

**ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP
APLIKASI MIFX DI PLAY STORE MENGGUNAKAN
PENDEKATAN VADER DAN METODE NAÏVE BAYES**

SKRIPSI

DISUSUN OLEH

NAZRI ADLANI

NPM. 2109010146



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2025

**ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP
APLIKASI MIFX DI PLAY STORE MENGGUNAKAN
PENDEKATAN VADER DAN METODE NAÏVE BAYES**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Informasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

NAZRI ADLANI

NPM. 2109010146

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2025

LEMBAR PENGESAHAN

Judul Skripsi : ANALISIS SENTIMEN ULASAN PENGGUNA
TERHADAP APLIKASI MIFX DI PLAY STORE
MENGUNAKAN PENDEKATAN VADER DAN
METODE NAÏVE BAYES

Nama Mahasiswa : NAZRI ADLANI

NPM : 2109010146

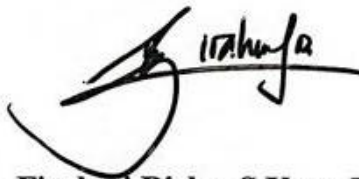
Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

Ketua Program Studi



(Dr. Firahmi Rizky, S.Kom., M.Kom.)
NIDN. 0116079201

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP APLIKASI MIFX DI PLAY STORE MENGGUNAKAN PENDEKATAN VADER DAN METODE NAÏVE BAYES

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 28 Oktober 2025

Yang membuat pernyataan



Nazri Adlani

NPM. 2109010146

**PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama	: Nazri Adlani
NPM	: 2109010146
Program Studi	: Sistem Informasi
Karya Ilmiah	: Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP APLIKASI MIFX
DI PLAY STORE MENGGUNAKAN PENDEKATAN VADER DAN
METODE NAÏVE BAYES**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 28 Oktober 2025

Yang membuat pernyataan



Nazri Adlani

NPM. 2109010146

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Nazri Adlani
Tempat dan Tanggal Lahir : Sidodadi, 28 Maret 2003
Alamat Rumah : Desa Meranti
Telepon/Faks/HP : 081260684425
E-mail : nazriadlani28@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : NEGERI 118170 KAMPUNG BANGUN TAMAT: 2015
SMP : NEGERI 3 BILAH HULU TAMAT: 2018
SMA : NEGERI 1 BILAH HULU TAMAT: 2021

KATA PENGANTAR



Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, karunia, serta kekuatan yang diberikan sehingga penulis dapat menyelesaikan tugas akhir dalam menempuh pendidikan Strata 1. Skripsi ini disusun dalam bentuk buku sederhana dengan judul: **“Analisis Sentimen Ulasan Pengguna terhadap Aplikasi MIFX di Play Store Menggunakan Pendekatan VADER dan Metode Naïve Bayes.”** Penyusunan skripsi ini merupakan salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara. Isi dan pembahasan dalam skripsi ini disusun berdasarkan hasil penelitian yang dilakukan penulis serta didukung oleh berbagai sumber literatur yang relevan. Penulis menyadari bahwa penyusunan skripsi ini tidak akan terselesaikan tanpa adanya bimbingan, dukungan, dan doa dari berbagai pihak.

Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi. Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Bapak Halim Maulana, S.T, M.Kom sebagai Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
4. Bapak Dr. Lutfi Basit, S.Sos, M.I.Kom sebagai Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
5. Ibu Dr.Firahmi Rizky, S.Kom.,M.Kom, selaku Ketua Program Studi Sistem Informasi
6. Bapak Mahardika Abdi Prawira Tanjung, S.Kom.,M.Kom, selaku Sekretaris Program Studi Sistem Informasi

7. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. selaku dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan berharga selama proses penyusunan skripsi ini.
8. Penulis menyampaikan terima kasih yang sebesar-besarnya kepada Ayahanda Amat Muksin dan Ibunda Jaitun atas segala doa, dukungan, dan pengorbanan yang telah diberikan selama penulis menempuh pendidikan. Kasih sayang dan semangat dari keduanya menjadi dorongan penting dalam proses penyelesaian skripsi ini.
9. Penulis mengucapkan terima kasih kepada abang dan kakak tercinta atas segala dukungan, perhatian, dan doa yang telah diberikan. Semangat dan motivasi dari kalian menjadi dorongan berharga dalam penyelesaian skripsi ini.
10. Penulis juga mengucapkan terima kasih kepada keluarga di perantauan yang sekaligus menjadi teman seperjuangan selama masa perkuliahan, yaitu Arya, Tigor, Athary (Aji), Zaini, Tio, Fiqri. Serta teman-teman yang tidak dapat disebutkan satu per satu atas kebersamaan dan dukungan yang telah membantu penulis menyelesaikan skripsi ini.
11. Semua pihak yang terlibat langsung ataupun tidak langsung yang tidak dapat penulis ucapkan satu-persatu yang telah membantu penyelesaian skripsi ini.
12. Penulis juga mengucapkan terima kasih kepada diri sendiri atas usaha, ketekunan, dan kesabaran selama proses penyusunan skripsi ini. Berbagai tantangan yang dihadapi telah menjadi pengalaman berharga yang mengajarkan arti konsistensi dan keyakinan diri dalam mencapai tujuan.

Medan,

Hormat Penulis



Nazri Adlani

NPM. 2109010146

ANALISIS SENTIMEN ULASAN PENGGUNA TERHADAP APLIKASI MIFX DI PLAY STORE MENGGUNAKAN PENDEKATAN VADER DAN METODE NAÏVE BAYES

ABSTRAK

Pertumbuhan aplikasi keuangan digital di Indonesia terus meningkat, termasuk aplikasi MIFX yang digunakan untuk aktivitas trading forex secara online. Banyaknya ulasan pengguna di Google Play Store menjadi sumber penting untuk menilai kualitas layanan dan pengalaman pengguna terhadap aplikasi tersebut. Namun, bentuk ulasan yang tidak terstruktur dan menggunakan bahasa informal membuat analisis manual menjadi sulit dilakukan. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap aplikasi MIFX dengan menggabungkan metode VADER dan Naïve Bayes. Data ulasan dikumpulkan menggunakan teknik web scraping dari Google Play Store, kemudian melalui tahapan preprocessing seperti cleaning, case folding, normalisasi, translasi ke Bahasa Inggris, tokenisasi, stopword removal, dan stemming. Metode VADER digunakan untuk memberikan label sentimen positif dan negatif, sedangkan Naïve Bayes digunakan untuk melakukan klasifikasi berdasarkan pembobotan TF-IDF. Hasil evaluasi menunjukkan bahwa model Naïve Bayes mampu mencapai akurasi sebesar 85,48%, dengan performa terbaik pada sentimen positif (precision 90%, recall 91%, F1-score 90%) dan performa lebih rendah pada sentimen negatif (precision 72%, recall 67%, F1-score 70%). Berdasarkan hasil pelabelan menggunakan VADER, dari 3.409 ulasan yang dianalisis, sebanyak 76,3% termasuk kategori positif dan 23,7% negatif. Hasil ini menunjukkan bahwa persepsi mayoritas pengguna terhadap aplikasi MIFX cenderung positif. Seluruh alur penelitian juga diimplementasikan ke dalam sebuah prototipe aplikasi web yang mampu menampilkan hasil analisis sentimen secara interaktif dan real-time sebagai bentuk implementasi sistem.

Kata Kunci: Analisis Sentimen; MIFX; VADER; Naïve Bayes.

SENTIMENT ANALYSIS OF USER REVIEWS ON THE MIFX APPLICATION IN THE PLAY STORE USING THE VADER APPROACH AND NAÏVE BAYES METHOD

ABSTRACT

The growth of digital financial applications in Indonesia continues to increase, including MIFX, an online trading platform for forex transactions. Numerous user reviews on the Google Play Store serve as valuable data sources to evaluate the service quality and user experience of the application. However, these reviews are often unstructured and written in informal language, making manual analysis difficult. This study aims to analyze user sentiments toward the MIFX application by combining the VADER and Naïve Bayes methods. Review data were collected using web scraping techniques from the Google Play Store, followed by several preprocessing stages, including cleaning, case folding, normalization, English translation, tokenization, stopword removal, and stemming. The VADER method was applied to assign positive and negative sentiment labels, while Naïve Bayes performed sentiment classification based on TF-IDF weighting. The evaluation results showed that the Naïve Bayes model achieved an accuracy of 85.48%, with the best performance on positive sentiment (precision 90%, recall 91%, F1-score 90%) and lower performance on negative sentiment (precision 72%, recall 67%, F1-score 70%). Based on VADER labeling, from 3,409 analyzed reviews, 76.3% were identified as positive and 23.7% as negative, indicating that most users have a positive perception of the MIFX application. The entire research workflow was also implemented into a functional web prototype capable of displaying sentiment analysis results interactively and in real time as a practical system implementation.

Keywords: Sentiment Analysis; MIFX; VADER; Naïve Bayes; TF-IDF.

DAFTAR ISI

DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR	xiv
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
BAB II.....	6
LANDASAN TEORI	6
2.1. MIFX	6
2.2. Play Store	7
2.3. Analisis Sentimen.....	8
2.4. Pre-Processing	11
2.5. Naive Bayes.....	13
2.6. VADER (Valence Aware Dictionary and sEntiment Reasoner).....	15
2.7. Algoritma TF-IDF	17
2.8. Confusion Matrix	18
2.9. Google Colab.....	19
2.10. Bahasa Pemrograman Python	20
BAB III	22
METODOLOGI PENELITIAN.....	22
3.1. Alur Penelitian.....	22
3.2. Pengumpulan Data	23
3.2.1 Sumber data.....	23
3.2.2 Teknik Pengambilan Data	23
3.2.3 Jenis Data	24
3.3. Preprocessing Data	24
3.3.1 Preprocessing untuk VADER	24
3.3.2 Preprocessing untuk Naive Bayes	26

3.4. Hipotesis Penelitian	29
3.4.1. Metode Vader (Valence Aware Dictionary dan sEntiment Reasoner) 29	
3.4.2. Metode Naive Bayes	31
3.5. Penelitian Terdahulu.....	34
3.6. Model Perancangan Sistem	37
3.6.1. Model SDLC Waterfall	37
3.6.2. Pemodelan UML (Unified Modeling Language).....	41
BAB IV	46
HASIL DAN PEMBAHASAN.....	46
4.1. Scraping Data	46
4.2. Hasil Pre-Processing Ringan	47
4.3. Labeling dengan VADER	52
4.4. Hasil Pre-Processing Lanjutan	54
4.5. Pembobotan TF-IDF.....	57
4.6. Split Data	58
4.7. Klasifikasi Dengan Naïve bayes.....	59
4.8. Hasil Evaluasi.....	61
4.9. Hasil Visualisasi	63
4.10. Implementasi Sistem.....	65
BAB V.....	71
PENUTUP.....	71
5.1. Kesimpulan.....	71
5.2. Saran	72
DAFTAR PUSTAKA	74

DAFTAR TABEL

Tabel 3. 1 Cleaning	25
Tabel 3. 2 Case Folding	25
Tabel 3. 3 Normalisasi Kata.....	26
Tabel 3. 4 Tokenizing	27
Tabel 3. 5 Stopword Removal/Filtering.....	28
Tabel 3. 6 Stemming	28
Tabel 3. 7 Penelitian Terdahulu	34
Tabel 3. 8 Perbedaan Vader dan Naive Bayes	36
Tabel 4. 1 Cofusion Matrix Tanpa Nilai	61
Tabel 4. 2 Confusion Matrix Keseluruhan.....	61
Tabel 4. 3 Confusion Matrix Kelas Positif	62
Tabel 4. 4 Confusion Matrix Kelas Negatif	62

DAFTAR GAMBAR

Gambar 3. 1 Flowchart Algoritma Vader	30
Gambar 3. 2 Flowchart Algoritma Naive Bayes	32
Gambar 3. 3 Model SDLC Waterfall	37
Gambar 3. 4 Tabel Struktur Basis Data	40
Gambar 3. 5 Diagram Use Case Prototipe Sistem	42
Gambar 3. 6 Diagram Activity	43
Gambar 3. 7 Sequence Diagram	44
Gambar 3. 8 Class Diagram	45
Gambar 4. 1 Hasil Scarping data (data Frame)	46
Gambar 4. 2 Data Hasil Scraping (CSV)	47
Gambar 4. 3 Import data dari file Csv ke DataFrame	48
Gambar 4. 4 Hasil Clening	48
Gambar 4. 5 Hasil Case Folding	49
Gambar 4. 6 Instalasi Kaggle	49
Gambar 4. 7 Kamus Kata Baku	50
Gambar 4. 8 Hasil Normalisasi Kata	50
Gambar 4. 9 Instalasi deep-translator	51
Gambar 4. 10 Data Hasil Translit Pada Kolom "english"	51
Gambar 4. 11 HasilPembersihan Data Duplikat	52
Gambar 4. 12 Instal Vader Sentiment	52
Gambar 4. 13 Hasil Klasifikasi Sentimen dengan VADER	53
Gambar 4. 14 Tabel Distribusi Sentimen VADER	53
Gambar 4. 15 Source Code dan Hasil Tokenizing	54
Gambar 4. 16 Source Code dan Hasil Stopword	55
Gambar 4. 17 Hasil Stemming	55
Gambar 4. 18 Diagram Frekuensi Kemunculan Kata	56
Gambar 4. 19 Word Cloud Stemming	56
Gambar 4. 20 Hasil TF IDF	57
Gambar 4. 21 TF-IDF bentuk Matrix	58
Gambar 4. 22 Source Code dan Hasil Split Data	58
Gambar 4. 23 Diagram Distribusi Data Training dan Testing	59
Gambar 4. 24 Hasil Classification Report	59
Gambar 4. 25 Confusion Matrix	60
Gambar 4. 26 Hasil Evaluasi	63
Gambar 4. 27 WordCloud Sentimen Positif	64
Gambar 4. 28 WordCloud Sentimen Negatif	64
Gambar 4. 29 Grafik Area Bertumpuk Tren Sentimen	65
Gambar 4. 30 Tampilan Halaman Utama	66
Gambar 4. 31 Tampilan Scraping	67
Gambar 4. 32 Tampilan Halaman Preprocessing	68
Gambar 4. 33 Tampilan Translit dan Pelabelan Vader	68
Gambar 4. 34 Diagram Lingkaran Distribusi Sentimen	69
Gambar 4. 35 Hasil Evaluasi Model Naïve Bayes	69
Gambar 4. 36 Tampilan tabel "data_ulasan" pada phpMyAdmin	70

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Dalam era digital saat ini, penggunaan aplikasi keuangan dan investasi berbasis mobile mengalami pertumbuhan yang sangat pesat. Salah satu sektor yang menunjukkan perkembangan signifikan adalah platform trading forex, yang memungkinkan pengguna untuk melakukan transaksi secara real-time melalui perangkat digital (Kusnawi & Anam, 2025). Di Indonesia, salah satu aplikasi trading yang cukup populer adalah MIFX, milik broker resmi Monex Investindo Futures. Aplikasi ini menyediakan layanan perdagangan aset seperti forex, emas, dan indeks dengan antarmuka yang ramah pengguna dan fitur yang terus diperbarui (Sofyanty et al., 2022).

Seiring dengan meningkatnya jumlah pengguna aplikasi MIFX, semakin banyak pula ulasan dan penilaian yang diberikan oleh pengguna melalui Google Play Store, yang merupakan salah satu platform distribusi aplikasi terbesar. Ulasan tersebut mencerminkan berbagai pengalaman pengguna, mulai dari kepuasan terhadap fitur, performa aplikasi, hingga keluhan terhadap layanan yang disediakan. Namun, ulasan-ulasan tersebut umumnya berbentuk teks bebas yang tidak terstruktur, menggunakan bahasa informal, serta sering kali mengandung istilah teknis dalam dunia trading, seperti buy, sell, margin call, dan sebagainya (Barik & Misra, 2024).

Bentuk ulasan yang demikian menyulitkan proses analisis secara manual, terutama dalam jumlah besar. Di sisi lain, ulasan pengguna merupakan sumber data yang sangat penting untuk mengetahui persepsi publik terhadap performa dan

kualitas layanan aplikasi. Oleh karena itu, diperlukan suatu pendekatan analitis yang dapat mengolah data tersebut secara efisien dan akurat.

Beberapa masalah yang sering ditemukan dalam ulasan mencakup gangguan saat membuka posisi (open position), kendala saat melakukan penarikan dana (withdrawal), lambatnya respon dari layanan pelanggan, serta error saat login atau saat mengeksekusi order. Keluhan-keluhan tersebut menunjukkan bahwa ada aspek dalam layanan aplikasi yang perlu diperbaiki, dan menganalisis sentimen dari ulasan ini dapat memberikan wawasan yang lebih dalam terkait kebutuhan serta kepuasan pengguna.

Oleh karena itu, penelitian ini dilandasi oleh kebutuhan untuk mengolah data teks ulasan pengguna secara efisien dan akurat, khususnya dalam konteks analisis sentiment. Untuk menjawab tantangan tersebut, penelitian ini menerapkan analisis sentimen menggunakan kombinasi dua pendekatan, yaitu VADER (Valence Aware Dictionary and sEntiment Reasoner) dan Naïve Bayes Classifier. VADER merupakan metode berbasis leksikon yang sangat cocok untuk menganalisis teks pendek dan informal, seperti komentar atau ulasan aplikasi dan Naïve Bayes merupakan algoritma pembelajaran mesin yang efisien untuk klasifikasi teks berdasarkan probabilitas (Susanti & Ilahi, 2024).

Kombinasi kedua metode ini diharapkan dapat saling melengkapi: VADER memberikan analisis berbasis kamus emosi, sedangkan Naïve Bayes mampu belajar dari pola data yang telah dilabeli. Meskipun analisis sentimen telah banyak diterapkan dalam konteks ulasan digital, penelitian yang secara spesifik menggabungkan kedua pendekatan ini untuk menganalisis ulasan pengguna pada aplikasi trading lokal seperti MIFX.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan metode analisis sentimen di Indonesia, khususnya pada aplikasi keuangan lokal. Selain itu, hasil analisis ini juga dapat menjadi masukan berharga bagi pengembang dan penyedia layanan dalam meningkatkan kualitas aplikasi MIFX berdasarkan persepsi nyata dari para penggunanya.

1.2. Rumusan Masalah

Banyaknya ulasan pengguna aplikasi MIFX di Google Play Store yang bersifat tidak terstruktur dan menggunakan bahasa informal menghadirkan tantangan dalam proses analisis sentimen secara manual. Oleh karena itu, diperlukan pendekatan otomatis yang mampu mengklasifikasikan sentimen secara akurat. Penelitian ini mengangkat permasalahan mengenai efektivitas kombinasi metode VADER dan Naïve Bayes dalam melakukan klasifikasi sentimen secara otomatis terhadap ulasan yang telah melalui proses translasi dari Bahasa Indonesia ke Bahasa Inggris. Selain itu, penelitian ini juga mengkaji permasalahan bagaimana hasil dari model klasifikasi tersebut dapat diimplementasikan ke dalam sebuah prototipe aplikasi web yang fungsional dan interaktif. Masalah utama yang dikaji adalah sejauh mana pendekatan gabungan ini dapat berfungsi secara efektif dalam melakukan klasifikasi sentimen, yang dievaluasi menggunakan metrik akurasi, presisi, recall, dan F1-score sebagai tolok ukur performa model.

1.3. Batasan Masalah

Adapun batasan masalah dalam melakukan penelitian ini adalah:

1. Data yang digunakan bersumber dari Google Play Store.

2. Analisa dilakukan kepada ulasan Aplikasi MIFX dan berbahasa Indonesia.
3. Hasil analisis berupa klasifikasi positif dan negative.
4. Metode yang digunakan adalah Vader dan Naive Bayes.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk menguji efektivitas kombinasi metode VADER dan Naïve Bayes dalam melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi MIFX di Google Play Store. Ulasan yang digunakan ditulis dalam Bahasa Indonesia dan mengandung gaya bahasa informal, sehingga perlu melalui proses translasi ke Bahasa Inggris sebelum dianalisis oleh VADER. Hasil label sentimen dari VADER digunakan sebagai acuan pelatihan dalam model Naïve Bayes. Evaluasi dilakukan untuk menilai performa kombinasi metode tersebut menggunakan metrik klasifikasi seperti akurasi, presisi, recall, dan F1-score guna memastikan ketepatan dan keandalan analisis yang dihasilkan. Selanjutnya, sebagai implementasi dari hasil penelitian, tujuan lainnya adalah membangun sebuah prototipe aplikasi web yang fungsional untuk mendemonstrasikan kemampuan model dalam menganalisis data baru secara langsung..

1.5. Manfaat Penelitian

Penelitian ini memberikan gambaran penerapan kombinasi pendekatan leksikon (VADER) dan pembelajaran mesin (Naive Bayes) dalam mengolah data ulasan berbahasa Indonesia secara otomatis. Metode yang digunakan menunjukkan potensi penerapan gabungan teknik dalam klasifikasi opini berbasis teks digital yang tidak terstruktur. Hasil dari pendekatan ini dapat memberikan pemahaman

terhadap penyebaran sentimen dan menjadi referensi awal bagi studi lanjutan yang tertarik mengeksplorasi integrasi pendekatan leksikon dan pembelajaran mesin dalam konteks serupa.

BAB II

LANDASAN TEORI

2.1. MIFX

MIFX Monex adalah PT Monex Investindo Futures yang merupakan perusahaan pialang berjangka yang memfasilitasi perdagangan forex, komoditas, dan indeks saham. Didirikan pada tahun 2000, dan para investor ini memerlukan broker untuk dapat melakukan jual beli mata uang asing. Mudahnya, pengertian broker forex adalah perusahaan penyedia layanan keuangan yang menjadi pihak perantara bagi kliennya untuk saling memperdagangkan mata uang asing (Monex Investindo Futures, 2025) (Alfiah, 2022).

Broker MIFX memiliki beberapa keunggulan seperti sudah mengantongi izin dari Bappebti dengan nomor izin 178/BAPPEBT/SI/I/2003, punya kantor oprasional atau jaringan yang luas diberbagai kota di Indonesia seperti Bandung, Denpasar, Surabaya dan lainnya., Anda bisa melakukan pendaftaran, deposit, penarikan dana, hingga melakukan transaksi jual/beli aset dengan lebih mudah karena bisa dilakukan 100% online (Febrian Surya Kristardi, 2024).

PT Monex Investindo Futures memberikan kemudahan bagi investor untuk memilih sendiri jenis mata uang yang akan diperdagangkan, transaksi yang diinginkan oleh nasabah (online melalui gadget atau laptop atau manual), waktu trading (short time, long time), orang yang akan melakukan transaksi (nasabah sendiri atau broker) hingga besaran margin yang dapat disesuaikan dengan kemampuan nasabah (Sofyanty et al., 2022).

Selain itu keunggulan dari aplikasi MIFX terletak pada fitur-fitur yang mendukung aktivitas trading secara real-time, seperti chart harga langsung dan

analisis pasar harian. Aplikasi ini juga dilengkapi dengan edukasi dan tutorial yang mendukung pengguna pemula untuk belajar mengenal dunia trading. Ketersediaan fitur tersebut menjadikan MIFX salah satu aplikasi yang cukup diminati oleh pengguna pasar modal dan forex Indonesia. Namun demikian, seiring meningkatnya jumlah pengguna, kualitas layanan dan fitur yang tersedia dalam aplikasi menjadi semakin penting untuk dievaluasi. Salah satu indikator yang dapat digunakan untuk mengetahui kepuasan pengguna adalah dengan menganalisis ulasan yang ditinggalkan oleh pengguna di platform Google Play Store. Melalui ulasan tersebut, dapat diperoleh gambaran mengenai kelebihan, kekurangan, serta saran perbaikan dari perspektif pengguna langsung (Monex Investindo Futures, 2025)

2.2. Play Store

Google Play, sebelumnya dikenal sebagai Android Market, adalah toko aplikasi resmi yang dikembangkan dan dikelola oleh Google untuk perangkat yang menjalankan sistem operasi Android. Diluncurkan pada tahun 2008, Google Play tidak hanya menyediakan aplikasi, tetapi juga menawarkan berbagai konten digital lainnya seperti musik, buku, film, dan perangkat keras tertentu. Google Play telah berkembang menjadi platform distribusi konten digital terbesar di dunia, yang digunakan oleh jutaan pengembang dan miliaran pengguna di seluruh dunia. Dengan lebih dari tiga juta aplikasi dan permainan yang tersedia, Google Play menjadi pusat ekosistem Android yang sangat vital (Siregar et al., 2025).

App markets Google Playstore menyediakan mekanisme baru untuk distribusi perangkat lunak, dimana pengguna memperoleh aplikasi langsung dari pengembang dan juga proses ulasan pengguna aplikasi dapat langsung dilakukan

melalui komentar. Ulasan pada Google Playstore memiliki rating dari 1 hingga 5. Namun sering kali pengguna memberikan rating tidak sesuai dengan ulasannya sehingga hal ini belum cukup menggambarkan kualitas dari aplikasi. Ulasan berbentuk kalimat lebih menggambarkan bagaimana tanggapan pengguna terhadap aplikasi. Hal ini dapat memungkinkan bahwa ulasan tersebut mempengaruhi pengunjung atau pengguna yang baru akan mengunduh aplikasi MIFX (Afifah Faadilah, 2020).

Analisis terhadap ulasan-ulasan ini menjadi penting untuk mengetahui sejauh mana aplikasi MIFX mampu memenuhi kebutuhan penggunanya. Penelitian ini menggunakan pendekatan analisis sentimen untuk memahami opini-opini tersebut secara sistematis, yang akan membantu pengembang dalam proses evaluasi dan pengambilan keputusan berbasis data.

2.3. Analisis Sentimen

Analisis sentimen adalah metode dalam bidang pemrosesan bahasa alami (Natural Language Processing/NLP) yang digunakan untuk mengevaluasi opini atau emosi yang terkandung dalam sebuah teks. Tujuan utama dari analisis ini adalah untuk mengidentifikasi sikap pengguna, apakah bersifat positif, negatif, atau netral terhadap suatu produk, layanan, atau fenomena tertentu. Teknik ini banyak dimanfaatkan di berbagai bidang, terutama dalam menganalisis data ulasan, media sosial, dan survei kepuasan pelanggan. Analisis sentimen telah memperluas penelitian NLP secara signifikan karena telah memperkenalkan banyak masalah penelitian yang menantang yang belum pernah dipelajari sebelumnya. Analisis sentimen merupakan masalah analisis semantik, tetapi sangat terfokus dan terbatas

karena sistem analisis sentimen tidak perlu sepenuhnya "memahami" setiap kalimat atau dokumen dan hanya perlu memahami beberapa aspek saja, misalnya opini positif dan negatif serta sasarannya. Terdapat beberapa pendekatan untuk melakukan analisis sentimen (Ernianti Hasibuan & Elmo Allistair Heriyanto, 2022).

Dalam konteks digital, analisis sentimen memiliki peran yang sangat strategis. Hal ini dikarenakan banyaknya opini pengguna yang tersedia secara daring, baik melalui aplikasi, situs web, maupun media sosial. Dengan teknik ini, perusahaan atau pengembang dapat mengevaluasi tanggapan masyarakat terhadap produk atau layanan yang mereka tawarkan tanpa harus melakukan survei manual yang memakan waktu dan biaya.

Ada beberapa tipe analisis sentimen sebagai berikut : (Sitio & Nadiyanti, 2022).

a) Analisis Sentimen Bertingkat (*Graded Sentiment Analysis*)

Analisis sentimen bertingkat merupakan tipe analisis yang memiliki penilaian spesifik. Tipe ini memungkinkan untuk memperluas kategori polaritas untuk menyertakan tingkat positif dan negatif yang berbeda yaitu: sangat positif, positif, netral, negatif, sangat negatif. Analisis sentimen bertingkat (fine-grained) dapat digunakan untuk menafsirkan bintang 5 dalam sebuah ulasan. Bintang 5 untuk menyatakan sangat positif, sedangkan bintang 1 untuk menyatakan sangat negatif.

b) Deteksi Emosi (*Emotion Detection*)

Analisis sentimen pendeteksi emosi digunakan untuk mendeteksi emosi, seperti kebahagiaan, frustrasi, kemarahan, dan kesedihan. Banyak dari sistem

pendeteksi emosi menggunakan Lexicon (sebuah daftar kata dan emosi yang disampaikan) atau algoritma machine learning yang kompleks.

c) Analisis Sentimen Berbasis Aspek (*Aspect-based Sentiment Analysis*)

Analisis sentimen berbasis aspek digunakan saat menganalisis sentimen teks untuk mengetahui aspek atau fitur tertentu yang disampaikan seseorang apakah itu positif, netral, atau negatif. Misalnya pada ulasan sebuah restoran, “Makanannya enak dan murah. Tapi pelayanan dan tempat masih sangat kurang nyaman”. Dari contoh ulasan tersebut dapat dikatakan dari aspek makanan mendapat penilaian yang positif, namun pada aspek pelayanan dan tempat mendapat penilaian negatif.

d) Analisis Sentimen Multibahasa (*Multilingual Sentiment Analysis*)

Analisis sentimen berbasis aspek digunakan saat menganalisis sentimen teks untuk mengetahui aspek atau fitur tertentu yang disampaikan seseorang apakah itu positif, netral, atau negatif. Misalnya pada ulasan sebuah restoran, “Makanannya enak dan murah. Tapi pelayanan dan tempat masih sangat kurang nyaman”. Dari contoh ulasan tersebut dapat dikatakan dari aspek makanan mendapat penilaian yang positif, namun pada aspek pelayanan dan tempat mendapat penilaian negatif.

Alasan pentingnya analisis sentimen karena manusia mulai mengekspresikan pikiran dan perasaan mereka lebih terbuka dari sebelumnya, analisis sentimen dengan cepat menjadi salah satu alat penting untuk memantau dan memahami sentimen di semua jenis data. Secara otomatis menganalisis feedback pelanggan, seperti pendapat mereka dalam tanggapan survei dan percakapan di media sosial, memungkinkan brand untuk mempelajari apa yang membuat

pelanggan senang atau tidak, sehingga brand dapat menyesuaikan produk dan layanan mereka untuk memenuhi kebutuhan pelanggan mereka (Sitio & Nadiyanti, 2022).

2.4. Pre-Processing

Dalam penelitian ini di terapkan text preprocessing untuk data yang akan digunakan dalam proses analisa sentimen, dimana data yang kita proses akan kita ambil informasi yang terkandung didalamnya. Dalam hal sentimen penulisannya yaitu negatif atau positif. Guna memudahkan dalam mengelola data maka data perlu kita berikan analisa sentimen secara manual dengan membaca maksud dari kalimat yang ada dalam sentiment tersebut, sehingga dapat diberikan penilaian bahwa sentiment tersebut merupakan sentiment negatif atau positif (R.Sanjaya, E.Tohidi, 2024).

Tahap preprocessing adalah tahap yang dipakai untuk menghapus kata atau teks yang tidak dibutuhkan dalam proses klasifikasi. Terdapat beberapa proses yang terdapat pada tahap preprocessing yang terdiri dari cleansing, remove duplicate, sleksi data, normalisasi, transform case, tokenizing, filtering, stopword, stemming, dan pemberian label (labeling) (Pramukti et al., 2022).

Dibawah ini merupakan penjelasan mengenai proses-proses yang ada pada tahap preprocessing.

1. Cleansing

Tahap cleansing atau tahap pembersihan data bertujuan untuk menghapus special karakter (simbol), nama username, retweet (RT), dan link yang terdapat didalam dataset.

2. Tokenizing

Tokenize merupakan tahap untuk memecah kalimat yang ada didalam data tweet (dataset), data yang asalnya berbentuk sebuah kalimat akan di ubah ke bentuk kata perkata.

3. Normalize

Normalisasi merupakan tahap untuk mengubah kata-kata yang terdapat pada dataset yang sebelumnya kata tersebut tidak baku menjadi kata dasar yang baku atau sesuai ejaan kata yang benar. Normalisasi dilakukan dikarenakan kebiasaan pengguna playstore yang menggunakan kata singkatan ataupun kata gaul dalam menulis sebuah ulasan.

4. Filtering

Filtering merupakan tahap menghilangkan kata-kata (setelah proses tokenize) yang tidak memiliki nilai dalam proses klasifikasi teks. Pada penelitian yang dilakukan, kata yang didalam dataset memiliki nilai kurang dari 4 karakter dan panjang melebihi 100 karakter akan hapus atau dihilangkan.

5. Stopwords

Stopwords Stopwords merupakan proses menghilangkan kata sambung yang ada didalam dataset. stopwords yang didapatkan bersumber dari library library NLTK (Natural Language Toolkit) dengan total kata berjumlah sekitar 700 kata.

6. Stemming

Stemming merupakan tahap akhir dalam tahap preprocessing, stemming adalah proses menghilangkan semua imbuhan (prefix dan suffixes) yang ada pada awal dan akhir kata (Pramukti et al., 2022).

2.5. Naive Bayes

Naive Bayes adalah algoritma probabilistik yang biasanya digunakan untuk masalah klasifikasi. Naive Bayes Classifier bukanlah algoritma tunggal, melainkan kumpulan algoritma klasifikasi di mana semuanya memiliki prinsip yang sama berdasarkan Teorema Bayes yang menggambarkan probabilitas suatu peristiwa berdasarkan pengetahuan sebelumnya atau probabilitas tertentu lainnya yang diketahui dari peristiwa itu khususnya dalam pengolahan teks. Algoritma ini bekerja berdasarkan prinsip Teorema Bayes, yang menghitung probabilitas suatu kelas berdasarkan nilai-nilai fitur yang diberikan. Istilah “naive” merujuk pada asumsi bahwa setiap fitur (misalnya kata dalam sebuah kalimat) dianggap independen terhadap fitur lainnya, meskipun dalam praktiknya hal ini tidak selalu berlaku (Ernianti Hasibuan & Elmo Allistair Heriyanto, 2022).

Dalam analisis sentimen, Naive Bayes digunakan untuk mengklasifikasikan teks ke dalam kategori seperti positif, negatif, atau netral. Model ini mempelajari pola kata dari data pelatihan, kemudian digunakan untuk memprediksi sentimen dari teks yang belum pernah dilihat sebelumnya. Kelebihan utama dari metode ini adalah kecepatan dan efisiensinya dalam memproses data dalam jumlah besar, serta kemudahan implementasinya di berbagai bahasa, termasuk Bahasa Indonesia.

Naïve Bayes termasuk merupakan algoritma klasifikasi sederhana yang menghitung sekumpulan probabilitas dengan menjumlahkan dan menggabungkan nilai dari kumpulan data yang diberikan serta didasarkan pada Teorema Bayes. Walaupun klasifikasi Algoritma Naïve Bayes dapat dikatakan sebagai algoritma klasifikasi yang sederhana, tetapi hasil Naïve Bayes Classifier merupakan metode

supervised learning, sehingga setiap data perlu diberikan label sebelum proses training dilakukan (H.Taufiqurrahman, F.Anggraeny, 2023).

Proses pelatihan Naive Bayes membutuhkan data berlabel, di mana setiap teks sudah diketahui kategorinya. Berdasarkan data tersebut, model akan membentuk pemahaman statistik tentang frekuensi kemunculan kata-kata tertentu dalam masing-masing kelas sentimen. Kemudian, model menghitung kemungkinan sebuah teks baru termasuk dalam kelas tertentu berdasarkan distribusi kata yang dimiliki.

Tingkat performa dari sistem klasifikasi yang dibuat menggunakan Naïve Bayes Classifier tergantung pada data yang dimiliki dan data yang diambil sebagai data Training. Jika data yang diambil sebagai data Training bisa memiliki semua ataupun sebagian besar data yang dimiliki, maka sistem klasifikasi mempunyai performa yang baik. Ketika suatu sistem klasifikasi yang dibuat mempunyai performa yang baik, maka sistem tersebut bisa diterapkan untuk melakukan klasifikasi terhadap data yang lebih banyak.

Naïve Bayes merupakan penyederhanaan dari Teorema Bayes yang digunakan untuk melakukan klasifikasi berdasarkan pendekatan probabilistik. Metode ini menghitung probabilitas suatu data termasuk ke dalam kelas tertentu berdasarkan data yang diamati sebelumnya.

Teorema Bayes dapat dirumuskan sebagai berikut:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)}$$

Dimana:

C : Kelas (misalnya: *positif* atau *Negatif*)

X : Data atau fitur yang diamati (misalnya: kata-kata dalam ulasan)

Keterangan:

- $P(C|X)$: Probabilitas suatu data termasuk ke dalam kelas C (posterior)
- $P(X|C)$: Probabilitas data X muncul bila diketahui kelasnya C (likelihood)
- $P(C)$: Probabilitas awal dari kelas C (prior)
- $P(X)$: Probabilitas keseluruhan dari data X (evidence)

Bayes Rules digunakan untuk menghitung posterior dan probabilitas dari data sebelumnya. Hasil akhirnya akan memberi informasi prior dan posterior untuk menghasilkan probabilitas menggunakan Bayes (Sitio & Nadiyanti, 2022).

2.6. VADER (Valence Aware Dictionary and sEntiment Reasoner)

VADER (Valence Aware Dictionary and Sentiment Reasoner) merupakan metode Lexicon based dengan teknik unsupervised. Vader Sentiment menggunakan leksikon yang dirancang khusus untuk analisis sentimen dalam teks berbahasa Inggris, dan algoritma ini mempertimbangkan intensitas kata-kata serta konteks penggunaannya yang menghasilkan class sentiment berupa positif, negatif dan netral dengan tambahan score total yaitu gabungan. Selain itu, Vader Sentiment juga memiliki kamus dengan 7.500 tokens yang mengandung kata berbahasa Inggris, kata slang, emoticons, serta sentimen yang terkait dengan singkatan, inisial dan akronim (Gultom et al., 2024).

Keunggulan VADER terletak pada sensitivitasnya terhadap nuansa bahasa, termasuk penggunaan tanda baca, huruf kapital, kata penekanan (seperti "sangat",

"tidak terlalu"), serta emoji dan singkatan. Hal ini membuat VADER sangat cocok untuk mengolah teks ulasan atau komentar yang sering kali tidak mengikuti kaidah bahasa formal. Selain itu, VADER memberikan output skor sentimen yang lebih terperinci, seperti skor positif, negatif, netral, dan skor komposit (compound).

Menurut (Abimanyu et al., 2022), Keuntungan menggunakan VADER polarity detection adalah tersedianya kamus yang berisi nilai dari setiap kata. Hasil Preprocess text akan di nilai berdasarkan lexicon apakah itu positif, negatif atau netral dan menambahkan skor total (compound). Beberapa perintah VADER yang menggunakan bahasa pemrograman python akan dikerjakan, dan VADER akan memanggil data lexicon dari server NLTK untuk menghitung polarity class sentimen.

Meskipun efektif, VADER memiliki keterbatasan utama yaitu hanya mendukung bahasa Inggris. Oleh karena itu, jika ulasan pengguna ditulis dalam Bahasa Indonesia, maka perlu dilakukan penerjemahan terlebih dahulu sebelum teks dapat dianalisis menggunakan VADER. Proses ini harus dilakukan dengan hati-hati agar makna asli dari ulasan tidak berubah secara signifikan.(Gultom et al., 2024).

Dalam penelitian ini, VADER digunakan sebagai pendekatan berbasis leksikon untuk memberikan alternatif klasifikasi sentimen terhadap ulasan pengguna. Hasil dari metode ini akan diperkuat dengan hasil dari metode Naive Bayes yang berbasis pembelajaran mesin.

Adapun proses penentuan polaritas kalimat didapatkan dari penyatuan attribute “compound” dari setiap kata yang tersedia. Kriteria pengelompokan positif, netral dan negatif yakni jika hasil compound lebih dari 0, 5 maka

dimasukkan kategori positif yang diwakilkan dengan angka 1 lalu jika hasil compound terletak diantara -0,5 dan 0,5 maka termasuk kategori netral yang diwakilkan dengan angka 0 dan yang terakhir jika hasil compound dibawah -0,5 maka termasuk kategori negatif yang diwakilkan dengan angka -1 (Abimanyu et al., 2022).

2.7. Algoritma TF-IDF

Algoritma TF-IDF (Term Frequency – inverse Document Frequency) adalah salah satu algoritma yang digunakan untuk menganalisa hubungan antara sebuah kalimat dengan sekumpulan dokumen dan metode untuk menghitung bobot setiap kata yang paling umum di gunakan (Musfiroh et al., 2021). Rumus untuk TF-IDF yaitu:

$$W_{td} = tf_{td} \left(\log \left(\frac{N}{DF_t} \right) + 1 \right)$$

Keterangan:

d : dokumen ke-d

t : kata ke-t dari kata kunci

d : bobot d sampai d terhadap kata t

tf : jumlah kata yang dicari pada sebuah dokumen

IDF : Inverse Document Frequency

N : jumlah total dokumen

df : jumlah dokumen yang berisi token / kata

Pada tahap TF-IDF ini akan dilakukan pembobotan term, yang mana merupakan proses pembobotan atau pemberian nilai terhadap setiap term yang ada pada setiap data ulasan yang telah di-preprocessing sebelumnya. Pemberian nilai

Bobot ini menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Hasil dari pembobotan kata ini kemudian akan menjadi masukan (input) pada proses klasifikasi (Gifari et al., 2022).

2.8. Confusion Matrix

Confusion Matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining. Confusion matrix digambarkan dengan tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan (Wiratama Putra et al., 2022).

Terdapat 4 metrik di dalam confusion matrix, yaitu:

1. True Positive (TP)

TP adalah jumlah prediksi yang berhasil diprediksi oleh model yang memprediksi kelas positif sebagai positif.

2. True Negative (TN)

TN adalah jumlah prediksi yang berhasil diprediksi oleh model yang memprediksi kelas negatif sebagai negative.

3. False Positive (FP)

FP adalah jumlah prediksi yang gagal diprediksi oleh model yang memprediksi kelas negatif sebagai positif.

4. False Negative (FN)

FN adalah jumlah prediksi yang gagal diprediksi oleh model yang memprediksi kelas positif sebagai negatif. Metrik-metrik diatas dapat digunakan untuk

menghitung nilai performa model seperti accuracy, precision, recall, dan F_1 score. Berikut ini adalah rumus-rumus nilai performa yang telah disebutkan,

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

$$recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall}$$

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

2.9. Google Colab

Google Colaboratory, atau yang lebih dikenal dengan Google Colab, merupakan layanan cloud dari Google yang memungkinkan pengguna untuk menulis dan menjalankan kode Python secara langsung di browser. Platform ini sangat populer di kalangan peneliti, pelajar, dan pengembang karena memberikan akses ke komputasi yang kuat, termasuk GPU dan TPU secara gratis, tanpa perlu instalasi perangkat lunak secara local (Gabriel Do Breviário et al., 2025).

Keunggulan utama Google Colab adalah integrasinya dengan Google Drive, sehingga memudahkan dalam penyimpanan dan kolaborasi dokumen notebook secara daring. Selain itu, berbagai pustaka Python untuk data science dan machine learning seperti pandas, scikit-learn, matplotlib, dan nltk sudah terpasang secara default, sehingga pengguna bisa langsung mengembangkan model analisis data (Vidhya, 2024).

Dalam penelitian ini, Google Colab dimanfaatkan sebagai platform untuk melakukan analisis sentimen, mulai dari pengumpulan dan pra-pemrosesan data,

pelatihan model Naive Bayes, penerapan VADER, hingga visualisasi hasil. Dengan menggunakan platform ini, proses eksperimen menjadi lebih efisien dan fleksibel, tanpa harus dibatasi oleh spesifikasi perangkat keras lokal.

Penggunaan Google Colab juga memungkinkan kolaborasi langsung dengan dosen pembimbing atau rekan peneliti, karena notebook dapat dibagikan dan dikerjakan bersama-sama secara real-time. Hal ini memberikan nilai tambah dalam konteks penelitian akademik dan pengembangan proyek berbasis data.

2.10. Bahasa Pemrograman Python

Menurut (Ziogas et al., 2021), Python adalah bahasa pemrograman tingkat tinggi yang dirancang dengan sintaksis sederhana dan mudah dibaca, menjadikannya pilihan utama dalam pengembangan aplikasi data science dan machine learning.

Juga merupakan Bahasa pemrograman interpretatif yang dapat digunakan diberbagai platform dengan perancangan yang berfokus pada tingkat keterbacaan kode dan menyediakan banyak modul, dimana hal tersebut menjadikan python bahasa pemrograman yang populer digunakan dikarenakan sifatnya yang fleksibel seperti sanggup digunakan untuk kebutuhan machine learning, data science, pengujian perangkat lunak, hingga artificial intelligence, dan lain-lain (Z.Aminullah, D.Arifianto, 2024).

Untuk keperluan analisis sentimen, Python menyediakan sejumlah library populer seperti nltk (Natural Language Toolkit), scikit-learn, vaderSentiment, dan textblob. Library tersebut memungkinkan pengguna untuk membersihkan data, mengekstrak fitur, membangun model, serta mengevaluasi hasil klasifikasi dengan

efisien. Python juga memiliki kemampuan integrasi dengan berbagai platform dan format data, seperti CSV, JSON, maupun API daring (Castro et al., 2024).

Dalam penelitian ini, Python digunakan sebagai bahasa utama untuk mengolah ulasan pengguna, melatih model Naive Bayes, serta mengimplementasikan VADER. Selain itu, Python juga dimanfaatkan untuk melakukan eksplorasi data dan membuat visualisasi yang membantu dalam memahami pola-pola sentimen dalam data. Fleksibilitas dan kekuatan Python menjadikannya alat yang ideal dalam melakukan analisis berbasis data secara sistematis.

Penggunaan Python dalam Google Colab semakin memperkuat efisiensi penelitian karena seluruh proses dapat dijalankan dalam satu platform terpusat tanpa perlu instalasi tambahan. Hal ini memungkinkan peneliti untuk lebih fokus pada analisis dan interpretasi hasil, dibandingkan dengan aspek teknis pengelolaan lingkungan kerja.

BAB III

METODOLOGI PENELITIAN

3.1. Alur Penelitian

Alur penelitian ini menjelaskan tahapan-tahapan yang dilakukan dalam analisis sentimen terhadap ulasan pengguna aplikasi MIFX yang diperoleh dari Google Play Store. Penelitian ini menggabungkan dua pendekatan: metode VADER sebagai labeling sentimen berbasis lexicon, dan Naive Bayes sebagai metode klasifikasi berbasis machine learning.

Alur penelitian ini dimulai dari pengumpulan data ulasan pengguna aplikasi MIFX yang diambil dari Google Play Store menggunakan teknik scraping. Data yang diperoleh berupa teks dalam Bahasa Indonesia. Selanjutnya, dilakukan preprocessing ringan terhadap data, yang mencakup tahapan cleaning (menghapus karakter tidak relevan), case folding (mengubah huruf menjadi huruf kecil), serta normalisasi kata (menyelaraskan kata tidak baku ke bentuk baku).

Setelah itu, data hasil preprocessing ringan diterjemahkan ke dalam Bahasa Inggris. Proses ini diperlukan karena metode VADER hanya dapat menganalisis teks dalam Bahasa Inggris. Hasil translasi ini digunakan sebagai input untuk proses labeling otomatis menggunakan VADER, yang menghasilkan label sentimen berupa positif, negatif, atau netral.

Dataset yang telah diberi label kemudian dilanjutkan ke tahapan analisis menggunakan metode Naive Bayes. Di tahap ini dilakukan preprocessing lanjutan untuk kebutuhan klasifikasi machine learning. Tahapan preprocessing lanjutan meliputi tokenizing (memecah kalimat menjadi kata), stopwords removal

(menghapus kata-kata umum yang tidak memiliki nilai sentimen), dan stemming (mengembalikan kata ke bentuk dasarnya).

Setelah teks siap, dilakukan ekstraksi fitur menggunakan metode TF-IDF (Term Frequency – Inverse Document Frequency) untuk mengubah teks menjadi representasi numerik. Selanjutnya dilakukan proses split data untuk membagi data menjadi data latih dan data uji. Data latih digunakan untuk melatih model Naive Bayes, sedangkan data uji digunakan untuk mengukur performa model.

Hasil dari model Naive Bayes kemudian dievaluasi menggunakan metrik seperti akurasi, presisi, recall, dan F1-score. Seluruh hasil evaluasi tersebut divisualisasikan dalam bentuk wordcloud, diagram, dan grafik, guna mempermudah interpretasi terhadap pola sentimen dari pengguna aplikasi MIFX.

3.2. Pengumpulan Data

3.2.1 Sumber data

Sumber data Penelitian ini diambil dari Google Play Store, Khususnya pada ulasan pengguna yang tersedia di halaman detail aplikasi MIFX. Data dikumpulkan dalam jangka waktu tertentu untuk sentimen yang akurat.

3.2.2 Teknik Pengambilan Data

- a) Web Scraping: Data utama dalam penelitian ini diperoleh melalui teknik web scraping, yaitu proses pengambilan data secara otomatis dari halaman web dengan bantuan program komputer. Penelitian ini memanfaatkan pustaka Python seperti google-play-scraper untuk mengambil data ulasan pengguna terhadap aplikasi MIFX di platform Google Play Store
- b) Penyimpanan Data: Data hasil scraping disimpan dalam format .csv untuk kemudian diproses lebih lanjut dalam tahap analisis sentimen. Semua data yang

dikumpulkan adalah data publik, dan penggunaannya dalam penelitian ini sepenuhnya untuk tujuan akademik, tanpa melibatkan informasi pribadi pengguna.

3.2.3 Jenis Data

Penelitian ini tidak menggunakan data primer karena seluruh data bersifat tekstual dan diperoleh secara publik melalui Google Play Store. Jenis data yang digunakan dalam penelitian ini adalah data sekunder. Data sekunder merupakan data yang diperoleh tidak secara langsung dari responden, melainkan berasal dari sumber yang sudah tersedia sebelumnya. Jadi fokus penelitian adalah analisis data yang sudah tersedia (data sekunder) dengan metode tekst mining.

3.3. Preprocessing Data

3.3.1 Preprocessing untuk VADER

Sebelum dilakukan analisis sentimen dengan metode VADER, data ulasan terlebih dahulu diproses melalui tahapan preprocessing ringan. Tahapan ini bertujuan untuk membersihkan dan menormalkan teks ulasan dalam Bahasa Indonesia agar siap diterjemahkan ke Bahasa Inggris. VADER sendiri hanya dapat bekerja pada teks berbahasa Inggris, sehingga proses translasi menjadi bagian penting dalam alur ini.

Langkah-langkah preprocessing sebelum VADER adalah sebagai berikut:

1. Cleaning

Tahap ini merupakan proses untuk menghapus URL, @mention, #hashtag, link, double whitespaces, emoticon, serta angka. Karakter-

karakter ini dihapus karena dianggap sebagai noise dalam data dan tidak diperlukan dalam analisis sentimen.

Tabel 3. 1 Cleaning

Sebelum	Sesudah
Bener emang diindo banyak iq 🔴 loss nyalahin broker , ga bisa trading mending nguli aja, ngapain maen forex 🔴	Bener emang diindo banyak iq loss nyalahin broker ga bisa trading mending nguli aja ngapain maen forex

2. Case Folding

Case Folding merupakan proses untuk penyeragaman bentuk huruf atau kata kedalam bentuk huruf kecil. Selain itu case folding juga bisa untuk menghilangkan tanda baca dan menghapus karakter spasi yang berlebihan. Pada proses ini berguna agar huruf besar dan huruf kecil tidak terdeteksi memiliki perbedaan arti. Adapun sebagai contoh kalimat ulasan,

Tabel 3. 2 Case Folding

Sebelum	Sesudah
Bagus.. Hanya saja buat agar bisa mode HP miring	bagus hanya saja buat agar bisa mode hp miring

3. Normalisasi Kata

Pada bagian spelling normalization berguna untuk melakukan perbaikan kata-kata yang disingkat maupun salah ejaan dengan bentuk

tertentu dengan maksud yang sama, sebagai contoh pada kata “tidak” memiliki banyak bentuk penulisan yaitu tak, enggak, tdak, tdk, dan sebagainya. Adapun contoh dalam kalimat ulasan.

Tabel 3. 3 Normalisasi Kata

Sebelum	Sesudah
perbaiki lupa sandi g bisa trus	perbaiki lupa sandi tidak bisa terus

Setelah teks dinormalisasi, dilakukan penerjemahan ke Bahasa Inggris menggunakan library translator otomatis seperti Google Translate API atau deep-translator. Langkah ini dilakukan agar teks bisa dianalisis oleh VADER. Setelah tahapan ini selesai, data siap untuk dianalisis menggunakan metode VADER guna menghasilkan label sentiment.

3.3.2 Preprocessing untuk Naive Bayes

Setelah data ulasan diberi label sentimen menggunakan VADER, tahap berikutnya adalah klasifikasi menggunakan metode Naive Bayes. Sebelum data masuk ke tahap pelatihan model, dilakukan **preprocessing lanjutan** agar data sesuai untuk diolah oleh algoritma machine learning. Berbeda dengan preprocessing sebelumnya, tahap ini bertujuan untuk menyederhanakan kata dan mengurangi noise dalam data teks.

Langkah-langkah preprocessing untuk Naive Bayes adalah sebagai berikut:

4. Tokenizing

Tokenizing atau tokenisasi merupakan proses memisahkan teks dalam dokumen menjadi potongan kata yang tidak saling berpengaruh

atau independen yang disebut dengan token. Tokenizing digunakan untuk mendapatkan potongan kata atau token yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. Pada proses ini memudahkan proses perhitungan keberadaan kata tersebut dalam dokumen maupun dalam menghitung frekuensi kemunculan kata tersebut dalam corpus. Adapun sebagai contoh kalimat ulasan,

Tabel 3. 4 Tokenizing

Sebelum	Sesudah
disarankan jangan pakai broker ini untuk kenyamanan masing masing	disarankan, jangan, pakai, broker, ini, untuk, kenyamanan, masing, masing

5. Stopword Removal/Filtering

Tahap filtering merupakan tahap dilakukannya penghilangan kata di dalam dokumen atau pengurangan dimensi kata di dalam corpus yang disebut stopwords. Stopwords digunakan untuk menghilangkan kata-kata yang tidak ada berpengaruh atau tidak mengurangi informasi didalam dokumen tersebut namun keberadaannya sering muncul. Contoh isi dalam bahasa indonesia berdasarkan data penelitian yaitu “saya”, “kamu”, “itu”, “untuk”, “tolong”, “mohon”, “segera”, “sangat” dan lain sebagainya.

Tabel 3. 5 Stopword Removal/Filtering

Sebelum	Sesudah
sebagai, pengguna, mobile, hp, sangat, kesulitan, untuk, menggambar, di, layar, chart, karena, layar, potrait, tidak, bias, dirubah, ke, tampilan, landscap	pengguna, mobile, hp, kesulitan, menggambar, layar, chart, layar, potrait, dirubah, tampilan, landscap

6. Stemming

Stemming adalah proses dalam Natural Language Processing (NLP) untuk mengubah kata berimbuhan menjadi bentuk dasarnya (kata dasar atau stem). Imbuhan yg akan dihapus seperti awalan (me-, di-, ke-, se-), akhiran (-kan, -an, -i), atau sisipan.

Tabel 3. 6 Stemming

Sebelum	Sesudah
pengguna, mobile, hp, kesulitan, menggambar, layar, chart, layar, potrait, dirubah, tampilan, landscap	guna mobile hp sulit gambar layar chart layar potrait rubah tampil landscap

3.4. Hipotesis Penelitian

3.4.1. Metode Vader (Valence Aware Dictionary dan sEntiment Reasoner)

Salah satu manfaat dari pendekatan ini adalah tidak memerlukan pelatihan model atau estimasi parameter, namun tetap mampu memberikan hasil analisis sentimen yang akurat, khususnya untuk teks pendek dan informal. Karena VADER menggunakan pendekatan leksikon yang dipadukan dengan aturan linguistik, metode ini sangat mudah diimplementasikan dan memberikan hasil cepat dalam klasifikasi sentimen (Leonardi et al., 2025)

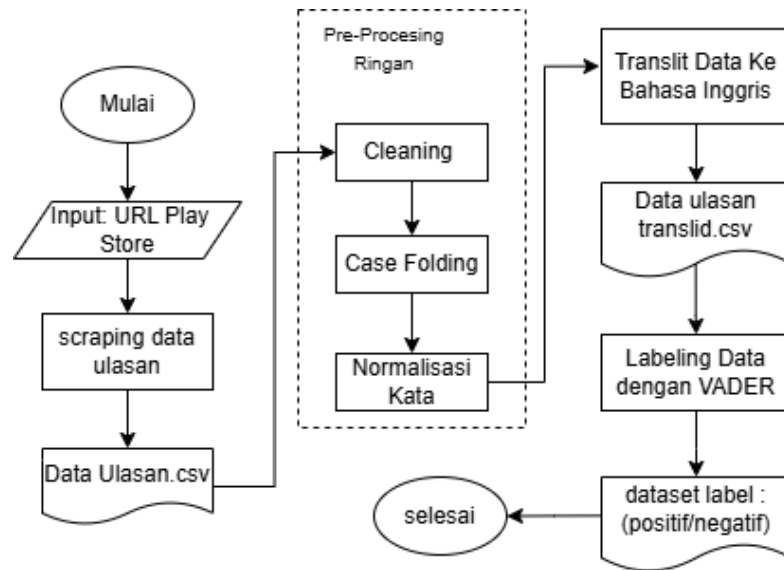
Vader Lexicon dipilih karena keandalannya dalam menganalisis sentimen teks pendek, seperti ulasan dan komentar. Metode ini memanfaatkan leksikon sentimen yang telah ditentukan sebelumnya, di mana setiap kata memiliki skor sentimen yang menggambarkan tingkat kepositifan atau kenegatifannya (Arifin & Febryanto, 2025).

Skor komposit VADER, yang dihitung dengan menjumlahkan skor valensi setiap kata dalam leksikon dan mengubahnya menjadi kisaran +1 untuk sentimen paling positif dan -1 untuk sentimen paling negatif, digunakan dalam penelitian ini.

Prosedur kerja VADER adalah sebagai berikut:

1. Mengidentifikasi kata-kata dari teks berdasarkan kamus sentimen yang telah disiapkan.
2. Menyesuaikan skor kata dengan konteks linguistik (misalnya penyangkalan atau penguatan makna).
3. Menghasilkan skor komposit dan klasifikasi sentimen (positif, netral, atau negatif)."

Pendekatan Vader pada penelitian ini diilustrasikan seperti gambar 3.1 dibawah ini



Gambar 3. 1 Flowchart Algoritma Vader

Keterangan :

1. Mulai

Titik awal proses analisis sentimen.

2. Scraping Ulasan dari Google Play Store

Mengambil ulasan aplikasi MIFX dalam Bahasa Indonesia melalui teknik scraping.

3. Preprocessing Ringan

Melakukan pembersihan awal data, mencakup:

- Case folding
- Cleaning karakter/symbol
- Normalisasi kata (misalnya: “gk” → “tidak”)

4. Translasi ke Bahasa Inggris

Ulasan diterjemahkan dari Bahasa Indonesia ke Bahasa Inggris agar kompatibel dengan VADER.

5. Analisis Sentimen dengan VADER

Menganalisis teks yang telah diterjemahkan menggunakan metode lexicon-based VADER, yang memberi skor sentimen dan mengklasifikasikan ulasan ke dalam:

- Positif
- Negatif

6. Output: Dataset Berlabel

Dataset yang telah diberi label sentimen secara otomatis digunakan sebagai input pada tahapan klasifikasi berikutnya (Naive Bayes).

7. Selesai

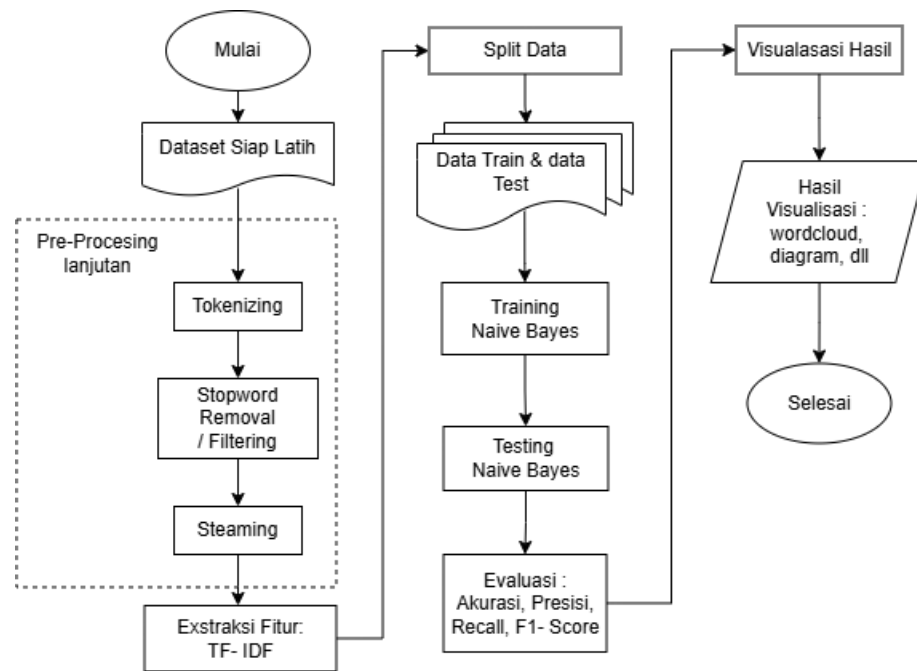
Menandakan akhir dari proses labeling otomatis.

3.4.2. Metode Naive Bayes

Dalam penelitian ini, pendekatan Naïve Bayes digunakan karena memiliki keunggulan dalam mengklasifikasikan teks secara efisien, termasuk dalam konteks analisis sentimen. Metode ini bekerja dengan menghitung kemungkinan sebuah ulasan termasuk dalam kategori tertentu, seperti positif, netral, atau negatif, berdasarkan pola distribusi kata-kata yang telah dipelajari dari data pelatihan. Berbeda dengan metode leksikon yang mengandalkan daftar kata dengan nilai sentimen tetap, Naïve Bayes memanfaatkan kemampuan belajar dari data historis untuk memahami kaitan antara kata-kata tertentu dan sentimen yang menyertainya.

Meskipun pendekatan ini mengasumsikan bahwa setiap kata dalam teks saling bebas atau independen, asumsi tersebut tidak menjadi hambatan signifikan karena Naïve Bayes tetap mampu menghasilkan hasil klasifikasi yang kompetitif, khususnya ketika diterapkan pada kumpulan data yang besar. Dibandingkan dengan

algoritma klasifikasi lainnya yang lebih kompleks, metode ini relatif lebih ringan secara komputasi, baik saat proses pelatihan maupun prediksi, namun tetap memberikan akurasi yang cukup baik. Untuk memberikan gambaran lebih jelas, ilustrasi proses kerja Naïve Bayes dapat dilihat pada Gambar 3.2 berikut.



Gambar 3. 2 Flowchart Algoritma Naive Bayes

Keterangan :

1. Mulai

Awal proses klasifikasi sentimen menggunakan metode machine learning.

2. Input Dataset Berlabel (hasil dari VADER)

Data ulasan yang telah diberi label (positif dan negatif) dari tahap sebelumnya.

3. Preprocessing Lanjutan

Tahapan pembersihan lanjutan untuk menyiapkan data teks bagi proses klasifikasi. Termasuk:

- Tokenizing

- Stopword removal
- Stemming

4. Ekstraksi Fitur: TF-IDF

Teks diubah menjadi representasi numerik agar bisa dibaca oleh algoritma machine learning. TF-IDF memberikan bobot pada setiap kata berdasarkan kemunculannya.

5. Split Data: Train & Test

Dataset dibagi menjadi dua:

- Data Train: digunakan untuk melatih model
- Data Test: digunakan untuk menguji performa model

6. Training Model Naive Bayes

Model Naive Bayes mempelajari pola dari data latih dan menghubungkannya dengan label sentimen.

7. Testing Model Naive Bayes

Model yang sudah dilatih diuji menggunakan data test untuk melihat seberapa akurat model mengenali sentimen baru.

8. Evaluasi Model

Menilai performa model dengan metrik evaluasi, seperti:

- Akurasi
- Presisi
- Recall
- F1-score
- Confusion matrix

9. Visualisasi Hasil

Menampilkan distribusi sentimen dan performa model dalam bentuk visual, seperti:

- Grafik batang
- Wordcloud
- Pie chart

10. Output: Hasil Visualisasi

Grafik dan analisis hasil klasifikasi yang bisa digunakan untuk pelaporan.

11. Selesai

Proses klasifikasi dan analisis berakhir.

3.5. Penelitian Terdahulu

Penelitian ini mengacu pada beberapa penelitian terdahulu yang memiliki keterkaitan baik dari segi topik, metode, maupun objek penelitian. Adapun beberapa penelitian terdahulu yang relevan adalah sebagai berikut:

Tabel 3. 7 Penelitian Terdahulu

No	Judul Penelitian	Peneliti & Tahun	Metode	Kelebihan & Kekurangan
1	Pengembangan Model Klasifikasi Sentimen Dengan Pendekatan Vader dan Algoritma Naive Bayes	Zendrato, Agus Dirgahayu Berutu, Sunneng Sandino Sumihar, Pieter	VADER dan Naïve Bayes	Kelebihan : Akurasi klasifikasi mencapai 78% Kekurangan: Recall untuk

	Terhadap Ulasan Aplikasi Indodax	Budiati, Haeni (2024)		sentimen positif rendah (44%)
2	Analisis Sentimen Data Ulasan Aplikasi MyPertamina Menggunakan VADER Dan Indonesian Sentiment Lexicon Dengan Algoritma Naive Bayes	Chrisna Barka (2024)	Vader, InSet (Indonesia Sentimen), dan Naive Bayes	Kelebihan: Akurasi tertinggi mencapai 89,64% (InSet + NB) Kekurangan: VADER hanya menghasilkan 81,51%, kurang efektif untuk bahasa Indonesia

Analisis dan Relevansi:

Penelitian oleh Chrisna Barka Ramadhan (2024) menunjukkan bahwa metode pelabelan berbasis leksikon lokal, seperti Indonesian Sentiment Lexicon (InSet), lebih unggul dibandingkan VADER dalam konteks bahasa Indonesia. Kombinasi InSet dan Naïve Bayes berhasil mencapai akurasi hingga 89,64%, sedangkan VADER hanya mencapai 81,51%.

Sementara itu, penelitian oleh Zendrato dkk. (2024) menggunakan pendekatan gabungan antara VADER dan Naïve Bayes. Hasilnya menunjukkan akurasi sebesar 78%, namun dengan kelemahan signifikan pada kemampuan mendeteksi sentimen positif (recall) rendah.

Untuk mendukung pemahaman terhadap pendekatan yang digunakan dalam penelitian ini, berikut disajikan perbedaan karakteristik antara metode Naïve Bayes dan VADER sebagai bagian dari pendekatan analisis sentimen:

Tabel 3. 8 Perbedaan Vader dan Naive Bayes

Naïve Bayes	VADER (Valence Aware Dictionary and sEntiment Reasoner)
Berorientasi pada probabilitas dan pembelajaran dari data yang dilabeli.	Berorientasi pada kamus kata (lexicon) dengan skor sentimen yang telah ditentukan.
Membutuhkan pra-pelabelan data untuk melatih model.	Tidak memerlukan pelabelan data karena berbasis kamus.
Menggunakan distribusi probabilitas untuk memprediksi sentimen.	Menggunakan skor sentimen dari kata-kata untuk menghitung total skor sentimen kalimat.
Dapat digunakan untuk berbagai bahasa (tergantung pelatihan model).	Secara default dioptimalkan untuk Bahasa Inggris. Butuh translasi jika datanya berbahasa lain.
Performa tergantung pada kualitas data latih dan representasi fitur seperti TF-IDF.	Performa tergantung pada cakupan dan kekuatan leksikon dalam mendeteksi nuansa sentimen.

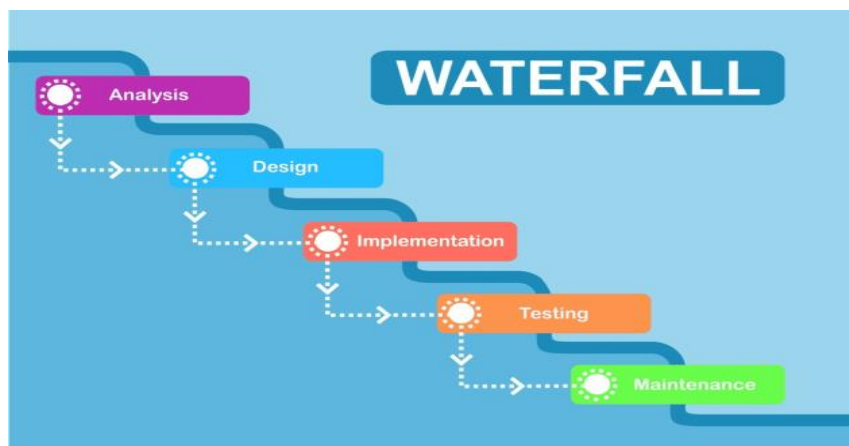
3.6. Model Perancangan Sistem

Pada sub-bab ini, diuraikan kerangka kerja perancangan dan pengembangan prototipe aplikasi web yang dibangun untuk mendemonstrasikan dan memvisualisasikan hasil dari model analisis sentimen. Perancangan sistem ini bertujuan untuk mentransformasikan alur penelitian menjadi sebuah produk perangkat lunak fungsional sebagai bukti konsep (proof of concept). Model pengembangan yang digunakan adalah Software Development Life Cycle (SDLC) dengan pendekatan Waterfall, sementara untuk pemodelan fungsionalitas sistem digunakan Unified Modeling Language (UML).

3.6.1. Model SDLC Waterfall

Model Model Waterfall dipilih karena sifatnya yang terstruktur dan sangat cocok untuk pengembangan prototipe dengan lingkup yang sudah pasti. Setiap fase diselesaikan sepenuhnya sebelum melanjutkan ke fase berikutnya.

Berikut adalah penjabaran detail dari setiap tahapan model Waterfall yang diterapkan dalam pengembangan prototipe ini:



Gambar 3. 3 Model SDLC Waterfall

1. Analisis Kebutuhan (Requirement Analysis) Tahap ini mengidentifikasi kebutuhan utama prototipe sebagai alat demonstrasi penelitian.

- Kebutuhan Fungsional:
 - Prototipe harus mampu memvisualisasikan setiap tahapan proses yang dilakukan dalam penelitian, mulai dari input data hingga output hasil analisis.
 - Prototipe harus dapat menyimpan data ulasan mentah dan hasil analisisnya ke dalam sebuah basis data untuk setiap sesi analisis.
 - Prototipe harus menampilkan metrik evaluasi model (Accuracy, Precision, Recall, F1-Score, dan Confusion Matrix) sesuai dengan hasil yang didapat pada BAB IV.
 - Prototipe harus menyajikan visualisasi data (Word Cloud dan grafik) yang identik dengan hasil penelitian.
- Kebutuhan Non-Fungsional:
 - Antarmuka Pengguna (UI): Antarmuka dirancang untuk keperluan demonstrasi, dengan fokus pada kejelasan alur dan visualisasi data.
 - Lingkungan Operasi: Prototipe dirancang untuk dijalankan di lingkungan lokal (localhost) pada komputer peneliti untuk keperluan presentasi atau sidang.

2. Perancangan Sistem (System Design) Pada tahap ini, kebutuhan diterjemahkan menjadi rancangan teknis prototipe.

- Perancangan Arsitektur: Sistem dirancang sebagai aplikasi web yang berjalan secara lokal. Backend dibangun menggunakan framework Flask dengan bahasa Python untuk memproses dan menyajikan data hasil

penelitian. Frontend dibangun menggunakan HTML, CSS, dan JavaScript untuk menampilkan visualisasi.

- Perancangan Basis Data: Untuk mengelola data ulasan dan hasil analisisnya, digunakan basis data relasional MySQL. Basis data ini dipilih karena mampu menangani data secara terstruktur dan sesuai dengan kebutuhan penelitian. Basis data dirancang menggunakan MySQL yang dijalankan melalui paket XAMPP. Pemilihan XAMPP bertujuan untuk memudahkan pengelolaan server lokal yang sudah terintegrasi dengan Apache dan phpMyAdmin, sehingga proses penyimpanan maupun pengujian data dapat dilakukan secara lebih efisien. Tabel utama yang digunakan adalah data_ulasan, yang berfungsi untuk menyimpan seluruh informasi terkait proses analisis setiap sesi, mulai dari ulasan mentah hingga hasil klasifikasi sentimen.
- Struktur Tabel data_ulasan:
 - id (INT): Primary Key, Auto Increment.
 - sesi_id (VARCHAR (255)): ID unik untuk setiap sesi analisis.
 - ulasan_asli (TEXT): Teks ulasan mentah sebelum diproses.
 - ulasan_bersih (TEXT): Teks setelah proses cleaning dan normalisasi.
 - ulasan_english (TEXT): Teks setelah ditranslasi ke Bahasa Inggris.
 - label_vader (VARCHAR (10)): Label sentimen ('positif'/'negatif'/'netral') dari VADER.

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
☐	1 id	int(11)			No	None		AUTO_INCREMENT	Change Drop More
☐	2 sesi_id	varchar(50)	utf8mb4_unicode_ci		No	None			Change Drop More
☐	3 tahap	int(11)			No	None			Change Drop More
☐	4 ulasan_asli	text	utf8mb4_unicode_ci		Yes	NULL			Change Drop More
☐	5 ulasan_bersih	text	utf8mb4_unicode_ci		Yes	NULL			Change Drop More
☐	6 ulasan_english	text	utf8mb4_unicode_ci		Yes	NULL			Change Drop More
☐	7 label_vader	varchar(20)	utf8mb4_unicode_ci		Yes	NULL			Change Drop More
☐	8 dibuat_pada	timestamp			No	current_timestamp()			Change Drop More

Gambar 3. 4 Tabel Struktur Basis Data

- Perancangan Antarmuka (UI/UX): Alur prototipe dirancang secara linear untuk menceritakan kembali proses penelitian: Halaman Awal (Input Data) → Halaman Hasil (Menampilkan semua metrik dan visualisasi yang diambil dari basis data).
3. Implementasi (Implementation) Tahap ini adalah proses penulisan kode berdasarkan rancangan.
 - Lingkungan Pengembangan: Kode program dikembangkan dan diuji menggunakan editor kode lokal.
 - Teknologi: Menggunakan Python 3.x dengan library Flask (web server), Pandas (manipulasi data), NLTK & Scikit-learn (analisis), VaderSentiment, dan mysql-connector-python (manajemen basis data). Frontend menggunakan HTML5, Tailwind CSS, Chart.js, dan JavaScript.
 4. Pengujian (Testing) Pengujian dilakukan oleh peneliti sendiri untuk memastikan prototipe berjalan sesuai harapan.
 - Pengujian Fungsional: Memastikan proses penyimpanan ke basis data MySQL, pengambilan data, dan penampilan data berjalan dengan benar.

- Pengujian Kesesuaian: Memverifikasi bahwa semua data yang ditampilkan dari basis data sama persis dengan yang dilaporkan dalam naskah skripsi.
5. Demonstrasi Sistem (System Demonstration) Tahap ini menggantikan "Deployment" dan "Maintenance". Tujuannya adalah sebagai berikut:
- Tujuan: Menunjukkan bahwa alur kerja dan model yang diusulkan dalam penelitian dapat diimplementasikan menjadi sebuah sistem yang nyata dan fungsional, lengkap dengan manajemen data.
 - Pelaksanaan: Prototipe dijalankan secara langsung (live demo) saat sidang skripsi atau presentasi penelitian untuk memberikan bukti konkret dari hasil kerja.

3.6.2. Pemodelan UML (Unified Modeling Language)

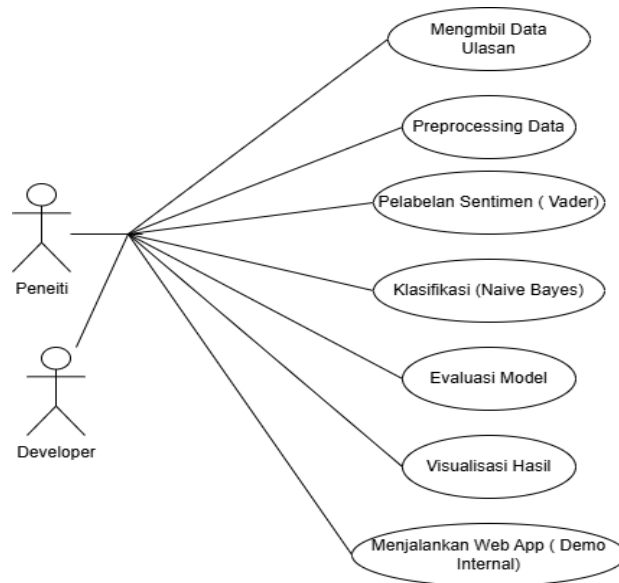
Untuk memvisualisasikan fungsionalitas prototipe, digunakan Use Case Diagram. Diagram ini menggambarkan interaksi antara aktor (dalam hal ini, Peneliti) dengan fitur-fitur utama prototipe.

A. Aktor

- Peneliti (Researcher): Merupakan aktor tunggal yang menjalankan dan berinteraksi dengan prototipe untuk mendemonstrasikan alur dan hasil penelitian.

B. Diagram Use Case

Diagram Use Case pada gambar 3.4 memvisualisasikan fungsionalitas utama dari prototipe aplikasi web yang dibangun, dengan Peneliti sebagai satu-satunya aktor yang berinteraksi dengannya. Diagram ini menunjukkan bagaimana Peneliti menggunakan prototipe sebagai alat untuk mendemonstrasikan keseluruhan proses penelitian secara interaktif.

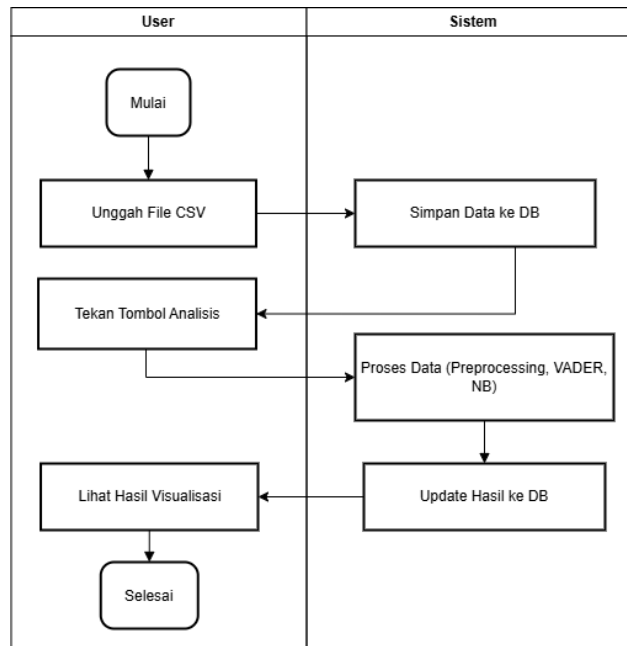


Gambar 3. 5 Diagram Use Case Prototipe Sistem

- Mendemonstrasikan Input Data: Peneliti mengunggah dataset ulasan, sistem memvalidasi dan menyimpan ke MySQL sebagai awal sesi analisis.
- Proses Analisis: Sistem otomatis menjalankan preprocessing, VADER, dan Naïve Bayes, lalu menyimpan hasil ke basis data.
- Hasil & Visualisasi: Sistem menampilkan dasbor berisi laporan klasifikasi, confusion matrix, serta visualisasi seperti word cloud.

C. Activity Diagram

Activity Diagram digunakan untuk memodelkan alur kerja (workflow) dari sebuah proses. Diagram berikut menggambarkan alur aktivitas utama dalam sistem aplikasi web, mulai dari pengunggahan data hingga penampilan hasil analisis.

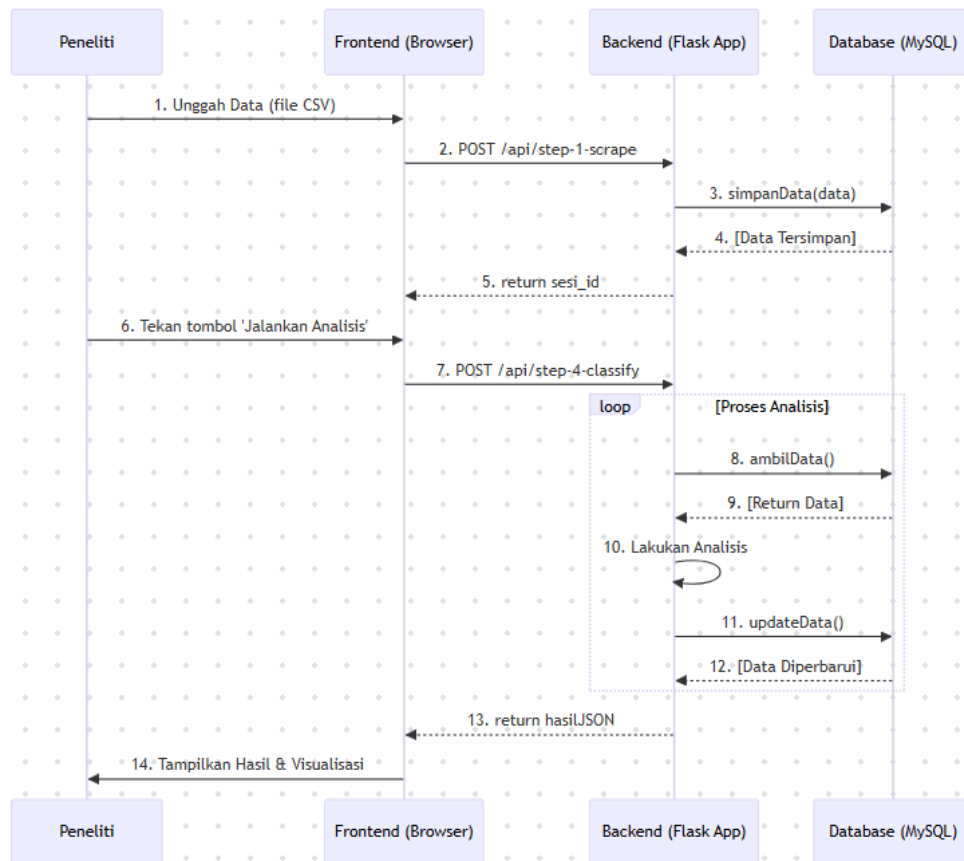


Gambar 3. 6 Diagram Activity

Diagram ini menjelaskan alur kerja interaktif antara Pengguna (User) dan Sistem. Proses diawali oleh Pengguna yang melakukan aksi "Unggah File CSV". Sistem kemudian merespons dengan "Simpan Data ke DB" dan menunggu aksi selanjutnya. Pengguna melanjutkan dengan "Tekan Tombol Analisis", yang memicu Sistem untuk menjalankan alur pemrosesan utama, mencakup "Proses Data (Preprocessing, VADER, NB)" dan "Update Hasil ke DB". Setelah proses di sisi sistem selesai, hasilnya dikirim kembali agar Pengguna dapat "Lihat Hasil Visualisasi". Alur kerja berakhir saat Pengguna selesai melihat hasil (Selesai).

D. Sequence Diagram

Sequence Diagram digunakan untuk menggambarkan interaksi antar objek atau komponen dalam urutan waktu. Diagram ini menunjukkan bagaimana Frontend (Browser), Backend (Server Flask), dan Database (MySQL) berkomunikasi saat Peneliti menjalankan proses analisis di aplikasi web.

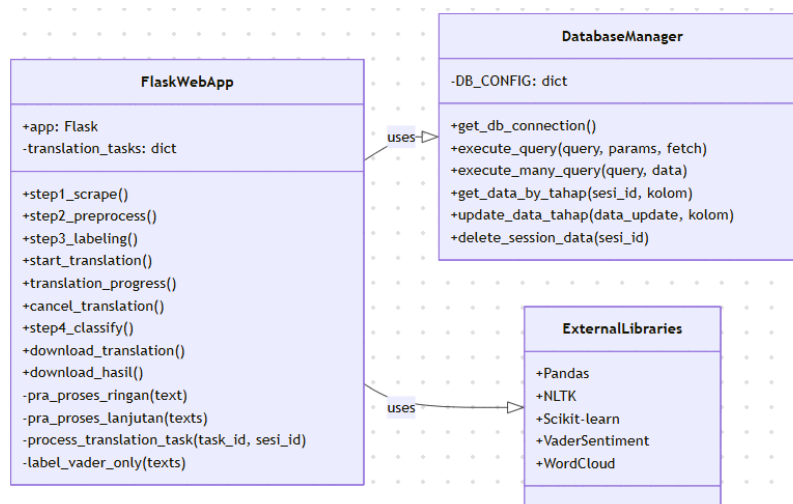


Gambar 3. 7 Sequence Diagram

Diagram ini memvisualisasikan alur komunikasi antar komponen sistem. Proses diawali saat Peneliti mengunggah file melalui Frontend, yang kemudian mengirimkannya ke Backend. Backend memproses dan menyimpan data ke Database, lalu mengembalikan sesi_id sebagai konfirmasi. Selanjutnya, saat Peneliti memulai analisis, Frontend mengirim permintaan ke Backend. Backend kemudian berulang kali mengambil data dari Database, menjalankan proses analisis, dan memperbarui hasilnya. Setelah selesai, Backend mengirimkan hasil akhir ke Frontend untuk ditampilkan kepada Peneliti.

E. Class Diagram

Class Diagram digunakan untuk menggambarkan struktur statis dari sistem, yaitu kelas-kelas perangkat lunak dan hubungan antar kelas tersebut. Diagram ini menunjukkan komponen-komponen utama dalam kode backend aplikasi web Anda.



Gambar 3. 8 Class Diagram

- **app.py (FlaskWebApp):** Merupakan kelas utama yang mengelola routing (URL endpoint) dan logika aplikasi. Kelas ini bergantung pada database.py untuk semua operasi database dan menggunakan berbagai library analisis.
- **database.py (DatabaseManager):** Kelas ini bertanggung jawab untuk semua interaksi dengan database MySQL. Ini berisi fungsi-fungsi untuk menghubungkan, mengambil (query), dan memanipulasi data di dalam tabel.
- **ExternalLibraries:** Ini bukan satu kelas tunggal, melainkan representasi dari berbagai library eksternal yang digunakan oleh app.py untuk melakukan tugas-tugas spesifik, seperti Pandas, Scikit-learn, dan VaderSentiment.
-

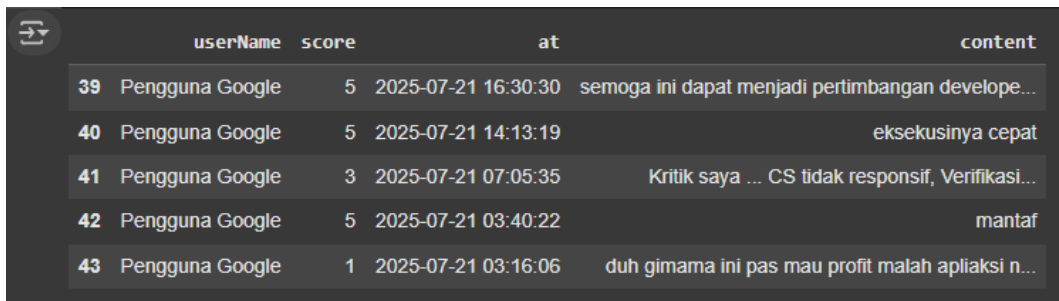
BAB IV

HASIL DAN PEMBAHASAN

4.1. Scraping Data

Scraping data adalah proses pengambilan data secara otomatis dari suatu sumber digital, seperti situs web atau aplikasi (Rizquina & Ratnasari, 2023). Pada tahap ini, peneliti menggunakan library google-play-scraper. Kode tersebut berfungsi untuk menginstal dan memanfaatkan tools scraping yang dapat mengambil data aplikasi secara otomatis. Selain itu, peneliti juga mempersiapkan tools analisis data dengan mengimpor library Pandas (untuk pengelolaan data) dan NumPy (untuk proses numerik).

Proses scraping dilakukan dengan mendefinisikan aplikasi target melalui ID com.mifxid.app, kemudian mengambil seluruh ulasan terbaru berbahasa Indonesia dari pengguna di wilayah Indonesia menggunakan parameter sort=Sort.NEWEST. Data hasil scraping selanjutnya disimpan ke dalam DataFrame dan difokuskan pada kolom userName, score, at, dan content. Ulasan tersebut diurutkan berdasarkan tanggal terbaru agar data yang diperoleh bersifat aktual dan terkini untuk keperluan analisis sentimen.



	userName	score	at	content
39	Pengguna Google	5	2025-07-21 16:30:30	semoga ini dapat menjadi pertimbangan develop...
40	Pengguna Google	5	2025-07-21 14:13:19	eksekusinya cepat
41	Pengguna Google	3	2025-07-21 07:05:35	Kritik saya ... CS tidak responsif, Verifikasi...
42	Pengguna Google	5	2025-07-21 03:40:22	mantaf
43	Pengguna Google	1	2025-07-21 03:16:06	duh gimama ini pas mau profit malah aplikasi n...

Gambar 4. 1 Hasil Scraping data (data Frame)

userName	score	at	content
Pengguna Googl	5	7/21/2025 16:30	semoga ini dapat menjadi pertimbangan developernya. untuk bagian ui desai tolong di tingkatkan terutama bagian tr
Pengguna Googl	5	7/21/2025 14:13	eksekusinya cepat
Pengguna Googl	3	7/21/2025 7:05	Kritik saya ... CS tidak responsif, Verifikasi live akun tdak sesuai jadwal, mencari pengaturannya susah , Sangat lemot
Pengguna Googl	5	7/21/2025 3:40	mantaf
Pengguna Googl	1	7/21/2025 3:16	duh gimama ini pas mau profit malah apliaksi ngeblank ðŸ˜ˆ jadi rugi
Pengguna Googl	5	7/19/2025 14:57	Broker yang Paling Terbaik ,Sampai ultah ajah di ucapin Trimakasih ðŸŒˆðŸŒˆ
Pengguna Googl	1	7/19/2025 11:38	nggak bisa masuk...
Pengguna Googl	5	7/19/2025 7:11	oke
Pengguna Googl	5	7/18/2025 11:40	bagus tools nya banyak, mau draw juga mudah udah di sediakan semua nice, minusnya ngelag signal nya
Pengguna Googl	5	7/18/2025 10:34	Cukup bagus. Tapi untuk penarikan dana cukup lama
Pengguna Googl	3	7/18/2025 4:23	Meskipun saya cuma menggunakan demonya tetapi tolong jumlah kerugian jangan sampai melebihi jumlah saldo inv
Pengguna Googl	5	7/17/2025 17:54	mantap , senang vbelajar masi pemula ðŸ™ smoga secara konsisten bisa lancar
Pengguna Googl	5	7/17/2025 17:22	aplikasi error,, keluar sendiri,, aplikasi tidak menanggapi... mungkin saatny pindah
Pengguna Googl	5	7/16/2025 17:15	sangat bagus
Pengguna Googl	1	7/16/2025 13:10	aplikasi apaan lemot bangt padahal jaringan bagus
Pengguna Googl	5	7/16/2025 8:42	coba aja dulu moga beneran menghasilkan uang lewat sini saya kasi bintang 5
Pengguna Googl	5	7/15/2025 22:29	aplikasi pernah barokah
Pengguna Googl	5	7/15/2025 14:05	Mantapp
Pengguna Googl	2	7/15/2025 13:44	saya takut main di mifx saat ini alasan nya stop los gak bisa di atur sama tp , tambah lagi loding apk gelag parah
Pengguna Googl	1	7/15/2025 12:56	apk tolol lemot mau entry ketinggalan trs TOL
Pengguna Googl	5	7/15/2025 9:53	goog

Gambar 4. 2 Data Hasil Scraping (CSV)

Sebanyak 4.559 data ulasan berhasil dikumpulkan dalam rentang waktu 9 November 2021 hingga 21 Juli 2025. Data mencakup kolom userName, score (rating), at (tanggal ulasan), dan content (isi ulasan). Namun, hanya kolom content yang digunakan dalam analisis

4.2. Hasil Pre-Processing Ringan

Pada tahap ini dilakukan pembersihan awal terhadap data ulasan pengguna aplikasi MIFX. Pre-processing ringan mencakup beberapa tahapan berikut:

- Case folding
- Cleaning
- Normalisasi

Tahapan ini bertujuan untuk merapikan data teks sehingga lebih siap untuk diproses lebih lanjut dan meningkatkan akurasi dalam analisis sentimen.

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4559 entries, 0 to 4558
Data columns (total 4 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   userName    4559 non-null   object  
 1   score       4559 non-null   int64   
 2   at          4559 non-null   object  
 3   content     4559 non-null   object  
dtypes: int64(1), object(3)
memory usage: 142.6+ KB

```

Gambar 4. 3 Import data dari file Csv ke DataFrame

Gambar 4.4 tersebut menunjukkan struktur data awal setelah file scrapping_NEWEST.csv dibaca ke dalam DataFrame. Dataset terdiri dari 4559 baris dan 4 kolom, yaitu userName, score, at, dan content, dengan tidak ada nilai yang hilang (non-null).

Cleaning

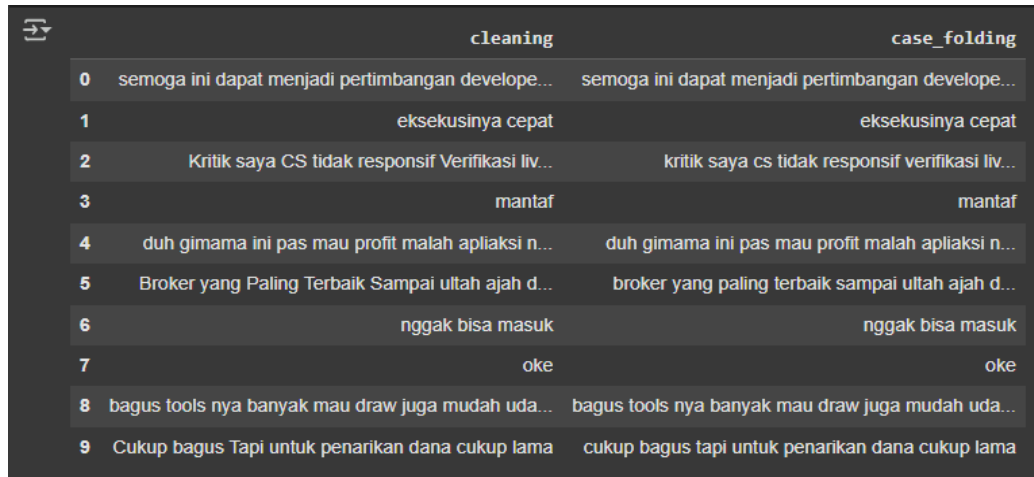
Setelah data didapatkan, data akan dibersihkan (Cleaning) dari karakter-karakter yang tidak dibutuhkan seperti emoji, simbol, angka, dan username.

	content	cleaning
0	semoga ini dapat menjadi pertimbangan develop...	semoga ini dapat menjadi pertimbangan develop...
1	eksekusinya cepat	eksekusinya cepat
2	Kritik saya ... CS tidak responsif, Verifikasi...	Kritik saya CS tidak responsif Verifikasi liv...
3	mantaf	mantaf
4	duh gimama ini pas mau profit malah aplikasi n...	duh gimama ini pas mau profit malah aplikasi n...
5	Broker yang Paling Terbaik ,Sampai ultah ajah ...	Broker yang Paling Terbaik Sampai ultah ajah d...
6	nggak bisa masuk...	nggak bisa masuk
7	oke	oke
8	bagus tools nya banyak, mau draw juga mudah ud...	bagus tools nya banyak mau draw juga mudah uda...
9	Cukup bagus. Tapi untuk penarikan dana cukup lama	Cukup bagus Tapi untuk penarikan dana cukup lama

Gambar 4. 4 Hasil Clening

Case Folding

Ulasan yang ada di Play Store menggunakan campuran huruf besar dan kecil. Agar seragam, semua huruf diubah menjadi huruf kecil (lowercase). Contohnya, kata “MIFX”, “Mifx”, dan “mifx” akan dianggap sama. Tahap ini penting untuk mengurangi redundansi data.



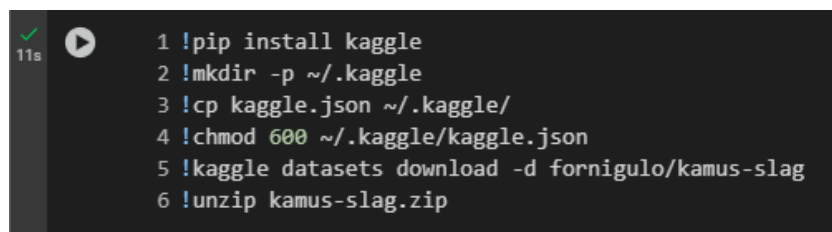
	cleaning	case_folding
0	semoga ini dapat menjadi pertimbangan develope...	semoga ini dapat menjadi pertimbangan develope...
1	eksekusinya cepat	eksekusinya cepat
2	Kritik saya CS tidak responsif Verifikasi liv...	kritik saya cs tidak responsif verifikasi liv...
3	mantaf	mantaf
4	duh gimama ini pas mau profit malah apliaksi n...	duh gimama ini pas mau profit malah apliaksi n...
5	Broker yang Paling Terbaik Sampai ultah ajah d...	broker yang paling terbaik sampai ultah ajah d...
6	nggak bisa masuk	nggak bisa masuk
7	oke	oke
8	bagus tools nya banyak mau draw juga mudah uda...	bagus tools nya banyak mau draw juga mudah uda...
9	Cukup bagus Tapi untuk penarikan dana cukup lama	cukup bagus tapi untuk penarikan dana cukup lama

Gambar 4. 5 Hasil Case Folding

Normalisasi

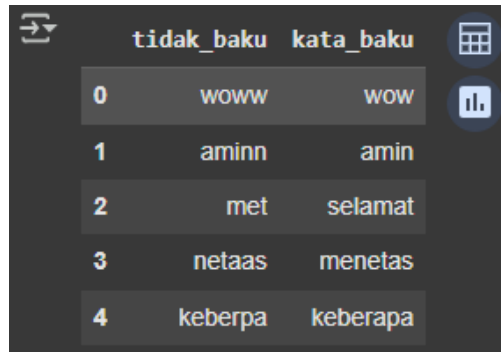
Selanjutnya dilakukan proses normalisasi, yaitu mengganti kata-kata tidak baku atau singkatan menjadi bentuk baku. Untuk mendukung proses ini, peneliti menggunakan kamus kata baku yang diunduh dari Kaggle.

<https://www.kaggle.com/datasets/fornigulo/kamus-slag>



```
1 !pip install kaggle
2 !mkdir -p ~/.kaggle
3 !cp kaggle.json ~/.kaggle/
4 !chmod 600 ~/.kaggle/kaggle.json
5 !kaggle datasets download -d fornigulo/kamus-slag
6 !unzip kamus-slag.zip
```

Gambar 4. 6 Instalasi Kaggle



	tidak_baku	kata_baku
0	wowww	wow
1	aminn	amin
2	met	selamat
3	netaas	menetas
4	keberpa	keberapa

Gambar 4. 7 Kamus Kata Baku

Gambar 4.9 menunjukkan kamus kata baku yang digunakan pada tahap preprocessing untuk melakukan normalisasi kata. Kamus ini berisi pasangan kata tidak baku beserta padanannya dalam bentuk baku, yang berfungsi untuk menyeragamkan penulisan kata pada teks ulasan. Proses ini penting dilakukan untuk mengurangi variasi penulisan kata yang berbeda namun memiliki makna yang sama, sehingga analisis sentimen dapat berjalan lebih akurat



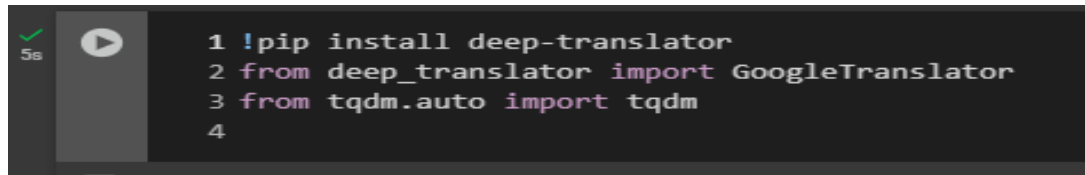
	case_folding	normalisasi
0	semoga ini dapat menjadi pertimbangan develop...	semoga ini dapat menjadi pertimbangan develop...
1	eksekusinya cepat	eksekusinya cepat
2	kritik saya cs tidak responsif verifikasi liv...	kritik saya cs tidak responsif verifikasi live...
3	mantaf	mantaf
4	duh gimama ini pas mau profit malah apliksi n...	duh gimama ini pas mau profit malah apliksi n...
5	broker yang paling terbaik sampai ultah ajah d...	broker yang paling terbaik sampai ultah saja d...
6	nggak bisa masuk	tidak bisa masuk
7	oke	oke

Gambar 4. 8 Hasil Normalisasi Kata

Translit Data ke Bahasa Inggris

Selanjutnya, dilakukan proses translasi teks ulasan ke dalam Bahasa Inggris. Translasi ini penting karena metode analisis sentimen VADER hanya bekerja optimal pada teks berbahasa Inggris. Proses penerjemahan dilakukan secara otomatis menggunakan library googletrans, sehingga seluruh isi ulasan yang

awalnya ditulis dalam Bahasa Indonesia dapat diubah ke Bahasa Inggris dan siap dianalisis menggunakan metode VADER.

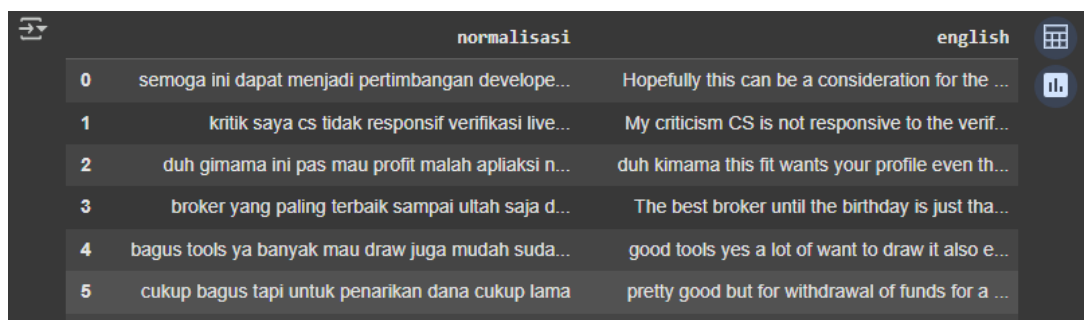


```

1 !pip install deep-translator
2 from deep_translator import GoogleTranslator
3 from tqdm.auto import tqdm
4

```

Gambar 4. 9 Instalasi deep-translator



	normalisasi	english
0	semoga ini dapat menjadi pertimbangan develope...	Hopefully this can be a consideration for the ...
1	kritik saya cs tidak responsif verifikasi live...	My criticism CS is not responsive to the verif...
2	duh gimama ini pas mau profit malah aplikasi n...	duh kimama this fit wants your profile even th...
3	broker yang paling terbaik sampai ultah saja d...	The best broker until the birthday is just tha...
4	bagus tools ya banyak mau draw juga mudah suda...	good tools yes a lot of want to draw it also e...
5	cukup bagus tapi untuk penarikan dana cukup lama	pretty good but for withdrawal of funds for a ...

Gambar 4. 10 Data Hasil Translit Pada Kolom “english”

Untuk keperluan penelitian, translasi ulasan ke Bahasa Inggris hanya dilakukan satu kali. Hasil translasi kemudian disimpan dan diunduh dalam bentuk file CSV. Hal ini dilakukan agar hasil pelabelan dan akurasi tidak berubah setiap kali dijalankan ulang, sehingga proses analisis tetap konsisten hingga akhir.

Hapus Data Duplikat

Sebelum data digunakan untuk analisis lebih lanjut, dilakukan pengecekan data duplikat. Dari total 4.559 data awal, hanya sebanyak 3.861 data yang tersisa setelah duplikat dihapus. Penghapusan dilakukan untuk memastikan bahwa tidak ada ulasan yang muncul lebih dari satu kali, agar hasil analisis tidak bias atau berulang.

```

<class 'pandas.core.frame.DataFrame'>
Index: 3861 entries, 0 to 4558
Data columns (total 7 columns):
#   Column          Non-Null Count  Dtype
---  -
0   at               3861 non-null   object
1   userName         3861 non-null   object
2   content          3861 non-null   object
3   cleaning         3838 non-null   object
4   case_folding     3838 non-null   object
5   normalisasi     3836 non-null   object
6   english          3836 non-null   object
dtypes: object(7)
memory usage: 241.3+ KB

```

Gambar 4. 11 HasilPembersihan Data Duplikat

4.3. Labeling dengan VADER

Selanjutnya, proses analisis sentimen dilakukan menggunakan metode VADER (Valence Aware Dictionary and sEntiment Reasoner).

```

Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl.metadata (572 bytes)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from vaderSentiment)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests)
Requirement already satisfied: certifi<2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests)
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
126.0/126.0 kB 5.3 MB/s eta 0:00:00
Installing collected packages: vaderSentiment
Successfully installed vaderSentiment-3.3.2

```

Gambar 4. 12 Instal Vader Sentiment

Data yang digunakan berasal dari file CSV hasil translasi, lalu dianalisis menggunakan fungsi `SentimentIntensityAnalyzer` dari pustaka `vaderSentiment`. Setiap baris teks akan diberi skor compound, yaitu skor keseluruhan sentimen dalam satu kalimat.

Nilai compound ini berada dalam rentang -1 hingga 1, dengan ketentuan sebagai berikut:

- Skor ≥ 0.05 dianggap positif
- Skor ≤ -0.05 dianggap negatif
- Skor di antara -0.05 dan 0.05 dianggap netral

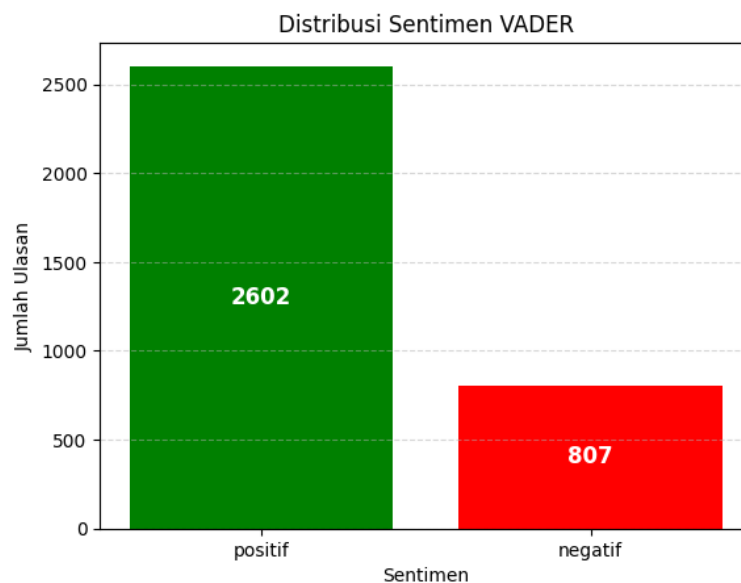
Setelah semua data diberi label, ulasan dengan kategori netral dihapus dalam penelitian ini, peneliti memilih untuk menghilangkan data berlabel netral, agar fokus analisis hanya pada sentimen positif dan negatif.

	english	compound	label
0	Hopefully this can be a consideration for the ...	0.9287	positif
1	My criticism CS is not responsive to the verif...	-0.7784	negatif
2	duh kimama this fit wants your profile even th...	0.1469	positif
3	The best broker until the birthday is just tha...	0.7717	positif
4	good tools yes a lot of want to draw it also e...	0.8910	positif
5	pretty good but for withdrawal of funds for a ...	0.4939	positif
6	Even though I only use the demo, please help t...	0.4215	positif
7	great happy vbale learning is still a beginner...	0.8885	positif
8	Error Application Own Own The application does...	-0.4019	negatif
9	very good	0.4927	positif

Jumlah Sentimen:	
label	
positif	2602
negatif	807
Name: count, dtype: int64	

Gambar 4. 13 Hasil Klasifikasi Sentimen dengan VADER

Dari hasil analisis sentimen menggunakan VADER, ditemukan bahwa mayoritas ulasan pengguna aplikasi MIFX bernada positif. Jumlah ulasan berlabel positif sebanyak 2.602, sedangkan yang negatif sebanyak 807.



Gambar 4. 14 Tabel Distribusi Sentimen VADER

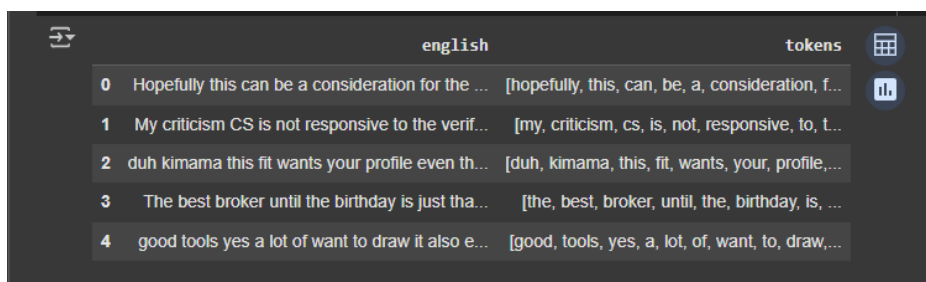
4.4. Hasil Pre-Processing Lanjutan

Setelah proses pelabelan selesai, data masuk ke tahap preprocessing lanjutan. Pada tahap ini dilakukan beberapa proses seperti tokenizing, stopwords removal, dan stemming. Tujuannya adalah untuk menyiapkan data teks agar bisa diolah lebih lanjut oleh model klasifikasi.

Hasil akhir dari tahap ini disimpan dalam kolom baru dan akan digunakan dalam proses pembobotan menggunakan TF-IDF.

Tokenizing

Tokenizing pada penelitian ini dilakukan untuk memecah setiap ulasan menjadi potongan kata. Proses ini diterapkan pada data ulasan yang sudah diterjemahkan ke dalam Bahasa Inggris.



	english	tokens
0	Hopefully this can be a consideration for the ...	[hopefully, this, can, be, a, consideration, f...
1	My criticism CS is not responsive to the verif...	[my, criticism, cs, is, not, responsive, to, t...
2	duh kimama this fit wants your profile even th...	[duh, kimama, this, fit, wants, your, profile,...
3	The best broker until the birthday is just tha...	[the, best, broker, until, the, birthday, is, ...
4	good tools yes a lot of want to draw it also e...	[good, tools, yes, a, lot, of, want, to, draw,...

Gambar 4. 15 Source Code dan Hasil Tokenizing

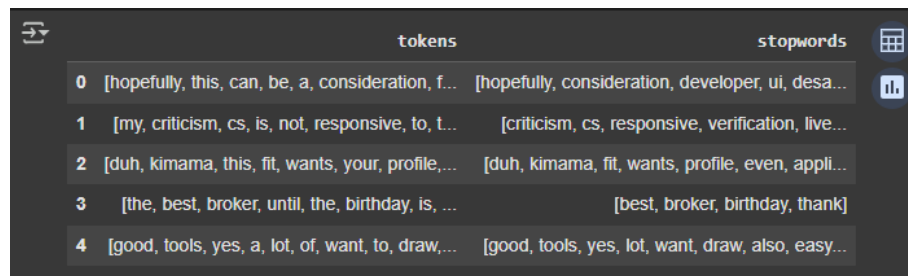
Hasil tokenizing disimpan dalam kolom “tokens” dan akan digunakan pada tahapan preprocessing lanjutan berikutnya, seperti stopwords removal dan stemming.

Stopword Removal/Filtering

Stopword removal pada penelitian ini dilakukan untuk menghilangkan kata-kata umum yang tidak memiliki pengaruh besar terhadap makna utama ulasan.

Proses ini diterapkan pada hasil tokenizing sebelumnya, menggunakan daftar stopwords Bahasa Inggris dari library nltk.

Hasil akhirnya disimpan dalam kolom stopwords, yang berisi kumpulan kata dari setiap ulasan yang telah dibersihkan dari kata-kata seperti “the”, “is”, “and”, dan lainnya.

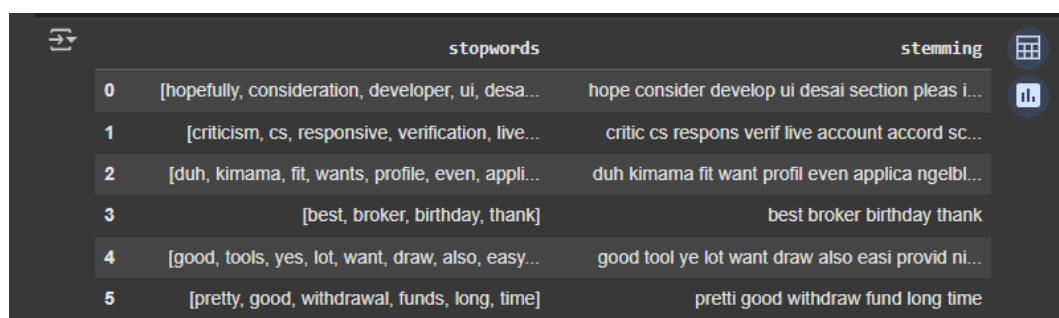


	tokens	stopwords
0	[hopefully, this, can, be, a, consideration, f...	[hopefully, consideration, developer, ui, desa...
1	[my, criticism, cs, is, not, responsive, to, t...	[criticism, cs, responsive, verification, live...
2	[duh, kimama, this, fit, wants, your, profile,...	[duh, kimama, fit, wants, profile, even, appli...
3	[the, best, broker, until, the, birthday, is, ...	[best, broker, birthday, thank]
4	[good, tools, yes, a, lot, of, want, to, draw,...	[good, tools, yes, lot, want, draw, also, easy...

Gambar 4. 16 Source Code dan Hasil Stopword

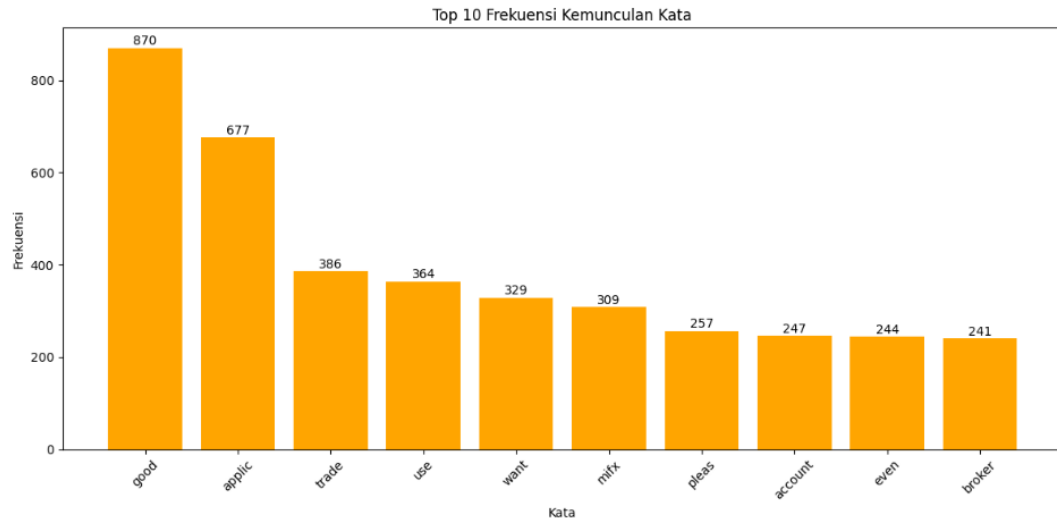
Stemming

Setelah proses sebelumnya selesai, tahap selanjutnya adalah stemming. Tujuannya adalah untuk menyamakan bentuk kata dengan cara memotongnya ke bentuk dasar. Proses ini menggunakan PorterStemmer dari library nltk, dan hasilnya disimpan dalam kolom stemming.



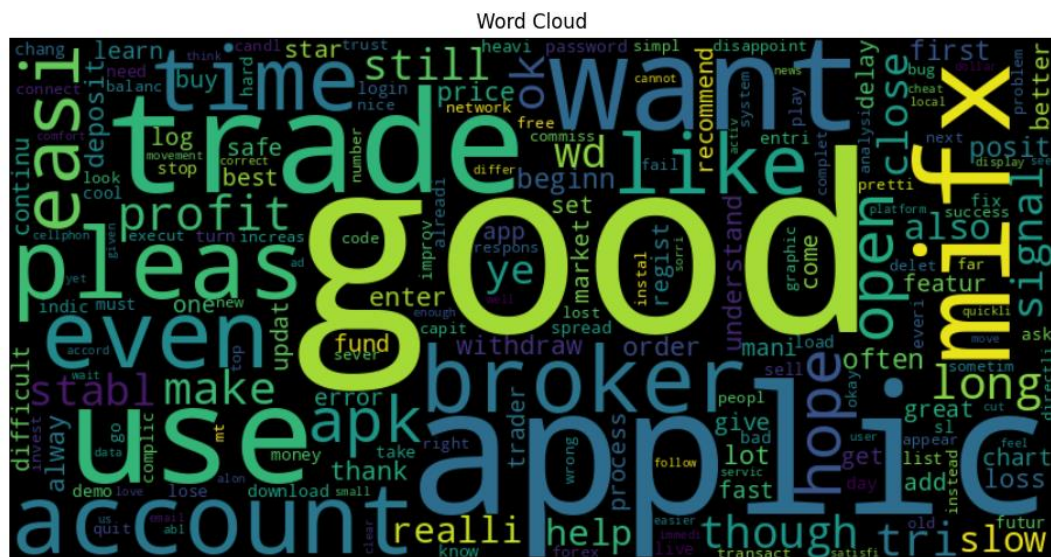
	stopwords	stemming
0	[hopefully, consideration, developer, ui, desa...	hope consider develop ui desai section pleas i...
1	[criticism, cs, responsive, verification, live...	critic cs respons verif live account accord sc...
2	[duh, kimama, fit, wants, profile, even, appli...	duh kimama fit want profil even applica ngelbl...
3	[best, broker, birthday, thank]	best broker birthday thank
4	[good, tools, yes, lot, want, draw, also, easy...	good tool ye lot want draw also easi provid ni...
5	[pretty, good, withdrawl, funds, long, time]	pretti good withdraw fund long time

Gambar 4. 17 Hasil Stemming



Gambar 4. 18 Diagram Frekuensi Kemunculan Kata

Gambar tersebut menampilkan 10 kata teratas yang paling sering muncul setelah dilakukan proses stemming pada data ulasan pengguna aplikasi MIFX. Kata "good" menjadi yang paling dominan dengan frekuensi 870, diikuti oleh "applic", "trade", dan "use". Kata-kata ini mencerminkan fokus ulasan yang banyak membahas kualitas aplikasi, aktivitas trading, serta pengalaman penggunaan secara umum.

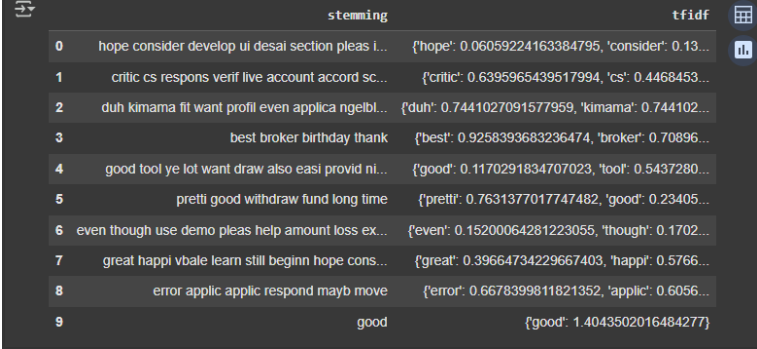


Gambar 4. 19 Word Cloud Stemming

Pada word cloud juga menampilkan hal yang sama, yaitu dominasi kata-kata dengan sentimen positif seperti “good”, “app”, “account”, “trade”, dan “please”. Kata-kata ini muncul paling sering dalam ulasan pengguna, ditandai dengan ukuran huruf yang lebih besar

4.5. Pembobotan TF-IDF

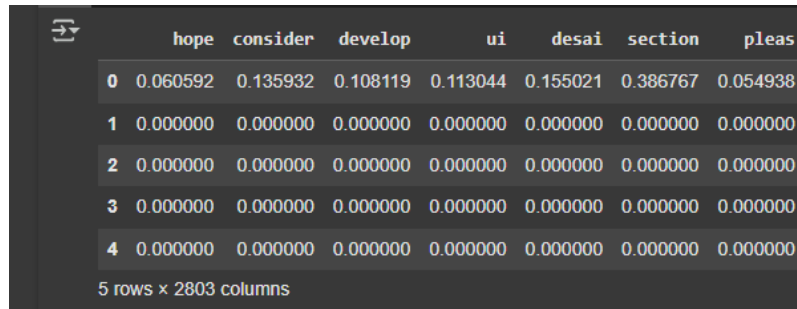
Pada tahap ini dilakukan perhitungan TF-IDF dari kolom stemming. Nilai IDF dihitung untuk mengukur kejarangan kata dalam seluruh dokumen, kemudian dikalikan dengan frekuensi kata (TF) sehingga diperoleh bobot TF-IDF. Bobot ini menunjukkan pentingnya kata pada setiap ulasan, di mana kata yang jarang muncul di banyak dokumen tetapi sering muncul di satu dokumen akan memiliki nilai lebih tinggi. Hasil akhirnya disimpan dalam kolom tfidf untuk keperluan analisis. Seperti pada Gambar 4.23.



	stemming	tfidf
0	hope consider develop ui desai section pleas i...	{'hope': 0.06059224163384795, 'consider': 0.13...
1	critic cs respons verif live account accord sc...	{'critic': 0.6395965439517994, 'cs': 0.4468453...
2	duh kimama fit want profil even applica ngelbl...	{'duh': 0.7441027091577959, 'kimama': 0.744102...
3	best broker birthday thank	{'best': 0.9258393683236474, 'broker': 0.70896...
4	good tool ye lot want draw also easi provid ni...	{'good': 0.1170291834707023, 'tool': 0.5437280...
5	pretti good withdraw fund long time	{'pretti': 0.7631377017747482, 'good': 0.23405...
6	even though use demo pleas help amount loss ex...	{'even': 0.15200064281223055, 'though': 0.1702...
7	great happi vbale learn still beginn hope cons...	{'great': 0.39664734229667403, 'happi': 0.5766...
8	error applic applic respond mayb move	{'error': 0.6678399811821352, 'applic': 0.6056...
9	good	{'good': 1.4043502016484277}

Gambar 4. 20 Hasil TF IDF

Hasil perhitungan TF-IDF ditampilkan dalam bentuk matriks, di mana setiap baris merepresentasikan dokumen dan setiap kolom adalah kata unik. Misalnya, pada dokumen pertama kata hope, consider, dan develop memiliki bobot tertentu, sedangkan kata yang tidak muncul akan bernilai nol. Representasi ini mengubah teks menjadi data numerik yang dapat digunakan untuk analisis sentimen



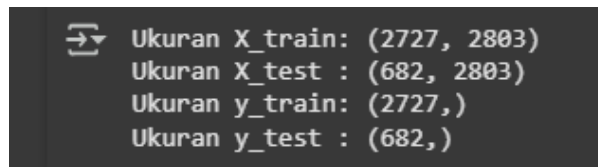
	hope	consider	develop	ui	desai	section	pleas
0	0.060592	0.135932	0.108119	0.113044	0.155021	0.386767	0.054938
1	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
2	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

5 rows x 2803 columns

Gambar 4. 21 TF-IDF bentuk Matrix

4.6. Split Data

Setelah proses pembobotan selesai, data dibagi menjadi dua bagian menggunakan metode train-test split. Pembagian dilakukan dengan rasio 80% untuk data latih dan 20% untuk data uji, agar model bisa dilatih dan dievaluasi secara terpisah.



```

Ukuran X_train: (2727, 2803)
Ukuran X_test : (682, 2803)
Ukuran y_train: (2727,)
Ukuran y_test : (682,)

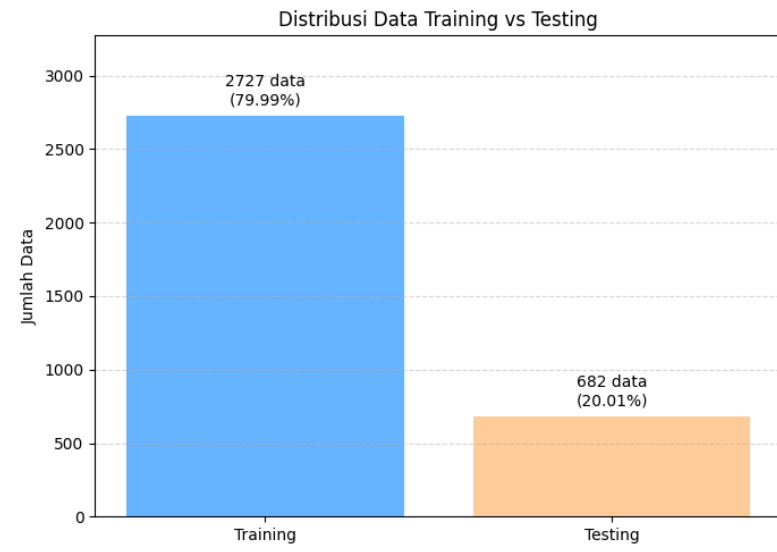
```

Gambar 4. 22 Source Code dan Hasil Split Data

Dari hasil pembagian pada Gambar 4.25, didapatkan:

- X_train berisi 2.727 data latih dengan 2.803 fitur (kata)
- X_test berisi 682 data uji dengan jumlah fitur yang sama
- y_train berisi 2.727 label
- y_test berisi 682 label

Pembagian ini bertujuan untuk menguji performa model secara objektif menggunakan data yang belum pernah dilihat sebelumnya saat pelatihan.



Gambar 4. 23 Diagram Distribusi Data Training dan Testing

4.7. Klasifikasi Dengan Naïve bayes

Pada bagian ini, dilakukan evaluasi terhadap model Naive Bayes yang telah dilatih dan diuji menggunakan data yang telah displit sebelumnya. Data X pada proses ini berasal dari hasil pembobotan TF-IDF pada kolom ulasan yang telah melalui tahap preprocessing, sedangkan data y berasal dari label sentimen yang dihasilkan oleh metode VADER. Hasil evaluasi menunjukkan akurasi model sebesar 85.48%, yang menggambarkan seberapa baik model dalam mengklasifikasikan sentimen pada data testing.

```

➡ Akurasi: 85.48 %

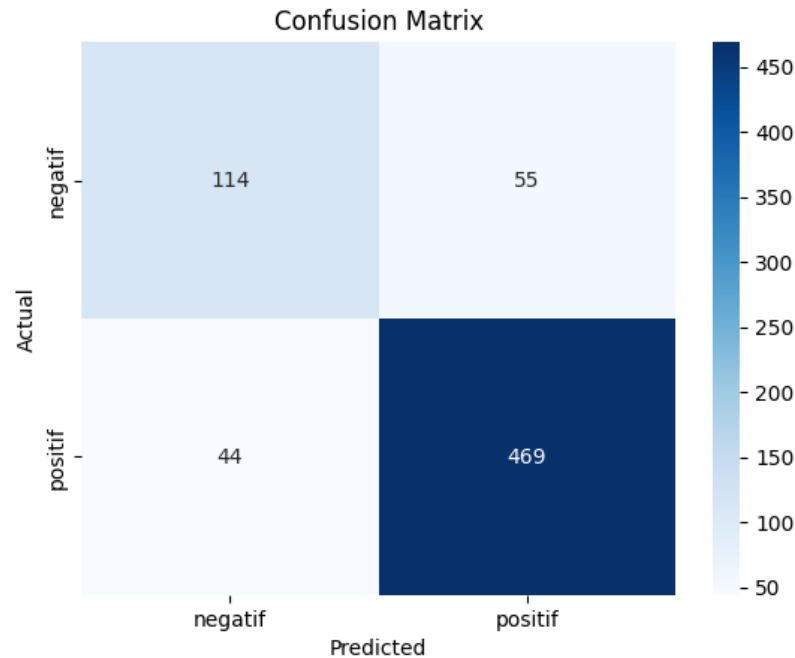
Classification Report:
      precision    recall  f1-score   support

negatif    0.72    0.67    0.70      169
positif    0.90    0.91    0.90      513

accuracy    0.85    0.85    0.85      682
macro avg   0.81    0.79    0.80      682
weighted avg 0.85    0.85    0.85      682
  
```

Gambar 4. 24 Hasil Classification Report

Gambar 4.28 adalah Hasil klasifikasi untuk model Naive Bayes, yang menunjukkan nilai precision, recall, dan F1-score per kelas



Gambar 4. 25 Confusion Matrix

Berdasarkan hasil pengujian model Naive Bayes pada data testing, diperoleh confusion matrix yang menunjukkan jumlah prediksi benar dan salah pada masing-masing kelas sentimen.

- Nilai True Positive (TP) sebanyak 469 menunjukkan data sentimen positif yang berhasil diprediksi dengan benar
- True Negative (TN) sebanyak 114 menunjukkan data sentimen negatif yang diprediksi dengan benar.
- False Positive (FP) sebanyak 55 adalah data negatif yang salah diprediksi positif
- False Negative (FN) sebanyak 44 adalah data positif yang salah diprediksi negative

4.8. Hasil Evaluasi

Akurasi model Naive Bayes pada data testing adalah 85.48%, yang dihitung menggunakan rumus berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} = \frac{469 + 114}{469 + 114 + 55 + 44} = \frac{583}{682} = 85.49\%$$

Ini menunjukkan bahwa model berhasil memprediksi dengan benar 85.48% dari data testing yang diberikan.

Clasification report

Peneliti menggunakan data confusion matrix yang sudah ada:

Tabel 4. 1 Cofusion Matrix Tanpa Nilai

	Prediksi Negatif	Prediksi positif
Aktual Negatif	TN	FP
Aktual Positif	FN	TP

Tabel 4. 2 Confusion Matrix Keseluruhan

	Prediksi Negatif	Prediksi positif
Aktual Negatif	114	55
Aktual Positif	44	469

1. Sudut Pandang Kelas Positif

(positif dianggap sebagai “positif” dalam perhitungan metrik)

- TP = 469 (positif yang benar diprediksi positif)
- FN = 44 (positif yang salah jadi negatif)
- FP = 55 (negatif yang salah jadi positif)

- **TN** = 114 (negatif yang benar diprediksi negatif)

Tabel 4. 3 Confusion Matrix Kelas Positif

	Prediksi Positif	Prediksi Negatif	total
Aktual Positif	TP=469	FN=44	513
Aktual Negatif	FP=55	TN=114	169

$$Presisi(Positif) = \frac{TP}{TP + FP} = \frac{469}{469 + 55} = \frac{469}{524} = 0.8950 = 89.50\%$$

$$Recal(Positif) = \frac{TP}{TP + FN} = \frac{469}{469 + 44} = \frac{469}{513} = 0.9143 = 91.43\%$$

F1 score (Positif)

$$F1 = 2 \times \frac{Presisi \times Recal}{Presisi + Recal} = 2 \times \frac{0.8950 \times 0.9143}{0.8950 + 0.9143} = 2 \times \frac{0.8182}{1.8093} = 0.9044$$

$$= 90.45\%$$

2. Sudut Pandang Kelas Negatif

(negatif dianggap sebagai “positif” dalam perhitungan metrik)

- **TP** = 114 (negatif yang benar diprediksi negatif)
- **FN** = 55 (negatif yang salah jadi positif)
- **FP** = 44 (positif yang salah jadi negatif)
- **TN** = 469 (positif yang benar diprediksi positif)

Tabel 4. 4 Confusion Matrix Kelas Negatif

	Prediksi Negatif	Prediksi positif	total
Aktual Negatif	TP=114	FN= 55	169
Aktual Positif	FP=44	TN= 469	513

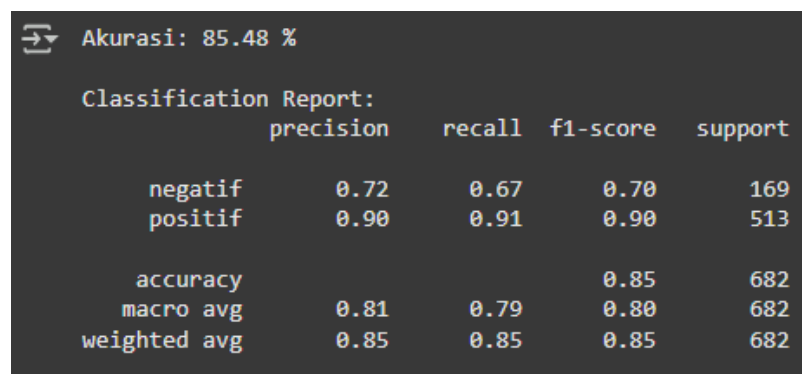
$$Presisi(Negatif) = \frac{TP}{TP + FP} = \frac{114}{114 + 44} = \frac{114}{158} = 0.7215 = 72.15\%$$

$$Recal(Negatif) = \frac{TP}{TP + FN} = \frac{114}{114 + 55} = \frac{114}{169} = 0.6745 = 67.45\%$$

F1 score (Negatif)

$$F1 = 2 \times \frac{Presisi \times Recal}{Presisi + Recal} = 2 \times \frac{0.7215 \times 0.6745}{0.7215 + 0.6745} = 2 \times \frac{0.4866}{1.396} = 0.6971$$

$$= 69.75\%$$



Classification Report:				
	precision	recall	f1-score	support
negatif	0.72	0.67	0.70	169
positif	0.90	0.91	0.90	513
accuracy			0.85	682
macro avg	0.81	0.79	0.80	682
weighted avg	0.85	0.85	0.85	682

Gambar 4. 26 Hasil Evaluasi

Model Naïve Bayes menghasilkan akurasi sebesar 85,48%. Berdasarkan interpretasi akurasi dalam penelitian (Hartmann et al., 2023), akurasi di atas 80% dikategorikan sebagai tinggi (high), sementara rentang 60–80% dipandang sedang (moderate), dan $\leq 60\%$ rendah (low). Dengan demikian, akurasi 85,48% yang diperoleh dalam penelitian ini dapat dikategorikan sebagai tinggi.

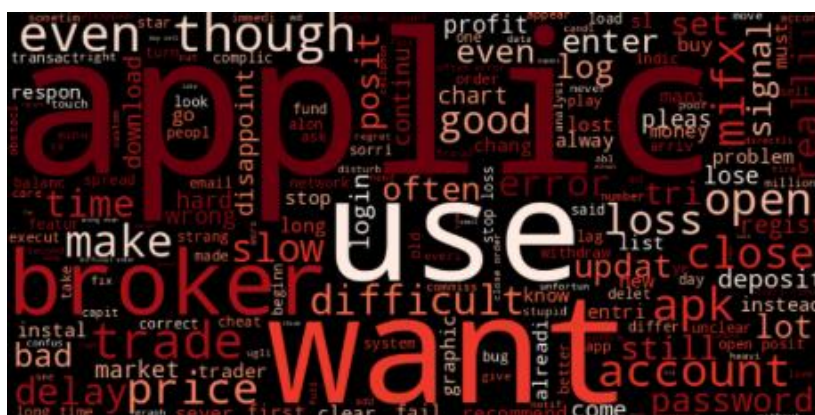
4.9. Hasil Visualisasi

Tahap akhir analisis dilakukan dengan membuat visualisasi Word Cloud untuk melihat kata-kata yang paling sering muncul pada masing-masing kategori sentimen.



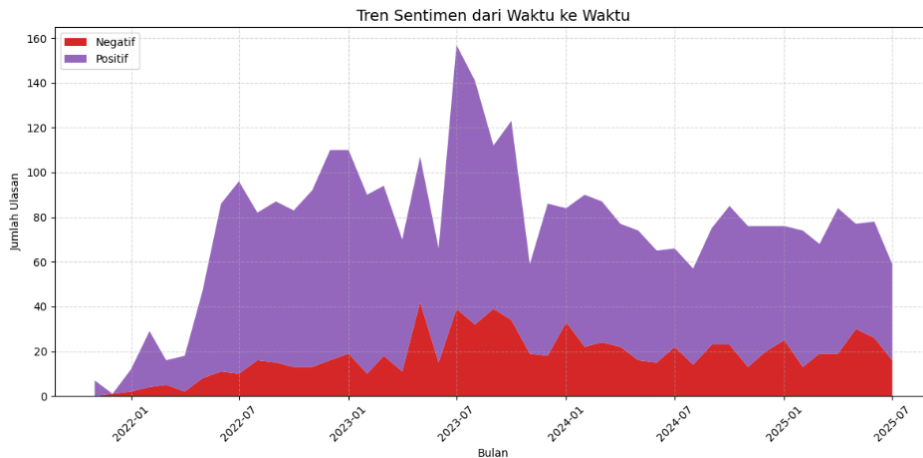
Gambar 4. 27 WordCloud Sentimen Positif

Gambar 4.31 adalah visualisasi WordCloud sentimen positif yang menampilkan kata-kata dominan seperti “good”, “applic”, “use”, “mifx”, dan “trade”.



Gambar 4. 28 WordCloud Sentimen Negatif

Gambar 4.32 merupakan isualisasi Word Cloud sentimen negatif yang memperlihatkan kata-kata dominan seperti “applic”, “use”, “want”, “broker”, dan “error”.



Gambar 4. 29 Grafik Area Bertumpuk Tren Sentimen

Grafik ini menunjukkan tren sentimen positif dan negatif dari awal 2022 hingga pertengahan 2025. Secara umum, sentimen positif selalu lebih dominan dibandingkan sentimen negatif sepanjang periode tersebut.

4.10. Implementasi Sistem

Sebagai bagian akhir dari penelitian, seluruh metodologi yang telah dijelaskan diimplementasikan ke dalam sebuah sistem aplikasi web yang cukup sederhana. Aplikasi web ini dibangun menggunakan Python dengan framework Flask sebagai backend untuk pemrosesan data, dan HTML, CSS, serta JavaScript untuk antarmuka pengguna (frontend). Sistem ini dirancang untuk memvisualisasikan setiap tahapan analisis sentimen secara interaktif, memungkinkan pengguna untuk mengikuti alur proses dari awal hingga akhir.

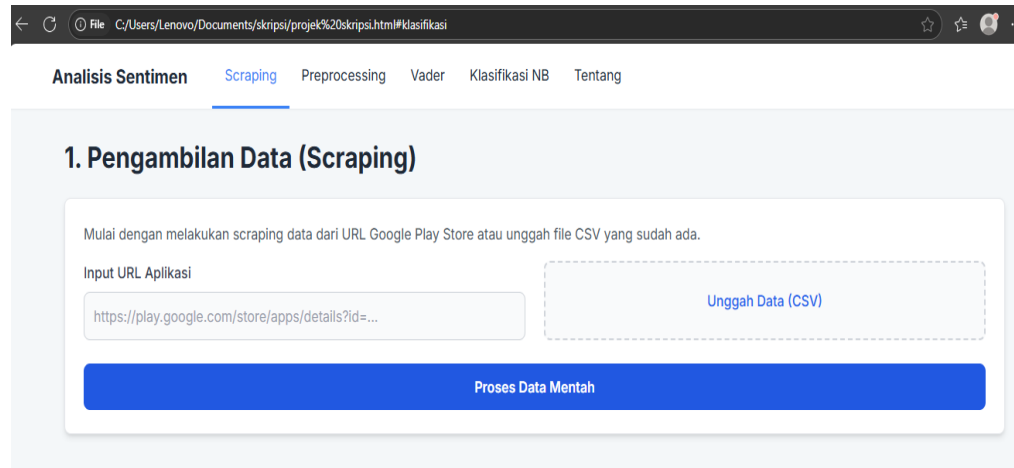
Berikut adalah tampilan dari setiap tahapan implementasi sistem:

A. Halaman Utama

Bagian ini merupakan tampilan awal aplikasi. Halaman utama menampilkan identitas penelitian berupa judul, nama peneliti, serta menyediakan navigasi menuju fitur-fitur analisis. Dari halaman ini, pengguna dapat berpindah ke

bagian scraping, preprocessing, pelabelan dengan VADER, klasifikasi Naïve Bayes, maupun informasi tambahan.

Tampilan halaman utama ditunjukkan pada Gambar 4.30

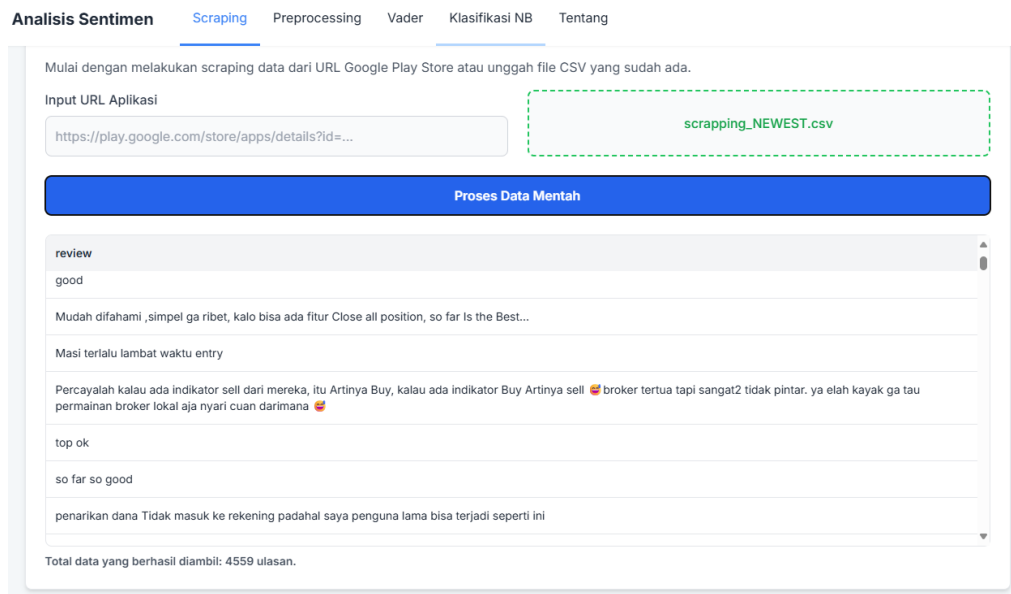


Gambar 4. 30 Tampilan Halaman Utama

B. Scraping Data

Fitur scraping digunakan untuk mengumpulkan ulasan pengguna aplikasi MIFX dari Google Play Store. Data yang berhasil diambil ditampilkan dalam bentuk tabel, lengkap dengan nama pengguna, rating, serta isi ulasan. Hasil pengambilan data juga dapat disimpan dalam format CSV agar lebih mudah diproses pada tahap selanjutnya.

Ilustrasi tampilan scraping dapat dilihat pada Gambar 4.2.

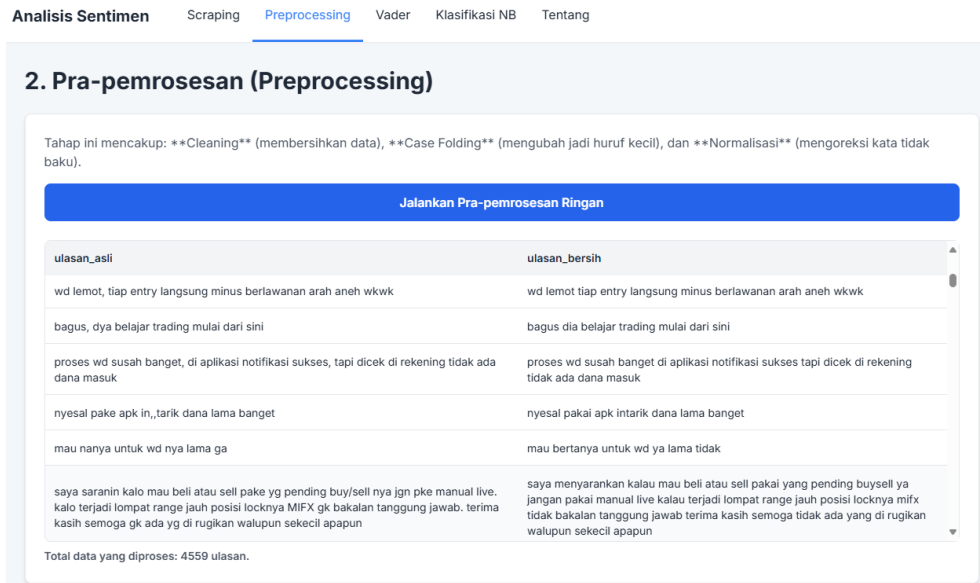


Gambar 4. 31 Tampilan Scraping

C. Preprocessing

Tahap preprocessing ditampilkan dalam bentuk halaman yang memperlihatkan hasil pembersihan teks ulasan. Proses ini mencakup case folding, cleaning, normalisasi, tokenisasi, stopwords removal, dan stemming. Pada tampilan aplikasi diperlihatkan perbandingan teks ulasan sebelum dan sesudah diproses sehingga perubahan dapat diamati secara langsung.

Contoh tampilan preprocessing dapat dilihat pada Gambar 4.3.

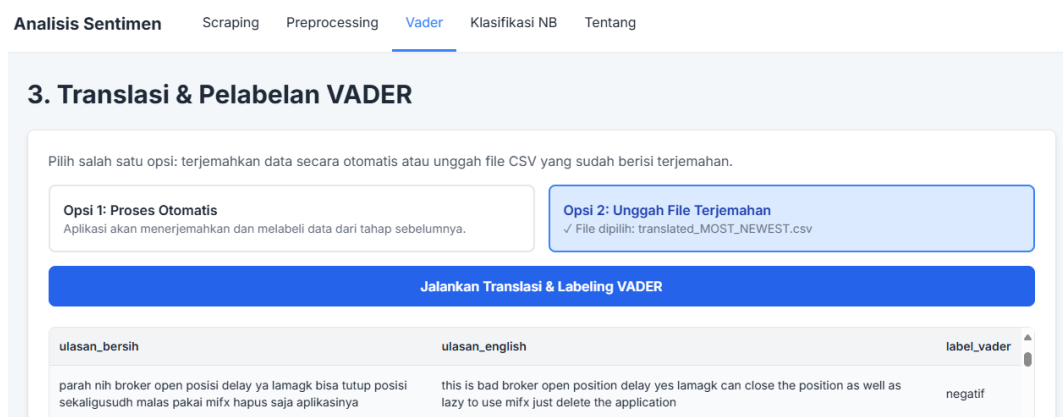


Gambar 4. 32 Tampilan Halaman Preprocessing

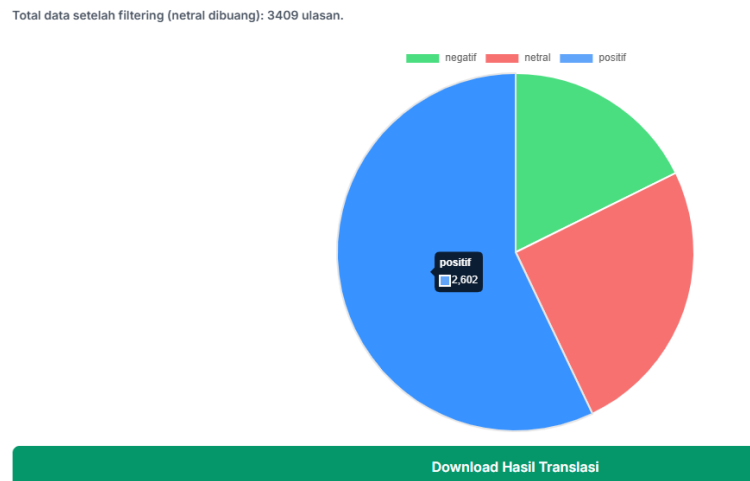
D. Pelabelan dengan Vader

Halaman ini menampilkan hasil pelabelan sentimen menggunakan metode VADER. Ulasan yang sudah diterjemahkan ke Bahasa Inggris dianalisis untuk mendapatkan skor sentimen, kemudian dikategorikan menjadi positif, negatif, atau netral. Output sistem berupa tabel yang berisi teks ulasan, nilai skor, dan label sentimennya.

Tampilan pelabelan VADER dapat dilihat pada Gambar 4.4.



Gambar 4. 33 Tampilan Translit dan Pelabelan Vader

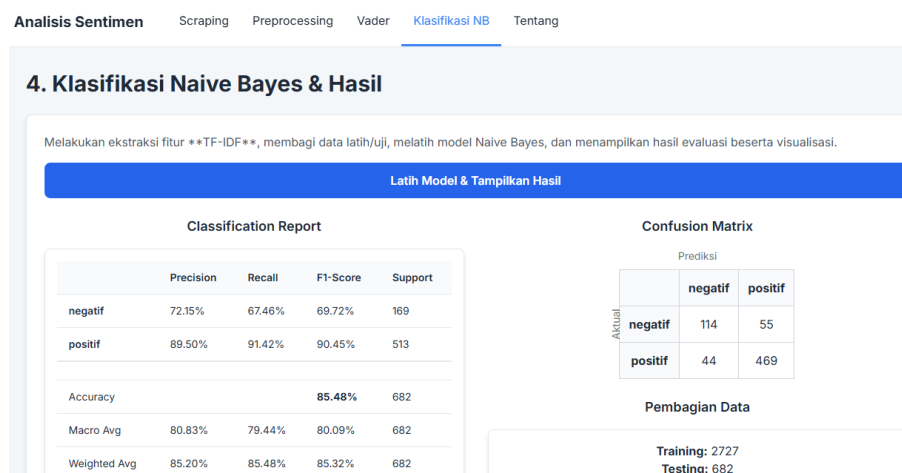


Gambar 4. 34 Diagram Lingkaran Distribusi Sentimen

E. Klasifikasi Naïve Bayes

Proses klasifikasi dengan algoritma Naïve Bayes juga disajikan melalui aplikasi. Data yang sudah diolah dengan TF-IDF digunakan untuk melatih model, kemudian hasilnya ditampilkan berupa evaluasi akurasi, precision, recall, dan F1-score. Selain itu, halaman ini juga memperlihatkan confusion matrix serta visualisasi tambahan berupa distribusi sentimen dan word cloud.

Tampilan klasifikasi ditunjukkan pada Gambar 4.5.



Gambar 4. 35 Hasil Evaluasi Model Naïve Bayes

F. Basis Data

sistem menyimpan hasil pengumpulan ulasan ke dalam tabel `data_ulasan` yang telah dirancang sebelumnya. Data yang masuk berasal dari proses scraping maupun dari dataset yang diunggah melalui sistem. Seluruh data yang tersimpan dapat diakses kembali melalui phpMyAdmin untuk memastikan proses penyimpanan berjalan dengan baik.

Showing rows 0 - 24 (4559 total, Query took 0.0005 seconds.)

SELECT * FROM `data\_ulasan`

Profiling

Edit inline

Edit

Explain SQL

Create PHP code

Refresh

1

>

>>

Number of rows: 25

Filter rows: Search this table

Sort by key: None

Extra options

		id	sesi_id	tahap	ulasan_asli	ulasan_bersih	ulasan_english	label_vader	dibuat_pada
☐	Edit Copy Delete	1	44051b78-b69c-419f-8902-2efbfa3abc97	0	semoga ini dapat menjadi pertimbangan developernya...	semoga ini dapat menjadi pertimbangan developernya...	Hopefully this can be a consideration for the deve...	positif	2025-08-22 02:05:00
☐	Edit Copy Delete	2	419f-8902-2efbfa3abc97	0	eksekusinya cepat	eksekusinya cepat	The execution is fast	netral	2025-08-22 02:05:00

Gambar 4. 36 Tampilan tabel "data_ulasan" pada phpMyAdmin

Gambar 4.xx memperlihatkan hasil implementasi tabel `data_ulasan` yang digunakan dalam penelitian ini. Pada tahap ini, sebanyak 4.559 data ulasan berhasil tersimpan di basis data setelah melalui proses pengumpulan

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah diuraikan pada bab-bab sebelumnya mengenai analisis sentimen ulasan pengguna aplikasi MIFX menggunakan metode VADER dan Naïve Bayes, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Kombinasi metode VADER sebagai labeling tool dan Naïve Bayes sebagai klasifikator terbukti efektif untuk melakukan analisis sentimen pada ulasan aplikasi MIFX. Model klasifikasi Naïve Bayes yang dibangun mampu mencapai tingkat akurasi sebesar 85,48% pada data pengujian. Nilai ini termasuk dalam kategori tinggi, karena berada pada ambang di atas 80% Hal ini menunjukkan bahwa model memiliki kemampuan yang baik dalam memprediksi sentimen ulasan secara otomatis.
2. Performa model menunjukkan perbedaan dalam mengenali sentimen positif dan negatif. Model lebih unggul dalam mengidentifikasi sentimen positif dengan nilai precision, recall, dan F1-score mencapai 90%, 91%, dan 90%. Sementara itu, untuk sentimen negatif, performanya lebih rendah dengan nilai masing-masing 72%, 67%, dan 70%.
3. Hasil pelabelan menggunakan VADER menunjukkan bahwa dari total 3.409 ulasan yang dianalisis (setelah data netral dihilangkan), sebanyak 2.602 ulasan (76,3%) teridentifikasi sebagai sentimen positif, dan 807 ulasan (23,7%)

sebagai sentimen negatif. Distribusi ini mengindikasikan bahwa persepsi mayoritas pengguna terhadap aplikasi MIFX cenderung positif.

4. Seluruh alur penelitian, mulai dari pengumpulan data, analisis, hingga evaluasi, berhasil diimplementasikan ke dalam sebuah prototipe aplikasi web yang fungsional. Prototipe ini mampu mendemonstrasikan setiap tahapan penelitian secara interaktif dan membuktikan bahwa model yang diusulkan dapat diterapkan dalam sebuah sistem nyata.

5.2. Saran

Berdasarkan kesimpulan yang telah dipaparkan, terdapat beberapa saran yang dapat menjadi pertimbangan untuk pengembangan penelitian sejenis di masa mendatang:

1. Pengembangan Dataset: Untuk meningkatkan kemampuan generalisasi model, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar dan beragam, baik dari rentang waktu yang lebih panjang maupun dari sumber platform lain selain Google Play Store.
2. Eksplorasi Algoritma Alternatif: Disarankan untuk melakukan studi komparatif dengan membandingkan performa algoritma Naïve Bayes dengan metode klasifikasi lain yang lebih kompleks, seperti Support Vector Machine (SVM), Random Forest, atau bahkan pendekatan deep learning seperti LSTM dan BERT yang mungkin dapat menangkap konteks kalimat dengan lebih baik.
3. Penyempurnaan Tahap Preprocessing: Proses preprocessing dapat diperkaya dengan teknik yang lebih lanjut, seperti lemmatization

(mengubah kata ke bentuk dasar leksikalnya) dan negation handling (penanganan kata-kata negasi seperti "tidak", "bukan") untuk meningkatkan akurasi, terutama dalam mendeteksi sentimen negatif.

4. Pengembangan Prototipe: Prototipe yang telah dibangun dapat dikembangkan lebih lanjut menjadi sebuah aplikasi dasbor pemantauan sentimen yang berjalan secara real-time, sehingga dapat dimanfaatkan secara praktis oleh pengembang aplikasi MIFX untuk memonitor umpan balik pengguna secara berkelanjutan.

DAFTAR PUSTAKA

- Abimanyu, D., Budianita, E., Cynthia, E. P., Yanto, F., & Yusra, Y. (2022). Analisis Sentimen Akun Twitter Apex Legends Menggunakan VADER. *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, 5(3), 423–431. <https://doi.org/10.32672/jnkti.v5i3.4382>
- Afifah Faadilah. (2020). *Analisis sentimen pada ulasan aplikasi tokopedia di google play store menggunakan metode long short term memory*.
- Alfiah, F. I. (2022). *No Title*. Duwitmu.Com. <https://duwitmu.com/investasi/mifx-indonesia-monex-review>
- Arifin, S., & Febryanto, B. A. (2025). *Analisis Sentimen Pengguna Aplikasi Merdeka Mengajar Menggunakan Metode Vader Lexicon Sentiment Analysis of Merdeka Mengajar Application Users Using the Vader Lexicon Method*. 15(1), 1–11.
- Barik, K., & Misra, S. (2024). Analysis of customer reviews with an improved VADER lexicon classifier. *Journal of Big Data*, 11(1). <https://doi.org/10.1186/s40537-023-00861-x>
- Castro, O., Bruneau, P., Sottet, J. S., & Torregrossa, D. (2024). Landscape of High-Performance Python to Develop Data Science and Machine Learning Applications. *ACM Computing Surveys*, 56(3), 1–30. <https://doi.org/10.1145/3617588>
- Ernianti Hasibuan, & Elmo Allistair Heriyanto. (2022). Analisis Sentimen Pada Ulasan Aplikasi Amazon Shopping Di Google Play Store Menggunakan Naive Bayes Classifier. *Jurnal Teknik Dan Science*, 1(3), 13–24. <https://doi.org/10.56127/jts.v1i3.434>

- Febrian Surya Kristardi. (2024). *Review Broker MIFX Terbaru*. Rankia.Id.
<https://rankia.id/review-mifx/>
- Gabriel Do Breviário, Á., Marques De Souza, J., Lucena, J. B., & Rago, L. F. (2025). Potentialities and Challenges Do Google Colab To Machine Learning and Big Data Analytics. *Current Scientific Journal*, 1759(7), 505.
<https://doi.org/10.5281/zenodo.15537694>
- Gifari, O. I., Adha, M., Freddy, F., & Durrand, F. F. S. (2022). Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine. *Journal of Information Technology*, 2(1), 36–40.
<https://doi.org/10.46229/jifotech.v2i1.330>
- Gultom, M., Marikros, J., & Rusli, W. (2024). *Penerapan Vader Sentiment untuk Mendeteksi Sentimen Bahasa Inggris berbasis Website*. 13–18.
- H.Taufiqurrahman, F.Anggraeny, M. H. (2023). *PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR PADA ANALISIS SENTIMEN ULASAN APLIKASI MYPERTAMINA*. 7(6), 3934–3939.
- Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing*, 40(1), 75–87.
<https://doi.org/10.1016/j.ijresmar.2022.05.005>
- Kusnawi, K., & Anam, K. (2025). Comparison ff Sentiment Labeling Using Textblob, Vader, and Flair in Public Opinion Analysis Post-2024 Presidential Inauguration with IndoBERT. *Jurnal Teknik Informatika (Jutif)*, 6(2), 803–818. <https://doi.org/10.52436/1.jutif.2025.6.2.4015>
- Leonardi, V. H., Ibrahim, A., Kurnia, R. D., & Afrina, M. (2025). *Analisis Sentimen*

- Ulasan Aplikasi Pembelajaran Bahasa Menggunakan Metode VADER.*
<https://doi.org/10.33364/algorithm/v.22-1.2285>
- Monex Investindo Futures. (2025). *Website Resmi MIFX*. PT Monex Investindo Futures. <https://www.mifx.com/>
- Musfiroh, D., Khaira, U., Utomo, P. E. P., & Suratno, T. (2021). Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 24–33. <https://doi.org/10.57152/malcom.v1i1.20>
- Pramukti, S. D., Nugroho, A., & Sunge, A. S. (2022). Analisis Sentimen Masyarakat Dengan Metode Naïve Bayes dan Particle Swarm Optimization. *Techno.Com*, 21(1), 61–74. <https://doi.org/10.33633/tc.v21i1.5332>
- R.Sanjaya, E.Tohidi, E. W. (2024). *MEDIA SOSIAL TWITTER MENGGUNAKAN ALGORITMA NAÏVE BAYES*. 8(1), 507–514.
- Rizquina, A. Z., & Ratnasari, C. I. (2023). Implementasi Web Scraping untuk Pengambilan Data Pada Website E-Commerce. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, 5(4), 377–383. <https://doi.org/10.47233/jteksis.v5i4.913>
- Siregar, M. Z., Elhanafi, A. M., Irwan, D., Informasi, S., Medan, U. H., Bayes, N., Play, G., & Media, S. (2025). *ANALISIS SENTIMEN TERHADAP ULASAN APLIKASI MEDIA SOSIAL DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES*. 9(2), 3373–3381. <https://ejournal.itn.ac.id/index.php/jati/article/download/12841/7415>
- Sitio, S. L. M., & Nadiyanti, R. (2022). Analisis Sentimen Kenaikan Harga BBM Pertamina Pada Media Sosial Menggunakan Metode Naïve Bayes Classifier. *Building of Informatics, Technology and Science (BITS)*, 4(3), 1224–1231.

<https://doi.org/10.47065/bits.v4i3.2311>

- Sofyanty, D., Yanti, V. A., Ong, D., & Suwantica, S. (2022). Manajemen Risiko dalam Transaksi Perdagangan Valuta Asing. *Jurnal Ilmiah Universitas Batanghari Jambi*, 22(2), 1236. <https://doi.org/10.33087/jiubj.v22i2.2418>
- Susanti, A. R., & Ilahi, E. N. (2024). Sentiment Analysis of User Reviews of E-commerce Applications: Case Study on the Shoppe Platform. *Journal of Social Science*, 5(4), 983–988. <https://doi.org/10.46799/jss.v5i4.885>
- Vidhya, A. (2024). *Google Colab Features, Best Practices Guide*. <https://www.analyticsvidhya.com/blog/2020/03/google-colab-machine-learning-deep-learning>
- Wiratama Putra, T., Triayudi, A., & Andrianingsih, A. (2022). Analisis Sentimen Pembelajaran Daring Menggunakan Metode Naïve Bayes, KNN, dan Decision Tree. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 6(1), 20–26. <https://doi.org/10.35870/jtik.v6i1.368>
- Z.Aminullah, D.Arifianto, H. (2024). *Lexicon-Based Approach Pada Analisis Sentimen Ulasan Airbnb Menggunakan Vader Sentiment Lexicon-Based Approach to Sentiment Analysis of Airbnb Reviews Using Vader Sentiment*. 5(4), 558–566.
- Ziogas, A. N., Schneider, T., Nun, T. Ben, Calotoiu, A., Matteis, T. De, Licht, J. D. F., Lavarini, L., & Hoefler, T. (2021). Productivity, portability, performance: Data-centric python. In *International Conference for High Performance Computing, Networking, Storage and Analysis, SC* (Vol. 1, Issue 1). Association for Computing Machinery. <https://doi.org/10.1145/3458817.3476176>

LAMPIRAN

1. Surat Penetapan Dosen Pembimbing



UMSU
Unggul | Cerdas | Terpercaya
(Mata kuliah yang akan diajarkan harus dan lengkap)

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA

FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/IIU/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
<http://fkti.umsu.ac.id> fkti@umsu.ac.id [fksi.umsu.ac.id](#) [umsu.ac.id](#) [umsu.ac.id](#) [umsu.ac.id](#) [umsu.ac.id](#)

PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 660/IL.3-AU/UMSU-09/F/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi : Sistem Informasi
Pada tanggal : 11 Juni 2025

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : Nazri Adlani
NPM : 2109010146
Semester : VIII (Delapan)
Program studi : Sistem Informasi
Judul Proposal / Skripsi : Analisis Sentimen Ulasan Pengguna terhadap Aplikasi MIFX di Play Store Menggunakan Metode VADER dan Naive Bayes

Dosen Pembimbing : Dr. Al-Khowarismi, M.Kom

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi dinyatakan " BATAL " bila tidak selesai sebelum Masa Kadaluarsa tanggal : 11 Juni 2026**
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh.

Ditetapkan di : Medan
Pada Tanggal : 15 Dzulhijjah 1446 H
11 Juni 2025 M

Dekan


Dr. Al-Khowarismi, M.Kom.
NIDN : 0117099201



Cc. File



2. Lembar Bimbingan Skripsi



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
 UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 69/SK/BAN-PT/Akred/PT/III/2019
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
 Website: www.umsu.ac.id Email: info@umsu.ac.id [umsu.ac.id](https://www.facebook.com/umsu.ac.id) [umsu.ac.id](https://www.instagram.com/umsu.ac.id) [umsu.ac.id](https://www.youtube.com/umsu.ac.id)

Berita Acara Pembimbingan Skripsi

Nama Mahasiswa : Nazri Adlani Program Studi : Sistem Informasi
 NPM : 2109010146 Konsentrasi :
 Nama Dosen Pembimbing : Dr. Al-Khawarizmi, M.Kom Judul Penelitian :

Item	Hasil Evaluasi	Tanggal	Paraf Dosen
	- Revisi pasca Simpro - perbaikan bab III	6/08/25	u
	- bab III - perbaikan model perancangan sistem	9/08/25	u
	- BAB IV - Revisi perhitungan manual	14/08/25	u
	- BAB IV - Tambahkan design visual	16/08/25	u
	- BAB IV - Revisi design web App	19/08/25	u
	- BAB V - Kesimpulan dan Saran	21/08/25	u
	- Bimbingan Jurnal	23/08/25	u
	- Acc Sidang	25/08/25	u

Diketahui oleh :
 Ketua Program Studi
 Sistem Informasi
 (.....)

Medan,

Disetujui oleh :

Dosen Pembimbing

(.....)



3. Surat Undangan Semprom

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
Nomor : 770/IL3-AU/UMSU-09/F/2025

UNDANGAN SEMINAR PROPOSAL
Fakultas : Ilmu Komputer dan Teknologi Informasi
Program Studi : Sistem Informasi
Hari/Tanggal : Rabu, 30 Juli 2025
Waktu /Tempat : 09.00/G
Pemimpin Seminar : Martiano, S.Pd., S.Kom., M.Kom.

No.	NPM	NAMA MAHASISWA	Dosen Pembimbing	Dosen Pembahas	JUDUL PROPOSAL
11	2109010146	Nazri Adlani	Dr. Al-Khowarizmi, M.Kom.	Dr. Marah Doly Nasution, S.Pd., M.Si.	Analisis Sentimen Ulasan Pengguna Terhadap Aplikasi Mifx Di Play Store Menggunakan Metode Vader Dan Naive Bayes
12	2109010139	Athary Zikry	Farid Akbar Siregar S.Kom., M.Kom.	Yoshida Sary, S.Kom, S.E, M.Kom.	Analisa Kinerja Algoritma Naive Bayes Dan SVM (Support Vector Machine) Dalam Sentimen Analisis Permainan Mobile E-Football
13	2109010142	Ade Satrio	Yoshida Sary, S.Kom, S.E, M.Kom.	Fatma Sari Hutagalung, S.Kom., M.Kom.	Prediksi Penjualan Kambing Menggunakan Metode Single Exponential Smoothing (SES)

NB: - Laki-laki berbusana hitam putih dan memakai dasi
Perempuan berbusana muslimah hitam putih

Medan, 03 Shafar 1447 H
28 Juli 2025 M

Dekan
Dr. Al-Khowarizmi, M.Kom.
NIDN: 4127099201





4. Surat Undangan Sidang Skripsi

UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
Nomor : 907/IL3-AU/UMSU-09/F/2025
HAL UJIAN MEJA HIJAU SARJANA (SI)

UNDANGAN PANGGILAN

Kepada Yang Terhormat
Bapak/Ibu Dosen Penguji Meja Hijau
di
Medan
Catatan :
*Harap datang tepat waktu karena ujian dalam bentuk tim (2 Orang) penguji I & II
*Dosen Penguji yang terlambat 30 menit akan diganti

Fakultas : Ilmu Komputer dan Teknologi Informasi
Program Studi : Sistem Informasi
Hari/Tanggal : Sabtu, 13 September 2025
Waktu /Tempat : 08.00-14.00 WIB/Gedung G

No.	NAMA MAHASISWA	JUDUL SKRIPSI	Dosen Pembimbing	Dosen Pembahas	Dosen Pengganti
1	Muhammad Rafli Pratama	2109010120	Sistem Informasi Manajemen: Pembinaan UMKM Desa Pematang Johar Menggunakan Metode Value Chain Analysis (VCA)	Dr. Al-Khowarizmi, M.Kom.	Fatma Sari Hutagalung, S.Kom., M.Kom.
2	Zaki Zain Rangku	2009010031	Klasifikasi Barang Promosi Dan Harga Diskon Di Alfamidi Turi Menggunakan Metode Random Forest Berbasis WEB	Fatma Sari Hutagalung, S.Kom., M.Kom.	Mahardika Abdi Prawira, S.Kom., M.Kom.
3	Nazri Adlani	2109010146	Analisis Sentimen Ulasan Pengguna Terhadap Aplikasi MIFX Di Play Store Menggunakan Metode Vader Dan Naive Bayes	Dr. Marah Doly Nasution, S.Pd., M.Si.	Dr. Zainal Aziz, M.Si.

Medan, 19 Rabi'ul Awwal 1447 H
11 September 2025 M

Asisten Pengambilan Berita Acara :
1. Suvia Agustin S.I.Kom
2. Andika Suras Saputra, S.M

Ditetapkan Oleh
a.n Rektor
Prof. Dr. MUHAMMAD ARIFIN, S.H., M.Hum.

Ketua Panitia Ujian
Dr. Al-Khowarizmi, M.Kom.

Sekretaris Wakil Dekan I
Halim Maulana, S.T., M.Kom.

5. LOA Jurnal



LETTER OF ACCEPTANCE (LoA)

Kepada Yth Bpk/Ibu/Sdr

Nazri Adlani, Al-Khowarizmi

Di

Tempat

Dengan ini kami sampaikan bahwa naskah dengan rincian berikut dinyatakan diterima untuk diterbitkan di dalam Hello World Jurnal Ilmu Komputer pada terbitan Volume 5 Nomor 2 Edisi Oktober 2025.

Judul	Analisis Sentimen Ulasan Pengguna Terhadap Aplikasi MIFX Di Play Store Menggunakan Pendekatan Vader dan Naïve Bayes
Penulis	Nazri Adlani, Al-Khowarizmi
Correspondent Email	nazriadlani28@gmail.com

Demikianlah surat keterangan ini kami buat untuk dapat digunakan seperlunya.



Medan, 21 Oktober 2025

Editor in Chief

Anjar Wanto, M.Kom.

Hello World Jurnal Ilmu Komputer
Ilmu Bersama Center
Email: helloworldjurnal@gmail.com



Hello World Jurnal Ilmu Komputer is licensed under a
[Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

6. Hasil Turnitin

sebelum turnitin.docx

ORIGINALITY REPORT

7 %

SIMILARITY INDEX

6 %

INTERNET SOURCES

3 %

PUBLICATIONS

0 %

STUDENT PAPERS

PRIMARY SOURCES

1

repository.umsu.ac.id

Internet Source

1 %

2

oulurepo.oulu.fi

Internet Source

<1 %

3

repository.upnjatim.ac.id

Internet Source

<1 %

4

etheses.uin-malang.ac.id

Internet Source

<1 %

5

ejurnal.seminar-id.com

Internet Source

<1 %

6

Thival Rayyani, Aisah Rini Susanti. "Laravel untuk Pengembangan Web Puskesmas Diskominfo Kota Bogor", Karimah Tauhid, 2024

Publication

<1 %

7

repository.uinjambi.ac.id

Internet Source

<1 %

8

anyflip.com

Internet Source

<1 %

9

eprints.upnyk.ac.id

Internet Source

<1 %

10

lirias.kuleuven.be

Internet Source

<1 %

11

eprints.itn.ac.id