelan menunjukkan bahwa dari 5.999 ulasan, terdapat 3.376 ulasan dengan sentimen positif (56,3%), 1.616 ulasan dengan sentimen negatif (26,9%), dan 1007 ulasan dengan sentimen netral (16,8%). Distribusi sentimen ini dirangkum pada Tabel 4.1.

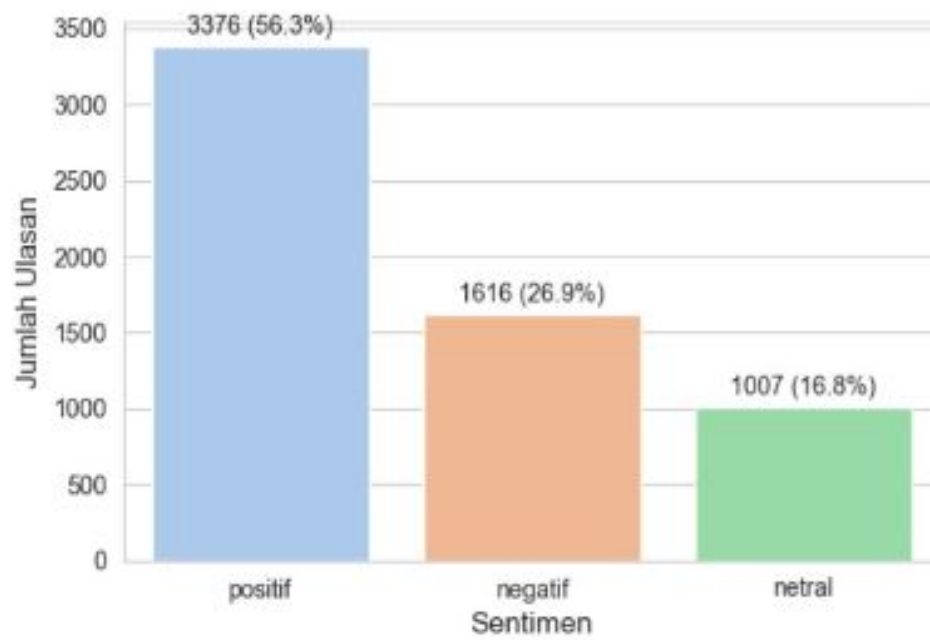
Tabel 4. 1 Hasil Pengumpulan Data

No	Nama	Ulasan	Sentimen
1	Bima Wiguna	ini knapa kok ga bisa di buka aplikasi nya. sistem sedang sibuk terus ini gimana solusinya	negatif
2	Fikri Alfarizi	Setelah diupdate kenapa gak bisa bagikan struk atau bukti transfer muter terus gak bisa bisa saya tungguin beberapa menit masih aja muter-muter, sebelumnya normal-normal aja	netral
3	Ucok Gurda	Tanggung jawab nya gimana ini yg punya aplikasi dana....di buka aja ga bisa pdhal ada orang ngirim lewat dana ke akun saya	positif

No	Nama	Ulasan	Sentimen
4	Doremi fasoL	Proses untuk transfer oke , tapi jika ada kendala dan bertanya ke cs diana , itu fast respon tapi pengerjaan untuk kendala terlalu lama mamakan waktu 18 hari , jadi saya kasi bintang segitu aja ya , terimakasih	positif
5	Indhel Tanjung	bagaimana cara masuk ke akun premium tapi nomor nya sudah tidak aktif untuk mingirim code OTP ?	netral
6	Ilwan Saleh	Setelah bertahun menggunakannya akun diproteksi dan tidak bisa dibuks	positif
7	sanz Askal	Membuat laporan aja susah untuk menghubungi CS nya, bahkan nomor Whatsapp yg dituju tidak di temukan, mohon untuk mengutamakan pelayanan Customer Service karna untuk menghubungi by phone sebagian merasa keberatan karena harus menyediakan pulsa dulu untuk telpon belum lagi tidak cukup dengan pulsa minimalis butuh pulsa ekstra yang mungkin bagi sebagian orang terlalu berat, kalau tidak ada niatan membantu konsumen gak usah bikin layanan pelanggan	negatif
8	Bims Laksana	untuk fitur dan ui sangat bagus, tapi kekurangannya tidak ada layanan bantuan cs live chat dan menyebabkan prosesnya lama	positif
9	the gemini	Dana cicil biasa gua pake malah gak dibisa dipake Uda dikirim email berapa x non	positif

No	Nama	Ulasan	Sentimen
	is love ever	tanggapan eh malah dibekukan bukan nya minta bantu untuk dipulihkan malah gak ada respon sama sekali saya rasa akan hengkang menggunakan dana	
10	asep harisun	Aplikasi yang tidak menyediakan solusi. Komplen nomor token ga muncul, katanya sudah ditangani dan masalah sudah selesai. Nyatanya di cek nomor token listrik masih belum muncul.	positif
...			
5.999	Pengguna Google	Halo, Tolong sampaikan ke developernya, masih ada beberapa bug yg saya temukan setelah saya mengoprek aplikasi ini, dan untuk systemnya sih lumayan, bikin puyeng otak, cmn akhirnya bisa masuk, tolong ya sampaikan ke developernya, *No System is Safe*  	negatif

Jumlah dan Persentase Sentimen Ulasan Pengguna Aplikasi Dana

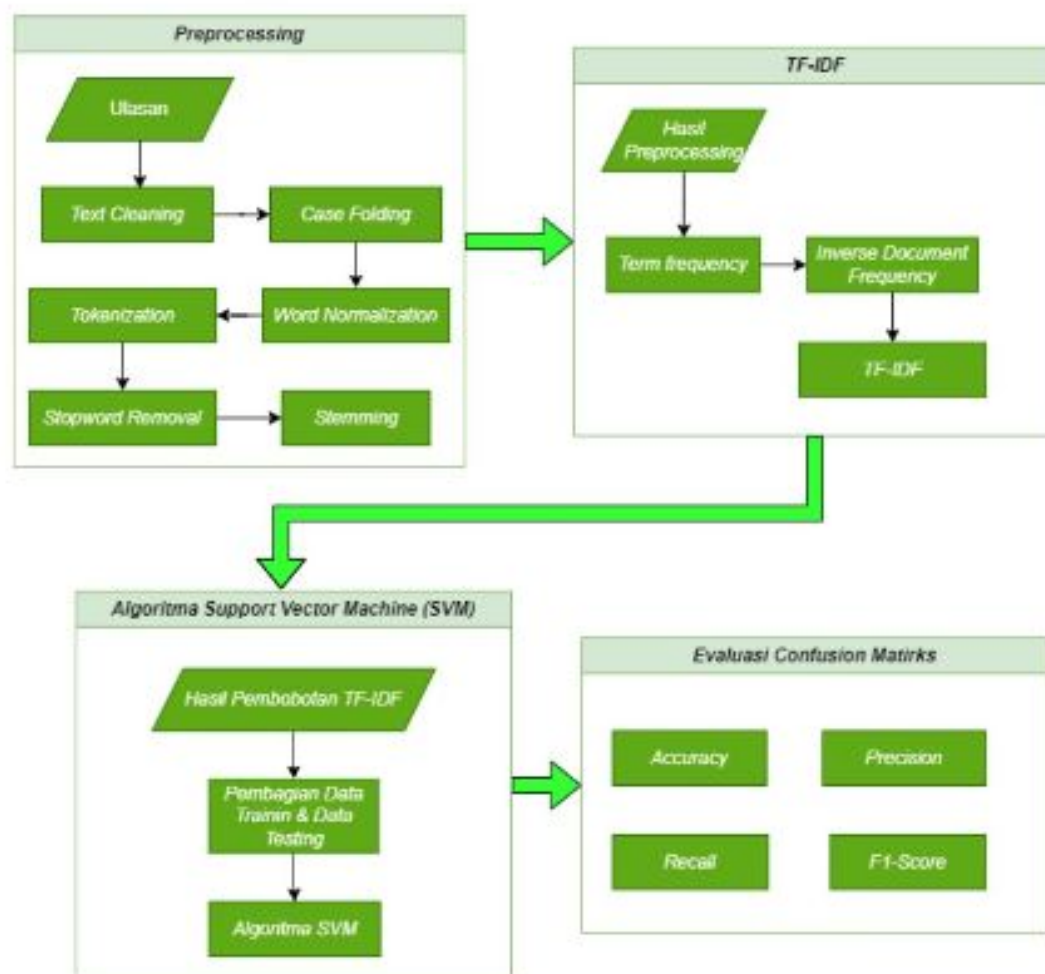


Gambar 4. 1 Hasil Pengumpulan data

4.2 Hasil Pemodelan

Pengembangan model dilakukan dengan melatih *algorithm Support Vector Machine* (SVM) menggunakan data ulasan yang telah melalui preprocessing dan pembobotan TF-IDF. Data dibagi menjadi training dan testing set untuk memastikan model dapat memprediksi sentimen secara akurat. peneliti menggunakan metode *Random Over Sampler* dari pustaka *imbalanced-learn*, yang menambah jumlah sampel pada kelas minoritas sehingga distribusi antar kelas lebih seimbang. Setelah data diseimbangkan, model diuji menggunakan data *testing*, dan evaluasi dengan metrik akurasi, presisi, *recall*, serta *F1-score*.

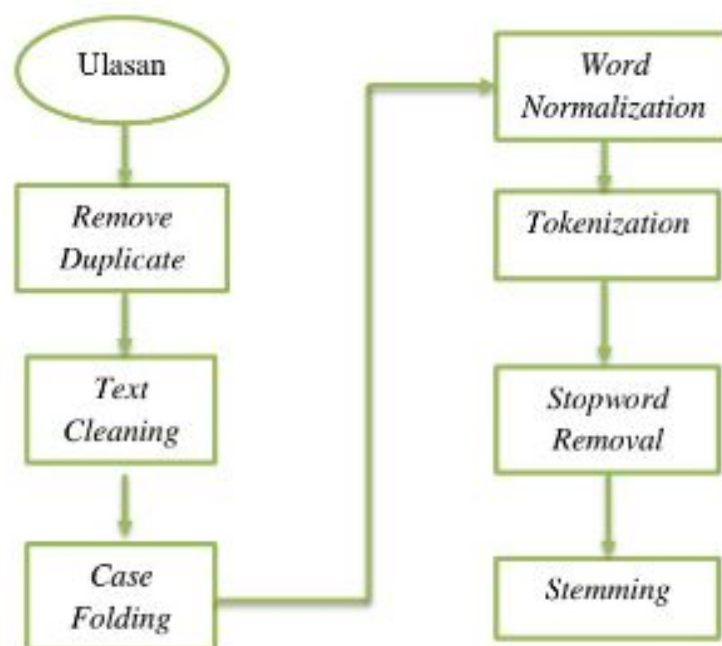
Tahapan Pemodelan *Support Vector Machine* (SVM) dapat dilihat pada gambar flowchart berikut:



Gambar 4. 2 Pemodelan SVM

4.2.1 Hasil Preprocessing

Peneliti melakukan *preprocessing* data ulasan melalui beberapa tahapan, seperti penghapusan data duplikat, pembersihan teks dari karakter yang tidak relevan seperti angka, tanda baca, dan URL, *case folding* untuk konsistensi huruf kecil, normalisasi kata tidak baku, *tokenisasi*, penghapusan *stopword*, dan *stemming* untuk mengubah kata ke bentuk dasar. Tahapan Preprocessing dapat dilihat melalui flowchart berikut:



Gambar 4. 3 Tahapan Preprocessing

Hasil dari tiap tahapan *preprocessing* ulasan pengguna menggunakan bahasa pemrograman Python ditampilkan sebagai berikut:

a. Text Cleaning

Tahapan *text cleaning* pada penelitian ini dilakukan untuk membersihkan data ulasan dari elemen-elemen yang tidak relevan. Proses ini mencakup penghapusan angka, emoji, URL, simbol, dan tanda baca. Selain itu, dilakukan penghapusan huruf dan kata yang berulang lebih dari dua kali untuk meningkatkan konsistensi teks. Hasil *text cleaning* dapat dilihat sebagai berikut:

Tabel 4. 2 Hasil Text Cleaning

No	Ulasan	Text Cleaning
1	ini knapa kok ga bisa di buka aplikasi nya. sistem sedang sibuk terus ini gimana solusinya	ini knapa kok ga bisa di buka aplikasi nya sistem sedang sibuk terus ini gimana solusinya
2	Setelah diupdate kenapa gak bisa bagikan struk atau bukti transfer muter terus gak bisa bisa saya tungguin beberapa menit masih aja muter-muter, sebelumnya normal-normal aja	Setelah diupdate kenapa gak bisa bagikan struk atau bukti transfer muter terus gak bisa saya tungguin beberapa menit masih aja muter sebelumnya normal aja
3	Tanggung jawab nya gimana ini yg punya aplikasi dana....di buka aja ga bisa pdhal ada orang ngirim lewat dana ke akun saya	Tanggung jawab nya gimana ini yg punya aplikasi dana di buka aja ga bisa pdhal ada orang ngirim lewat dana ke akun saya
4	Proses untuk transfer oke , tapi jika ada kendala dan bertanya ke cs diana , itu fast respon tapi pengerjaan untuk kendala terlalu lama mamakan waktu 18 hari , jadi saya kasi bintang segitu aja ya , terimakasih	Proses untuk transfer oke tapi jika ada kendala dan bertanya ke cs diana itu fast respon tapi pengerjaan untuk kendala terlalu lama mamakan waktu hari jadi saya kasi bintang segitu aja ya terimakasih
5	bagaimana cara masuk ke akun premium tapi nomor nya sudah tidak aktif untuk mingirim code OTP ?	bagaimana cara masuk ke akun premium tapi nomor nya sudah tidak aktif untuk mingirim code OTP

No	Ulasan	<i>Text Cleaning</i>
6	Setelah bertahun menggunakan akun diproteksi dan tidak bisa dibuks	Setelah bertahun menggunakannya akun diproteksi dan tidak bisa dibuks
7	Membuat laporan aja susah untuk menghubungi CS nya, bahkan nomor Whatsapp yg dituju tidak di temukan, mohon untuk mengutamakan pelayanan Customer Service karna untuk menghubungi by phone sebagian merasa keberatan karena harus menyediakan pulsa dulu untuk telpon belum lagi tidak cukup dengan pulsa minimalis butuh pulsa ekstra yang mungkin bagi sebagian orang terlalu berat, kalau tidak ada niatan membantu konsumen gak usah bikin layanan pelanggan	Membuat laporan aja susah untuk menghubungi CS nya bahkan nomor Whatsapp yg dituju tidak di temukan mohon untuk mengutamakan pelayanan Customer Service karna untuk menghubungi by phone sebagian merasa keberatan karena harus menyediakan pulsa dulu untuk telpon belum lagi tidak cukup dengan pulsa minimalis butuh pulsa ekstra yang mungkin bagi sebagian orang terlalu berat kalau tidak ada niatan membantu konsumen gak usah bikin layanan pelanggan
8	untuk fitur dan ui sangat bagus, tapi kekurangannya tidak ada layanan bantuan cs live chat dan menyebabkan prosesnya lama	untuk fitur dan ui sangat bagus tapi kekurangannya tidak ada layanan bantuan cs live chat dan menyebabkan prosesnya lama
9	Dana cicil biasa gua pake malah gak dibisa dipake Uda dikirim email berapa x non tanggapan ehh malah dihekukan bukan nya minta bantu untuk dipulihkan	Dana cicil biasa gua pake malah gak dibisa dipake Uda dikirim email berapa x non tanggapan ehh malah dihekukan bukan nya minta bantu untuk dipulihkan malah gak ada

No	Ulasan	Text Cleaning
	malah gak ada respon sama sekali saya rasa akan hengkang menggunakan dana	respon sama sekali saya rasa akan hengkang menggunakan dana
10	Aplikasi yang tidak menyediakan solusi. Komplen nomor token ga muncul, katanya sudah ditangani dan masalah sudah selesai. Nyatanya di cek nomor token listrik masih belum muncul.	Aplikasi yang tidak menyediakan solusi Komplen nomor token ga muncul katanya sudah ditangani dan masalah sudah selesai Nyatanya di cek nomor token listrik masih belum muncul

b. case folding

Tahapan *case folding* dilakukan untuk mengubah seluruh teks dalam data ulasan menjadi huruf kecil. Proses ini bertujuan untuk menyamakan format teks sehingga variasi huruf besar dan kecil tidak memengaruhi analisis. Sebagai contoh, kata "Dana" dan "dana" akan dianggap sama setelah *case folding*. Hasil *case folding* ditampilkan sebagai berikut:

Tabel 4. 3 Hasil Case folding

No	Text Cleaning	Case Folding
1	ini knapa kok ga bisa di buka aplikasi nya sistem sedang sibuk terus ini gimana solusinya	ini knapa kok ga bisa di buka aplikasi nya sistem sedang sibuk terus ini gimana solusinya
2	Setelah diupdate kenapa gak bisa bagikan struk atau bukti transfer muter terus gak bisa saya tungguin	setelah diupdate kenapa gak bisa bagikan struk atau bukti transfer muter terus gak bisa saya tungguin

No	<i>Text Cleaning</i>	<i>Case Folding</i>
	beberapa menit masih aja muter sebelumnya normal aja	beberapa menit masih aja muter sebelumnya normal aja
3	Tanggung jawab nya gimana ini yg punya aplikasi dana di buka aja ga bisa pdhal ada orang ngirim lewat dana ke akun saya	tanggung jawab nya gimana ini yg punya aplikasi dana di buka aja ga bisa pdhal ada orang ngirim lewat dana ke akun saya
4	Proses untuk transfer oke tapi jika ada kendala dan bertanya ke cs diana itu fast respon tapi pengerjaan untuk kendala terlalu lama mamakan waktu hari jadi saya kasi bintang segitu aja ya terimakasih	proses untuk transfer oke tapi jika ada kendala dan bertanya ke cs diana itu fast respon tapi pengerjaan untuk kendala terlalu lama mamakan waktu hari jadi saya kasi bintang segitu aja ya terimakasih
5	bagaimana cara masuk ke akun premium tapi nomor nya sudah tidak aktif untuk mingirim code OTP	bagaimana cara masuk ke akun premium tapi nomor nya sudah tidak aktif untuk mingirim code otp
6	Setelah bertahun menggunakannya akun diproteksi dan tidak bisa dibuks	setelah bertahun menggunakannya akun diproteksi dan tidak bisa dibuks
7	Membuat laporan aja susah untuk menghubungi CS nya bahkan nomor Whatsapp yg dituju tidak di temukan mohon untuk mengutamakan pelayanan Customer Service karna untuk	membuat laporan aja susah untuk menghubungi cs nya bahkan nomor whatsapp yg dituju tidak di temukan mohon untuk mengutamakan pelayanan customer service karna untuk

No	<i>Text Cleaning</i>	<i>Case Folding</i>
	menghubungi by phone sebagian merasa keberatan karena harus menyediakan pulsa dulu untuk telpon belum lagi tidak cukup dengan pulsa minimalis butuh pulsa ekstra yang mungkin bagi sebagian orang terlalu berat kalau tidak ada niatan membantu konsumen gak usah bikin layanan pelanggan	menghubungi by phone sebagian merasa keberatan karena harus menyediakan pulsa dulu untuk telpon belum lagi tidak cukup dengan pulsa minimalis butuh pulsa ekstra yang mungkin bagi sebagian orang terlalu berat kalau tidak ada niatan membantu konsumen gak usah bikin layanan pelanggan
8	untuk fitur dan ui sangat bagus tapi kekurangannya tidak ada layanan bantuan cs live chat dan menyebabkan prosesnya lama	untuk fitur dan ui sangat bagus tapi kekurangannya tidak ada layanan bantuan cs live chat dan menyebabkan prosesnya lama
9	Dana cicil biasa gua pake malah gak dibisa dipake Uda dikirim email berapa x non tanggapan ehh malah dibekukan bukan nya minta bantu untuk dipulihkan malah gak ada respon sama sekali saya rasa akan hengkang menggunakan dana	dana cicil biasa gua pake malah gak dibisa dipake uda dikirim email berapa x non tanggapan ehh malah dibekukan bukan nya minta bantu untuk dipulihkan malah gak ada respon sama sekali saya rasa akan hengkang menggunakan dana
10	Aplikasi yang tidak menyediakan solusi Komplek nomor token ga muncul katanya sudah ditangani dan masalah sudah selesai Nyatanya di cek nomor token listrik masih belum muncul	aplikasi yang tidak menyediakan solusi komplek nomor token ga muncul katanya sudah ditangani dan masalah sudah selesai nyatanya di cek nomor token listrik masih belum muncul

c. *Word normalization*

Tahapan normalisasi kata bertujuan untuk mengubah kata-kata tidak baku atau slang menjadi bentuk bakunya. Proses ini penting untuk menyamakan variasi penulisan kata yang memiliki makna sama, misalnya kata “gak” diubah menjadi “tidak” atau “cpt” menjadi “cepat.”. hasil normalisasi kata ditampilkan sebagai berikut:

Tabel 4. 4 Hasil Normalisasi Kata

No	<i>Case Folding</i>	<i>Word Normalization</i>
1	ini knapa kok ga bisa di buka aplikasi nya sistem sedang sibuk terus ini gimana solusinya	ini knapa kok tidak bisa di buka aplikasi ya sistem sedang sibuk terus ini bagaimana solusinya
2	setelah diupdate kenapa gak bisa bagikan struk atau bukti transfer muter terus gak bisa saya tungguin beberapa menit masih aja muter sebelumnya normal aja	setelah diupdate kenapa tidak bisa bagikan struk atau bukti transfer muter terus tidak bisa saya tungguin beberapa menit masih saja muter sebelumnya normal saja
3	tanggung jawab nya gimana ini yg punya aplikasi dana di buka aja ga bisa pdhal ada orang ngirim lewat dana ke akun saya	tanggung jawab ya bagaimana ini yang punya aplikasi dana di buka saja tidak bisa padahal ada orang mengirim lewat dana ke akun saya
4	proses untuk transfer oke tapi jika ada kendala dan bertanya ke cs diana itu fast respon tapi pengerjaan untuk kendala terlalu lama mamakan waktu hari jadi	proses untuk transfer oke tapi jika ada kendala dan bertanya ke cs diana itu fast respon tapi pengerjaan untuk kendala terlalu lama mamakan waktu hari jadi saya

No	<i>Case Folding</i>	<i>Word Normalization</i>
	saya kasi bintang segitu aja ya terimakasih	kasih bintang segitu saja ya terimakasih
5	bagaimana cara masuk ke akun premium tapi nomor nya sudah tidak aktif untuk mengirim code otp	bagaimana cara masuk ke akun premium tapi nomor ya sudah tidak aktif untuk mengirim code otp
6	setelah bertahun menggunakannya akun diproteksi dan tidak bisa dibuks	setelah bertahun menggunakannya akun diproteksi dan tidak bisa dibuks
7	membuat laporan aja susah untuk menghubungi cs nya bahkan nomor whatsapp yg dituju tidak di temukan mohon untuk mengutamakan pelayanan customer service karna untuk menghubungi by phone sebagian merasa keberatan karena harus menyediakan pulsa dulu untuk telpon belum lagi tidak cukup dengan pulsa minimalis butuh pulsa ekstra yang mungkin bagi sebagian orang terlalu berat kalau tidak ada niatan membantu konsumen gak usah bikin layanan pelanggan	membuat laporan saja susah untuk menghubungi cs ya bahkan nomor whatsapp yang dituju tidak di temukan mohon untuk mengutamakan pelayanan customer service karena untuk menghubungi by phone sebagian merasa keberatan karena harus menyediakan pulsa dulu untuk telpon belum lagi tidak cukup dengan pulsa minimalis butuh pulsa ekstra yang mungkin bagi sebagian orang terlalu berat kalau tidak ada niatan membantu konsumen tidak usah bikin layanan pelanggan
8	untuk fitur dan ui sangat bagus tapi kekurangannya tidak ada layanan	untuk fitur dan ui sangat bagus tapi kekurangannya tidak ada layanan

No	Case Folding	Word Normalization
	bantuan cs live chat dan menyebabkan prosesnya lama	bantuan cs live chat dan menyebabkan prosesnya lama
9	dana cicil biasa gua pake malah gak dibisa dipake uda dikirim email berapa x non tanggapan eh malah dibekukan bukan nya minta bantu untuk dipulihkan malah gak ada respon sama sekali saya rasa akan henggang menggunakan dana	dana cicil biasa gua pakai malah tidak dibisa dipakai sudah dikirim email berapa kali non tanggapan eh malah dibekukan bukan ya meminta bantu untuk dipulihkan malah tidak ada respon sama sekali saya rasa akan henggang menggunakan dana
10	aplikasi yang tidak menyediakan solusi komplek nomor token ga muncul katanya sudah ditangani dan masalah sudah selesai nyatanya di cek nomor token listrik masih belum muncul	aplikasi yang tidak menyediakan solusi komplek nomor token tidak muncul katanya sudah ditangani dan masalah sudah selesai nyatanya di cek nomor token listrik masih belum muncul

d. Tokenization

Tahapan *tokenization* bertujuan untuk memecah teks ulasan menjadi unit-unit kecil berupa kata atau token. Proses ini membantu dalam analisis teks dengan mengubah kalimat menjadi sekumpulan kata yang dapat diolah lebih lanjut oleh algoritma pemodelan. Misalnya, kalimat "Aplikasi ini sangat membantu" diubah menjadi ["Aplikasi", "ini", "sangat", "membantu"]. Hasil *tokenization* di tampilkan seperti berikut

Tabel 4. 5 Hasil Tokenizing

No	Word Normalization	Tokenization
1	ini knapa kok tidak bisa di buka aplikasi ya sistem sedang sibuk terus ini bagaimana solusinya	['ini', 'knapa', 'kok', 'tidak', 'bisa', 'di', 'buka', 'aplikasi', 'ya', 'sistem', 'sedang', 'sibuk', 'terus', 'ini', 'bagaimana', 'solusinya']
2	setelah diupdate kenapa tidak bisa bagikan struk atau bukti transfer muter terus tidak bisa saya tungguin beberapa menit masih saja muter sebelumnya normal saja	['setelah', 'diupdate', 'kenapa', 'tidak', 'bisa', 'bagikan', 'struk', 'atau', 'bukti', 'transfer', 'muter', 'terus', 'tidak', 'bisa', 'saya', 'tungguin', 'beberapa', 'menit', 'masih', 'saja', 'muter', 'sebelumnya', 'normal', 'saja']
3	tanggung jawab ya bagaimana ini yang punya aplikasi dana di buka saja tidak bisa padahal ada orang mengirim lewat dana ke akun saya	['tanggung', 'jawab', 'ya', 'bagaimana', 'ini', 'yang', 'punya', 'aplikasi', 'dana', 'di', 'buka', 'saja', 'tidak', 'bisa', 'padahal', 'ada', 'orang', 'mengirim', 'lewat', 'dana', 'ke', 'akun', 'saya']
4	proses untuk transfer oke tapi jika ada kendala dan bertanya ke cs diana itu fast respon tapi pengerjaan untuk kendala terlalu lama mamakan waktu hari jadi saya kasih bintang segitu saja ya terimakasih	['proses', 'untuk', 'transfer', 'oke', 'tapi', 'jika', 'ada', 'kendala', 'dan', 'bertanya', 'ke', 'cs', 'diana', 'itu', 'fast', 'respon', 'tapi', 'pengerjaan', 'untuk', 'kendala', 'terlalu', 'lama', 'mamakan', 'waktu', 'hari', 'jadi', 'saya', 'kasih', 'bintang', 'segitu', 'saja', 'ya', 'terimakasih']

No	Word Normalization	Tokenization
5	bagaimana cara masuk ke akun premium tapi nomor ya sudah tidak aktif untuk mingirim code otp	['bagaimana', 'cara', 'masuk', 'ke', 'akun', 'premium', 'tapi', 'nomor', 'ya', 'sudah', 'tidak', 'aktif', 'untuk', 'mingirim', 'code', 'otp']
6	setelah bertahun menggunakannya akun diproteksi dan tidak bisa dibuks	['setelah', 'bertahun', 'menggunakannya', 'akun', 'diproteksi', 'dan', 'tidak', 'bisa', 'dibuks']
7	membuat laporan saja susah untuk menghubungi cs ya bahkan nomor whatsapp yang dituju tidak di temukan mohon untuk mengutamakan pelayanan customer service karena untuk menghubungi by phone sebagian merasa keberatan karena harus menyediakan pulsa dulu untuk telpon belum lagi tidak cukup dengan pulsa minimalis butuh pulsa ekstra yang mungkin bagi sebagian orang terlalu berat kalau tidak ada niatan membantu konsumen tidak usah bikin layanan pelanggan	['membuat', 'laporan', 'saja', 'susah', 'untuk', 'menghubungi', 'cs', 'ya', 'bahkan', 'nomor', 'whatsapp', 'yang', 'dituju', 'tidak', 'di', 'temukan', 'mohon', 'untuk', 'mengutamakan', 'pelayanan', 'customer', 'service', 'karena', 'untuk', 'menghubungi', 'by', 'phone', 'sebagian', 'merasa', 'keberatan', 'karena', 'harus', 'menyediakan', 'pulsa', 'dulu', 'untuk', 'telpon', 'belum', 'lagi', 'tidak', 'cukup', 'dengan', 'pulsa', 'minimalis', 'butuh', 'pulsa', 'ekstra', 'yang', 'mungkin', 'bagi', 'sebagian', 'orang', 'terlalu', 'berat', 'kalau', 'tidak', 'ada', 'niatan', 'membantu', 'konsumen', 'tidak', 'usah', 'bikin', 'layanan', 'pelanggan']
8	untuk fitur dan ui sangat bagus tapi kekurangannya tidak ada layanan	['untuk', 'fitur', 'dan', 'ui', 'sangat', 'bagus', 'tapi', 'kekurangannya',

No	Word Normalization	Tokenization
	bantuan cs live chat dan menyebabkan prosesnya lama	'tidak', 'ada', 'layanan', 'bantuan', 'cs', 'live', 'chat', 'dan', 'menyebabkan', 'prosesnya', 'lama']
9	dana cicil biasa gua pakai malah tidak dibisa dipakai sudah dikirim email berapa kali non tanggapan eh malah dibekukan bukan ya meminta bantu untuk dipulihkan malah tidak ada respon sama sekali saya rasa akan henggang menggunakan dana	['dana', 'cicil', 'biasa', 'gua', 'pakai', 'malah', 'tidak', 'dibisa', 'dipakai', 'sudah', 'dikirim', 'email', 'berapa', 'kali', 'non', 'tanggapan', 'eh', 'malah', 'dibekukan', 'bukan', 'ya', 'meminta', 'bantu', 'untuk', 'dipulihkan', 'malah', 'tidak', 'ada', 'respon', 'sama', 'sekali', 'saya', 'rasa', 'akan', 'henggang', 'menggunakan', 'dana']
10	aplikasi yang tidak menyediakan solusi komplek nomor token tidak muncul katanya sudah ditangani dan masalah sudah selesai nyatanya di cek nomor token listrik masih belum muncul	['aplikasi', 'yang', 'tidak', 'menyediakan', 'solusi', 'komplek', 'nomor', 'token', 'tidak', 'muncul', 'katanya', 'sudah', 'ditangani', 'dan', 'masalah', 'sudah', 'selesai', 'nyatanya', 'di', 'cek', 'nomor', 'token', 'listrik', 'masih', 'belum', 'muncul']

e. Stopword removal

Tahapan stopword removal bertujuan untuk menghapus kata-kata yang dianggap tidak memberikan kontribusi signifikan dalam analisis, seperti "dan," "yang," atau "dengan." Kata-kata ini umumnya tidak memiliki makna penting dalam konteks sentimen ditampilkan sebagai berikut:

Tabel 4. 6 Hasil Stopword Removal

No	Tokenization	Stopword Removal
1	['ini', 'knapa', 'kok', 'tidak', 'bisa', 'di', 'buka', 'aplikasi', 'ya', 'sistem', 'sedang', 'sibuk', 'terus', 'ini', 'bagaimana', 'solusinya']	['knapa', 'buka', 'aplikasi', 'sistem', 'sibuk', 'solusinya']
2	['setelah', 'diupdate', 'kenapa', 'tidak', 'bisa', 'bagikan', 'struk', 'atau', 'bukti', 'transfer', 'muter', 'terus', 'tidak', 'bisa', 'saya', 'tungguin', 'beberapa', 'menit', 'masih', 'saja', 'muter', 'sebelumnya', 'normal', 'saja']	['diupdate', 'bagikan', 'struk', 'bukti', 'transfer', 'muter', 'tungguin', 'menit', 'muter', 'normal']
3	['tanggung', 'jawab', 'ya', 'bagaimana', 'ini', 'yang', 'punya', 'aplikasi', 'dana', 'di', 'buka', 'saja', 'tidak', 'bisa', 'padahal', 'ada', 'orang', 'mengirim', 'lewat', 'dana', 'ke', 'akun', 'saya']	['tanggung', 'aplikasi', 'dana', 'buka', 'orang', 'mengirim', 'dana', 'akun']
4	['proses', 'untuk', 'transfer', 'oke', 'tapi', 'jika', 'ada', 'kendala', 'dan', 'bertanya', 'ke', 'cs', 'diana', 'itu', 'fast', 'respon', 'tapi', 'pengerjaan', 'untuk', 'kendala', 'terlalu', 'lama', 'mamakan', 'waktu', 'hari', 'jadi', 'saya', 'kasih', 'bintang', 'segitu', 'saja', 'ya', 'terimakasih']	['proses', 'transfer', 'oke', 'kendala', 'cs', 'diana', 'fast', 'respon', 'pengerjaan', 'kendala', 'mamakan', 'kasih', 'bintang', 'segitu', 'terimakasih']

No	Tokenization	Stopword Removal
5	['bagaimana', 'cara', 'masuk', 'ke', 'akun', 'premium', 'tapi', 'nomor', 'ya', 'sudah', 'tidak', 'aktif', 'untuk', 'mingirim', 'code', 'otp']	['masuk', 'akun', 'premium', 'nomor', 'aktif', 'mingirim', 'code', 'otp']
6	['setelah', 'bertahun', 'menggunakannya', 'akun', 'diproteksi', 'dan', 'tidak', 'bisa', 'dibuks']	['bertahun', 'menggunakannya', 'akun', 'diproteksi', 'dibuks']
7	['membuat', 'laporan', 'saja', 'susah', 'untuk', 'menghubungi', 'cs', 'ya', 'bahkan', 'nomor', 'whatsapp', 'yang', 'dituju', 'tidak', 'di', 'temukan', 'mohon', 'untuk', 'mengutamakan', 'pelayanan', 'customer', 'service', 'karena', 'untuk', 'menghubungi', 'by', 'phone', 'sebagian', 'merasa', 'keberatan', 'karena', 'harus', 'menyediakan', 'pulsa', 'dulu', 'untuk', 'telpon', 'belum', 'lagi', 'tidak', 'cukup', 'dengan', 'pulsa', 'minimalis', 'butuh', 'pulsa', 'ekstra', 'yang', 'mungkin', 'bagi', 'sebagian', 'orang', 'terlalu', 'berat', 'kalau', 'tidak', 'ada', 'niatan', 'membantu', 'konsumen', 'tidak', 'usah', 'bikin', 'layanan', 'pelanggan']	['laporan', 'susah', 'menghubungi', 'cs', 'nomor', 'whatsapp', 'dituju', 'temukan', 'mohon', 'mengutamakan', 'pelayanan', 'customer', 'service', 'menghubungi', 'by', 'phone', 'keberatan', 'menyediakan', 'pulsa', 'telpon', 'pulsa', 'minimalis', 'butuh', 'pulsa', 'ekstra', 'orang', 'berat', 'niatan', 'membantu', 'konsumen', 'bikin', 'layanan', 'pelanggan']
8	['untuk', 'fitur', 'dan', 'ui', 'sangat', 'bagus', 'tapi', 'kekurangannya']	['fitur', 'ui', 'bagus', 'kekurangannya', 'layanan']

No	Tokenization	Stopword Removal
	'tidak', 'ada', 'layanan', 'bantuan', 'cs', 'live', 'chat', 'dan', 'menyebabkan', 'prosesnya', 'lama']	'bantuan', 'cs', 'live', 'chat', 'menyebabkan', 'prosesnya']
9	['dana', 'cicil', 'biasa', 'gua', 'pakai', 'malah', 'tidak', 'dibisa', 'dipakai', 'sudah', 'dikirim', 'email', 'berapa', 'kali', 'non', 'tanggapan', 'eh', 'malah', 'dibekukan', 'bukan', 'ya', 'meminta', 'bantu', 'untuk', 'dipulihkan', 'malah', 'tidak', 'ada', 'respon', 'sama', 'sekali', 'saya', 'rasa', 'akan', 'hengkang', 'menggunakan', 'dana']	['dana', 'cicil', 'gua', 'pakai', 'dibisa', 'dipakai', 'dikirim', 'email', 'kali', 'non', 'tanggapan', 'eh', 'dibekukan', 'bantu', 'dipulihkan', 'respon', 'hengkang', 'dana']
10	['aplikasi', 'yang', 'tidak', 'menyediakan', 'solusi', 'komplen', 'nomor', 'token', 'tidak', 'muncul', 'katanya', 'sudah', 'ditangani', 'dan', 'masalah', 'sudah', 'selesai', 'nyatanya', 'di', 'cek', 'nomor', 'token', 'listrik', 'masih', 'belum', 'muncul']	['aplikasi', 'menyediakan', 'solusi', 'komplen', 'nomor', 'token', 'muncul', 'ditangani', 'selesai', 'cek', 'nomor', 'token', 'listrik', 'muncul']

f. Stemming

Tahapan stemming bertujuan untuk mengubah kata-kata dalam ulasan ke bentuk dasarnya. Misalnya, kata "bermain" akan diubah menjadi "main," atau "menjalankan" menjadi "jalan." Proses ini membantu mengurangi variasi kata yang memiliki makna sama sehingga analisis lebih konsisten dan akurat. Hasil stemming ditampilkan sebagai berikut:

Tabel 4. 7 Hasil Stemming

No	<i>Stopword Removal</i>	<i>Stemming</i>
1	['knapa', 'buka', 'aplikasi', 'sistem', 'sibuk', 'solusinya']	knapa buka aplikasi sistem sibuk solusi
2	['diupdate', 'bagikan', 'struk', 'bukti', 'transfer', 'muter', 'tungguin', 'menit', 'muter', 'normal']	diupdate bagi struk bukti transfer muter tungguin menit muter normal
3	['tanggung', 'aplikasi', 'dana', 'buka', 'orang', 'mengirim', 'dana', 'akun']	tanggung aplikasi dana buka orang kirim dana akun
4	['proses', 'transfer', 'oke', 'kendala', 'cs', 'diana', 'fast', 'respon', 'pengerjaan', 'kendala', 'mamakan', 'kasih', 'bintang', 'segitu', 'terimakasih']	proses transfer oke kendala cs ana fast respon kerja kendala mama kasih bintang segitu terimakasih
5	['masuk', 'akun', 'premium', 'nomor', 'aktif', 'mingirim', 'code', 'otp']	masuk akun premium nomor aktif mingirim code otp
6	['bertahun', 'menggunakannya', 'akun', 'diproteksi', 'dibuks']	tahun guna akun proteksi dibuks
7	['laporan', 'susah', 'menghubungi', 'cs', 'nomor', 'whatsapp', 'dituju', 'temukan', 'mohon', 'mengutamakan', 'pelayanan', 'customer', 'service', 'menghubungi', 'by', 'phone', 'keberatan', 'menyediakan', 'pulsa', 'telpon', 'pulsa', 'minimalis', 'butuh', 'pulsa', 'ekstra', 'orang', 'berat', 'niatan']	lapor susah hubung cs nomor whatsapp tuju temu mohon utama layan customer service hubung by phone berat sedia pulsa telpon pulsa minimal butuh pulsa ekstra orang berat niat

No	Stopword Removal	Stemming
	'membantu', 'konsumen', 'bikin', 'layanan', 'pelanggan']	bantu konsumen bikin layan langgan
8	['fitur', 'ui', 'bagus', 'kekurangannya', 'layanan', 'bantuan', 'cs', 'live', 'chat', 'menyebabkan', 'prosesnya']	fitur ui bagus kurang layan bantu cs live chat sebab proses
9	['dana', 'cicil', 'gua', 'pakai', 'dibisa', 'dipakai', 'dikirim', 'email', 'kali', 'non', 'tanggapan', 'eh', 'dibekukan', 'bantu', 'dipulihkan', 'respon', 'hengkang', 'dana']	dana cicil gua pakai bisa pakai kirim email kali non tanggap eh beku bantu pulih respon hengkang dana
10	['aplikasi', 'menyediakan', 'solusi', 'komplen', 'nomor', 'token', 'muncul', 'ditangani', 'selesai', 'cek', 'nomor', 'token', 'listrik', 'muncul']	aplikasi sedia solusi komplek nomor token muncul tangan selesai cek nomor token listrik muncul

4.2.2 Hasil Pembobotan TF-IDF

Setelah proses preprocessing, tahap selanjutnya adalah pembobotan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF digunakan untuk menghitung bobot atau tingkat kepentingan setiap kata dalam dokumen ulasan. *Term Frequency* (TF) mengukur seberapa sering sebuah kata muncul dalam satu ulasan, sedangkan *Inverse Document Frequency* (IDF) mengurangi bobot kata-kata yang sering muncul di banyak ulasan, sehingga lebih menonjolkan kata-kata yang spesifik dan relevan dengan ulasan tertentu.

Pembobotan kata dilakukan menggunakan dokumen D1, D2, D3, D4, D6, D7, D8, D9, dan D10.

1) *Term Frequency* (TF)

Pada tahap ini, dilakukan perhitungan kemunculan setiap term dalam dokumen ulasan. *Term Frequency* menunjukkan seberapa sering sebuah kata muncul dalam sebuah dokumen. Berikut adalah hasil perhitungan TF dari beberapa sampel dokumen ulasan (D1 – D10):

Tabel 4. 8 Sample Term dan Perhitungan TF

No	Terms	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
1	admin	0	0	0	1	0	0	0	0	0	0
2	alfamat	0	0	0	2	0	0	0	0	0	0
3	aman	2	0	0	0	0	0	0	0	0	0
4	apk	0	0	0	0	0	1	0	2	0	0
5	aplikasi	0	0	1	1	0	0	1	0	1	0
6	ataw	1	0	0	0	0	0	0	0	0	0
7	ayo	0	0	0	0	0	0	0	1	0	0
8	baca	0	0	0	0	0	0	0	1	0	0
9	bagus	0	0	0	1	0	0	1	2	0	0
10	baik	1	0	0	0	0	0	0	1	0	0
11	balas	0	0	0	0	0	0	0	0	0	1
12	banget	0	0	1	0	0	0	0	0	0	0
13	bayar	0	0	0	0	0	0	0	0	0	1
14	beli	0	0	0	0	0	0	0	0	0	1
15	biaya	0	0	0	0	0	0	3	0	0	0

No	Terms	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
35	isi	0	0	0	0	0	0	1	0	0	0
36	jamin	1	0	0	0	0	0	0	0	0	0
37	kali	0	0	0	1	0	0	1	0	0	0
38	kecewa	0	0	0	0	0	0	0	0	0	2
39	kena	0	0	0	1	0	0	2	0	0	0
40	kendala	0	0	0	0	0	0	0	0	1	0
41	konfirmasi	1	0	0	0	0	0	0	0	0	0
42	konsumen	0	0	0	0	0	0	1	0	0	0
43	langkah	0	1	0	0	0	0	0	0	0	0
44	layar	0	0	0	0	0	0	0	0	1	0
45	live	0	0	0	0	0	0	0	0	0	1
46	login	0	1	0	0	0	0	0	0	0	0
47	loh	0	0	0	0	0	0	0	1	0	0
48	lot	0	0	1	0	0	0	0	0	0	0
49	masuk	0	0	0	0	0	1	0	0	0	0
50	min	0	0	0	1	0	0	0	0	0	0
51	minimal	0	0	0	1	0	0	0	0	0	0
52	muncul	0	0	0	0	0	0	0	0	0	1
53	namum	0	0	0	0	1	0	0	0	0	0

No	Terms	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
54	nasib	1	0	0	0	0	0	0	0	0	0
55	nomer	0	0	1	0	0	0	0	0	0	0
56	nomor	0	0	0	0	0	0	0	0	0	1
57	orang	1	0	0	0	0	0	0	0	0	0
58	perhati	0	0	0	0	0	0	1	0	0	0
59	pikir	0	0	0	0	0	0	1	0	0	0
60	play	0	0	0	0	0	0	0	0	1	0
61	playstore	0	0	0	0	0	1	0	0	0	0
62	pulsa	0	0	0	1	0	0	0	0	0	0
63	ribu	0	0	0	1	0	0	0	0	0	0
64	robot	0	0	0	0	0	0	0	0	0	1
65	saldo	0	0	0	0	0	0	1	0	0	0
66	sayang	0	0	0	0	1	0	0	0	0	0
67	scroll	0	0	0	0	0	0	0	1	0	0
68	sedia	0	0	0	0	1	0	0	0	0	0
69	sistim	1	0	0	0	0	0	0	0	0	0
70	store	0	0	0	0	0	0	0	0	1	0
71	susah	0	0	1	0	0	0	0	0	0	0
72	tambah	0	0	0	2	0	0	0	0	0	0

No	Terms	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
73	tanggung	0	0	0	0	0	0	0	0	0	1
74	tarik	0	0	0	3	0	0	0	0	0	0
75	token	0	0	0	1	0	0	0	0	0	2
76	tolong	1	0	0	1	0	0	1	0	0	0
77	toup	0	0	0	1	0	0	0	0	0	0
78	transaksi	2	0	0	0	0	0	0	0	0	0
79	transfer	0	0	0	0	0	0	1	0	0	0
80	tunai	0	0	0	3	0	0	0	0	0	0
81	uang	1	0	0	0	0	0	0	0	0	0
82	verifikasi	0	0	0	0	1	0	0	0	0	0
Panjang Dokumen		19	4	6	26	8	6	17	11	9	14

Selanjutnya setelah mendapatkan hasil dari Term Frequency (TF), langkah selanjutnya Tahap normalisasi Term Frequency (TF) yang dilakukan untuk menghindari bias pada dokumen dengan panjang yang berbeda. Proses ini menghasilkan nilai TF yang lebih seimbang dengan cara membagi frekuensi kemunculan setiap term dengan jumlah total kata dalam dokumen tersebut. Sebagai contoh, term "bagus" memiliki bobot normalisasi tertinggi 0.600 pada Doc5 karena muncul 3 kali dalam dokumen yang hanya terdiri dari 5 kata. Sementara term "proses" mendapat bobot 0.273 di Doc1 meski muncul 3 kali, karena dokumen tersebut lebih panjang (11 kata). Hasil dari TF Normalisasi dapat dilihat sebagai berikut:

Tabel 4. 9 Sampel Perhitungan TF Normalisasi

Terms	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
admin	0	0	0	0,038	0	0	0	0	0	0
alfamat	0	0	0	0,077	0	0	0	0	0	0
aman	0,105	0	0	0	0	0	0	0	0	0
apk	0	0	0	0	0	0,167	0	0,182	0	0
aplikasi	0	0	0,167	0,038	0	0	0,059	0	0,111	0
ataw	0,053	0	0	0	0	0	0	0	0	0
ayo	0	0	0	0	0	0	0	0,091	0	0
baca	0	0	0	0	0	0	0	0,091	0	0
bagus	0	0	0	0,038	0	0	0,059	0,182	0	0
baik	0,053	0	0	0	0	0	0	0,091	0	0
balas	0	0	0	0	0	0	0	0	0	0,071
banget	0	0	0,167	0	0	0	0	0	0	0
bayar	0	0	0	0	0	0	0	0	0	0,071
beli	0	0	0	0	0	0	0	0	0	0,071
biaya	0	0	0	0	0	0	0,176	0	0	0
bohong	0,053	0	0	0	0	0	0	0	0	0
bulan	0	0	0	0,038	0	0	0	0	0	0
capek	0	0	0	0	0	0,167	0	0	0	0
cari	0	0	0	0	0	0	0	0	0,111	0
chat	0	0	0	0	0	0	0	0	0	0,071
cicil	0	0	0	0	0,25	0	0	0	0	0
cuman	0	0	0	0	0	0	0	0,091	0	0
dana	0,053	0,25	0	0	0,25	0	0,118	0	0,111	0
danaviznya	0,053	0	0	0	0	0	0	0	0	0
data	0	0	0	0,038	0	0	0	0	0	0
deh	0	0	0	0	0	0,167	0	0	0	0
fitur	0,053	0	0	0	0	0	0	0	0	0
ganti	0	0	0,167	0	0	0	0	0	0	0

Terms	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
perhati	0	0	0	0	0	0	0,059	0	0	0
pikir	0	0	0	0	0	0	0,059	0	0	0
play	0	0	0	0	0	0	0	0	0,111	0
playstore	0	0	0	0	0	0,167	0	0	0	0
pulsa	0	0	0	0,038	0	0	0	0	0	0
ribu	0	0	0	0,038	0	0	0	0	0	0
robot	0	0	0	0	0	0	0	0	0	0,071
saldo	0	0	0	0	0	0	0,059	0	0	0
sayang	0	0	0	0	0,125	0	0	0	0	0
scroll	0	0	0	0	0	0	0	0,091	0	0
sedia	0	0	0	0	0,125	0	0	0	0	0
sistim	0,053	0	0	0	0	0	0	0	0	0
store	0	0	0	0	0	0	0	0	0,111	0
susah	0	0	0,167	0	0	0	0	0	0	0
tambah	0	0	0	0,077	0	0	0	0	0	0
tanggung	0	0	0	0	0	0	0	0	0	0,071
tarik	0	0	0	0,115	0	0	0	0	0	0
token	0	0	0	0,038	0	0	0	0	0	0,143
tolong	0,053	0	0	0,038	0	0	0,059	0	0	0
toup	0	0	0	0,038	0	0	0	0	0	0
transaksi	0,105	0	0	0	0	0	0	0	0	0
transfer	0	0	0	0	0	0	0,059	0	0	0
tunai	0	0	0	0,115	0	0	0	0	0	0
uang	0,053	0	0	0	0	0	0	0	0	0
verifikasi	0	0	0	0	0,125	0	0	0	0	0
Panjang Doc	19	4	6	26	8	6	17	11	9	14

2) Tahap Inverse Document Frequency (IDF)

Pada tahap ini, dilakukan perhitungan bobot IDF untuk setiap term yang muncul dalam kumpulan dokumen ulasan dana. Proses perhitungan dilakukan dengan menggunakan formula logaritmik yang membandingkan jumlah total dokumen dengan jumlah dokumen yang mengandung term tertentu. Hasil Perhitungan IDF dari D1-D10 adalah sebagai berikut:

Tabel 4. 10 Sampel Perhitungan DF dan IDF

No	Terms	DF	IDF
1	admin	1	1,7047
2	alfamat	1	1,7047
3	aman	1	1,7047
4	apk	2	1,2993
5	aplikasi	4	0,7885
6	ataw	1	1,7047
7	ayo	1	1,7047
8	baca	1	1,7047
9	bagus	3	1,0116
10	baik	2	1,2993
11	balas	1	1,7047
12	banget	1	1,7047
13	bayar	1	1,7047
14	beli	1	1,7047
15	biaya	1	1,7047
16	bohong	1	1,7047
17	bulan	1	1,7047
18	capek	1	1,7047
19	cari	1	1,7047
20	chat	1	1,7047
21	cicil	1	1,7047
22	cuman	1	1,7047

No	Terms	DF	IDF
23	dana	5	0,6061
24	danaviznya	1	1,7047
25	data	1	1,7047
26	deh	1	1,7047
27	fitur	1	1,7047
28	ganti	1	1,7047
29	gratis	1	1,7047
30	hasil	2	1,2993
31	hilang	2	1,2993
32	hp	2	1,2993
33	ikut	1	1,7047
34	informasi	1	1,7047
35	isi	1	1,7047
36	jamin	1	1,7047
37	kali	2	1,2993
38	kecewa	1	1,7047
39	kena	2	1,2993
40	kendala	1	1,7047
41	konfirmasi	1	1,7047
42	konsumen	1	1,7047
43	langkah	1	1,7047
44	layar	1	1,7047
45	live	1	1,7047
46	login	1	1,7047
47	loh	1	1,7047
48	lot	1	1,7047
49	masuk	1	1,7047
50	min	1	1,7047
51	minimal	1	1,7047

No	Terms	DF	IDF
52	muncul	1	1,7047
53	namum	1	1,7047
54	nasib	1	1,7047
55	nomer	1	1,7047
56	nomor	1	1,7047
57	orang	1	1,7047
58	perhati	1	1,7047
59	pikir	1	1,7047
60	play	1	1,7047
61	playstore	1	1,7047
62	pulsa	1	1,7047
63	ribu	1	1,7047
64	robot	1	1,7047
65	saldo	1	1,7047
66	sayang	1	1,7047
67	scroll	1	1,7047
68	sedia	1	1,7047
69	sistim	1	1,7047
70	store	1	1,7047
71	susah	1	1,7047
72	tambah	1	1,7047
73	tanggung	1	1,7047
74	tarik	1	1,7047
75	token	2	1,2993
76	tolong	3	1,0116
77	toup	1	1,7047
78	transaksi	1	1,7047
79	transfer	1	1,7047
80	tunai	1	1,7047

No	Terms	DF	IDF
81	uang	1	1,7047
82	verifikasi	1	1,7047

3) Perhitungan TF-IDF

Tahap akhir dalam pembobotan kata adalah menghitung TF-IDF dengan mengalikan nilai *Term Frequency* (TF) normalisasi dengan *Inverse Document Frequency* (IDF). Proses ini menghasilkan bobot akhir yang menggambarkan kepentingan suatu term dalam dokumen dan seluruh kumpulan dokumen. Hasil Perhitungan TF-IDF dapat dilihat sebagai berikut:

Tabel 4. 11 Sampel Perhitungan TF-IDF

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
0	0	0	0,066	0	0	0	0	0	0
0	0	0	0,131	0	0	0	0	0	0
0,18	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0,217	0	0,236	0	0
0	0	0,131	0,03	0	0	0,046	0	0,088	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0,155	0	0
0	0	0	0	0	0	0	0,155	0	0
0	0	0	0,039	0	0	0,059	0,184	0	0
0,068	0	0	0	0	0	0	0,118	0	0
0	0	0	0	0	0	0	0	0	0,122
0	0	0,284	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,122
0	0	0	0	0	0	0	0	0	0,122
0	0	0	0	0	0	0,301	0	0	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0	0,066	0	0	0	0	0	0

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
0	0	0	0	0	0,284	0	0	0	0
0	0	0	0	0	0	0	0	0,189	0
0	0	0	0	0	0	0	0	0	0,122
0	0	0	0	0,426	0	0	0	0	0
0	0	0	0	0	0	0	0,155	0	0
0,032	0,152	0	0	0,152	0	0,071	0	0,067	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0	0,066	0	0	0	0	0	0
0	0	0	0	0	0,284	0	0	0	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0,284	0	0	0	0	0	0	0
0	0	0	0,131	0	0	0	0	0	0
0,068	0	0	0	0	0	0	0	0	0,093
0,068	0	0	0	0	0	0	0	0,144	0
0	0	0	0	0	0,217	0	0	0,144	0
0	0,426	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0,155	0	0
0	0	0	0	0	0	0,1	0	0	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0	0,05	0	0	0,076	0	0	0
0	0	0	0	0	0	0	0	0	0,244
0	0	0	0,05	0	0	0,153	0	0	0
0	0	0	0	0	0	0	0	0,189	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0,1	0	0	0
0	0,426	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0,189	0
0	0	0	0	0	0	0	0	0	0,122
0	0,426	0	0	0	0	0	0	0	0

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
0	0	0	0	0	0	0	0,155	0	0
0	0	0,284	0	0	0	0	0	0	0
0	0	0	0	0	0,284	0	0	0	0
0	0	0	0,066	0	0	0	0	0	0
0	0	0	0,066	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,122
0	0	0	0	0,213	0	0	0	0	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0,284	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,122
0,09	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0,1	0	0	0
0	0	0	0	0	0	0,1	0	0	0
0	0	0	0	0	0	0	0	0,189	0
0	0	0	0	0	0,284	0	0	0	0
0	0	0	0,066	0	0	0	0	0	0
0	0	0	0,066	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,122
0	0	0	0	0	0	0,1	0	0	0
0	0	0	0	0,213	0	0	0	0	0
0	0	0	0	0	0	0	0,155	0	0
0	0	0	0	0,213	0	0	0	0	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0,189	0
0	0	0,284	0	0	0	0	0	0	0
0	0	0	0,131	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0,122
0	0	0	0,197	0	0	0	0	0	0
0	0	0	0,05	0	0	0	0	0	0,186

D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
0,053	0	0	0,039	0	0	0,059	0	0	0
0	0	0	0,066	0	0	0	0	0	0
0,18	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0,1	0	0	0
0	0	0	0,197	0	0	0	0	0	0
0,09	0	0	0	0	0	0	0	0	0
0	0	0	0	0,213	0	0	0	0	0

4.2.3 Klasifikasi *Support Vector Machine* (SVM)

Dataset yang telah melalui tahap *preprocessing* dan *feature extraction* selanjutnya digunakan dalam proses pembelajaran menggunakan metode *Support Vector Machine* (SVM). Proses ini memanfaatkan fungsi *Support Vector Classifier* (SVC) dengan kernel *linear* untuk memisahkan data dalam ruang fitur secara optimal. Berbagai nilai parameter C dievaluasi, yaitu $C = [0.1, 0.25, 0.5, 0.75, 1]$ guna mengukur pengaruhnya terhadap performa model. Adapun kode program yang di gunakan adalah sebagai berikut:

```

c_values = [0.1,0.25,0.5,0.75,1]
results = []
for c in c_values:
    print(f"\nEvaluasi Kinerja SVM dengan C={c}")
    model = SVC(kernel='linear',gamma='auto', C=c, probability=True)
    model.fit(X_train_tfidf, y_train)

    predictions = model.predict(X_test_tfidf)
    probabilities = model.predict_proba(X_test_tfidf)

    accuracy = accuracy_score(y_test, predictions)
    precision = precision_score(y_test, predictions, average='weighted')
    recall = recall_score(y_test, predictions, average='weighted')
    f1 = f1_score(y_test, predictions, average='weighted')
    results.append({
        'C': c,
        'Accuracy': accuracy,
        'Precision': precision,
        'Recall': recall,
        'F1 Score': f1
    })
print(f"Accuracy: {accuracy:.2%}")
print(f"Precision: {precision:.2%}")
print(f"Recall: {recall:.2%}")
print(f"F1 Score: {f1:.2%}")

```

Gambar 4. 4 Source Code Klasifikasi SVC

Kemudian di tentukan nilai akurasi tertinggi dari parameter C yang ditampilkan sebagai berikut:

Tabel 4. 12 Akurasi SVM berdasarkan C value

C	Accuracy
0.1	0.690833
0.25	0.758333
0.50	0.775000
0.75	0.790000
1.00	0.797500

Hasil akurasi di atas menunjukkan performa model SVM dengan berbagai nilai parameter C, yang mempengaruhi tingkat toleransi model terhadap kesalahan klasifikasi. Dari tabel, terlihat bahwa akurasi meningkat seiring dengan kenaikan

nilai C , dimulai dari 69,08% pada $C = 0.1$ hingga mencapai 79,75% pada $C = 1.0$. Laporan Hasil klasifikasi model SVM dapat dilihat dibawah ini:

Akurasi: 79.75 %				
Classification Report:				
	precision	recall	f1-score	support
negatif	0.82	0.80	0.81	345
netral	0.50	0.30	0.38	204
positif	0.84	0.95	0.89	651
accuracy			0.80	1200
macro avg	0.72	0.68	0.69	1200
weighted avg	0.77	0.80	0.78	1200

Gambar 4. 5 hasil klasifikasi sebelum balancing data

Model SVM dengan $C = 1.0$ menunjukkan kinerja dengan akurasi 79,75%. Namun, performa pada kelas "netral" masih kurang optimal dibandingkan kelas lainnya, yang dapat disebabkan oleh kurangnya jumlah data "netral" atau kesulitan dalam membedakan pola dari data tersebut.

4.2.4 Penanganan *Imbalanced Data* dengan *Random Over Sampler*

Untuk mengatasi ketidakseimbangan data pada kelas "netral" yang berdampak pada performa model, digunakan metode *Random Over Sampler* dari pustaka *imbalanced-learn*. Teknik ini bertujuan untuk menambah jumlah sampel pada kelas minoritas dengan cara membuat salinan data secara acak, sehingga distribusi antar kelas menjadi lebih seimbang. Adapun kode program yang digunakan dapat dilihat di bawah ini.

```

from imblearn.over_sampling import RandomOverSampler
import pandas as pd

# Melihat distribusi data awal
print(df['sentimen'].value_counts())
print('=')

X = df.drop(columns=['sentimen'])
y = df['sentimen']

# Melakukan oversampling menggunakan RandomOverSampler
ros = RandomOverSampler(random_state=42)
X_resampled, y_resampled = ros.fit_resample(X, y)

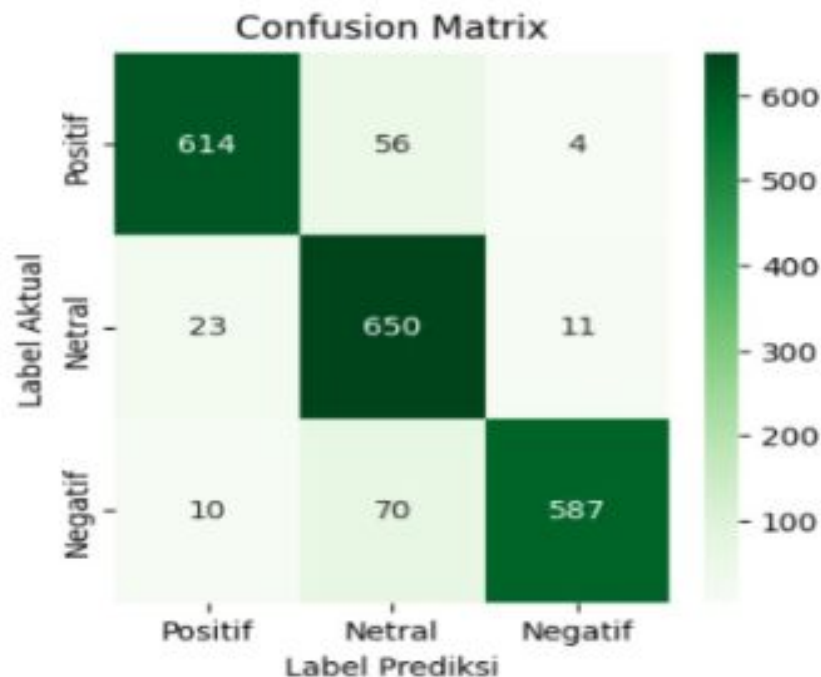
df_balanced = pd.concat([X_resampled, y_resampled], axis=1)
# Melihat distribusi data setelah balancing
print(df_balanced['sentimen'].value_counts())
df_balanced.head()

```

Gambar 4. 6 Source code Random Over Sampler

4.3 Uji Model Confusion Matriks

Pada tahap evaluasi model, *confusion matrix* digunakan untuk mengukur kinerja model klasifikasi dalam membedakan data pada masing-masing kelas. *Confusion matrix* memberikan gambaran yang jelas tentang sejauh mana model dapat mengklasifikasikan data ke dalam kelas yang benar, serta mendeteksi adanya kesalahan klasifikasi, baik pada kelas mayoritas maupun kelas minoritas. Dalam penelitian ini, *confusion matrix* diterapkan untuk menguji model klasifikasi yang telah diterapkan setelah proses *Random Over Sampling*.



Gambar 4. 7 Confusion Matriks

peneliti melakukan perhitungan secara manualisasi untuk menghitung nilai performa dari model *Support Vector Machine* pada nilai akurasi yang dihitung dari tabel *multiclass confusion matrix* di atas. Sehingga, hasil perhitungan nilai akurasi secara manualisasi dapat dijabarkan sebagai berikut :

$$akurasi = \frac{638 + 579 + 621}{638 + 68 + 8 + 63 + 579 + 73 + 1 + 24 + 621} \times 100 \quad (4. 1)$$

$$akurasi = \frac{1838}{2025} = 0,90765 \times 100 = 90,76 \%$$

Total keseluruhan dari nilai *precision*, *recall* serta *f1-score* dapat dilihat pada laporan klasifikasi model SVM yang telah melalui teknik *Random Over Sampler* di gambar berikut.

Classification Report:				
	precision	recall	f1-score	support
negatif	0.95	0.91	0.93	674
netral	0.84	0.95	0.89	684
positif	0.98	0.88	0.93	667
accuracy			0.91	2025
macro avg	0.92	0.91	0.92	2025
weighted avg	0.92	0.91	0.91	2025

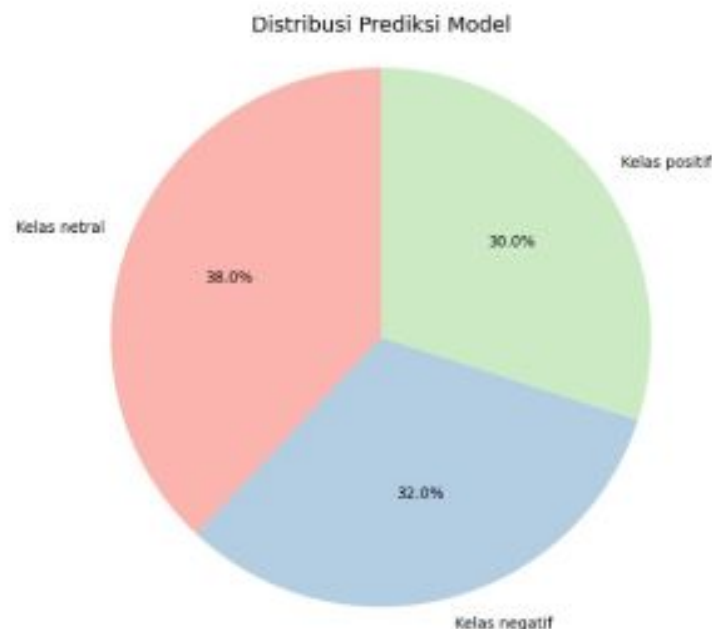
Gambar 4. 8 Hasil Klasifikasi setelah *balancing* data

4.4 Visualisasi Hasil Penelitian

Pada tahap ini akan menyajikan hasil penelitian dalam bentuk visualisasi untuk memberikan gambaran yang lebih jelas mengenai hasil analisis model yang diterapkan. Visualisasi yang digunakan oleh peneliti meliputi distribusi prediksi model, *wordcloud*, dan distribusi probabilitas sampel

4.4.1 Distribusi Prediksi Model

Distribusi prediksi model bertujuan untuk menunjukkan seberapa sering setiap label diklasifikasikan oleh model. Hal ini memberikan informasi tentang pola hasil klasifikasi, misalnya distribusi antara sentimen positif, negatif, dan netral



Gambar 4. 9 Distribusi Prediksi label

Berdasarkan *pie chart*, distribusi prediksi model menunjukkan bahwa kelas netral memiliki proporsi terbesar dengan 38%, diikuti oleh kelas negatif sebesar 32%, dan kelas positif sebesar 30%. Distribusi ini relatif seimbang, meskipun kelas netral mendominasi, yang mengindikasikan bahwa model lebih sering mengkategorikan data sebagai netral dibandingkan positif atau negatif.

4.4.2 WordCloud

Memvisualisasikan kata-kata yang sering muncul dalam data berdasarkan label sentimen.



Gambar 4. 10 Wordcloud

BAB V PEMBAHASAN

5.1 Pembahasan Model

Pada pengembangan model penelitian ini, peneliti menggunakan *Jupyter Notebook* sebagai platform utama untuk pengolahan data dan pelatihan model *Support Vector Machine* (SVM). Bahasa pemrograman yang digunakan adalah *Python*, yang dipilih karena memiliki beragam pustaka (*library*) yang mendukung pengolahan teks dan implementasi algoritma pembelajaran mesin (*machine learning*), seperti *NumPy*, *Pandas*, *Scikit-learn*, dan *Sastrawi*. Model yang telah dilatih kemudian disimpan dalam bentuk file menggunakan pustaka *joblib* atau *pickle* untuk mempermudah proses implementasi dan evaluasi di tahap berikutnya.

5.1.1 Preprocessing

Pada tahap preprocessing, Proses ini dilakukan menggunakan bahasa pemrograman *Python* dengan pustaka seperti *Pandas* untuk manipulasi data, *Sastrawi* untuk stemming, dan *Scikit-learn* untuk transformasi teks. Tahapan preprocessing meliputi konversi teks menjadi huruf kecil (*lowercasing*), penghapusan karakter khusus seperti tanda baca dan angka, tokenisasi untuk memisahkan setiap kata, penghapusan kata-kata tidak penting (*stopwords*), serta *stemming* untuk mengembalikan kata ke bentuk dasar. Berikut adalah tahapan *preprocessing* yang dilakukan dalam pengembangan model di penelitian ini, berupa hasil output dari *Jupyter Notebook*:

1) Text Cleaning

komentar	score	clean_text
Tolong perbaiki sistim ke amanan bertransaksi ...	1	Tolong perbaiki sistim ke amanan bertransaksi ...
Saya udah ikutin langkah nya, tapi saya tetap ...	3	Saya udah ikutin langkah nya tapi saya tetap ...
Kenapa aplikasi nya lemot ya mau ganti nomer s...	3	Kenapa aplikasi nya lemot ya mau ganti nomer s...
Aplikasi nya bagus buat toup pulsa dan data da...	4	Aplikasi nya bagus buat toup pulsa dan data da...
Dana cicil sudah tersedia dan sudah terverifik...	3	Dana cicil sudah tersedia dan sudah terverifik...

Gambar 5. 1 Tahap Cleantext

Tahapan *text cleaning* adalah langkah awal dalam *preprocessing* data teks yang bertujuan untuk membersihkan teks dari elemen-elemen yang tidak relevan atau mengganggu analisis.

2) Case Folding

clean_text	case_folding
tolong perbaiki sistim ke amanan bertransaksi ...	tolong perbaiki sistim ke amanan bertransaksi ...
Saya udah ikutin langkah nya tapi saya tetap ...	saya udah ikutin langkah nya tapi saya tetap ...
Kenapa aplikasi nya lemot ya mau ganti nomer s...	kenapa aplikasi nya lemot ya mau ganti nomer s...
Aplikasi nya bagus buat toup pulsa dan data da...	aplikasi nya bagus buat toup pulsa dan data da...
Dana cicil sudah tersedia dan sudah terverifik...	dana cicil sudah tersedia dan sudah terverifik...
ini kenapa ga kelihatan di hp apk nya harus ma...	ini kenapa ga kelihatan di hp apk nya harus ma...
Aplikasi sudah bagus mungkin yg diperhatikan k...	aplikasi sudah bagus mungkin yg diperhatikan k...

Gambar 5. 2 Case Folding

Tahap *case folding* adalah proses mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) untuk memastikan konsistensi dalam analisis data teks.

3) Word Normalization

	case_folding	normalisasi
0	tolong perbaiki sistim ke amanan bertransaksi ...	tolong perbaiki sistim ke amanan bertransaksi ...
1	saya udah ikutin langkah nya tapi saya tetap ...	saya sudah mengikuti langkah ya tapi saya teta...
2	kenapa aplikasi nya lemot ya mau ganti nomer s...	kenapa aplikasi ya lemot ya mau ganti nomer su...
3	aplikasi nya bagus buat toup pulsa dan data da...	aplikasi ya bagus buat toup pulsa dan data dan...
4	dana cicil sudah tersedia dan sudah terverifik...	dana cicil sudah tersedia dan sudah terverifik...

Gambar 5. 3 Normalisasi Kata

Tahap *word normalization* adalah proses mengubah kata-kata tidak baku atau slang menjadi kata baku sesuai dengan standar bahasa yang digunakan, seperti KBBI untuk bahasa Indonesia.

4) *Tokenization*

	normalisasi	tokenize
0	tolong perbaiki sistim ke amanan bertransaksi ...	[tolong, perbaiki, sistim, ke, amanan, bertran...
1	saya sudah mengikuti langkah ya tapi saya teta...	[saya, sudah, mengikuti, langkah, ya, tapi, sa...
2	kenapa aplikasi ya lemot ya mau ganti nomer su...	[kenapa, aplikasi, ya, lemot, ya, mau, ganti, ...
3	aplikasi ya bagus buat toup pulsa dan data dan...	[aplikasi, ya, bagus, buat, toup, pulsa, dan, ...
4	dana cicil sudah tersedia dan sudah terverifik...	[dana, cicil, sudah, tersedia, dan, sudah, ter...

Gambar 5. 4 Tokenization

Tahap *tokenization* adalah proses memecah teks menjadi unit-unit kecil yang disebut token, biasanya berupa kata, frasa, atau simbol. Pada tahap ini, setiap kalimat diurai menjadi kumpulan kata-kata sehingga lebih mudah untuk dianalisis secara komputasional.

5) *Stopword Removal*

	tokenize	no_stopword
0	[tolong, perbaiki, sistim, ke, amanan, bertran...	[tolong, perbaiki, sistim, amanan, bertransaks...
1	[saya, sudah, mengikuti, langkah, ya, tapi, sa...	[mengikuti, langkah, login, dana]
2	[kenapa, aplikasi, ya, lemot, ya, mau, ganti, ...	[aplikasi, lemot, ganti, nomer, susah, banget]
3	[aplikasi, ya, bagus, buat, toup, pulsa, dan, ...	[aplikasi, bagus, toup, pulsa, data, tarik, tu...
4	[dana, cicil, sudah, tersedia, dan, sudah, ter...	[dana, cicil, tersedia, terverifikasi, namum, ...

Gambar 5. 5 Stopword removal

Tahap *stopword removal* adalah proses menghapus kata-kata umum yang sering muncul dalam teks tetapi dianggap tidak memiliki makna signifikan untuk analisis, seperti "dan," "yang," "di," atau "ke." Kata-kata ini disebut *stopword* karena tidak memberikan kontribusi penting dalam memahami konteks atau makna utama suatu teks.

6) Stemmer

	no_stopword	stem_text
0	[tolong, perbaiki, sistim, amanan, bertransaks...	tolong baik sistim aman transaksi orang nasib ...
1	[mengikuti, langkah, login, dana]	ikut langkah login dana
2	[aplikasi, lemot, ganti, nomer, susah, banget]	aplikasi lot ganti nomer susah banget
3	[aplikasi, bagus, toup, pulsa, data, tarik, tu...	aplikasi bagus toup pulsa data tarik tunai min...
4	[dana, cicil, tersedia, terverifikasi, namum, ...	dana cicil sedia verifikasi namum dana cicil s...

Gambar 5. 6 Stemming

Tahap *stemming* adalah proses mengubah kata berimbuhan (*prefix*, *suffix*, atau *infix*) menjadi bentuk dasar atau root word-nya. Tujuan utama stemming adalah menyederhanakan kata sehingga variasi kata yang memiliki makna serupa dapat disamakan.

5.1.2 Pembobotan Kata

Tahap pembobotan pada penelitian ini menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk merepresentasikan data teks dalam bentuk vektor numerik. Proses pembobotan dilakukan dengan pustaka *TfidfVectorizer* dari *Scikit-learn* pada *Python*, menghasilkan matriks TF-IDF sebagai berikut.

1) Term frequency (TF)

No	Terms	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
0	1 admin	0	0	0	1	0	0	0	0	0	0
1	2 alfamat	0	0	0	2	0	0	0	0	0	0
2	3 aman	2	0	0	0	0	0	0	0	0	0
3	4 apk	0	0	0	0	0	1	0	2	0	0
4	5 aplikasi	0	0	1	1	0	0	1	0	1	0
5	6 ataw	1	0	0	0	0	0	0	0	0	0
6	7 ayo	0	0	0	0	0	0	0	1	0	0
7	8 baca	0	0	0	0	0	0	0	1	0	0
8	9 bagus	0	0	0	1	0	0	1	2	0	0
9	10 baik	1	0	0	0	0	0	0	1	0	0
10	11 balas	0	0	0	0	0	0	0	0	0	1
11	12 banget	0	0	1	0	0	0	0	0	0	0

Gambar 5. 7 Pembobotan TF

Gambar di atas menunjukkan representasi *term frequency* (TF) dari beberapa dokumen (D1 hingga D10) berdasarkan kemunculan kata-kata tertentu (terms). TF digunakan untuk menghitung frekuensi kemunculan kata dalam masing-masing dokumen sebagai bagian dari proses analisis teks. Dari tabel, terlihat bahwa setiap term memiliki distribusi frekuensi yang bervariasi di tiap dokumen. Misalnya, term "aman" memiliki frekuensi tertinggi pada D3 dengan nilai 2, sedangkan term "balas" hanya muncul di D10.

2) TF-IDF

No	Terms	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
0	1 admin	0.000000	0.000000	0.000000	0.056000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
1	2 alfamat	0.000000	0.000000	0.000000	0.131000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	3 aman	0.180000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	4 apk	0.000000	0.000000	0.000000	0.000000	0.000000	0.217000	0.000000	0.236000	0.000000	0.000000
4	5 aplikasi	0.000000	0.000000	0.131000	0.030000	0.000000	0.000000	0.040000	0.000000	0.088000	0.000000
5	6 atsw	0.090000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	7 ayo	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.155000	0.000000	0.000000
7	8 baca	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.155000	0.000000	0.000000
8	9 bagus	0.000000	0.000000	0.000000	0.035000	0.000000	0.000000	0.059000	0.184000	0.000000	0.000000
9	10 bwik	0.060000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.110000	0.000000	0.000000
10	11 balas	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.122000
11	12 banget	0.000000	0.000000	0.284000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Gambar 5. 8 Hasil TF-IDF

Gambar tersebut menampilkan hasil perhitungan *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk setiap term di dokumen (D1 hingga D10). TF-IDF digunakan untuk menilai tingkat kepentingan sebuah term dalam suatu dokumen relatif terhadap kumpulan dokumen lainnya. Nilai TF-IDF yang lebih tinggi menunjukkan bahwa term tersebut lebih penting dalam dokumen tertentu dibandingkan dokumen lainnya. Sebagai contoh, term "aman" memiliki nilai TF-IDF tertinggi di D3 sebesar 0,180000, menunjukkan bahwa term ini sangat relevan di dokumen tersebut. Selain itu, term "banget" memiliki nilai TF-IDF signifikan di D4 (0,284000), menekankan bahwa kata ini sangat penting di dokumen D4 dibanding dokumen lainnya.

5.1.3 Klasifikasi

Pada tahap klasifikasi, data dibagi dengan rasio 80:20 untuk pelatihan dan pengujian. Untuk mengatasi ketidakseimbangan kelas, digunakan *Random Over Sampler*. Model *Support Vector Classifier* (SVC) diuji dengan berbagai nilai parameter C untuk menentukan trade-off antara margin dan kesalahan klasifikasi. Berikut adalah hasil pada gambar 5.9 yang menunjukkan contoh prediksi menggunakan model terbaik

```

Sample Predictions with Best Model:

Sample 1
Teks: aplikasi upgrade premium tolong saran aplikasi mudah transaksi mohon bantu
Label Sebenarnya: netral
Prediksi: netral
Probabilitas:
  Kelas Probabilitas
negatif 0.035225
netral 0.936770
positif 0.028005

Sample 2
Teks: buruk upgrade premium transaksi sulit tolong baik layan orglain susah
Label Sebenarnya: netral
Prediksi: netral
Probabilitas:
  Kelas Probabilitas
negatif 0.066116
netral 0.849169
positif 0.084715

Sample 3
Teks: banget ganggu pemberitahuan apa moga baik
Label Sebenarnya: negatif
Prediksi: negatif
Probabilitas:
  Kelas Probabilitas
negatif 0.986753
netral 0.010203
positif 0.003044

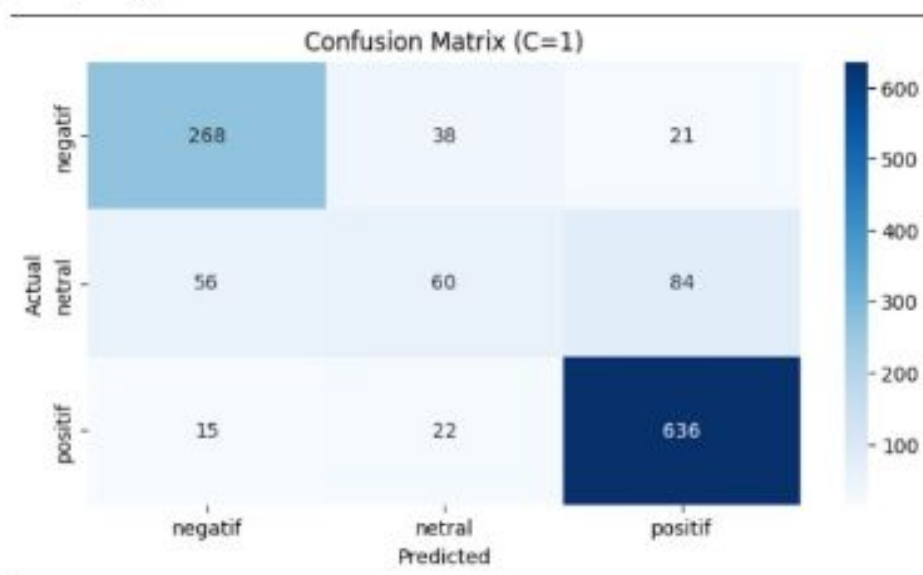
```

Gambar 5. 9 Sample Prediksi

5.1.4 Evaluasi

Pada tahap evaluasi, dilakukan perbandingan performa model berdasarkan laporan klasifikasi dan *confusion matrix* sebelum dan sesudah dilakukan *balancing* data menggunakan *Random Over Sampler*. Laporan klasifikasi mencakup metrik-metrik seperti akurasi, presisi, recall, dan F1-score yang memberikan gambaran mengenai kualitas prediksi model terhadap masing-masing kelas sentimen.

Sebelum dilakukan *balancing* data, laporan klasifikasi menunjukkan akurasi sebesar 79.75% dengan presisi dan recall yang lebih rendah pada kelas netral. Confusion matrix sebelum *balancing* memperlihatkan bahwa model cenderung lebih banyak memprediksi kelas positif, sehingga menghasilkan nilai false netral yang cukup tinggi.

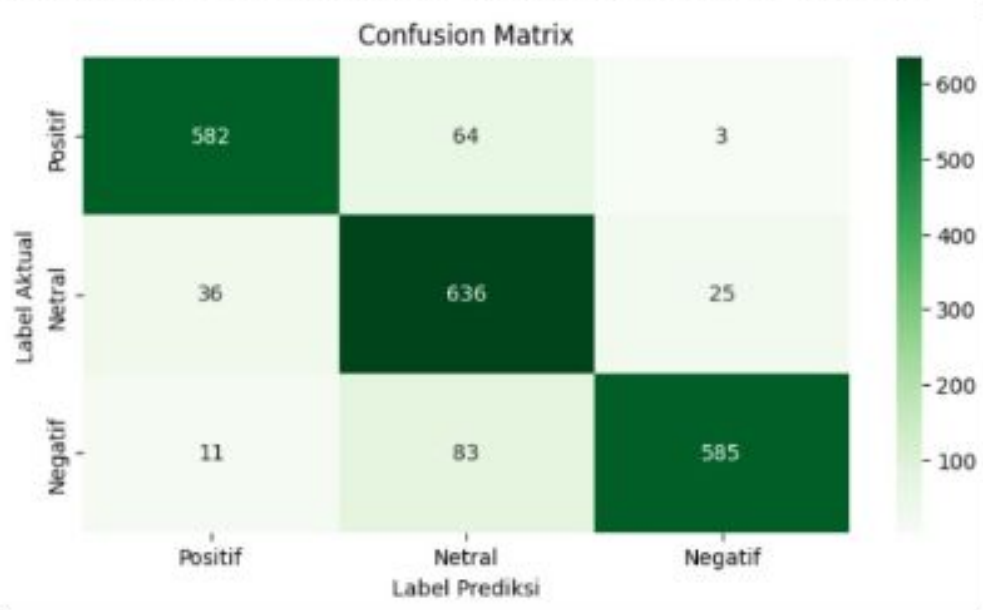


Gambar 5. 10 Confusion Matriks (sebelum balancing)

Classification Report:				
	precision	recall	f1-score	support
negatif	0.82	0.80	0.81	345
netral	0.50	0.30	0.38	204
positif	0.84	0.95	0.89	651
accuracy			0.80	1200
macro avg	0.72	0.68	0.69	1200
weighted avg	0.77	0.80	0.78	1200

Gambar 5. 11 Laporan Klasifikasi (sebelum balancing)

Setelah dilakukan *balancing* data dengan *Random Over Sampler*, hasil laporan klasifikasi menunjukkan peningkatan yang signifikan, dengan akurasi sebesar 90.42 %. Begitu pula dengan metrik presisi, recall, dan F1-score yang lebih seimbang antara kelas positif, negatif dan netral. *Confusion matrix* setelah *balancing* memperlihatkan distribusi prediksi yang lebih merata, dengan penurunan jumlah false netral dan peningkatan prediksi yang benar pada ketiga kelas.



Gambar 5. 12 Confusion Matrix (sesudah balancing)

Classification Report:				
	precision	recall	f1-score	support
negatif	0.93	0.90	0.91	649
netral	0.81	0.91	0.86	697
positif	0.95	0.86	0.91	679
accuracy			0.89	2025
macro avg	0.90	0.89	0.89	2025
weighted avg	0.90	0.89	0.89	2025

Gambar 5. 13 Laporan Klasifikasi (sesudah balancing)

5.2 Implementasi Model

Setelah model dilatih, model tersebut diimplementasikan menggunakan *Streamlit* untuk membuat aplikasi web interaktif. Aplikasi ini memungkinkan pengguna untuk memasukkan teks ulasan, yang kemudian diproses dan diklasifikasikan sentimennya oleh model SVC yang telah dilatih. Hasil klasifikasi dan probabilitas ditampilkan secara langsung di aplikasi, memungkinkan analisis sentimen secara *real-time*.

Berikut merupakan visualisasi dari implementasi model yang telah dilatih sebelumnya pada *streamlit* dengan contoh komentar pertama.

Masukkan Komentar untuk Dianalisis

Semakin diupdate, aplikasinya makin berat dan lemot. Dulu kirim uang ke bank bisa, tapi sekarang loading terus tanpa akhir. Sudah 2 bulan saya nggak bisa transfer karena loading sampai berjam-jam. Masa iya cuma bisa buat beli pulsa doang? Parah!

 Analisis

 Reset

Hasil Analisis

 **NEGATIF**

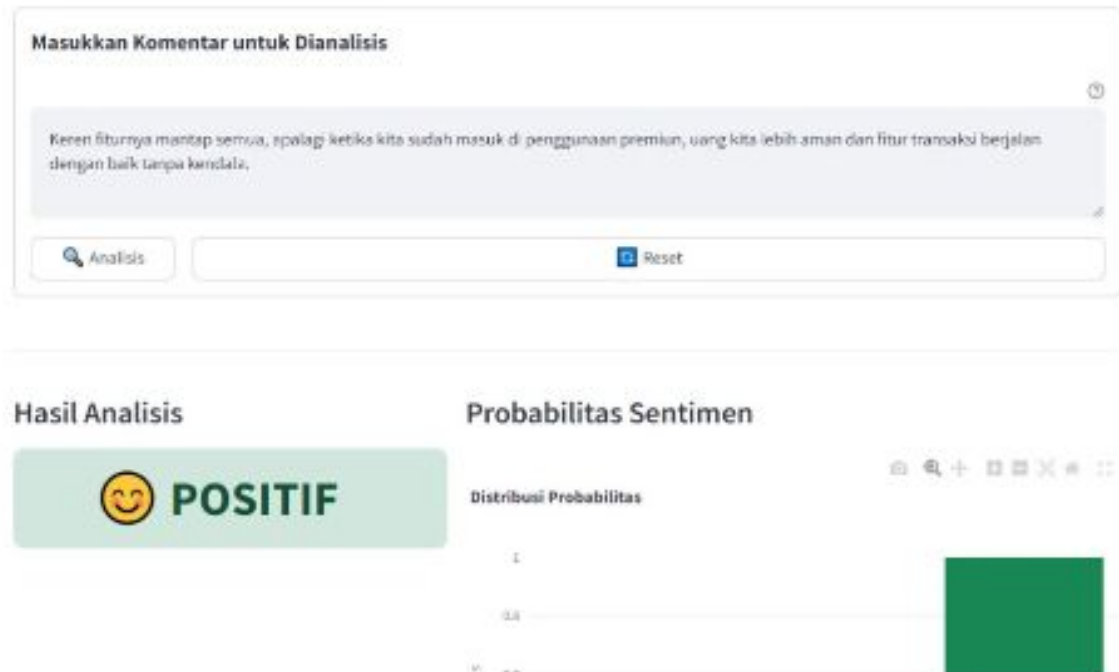
Probabilitas Sentimen

Distribusi Probabilitas

1

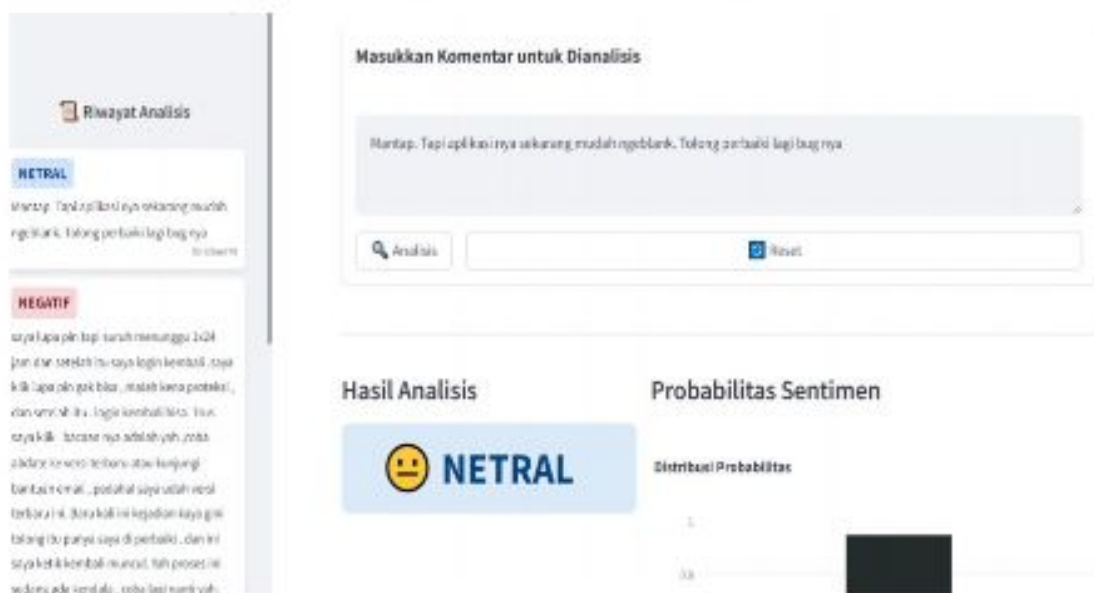
Gambar 5. 14 Hasil Implementasi Model 1

Berikut merupakan visualisasi dari implementasi model yang telah dilatih sebelumnya pada *streamlit* dengan contoh komentar kedua.



Gambar 5. 15 Hasil Implementasi Model 2

Berikut merupakan visualisasi dari implementasi model yang telah dilatih sebelumnya pada *streamlit* dengan contoh komentar Ketiga.



Gambar 5. 16 Hasil Implementasi Model 3

BAB VI PENUTUP

6.1 Kesimpulan

Berdasarkan penelitian yang dilakukan, penulis mengambil kesimpulan sebagai berikut:

1. Penelitian ini berhasil mengembangkan sistem klasifikasi sentimen yang efisien dan akurat untuk ulasan pengguna aplikasi DANA di Google Play menggunakan metode *Support Vector Machine* (SVM). Sistem ini dilengkapi dengan tahapan *preprocessing* data, pembobotan TF-IDF, dan *balancing* data menggunakan *Random Over Sampler*. Sebelum proses *balancing* data, model menghasilkan akurasi sebesar 79.75%, sedangkan setelah *balancing* data, akurasi model meningkat menjadi 90.47%, yang menunjukkan peningkatan performa model yang signifikan.
2. Metode SVM berhasil diterapkan untuk meningkatkan akurasi dan efektivitas klasifikasi sentimen. Dengan dukungan implementasi berbasis web menggunakan *Streamlit*, sistem dapat memberikan hasil klasifikasi sentimen secara *real-time* kepada pengguna, membuktikan keandalan SVM dalam menghasilkan model klasifikasi yang unggul.

6.2 Saran

Untuk pengembangan lebih lanjut, disarankan untuk menggunakan dataset yang lebih besar dan lebih beragam guna meningkatkan akurasi model, serta mengeksplorasi algoritma lain seperti Deep Learning atau Ensemble Methods untuk menangani kompleksitas data yang lebih tinggi. Selain itu, penelitian ini dapat diterapkan pada platform atau aplikasi lain di luar aplikasi Dana, seperti analisis sentimen pada media sosial, dengan menyesuaikan teknik *preprocessing* dan klasifikasi yang digunakan. Penerapan model ini secara lebih luas dapat memberikan manfaat signifikan dalam memahami sentimen pengguna secara lebih mendalam.

DAFTAR PUSTAKA

- [1] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," ... *Teknol. Inf. dan ...*, vol. 6, no. 9, pp. 4305–4313, 2022, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [2] V. H. Juli, "Technologica," vol. 3, no. 2, pp. 64–74, 2024.
- [3] D. E. Sondakh, S.Kom, M.T, Ph.D, S. W. Taju, M. G. Tene, and A. E. T. Pangaila, "Sistem Analisis Sentimen Ulasan Aplikasi Belanja Online Menggunakan Metode Ensemble Learning," *CogITO Smart J.*, vol. 9, no. 2, pp. 280–291, 2023, doi: 10.31154/cogito.v9i2.525.280-291.
- [4] A. H. Heryanti, "Pengaruh Kualitas Layanan Aplikasi Dana terhadap Kepuasan Pelanggan dalam Melakukan Transaksi Secara Online sebagai Alat Pembayaran Elektronik (E-Payment)," *J. Educ.*, vol. 5, no. 3, pp. 8080–8096, 2023, doi: 10.31004/joe.v5i3.1595.
- [5] E. Daniati and H. Utama, "Analisis Sentimen Dengan Pendekatan Ensemble Learning Dan Word Embedding Pada Twitter," *J. Inf. Syst. Manag.*, vol. 4, no. 2, pp. 125–131, 2023, doi: 10.24076/joism.2023v4i2.973.
- [6] F. F. Mailoa, "Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia," *J. Inf. Syst. Public Heal.*, vol. 6, no. 1, p. 44, 2021, doi: 10.22146/jisph.44455.
- [7] Revou.co.id, "Ensemble Learning," 2024, [Online]. Available: <https://revou.co/kosakata/ensemble-learning>
- [8] I. S. K. Idris, Y. A. Mustofa, and I. A. Salihi, "Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)," *Jambura J. Electr. Electron. Eng.*, vol. 5, no. 1, pp. 32–35, 2023, doi: 10.37905/jjee.v5i1.16830.
- [9] bisnis tekno, "Apa itu aplikasi dompet digital, kelebihan dan kekurangannya," 29 januari 2024. [Online]. Available: <https://teknologi.bisnis.com/read/20240129/84/1736046/apa-itu-aplikasi-dompet-digital-kelebihan-serta-kekurangannya>

- [10] 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.
- [11] Antaranews.com, "Apa itu aplikasi dompet digital DANA dan berbagai manfaatnya," 19 agustus 2024. [Online]. Available: <https://www.antaranews.com/berita/4270503/apa-itu-aplikasi-dompet-digital-dana-dan-berbagai-manfaatnya>
- [12] Dibimbing.com, "Klasifikasi: Definisi, Model, Langkah-Langkah, & Fungsinya," 22 agustus 2024. [Online]. Available: <https://dibimbing.id/en/blog/detail/klasifikasi-definisi-model-langkah-langkah-fungsinya>
- [13] P. B. N. Setio, D. R. S. Saputro, and Bowo Winarno, "Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5," *Prism. Pros. Semin. Nas. Mat.*, vol. 3, pp. 64–71, 2020.
- [14] D. Atika, A. Ari Aldino, S. Informasi, J. Pagar Alam No, L. Ratu, and K. Kedaton, "Term Frequency-Inverse Document Frequency Support Vector Machine Untuk Analisis Sentimen Opini Masyarakat Terhadap Tekanan Mental Pada Media Sosial Twitter," *J. Teknol. dan Sist. Inf.*, vol. 3, no. 4, p. page-page, 2022, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [15] D. Darwis, E. S. Pratiwi, and A. F. O. Pasaribu, "Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia," *Eduatic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 1–11, 2020, doi: 10.21107/edutic.v7i1.8779.

LAMPIRAN

```
import streamlit as st

import pandas as pd

from pandas import DataFrame

import seaborn as sn

import matplotlib.pyplot as plt

import numpy as np

import nltk

# nltk.download('stopwords')

from nltk.corpus import words, stopwords

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

import string

import re

from sklearn import model_selection, metrics

from sklearn.svm import LinearSVC, SVC

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score,
confusion_matrix

from sklearn.model_selection import train_test_split

from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer,
TfidfVectorizer

from PIL import Image

from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator

from collections import Counter

import plotly.graph_objects as go

import plotly.express as px

import joblib
```



```
import streamlit as st

import pandas as pd

from pandas import DataFrame

import seaborn as sn

import matplotlib.pyplot as plt

import numpy as np

import nltk

from nltk.corpus import words, stopwords

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory

import string

import re

import uuid

from sklearn import model_selection, metrics

from sklearn.svm import LinearSVC, SVC

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score,
confusion_matrix

from sklearn.model_selection import train_test_split

from sklearn.feature_extraction.text import CountVectorizer, TfidfTransformer,
TfidfVectorizer

from PIL import Image

from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator

from collections import Counter

import plotly.graph_objects as go

import plotly.express as px

import joblib

import uuid
```

```
# Function to save analysis history

def save_analysis_history(text, prediction, probabilities):

    # Create a unique ID for each analysis

    analysis_id = str(uuid.uuid4())[:8]

    # Prepare the history entry

    history_entry = {

        'id': analysis_id,

        'text': text,

        'sentiment': prediction,

        'probabilities': probabilities

    }

    # Retrieve existing history or initialize if not exists

    if 'analysis_history' not in st.session_state:

        st.session_state.analysis_history = []

    # Add new entry to the beginning of the list

    st.session_state.analysis_history.insert(0, history_entry)

    # Limit history to last 10 entries

    st.session_state.analysis_history = st.session_state.analysis_history[:10]

# Function to display history in sidebar with attractive styling

def display_analysis_history():
```

Sidebar history styling

```
st.sidebar.markdown("""
```

```
<style>
```

```
.history-container {
```

```
    background-color: #f0f2f6;
```

```
    border-radius: 10px;
```

```
    padding: 10px;
```

```
    margin-bottom: 10px;
```

```
}
```

```
.history-header {
```

```
    font-weight: bold;
```

```
    color: #333;
```

```
    margin-bottom: 10px;
```

```
    text-align: center;
```

```
}
```

```
.history-item {
```

```
    background-color: white;
```

```
    border-radius: 8px;
```

```
    padding: 10px;
```

```
    margin-bottom: 8px;
```

```
    box-shadow: 0 2px 4px rgba(0,0,0,0.1);
```

```
}
```

```
.sentiment-badge {
```

```
    display: inline-block;
```

```
    padding: 3px 8px;
```

```
        border-radius: 5px;

        font-weight: bold;

        margin-bottom: 5px;
    }

    .sentiment-positif {

        background-color: #d1e7dd;

        color: #0f5132;
    }

    .sentiment-netral {

        background-color: #cfe2ff;

        color: #084298;
    }

    .sentiment-negatif {

        background-color: #f8d7da;

        color: #842029;
    }

    .no-analysis {

        text-align: center;

        color: #6c757d;

        padding: 20px;

        background-color: #f8f9fa;

        border-radius: 10px;
    }
</style>

"""', unsafe_allow_html=True)
```

```

# History container

st.sidebar.markdown('<div class="history-container">', unsafe_allow_html=True)

st.sidebar.markdown('<h3 class="history-header"> 📖 Riwayat Analisis</h3>',
unsafe_allow_html=True)


# Always ensure analysis_history exists in session state
if 'analysis_history' not in st.session_state:

    st.session_state.analysis_history = []


# Check if there's any history
if not st.session_state.analysis_history:

    st.sidebar.markdown("""
    <div class="no-analysis">

        <p>🤖 Belum ada ulasan yang dianalisis</p>

        <small>Mulai analisis sentimen untuk melihat riwayat</small>

    </div>

    """, unsafe_allow_html=True)
else:

    # Display each history entry
    for entry in st.session_state.analysis_history:

        st.sidebar.markdown(f"""
        <div class="history-item">

            <div class="sentiment-badge sentiment-{entry['sentiment'].lower()}">

                {entry['sentiment'].upper()}

            </div>

```



```

        <br>

        <small>{entry['text']}</small>

        <div style="text-align:right; font-size:10px; color:gray;">ID: {entry['id']}</div>
    </div>

    ""', unsafe_allow_html=True)

st.sidebar.markdown('</div>', unsafe_allow_html=True)

# Clear History
if st.session_state.analysis_history:
    if st.sidebar.button('🗑️ Hapus Riwayat'):
        del st.session_state.analysis_history
        st.experimental_rerun()

st.set_page_config(
    page_title="ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI DANA",
    page_icon="📊",
    layout="wide",
    initial_sidebar_state="expanded",
    menu_items=None
)

st.image('images/dana_2.png',use_column_width=True)

tab1,tab2 = st.tabs(['Data Sentimen','Implementasi Model'])

```

```
# Pra Proses Data
```

```
stopword = set(stopwords.words('indonesian'))
```

```
stopword.update(['ya','nya'])
```

```
factory = StemmerFactory()
```

```
stemmer = factory.create_stemmer()
```

```
# Fungsi Untuk Tahap Preprocessing
```

```
import re
```

```
import unicodedata
```

```
def clean_text_advanced(text):
```

```
    # Menghapus angka
```

```
    text = re.sub(r'\d+', '', text)
```

```
    # Menghapus emoji
```

```
    text = ''.join(c for c in text if not unicodedata.decomposition(c))
```

```
    # Menghapus URL
```

```
    text = re.sub(r'https?://\S+', '', text)
```

```
    # Menghapus simbol dan tanda baca
```

```
    text = re.sub(r'[^\w\s]', '', text)
```

```
    # Menghapus huruf berulang lebih dari 2 kali dalam setiap kata
```

```
    text = re.sub(r'(\w)\1{2,}', r'\1', text)
```

```
    # Menghapus kata berulang (misalnya 'mantap mantap' menjadi 'mantap')
```

```
    text = re.sub(r'\b(\w+)(?:\s+\1\b)+', r'\1', text)
```

```
    return text.strip()
```

```
def to_lower(text):
```

```
    text = text.lower()
```

```
    return text
```

```
def replace_taboo_word(text, kamus_slag):
```

```
    if isinstance(text, str):
```

```
        words = text.split()
```

```
        replaced_words = []
```

```
        kalimat_baku = []
```

```
        kata_diganti = []
```

```
        kata_tidak_baku_hash = []
```

```
        for word in words:
```

```
            if word in kamus_slag:
```

```
                baku_word = kamus_slag[word]
```

```
                if isinstance(baku_word, str) and all(char.isalpha() for char in baku_word):
```

```
                    replaced_words.append(baku_word)
```

```
                    kalimat_baku.append(baku_word)
```

```
                    kata_diganti.append(word)
```

```
                    kata_tidak_baku_hash.append(word)
```

```
            else:
```

```
                replaced_words.append(word)
```

```
        replaced_text = ' '.join(replaced_words)
```

```
    else:
```

```

    replaced_text = "

    kalimat_baku = []

    kata_diganti = []

    kata_tidak_baku_hash = []

    return replaced_text, kalimat_baku, kata_diganti, kata_tidak_baku_hash

# initialize kamus slang
kamus_data = pd.read_excel('model/kamuskatabaku.xlsx')
kamus_tidak_baku = dict(zip(kamus_data['tidak_baku'],kamus_data['kata_baku']))

def remove_punct(text):
    text = re.sub(r'^a-zA-Z0-9',' ',str(text))
    text = re.sub(r'\b\w{1,2}\b',"",text)
    text = re.sub(r'\s\s+', ' ',text)
    return text

def tokenization(text):
    tokens = text.split()
    return tokens

def remove_stopwords(text):
    text = [word for word in text if word not in stopwords]
    return text

def stem_text(words):
    """Melakukan stemming pada list kata."""

```

```

return ' '.join([stemmer.stem(word) for word in words])

# Load data
df = pd.read_excel('hasil_labelling.xlsx')
jmlData = df['komentar'].count()

# Membagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(
    df['komentar'],
    df['sentimen'],
    test_size=0.2,
)

data_uji = pd.DataFrame({
    'Komentar Hasil Preprocessing': X_test,
    'Sentimen': y_test
})

jmlDataUji = len(X_test)
jmlDataLatih = len(X_train)

with tab1:
    df = pd.read_excel('hasil_labelling.xlsx')
    jmlData = len(df['komentar'])

st.header('DATA SENTIMEN')

```



```

st.write(f'Data Ulasan Pengguna Aplikasi Dana: {jmlData} ulasan')

st.write(f'Jumlah Data Latih: {jmlDataLatih} ulasan')

st.write(f'Jumlah Data Uji: {jmlDataUji} ulasan')

st.dataframe(data_uji, use_container_width=True, height=400)

jml_sentimen = df.sentimen.value_counts()

# diagram batang

sn.set_style('whitegrid')

fig,ax = plt.subplots(figsize=(6,4))

ax = sn.barplot(
    x=jml_sentimen.index,
    y=jml_sentimen.values,
    palette='pastel',
    hue=jml_sentimen.index,
    dodge=False,
    legend=False
)

plt.title('Jumlah Sentimen Ulasan Pengguna Aplikasi dana',fontsize=14, pad=20)

plt.xlabel('Sentimen', fontsize=12)

plt.ylabel('Jumlah Ulasan', fontsize=12)

for i, count in enumerate(jml_sentimen.values):
    ax.text(i, count+0.10, str(count), ha='center', va='bottom')

st.pyplot(fig)

```

with tab2:

```
@st.cache_resource
```

```
def load_model_and_vectorizer():
```

```
    svm_model = joblib.load('model/svm_model.pkl')
```

```
    tfidf_vectorizer = joblib.load('model/tfidf_vectorizer.pkl')
```

```
    return svm_model, tfidf_vectorizer
```

```
svm_model, tfidf_vectorizer = load_model_and_vectorizer()
```

```
st.header('🗨 Analisis Sentimen Komentar')
```

```
with st.form(key='sentiment_analysis_form'):
```

```
    st.markdown("""
```

```
    <style>
```

```
    .big-font {
```

```
        font-size:20px !important;
```

```
        font-weight: bold;
```

```
    }
```

```
    </style>
```

```
    <p class='big-font'>Masukkan Komentar untuk Dianalisis</p>
```

```
    """, unsafe_allow_html=True)
```

```
    text = st.text_area('', height=100,
```

```
                        placeholder='Contoh: Aplikasi ini sangat membantu untuk transaksi  
sehari-hari...')
```

```
        help="Masukkan komentar yang ingin dianalisis  
sentimennya",label_visibility="hidden")
```

```
col1, col2 = st.columns([1, 6])
```

```
with col1:
```

```
    predict_button = st.form_submit_button('🔍 Analisis',  
use_container_width=True)
```

```
with col2:
```

```
    clear_button = st.form_submit_button('🔄 Reset', use_container_width=True)
```

```
if clear_button:
```

```
    text = ""
```

```
    st.rerun()
```

```
if predict_button:
```

```
    if text.strip() != "":
```

```
        with st.spinner('Menganalisis sentimen...'):
```

```
            try:
```

```
                def pre_proses(text):
```

```
                    text = clean_text_advanced(text)
```

```
                    text = to_lower(text)
```

```
                    text, _, _ = replace_taboo_word(text, kamus_tidak_baku)
```

```
                    text = tokenization(text)
```

```
                    text = remove_stopwords(text)
```

```
                    processed_text = stem_text(text)
```

```

        return processed_text

    processed_text = pre_proses(text) # Stemming teks

    # Transform input text
    text_vector = tfidf_vectorizer.transform([processed_text])

    # Predict
    prediction = svm_model.predict(text_vector)[0]
    probabilities = svm_model.predict_proba(text_vector)[0]
    save_analysis_history(text, prediction, probabilities)

    # Display results
    st.markdown("----")

    # Hasil Prediksi dengan styling
    result_container = st.container()

    with result_container:

        col1, col2 = st.columns([2, 3])

        with col1:

            st.markdown("### Hasil Analisis")

            if prediction == "positif":

                st.markdown(f"<h1 style='color: #0f5132; background-color: #d1e7dd; padding: 20px; border-radius: 10px; text-align: center;'>😊 POSITIF</h1>",
unsafe_allow_html=True)

            elif prediction == 'netral':

                st.markdown(f"<h1 style='color: #004280; background-color:rgb(102 169 230 / 24%); padding: 20px; border-radius: 10px; text-align: center;'>😐 NETRAL</h1>", unsafe_allow_html=True)

```

```

else:

    st.markdown(f"<h1 style='color: #842029; background-color: #f8d7da;
padding: 20px; border-radius: 10px; text-align: center;'>☹️ NEGATIF</h1>",
unsafe_allow_html=True)

```

```

with col2:

```

```

    st.markdown("### Probabilitas Sentimen")

```

```

    prob_df = pd.DataFrame({

        'Sentimen': ['Negatif', 'Netral', 'Positif'],

        'Probabilitas': probabilities

    })

```

```

fig = go.Figure(data=[

    go.Bar(

        x=prob_df['Sentimen'],

        y=prob_df['Probabilitas'],

        marker_color=['#dc3545', '#232b2b', '#198754']

    )

])

```

```

fig.update_layout(

    title="Distribusi Probabilitas",

    yaxis_title="Probabilitas",

    yaxis=dict(range=[0, 1])

)

```

```

st.plotly_chart(fig, use_container_width=True)

```

```

# Teks asli dan hasil preprocessing

with st.expander("Lihat Detail Preprocessing"):

    col1, col2 = st.columns(2)

    with col1:

        st.markdown("***Teks Asli:***")

        st.info(text)

    with col2:

        st.markdown("***Hasil Preprocessing:***")

        st.info(processed_text)


# Kata-kata yang berpengaruh

if hasattr(svm_model, 'coef_'):

    feature_names = tfidf_vectorizer.get_feature_names_out()

    coefficients = svm_model.coef_.toarray()[0]


top_negative = pd.DataFrame({

    'Kata': feature_names[coefficients.argsort()[::-5][::-1]],

    'Pengaruh': coefficients[coefficients.argsort()[::-5][::-1]]

})

top_positive = pd.DataFrame({

    'Kata': feature_names[coefficients.argsort()[:5]],

    'Pengaruh': coefficients[coefficients.argsort()[:5]]

})

```



```

with st.expander("Lihat Kata-kata yang Berpengaruh"):

    col1, col2 = st.columns(2)

    with col1:

        st.markdown("***Top Kata Positif***")

        st.dataframe(top_positive, use_container_width=True)

    with col2:

        st.markdown("***Top Kata Negatif***")

        st.dataframe(top_negative, use_container_width=True)

except Exception as e:

    st.error(f"Terjadi kesalahan dalam pemrosesan: {str(e)}")

else:

    st.warning('⚠ Mohon masukkan komentar terlebih dahulu')

display_analysis_history()

# Informasi tambahan

with st.expander("📖 Tentang Analisis Sentimen"):

    st.markdown("""

    ***Bagaimana Ini Bekerja?***

    1. Komentar yang Anda masukkan akan diproses melalui beberapa tahap preprocessing

    2. Teks yang sudah diproses akan diubah menjadi vektor menggunakan TF-IDF

    3. Model SVM yang sudah dilatih akan memprediksi sentimen dari vektor tersebut

    4. Hasil prediksi akan ditampilkan beserta probabilitasnya

    ***Tips:***

    - Masukkan komentar yang cukup panjang untuk hasil yang lebih akurat

```

- Analisis bisa mendeteksi sentimen positif dan negatif
- Hasil preprocessing bisa dilihat untuk transparansi analisis

""")



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
LEMBAGA PENELITIAN

Jl. Achmad Nadjamuddin No.17, Kampus Unisan Gorontalo Lt.I Kota Gorontalo 96128
Website: lemlitunisan.ac.id, Email: lembagapenelitian@unisan.ac.id

SURAT KETERANGAN

Nomor : 54/PIP/B.04/LP-UIG/2024

Yang bertandatangan di bawah ini:

Nama : Dr. Rahmisyari, ST., SE., MM
NIDN : 0929117202
Pangkat Akademik : Lektor Kepala
Jabatan : Ketua Lembaga Penelitian Universitas Ichsan Gorontalo

Menerangkan bahwa:

Nama : ANDRES FERNANDA RAHIM
NIM : T3119002
Fakultas : Ilmu Komputer
Program Studi : Teknik Informatika
Judul Penelitian : PENGGUNAAN TEKNIK ENSEMBLE LEARNING UNTUK ANALISIS SENTIMEN PADA APLIKASI DANA DI GOOGLE PLAY STORE

Adalah benar telah melakukan pengambilan data penelitian dalam rangka Penyusunan Proposal/Skripsi pada Universitas Ichsan Gorontalo.

Dikeluarkan di Gorontalo
Tanggal, 25/11/2024

Ketua Lembaga Penelitian

Dr. Rahmisyari, ST., SE., MM
NIDN: 0929117202



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
FAKULTAS ILMU KOMPUTER
UPT. PERPUSTAKAAN FAKULTAS
SK. MENDIKNAS RI NO. 84/D/0/2001

Jl. Achmad Nadjamuddin No.17 Telp(0435) 829975 Fax. (0435) 829976 Gorontalo

SURAT KETERANGAN BEBAS PUSTAKA

No : 002/Perpustakaan-Fikom/XI/2024

Perpustakaan Fakultas Ilmu Komputer (FIKOM) Universitas Ichsan Gorontalo dengan ini menerangkan bahwa :

Nama Anggota : Andres Fernanda Rahim
No. Induk : T3119002
No. Anggota : M202467

Terhitung mulai hari, tanggal : Sabtu, 16 November 2024, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di pergunakan sebagaimana mestinya.



Gorontalo, 16 November 2024

**Mengetahui,
Kepala Perpustakaan**

Apriyanto Alhamad, M.Kom

NIDN : 0924048601



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI

UNIVERSITAS ICHSAN GORONTALO

FAKULTAS ILMU KOMPUTER

SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001

Jl. Achmad Najamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT REKOMENDASI BEBAS PLAGIASI

No. 546/FIKOM-UIG/R/XII/2024

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom
NIDN : 0928028101
Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

Nama Mahasiswa : Andres Fernanda Rahim
NIM : T3119002
Program Studi : Teknik Informatika (S1)
Fakultas : Fakultas Ilmu Komputer
Judul Skripsi : Penggunaan Metode Support Vector Machine Untuk Analisis Sentimen Pada Ulasan Pengguna Aplikasi Dana Pada Google Play Store

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi **Turnitin** untuk judul skripsi di atas diperoleh hasil *Similarity* sebesar **16%**, berdasarkan Peraturan Rektor No. 32 Tahun 2019 tentang Pendeteksian Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ichsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan **BEBAS PLAGIASI** dan layak untuk diujikan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.

Mengetahui
Dekan,

Irvan Abraham Salihi, M.Kom
NIDN. 0928028101

Gorontalo, 10 Desember 2024

Tim Verifikasi


Zulfrianto M. Lamasigi, M.Kom
NIDN. 0914089101

Terlampir :
Hasil Pengecekan Turnitin

Fikom03 Unisan

PENGUNAAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI D...

-  FAKULTAS ILMU KOMPUTER
-  Fak. Ilmu Komputer
-  LL Dikti IX Turnitin Consortium

Document Details

Submission ID

trn:oid::1:3105762577

Submission Date

Dec 7, 2024, 2:17 PM GMT+7

Download Date

Dec 7, 2024, 2:18 PM GMT+7

File Name

SKRIPSI_Andres_Fernanda_Rahim_andresrahim199gmail.com.pdf

File Size

1.6 MB

86 Pages

14,883 Words

83,293 Characters

16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography
- Quoted Text

Top Sources

- 0%  Internet sources
- 6%  Publications
- 12%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

BIODATA MAHASISWA
UNIVERSITAS ICHSAN GORONTALO
JURUSAN TEKNIK INFORMATIKA

Nama : Andres Fernanda Rahim
Tempat/Tgl lahir : Gorontalo, 14 Juli 2002
Jenis Kelamin : Pria
Gol Darah : O
Alamat : Jl.Raja Bulango
RT/RW : 000/000
Kel/Desa : Desa Bandungan
Kec : Bulango Utara
Agama : Islam
Status Perkawinan : Belum Menikah
Pekerjaan : Mahasiswa
Kewarganegaraan : WNI
Riwayat Kependidikan :
TK : TK Citra
SD : SDN 6 Bulango Utara
SMP : SMPN 3 Telaga
SMA : SMKN 1 Bulango Utara
Perguruan Tinggi : Universitas Ichsan Gorontalo

