

**PENERAPAN METODE KLASIFIKASI SUPPORT VECTOR
MACHINE DALAM ANALISIS SENTIMEN ULASAN
PENGGUNA LAYANAN GORIDE PADA APLIKASI
GOJEK DI GOOGLE PLAY STORE**

SKRIPSI



NIM : 2113000046

NAMA : MUH. ADREANSYAH PRATAMA LUBIS

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS POTENSI UTAMA
MEDAN
2025**

KATA PENGANTAR



Assalamu‘alaikum Wr. Wb.

Alhamdulillahirobbil’alamin puji dan syukur penulis panjatkan kehadirat Allah SWT, karena atas rahmat dan hidayah-Nya penulis dapat menyelesaikan skripsi ini dengan judul **“Penerapan Metode Klasifikasi Support Vector Machine Dalam Analisis Sentimen Ulasan Pengguna Layanan Goride Pada Aplikasi Gojek Di Google Playstore”**. Shalawat beserta salam senantiasa tercurah kepada junjungan Nabi Besar Muhammad SAW, beserta keluarga dan sahabatnya yang telah menuntun islam kejalan yang benar.

Maka dengan selesainya penyusunan Skripsi ini, penulis mengucapkan terima kasih yang sebesar – besarnya untuk seluruh pihak yang membantu. Pada kesempatan ini penulis mengucapkan terima kasih kepada :

1. Ibu Ria Eka Sari, M.Kom, selaku Dosen Pembimbing I, yang telah memberikan waktu, arahan, saran, serta motivasi sehingga penulis dapat menyelesaikan Skripsi ini dengan baik.
2. Ibu Hj. Nuriandy, BA., selaku Pembina Yayasan Universitas Potensi Utama Medan.
3. Bapak Dr. Bob Subhan Riza, ST, M.Kom., selaku Ketua Yayasan

Universitas Potensi Utama Medan.

4. Ibu Prof. Dr. Rika Rosnelly, S.Kom., M.Kom., selaku Rektor Universitas Potensi Utama.
5. Ibu Dr. Lili Tanti, M.Kom, selaku Wakil Rektor I Universitas Potensi Utama.
6. Ibu Ratih Puspasari, M.Kom, Selaku Dekan Fakultas Teknik Dan Ilmu Komputer Universitas Potensi Utama.
7. Ibu Mas Ayoe Elhias Nasution, M.Kom, selaku Ketua Program Studi Sistem Informasi Universitas Potensi Utama.
8. Kepada Kedua orang tua saya tercinta, terutama kepada Ibu saya tercinta yang telah membimbing dan telah memberikan dorongan dan bantuan doa maupun material sehingga penulis dapat menyelesaikan penulisan skripsi ini.

Dalam penulisan skripsi ini penulis menyadari bahwa masih terdapat kekurangan, oleh karena itu penulis mengharapkan kritikan dan saran yang bersifat membangun untuk kesempurnaan skripsi ini agar lebih bermanfaat bagi penulis dan bagi kita semua.

Medan, 11 Februari 2025

Penulis,

Muhammad Adreansyah Pratama

DAFTAR ISI

KATA PENGANTAR	i
DAFTAR ISI	iii
DAFTAR GAMBAR	vii
DAFTAR TABEL	ix
BAB I PENDAHULUAN.....	1
I.1. Latar Belakang.....	1
I.2. Ruang Lingkup Permasalahan	4
I.2.1. Identifikasi Masalah	4
I.2.2. Perumusan Masalah	4
I.2.3. Batasan Masalah.....	5
I.3. Tujuan dan Manfaat Penelitian	6
I.3.1. Tujuan.....	6
I.3.2. Manfaat.....	6
I.4. Kontribusi Penelitian	7
I.5. Sistematika Penulisan	8
BAB II TINJAUAN PUSTAKA	10
II.1. Penelitian Terkait.....	10
II.2. Analisis Sentimen.....	12
II.3. Gojek	13
II.4. Text Mining	14
II.5. Cross Industri Standard Process - Data Mining (CRISP-DM).....	15
II.5.1. Business Understanding.....	16
II.5.2. Data Understanding	16

II.5.3. Data Preparation.....	16
II.5.4. Modelling.....	16
II.5.5. Evaluation	17
II.5.6. Deployment.....	17
II.6. Preprocessing Data	17
II.6.1. Casefolding	17
II.6.2. Tokenizing.....	18
II.6.3. Stopword.....	18
II.6.4. Stemming.....	18
II.7. TF-IDF.....	18
II.8. Support Vector Machine (SVM).....	20
II.8.1. Confusion Matrix	23
II.9. Python Programming.....	25
II.9.1. Library Python	25
II.10. Tools	26
II.10.1. Pycharm	26
II.10.2. Visual Studio Code	26
II.10.3. Streamlit.....	26
BAB III METODOLOGI PENELITIAN	28
III.1. Gambaran Umum Objek Penelitian	28
III.2. Desain Penelitian	28
III.2.1. Business Understanding	30
III.2.2. Data Understanding	30
III.2.3. Data Preparation	30
III.2.4. Modelling	32

III.2.5. Evaluation.....	33
III.2.6. Deployment	33
III.3. Metode Penelitian	33
III.3.1. Jenis Penelitian dan Metode yang digunakan.....	33
III.3.2. Jenis dan Sumber Data	34
III.4. Variabel Penelitian	34
III.5. Teknik Pengumpulan Data	35
III.5.1. Studi Literatur.....	35
III.5.2. Web Scrapping.....	36
III.6. Teknik Analisis Data	37
III.7. Rancangan Aplikasi	38
III.7.1. Rancangan Use Case Diagram	39
III.7.2. Rancangan User Interface.....	39
III.8. Bahan dan Alat Penelitian.....	42
III.8.1. Perangkat Keras (Hardware)	42
III.8.2. Perangkat Lunak (Software).....	42
BAB IV HASIL DAN PEMBAHASAN	44
IV.1. Implementasi Alur Kerja CRISP-DM	44
IV.1.1. Hasil Business Understanding	44
IV.1.2. Hasil Data Understanding	45
IV.1.3. Hasil Data Preparation	46
IV.1.4. Hasil Modelling	58
IV.1.5. Hasil Evaluation.....	61
IV.1.6. Hasil Deployment	64
IV.2. Pembahasan.....	68

IV.2.1. Uji Sistem Aplikasi SentimenGo	68
IV.2.2. Perbandingan Performa Model Algoritma SVM dengan Algoritma lain	70
IV.2.3. Visualisasi Diagram dan Wordcloud Sentiment Ulasan	71
BAB V KESIMPULAN DAN SARAN	74
V.1. Kesimpulan.....	74
V.2. Saran	75
DAFTAR PUSTAKA.....	76

DAFTAR GAMBAR

Gambar II. 1 Sentimen Analysis Methods	12
Gambar II. 2 Siklus Proses Metodologi CRISP-DM	15
Gambar II. 3 SVM Linier dan Hyperplane Optimal	20
Gambar II. 4 SVM Non-Linier Transformation.....	22
Gambar III. 1 Desain Penelitian.....	29
Gambar III. 2 Dataset Ulasan Pengguna Goride.....	36
Gambar III. 3 Use Case Diagram Aplikasi.....	39
Gambar III. 4 User Interface Login.....	39
Gambar III. 5 User Interface Registrasi	40
Gambar III. 6 Interface Forgot Password.....	40
Gambar III. 7 User Interface Dashboard Analisis Sentimen.....	41
Gambar III. 8 User Interface Analisis Text	41
Gambar III. 9 User Interface Prediksi Sentimen Text.....	42
Gambar IV. 1 Proses Scrapping Data Ulasan Pengguna Goride	45
Gambar IV. 2 Contoh Dataset Ulasan Pengguna Layanan Goride.....	46
Gambar IV. 3 Preprocessing Data – Case Folding	47
Gambar IV. 4 Preprocessing Data - Cleansing	48
Gambar IV. 5 Preprocessing Data – Normalisasi Slang.....	49
Gambar IV. 6 Preprocessing Data – Remove Repeated Characters	50
Gambar IV. 7 Preprocessing Data - Tokenisasi.....	51
Gambar IV. 8 Preprocessing Data – Stopword Removal	52
Gambar IV. 9 Preprocessing Data - Stemming	53
Gambar IV. 10 Pelabelan Dataset Dengan Model Pre-trained BERT	55
Gambar IV. 11 Proses Data Preparation – Feature Extraction	56
Gambar IV. 12 Proses Splitting Data	59
Gambar IV. 13 Proses Pengaturan Hyperparamter Tuning dengan GridSearchCV	
Gambar IV. 12 Proses Splitting Data.....	59
Gambar IV. 13 Proses Pengaturan Hyperparamter Tuning dengan GridSearchCV	60

Gambar IV. 14 Hasil Modelling Algoritma Support Vector Machine	61
Gambar IV. 15 Visual Confution Matrix Algoritma SVM	62
Gambar IV. 16 Tampilan Halaman Daftar AkunGambar IV. 15 Visual Confution Matrix Algoritma SVM	62
Gambar IV. 16 Tampilan Halaman Daftar Akun	65
Gambar IV. 17 Tampilan Halaman Masuk Akun	65
Gambar IV. 18 Tampilan Halaman Reset Kata Sandi Akun	66
Gambar IV. 19 Tampilan Halaman Dashboard Ringkasan	67
Gambar IV. 20 Tampilan Halaman Analisis Data	67
Gambar IV. 21 Halaman Prediksi Sentimen	68
Gambar IV. 22 ROC-AUC Perbandingan Performa Model Algoritma.....	71
Gambar IV. 23 Visualisasi Wordcloud Sentimen Positif dan Negatif	72
Gambar IV. 24 Visualisasi Diagram (Bigram) Frekuensi Kata Sentimen Negatif dan Positif	72

DAFTAR TABEL

Tabel II. 1 Tabel Confusion Matrix	24
Tabel II. 1 Tabel Confusion Matrix	24
Tabel IV. 1 Contoh Hasil Case Folding.....	47
Tabel IV. 2 Contoh Hasil Cleansing	48
Tabel IV. 3 Contoh Hasil Normalisasi Slang.....	49
Tabel IV. 4 Contoh Hasil Remove Repeated Characters.....	50
Tabel IV. 5 Contoh Hasil Tokenisasi	51
Tabel IV. 6 Contoh Hasil Stopword Removal	52
Tabel IV. 7 Contoh Hasil Stemming.....	54
Tabel IV. 8 Contoh Hasil Dataset Yang Telah Dipreprocessing	54
Tabel IV. 9 Contoh Hasil Pelabelan Dataset	56
Tabel IV. 10 Paramater Feature Extraction TF_IDF	57
Tabel IV. 11 Contoh Hasil Feature Extraction	58
Tabel IV. 12 Skema Pembagian Data	59
Tabel IV. 13 Hyperparamter dan Nilai Parameter Terbaik Model SVM.....	60
Tabel IV. 14 Confution Matrix Model SVM	62
Tabel IV. 15 Tabel Black Box Testing Case Aplikasi SentimenGo.....	69
Tabel IV. 16 Hasil Perbandingn Performa Model Algoritma	70

BAB I

PENDAHULUAN

I.1. Latar Belakang

PT. Aplikasi Karya Anak Bangsa atau Gojek Indonesia adalah perusahaan teknologi asal indonesia yang di dirikan pada tahun 2010 oleh Nadim Makarim, sebagai salah satu *super apps* yang mendorong inovasi dalam sektor transportasi dengan memunculkan aplikasi transportasi *online*. Gojek menawarkan berbagai layanan untuk memenuhi kebutuhan penggunanya seperti *GoRide*, *GoCar*, *GoSend*, *GoFood* dsb. Dibandingkan dengan layanan serupa dari perusahaan pesaing, Gojek masih menduduki peringkat pertama (Indra et.al 2020). Hal ini didukung berdasarkan survei *Institute for Development of Economics and Finance (INDEF)* pada tahun 2022, Gojek mendominasi pasar transportasi daring dengan tingkat penggunaan mencapai 82%, diikuti oleh Grab sebesar 53% (Katadata.co.id, 2022). Data ini menunjukkan preferensi masyarakat terhadap Gojek sebagai solusi mobilitas berbasis aplikasi yang telah menjadi integral dari kehidupan sehari-hari masyarakat modern. Sebagai salah satu layanan unggulan Gojek, *GoRide* menawarkan solusi transportasi roda dua yang cepat dan ekonomis untuk memenuhi kebutuhan mobilitas masyarakat perkotaan. Namun, dengan semakin tingginya jumlah pengguna, evaluasi terhadap kualitas layanan menjadi semakin penting.

Pada *Google Play Store* menunjukkan bahwa Gojek telah diunduh lebih dari 190 juta kali, dengan rating 4,6 dan lebih dari 6 juta ulasan. Setiap pengguna yang

mengunduh aplikasi Gojek memiliki kesempatan untuk memberikan ulasan melalui fitur ulasan yang tersedia. Ulasan ini dapat diakses secara terbuka dan dapat menjadi sumber informasi berharga jika dikelola dengan baik. Mencatat bahwa ulasan pengguna yang telah di input oleh pengguna aplikasi di *PlayStore* mencakup berbagai aspek, mulai dari ulasan positif hingga keluhan, kritik, atau saran (Watrianthos et.al 2019). Data yang dihasilkan dari ulasan ini, jika diolah dengan baik, dapat memberikan masukan yang konstruktif untuk perbaikan dan pengembangan aplikasi, karena kritik dan saran tersebut mencerminkan kebutuhan langsung dari pengguna, namun dengan jumlahnya yang besar dan sifatnya yang tidak terstruktur membuat analisis manual menjadi kurang efisien dan memakan waktu sehingga dapat mempengaruhi insight yang diperoleh.

Oleh karena itu, pendekatan berbasis teknologi seperti analisis sentimen dengan algoritma *Machine Learning* diperlukan untuk memahami opini pelanggan secara mendalam. Dengan teknologi ini, Gojek dapat mengidentifikasi area yang perlu ditingkatkan dan mengambil langkah strategis untuk mempertahankan keunggulan kompetitifnya di pasar terutama pada *GoRide* mengingat layanan tersebut merupakan layanan yang paling banyak digunakan pengguna aplikasi Gojek.

Analisis sentimen atau *opinion mining* adalah proses untuk mengidentifikasi dan mengkategorikan opini atau emosi pengguna terhadap suatu layanan, apakah positif, negatif, atau netral. Sumber data untuk analisis ini meliputi ulasan, diskusi forum, blog, mikroblog, hingga media sosial seperti *Twitter*. Sejumlah besar data berbasis opini ini disimpan dalam bentuk digital. Untuk topik tertentu atau opini

tertentu, analisis sentimen yang terkait dengan proses penambangan data bekerja untuk menghasilkan keluaran yang relevan (Mehta & Pandya, 2020).

Support Vector Machine (SVM) adalah salah satu metode klasifikasi dalam *Machine Learning* yang bekerja dengan memanfaatkan *Hyperplane* (garis pemisah) untuk memisahkan data dalam kelompok tertentu. Dalam penelitian ini, *SVM* dapat memudahkan penelitian ini dalam mengelompokkan ulasan pengguna layanan *Goride* ke dalam kategori positif atau negatif, sehingga mempermudah analisis sentimen secara akurat dan efisien.

Berdasarkan penelitian yang telah dilakukan (Alamsyah et al., 2024) dalam artikel yang berjudul ” Penerapan Algoritma *Support Vector Machine* Untuk Analisis Sentimen Data Ulasan Aplikasi Binance Pada Google Play Store ”, dapat disimpulkan hasil dari penelitian tersebut menunjukkan model *SVM* mampu mencapai akurasi sebesar 87,24% dalam klasifikasi sentimen ulasan pengguna aplikasi Binance. Selain itu, hasil evaluasi menunjukkan precision sebesar 85%, recall sebesar 87% dan F1-score sebesar 85%.

Berdasarkan literatur selanjutnya (Idris et al., 2023) yang berjudul ” Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Suport Vector Machine (*SVM*) ” menunjukkan bahwa metode *SVM* memiliki performa yang tinggi dalam klasifikasi sentimen. Hasil penelitian tersebut mengungkapkan bahwa algoritma *SVM* mampu mencapai tingkat akurasi sebesar 89%.

Menurut penelitian sebelumnya (Iskandar & Nataliani, 2021) telah melakukan perbandingan metode dalam melakukan analisis sentimen, penelitian tersebut berjudul ” Perbandingan *Naïve Bayes*, *SVM*, dan k-NN untuk Analisis

Sentimen Gadget Berbasis Aspek ” hasil penelitiannya menunjukkan bahwa keakuratan penerapan antara metode *Naïve Bayes*, *SVM*, dan *KNN* dalam analisis sentimen model klasifikasi *Support Vector Machine (SVM)* menunjukkan hasil terbaik dengan menghasilkan rata-rata akurasi sebesar 96,43%.

Berdasarkan latar belakang yang telah dijelaskan sebelumnya serta merujuk pada penelitian-penelitian terdahulu serta pedoman literatur terkait analisis sentimen, maka penulis tertarik untuk mengambil judul ” **Penerapan Metode Klasifikasi Support Vector Machine Dalam Analisis Sentimen Ulasan Pengguna Layanan Goride Pada Aplikasi Gojek Di Google Play Store ”.**

I.2. Ruang Lingkup Permasalahan

I.2.1. Identifikasi Masalah

Adapun identifikasi masalah dalam peneltian ini adalah :

1. Tidak adanya sistem berbasis machine learning yang dapat secara otomatis menganalisis sentimen ulasan pengguna untuk memahami opini pelanggan secara mendalam.
2. Ulasan pengguna layanan *GoRide* pada Google Playstore yang sangat banyak dan tidak terstruktur menyebabkan kesulitan dalam menganalisisnya secara manual.
3. Belum ada langkah strategis yang menggunakan analisis untuk meningkatkan kualitas layanan *GoRide* yang diambil berdasarkan hasil analisis sentimen.

I.2.2. Perumusan Masalah

Berdasarkan identifikasi masalah tersebut, maka dapat dirumuskan permasalahan melalui penulisan laporan penelitian ini yaitu :

1. Bagaimana metode klasifikasi *Support Vector Machine (SVM)* yang diterapkan dalam sistem berbasis analisis sentimen dapat memprediksi sentimen terhadap ulasan pengguna layanan *Goride* pada aplikasi Gojek di Google Play Store sebagai sentimen positif atau negatif?
2. Bagaimana cara mengelolah data ulasan pengguna layanan *GoRide* menjadi terstruktur agar dapat digunakan dalam analisis sentimen?
3. Bagaimana hasil analisis sentimen dapat digunakan untuk memberikan wawasan strategis yang mendukung peningkatan kualitas layanan GoRide?

I.2.3. Batasan Masalah

Adapun yang menjadi batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini menggunakan metode klasifikasi *Support Vector Machine (SVM)* untuk analisis sentimen.
2. Data yang digunakan adalah ulasan pengguna *GoRide* yang diperoleh dari Google Play Store yang disimpan dalam format .csv.
3. Proses analisis mencakup pengumpulan data, preprocessing, ekstraksi fitur, pelatihan model dan evaluasi performa model.
4. Output penelitian berupa hasil klasifikasi sentimen dalam kategori positif dan negatif yang divisualisasikan menggunakan *Streamlit* berbasis web.
5. Penelitian ini hanya berfokus pada ulasan pengguna layanan *GoRide* dan tidak mencakup layanan lain dalam aplikasi Gojek.
6. Data yang digunakan adalah ulasan dalam bahasa indonesia dengan kata kunci terkait *GoRide*.

7. Model klasifikasi dievaluasi menggunakan metrik akurasi seperti presisi, recall dan *FI-Score*.
8. Penelitian ini tidak membahas integrasi sistem ke dalam aplikasi Gojek secara langsung.

I.3. Tujuan dan Manfaat Penelitian

I.3.1. Tujuan

Adapun yang menjadi tujuan dalam penelitian ini adalah :

1. Menerapkan metode klasifikasi *Support Vector Machine (SVM)* dalam analisis sentimen untuk memprediksi sentimen ulasan pengguna layanan *GoRide* pada aplikasi Gojek di Google Play Store ke dalam kategori positif, negatif atau netral.
2. Dapat menyajikan sistem berbasis sentimen menggunakan framework *Streamlit* untuk memvisualisasikan hasil analisis secara interaktif dan informatif sehingga memudahkan interpretasi data serta memberikan insight kepada Gojek Indonesia terhadap layanan Goride.

I.3.2. Manfaat

Adapun yang menjadi manfaat dalam penelitian ini adalah :

1. Memberikan wawasan strategis kepada Gojek dalam memahami umpan balik pengguna layanan *GoRide* sebagai bahan pertimbangan untuk meningkatkan kualitas layanan tersebut.
2. Menyediakan referensi bagi perusahaan dalam mengembangkan inovasi teknologi berbasis machine learning dan berbasis data dalam meningkatkan layanan transportasi daring.

3. Menghasilkan sistem yang memberikan visualisasi interaktif melalui framework *Streamlit* sehingga mempermudah pemangku kepentingan dalam memahami informasi dari hasil analisis.

I.4. Kontribusi Penelitian

Berdasarkan penelitian yang dilakukan oleh (Dwianto & Sadikin, 2021) yang berjudul "**Analisis sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine**", metode *SVM* memberikan hasil yang lebih baik dibandingkan *Naïve Bayes* dalam menganalisis sentimen pada data dari media sosial Twitter. Untuk layanan Gojek, *SVM* menghasilkan akurasi sebesar 69,50% dengan precision positif 73,90% dan recall positif 89,84%. Penelitian ini menggarisbawahi pentingnya analisis sentimen untuk mengevaluasi tingkat kepuasan pengguna terhadap layanan transportasi daring dan mengoptimalkan strategi pelayanan berdasarkan data.

Berdasarkan penelitian yang dilakukan oleh (Idris et al., 2023) yang berjudul "**Analisis Sentimen Terhadap Pengguna Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM)**", metode *SVM* menunjukkan performa yang sangat baik dalam klasifikasi sentimen. Penelitian tersebut berhasil mencapai tingkat akurasi sebesar 98% dengan F1-score sebesar 0,98. Penelitian ini menggunakan data ulasan aplikasi Shopee yang dikumpulkan melalui metode scrapping. Hasil tersebut menunjukkan bahwa metode *SVM* dapat memberikan hasil yang sangat akurat dalam memisahkan

sentimen positif dan negatif, sekaligus membuktikan keunggulan algoritma *SVM* dibandingkan metode lainnya dalam analisis sentimen berbasis teks.

Pada penelitian yang penulis lakukan saat ini yang berjudul ”**Penerapan Metode Klasifikasi Support Vector Machine Dalam Analisis Sentimen Ulasan Pengguna Layanan Goride Pada Aplikasi Gojek Di Google Playstore**” berfokus pada penerapan algoritma *Support Vector Machine (SVM)* dalam menganalisis sentimen ulasan pengguna layanan GoRide. Berbeda dari penelitian sebelumnya, penelitian ini menerapkan preprocessing data (casefolding, tokenizing, stopword removal, stemming) dan metode TF-IDF untuk meningkatkan kualitas data dan fitur. Hasil analisis divisualisasikan secara interaktif berbasis web menggunakan framework *Streamlit*, memudahkan interpretasi oleh pemangku kepentingan. Dengan evaluasi akurasi, presisi, recall, dan F1-score, penelitian ini diharapkan memberikan insight strategis berbasis data untuk meningkatkan kualitas layanan *GoRide* sekaligus memperluas penerapan teknologi analisis sentimen berbasis machine learning.

I.5. Sistematika Penulisan

Penelitian ini disusun berdasarkan sistematika penulisan yang telah ditetapkan. Adapun sistematika penulisan yang digunakan oleh penulis dalam penelitian ini sebagai berikut:

BAB I : PENDAHULUAN

Pada bab ini berisi latar belakang, identifikasi permasalahan, tujuan dan manfaat serta kontribusi penelitian. Bab ini memberikan gambaran mengenai isu yang diteliti serta ruang lingkup penelitian yang dilakukan.

BAB II : TINJAUAN PUSTAKA

Pada bab ini membahas teori-teori yang mendukung penelitian, termasuk konsep analisis sentimen, metode klasifikasi, serta penelitian terdahulu yang relevan dengan penelitian saat ini.

BAB III : METODOLOGI PENELITIAN

Pada bab ini menjelaskan metode yang digunakan dalam penelitian, mencakup teknik pengumpulan data, proses pengelolahan data dan algoritma klasifikasi yang ditetapkan. Bab ini juga menguraikan tahapan penelitian agar dapat dilakukan secara sistematis.

BAB IV : HASIL DAN PEMBAHASAN

Pada bab ini menyajikan hasil penelitian serta analisis terhadap data yang diolah. Bab ini juga membahas evaluasi performa model dan interpretasi hasil yang diperoleh dalam penelitian.

BAB V : KESIMPULAN DAN SARAN

Pada bab ini diuraikan sebuah kesimpulan mengenai temuan utama dalam penelitian dan menjawab tujuan penelitian yang telah rumuskan. Bab ini memberikan saran untuk penelitian selanjutnya untuk pengembangan lebih lanjut.

BAB II

TINJAUAN PUSTAKA

II.1. Penelitian Terkait

Adapun penelitian terkait yang akan digunakan sebagai sumber acuan yang relevan dan terkini yaitu oleh :

Berdasarkan Penelitian yang dilakukan (Iskandar & Nataliani, 2021) yang berjudul ” **Perbandingan Naive Bayes, SVM dan K-NN untuk Analisis Sentimen Gadget Berbasis Aspek**”, ditemukan bahwa algoritma *Support Vector Machine (SVM)* memberikan hasil terbaik dengan rata-rata akurasi sebesar 96,43% dibandingkan dengan *Naive Bayes* (83%54) dan *K-Nearest Neighbor* (59,68%). Selain itu dalam penelitian ini menunjukkan bahwa aspek desain cenderung menghasilkan sentimen positif sedangkan aspek harga, spesifikasi dan citra merek cenderung memiliki sentimen negatif.

Berdasarkan penelitian yang telah dilakukan (Alamsyah et al., 2024) dalam artikel yang berjudul ” **Penerapan Algoritma Support Vector Machine Untuk Analisis Sentimen Data Ulasan Aplikasi Binance Pada Google Play Store** ”, dapat disimpulkan hasil dari penelitian tersebut menunjukkan model *SVM* mampu mencapai akurasi sebesar 87,24% dalam klasifikasi sentimen ulasan pengguna aplikasi Binance. Selain itu, hasil evaluasi menunjukkan precision sebesar 85%, recall sebesar 87% dan F1-score sebesar 85%.

Berdasarkan penelitian yang dilakukan (Hakim et al., 2023) yang berjudul "**Analisis Perbandingan Algoritma Machine Learning Terhadap Sentimen Analisis Pemindahan Ibu Kota Negara**", hasil penelitian tersebut menunjukkan bahwa algoritma *Support Vector Machine (SVM)* memberikan performa terbaik dibandingan *Naive Bayes* dan *Random Forest* dalam menganalisis sentimen data dari Twitter. Dengan menambahkan teknik *Synthetic Minority Oversampling Technique (SMOTE)* untuk menangani ketidakseimbangan data, *SVM* mencapai akurasi tertinggi sebesar 82,82%.

Berdasarkan penelitian yang dilakukan (Utami et al., 2023) yang berjudul "**Algoritme Support Vector Mechine untuk Analisis Sentimen Berbasis Aspek Ulasan Game Online Mobile Legends : Bang-Bang**", ditemukan bahwa algoritma *Support Vector Machine (SVM)* mampu mengklasifikasikan sentimen ulasan pengguna aplikasi Mobile Legends secara efektif dengan nilai akurasi diatas 80% untuk semua aspek, termasuk gameplay, performa, visualisasi dan player. Selain itu , penelitian ini menunjukkan bahwa aspek gameplay, performa dan player cenderung memiliki sentimen negatif, sementara aspek visualisasi lebih dominan dengan sentimen positif dengan menunjukkan kekuatan visual aplikasi.

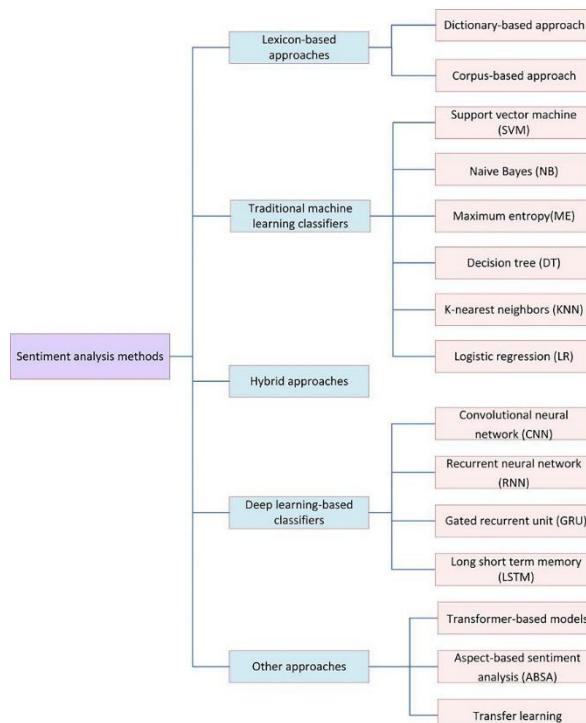
Berdasarkan penelitian yang dilakukan oleh (Reza Pahlawan et al., 2024) yang berjudul "**Review Komprehensif Penggunaan Data Imbalance Pada Metode Klasifikasi Dalam Machine Learning**", di temukan bahwa data tidak seimbang (imbalance data) menjadi salah satu tantangan utama dalam menggunakan metode klasifikasi, termasuk *Support Vector Machine (SVM)*. Pada penelitian tersebut menunjukkan hasil bahwa *SVM* memiliki performa tertinggi

dalam pengelolahan teks meskipun dalam penerapannya terhadap data yang tidak seimbang, *Support Vector Machine* memiliki persentase penggunaan sebesar 25% dibandingkan metode lain seperti *KNN* dan *Random Forest*. Ketangguhan dan Fleksibilitas *SVM* membuat metode tersebut cocok untuk berbagai tugas klasifikasi teks, termasuk analisis sentimen, kategorisasi topik dan klasifikasi dokumen.

II.2. Analisis Sentimen

Analisis sentimen merupakan suatu proses terstruktur yang bertujuan untuk mengevaluasi opini, emosi atau sikap yang terkandung dalam teks. Proses ini berfungsi untuk menganalisis persepsi individu terhadap suatu subjek yang dapat dikategorikan ke dalam sentimen positif, negatif atau netral (Mao et al., 2024).

Berikut gambar diagram struktur Analisis Sentimen yang terlampir dalam gambar 2.1.



Gambar II. 1 Sentimen Analysis Methods
 (Sumber : Yanying Mao, et.al 2024)

Analisis sentimen diambil berdasarkan data tekstual dari sumber terbuka yang dapat diakses secara umum seperti aplikasi, sosial media ataupun website dimana pengguna dapat mengekspresikan pendapat mereka kedalam sebuah teks (Mehta & Pandya, 2020). Analisis sentimen adalah ulasan yang memiliki arti positif, negatif atau netral yang kemudian dilakukan klasifikasi sebagai proses pengidentifikasian data dalam bentuk teks sehingga memperoleh hasil prediksi teks tersebut dapat dikategorikan kedalam sebuah sentimen.

Dengan hal itu, Tujuan analisis sentimen yaitu untuk mengekstraksi wawasan yang memiliki makna dari sejumlah data besar yang ada dalam dokumen atau teks terkait opini pengguna untuk memahami sentimen publik terkait sebuah produk, layanan dsb. Dengan maksud untuk mengklasifikasi tanggapan positif, negatif atau netral (Dashtipour, K, et.al 2021). Analisis sentimen menjadi salah satu topik penelitian yang menggunakan *Natural Language Processing (NLP)* dan merupakan salah satu bidang yang banyak diteliti secara aktif saat ini. Selain berperan penting dalam *NLP*, analisis sentimen terus menunjukkan kemajuan pesat dengan penambahan berbagai metode yang dikembangkan untuk mendukung penerapannya (Haviana & Petro, 2022).

II.3. Gojek

PT Aplikasi Karya Anak Bangsa atau Gojek Indonesia adalah perusahaan asal indonesia yang didirikan oleh Nadiem Makarim pada tahun 2010. Pada awal didirikan Gojek hanya memiliki 20 pengemudi yang beroperasi melalui *call center* untuk menghubungin penumpang dengan ojek. Seiring waktu, Gojek meluncurkan aplikasi seluler untuk perangkat Android dan iOS dengan memfasilitasi pemesanan

layanan yang lebih mudah. Gojek telah memperluas operasionalnya ke lebih 50 kota di Indonesia (Harimurti & Sofyan, 2022).

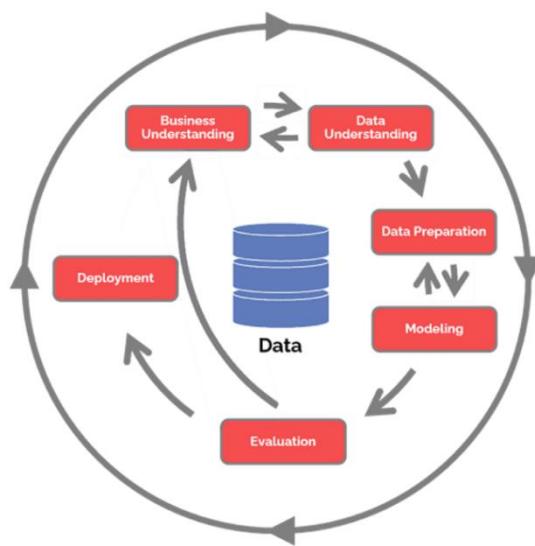
Gojek saat ini telah menjadi super apps yang menyediakan berbagai layanan untuk memenuhi kebutuhan penggunanya sehari-hari. Gojek menyediakan lebih dari 20 fitur atau layanan, yaitu *GoRide*, *GoCar*, *GoBlueBird*, *GoSend*, *GoBox*, *GoFood*, *GoMart*, *GoPay*, *Gopay Pinjam*, *Gopay Later*, *GoBills*, *GoPulsa*, *GoGive*, *GoGames*, *GoPlay*, *GoTix*, *GoMessage*, *GoClean*, *GoAuto*, *GoGlam*, *GoLaundry*, *GoDaily*, *Gofix* dan *GoMed*. Dengan keunggulan layanan Gojek tersebut pengguna mendapatkan kemudahan sehingga dapat memesan layanan Gojek kapan saja dimana saja hanya melalui aplikasi dan akses internet, karena pengemudi akan datang menjemput dan mengantar pelanggan langsung ke tempat tujuan (Indra et al., 2020).

II.4. Text Mining

Text mining ialah proses ekstraksi informasi berharga dari berbagai sumber data tekstual seperti dokumen CSV, PDF, kutipan teks dsb. Sebagai pembeda text mining dan data mining yakni teks mining merujuk pada teknologi yang mampu menganalisis data teks yang bersifat semi-struktural maupun tidak terstruktur sedangkan data mining berfokus pada pengolahan data terstruktur (Rokhman et al., 2021). Sehingga pada text mining proses ini berfokus pada pengelolahan dokumen atau teks dalam jumlah besar untuk mengidentifikasi pola dan hubungan didalamnya. Dengan menggunakan teknik klasifikasi dan kategorisasi, text mining bertujuan untuk menemukan informasi yang bermanfaat sesuai dengan tujuan tertentu (Joergensen E Munthe et al., 2022).

II.5. Cross Industri Standard Process - Data Mining (CRISP-DM)

CRISP-DM (Cross Industry Standard Process – Data Mining) merupakan metodologi standar industri yang dirancang untuk memandu proses eksplorasi, analisis dan ekstraksi pengetahuan dari data secara sistematis untuk memperoleh wawasan yang bermakna serta mendukung pengambilan keputusan (Hakim et al., 2023). Model proses data mining ini pertama kali dikembangkan pada tahun 1996 oleh lima perusahaan yaitu Integral Solutions Ltd, Teradata, Daimler AG, NCR Corporatio, dan OHRA. Framework ini terus disempurnakan oleh ratusan organisasi dan perusahaan di Eropa hingga menjadi metodologi standar non-proprietary dalam bidang data minig (Adhi Putra, 2021). Berikut Siklus proses dari metodologi CRISP-DM dapat dilihat pada Gambar II.2.



Gambar II. 2 Siklus Proses Metodologi CRISP-DM
(Sumber : Z.S. Sudar et al., 2024)

II.5.1. Business Understanding

Fase pemahaman bisnis bertujuan untuk mengidentifikasi tujuan utama dari proyek data mining dengan memahami kebutuhan bisnis dan permasalahan yang dihadapi. Pada tahap ini dilakukan analisis terhadap konteks bisnis, penetapan sasaran analisis serta definisi kriteria keberhasilan proyek (Hakim et al., 2023).

II.5.2. Data Understanding

Fase pemahaman data mencakup proses pengumpulan data dari berbagai sumber yang relevan, eksplorasi awal untuk memahami struktur dan karakteristik data, serta analisis kualitas data (Hakim et al., 2023).

II.5.3. Data Preparation

Fase persiapan data, yang juga dikenal sebagai data munging merupakan tahap dimana data yang relevan dikumpulkan dan diolah agar siap untuk diproses data mining. Proses ini mencakup berbagai langkah seperti pra-pemrosesan data, pemilihan tabel, pembuatan fitur, transformasi atribut, klasifikasi, normalisasi, perhilangan noise, serta pengambilan sampel. Keberhasilan tahap ini sangat bergantung pada kualitas serta karakteristik data mentah yang digunakan, Bahkan sekitar 50-80% dari keseluruhan usaha dalam proyek data mining umumnya dihabiskan pada tahap ini.

II.5.4. Modelling

Fase pemodelan bertujuan untuk menerapkan berbagai teknik analisis data dan machine learning guna membangun model yang mampu memprediksi atau mengklasifikasi data sesuai dengan tujuan bisnis. Dalam tahap ini, dilakukan pemilihan algoritma yang sesuai pelatihan model menggunakan dataset yang telah

dipersiapkan serta penyesuaian parameter untuk meningkatkan performa model (Hakim et al., 2023).

II.5.5. Evaluation

Fase evaluasi bertujuan untuk menilai performa model berdasarkan metrik yang telah ditentukan seperti akurasi, presisi, recall dan nilai F1-score tergantung pada jenis permasalahan yang dianalisis (Hakim et al., 2023).

II.5.6. Deployment

Fase *Deployment* adalah tahapan akhir dalam proses *CRISP-DM* setelah evaluasi model, dimana model yang terpilih diimplementasikan ke dalam sistem operasional. Implementasi ini dapat berupa visualisasi model dalam bentuk dashboard atau otomatisasi proses bisnis berbasis data, Tujuan utama dari tahapan *deployment* ini adalah untuk membuat visualisasi data (Pramudita et al., 2024).

II.6. Preprocessing Data

Preprocessing adalah proses menyusun kembali dan membersihkan data yang sudah ada sehingga terlihat lebih rapi. Tujuan dari *preprocessing* ini adalah untuk menghasilkan data ulasan yang lebih ringkas dan lebih mudah untuk dilanjutkan ke tahap berikutnya (Alamsyah et al., 2024).

Berikut adalah beberapa tahapan dari preprocessing :

II.6.1. Casefolding

Casefolding adalah proses mengubah semua huruf teks menjadi huruf kecil atau besar secara bersamaan, sehingga dapat menyederhanakan teks dan membuat

komputer lebih mudah memprosesnya. Misalnya kata ” Saya Sangat Semangat Hari Ini” menjadi ” saya sangat semangat hari ini” (Rizki et al., 2021).

II.6.2. Tokenizing

Tokenizing adalah proses membagi teks menjadi unit-unit yang lebih kecil, yang disebut token. Token ini dapat berupa kata, frasa atau bahkan karakter tunggal. Misalnya ” Kucing itu makan ikan ” akan dibagi menjadi token -token ” Kucing”, ”itu”, ”sangat”, ”lucu” (Rizki et al., 2021).

II.6.3. Stopword

Stopword sebuah kata-kata yang sangat umum dalam bahasa dan tidak memberikan banyak informasi yang berarti dalam *analisis text*, jika kata-kata tersebut dipertahankan dapat menyebabkan *noise* dalam data sehingga diperlukannya tahap *stopword* dalam *preprocessing*. Contoh *Stopword* dalam bahasa indonesia yaitu ”dan”, ”atau”, ”yang”, ”pada” dsb (Rizki et al., 2021).

II.6.4. Stemming

Stemming merupakan proses mereduksi kata menjadi bentuk dasarnya atau root word. Tujuannya adalah untuk mengurangi variasi bentuk kata yang sama. Misalnya kata ”berlari”, ”lari”, dan ”pelari” kata tersebut akan direduksi menjadi ”lari” (Maulana et al., 2024).

II.7. TF-IDF

TF-IDF (Term Frecuency – Invers Document Frecuency) merupakan teknik yang sering digunakan untuk memberikan bobot pada fitur dan menunjukkan tingkat akurasi serta efektivitas yang tinggi. *TF-IDF* menilai pentingnya suatu kata

dalam konteks sebuah dokumen atau kumpulan dokumen sehingga semakin sering sebuah kata muncul dalam satu dokumen, nilai kontribusinya akan tinggi. Sebaliknya, jika kata tersebut sering muncul di banyak dokumen maka kontribusinya akan berkurang. *TF-IDF* dapat dirumuskan :

a. Rumus *Term Frequency (TF)*

TF mengukur seberapa sering suatu kata muncul dalam sebuah dokumen.

$$TF_{t,d} = \frac{f_{t,d}}{\sum_{t'} f_{t',d}}$$

Keterangan :

$TF_{t,d}$ = Frekuensi kata t dalam dokumen d

$f_{t,d}$ = Jumlah kemunculan kata t dalam dokumen d

$\sum_{t'} f_{t',d}$ = Total jumlah kata dalam dokumen d

b. Rumus Invers Document Frecuency (IDF)

Invers Document Frequency (IDF) menghitung seberapa penting suatu kata dengan mempertimbangkan semua dokumen dalam koleksi.

$$IDF_t = \log \frac{N}{n_t}$$

Keterangan :

IDF_t = *Inverse Document Frequency untuk kata t*

N = *Total dokumen dalam koleksi*

n_t = *Jumlah dokumen yang mengandung kata t*

c. Rumus Term Frequency - Invers Document Frecuency (TF-IDF)

TF-IDF adalah gabungan dari TF dan IDF untuk memberikan bobot akhir kata.

$$W_{t,d} = TF_{t,d} \cdot IDF_t$$

Keterangan :

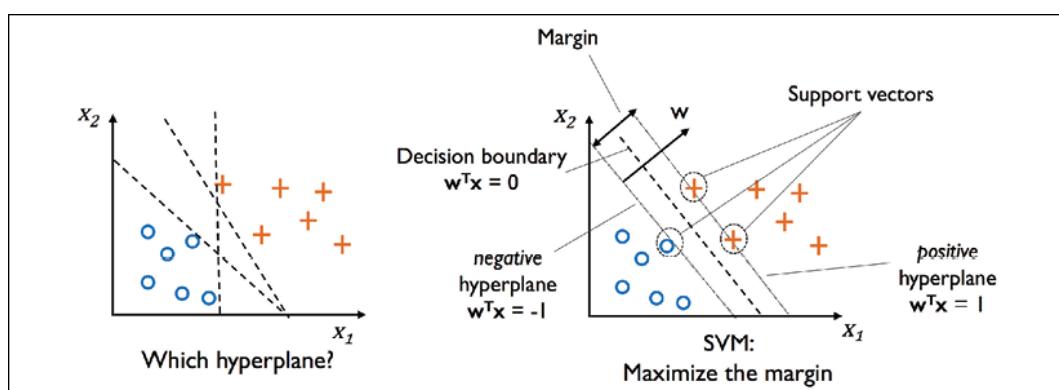
$W_{t,d}$ = Bobot TF-IDF untuk kata t dalam dokumen d

$TF_{t,d}$ = Term Frequency untuk kata t dalam dokumen d

IDF_t = Inverse Document Frequency untuk kata t

II.8. Support Vector Machine (SVM)

Support vector machine (SVM) pertama kali diperkenalkan oleh Vapnik, Guyon, dan Boser pada tahun 1992 sebagai metode klasifikasi yang dirancang untuk menemukan hyperplane terbaik sebagai pemisah antara dua kelas. Metode ini bekerja dengan memetakan data ke dalam ruang fitur berdimensi tinggi menggunakan pendekatan linier maupun non-linier. Pada setiap data direpresentasikan sebagai vektor n-dimensi yang dikelompokkan ke salah satu dari dua kelas. SVM memilih hyperplane dengan margin terbesar untuk memastikan antar kelas yang optimal, sehingga memberikan hasil klasifikasi yang lebih akurat (Rizki et al., 2021).



Gambar II. 3 SVM Linier dan Hyperplane Optimal
(Sumber: Raschka & Marjalili, 2019)

Hyperplane optimal dalam support vector machine (SVM) merupakan decision boundary terbaik yang memisahkan data dari dua kelas (+1 dan -1) dengan margin maksimum. Tujuan utama SVM untuk menemukan hyperplane yang tidak hanya memisahkan kedua kelas, tetapi juga memberikan jarak terbesar (margin) antara data dari kedua kelas yang paling dekat dengan hyperplane, yang disebut sebagai support vectors. Hyperplane secara umum didefinisikan oleh persamaan berikut.

$$w \cdot x + b = 0$$

Persamaan $w \cdot x + b = 0$ menggambarkan posisi geometris hyperplane dalam ruang fitur. Dalam persamaan ini, w adalah vektor bobot yang menentukan orientasi hyperplane, x adalah vektor fitur yang merepresentasikan data input, dan b adalah baris yang menggeser hyperplane dari titik asal. Nilai pada $w \cdot x + b$ menentukan posisi relatif data x terhadap hyperplane.

- Jika $w \cdot x + b > 0$, maka data x berada di sisi positif hyperplane.
- Jika $w \cdot x + b < 0$, maka data x berada di sisi negatif hyperplane.
- Jika $w \cdot x + b = 0$, maka data x berada tepat di hyperplane.

Sehingga pada kasus klasifikasi dengan SVM pada dataset linier data dibagi menjadi dua kelas sehingga margin dapat dimaksimalkan sedemikian rupa sehingga dua hyperplane untuk kelas positif dan negatif sejajar satu sama lain dengan aturan posisi secara konseptual sebagai berikut.

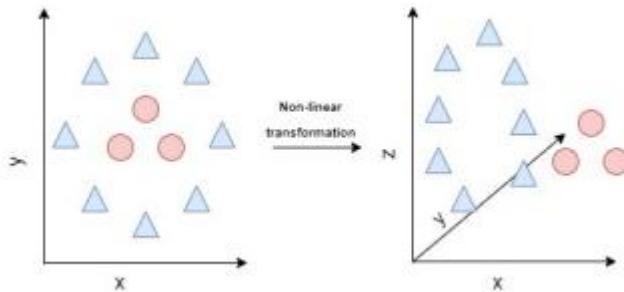
$$w \cdot x_i + b \geq 1$$

$$w \cdot x_i + b \leq -1$$

Subjek pada constraint:

$$y_i(w \cdot x_i + b) \geq 1, \forall i$$

Dimana y_i adalah label kelas (+1 untuk kelas positif dan -1 untuk kelas negatif) sedangkan x_i adalah data input.



Gambar II. 4 SVM Non-Linier Transformation
(Sumber: Mohri et al., 2018)

Pada kasus data yang tidak dapat dipisahkan secara linier (non-linier), Support Vector Machine (SVM) menggunakan soft margin yang memperbolehkan beberapa data melanggar margin untuk memperoleh hyperplane optimal sebagai berikut:

$$\min_{w,b,\gamma} \frac{1}{2} |w|^2 + C \sum_{i=1}^l \gamma_i$$

Subjek pada constraint :

$$y_i(w \cdot x_i + b) \geq 1 - \gamma_i, \quad \gamma_i \geq 0, \quad \forall i = 1, 2, \dots, l$$

Parameter C adalah nilai regulasi yang mengontrol trade-off antara margin maksimum dan penalti untuk data yang melanggar margin dalam Support Vector Machine (SVM). Dengan fungsi soft margin, model tetap dapat bekerja meskipun data tidak sepenuhnya terpisah secara linier, memungkinkan fleksibilitas dalam menangani dataset yang memiliki noise atau tumpang tindih antar kelas. Untuk data non-linier, SVM menggunakan kernel trick untuk memetakan data ke ruang

dimensi lebih tinggi tanpa harus memetakan data secara eksplisit. Berikut fungsi yang digunakan mengatasi masalah dataset non-linier.

$$K(x_i, x_j) = \phi(x_i) \cdot \phi(x_j)$$

Berikut Kernel yang sering digunakan Support Vector Machine dalam klasifikasi data non-linier :

1. Linear Kernel

$$K(x_i, x_j) = x_i \cdot x_j$$

2. Polynomial Kernel

$$K(x_i, x_j) = (x_i \cdot x_j + c)^d$$

3. RBF (Gaussian) Kernel

$$K(x_i, x_j) = (x_i \cdot x_j + c)^d$$

4. Sigmoid Kernel

$$K(x_i, x_j) = \tanh(x_i \cdot x_j + c)$$

II.8.1. Confusion Matrix

Confusion Matrix ialah sebuah tabel yang digunakan sebagai alat ukur untuk mengevaluasi kinerja model klasifikasi dalam *machine learning*. Tabel tersebut membandingkan prediksi yang dihasilkan oleh model yang dikelolah sistem dengan nilai aktual dari data. Dalam melakukan klasifikasi pada data, *confusion matrix* membantu dalam memahami bagaimana model membuat kesalahan dan seberapa baik model tersebut dalam mengenali pola-pola dalam data. Berikut merupakan contoh dari tabel confusion matrix (Pranata et al., 2022).

Tabel II. 1 Tabel Confusion Matrix

Klasifikasi	Prediksi Klasifikasi	
	Positive	Negatif
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

Keterangan :

1. *True Positive (TP)*, Total data yang sebenarnya positive dan diklasifikasikan sebagai positif oleh model.
2. *False Positive (FP)*, Total data yang sebenarnya negatif tetapi diklasifikasikan positif oleh model.
3. *True Negative (TN)*, Total data yang sebenarnya negatif dan diklasifikasikan sebagai negatif oleh model.
4. *False Negative (FN)*, Total data yang sebenarnya positif, tetapi diklasifikasikan data positif sebagai negative.

Berdasarkan penjelasan confusion matrik, terdapat beberapa performance matriks yang dapat digunakan dalam mengevaluasi model sehingga dapat mengukur efektivitas model yang diimplementasikan tersebut, diantaranya :

- a. *Accuracy* : perbandingan dari total prediksi yang benar dibandingkan dengan total prediksi yang dilakukan.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$

- b. *Precision* : rasio dari prediksi positif yang benar dibandingkan dengan total prediksi positif yang dilakukan model.

$$\text{Precision} = \frac{TP}{TP+FP}$$

- c. Recall (Sensitivity/True Positive Rate : Rasio antara total prediksi positif yang akurat dengan total jumlah kasus positif yang sebenarnya yang diprediksi oleh model.

$$\text{Recall} = \frac{TP}{TP+FN}$$

- d. *F1-Score* : rata-rata harmonic dari precision dan recall.

$$F1\text{-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

II.9. Python Programming

Python merupakan salah satu bahasa pemrograman yang sangat populer karena kemudahan penggunaan dan fleksibilitasnya dalam berbagai bidang termasuk ilmu data, pengembangan web, machine learning dan kecerdasan buatan. Dengan kelebihan tersebut, Python dapat digunakan diberbagai bidang dengan bantuan library yang memudahkan pengguna dalam menuliskan kode yang lebih kompleks tanpa harus membuatnya dari awal.

II.9.1. Library Python

Library python adalah kumpulan kode yang telah dituliskan sebelumnya sehingga dapat digunakan kembali untuk menyelesaikan berbagai tugas pemrograman. Library python dapat mencakup berbagai fungsi seperti mengolah data, pengelolahan gambar, analisis statistik hingga pengembangan web. Oleh karena itu, dalam analisis sentimen diperlukan pustaka (library) yang dapat membantu dalam memproses dan menganalisis text secara efisien (Mahawardana et al., 2022). Berikut pustaka yang umum digunakan untuk analisis sentimen antara lain :

- a) Scikit-learn
- b) Natural Language Toolkit (NLTK)
- c) Sastrawi
- d) Numpy dan Pandas
- e) Seaborn and Matplotlib

II.10. Tools

II.10.1. Pycharm

Pycharm adalah Integrated Development Environment (IDE) yang dikembangkan oleh JetBrains khusus untuk pemrograman Python. Pycharm menyediakan berbagai fitur yang mendukung pengembangan yang berfokus pada bahasa pemrograman Python seperti sintaksis penyorotan, penyelesaian otomatis, debugging, manajemen proyek serta alat untuk pengujian dan pengelolahan basis data.

II.10.2. Visual Studio Code

Visual Studio Code (VS Code) adalah kode editor sumber terbuka yang dikembangkan oleh Microsoft. Visual Code menawarkan banyak fitur yang membuatnya populer dikalangan pengembang, termasuk dukungan untuk banyak bahasa pemrograman serta banyak ekstensi yang tersedia untuk memenuhi berbagai kebutuhan pengembang.

II.10.3. Streamlit

Streamlit adalah framework open-source yang memungkinkan pengembang, terutama data analyst untuk membangun aplikasi web interaktif dengan cepat menggunakan Python. Dengan *Streamlit*, pengguna dapat mengubah skrip Python

menjadi aplikasi web tanpa perlu memiliki pengetahuan mendalam tentang pengembangan web. *Streamlit* berfungsi untuk visualisasi data, presentasi model machine learning dan pembuatan dashboard interaktif.

BAB III

METODOLOGI PENELITIAN

III.1. Gambaran Umum Objek Penelitian

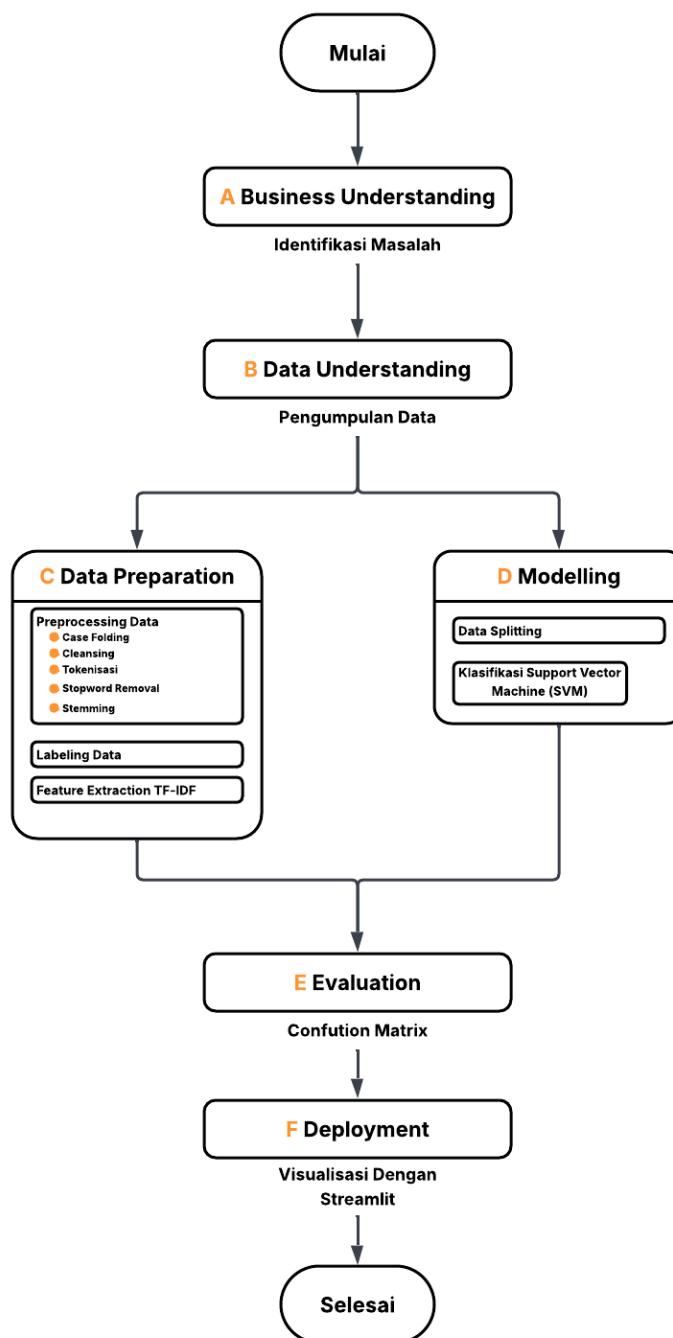
Penelitian ini berfokus pada layanan *Goride* yang merupakan salah satu layanan utama dari aplikasi Gojek. *Goride* menyediakan layanan transportasi berbasis kendaraan roda dua yang memungkinkan pengguna untuk berpergian dengan cepat dan efisien. Dengan jumlah pengguna yang semakin meningkat, terdapat banyak ulasan pengguna yang tersedia di *Google Playstore*. Ulasan ini mencerminkan kepuasan, pengalaman serta keluhan pengguna yang dapat menjadi sumber informasi berharga dalam meningkatkan kualitas layanan.

Sugiyono (2018) menjelaskan bahwa objek penelitian merupakan suatu sasaran yang dijadikan fokus dalam penelitian ilmiah, yang bertujuan untuk memperoleh informasi berupa data yang objektif, valid dan reliabel yang berkaitan dengan variabel yang ditentukan. Objek dalam penelitian ini adalah dataset yang berisi kumpulan ulasan dari para pengguna aplikasi Gojek. Ulasan – ulasan ini terkait dengan layanan Goride yang telah dikumpulkan dari website *Google Playstore*.

III.2. Desain Penelitian

Dalam penelitian ini penulis merancang desain penelitian yang mencakup beberapa tahapan untuk mencapai tujuan yang telah ditetapkan. Desain penelitian ini dibangun dengan menerapkan metodologi *Cross Industry Standard Process for*

Data Mining (CRISP-DM) yang kemudian dikembangkan lebih lanjut oleh penulis. Rincian tahapan penelitian dapat dilihat pada Gambar III.1 berikut.



Gambar III. 1 Desain Penelitian

III.2.1. Business Understanding

Tahap ini dilakukan untuk memahami kebutuhan bisnis dan mendefinisikan masalah yang akan dipecahkan melalui analisis data. Rencana penelitian ini nantinya berfokus pada analisis sentimen pengguna layanan *GoRide* untuk memberikan insight yang dapat meningkatkan kualitas layanan dan kepuasan pengguna. Dengan memahami pola sentimen dalam pengguna, Gojek diharapkan dapat mengambil keputusan strategis berbasis data untuk perbaikan layanan.

III.2.2. Data Understanding

Pada tahap ini, data yang diperoleh dieksplorasi untuk memahami karakteristiknya secara mendalam. Data ulasan pengguna diambil dari *Google Play Store* meliputi user, ulasan teks, rating untuk melihat sentimen dominan, distribusi panjang teks ulasan dan periode waktu ulasan yang paling aktif. Tujuannya adalah untuk mendapatkan wawasan awal mengenai polaritas data sekaligus mengidentifikasi potensi masalah berdasarkan sentimen yang dihasilkan.

III.2.3. Data Preparation

Data preparation adalah tahapan penting untuk memastikan bahwa data mentah yang tidak terstruktur menjadi siap guna dalam pemodelan. Proses ini mencakup tiga langkah utama yaitu :

III.2.3.1 Preprocessing Data

Tahap Preprocessing bertujuan untuk membersikan teks ulasan agar lebih mudah diolah oleh algoritma. Proses ini meliputi *casefolding* (mengubah teks menjadi huruf kecil), *cleansing* (membersihkan teks), *tokenizing* (memecah teks menjadi kata-kata), *stopword removal* (menghapus kata-kata umum yang

tidak bermakna), dan *steaming* (mengembalikan kata ke bentuk dasarnya).

Hasil dari tahap ini membuat teks yang lebih terstruktur dan relevan untuk dianalisis. Tahap preprocessing data yang akan disajikan pada gambar berikut:

1. *Case Folding*

Case folding adalah proses normalisasi teks dalam *preprocessing data* yang mengubah seluruh huruf dalam teks menjadi huruf kecil untuk memastikan konsistensi dalam analisis.

2. *Tokenisasi*

Tokenisasi merupakan proses membagi teks menjadi unit-unit yang lebih kecil yang disebut token. Yang biasanya berupa kata, frasa atau simbol. Proses ini bertujuan untuk mempermudah analisis teks dengan memisahkan setiap elemen yang memiliki makna dalam teks

3. *Stopword Removal*

Stopword removal adalah proses menghapus kata-kata yang tidak memiliki makna signifikan dalam analisis sentimen seperti kata "ini", "dan", "ke" atau "sama". Kata-kata ini biasanya tidak berkontribusi dalam memahami konteks teks.

4. *Stemming*

Stemming adalah proses mengubah teks berimbahan menjadi bentuk dasar untuk menghindari variasi kata yang dapat memengaruhi analisis. Misalnya kata "driver-nya" diubah menjadi "driver" dan "jalan-jalan" menjadi "jalan".

III.2.3.2 Data Labeling

Pada tahap ini, data diberi label berdasarkan kategori sentimen : positif atau negatif. Labeling dilakukan dengan menggunakan anotasi manual jika data ground-truth tidak tersedia.

III.2.3.3 Feature Extraction

Langkah ini mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma pembelajaran mesin. Metode TF-IDF (Term Frequency-Inverse Document Frequency) digunakan untuk menentukan bobot setiap kata dalam ulasan berdasarkan frekuensinya dalam satu dokumen relatif terhadap keseluruhan dataset. Representasi vektor ini memungkinkan algoritma memahami relevansi dan signifikansi kata-kata tertentu dalam analisis sentimen. Berikut penulis sajikan proses Feature Extraction menggunakan TF-IDF berdasarkan data yang telah melalui tahap preprocessing sebelumnya.

III.2.4. Modelling

Tahap pemodelan mencakup pengembangan model pembelajaran mesin dengan menerapkan algoritma klasifikasi Support Vector Machine (SVM) untuk menentukan sentimen ulasan pengguna. Algoritma SVM dipilih karena kemampuannya dalam menangani data berdimensi tinggi serta menghasilkan hyperplane optimal yang mampu membedakan kategori sentimen secara efektif. Selain itu diperlukan penyesuaian parameter (hyperparameter tuning) untuk menentukan kombinasi parameter terbaik seperti Kernel, C, Gamma guna meningkatkan akurasi model dalam proses klasifikasi secara terkomputerisasi.

III.2.5. Evaluation

Pada tahap evaluasi, performa model diukur untuk melihat kinerja algoritma pemodelan yang telah diaplikasikan sebelumnya menggunakan data uji berdasarkan metrik seperti *accuracy*, *presisi*, *recall*, dan *F1-score*. Evaluasi ini memberikan gambaran seberapa baik model mengklasifikasikan pola sentimen dan memprediksi kategori sentimen dengan akurat.

III.2.6. Deployment

Tahap akhir adalah implementasi model dalam bentuk aplikasi visualisasi untuk memudahkan pemangku kepentingan dalam memahami hasil klasifikasi dengan *Support Vector Machine*. Framework *Streamlit* digunakan untuk membuat antarmuka yang intuitif, memungkinkan hasil prediksi sentimen ditampilkan dalam bentuk grafik, diagram, atau laporan. *Insight* dari visualisasi ini diharapkan menjadi acuan bagi Gojek untuk merumuskan strategi peningkatan layanan berdasarkan data sentimen pengguna tanpa dikuti secara khusus bagi penyedia aplikasi Gojek Indonesia.

III.3. Metode Penelitian

III.3.1. Jenis Penelitian dan Metode yang digunakan

Jenis penelitian ini adalah penelitian kuantitatif dengan pendekatan deskriptif. Pendekatan kuantitatif dipilih karena proses penelitian mulai dari pengumpulan hingga visualisasi data berbasis pada analisis numerik (Sugiyono, 2018). Pendekatan deskriptif diterapkan melalui deduksi dimana penelitian diawali dengan kajian teori yang kemudian dikonfirmasi melalui observasi untuk menilai keabsahannya. Hasil penelitian disajikan secara deskriptif guna memberikan

pemahaman yang jelas mengenai data yang diperoleh serta menjawab pertanyaan penelitian.

Metode yang digunakan dalam penelitian ini adalah pendekatan *CRISP-DM* sebagai kerangka kerja tahapan analisis data dengan algoritma *Support Vector Machine (SVM)* sebagai metode utama dalam proses klasifikasi sentimen pada tahap *modeling*. Pendekatan ini dinilai relevan dengan penelitian karena aplikasi Gojek berfokus pada transportasi digital. *CRISP-DM* terdiri dari enam tahapan yaitu *business understanding*, *data understanding*, *data preparation*, *modelling*, *evaluation* dan *deployment* sehingga membentuk alur penelitian yang secara sistematis dan terstruktur.

III.3.2. Jenis dan Sumber Data

Jenis data yang digunakan dalam peneltian ini adalah data sekunder, yaitu data yang dikumpulkan dan diolah oleh peneliti secara mandiri untuk memahami tanggapan serta persepsi pengguna terhadap layanan *Goride di Google Playstore*. Pengambilan data dilakukan menggunakan teknik web scrapping dengan total 659 ulasan. Rentang waktu pengambilan data mencakup periode Januari hingga Desember 2024 sehingga memungkinkan analisis sentiment yang lebih bagus berdasarkan pengalaman pengguna dalam jangka waktu tersebut.

III.4. Variabel Penelitian

Berdasarkan penelitian yang akan dilakukan penulis, dalam penelitian ini terdapat dua varibel utama yang digunakan yaitu variabel bebas (independent) dan variabel terikat (dependent). Variabel independen dalam penelitian ini adalah ulasan atau komentar yang di-post user aplikasi Gojek terhadap layanan *Goride* di situs

Google Playstore. Variabel tersebut berifat dependen karena akan dianalisis untuk menentukan apakah termasuk kedalam sentimen positif atau negatif. Sementara itu, variabel dependen dalam penelitian ini adalah sentimen yang terkandung dalam ulasan pengguna yang dapat berubah tergantung pada isi ulasan pengguna. Sentimen ini kemudian akan digunakan dalam pemodelan klasifikasi menggunakan algoritma *Support Vector Machine (SVM)* untuk mengklasifikasikan sentimen.

III.5. Teknik Pengumpulan Data

Dalam penelitian ini penulis melakukan teknik pengumpulan data untuk memperoleh informasi yang lengkap dan dapat dijadikan referensi permanen dalam penelitian. Penulis memanfaatkan berbagai sumber daya serta studi literatur penelitian terkait yang diperoleh dari berbagai sumber seperti perpustakaan, jurnal, buku, website dan sumber digital lainnya.

III.5.1. Studi Literatur

Studi literatur dilakukan dengan meninjau berbagai referensi yang relevan dengan analisis sentimen dan metode *Support Vector Machine (SVM)*. Referensi yang digunakan mencakup jurnal, buku, laporan penelitian serta publikasi online yang berkaitan dengan analisis sentiment dan machine learning. Studi literatur ini bertujuan untuk memahami metode yang telah digunakan dalam penelitian sebelumnya dan membangun dasar teori yang kuat bagi penelitian ini. Terkait referensi yang penulis gunakan dapat dilihat melalui daftar pustaka.

III.5.2. Web Scrapping

Sumber data dalam penelitian ini dikumpulkan dari *Google Playstore* dengan teknik *Web Scrapping*. Data yang diambil mencakup ulasan teks dari pengguna layanan *Goride*, rating bintang, serta tanggal ulasan. Pengambilan data dilakukan menggunakan bahasa pemrograman *Python* beserta berbagai library pendukung seperti *Google Play Scrapper* yang memanfaatkan *API Google Play Store* untuk mengekstrak informasi dari halaman web. Data yang diperoleh kemudian disimpan kedalam dataset dengan format CSV agar lebih mudah dianalisis pada tahap berikutnya. Gambaran sampel data yang telah dikumpulkan dapat dilihat seperti berikut :

	date	username	ulasan	rating
0	2024-12-31 18:10:06	Adhi Bayu	Pas urgent tapi tidak dapat memesan...	1
1	2024-12-31 12:46:20	nadia safitri	ini goride kenapa ya sekarang, saya...	2
2	2024-12-28 22:15:24	Handini Putri	Mahal, udah gak pernah dapet vouche...	2
3	2024-12-28 19:58:13	Hadi Efendy	Liatt baik baik anjeeeng, tujuan ga...	1
4	2024-12-28 11:16:25	Ra Radyva	Aman selalu Kenyamanan penumpang se...	5
5	2024-12-28 09:54:48	Anisa meila Wati	Goride bosok !! Bukan masalah nomin...	1
6	2024-12-27 22:00:44	PAUL GTEP	Sekarang jumlah driver disekitar lo...	4
7	2024-12-27 14:17:55	Sa Papua	Makin ga jelas ni goride paket hema...	1
8	2024-12-27 13:37:08	badak kecil	Goride hemat jangan di taro bawah d...	1
9	2024-12-26 08:39:26	Shinta Panduri	kok aplikasi goride rada membingung...	4

Gambar III. 2 Dataset Ulasan Pengguna Goride

Merujuk pada Gambar III.2 dataset yang digunakan terdiri dari atas teks ulasan yang diperoleh dari platform digital melalui proses penyaringan menggunakan kata kunci "Goride" untuk memperoleh ulasan yang spesifik. Namun demikian, tidak keseluruhan data teks tersebut akan digunakan secara langsung dalam proses pembelajaran mesin. Sebelum digunakan data terlebih dahulu akan melalui tahapan seleksi bahasa, pra-pemrosesan (preprocessing) serta eliminasi

terhadap teks yang tidak mengandung opini. Langkah-langkah ini dilakukan untuk memastikan bahwa data yang dianalisis benar-benar merepresentasikan opini pengguna, sehingga model yang dibangun dapat menjalankan analisis sentimen secara lebih akurat mengenai tahapan pemrosesan data teks akan disampaikan pada Bab IV.

Tabel III. 1 Metadata Ulasan Pengguna Layanan Goride

Nama Kolom	Tipe Data
Id Ulasan	Integer (Int)
Text	String
Label	String

Tabel III.6 menyajikan metadata atau deskripsi umum dari dataset yang akan digunakan dalam penelitian ini. Setiap baris pada tabel tersebut merepresentasikan fitur-fitur yang terdapat dalam data, termasuk jenis atau tipe data. Oleh karena itu, peneliti hanya memfokuskan penggunaan data pada fitur teks ulasan dan label sentimen sebagai komponen utama dalam pemodelan.

III.6. Teknik Analisis Data

Dalam penelitian ini teknik analisis data dilakukan menggunakan tools PyCharm dan Excel untuk mendukung analisis dan pengujian data. Teknik analisis data yang akan dilakukan di penelitian ini terdiri dari :

1. Analisis Deskriptif, teknik ini digunakan untuk memberikan gambaran umum mengenai distibusi sentimen pengguna terhadap layanan Goride. Analisis ini mencakup statistik dasar seperti jumlah ulasan positif dan negatif.

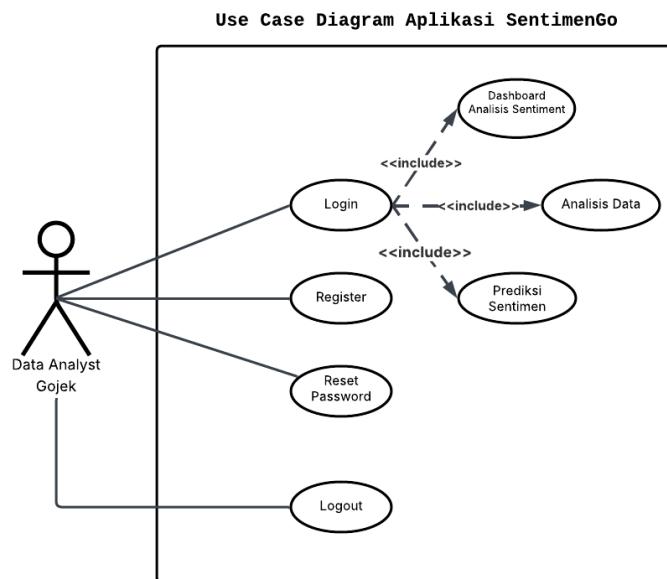
2. Text Mining, teknik ini digunakan untuk menemukan pola dan hubungan antar kata dalam ulasan pengguna.
3. Klasifikasi Sentimen berbasis Machine Learning, proses klasifikasi sentimen dilakukan dengan pendekatan berbasis pembelajaran mesin menggunakan algoritma Support Vector Machine (SVM). Model ini dilatih menggunakan fitur yang diekstraksi dari ulasan pengguna dengan teknik TF-IDF untuk mengubah teks menjadi representasi numerik yang dapat dianalisis oleh algoritma klasifikasi.
4. Evaluasi Model, Model Klasifikasi yang digunakan dalam penelitian ini dievaluasi menggunakan beberapa metrik evaluasi yaitu confusion matrix, accuracy, Precision, Recall dan F1-Score.
5. Visualisasi Data, untuk memudahkan pemahaman pengguna dari hasil analisis maka data divisualisasikan dalam bentuk word cloud, bar chart dan pie chart. Visualisasi ini membantu dalam mengidentifikasi kata-kata kunci yang sering muncul dalam ulasan serta memahami distribusi sentimen secara keseluruhan.

III.7. Rancangan Aplikasi

Pada tahap akhir dari alur CRISP-DM yaitu deployment, model yang telah dibuat akan didistribusikan agar dapat dimanfaatkan oleh pengguna. Berikut penulis sajikan rancangan usecase dan rancangan user interface sistem berbasis website yang akan diimplementasikan pada Bab IV :

III.7.1. Rancangan Use Case Diagram

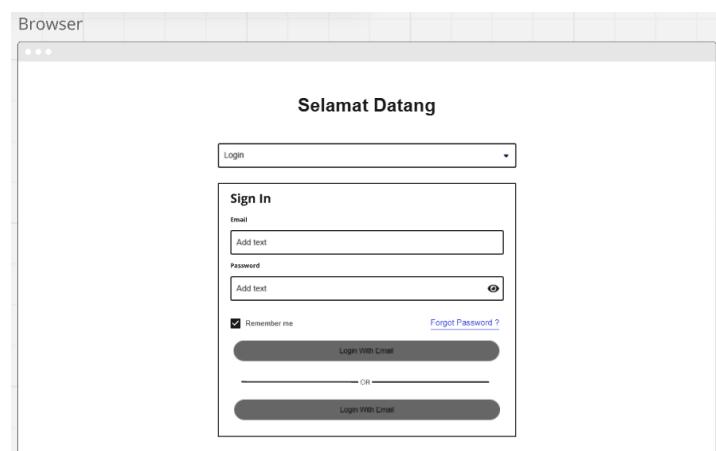
1. Use Case Diagram Aplikasi



Gambar III. 3 Use Case Diagram Aplikasi

III.7.2. Rancangan User Interface

1. User Interface Halaman Login



Gambar III. 4 User Interface Login

2. User Interface Halaman Registrasi

The screenshot shows a web browser window with a light gray header bar containing three dots on the left and a title bar with the text "Selamat Datang". Below the title bar is a dropdown menu with the option "Register". The main content area is titled "Create Account". It contains fields for "First Name" and "Last Name", both with placeholder text "Add text". There is also a "Email" field with placeholder text "Add text" and a "Password" field with placeholder text "Add text". To the right of the password field is a "Confirm Password" field with placeholder text "Add text". Below these fields is a checkbox labeled "I agree to the Terms of Service and Privacy Policy" with a checked mark. At the bottom of the form is a dark gray rectangular button with the text "Create Account" in white.

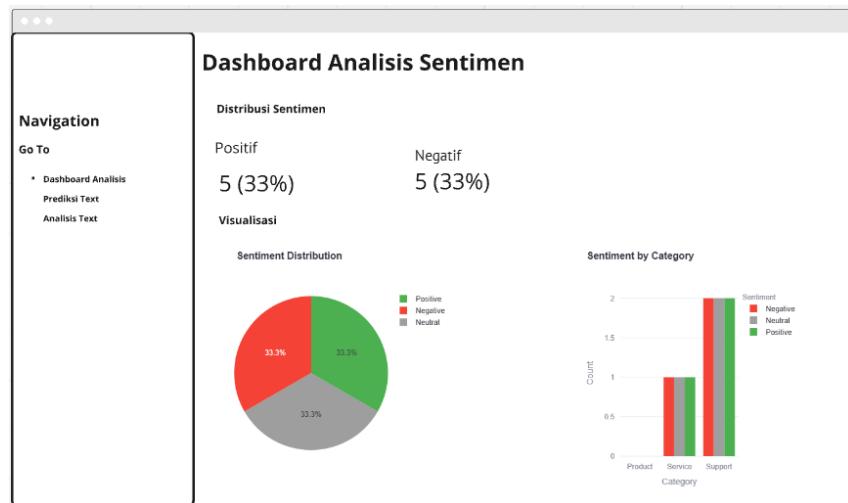
Gambar III. 5 User Interface Registrasi

3. User Interface Halaman Forgot Password

The screenshot shows a web browser window with a light gray header bar containing three dots on the left and a title bar with the text "Selamat Datang". Below the title bar is a dropdown menu with the option "Reset Password". The main content area is titled "Reset Password". It contains a text input field with placeholder text "Enter your email address below and we'll send you instructions to reset your password". Below this is a "Password" field with placeholder text "Add text". At the bottom of the form is a dark gray rectangular button with the text "Seri Reset Link" in white.

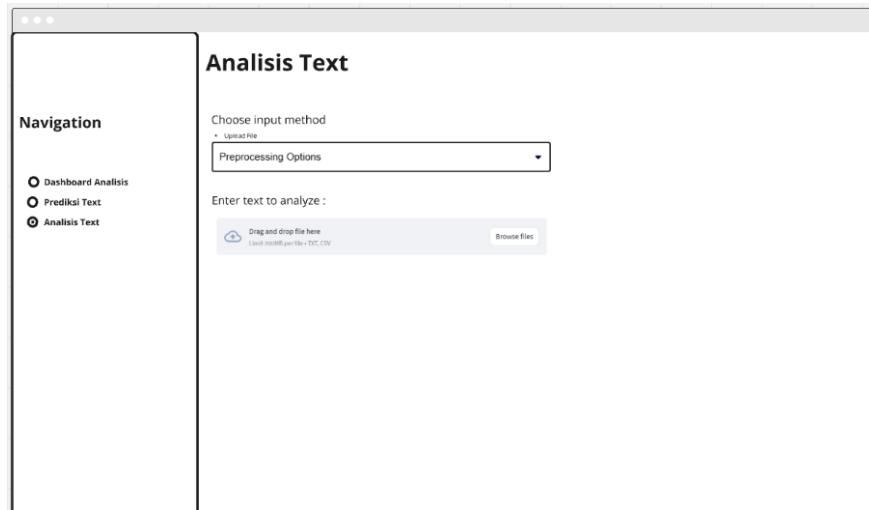
Gambar III. 6 Interface Forgot Password

4. User Interface Halaman Dashboard



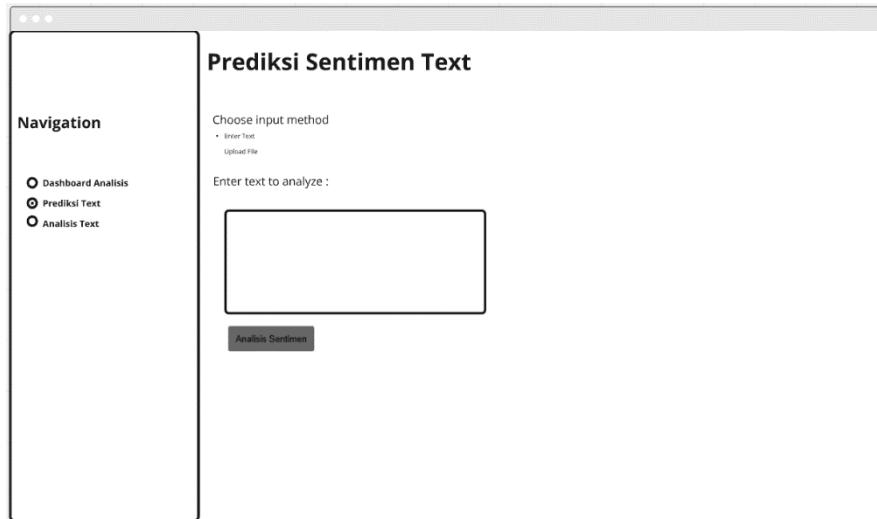
Gambar III. 7 User Interface Dashboard Analisis Sentimen

5. User Interface Halaman Analisis Text



Gambar III. 8 User Interface Analisis Text

6. User Interface Halaman Prediksi Text



Gambar III. 9 User Interface Prediksi Sentimen Text

III.8. Bahan dan Alat Penelitian

III.8.1. Perangkat Keras (Hardware)

Perangkat keras yang digunakan dalam penelitian ini mencakup spesifikasi laptop yang mendukung proses pengolahan data dan implementasi model. Rincian spesifikasi perangkat keras disajikan dalam Tabel III.7.

Tabel III. 2 Spesifikasi Perangkat Keras

Komponen	Spesifikasi
Device	VivoBook ASUS Laptop M1403QA_M1403QA
Processor	AMD Ryzen 7 5800HS dengan Radeon Graphics (16 CPU), ~3.2GHz
Memory	16GB
Graphic	AMD Radeon (TM) Graphic
Storage	SSD 512GB

III.8.2. Perangkat Lunak (Software)

Perangkat lunak yang digunakan dalam penelitian ini meliputi sistem operasi, bahasa pemrograman dan berbagai tools pendukung. Rincian perangkat lunak yang digunakan disajikan dalam Tabel III.8.

Tabel III. 3 Spesifikasi Perangkat Lunak

Perangkat Lunak	Keterangan
<i>Sistem Operasi</i>	Windows 11 Home Single Language 64-bit (10.0, Build 26100)
<i>Bahasa Pemrograman</i>	Python
<i>Aplikasi</i>	<ul style="list-style-type: none"> • PyCharm for Python • Microsoft Excel • Microsoft Visual Code
<i>Web Browser</i>	<ul style="list-style-type: none"> • Microsoft Edge • Google Chrome
<i>Library</i>	<ul style="list-style-type: none"> • Streamlit • Matplotlib • Scikit-learn • NTLK • Sastrawi • Google Play Scrapper .etc

BAB IV

HASIL DAN PEMBAHASAN

IV.1. Implementasi Alur Kerja CRISP-DM

IV.1.1. Hasil Business Understanding

Berdasarkan latar belakang yang telah diuraikan sebelumnya, jumlah pengguna layanan Goride terus mengalami peningkatan dari waktu ke waktu, diikuti oleh pertumbuhan signifikan dalam jumlah ulasan yang diberikan. Ulasan-ulasan tersebut memiliki nilai penting sebagai sumber informasi untuk mengevaluasi layanan, memahami tingkat kepuasan pelanggan serta mengidentifikasi kritik dan saran dari pengguna. Namun besarnya volume dan kompleksitas data ulasan membuat analisis secara manual menjadi tidak efisien untuk menggali persepsi pengguna secara mendalam.

Sebagai solusi, pendekatan analisis sentimen dalam algoritma Support Vector Machine (SVM) digunakan karena kemampuannya dalam melakukan klasifikasi data secara presisi. Dengan mengelompokkan ulasan menjadi sentimen positif dan negatif, hasil analisis dapat memberikan wawasan yang relevan dalam pengambilan keputusan strategis. Model ini nantinya akan diimplementasikan ke dalam sebuah sistem, sehingga diharapkan dapat mempermudah visualisasi hasil analisis dan memungkinkan pemahaman yang lebih cepat serta sistematis terhadap opini publik mengenai layanan goride.

IV.1.2. Hasil Data Understanding

Proses pemahaman data dalam penelitian ini diawali dengan pengambilan data melalui teknik web scrapping menggunakan library google-play-scraper dengan python yang difokuskan pada layanan gojek dalam aplikasi gojek dengan melakukan penyaringan ulasan yang secara spesifik mengandung kata kunci "goride" atau "GORIDE". Berikut adalah kode program scrapping yang dikembangkan dalam penelitian ini:

```

import pandas as pd
from google_play_scraper import Sort, reviews
from datetime import datetime
import time

def scrape_goride_reviews(year=2023, month=4):
    # Konfigurasi
    keywords = ["goride", "gojek", "go ride", "go ride", "GO RIDE", "go ride", "ride",
    "ride@ id", batch_size, lang, country, "com.gojek.app", 200, "id", "id"]

    # Validasi dan setup tanggal
    if not 1 <= month <= 12:
        print("Month harus antara 1-12")
        return None

    start_date = datetime(year, month, 1)
    end_date = datetime(year+1, 1, 1) if month == 12 else datetime(year, month+1, 1)

    print(f"Scraping Goride reviews untuk {start_date.strftime('%B %Y')} ...")

    # Scraping loop
    all_reviews, token, batch_count, start_time = [], None, 0, time.time()

    while True:
        batch_count += 1
        try:
            print(f"Batch {batch_count}: ...")
            result, token = reviews(app_id, lang=lang, country=country,
                                   sort=Sort.MOST_RELEVANT, count=batch_size,
                                   continuation_token=token)

            batch_found = 0
            for review in result:
                if review['at'] < start_date:
                    token = None # Stop scraping
                    break

                if (start_date < review['at'] < end_date and
                    review['content'] and
                    any(kw in review['content'] for kw in keywords)):
                    all_reviews.append({
                        'review_id': review['reviewId'],
                        'user_name': review['username'],
                        'rating': review['score'],
                        'content': review['content'],
                        'date': review['at'].strftime("%d-%m-%Y"),
                        'likes': review['thumbUpCount'],
                        'app_version': review.get('reviewCreatedVersion', '')
                    })
                    batch_found += 1

            print(f"Found {batch_found} matching reviews (Total: {len(all_reviews)})")
            if not token:
                break

            except Exception as e:
                print(f"Error: {e}")
                break

        # Hasil dan simpan
        if not all_reviews:
            print("Tidak ada ulasan ditemukan")
            return None

        df = pd.DataFrame(all_reviews)
        filename = f"ulasan_goride_{year}_{month:02}.csv"
        df.to_csv(filename, index=False, encoding='utf-8')

        print(f"Menyimpan {len(df)} reviews dalam {filename} (start_time:{start_time})")
        print(f"Saved: {filename}")
        return df

    # Jalankan scraper
    if __name__ == "__main__":
        reviews_data = scrape_goride_reviews()
        if reviews_data is not None:
            print(f"\nReviews: {reviews_data.head()}")

```

Gambar IV. 1 Proses Scrapping Data Ulasan Pengguna Goride

Implementasi program scrapping berhasil mengekstraksi ulasan-ulasan yang relevan dalam periode Januari hingga Desember 2024 dan menyimpan seluruh data yang terkumpul dalam format dataset CSV. Dataset yang dihasilkan memiliki

struktur data yang terdiri dari beberapa atribut meliputi review_id, username, review dan date. Berikut dataset yang telah diperoleh dari hasil scraping :

	date	username	ulasan	rating
0	2024-12-31 18:10:06	Adhi Bayu	Pas urgent tapi tidak dapat memesan...	1
1	2024-12-31 12:46:20	nadia safitri	ini goride kenapa ya sekarang, saya...	2
2	2024-12-28 22:15:24	Handini Putri	Mahal, udah gak pernah dapet vouche...	2
3	2024-12-28 19:58:13	Hadi Efendy	liatt baik baik anjeeeng, tujuan ga...	1
4	2024-12-28 11:16:25	Ra Radyva	Aman selalu Kenyamanan penumpang se...	5
5	2024-12-28 09:54:48	Anisa meila Wati	Goride bosok !! Bukan masalah nomin...	1
6	2024-12-27 22:00:44	PAUL GTEP	Sekarang jumlah driver disekitar lo...	4
7	2024-12-27 14:17:55	Sa Papua	Makin ga jelas ni goride paket hema...	1
8	2024-12-27 13:37:08	badak kecil	Goride hemat jangan di taro bawah d...	1
9	2024-12-26 08:39:26	Shinta Panduri	kok aplikasi goride rada membingung...	4

Gambar IV. 2 Contoh Dataset Ulasan Pengguna Layanan Goride

IV.1.3. Hasil Data Preparation

Proses data preparation berfokus pada pengolahan data mentah menjadi bentuk yang dapat digunakan dalam pemodelan. Kegiatan utamanya mencakup preprocessing datam pemberian label pada data serta ekstraksi fitur.

IV.1.3.1 Preprocessing Data

1. Case Folding

Pada tahap preprocessing data, case folding dilakukan untuk mengubah seluruh huruf dalam teks menjadi huruf kecil. Tujuan dari langkah ini adalah untuk menyeragamkan penulisan kata agar tidak terjadi perbedaan makna dari variasi huruf dan menjaga konsistensi data.

Berikut proses dan hasil case folding dalam preprocessing data:

```

# === Langkah 1: Case Folding ===
def case_folding(text):
    before = text
    after = text.lower()
    print_step("Case Folding", before, after)
    return after

df['casefolding'] = df['ulasan'].astype(str).apply(case_folding)
# Menampilkan DataFrame dengan kolom case folding
print("\n==== DataFrame setelah Case Folding ===")
df[['ulasan', 'casefolding']].rename(columns={'ulasan': 'sebelum case folding',
                                              'casefolding': 'setelah case folding'}).head(5)

```

Gambar IV. 3 Preprocessing Data – Case Folding

Tabel IV. 1 Contoh Hasil Case Folding

no	ulasan	setelah_casefolding
0	Aneh, voucher goride gocar gamuncul semua padahal saya punya banyak voucher yang muncul cuma voucher gosend diperbaiki deh tolong	aneh, voucher goride gocar gamuncul semua padahal saya punya banyak voucher yang muncul cuma voucher gosend diperbaiki deh tolong
1	Kok buka aplikasi goride hilang2 mulu...	kok buka aplikasi goride hilang2 mulu...
2	Gojek keren bisa pesen sekalian 2 goride satu waktu makasih gojek 😊😊😊	gojek keren bisa pesen sekalian 2 goride satu waktu makasih gojek 😊😊😊
3	Sekarang udah mahal goride dan gofood	sekarang udah mahal goride dan gofood
4	Kesalahan dari sistem gojek saldo Gopay kepotong otomatis ketika melakukan transaksi order GoRide GoRide nya BATAL tp saldo Gopay kepotong dan gak balik lagi. Udh bikin laporan dan disuruh tuunggu 2x24 jam tp saldo ga masuk lagi ke akun gue. Hadeuuu rumit	kesalahan dari sistem gojek saldo gopay kepotong otomatis ketika melakukan transaksi order goride goride nya batal tp saldo gopay kepotong dan gak balik lagi. udh bikin laporan dan disuruh tuunggu 2x24 jam tp saldo ga masuk lagi ke akun gue. hadeuuu rumit

2. Cleansing

Cleansing dalam penelitian ini dilakukan untuk membersihkan data dari elemen-elemen yang tidak diperlukan seperti angka, tanda baca, simbol, URL, karakter khusus, normalisasi slang, serta remove repeat character. Proses ini dilakukan untuk meningkatkan kualitas informasi yang terkandung dalam data.

```
# === Langkah 2: Cleansing ===
def clean_text(text):
    before = text
    after = re.sub(r'http\S+|www\S+', '', text)
    after = re.sub(r'[^a-zA-Z\s]', ' ', after)
    after = re.sub(r'\s+', ' ', after).strip()
    print_step("Cleansing", before, after)
    return after

df['cleansing'] = df['casefolding'].apply(clean_text)
print("\n== DataFrame setelah Cleansing ==")
df[['casefolding', 'cleansing']].rename(columns={'casefolding': 'sebelum_cleansing',
                                                'cleansing': 'setelah_cleansing'}).head(5)
```

*Gambar IV. 4 Preprocessing Data - Cleansing**Tabel IV. 2 Contoh Hasil Cleansing*

no	sebelum_cleansing	setelah_cleansing
0	aneh, voucher goride gocar gamuncul semua padahal saya punya banyak voucher yang muncul cuma voucher gosend diperbaiki deh tolong	aneh voucher goride gocar gamuncul semua padahal saya punya banyak voucher yang muncul cuma voucher gosend diperbaiki deh tolong
1	kok buka aplikasi goride hilang2 mulu...	kok buka aplikasi goride hilang mulu
2	gojek keren bisa pesen sekalian 2 goride satu waktu makasih gojek 😊😊😊	gojek keren bisa pesen sekalian goride satu waktu makasih gojek
3	sekarang udah mahal goride dan gofood	sekarang udah mahal goride dan gofood
4	kesalahan dari sistem gojek saldo gopay kepotong otomatis ketika melakukan transaksi order goride goride nya batal tp saldo gopay kepotong dan gak balik lagi. udh bikin laporan dan disuruh tuunggu 2x24 jam tp saldo ga masuk lagi ke akun gue. hadeuuu rumit	kesalahan dari sistem gojek saldo gopay kepotong otomatis ketika melakukan transaksi order goride goride nya batal tp saldo gopay kepotong dan gak balik lagi udh bikin laporan dan disuruh tuunggu x jam tp saldo ga masuk lagi ke akun gue hadeuuu rumit

Selanjutnya tahap dari cleansing normalisasi slang untuk menyesuaikan kata non-formal menjadi formal. Penelitian ini menggunakan bantuan kamus slang dari indonlp yang diambil dari github. Berikut proses normalisasi slang:

```

# === Langkah 3: Normalisasi Slang ===
def normalize_slang(text):
    before = text
    words = text.split()
    normalized = [slang_dict.get(word, word) for word in words]
    after = ' '.join(normalized)
    print_step("Normalize Slang", before, after)
    return after

df['normalized'] = df['cleansing'].apply(normalize_slang)
print("\n== DataFrame setelah Normalize Slang ==")
df[['cleansing', 'normalized']].rename(columns={'cleansing': 'sebelum normalize
    slang', 'normalized': 'setelah normalize slang'}).head(5)

```

Gambar IV. 5 Preprocessing Data – Normalisasi Slang**Tabel IV. 3 Contoh Hasil Normalisasi Slang**

no	sebelum normalize slang	setelah normalize slang
0	aneh voucher goride gocar gamuncul semua padahal saya punya banyak voucher yang muncul cuma voucher gosend diperbaiki deh tolong	aneh voucher goride gocar tidak muncul semua padahal saya punya banyak voucher yang muncul hanya voucher gosend diperbaiki deh tolong
1	kok buka aplikasi goride hilang mulu	kenapa buka aplikasi goride hilang melulu
2	gojek keren bisa pesen sekalian goride satu waktu makasih gojek	gojek keren bisa pesan sekalian goride satu waktu terima kasih gojek
3	sekarang udah mahal goride dan gofood	sekarang sudah mahal goride dan gofood
4	kesalahan dari sistem gojek saldo gopay kepotong otomatis ketika melakukan transaksi order goride goride nya batal tp saldo gopay kepotong dan gak balik lagi udh bikin laporan dan disuruh tuunggu x jam tp saldo ga masuk lagi ke akun gue hadeuuu rumit	kesalahan dari sistem gojek saldo gopay kepotong otomatis ketika melakukan transaksi order goride goride nya batal tapi saldo gopay kepotong dan tidak balik lagi sudah bikin laporan dan disuruh tuunggu x jam tapi saldo tidak masuk lagi ke akun saya aduh rumit
5	gojek udh bagus sih cuma lebih banyakin lagi promo nya trus juga kalo pesen goride padahal jarak nya deket bgt tpi kenapa hrs kena harga yang setara sma jarak jauh	gojek sudah bagus hanya lebih banyakin lagi promo nya terus juga kalau pesan goride padahal jarak nya dekat banget tapi kenapa harus terkena harga yang setara sma jarak jauh

Tahap terakhir dari cleansing yaitu remove repeat karakter yang mengharus menghapus karakter berulang. Berikut proses remove repeat karakter :

```
# === Langkah 4: Remove Repeated Characters ===
def remove_repeated(text):
    before = text
    after = re.sub(r'(\w)\1{2,}', r'\1\1', text)
    print_step("Remove Repeated Characters", before, after)
    return after

df['no_repeated'] = df['normalized'].apply(remove_repeated)
print("\n==== DataFrame setelah Remove Repeated Characters ===")
df[['normalized', 'no_repeated']].rename(columns={'normalized': 'sebelum remove repeated', 'no_repeated': 'setelah remove repeated'}).head(5)
```

Gambar IV. 6 Preprocessing Data – Remove Repeated Characters

Tabel IV. 4 Contoh Hasil Remove Repeated Characters

no	sebelum remove repeated	setelah remove repeated
0	aneh voucher goride gocar tidak muncul semua padahal saya punya banyak voucher yang muncul hanya voucher gosend diperbaiki deh tolong	aneh voucher goride gocar tidak muncul semua padahal saya punya banyak voucher yang muncul hanya voucher gosend diperbaiki deh tolong
1	kenapa buka aplikasi goride hilang melulu	kenapa buka aplikasi goride hilang melulu
2	gojek keran bisa pesan sekalian goride satu waktu terima kasih gojek	gojek keran bisa pesan sekalian goride satu waktu terima kasih gojek
3	sekarang sudah mahal goride dan gofood	sekarang sudah mahal goride dan gofood
4	kesalahan dari sistem gojek saldo gopay kepotong otomatis ketika melakukan transaksi order goride goride nya batal tapi saldo gopay kepotong dan tidak balik lagi sudah bikin laporan dan disuruh tuunggu x jam tapi saldo tidak masuk lagi ke akun saya aduh rumit	kesalahan dari sistem gojek saldo gopay kepotong otomatis ketika melakukan transaksi order goride goride nya batal tapi saldo gopay kepotong dan tidak balik lagi sudah bikin laporan dan disuruh tuunggu x jam tapi saldo tidak masuk lagi ke akun saya aduh rumit
5	gojek sudah bagus hanya lebih banyakin lagi promo nya terus juga kalau pesan goride padahal jarak nya dekat banget tapi kenapa harus terkena harga yang setara sma jarak jauh	gojek sudah bagus hanya lebih banyakin lagi promo nya terus juga kalau pesan goride padahal jarak nya dekat banget tapi kenapa harus terkena harga yang setara sma jarak jauh

3. Tokenisasi

Proses tokenisasi dalam penelitian ini merupakan pemisah teks menjadi satuan-satuan kecil yang disebut token berupa kata atau frasa yang dipisah. Tahap ini bertujuan untuk mengubah data teks menjadi format yang terstruktur untuk pelatihan model. Berikut proses dan hasil dari tokenisasi:

```

# === Langkah 5: Tokenization ===
def tokenize_text(text):
    before = text
    after = re.findall(r'\b\w+\b', text)
    print_step("Tokenization", before, after)
    return after

df['tokenized'] = df['no_repeated'].apply(tokenize_text)
print("\n== DataFrame setelah Tokenization ==")
# Pakai repr() agar elemen list terlihat sebagai string Python valid
df_token_display = df[['no_repeated', 'tokenized']].copy()
df_token_display['tokenized'] = df_token_display['tokenized'].apply(repr)
df_token_display.rename(columns={'no_repeated': 'sebelum tokenization', 'tokenized': 'setelah tokenization'}).head(5)

```

Gambar IV. 7 Preprocessing Data - Tokenisasi**Tabel IV. 5 Contoh Hasil Tokenisasi**

no	sebelum tokenization	setelah tokenization
0	aneh voucher goride goCAR tidak muncul semua padahal saya punya banyak voucher yang muncul hanya voucher gosend diperbaiki deh tolong	['aneh', 'voucher', 'goride', 'goCAR', 'tidak', 'muncul', 'semua', 'padahal', 'saya', 'punya', 'banyak', 'voucher', 'yang', 'muncul', 'hanya', 'voucher', 'gosend', 'diperbaiki', 'deh', 'tolong']
1	kenapa buka aplikasi goride hilang melulu	['kenapa', 'buka', 'aplikasi', 'goride', 'hilang', 'melulu']
2	gojek keren bisa pesan sekalian goride satu waktu terima kasih gojek	['gojek', 'keren', 'bisa', 'pesan', 'sekalian', 'goride', 'satu', 'waktu', 'terima', 'kasih', 'gojek']
3	sekarang sudah mahal goride dan gofood	['sekarang', 'sudah', 'mahal', 'goride', 'dan', 'gofood']
4	kesalahan dari sistem gojek saldo gopay kepotong otomatis ketika melakukan transaksi order goride goride nya batal tapi saldo gopay kepotong dan tidak balik lagi sudah bikin laporan dan disuruh tuunggu x jam tapi saldo tidak masuk lagi ke akun saya aduh rumit	['kesalahan', 'dari', 'sistem', 'gojek', 'saldo', 'gopay', 'kepotong', 'otomatis', 'ketika', 'melakukan', 'transaksi', 'order', 'goride', 'goride', 'nya', 'batal', 'tapi', 'saldo', 'gopay', 'kepotong', 'dan', 'tidak', 'balik', 'lagi', 'sudah', 'bikin', 'laporan', 'dan', 'disuruh', 'tuunggu', 'x', 'jam', 'tapi', 'saldo', 'tidak', 'masuk', 'lagi', 'ke', 'akun', 'saya', 'aduh', 'rumit']
5	gojek sudah bagus hanya lebih banyakkin lagi promo nya terus juga kalau pesan goride padahal jarak nya dekat banget tapi kenapa harus terkena harga yang setara sama jarak jauh	['gojek', 'sudah', 'bagus', 'hanya', 'lebih', 'banyakkin', 'lagi', 'promo', 'nya', 'terus', 'juga', 'kalau', 'pesan', 'goride', 'padahal', 'jarak', 'nya', 'dekat', 'banget', 'tapi', 'kenapa', 'harus', 'terkena', 'harga', 'yang', 'setara', 'sma', 'jarak', 'jauh']

4. Stopword Removal

Stopword removal dalam preprocessing data yaitu penghapusan kata-kata umum yang memiliki frekuensi tinggi namun tidak memberikan makna penting dalam analisis seperti kata hubung atau kata depan. Berikut proses stopword removal :

```

# === Langkah 6: Stopword Removal ===
def remove_stopwords(tokens):
    if isinstance(tokens, str):
        tokens = re.findall(r'\b\w+\b', tokens)
    before = tokens.copy()
    after = [word for word in tokens if word not in stopword_custom]
    print_step("Stopword Removal", before, after)
    return after

# Terapkan fungsi ke DataFrame
df['no_stopword'] = df['tokenized'].apply(remove_stopwords)

print("\n== DataFrame setelah Stopword Removal ==")
df_stopword_display = df[['tokenized', 'no_stopword']].copy()
df_stopword_display['tokenized'] = df_stopword_display['tokenized'].apply(repr)
df_stopword_display['no_stopword'] = df_stopword_display['no_stopword'].apply(repr)

# Rename dan tampilkan 10 baris pertama
df_stopword_display.rename(columns={
    'tokenized': 'sebelum_no_stopword',
    'no_stopword': 'setelah_no_stopword'
}).head(5)

```

Gambar IV. 8 Preprocessing Data – Stopword Removal**Tabel IV. 6 Contoh Hasil Stopword Removal**

no	sebelum no_stopword	setelah no_stopword
0	['aneh', 'voucher', 'goride', 'gocar', 'tidak', 'muncul', 'semua', 'padahal', 'saya', 'punya', 'banyak', 'voucher', 'yang', 'muncul', 'hanya', 'voucher', 'gosend', 'diperbaiki', 'deh', 'tolong']	['aneh', 'voucher', 'goride', 'gocar', 'tidak', 'muncul', 'voucher', 'muncul', 'voucher', 'gosend', 'diperbaiki', 'tolong']
1	['kenapa', 'buka', 'aplikasi', 'goride', 'hilang', 'melulu']	['buka', 'aplikasi', 'goride', 'hilang', 'melulu']
2	['gojek', 'keren', 'bisa', 'pesan', 'sekalian', 'goride', 'satu', 'waktu', 'terima', 'kasih', 'gojek']	['gojek', 'keren', 'bisa', 'pesan', 'goride', 'terima', 'kasih', 'gojek']
3	['sekarang', 'sudah', 'mahal', 'goride', 'dan', 'gofood']	['mahal', 'goride', 'gofood']
4	['kesalahan', 'dari', 'sistem', 'gojek', 'saldo', 'gopay', 'kepotong', 'otomatis', 'ketika', 'melakukan', 'transaksi', 'order', 'goride', 'goride', 'nya', 'batal', 'tapi', 'saldo', 'gopay', 'kepotong', 'dan', 'tidak', 'balik', 'lagi', 'sudah', 'bikin', 'laporan', 'dan', 'disuruh', 'tuunggu', 'x', 'jam', 'tapi', 'saldo', 'tidak', 'masuk', 'lagi', 'ke', 'akun', 'saya', 'aduh', 'rumit']	['kesalahan', 'sistem', 'gojek', 'saldo', 'gopay', 'kepotong', 'otomatis', 'transaksi', 'order', 'goride', 'goride', 'nya', 'batal', 'saldo', 'gopay', 'kepotong', 'tidak', 'bikin', 'laporan', 'disuruh', 'tuunggu', 'x', 'jam', 'saldo', 'tidak', 'masuk', 'akun', 'aduh', 'rumit']
5	['gojek', 'sudah', 'bagus', 'hanya', 'lebih', 'banyakin', 'lagi', 'promo', 'nya', 'terus', 'juga', 'kalau', 'pesan', 'goride', 'padahal', 'jarak', 'nya', 'dekat', 'banget', 'tapi', 'kenapa', 'harus', 'terkena', 'harga', 'yang', 'setara', 'sma', 'jarak', 'jauh']	['gojek', 'bagus', 'lebih', 'banyakin', 'promo', 'nya', 'pesan', 'goride', 'jarak', 'nya', 'banget', 'terkena', 'harga', 'setara', 'sma', 'jarak']

5. Stemming

Langkah terakhir dalam preprocessing data adalah stemming yang bertujuan mengubah kata dalam bentuk dasarnya. Dengan menghilangkan imbuhan kata memungkinkan sistem mengenali kata yang memiliki makna dasar yang sama sehingga dapat meningkatkan efisiensi dalam pemodelan data.

```
# === Langkah 7: Stemming ===
def stem_tokens(tokens):
    if isinstance(tokens, str):
        tokens = tokens.split()
    before = tokens.copy()
    after = [stemmer.stem(word) for word in tokens]
    print_step("Stemming", before, after)
    return after

# Terapkan proses stemming ke kolom 'no_stopword'
df['stemmed'] = df['no_stopword'].apply(stem_tokens)

print("\n==== DataFrame setelah Stemming ===")

# Format tampilan agar sama seperti langkah stopwords removal
df_stemming_display = df[['no_stopword', 'stemmed']].copy()
df_stemming_display['no_stopword'] = df_stemming_display['no_stopword'].apply(repr)
df_stemming_display['stemmed'] = df_stemming_display['stemmed'].apply(repr)

# Rename kolom dan tampilkan 10 baris pertama
df_stemming_display = df_stemming_display.rename(columns={
    'no_stopword': 'sebelum stemming',
    'stemmed': 'setelah stemming'
})
df_stemming_display.head(5)
```

Gambar IV. 9 Preprocessing Data - Stemming

Tabel IV. 7 Contoh Hasil Stemming

no	sebelum stemming	setelah stemming
0	['aneh', 'voucher', 'goride', 'gocar', 'tidak', 'muncul', 'voucher', 'muncul', 'voucher', 'gosend', 'diperbaiki', 'tolong']	['aneh', 'voucher', 'goride', 'gocar', 'tidak', 'muncul', 'voucher', 'muncul', 'voucher', 'gosend', 'baik', 'tolong']
1	['buka', 'aplikasi', 'goride', 'hilang', 'melulu']	['buka', 'aplikasi', 'goride', 'hilang', 'melulu']
2	['gojek', 'keren', 'bisa', 'pesan', 'goride', 'terima', 'kasih', 'gojek']	['gojek', 'keren', 'bisa', 'pesan', 'goride', 'terima', 'kasih', 'gojek']
3	['mahal', 'goride', 'gofood']	['mahal', 'goride', 'gofood']
4	['kesalahan', 'sistem', 'gojek', 'saldo', 'gopay', 'kepotong', 'otomatis', 'transaksi', 'order', 'goride', 'goride', 'nya', 'batal', 'saldo', 'gopay', 'kepotong', 'tidak', 'bikin', 'laporan', 'disuruh', 'tuunggu', 'x', 'jam', 'saldo', 'tidak', 'masuk', 'akun', 'aduh', 'rumbit']	['salah', 'sistem', 'gojek', 'saldo', 'gopay', 'potong', 'otomatis', 'transaksi', 'order', 'goride', 'goride', 'nya', 'batal', 'saldo', 'gopay', 'potong', 'tidak', 'bikin', 'lapor', 'suruh', 'tuunggu', 'x', 'jam', 'saldo', 'tidak', 'masuk', 'akun', 'aduh', 'rumbit']
5	['gojek', 'bagus', 'lebih', 'banyakin', 'promo', 'nya', 'pesan', 'goride', 'jarak', 'nya', 'banget', 'terkena', 'harga', 'setara', 'sma', 'jarak']	['gojek', 'bagus', 'lebih', 'banyakin', 'promo', 'nya', 'pesan', 'goride', 'jarak', 'nya', 'banget', 'kena', 'harga', 'tara', 'sma', 'jarak']

Berdasarkan hasil preprocessing yang telah dilakukan pada Data Preparation diperoleh dataset yang telah melalui berbagai tahap preprocessing data sehingga dapat digunakan untuk analisis lebih lanjut. Berikut merupakan tampilan data sebelum dan sesudah melalui proses preprocessing:

Tabel IV. 8 Contoh Hasil Dataset Yang Telah Dipreprocessing

no	sebelum preprocessing	setelah preprocessing
0	Aneh, voucher goride gocar gamuncul semua padahal saya punya banyak voucher yang muncul cuma voucher gosend diperbaiki deh tolong	aneh voucher goride tidak muncul voucher muncul voucher baik tolong
1	Kok buka aplikasi goride hilang2 mulu...	buka aplikasi goride hilang melulu
2	Gojek keren bisa pesen sekalian 2 goride satu waktu makasih gojek 🎉🎉	gojek keren bisa pesan goride terima kasih gojek
3	Sekarang udah mahal goride dan gofood	mahal goride
4	Kesalahan dari sistem gojek saldo Gopay kepotong otomatis ketika melakukan transaksi order GoRide GoRide nya BATAL tp saldo Gopay kepotong dan gak balik lagi. Udh bikin laporan dan disuruh tuunggu 2x24 jam tp saldo ga masuk lagi ke akun gue. Hadeuuu rumit	salah sistem gojek saldo gopay potong otomatis transaksi order goride goride nya batal saldo gopay potong tidak bikin lapor suruh tuunggu x jam saldo tidak masuk akun aduh rumit
5	Gojek udh bagus sih, cuma lebih banyakin lagi promo nya, trus juga kalo pesen goride padahal jarak nya deket bgt tpi kenapa hrs kena harga yang setara sama jarak jauh.	gojek bagus lebih banyakin promo nya pesan goride jarak nya banget kena harga tara sama jarak

IV.1.3.2 Labeling Data

Pemberian label pada dataset ulasan dilakukan dengan pendekatan semiotomatis. Dalam tahapan pelabelan ini, model pre-trained BERT dimanfaatkan untuk mengidentifikasi kategori sentimen pada setiap

ulasan, sebagai langkah awal sebelum digunakan dalam proses pelatihan model SVM. Untuk memastikan relevansi hasil pelabelan penulis melakukan verifikasi label secara manual berdasarkan konteks sebenarnya dari masing-masing ulasan. Berikut adalah proses dan hasil pelabelan dataset menggunakan model pre-trained BERT.

```
# STEP 1: Import library
import pandas as pd
from transformers import pipeline, AutoTokenizer, AutoModelForSequenceClassification

# STEP 2: Load IndoBERT model dan tokenizer
model_name = "mdhugol/indonesia-bert-sentiment-classification"
tokenizer = AutoTokenizer.from_pretrained(model_name)
model = AutoModelForSequenceClassification.from_pretrained(model_name)

# STEP 3: Load dataset CSV
df = pd.read_csv("goride_review.csv")
print(f"Jumlah ulasan: {len(df)}")
df.head()

# STEP 4: Buat pipeline analisis sentimen
sentiment_pipeline = pipeline("sentiment-analysis", model=model, tokenizer=tokenizer)

# STEP 5: Lakukan pseudo-labeling (berlabel otomatis)
# Proses batch untuk menghindari error terlalu banyak permintaan
batch_size = 100
labels = []

for i in range(0, len(df), batch_size):
    batch = df["review_text"].iloc[i:i+batch_size].tolist()
    preds = sentiment_pipeline(batch)
    labels.extend([pred["label"] for pred in preds])

df["pseudo_label"] = labels
```

Gambar IV. 10 Pelabelan Dataset Dengan Model Pre-trained BERT

Tabel IV. 9 Contoh Hasil Pelabelan Dataset

no	ulasan	sentiment
0	aneh voucher goride tidak muncul voucher muncul voucher baik tolong	Negative
1	buka aplikasi goride hilang melulu	Negative
2	gojek keren bisa pesan goride terima kasih gojek	Positive
3	mahal goride	Negative
4	salah sistem gojek saldo gopay potong otomatis transaksi order goride goride nya batal saldo gopay potong tidak bikin lapor suruh tuunggu x jam saldo tidak masuk akun aduh rumit	Negative

IV.1.3.3 Feature Extraction

Feature extraction merupakan bagian dari tahap pengolahan data dimana ulasan yang telah melalui tahap preprocessing dikonversi dalam bentuk numerik menggunakan teknik TF-IDF. Metode ini digunakan untuk menghasilkan representasi fitur yang dapat diproses dalam model klasifikasi. Proses ini diimplementasikan menggunakan TfidfVectorizer dengan beberapa parameter khusus. Berikut proses feature extraction yang diterapkan dalam penelitian ini

```
from sklearn.feature_extraction.text import TfidfVectorizer

# TF-IDF Feature Extraction
tfidf = TfidfVectorizer(
    ngram_range=(1,2),
    max_features=1000,
    min_df=2,
    max_df=0.95,
    sublinear_tf=True
)

X_tfidf = tfidf.fit_transform(df['preprocessing_text'])
```

Gambar IV. 11 Proses Data Preparation – Feature Extraction

Beberapa parameter yang digunakan dalam proses feature extraction akan diuraikan pada tabel berikut :

Tabel IV. 10 Paramater Feature Extraction TF_IDF

Parameter	Nilai
ngram_range	(1,2)
max_features	1000
min_df	2
max_df	0.95
Sublinear_tf	True

Setiap parameter pada ekstraksi fitur dengan *TF-IDF* digunakan untuk mengoptimalkan representasi fitur seperti *ngram_range*(1,2) untuk menangkap *unigram* dan *bigram*, *max_features=1000* membatasi jumlah fitur agar efisien, *min_df=2* menghapus kata yang jarang muncul, *max_df=0.95* membuang kata terlalu umum dan *sublinear_linear=true* memberi bobot pada teks yang memiliki frekuensi kemunculan tinggi tidak terlalu mendominasi. Hasil dari proses ini menghasilkan 1000 fitur teks yang dapat digunakan pada tahap pemodelan. Berikut hasil proses feature extraction yang disajikan pada tabel.

Tabel IV. 11 Contoh Hasil Feature Extraction

	Doc_ID	Term	TF-IDF
0	0	muncul	0.587920
1	0	voucher	0.526966
2	0	aneh	0.347235
3	0	voucher goride	0.302120
4	0	goride tidak	0.254053
5	0	tolong	0.200205
6	0	baik	0.199515
7	0	tidak	0.121079
8	0	goride	0.075609
9	1	aplikasi goride	0.478957
...

IV.1.4. Hasil Modelling

Pemodelan algoritma merupakan bagian utama dari proses analisis sentimen yang bertujuan membangun model klasifikasi berdasarkan data ulasan yang telah melalui tahap preprocessing. Pada penelitian ini proses pemodelan diawali dengan pembagian data (data splitting) untuk memisahkan data latih dan uji kemudian dilanjutkan dengan penerapan algoritma Support Vector Machine (SVM untuk melatih model dalam mengenali pola sentimen yang terdapat dalam data.

IV.1.4.1 Data Splitting

Tahap awal dalam pemodelan adalah pembagian data (data splitting). Proses ini dilakukan untuk memisahkan dataset menjadi set data latih (training set) dan set data uji (testing set) untuk memvalidasi model secara objektif dan mencegah terjadinya overfitting. Proses data

splitting yang dapat dilihat pada Gambar IV.1 dan proporsi pembagian data yang diterapkan dalam penelitian ini disajikan pada tabel IV.II

```
X = df['ulasan_bersih']
y = df['label']

# Bagi data menjadi 90% data latih dan 10% data uji
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.1, stratify=y, random_state=42
)

print(f"Jumlah data latih: {len(X_train)}")
print(f"Jumlah data uji: {len(X_test)}")
✓ [3] 18ms
```

Jumlah data latih: 593
Jumlah data uji: 66

Gambar IV. 12 Proses Splitting Data

Tabel IV. 12 Skema Pembagian Data

Komponen	Jumlah Data	Persentase
Data Latih (<i>Train</i>)	593	90%
Data Uji (<i>Test</i>)	66	10%
Total Data	659	100%

IV.1.4.2 Klasifikasi Support Vector Machine

Klasifikasi SVM dipilih untuk membangun model prediksi sentiment dengan penyetelan parameter yang sesuai. Model dilatih dengan set data latih yang telah melalui preprocessing feature extraction untuk mempelajari pola karakteristik yang membedakan sentimen disetiap ulasan. Untuk mencapai performa optimal dilakukan proses hyperparameter tuning dengan metode GridSearchCV. Tabel VI.13 menunjukkan hyperparameter yang akan diuji serta hasil tuning hyperparameter terbaik yang diperoleh serta proses penyetelan

hyperparameter tuning untuk algoritma support vector machine terhadap dataset ulasan pengguna layanan goride pada Gambar IV.13

```
# Membuat pipeline untuk menggabungkan langkah-langkah pemodelan
pipeline = ImbPipeline([
    ('tfidf', TfidfVectorizer(ngram_range=(1, 2),
        max_features=1000,
        min_df=1,
        max_df=1.0,
        sublinear_tf=True)),
    ('smote', SMOTE(random_state=42)),
    ('svm', SVC(probability=True, random_state=42))
])

# Menentukan parameter yang akan diuji untuk menemukan kombinasi terbaik
param_grid = {
    'svm__C': [0.1, 1, 10],
    'svm__kernel': ['linear', 'rbf'],
    'svm__gamma': ['scale', 'auto']
}

# Membuat objek GridSearchCV untuk melakukan pencarian hiperparameter dengan cross-validation
grid_search = GridSearchCV(
    pipeline, param_grid=param_grid,
    scoring='f1', cv=5, verbose=1, n_jobs=-1
)

# Melatih model dengan data latih
print("Memulai pencarian hiperparameter terbaik...")
grid_search.fit(X_train, y_train)
```

Gambar IV. 15 Proses Pengaturan Hyperparamter Tuning dengan GridSearchCV

Tabel IV. 13 Hyperparamter dan Nilai Parameter Terbaik Model SVM

Hyperparameter	Nilai	Nilai Parameter Terbaik
C	[0.1, 1, 10]	0.1
Kernel	['linier','rbf']	linier
Gamma	['scale','auto']	scale

Berdasarkan tiga hyperparameter utama yang di uji meliputi C, kernel dan gamma dengan ruang pencarian nilai yang telah ditentukan berdasarkan referensi dan eksperimen awal. Hasil pengujian konfigurasi parameter terbaik yang memberikan kinerja klasifikasi optimal diperoleh pada nilai C = 0.1, kernel = linier dan gamma = scale yang kemudian digunakan sebagai parameter akhir dalam pelatihan model SVM.

Pemodelan algoritma Support Vector Machine (SVM) dengan parameter terbaik dapat dilihat pada Gambar IV.14.

```
# Modelling SVM dengan parameter terbaik
best_svm_model = SVC(
    C=0.1,
    kernel='linear',
    gamma='scale',
    probability=True,
    random_state=42
)

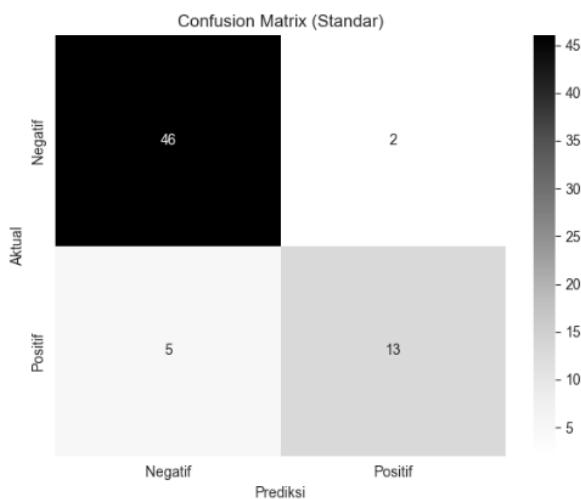
best_svm_model.fit(X_train_balanced, y_train_balanced)
y_pred_final = best_svm_model.predict(X_test_tfidf)
```

Gambar IV. 16 Hasil Modelling Algoritma Support Vector Machine

IV.1.5. Hasil Evaluation

Evaluasi performa dilakukan terhadap model Support Vector Machine (SVM) yang telah dikonfiguras dengan hyperparameter terbaik. Tahap ini bertujuan untuk mengukur secara objektif kemampuan model dalam melakukan klasifikasi sentimen pada ulasan pengguna layanan Goride dengan menggunakan data uji yang independen. Kinerja model diukur secara kuantitatif melalui metrik accuracy, precision, recall dan f-1 score. Dasar dari perhitungan matrik tersebut adalah Confution Matrix, hasil evaluasi confution matrix pada algoritma Support Vector

Machine (SVM) dapat dilihat dalam bentuk visual pada gambar IV.15 dan dalam bentuk matriks 2 dimensi pada tabel .



Gambar IV. 17 Visual Confution Matrix Algoritma SVM

Tabel IV. 14 Confution Matrix Model SVM

	pred.negative	pred.positive
act.negative	46 (TN)	2 (FN)
act.positive	5 (FP)	13 (TP)

Dari tabel confution matrix diatas dapat diperoleh rincian prediksi model. Diperoleh nilai True Positive (TP) sebanyak 13, True Negative (TN) sebanyak 46, False Positive (FP) sebanyak 2 dan False Negative (FN) sebanyak 5. Berdasarkan nilai-nilai tersebut dapat dilakukan perhitungan manual untuk setiap metrik evaluasi sebagai berikut :

- Accuracy

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

$$= \frac{13 + 46}{13 + 46 + 2 + 5}$$

$$= \frac{59}{66} \approx 0,89$$

- Precision

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$= \frac{13}{13 + 2}$$

$$= \frac{13}{15} \approx 0.87$$

- Recall

$$\text{Recall} = \frac{TP}{TP + FN}$$

$$= \frac{13}{13 + 5}$$

$$= \frac{13}{18} \approx 0.72$$

- F1-Score

$$\begin{aligned} F_{\text{1-Score}} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \\ &= 2 \times \frac{0.87 \times 0.72}{0.87 + 0.72} \approx 0.79 \end{aligned}$$

Hasil perhitungan metrik evaluasi menunjukkan performa model yang efektif dengan akurasi keseluruhan mencapai 89%. Dengan spesifik untuk kelas "Positive" model memperoleh nilai precision 87%, recall 72% dan F1-Score 79% yang menginfikasikan kemampuan model yang baik dalam melakukan klasifikasi set data ulasan.

IV.1.6. Hasil Deployment

Tahap deployment merupakan fase akhir dalam kerangka kerja CRISP-DM dimana model klasifikasi Support Vector Machine (SVM) yang telah dikembangkan dan dievaluasi diimplementasikan dalam bentuk aplikasi berbasis website bernama SentimenGo.

SentimenGo merupakan hasil pengembangan aplikasi dalam penelitian ini yang berfungsi sebagai platform aplikasi interaktif yang dirancang untuk mempermudah masyarakat dan pemangku kepentingan dalam memahami opini publik terhadap layanan goride. Aplikasi ini menyajikan hasil analisis sentimen dalam bentuk visual data yang informatif dan mudah diinterpretasikan. Antarmuka sistem ini dirancang secara sistematis berdasarkan konsep desain yang telah dijelaskan pada subbab III.7.2 untuk memastikan keselarasan antara implementasi aktual dan rancangan yang telah ditetapkan sebelumnya.

IV.1.6.1 Implementasi Autentikasi Aplikasi SentimenGo

Berdasarkan rancangan pada Gambar III.3, Gambar III.4 dan Gambar III.5 sistem autentikasi pengguna berhasil diimplementasikan. Fitur ini berfungsi untuk memastikan pengguna yang memiliki kredensial valid yang dapat mengakses lingkungan analisis data.

1. Tampilan halaman masuk

Halaman daftar berfungsi sebagai akses awal pengguna untuk membuat akun personal dengan mendaftarkan email dan kata sandi sebagaimana ditampilkan pada Gambar.16.

Pilih metode autentikasi

Daftar

Daftar

Nama Depan

Nama Belakang

Email

Kata Sandi

Konfirmasi Kata Sandi

Saya setuju dengan Syarat Layanan dan Kebijakan Privasi

Buat Akun

Tips Berguna

Gambar IV. 20 Tampilan Halaman Daftar Akun

2. Tampilan halaman masuk

Halaman masuk berfungsi sebagai proses verifikasi bagi pengguna terdaftar untuk mengakses tools aplikasi analisis sentimen. Berikut tampilan halaman masuk pada Gambar.17.

Pilih metode autentikasi

Masuk

Masuk

Email

Kata Sandi

Ingat saya

Lanjutkan dengan Email

ATAU

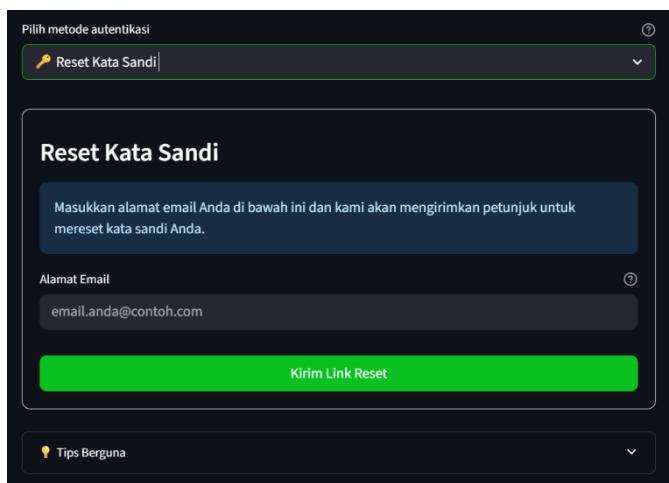
Lanjutkan dengan Google

Tips Berguna

Gambar IV. 21 Tampilan Halaman Masuk Akun

3. Tampilan halaman reset password

Halaman reset password dirancang khusus untuk membantu user yang terkendala terkait kridensial akun untuk melakukan autentikasi apabila terjadi lupa password akun. Berikut tampilan reset password pada Gambar.18.



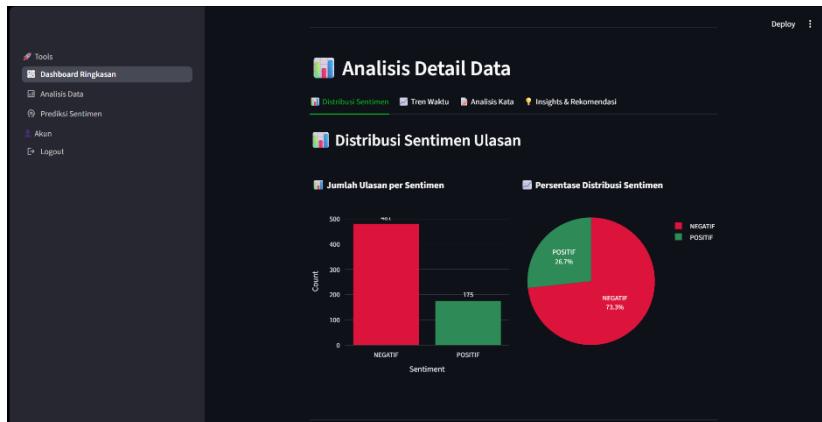
Gambar IV. 22 Tampilan Halaman Reset Kata Sandi Akun

IV.1.6.2 Implementasi Tools Aplikasi SentimenGo

Deployment inti aplikasi SentimenGo difokuskan pada tiga halaman utama yang masing-masing memiliki fungsi yang spesifik dalam menangani proses analisis sentimen ulasan Goride.

1. Tampilan Halaman Dashboard Ringkasan

Halaman ini merupakan dashboard analisis sentimen yang dijadikan dengan visualisasi melalui diagram batang dan pie char terkait proporsi sentimen sehingga dapat menyajikan ringkasan distribusi sentimen ulasan positif atau negatif. Lebih jelasnya tampilan dashboard ringkasan dapat dilihat pada Gambar.19.



Gambar IV. 23 Tampilan Halaman Dashboard Ringkasan

2. Tampilan Halaman Analisis Data

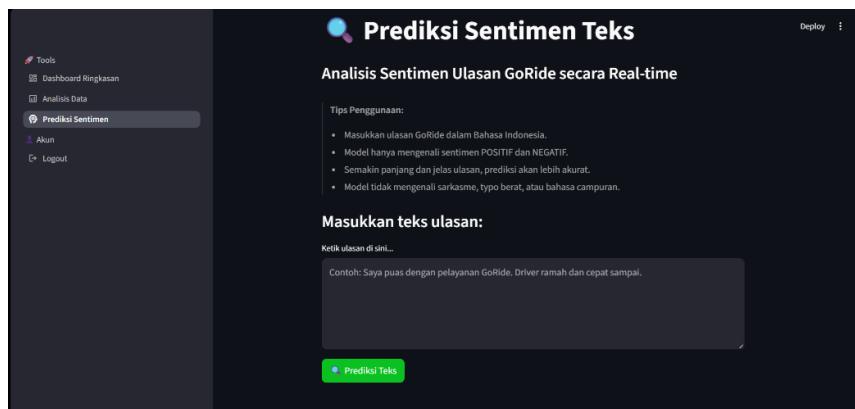
Halaman Analisis Data merupakan fitur yang dibuat untuk mengakomodasi kebutuhan analisis yang lebih dinamis dan spesifik yang memungkinkan pengguna untuk mengunggah dataset terbaru ulasan goride dalam format .csv dan sistem akan otomatis melakukan analisis. Berikut halaman analisis data dapat dilihat pada Gambar.20.



Gambar IV. 24 Tampilan Halaman Analisis Data

3. Tampilan Halaman Prediksi Sentimen

Halaman Prediksi Sentimen merupakan fitur demonstrasi langsung dari penggunaan model SVM secara real-time. Dengan menyediakan kolom input bagi pengguna untuk memasukkan satu ulasan kemudian sistem akan memproses dan menampilkan hasil prediksi sentimen yang mana bagian ini berfungsi sebagai alat validasi model. Berikut tampilan halaman prediksi sentimen dilihat pada Gambar.21.



Gambar IV. 25 Halaman Prediksi Sentimen

IV.2. Pembahasan

IV.2.1. Uji Sistem Aplikasi SentimenGo

Untuk menguji proses implementasi aplikasi perlu dilakukan serangkaian pengujian fungsionalitas sistem secara menyeluruh untuk memastikan setiap fungsi berjalan dengan baik. Pengujian ini menggunakan metode Black Box Testing yang perfokus pada perspektif pengguna akhir tanpa melihat struktur kode internal. Pendekatan ini memastikan bahwa aplikasi tidak hanya berkerja secara teknis tetapi

dapat digunakan dalam skenario nyata. Rincian dari setiap skenario pengujian disajikan secara lengkap pada Tabel IV.15.

Tabel IV. 15 Tabel Black Box Testing Case Aplikasi SentimenGo

No	Skenario Pengujian	Test Case	Hasil yang Diharapkan	Hasil Pengujian	Kesimpulan
1	Autentikasi Masuk	Login dengan email dan password valid	Pengguna berhasil login dan diarahkan ke dashboard	Pengguna berhasil login dan berhasil diarahkan ke dashboard	Valid
2	Autentikasi Masuk	Login dengan email salah	Pesan error "Email tidak ditemukan" muncul	Pesan error muncul	Valid
3	Autentikasi Masuk	Login dengan password salah	Pesan error "Kata sandi salah" muncul	Pesan error muncul	Valid
4	Autentikasi Masuk	Login dengan email belum verifikasi	Pesan verifikasi email muncul	Pesan " Email Anda belum diverifikasi!" muncul	Valid
5	Daftar Akun	Registrasi dengan data valid	Akun berhasil dibuat dan email verifikasi dikirim	Email berhasil dibuat dan email verifikasi berhasil dikirimkan	Valid
6	Daftar Akun	Registrasi dengan email sudah terdaftar	Pesan error "Email sudah terdaftar" muncul	Pesan error muncul	Valid
7	Daftar Akun	Registrasi dengan password lemah	Pesan error "Kata sandi terlalu lemah" muncul	Pesan error muncul	Valid
8	Reset Password	Reset password dengan email terdaftar	Email reset password dikirim	Email berhasil dikirimkan untuk reset password	Valid
9	Reset Password	Reset password dengan email tidak terdaftar	Pesan error "Email tidak ditemukan" muncul	Pesan error muncul	Valid
10	Analisis Data	Upload file CSV valid dan proses analisis	Hasil analisis dan visualisasi muncul	Hasil analisis berhasil muncul dan visualisasi berhasil muncul	Valid
11	Analisis Data	Upload file CSV tanpa kolom teks	Pesan error validasi muncul	Pesan error validasi "Silakan pilih kolom teks dan konfirmasi." berhasil muncul	Valid
12	Prediksi Sentimen	Input teks ulasan valid	Hasil prediksi sentimen dan confidence muncul	Hasil prediksi dan confidence berhasil muncul	Valid
13	Prediksi Sentimen	Input teks kosong	Pesan error input kosong muncul	Pesan error " Silakan masukkan teks terlebih dahulu untuk diprediksi." muncul	Valid

Berdasarkan data pengujian dari tabel diatas dapat ditarik kesimpulan bahwa seluruh 13 kasus uji yang dijalankan dinyatakan "Valid" dengan skenario 1 sampai 9 terkait autentikasi aplikasi dan skenario 10-13 terkait tools aplikasi. Dari hasil tersebut mengindikasikan bahwa setiap fungsionalitas yang diuji telah beroperasi sesuai dengan hasil yang diharapkan.

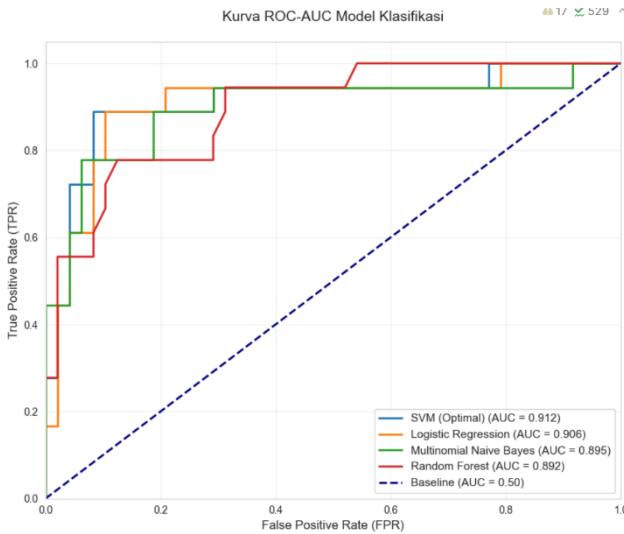
IV.2.2. Perbandingan Performa Model Algoritma SVM dengan Algoritma lain

Untuk memvalidasi efektivitas pemilihan algoritma Support Vector Machine (SVM) dalam penelitian ini dilakukan analisis perbandingan dengan beberapa algoritma machine learning dengan tipe supervised learning yang banyak digunakan pada penelitian terkait klasifikasi sentimen. Algoritma perbandingan yang diuji meliputi Logistic Regression, Multinomial Naive Bayes dan Random Forest dan untuk memastikan perbandingan dilakukan secara objektif, setiap algoritma yang dilatih menggunakan teknik evaluasi serta konfigurasi yang sama. Hasil evaluasi setiap model pada data uji dapat dilihat pada Tabel. Berikut.

Tabel IV. 16 Hasil Perbandingn Performa Model Algoritma

Model Algoritma	Accuracy	Precision	Recall	F1-Score
Support Vector Machine	0.89	0.86	0.72	0.85
Logistic Regression	0.84	0.83	0.55	0.78
Naive Bayes	0.86	0.80	0.66	0.81
Random Forest	0.83	0.70	0.66	0.78

Gambar.22 memberikan perspetif visual yang menggambarkan performa masing-masing model dalam melakukan klasifikasi dan pada kurva ini dapat dilihat sejauh mana setiap model mampu membedakan ulasan positif dan ulasan negatif. Nilai AUC (Area Under the Curve) yang menyertai setiap kurva merepresentasikan skor agregat dari kemampuan diskriminatif model.



Gambar IV. 26 ROC-AUC Perbandingan Performa Model Algoritma

Berdasarkan hasil visualisasi kurva ROC-AUC diatas dapat disimpulkan bahwa algoritma Support Vector Machine (SVM) menunjukkan performa klasifikasi terbaik dengan nilai AUC tertinggi sebesar 0,912 melampaui algoritma Logistic Regression dengan nilai AUC sebesar 0.906, Naive bayes AUC = 0.985 dan Random Forest dengan nilai AUC 0.892. Hal ini memperkuat pemilihan algoritma SVM dalam penelitian ini terbukti karena memiliki kemampuan yang unggul dalam mengklasifikasikan ulasan positif dan negatif pada dataset ulasan pengguna layanan goride.

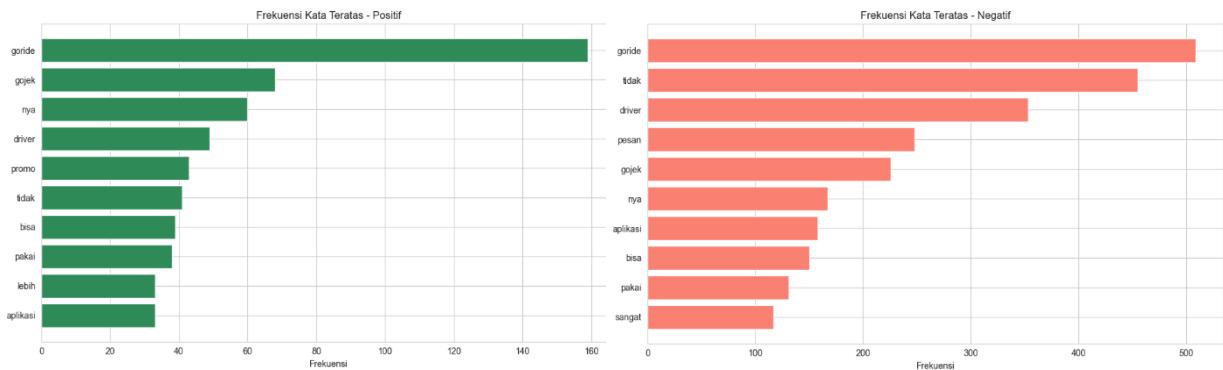
IV.2.3. Visualisasi Diagram dan Wordcloud Sentiment Ulasan

Sebagai bagian dari tahapan interpretasi hasil analisis sentimen visualisasi dalam bentuk diagram dan wordcloud digunakan untuk menggambarkan kata-kata yang paling sering muncul pada masing-masing kategori sentimen baik positif atau negatif. Tujuan dari visualisasi ini adalah untuk memperoleh pemahaman yang lebih mendalam terhadap karakteristik opini pengguna berdasarkan frekuensi kemunculan kata dalam ulasan. Dengan melihat frekuensi kemunculan kata pada

diagram dan konteks kata dalam wordcloud dalam visual dapat menjadi alat bantu yang penting dalam menafsirkan persepsi dan pengalaman pengguna terhadap layanan goride.



Gambar IV. 27 Visualisasi Wordcloud Sentimen Positif dan Negatif



Gambar IV. 28 Visualisasi Diagram (Bigram) Frekuensi Kata Sentimen Negatif dan Positif

Berdasarkan hasil visualisasi wordcloud sentimen pada Gambar IV.23 dan visualisasi frekuensi kata sentimen pada Gambar IV.24 persepsi penggunaan layanan GoRide dapat diidentifikasi secara jelas. Pada sentimen positif aspek yang paling menonjol berkaitan dengan nilai ekonomis dan efisiensi operasional layanan. Hal ini dibuktikan dari tingginya frekuensi kemunculan kata seperti "promo", "bantu" dan "baik" yang menunjukkan kepuasan dalam penggunaan layanan. Dominasi kata "goride", "gojek" dan "driver" pada wordcloud positif mengindikasi bahwa pengalaman menyeluruh baik dari sisi aplikasi maupun mitra

pengemudi yang memberikan kontribusi besar terhadap persepsi positif pengguna. Sementara itu pada sentimen negatif ditemukan berbagai permasalahan yang berkaitan dengan ketidakefesienan layanan dan gangguan teknis. Kata-kata seperti "lama", "tunggu", "batal" dan "susah" secara konsisten muncul yang mencerminkan keluhan pengguna terhadap waktu tunggu yang lama dan kesulitan dalam proses pemesanan. Frekuensi kata yang tinggi juga terlihat pada kata "aplikasi" dan "driver" diwordcloud negatif menegaskan bahwa permasalahan pada platform serta performa pengemudi menjadi faktor utama penyebab ketidakpuasan pengguna.

Dengan demikian hasil penelitian ini berhasil menjawab rumusan masalah penelitian mengenai identifikasi aspek dominan dari sentimen positif dan negatif utama dalam ulasan pengguna layanan goride. Temuan ini memberikan wawasan strategis bagi Gojek Indonesia untuk meningkatkan kualitas layanan secara menyeluruh. Dari sisi positif promosi yang sering ditawarkan Gojek serta performa driver yang baik terbukti menjadi faktor pendorong kepuasan dan ketertarikan pengguna menggunakan layanan goride. Sedangkan sentimen negatif menyoroti perlunya perbaikan pada aspek waktu tunggu dan kemudahan dalam proses pemesanan serta aktivitas user di platform aplikasi Gojek sebagai area krusial yang perlu diatasi. Dengan memperbaiki area-area tersebut Gojek Indonesia berpotensi besar untuk meningkatkan kualitas layanan, menjaga keunggulan kompetitif platform serta meningkatkan loyalitas pengguna di tengah persaingan layanan transportasi digital yang semakin ketat.

BAB V

KESIMPULAN DAN SARAN

V.1. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, maka diperoleh kesimpulan sebagai berikut :

1. Metode klasifikasi *Support Vector Machine (SVM)* yang diterapkan melalui kerangka kerja *CRISP-DM* terbukti berhasil mengolah 659 data ulasan goride yang tidak terstruktur dan secara efektif mengklasifikasikan kedalam kategori sentimen positif dan negatif.
2. Metode Support Vector Machine (SVM) yang dikembangkan menunjukkan performa klasifikasi yang baik dengan hasil evaluasi. Mencakup tingkat accuracy 89%, precision 86%, recall 72% dan F1-Score 85%. Performa algoritma SVM terbukti lebih unggul dibandingkan dengan model algoritma klasifikasi seperti *Logistic Regression*, *Multinomial Naive Bayes* serta *Decision Tree* dan model ini juga berhasil diimplementasikan menjadi aplikasi fungsional bernama "SentimenGo" untuk merepresentasikan visual analisis dan prediksi sentimen.
3. Hasil analisis sentimen berhasil mengidentifikasi aspek-aspek kunci yang mempengaruhi persepsi pengguna berdasarkan visual diagram dan worcloud. Sentiment positif sangat dipengaruhi oleh aspek penawaran harga dan efisiensi layanan yang ditandai dengan kata seperti "promo" dan "bantu". Sebaliknya sentimen negatif berpusat pada aspek waktu dan

kendala teknik aplikasi yang tercermin dari kata seperti "lama", "susah", "aplikasi" dan "batal"

V.2. Saran

Berdasarkan temuan dan kesimpulan yang dipaparkan, beberapa saran yang diharapkan dapat memberikan kontribusi praktis bagi pengembangan objek penelitian serta menjadi acuan bagi peneliti selanjutnya :

1. Eksplorasi model algoritma deep learning seperti *LSTM*, *XGBoost* dan model *pre-trained IndoBert* perlu ditingkatkan untuk memperoleh hasil analisis sentimen yang lebih optimal untuk ulasan bahasa indonesia.
2. Untuk mencegah *overfitting* dengan meningkatkan jumlah dataset dan menyeimbangkan proporsi label. Penelitian selanjutnya bisa memperoleh data dari platform lain seperti *kaggle*, *instagram* atau *twitter*.
3. Melakukan eksperimen analisis sentimen multi-bahasa untuk hasil analisis yang bervariasi dan menemukan teknik modelling yang efektif.
4. Merancang dashboard analisis sentiment real-time dengan integrasi *API Google* atau *Twitter* untuk menampilkan hasil yang selalu diperbarui.

DAFTAR PUSTAKA

- Adhi Putra, A. D. (2021). Sentiment Analysis on User Reviews of the Bibit and Bareksa Application with the KNN Algorithm. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(2), 636–646.
- Alamsyah, A. S., Budiman, S. N., & Romadhona, R. D. (2024). Penerapan Algoritma Support Vector Machine Untuk Analisis Sentimen Data Ulasan Aplikasi Binance Pada Google Play Store. *Journal Of Social Science Research*, 4(5), 6145–6158.
- Dashtipour, K., Gogate, M., Adeel, A., Larijani, H., & Hussain, A. (2021). Sentiment Analysis of English Movie Reviews using Deep Learning. *ACM International Conference Proceeding Series*, 23(5), 62–69.
<https://doi.org/10.1145/3677779.3677789>
- Dwianto, E., & Sadikin, M. (2021). Analisis Sentimen Transportasi Online pada Twitter Menggunakan Metode Klasifikasi Naïve Bayes dan Support Vector Machine. *Format : Jurnal Ilmiah Teknik Informatika*, 10(1), 94.
<https://doi.org/10.22441/format.2021.v10.i1.009>
- Hakim, A. R. ... Syarif, A. R. (2023). Analisis Perbandingan Algoritma Machine Learning Terhadap Sentimen Analis Pemindahan Ibu Kota Negara. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 7(2), 179–185.
<https://doi.org/10.35870/jtik.v7i2.701>
- Harimurti, C., & Sofyan, M. (2022). The Merging of Gojek and Tokopedia Forms Go To. Is This a Threat to National Banking and Fintech? *Management, Business and Social Science (IJEMBIS) Peer-Reviewed-International Journal*, 2(3).
<https://cvodis.com/ijembis/index.php/ijembishttps://cvodis.com/ijembis/index.php/ijembis/article/view/380>
- Haviana, S. F. C., & Poetro, B. S. W. (2022). Deep Learning Model for Sentiment

- Analysis on Short Informal Texts. *Indonesian Journal of Electrical Engineering and Informatics*, 10(1), 82–89.
<https://doi.org/10.52549/ijeei.v10i1.3181>
- Idris, I. S. K., Mustofa, Y. A., & Salihi, I. A. (2023). Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM). *Jambura Journal of Electrical and Electronics Engineering*, 5(1), 32–35. <https://doi.org/10.37905/jjeee.v5i1.16830>
- Indra, N., Tusholihah, M., Setianingsih, A., & Lusiana, D. (2020). Peluang dan Tantangan Perencanaan Pembentukan Koperasi Mitra Gojek di Era Milenial. *Jurnal Ilmiah Akuntansi Dan Keuangan*, 2(2), 223–253.
- Iskandar, J. W., & Nataliani, Y. (2021). Perbandingan Naïve Bayes, SVM, dan k-NN untuk Analisis Sentimen Gadget Berbasis Aspek. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(6), 1120–1126.
<https://doi.org/10.29207/resti.v5i6.3588>
- Joergensen E Munthe, C., Astuti Hasibuan, N., & Hutabarat, H. (2022). Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace. *Resolusi : Rekayasa Teknik Informatika Dan Informasi*, 2(3), 110–115. <https://doi.org/10.30865/resolusi.v2i3.309>
- Katadata.co.id. (2022). *Riset INDEF: Gojek Pimpin Pasar Taksi dan Ojek Online Tahun Ini*. 2022-12-07.
<https://katadata.co.id/digital/startup/63903b6aec7e5/riset-indef-gojek-pimpin-pasar-taksi-dan-ojek-online-tahun-ini>
- Mahawardana, P. P. O., Imawati, I. A. P. F., & Dika, I. W. (2022). Analisis Sentimen Berdasarkan Opini dari Media Sosial Twitter terhadap “Figure Pemimpin” Menggunakan Python. *Jurnal Manajemen Dan Teknologi Informasi*, 12(2), 50–56.
<https://ojs.mahadewa.ac.id/index.php/jmti/article/view/2111>
- Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications,

and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, 36(4).

<https://doi.org/10.1016/j.jksuci.2024.102048>

Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 4(2), 375–384.

<https://doi.org/10.57152/malcom.v4i2.1206>

Mehta, P., & Pandya, S. (2020). A Review On Sentiment Analysis Methodologies, Practices And Applications. *International Journal of Scientific and Technology Research*, 9(2), 601–609.

Pramudita, D., Akbar, Y., & Wahyudi, T. (2024). MALCOM: Indonesian Journal of Machine Learning and Computer Science Sentiment Analysis of the Indonesian Smart College Card Program on Social Media X Using the Naive Bayes Algorithm Analisis Sentimen Terhadap Program Kartu Indonesia Pintar Kuliah Pada Med. *Malcom*, 4(October), 1420–1430.

Pranata, A., Budianita, E., Yusra, & Cynthia, E. P. (2022). Klasifikasi Sentimen Terhadap Maxim Menggunakan Algoritma SVM Pada Media Sosial Twitter. *Klasifikasi Sentimen Terhadap Maxim Menggunakan Algoritma SVM Pada Media Sosial Twitter*, 5(3), 332–341.

<https://ojs.serambimekkah.ac.id/jnkti/article/view/4253>

Reza Pahlawan, M., Setyanto, A., & Rudyanto Arief, M. (2024). Review Komprehensif Penggunaan Data Imbalance Pada Metode Klasifikasi Dalam Machine Learning. *Journal Homepage: Journal of Electrical Engineering and Computer (JEECOM)*, xx, No. xx(xx).

<https://doi.org/10.33650/jecom.v4i2>

Rizki, M. F., Auliasari, K., & Primaswara Prasetya, R. (2021). Analisis Sentiment Cyberbullying Pada Sosial Media Twitter Menggunakan Metode Support Vector Machine. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 5(2), 548–

556. <https://doi.org/10.36040/jati.v5i2.3808>

Rokhman, K. A., Berlilana, B., & Arsi, P. (2021). Perbandingan Metode Support Vector Machine Dan Decision Tree Untuk Analisis Sentimen Review Komentar Pada Aplikasi Transportasi Online. *Journal of Information System Management (JOISM)*, 3(1), 1–7.
<https://doi.org/10.24076/joism.2021v3i1.341>

Utami, M. A. A. T., Silvianti, P., & Masjkur, M. (2023). Algoritme Support Vector Machine untuk Analisis Sentimen Berbasis Aspek Ulasan Game Online Mobile Legends: Bang-Bang. *Xplore: Journal of Statistics*, 12(1), 63–77.
<https://doi.org/10.29244/xplore.v12i1.1064>

