



Pengembangan Model Klasifikasi Sentimen Dengan Pendekatan Vader dan Algoritma Naive Bayes Terhadap Ulasan Aplikasi Indodax

Agus Dirgahayu Zendrato*, Sunneng Sandino Berutu, Yo'el Pieter Sumihar, Haeni Budiati

Fakultas Ilmu Sains Dan Komputer, Program Studi Informatika, Universitas Kristen Immanuel, Yogyakarta
Jl. Ukrim No.KM. 11, Kadirojo I, Purwomartani, Kec. Kalasan, Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia
Email: ^{1,*}dirgazendrato1708@gmail.com, ²sandinoberutu@ukrimuniversity.ac.id, ³pieter.haro@ukrimuniversity.ac.id,
⁴heni@ukrimuniversity.ac.id

Email Penulis Korespondensi: dirgazendrato1708@gmail.com

Submitted: 26/03/2024; Accepted: 30/04/2024; Published: 30/04/2024

Abstrak—Aplikasi perdagangan kriptokurensi seperti Indodax telah berkembang pesat, pemahaman tentang sentimen pengguna terhadap platform masih kurang, sehingga menarik untuk dianalisis setimen pengguna terhadap platform tersebut. Untuk mengukur sentimen, penelitian ini mengusulkan pendekatan gabungan metode Vader dan Naive Bayes. Data yang digunakan adalah kumpulan komentar pengguna di platform google play store terkait dengan pengalaman pengguna menggunakan Indodax. Metode Vader digunakan untuk menganalisis sentimen secara langsung dari teks komentar, sementara Naive Bayes diadopsi untuk meningkatkan akurasi dalam klasifikasi sentimen. Proses analisis sentimen melibatkan berbagai langkah, mulai dari persiapan data, pra-pemrosesan data, pelabelan data pelatihan dan pengujian serta evaluasi kinerja model Naive Bayes. Pada tahap analisis sentimen dengan metode Vader Sentiment, kategori positif memperoleh persentasi tertinggi sebesar 63,5%, disusul kategori netral sebesar 18,9% dan negatif sebesar 17,6%. Sementara itu, berdasarkan evaluasi kinerja model Naive Bayes diperoleh nilai akurasi sebesar 78% sedangkan nilai precision tertinggi dicapai oleh kategori sentimen negatif sebesar 80% dan recall pada kategori sentimen positif sebesar 44%.

Kata Kunci: Naive Bayes; Klasifikasi; Vader Sentimen; Sentimen; Indodax

Abstract—Cryptocurrency trading applications such as Indodax have grown rapidly, the understanding of user sentiment towards the platform is still lacking, so it is interesting to analyze user sentiment towards the platform. To measure sentiment, this research proposes a combined approach of Vader and Naive Bayes methods. The data used is a collection of user comments on the google play store platform related to user experience using Indodax. The Vader method is used to analyze sentiment directly from the comment text, while Naive Bayes is adopted to improve accuracy in sentiment classification. The sentiment analysis process involves various steps, starting from data preparation, data pre-processing, labeling of training and testing data and performance evaluation of the Naive Bayes model. At the sentiment analysis stage with the Vader Sentiment method, the positive category obtained the highest percentage of 63.5%, followed by the neutral category at 18.9% and negative at 17.6%. Meanwhile, based on the performance evaluation of the Naive Bayes model, the accuracy value is 78% while the highest precision value is achieved by the negative sentiment category at 80% and recall in the positive sentiment category at 44%.

Keywords: Naive Bayes; Classification; Vader Sentiment; Sentiment; Indodax

1. PENDAHULUAN

Dalam beberapa dekade terakhir, kemajuan teknologi dan konektivitas internet telah mengubah secara signifikan cara masyarakat memperdagangkan kripto atau bursa cryptocurrency yang berbasis online khususnya Indonesia. Salah satu aplikasi Indodax yang merupakan salah satu platform perdagangan cryptocurrency terkemuka di Indonesia. Indodax telah menjadi destinasi utama bagi para pengguna yang tertarik untuk membeli, menjual, dan melakukan perdagangan berbagai jenis aset kripto seperti Bitcoin, Ethereum, dan Ripple. Dengan peningkatan minat dan adopsi cryptocurrency di Indonesia, platform Indodax telah menjadi pusat penting bagi para investor, pedagang, dan pengguna kripto lainnya.

Fitur-fitur yang disediakan oleh aplikasi Indodax mencakup kemampuan untuk membeli dan menjual cryptocurrency dengan mata uang fiat lokal, analisis pasar, grafik harga, dan dompet digital yang aman. Selain itu, platform ini juga terus berinovasi dengan memperkenalkan fitur-fitur baru seperti perdagangan berjangka (futures trading), staking, dan layanan keuangan terkait lainnya.

Namun, seiring dengan pertumbuhan dan perkembangan aplikasi Indodax, muncul juga berbagai tantangan dan isu yang perlu dipertimbangkan. Pertama, keamanan dan keandalan platform menjadi perhatian utama, terutama dengan meningkatnya risiko keamanan dalam industri cryptocurrency. Selain itu, pengalaman pengguna dan kepuasan pelanggan juga menjadi faktor kritis dalam menilai kualitas platform ini.

Urgensinya terletak pada pentingnya menjaga kepercayaan pengguna, meningkatkan kualitas layanan, dan memperkuat posisi kompetitif di pasar yang semakin sibuk. Dalam konteks ini, analisis sentimen terhadap aplikasi Indodax menjadi penting untuk memahami respons pengguna terhadap platform ini dan mengidentifikasi area-area yang memerlukan perbaikan atau inovasi lebih lanjut.

Pada penelitian sebelumnya telah menunjukkan bahwa Support Vector Machine (SVM), sebuah model machine learning, berhasil mengklasifikasikan sentimen dengan baik, seperti yang didokumentasikan dalam referensi [1], [2]. Beberapa penelitian juga telah berhasil mengaplikasikan metode ini pada data ulasan film [3] dan tweet di Twitter, dengan hasil yang memuaskan, tetapi belum ada penelitian yang secara komprehensif membandingkan kinerja Naive Bayes dalam konteks yang sama.

Kemudian, belum ada penelitian yang secara khusus menerapkan metode ini untuk memahami sentimen pengguna terhadap aplikasi perdagangan kriptokurensi seperti Indodax. Penelitian ini akan mengambil pendekatan yang berbeda dengan menggunakan model Naive Bayes. Fokusnya adalah melakukan analisis sentimen dengan menggunakan data ulasan dari aplikasi Indodax dan pelabelan sentimen menggunakan metode Vader Sentiment. Dan tujuan dari studi ini adalah untuk mengeksplorasi kemampuan algoritma Naive Bayes diberbagai konteks data yang berbeda, dan melihat apakah metode ini dapat memberikan hasil yang komparatif dengan pendekatan SVM.

Dengan demikian, penelitian ini tidak hanya akan menambahkan wawasan tentang berbagai model machine learning dalam analisis sentimen, tetapi akan menambah pandangan yang lebih sederhana mengenai Naive Bayes dalam mengklasifikasikan sentimen ulasan Indodax yang telah dilabeli dengan metode Vader Sentiment.

Dalam konteks ini, analisis sentimen terhadap aplikasi Indodax menjadi penting untuk memahami bagaimana pengguna merespons platform ini, serta untuk mengidentifikasi area-area yang memerlukan perbaikan atau inovasi lebih lanjut. Dengan demikian, latar belakang ini memberikan konteks yang diperlukan untuk memahami analisis sentimen yang akan dilakukan terhadap aplikasi Indodax.

Ketidakpercayaan yang terjadi pada konsumen terhadap penjual menjadi suatu tantangan yang tak dapat diabaikan oleh para pelaku bisnis, karena hal tersebut memaksa mereka untuk menyusun pengujian yang mendalam terhadap produk dan layanan yang mereka tawarkan. Pelaku bisnis dihadapkan pada dilema yang kompleks; di satu sisi, mereka membutuhkan penanda permasalahan yang timbul dari komentar pengguna aplikasi, namun di sisi lain, proses membaca dan mengklasifikasikan setiap ulasan tersebut menambah durasi yang tidak terkoordinasi dan tidak efektif. Untuk mengatasi kompleksitas masalah ini, diperlukan adanya sistem yang mampu mengotomatiskan proses pengklasifikasian opini pelanggan ke pengelompokan sentimen negatif, positif dan netral. Sistem semacam ini, yang dikenal sebagai analisis sentimen, menjadi landasan utama dalam membantu para pelaku bisnis untuk memahami persepsi dan tanggapan pelanggan secara lebih efektif.

Analisis sentimen opini masyarakat, yang merupakan pendekatan untuk memahami dan mengukur pandangan, perasaan, dan sikap masyarakat terhadap aplikasi perdagangan kriptocurrency, telah menjadi sangat relevan dalam lingkup ini. Pelabelan analisis sentimen seperti Vader (Valence Aware Dictionary and sEntiment Reasoner) telah terbukti efektif dalam mengekstrak sentimen positif, negatif, dan netral dari teks yang dihasilkan oleh masyarakat, termasuk ulasan, komentar, dan pendapat di media sosial dan platform daring lainnya [4]

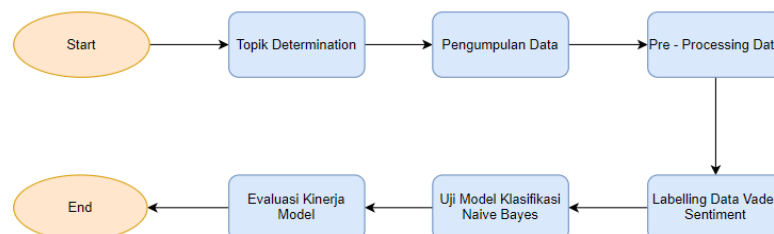
Tujuan dari penelitian untuk menyelidiki dan mengimplementasikan metode klasifikasi analisis sentimen opini masyarakat menggunakan Vader pada aplikasi perdagangan kriptocurrency. Analisis sentimen opini masyarakat akan memberikan wawasan berharga tentang bagaimana platform belanja online dipandang oleh konsumen. Dalam studi ini juga meneliti penyebab yang memengaruhi sentimen masyarakat terhadap aplikasi perdagangan kriptocurrency, seperti kualitas layanan, jenis order, analisis pasar, grafik harga, dompet digital, dan pengalaman pengguna.

Melalui pemahaman lebih dalam tentang opini dan sentimen masyarakat, penelitian ini dapat memberikan manfaat berupa panduan perbaikan bagi penyedia aplikasi perdagangan kriptocurrency, membantu mereka meningkatkan layanan mereka, mempertahankan kepercayaan konsumen, dan memperkuat posisi kompetitif mereka di pasar yang semakin sibuk. Selain itu, penelitian ini juga akan berkontribusi pada pengembangan metode analisis sentimen opini masyarakat yang lebih baik dan lebih efektif.

2. METODOLOGI PENELITIAN

2.1 Alur Penelitian

Didalam penelitian ini, memiliki alur serangkaian dalam menganalisis sentimen pada ulasan aplikasi indodax melalui Google Play Store. Langkah – langkahnya mencakup penentuan topik masalah (topic determination) untuk menjadi sebuah acuan dalam mengumpulkan data, kemudian pengumpulan data yang akan dianalisis. Setelah pengumpulan data, maka akan di pre-processing untuk siap diproses, setelah dipre-processing data akan dilabeli untuk menentukan sentimennya. Setelah tahap pelabelan, masuk dalam tahap klasifikasi dan evaluasi membentuk model Naive Bayes dan mengevaluasi kinerjanya untuk mengukur keakuratan analisisnya. Terakhir, memberikan umpan balik kepada pengguna untuk memberikan kepuasan dalam analisi penelitian ini [5]. Gambar 1 berikut merupakan diagram yang menggambarkan alur penelitian analisis data.



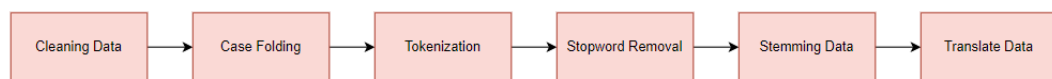
Gambar 1. Langkah Penelitian

2.2 Pengambilan Data

Sumber yang relevan dengan penelitian ini harus dikumpulkan dengan cermat. Ini bisa mencakup ulasan dan komentar pengguna tentang aplikasi Indodax di platform Google Play Store. Kualitas data yang baik sangat penting untuk memastikan hasil penelitian yang akurat. Proses crawling data dapat dilakukan secara manual dan otomatis menggunakan tools yang ada. Crawling data pada google play store menggunakan URL (link) aplikasi yang ingin di cari datanya kemudian di copy. Dalam hal ini, data yang dikumpulkan berbentuk ulasan review dari aplikasi indodax, data berhasil dikumpulkan sebanyak 3164 data ulasan dari rentang waktu 1 Januari 2022 sampai dengan 28 Februari 2024.

2.3 Pre - Processing

Sebelum pengolahan data adalah langkah awal dalam proses analisis data, yang bertujuan untuk mengubah data awal yang tidak terstruktur menjadi format yang lebih terstruktur, membersihkannya, dan menjadikannya siap untuk analisis lebih lanjut. Tujuan utamanya adalah untuk menyiapkan data agar dapat digunakan secara efektif dalam algoritma pembelajaran mesin atau analisis statistik



Gambar 2. Pre-Processing

Gambar 2 menunjukkan diagram alur yang menjelaskan langkah dalam setiap pra pemrosesan data sebelum masuk kedalam proses pengolahan data.

2.4 Pelabelan Data (Vader)

Pelabelan menggunakan VADER (Valence Aware Dictionary and sEntiment Reasoner) merupakan proses pemberian label sentimen pada teks menggunakan algoritma yang telah diprogram sebelumnya dan menggunakan kamus leksikon yang berisi 1773 kata sentimen yang telah diberi bobot, dan dari setiap data akan dihitung nilai polaritasnya berdasarkan bobot atau nilai dari setiap kata dalam suatu kalimat. Algoritma ini terutama digunakan untuk menganalisis sentimen dalam teks, mengidentifikasi review yang tergabung dalam kategori negatif, netral dan positif berdasarkan ulasan yang digunakan dan intensitasnya.

Untuk menguji kemampuan VADER dalam menganalisis sentimen pada suatu topik, pertama-tama dilakukan penggunaan VADER pada data ulasan online, kemudian dilanjutkan dengan penggunaan pada data vlog berbahasa Inggris asli. Selanjutnya, VADER juga diterapkan pada data vlog yang sudah diubah kebahasa asing untuk mengevaluasi kemampuannya dalam menganalisis sentimen pada data yang telah diterjemahkan.

VADER bekerja dengan cara mengukur polaritas dan intensitas emosi yang terkandung dalam teks. Algoritma ini berdasarkan pada penggunaan kamus yang telah ditetapkan sebelumnya, di mana setiap kata dikelompokkan ke dalam kategori sentimen tertentu (netral, positif, negatif). serta bobot yang menunjukkan tingkat intensitas sentimen. Kemudian, dengan menganalisis kata-kata yang muncul dalam teks dan mempertimbangkan bobot serta urutan kata-kata tersebut, VADER memberikan skor sentimen keseluruhan untuk teks tersebut .

2.5 Implementasi Model Dan Pembagian Data

Naïve Bayes merupakan model data yang akan diterapkan dalam penelitian. Algoritma ini adalah salah satu metode analisis statistik yang menggunakan probabilitas Bayesian untuk memproses data numerik. Model ini menggunakan teknik statistik yang berfungsi untuk memperkirakan anggota yang masuk dalam kelas berdasarkan probabilitasnya. Pada model klasifikasi Naïve Bayes yang sederhana, merupakan asumsi dalam model klasifikasi Naïve Bayes bahwa nilai atribut suatu kelas tidak saling memengaruhi atau dipengaruhi oleh nilai atribut lainnya. Pernyataan itu disebut sebagai class conditional independence, yang digunakan untuk mempermudah proses perhitungan. Secara sederhana, naive Bayesian classifier mengasumsikan bahwa karakter setiap kata dalam kalimat tidak dipengaruhi oleh kata lainnya. Dengan kata lain, analisis sentimen memiliki suatu proses setiap kata yang muncul diberikan bobot sendiri, dan total bobot dari semua kata tersebut kemudian dihitung dalam rumus kalimat tersebut masuk dalam pengelompokan negatif, positif, netral. [6]

Setelah prose pemodelan, dataset pada algoritma akan dibelah menjadi dua kategori, menjadi data uji dan data data. Data latih akan dimanfaatkan untuk melatih algoritma guna mencari model yang paling sesuai, sementara data uji akan berfungsi dalam mengevaluasi dan menguji performa model yang telah diperoleh pada tahap pengujian. Proses pengujian dilakukan dengan membagi dataset secara proporsional, di mana sekitar 80% (0.8) dari data digunakan sebagai data latih, sementara sisanya, sekitar 20% (0.2), digunakan sebagai data uji.

2.6 Pengujian Confusion Matrix

Confusion matrix merupakan adalah library yang berfungsi dalam data mining untuk mengukur akurasi suatu model klasifikasi. Confusion matrix juga dikenal sebagai error matrix. Ini menyediakan informasi tentang hasil

klasifikasi yang sebenarnya dan yang diprediksi pengelompokan dalam sistem. Evaluasi kinerja pada sistem akan dilakukan dengan menggunakan data dalam bentuk matriks.

Keterangan :

- a) TP: yaitu model dimana jumlah data positif yang diprediksi benar.
- b) TN: merupakan jumlah data negatif yang diprediksi benar oleh model.
- c) FP: adalah data positif diprediksi salah, tetapi jumlah data negatif yang seharusnya diprediksi benar, oleh model.
- d) FN: jumlah data negatif yang seharusnya diprediksi salah, namun benar diprediksi sebagai data positif oleh model.

Tabel 1. Pengujian

Kelas	Terklasifikasi Positif	Terklasifikasi Negatif
Positive	True Positive(TP)	False Negative(FN)
Negatives	False Positive (FP)	True Negative(TN)

Didalam tabel 1 menunjukkan performance metrics yang sering digunakan adalah: accuracy, precision, dan recall. Accuracy Berdasarkan nilai TN, FP, False Negatif (FN), dan True Positive (TP) dapat diperoleh nilai accuracy, precision, recall dan f-measure.

- a. Akurasi (Accuracy) : Merupakan proporsi dari total data yang diprediksi dengan benar oleh model, baik positif maupun negatif..

$$\text{Rumus} = \frac{TP + TN}{TP+TN+FP+FN} \times 100\% \tag{1}$$

- b. Presisi (Precision) : Merupakan proporsi dari data positif yang diprediksi benar oleh model dari total data yang diprediksi positif.

$$\text{Rumus} = \frac{TP}{TP+FP} \times 100\% \tag{2}$$

- c. Recall : ini mengukur seberapa banyak dari semua kasus positif yang diidentifikasi oleh model. Perbandingan recall merupakan dengan keseluruhan data yang benar positif maka rasio prediksi benar positif.

$$\text{Rumus} = \frac{TP}{TP+ FN} \times 100\% \tag{3}$$

- d. F1-score : Ini adalah perbandingan harmonik dari recall dan presisi, memberikan keseimbangan antara keduanya.

$$\text{Rumus} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% \tag{4}$$

3. HASIL DAN PEMBAHASAN

3.1 Pengambilan Data

Pengumpulan data adalah langkah awal yang perlu dilakukan dalam analisis sentimen. Dalam proses pengumpulan data, data dikumpulkan melalui Google Play Store kemudian langsung mengambil link id aplikasi yang akan dianalisis. Data yang diambil berupa ulasan, user, dan at, kemudian rentang waktu dalam pengambilan data adalah 1 Januari 2022 sampai 28 Februari 2024 dan data berhasil dikumpulkan sebanyak 3164 data ulasan.

index	at	userName	content
0	2022-01-05 11:22:22	Yokie Tokie	Kurang lengkap untuk coin coin lainnya seperti usdt,bnb atau lain lah. Contohnya seperti binance dan tokocrypt dia lengkap dan sama tambahkan fitur swab Tolong tambahkan lagi untuk updatenya
1	2022-01-16 03:14:21	Ridhwan Fathroni	Hallo indodax... Kenapa pas gua lagi butuh, apk error... Password sudah benar. Kok minta reset. Tolong dong...
2	2022-01-22 00:11:33	Ikhwan Indarto	2x saya mengalami kerugian karena walaupun saya ambil limit order tiba tiba aset saya langsung terjual semua seperti seperti saat menggunakan market order. Banyak di temukan bug. Setiap ingin transaksi selalu di arah kan ke btc/ldr di refresh berkali kali juga tidak bisa. Aplikasi sering ngehang dan ngeframe.
3	2022-01-23 06:40:32	Yudha Banawi	Aplikasi ringan, mudah digunakan, pilihan koin IDR lebih banyak. Hanya kurang cakep dikit pada tampilan trade simplenya.
4	2022-01-29 15:50:08	Septian Trividian	Hufft... harga coin sudah di turunkan tetapi tiba2 pas di tekan beli, harga coin berubah menjadi harga actual nya, udah gitu kena fee lagi udah gitu harga coin nya pas langsung turun juga wkwkw sungguh malang, tolong di perbaiki ya team it nya sesuaikan kualitas dengan fee yang di bebankan
5	2022-02-01 06:56:28	silas paelongan	Secara umum aplikasinya sudah baik... Tapi di fitur chart sangat kurang bahkan ketinggalan karna opsi simpan chart tidak tersedia. Padahal sumbernya dari trading view tapi malah dihilangkan. Padahal chart itu fitur utama.
6	2022-02-08 01:21:52	Xxx Divey	Biaya penarikan bukan di turunkin malah makin mahal. Kalo seperti ini terus pengguna bukan tambah banyak malah banyak yang kabur. Itu pasti bos Kalo gini terus mending pindah lagi ke wallet tetangga. Ngapain pake yang mahal kalo ada yang lebih murah. Kecewa Kalo udah turun baru gua kasih 5 bintang
7	2022-02-16 04:08:46	Divi Rantikunty	Transaksi/deposit dari other wallet tgl 2,4,6,7,10,14 dan 16 feb 2022 yg masuk hanya tgl 2,4 dan 16 februari saja. Yg lain kabur entah kemana, pdhl di wallet asal statusnya "payed" semua. Untungnya cuma recehan LTC doang, gimana kalau sampai withdraw 1 BTC ke indodax dan lenyap? Tolong diperbaiki lg systemnya, biar transaksi makin lancar dan tanpa rasa waswas
8	2022-02-16 05:15:35	Awang Gembuk	menurut saya kalau trading di indodax sangat mudah dan seru apa lagi skrg makin banyak listing coin yang baru"dan indodax cocok banget buat yang masih permula trading dan membuat kita semakin seru trading di indodax adalah sangat mudah dipahami. #8tahunindodax
9	2022-02-16 05:18:45	monz gamz	Pengalaman pake Indodax selama 1th terakhir ga ada masalah si. Blm ada kendala apapun. Cuma hanya sebatas saran untuk fee WD kalo bisa di perkecil, karna untuk saat ini sudah banyak saingan. (Akun : Suparmono) #8TahunIndodax

Gambar 3. Tampilan Hasil Crawling Data

Pada gambar 3 diatas merupakan sebuah contoh tampilan hasil dari proses crawling data, yang terdiri dari waktu ulasan dikirim, serta nama pengguna yang memberikan ulasan.



3.2 Pre – Processing

Setelah data dikumpulkan, kemudian dipre-processing agar siap untuk analisis. Berikut adalah teknik pra-pemrosesan yang akan digunakan.

3.2.1 Cleaning Data

Untuk proses data yang telah dicrawling akan dimasukkan ke dalam cleaning data untuk membersihkan data yang terdapat url, emoticon, retweet, simbol, hash-tag serta spasi yang tidak dibutuhkan. Tabel 2 berikut adalah proses sebelum dan sesudah data masuk kedalam tahap pembersihan.

Tabel 2. Cleaning Data

Sebelum	Sesudah
Kurang lengkap untuk coin coin lainnya seperti usdt, bnb atau lain lah. Contohnya seperti binance dan tokocrypt dia lengkap dan sama tambahkan filter swab Tolong tambahkan lagi untuk updatenya Halo indodax... Kenapa pas gua lagi butuh, apk error... Password sudah benar. Kok minta reset. Tolong dong... 2x saya mengalami kerugian karena walaupun saya ambil limit order tiba tiba aset saya langsung terjual semua seperti seperti saat menggunakan market order. Banyak di temukan bug. Setiap ingin transaksi selalu di arah kan ke btc/idr di refresh berkali kali juga tidak bisa. Aplikasi sering ngehang dan ngeframe.	Kurang lengkap untuk coin coin lainnya seperti usdt, bnb atau lain lah Contohnya seperti binance dan tokocrypt dia lengkap dan sama tambahkan filter swab Tolong tambahkan lagi untuk updatenya Halo indodax Kenapa pas gua lagi butuh apk error Password sudah benar Kok minta reset Tolong dong x saya mengalami kerugian karena walaupun saya ambil limit order tiba tiba aset saya langsung terjual semua seperti seperti saat menggunakan market order Banyak di temukan bug Setiap ingin transaksi selalu di arah kan ke btc/idr di refresh berkali kali juga tidak bisa Aplikasi sering ngehang dan ngeframe

3.2.2 Case Folding (CF)

Tahapan Case Folding bertujuan untuk semua karakter kata akan diubah menjadi huruf kecil atau huruf besar agar konsisten dalam analisis.

Tabel 3. Case Folding

Sebelum	Sesudah
Kurang lengkap untuk coin coin lainnya seperti usdt, bnb atau lain lah Contohnya seperti binance dan tokocrypt dia lengkap dan sama tambahkan filter swab Tolong tambahkan lagi untuk updatenya Halo indodax Kenapa pas gua lagi butuh apk error Password sudah benar Kok minta reset Tolong dong x saya mengalami kerugian karena walaupun saya ambil limit order tiba tiba aset saya langsung terjual semua seperti seperti saat menggunakan market order Banyak di temukan bug Setiap ingin transaksi selalu di arah kan ke btc/idr di refresh berkali kali juga tidak bisa Aplikasi sering ngehang dan ngeframe	kurang lengkap untuk coin coin lainnya seperti usdt, bnb atau lain lah contohnya seperti binance dan tokocrypt dia lengkap dan sama tambahkan filter swab tolong tambahkan lagi untuk updatenya hallo indodax kenapa pas gua lagi butuh apk error password sudah benar kok minta reset tolong dong x saya mengalami kerugian karena walaupun saya ambil limit order tiba tiba aset saya langsung terjual semua seperti seperti saat menggunakan market order banyak di temukan bug setiap ingin transaksi selalu di arah kan ke btc/idr di refresh berkali kali juga tidak bisa aplikasi sering ngehang dan ngeframe

Pada tabel 3 menunjukkan dalam tahap case folding proses sebelum memiliki huruf yang besar dan proses sesudah mengubah menjadi huruf kecil. dalam tahap case folding

3.2.3 Tokenization

Tahap tokenization adalah tahapan pemecahan seluruh isi teks ulasan menjadi potongan-potongan kata (termed word) yang berdiri secara tunggal. Tokenization bisa berupa , tweet/content kata-kata, frasa, kalimat, atau bahkan karakter dengan tujuan untuk proses analisis teks lebih lanjut. Pada tabel 4 berikut adalah sebelum dan sesudah diolah dalam tahap tokenization kata yang telah dipisah-pisah.

Tabel 4. Tokenization

Sebelum	Sesudah
kurang lengkap untuk coin coin lainnya seperti usdt, bnb atau lain lah contohnya seperti binance dan tokocrypt dia lengkap dan sama tambahkan filter swab tolong tambahkan lagi untuk updatenya	['kurang', 'lengkap', 'untuk', 'coin', 'coin', 'lainnya', 'seperti', 'usdt', 'bnb', 'atau', 'lain', 'lah', 'contohnya', 'seperti', 'binance', 'dan', 'tokocrypt', 'dia', 'lengkap', 'dan', 'sama', 'tambahan', 'filter', 'swab', 'tolong', 'tambahkan', 'lagi', 'untuk', 'updatenya']



Sebelum	Sesudah
hallo indodax kenapa pas gua lagi butuh apk error password sudah benar kok minta reset tolong dong	['hallo', 'indodax', 'kenapa', 'pas', 'gua', 'lagi', 'butuh', 'apk', 'error', 'password', 'sudah', 'benar', 'kok', 'minta', 'reset', 'tolong', 'dong']
x saya mengalami kerugian karena walaupun saya ambil limit order tiba tiba aset saya langsung terjual semua seperti seperti saat menggunakan market order banyak di temukan bug setiap ingin transaksi selalu di arah kan ke btcidr di refresh berkali kali juga tidak bisa aplikasi sering ngehang dan ngeframe	['x', 'saya', 'mengalami', 'kerugian', 'karena', 'walaupun', 'saya', 'ambil', 'limit', 'order', 'tiba', 'tiba', 'aset', 'saya', 'langsung', 'terjual', 'semua', 'seperti', 'seperti', 'saat', 'menggunakan', 'market', 'order', 'banyak', 'di', 'temukan', 'bug', 'setiap', 'ingin', 'transaksi', 'selalu', 'di', 'arah', 'kan', 'ke', 'btcidr', 'di', 'refresh', 'berkali', 'kali', 'juga', 'tidak', 'bisa', 'aplikasi', 'sering', 'ngehang', 'dan', 'ngeframe']

3.2.4 Stopword Removal

Proses ini adalah tahapan penghapusan kata-kata yang tidak memiliki makna atau kata-kata yang kurang memiliki arti yang akan mempengaruhi proses sentimen, seperti "dan", "atau", "di", dan sebagainya. Pada tabel 5 berikut adalah proses perubahan data sebelum dan sesudah stopwords removal dilakukan.

Tabel 5. Stopword Removal

Sebelum	Sesudah
['kurang', 'lengkap', 'untuk', 'coin', 'coin', 'lainnya', 'seperti', 'usdtbnb', 'atau', 'lain', 'lah', 'contohnya', 'seperti', 'binance', 'dan', 'tokocrypt', 'dia', 'lengkap', 'dan', 'sama', 'tambahan', 'filitur', 'swab', 'tolong', 'tambahin', 'lagi', 'untuk', 'updatenya']	['lengkap', 'coin', 'coin', 'usdtbnb', 'contohnya', 'binance', 'tokocrypt', 'lengkap', 'tambahan', 'filitur', 'swab', 'tolong', 'tambahin', 'updatenya']
['hallo', 'indodax', 'kenapa', 'pas', 'gua', 'lagi', 'butuh', 'apk', 'error', 'password', 'sudah', 'benar', 'kok', 'minta', 'reset', 'tolong', 'dong']	['hallo', 'indodax', 'pas', 'gua', 'butuh', 'apk', 'error', 'password', 'reset', 'tolong']
['x', 'saya', 'mengalami', 'kerugian', 'karena', 'walaupun', 'saya', 'ambil', 'limit', 'order', 'tiba', 'tiba', 'aset', 'saya', 'langsung', 'terjual', 'semua', 'seperti', 'seperti', 'saat', 'menggunakan', 'market', 'order', 'banyak', 'di', 'temukan', 'bug', 'setiap', 'ingin', 'transaksi', 'selalu', 'di', 'arah', 'kan', 'ke', 'btcidr', 'di', 'refresh', 'berkali', 'kali', 'juga', 'tidak', 'bisa', 'aplikasi', 'sering', 'ngehang', 'dan', 'ngeframe']	['mengalami', 'kerugian', 'ambil', 'limit', 'order', 'aset', 'langsung', 'terjual', 'market', 'order', 'temukan', 'bug', 'transaksi', 'arah', 'btcidr', 'refresh', 'berkali', 'kali', 'aplikasi', 'ngehang', 'ngeframe']

3.2.5 Stemming Data

Tahap stemming data adalah mengonversi semua ulasan ke bentuk kata dasarnya atau berimbuhan. Ini bertujuan mengurangi variasi kata yang memiliki akar yang sama. Misalnya, mengubah "berlari", "berlari-lari", dan "lari" menjadi "lari". Berikut tabel 6 sebelum dan sesudah stemming data dilakukan:

Tabel 6 Stemming Data

Sebelum	Sesudah
['lengkap', 'coin', 'coin', 'usdtbnb', 'contohnya', 'binance', 'tokocrypt', 'lengkap', 'tambahan', 'filitur', 'swab', 'tolong', 'tambahin', 'updatenya']	lengkap coin coin usdtbnb contoh binance tokocrypt lengkap tambah filitur swab tolong tambahin updatenya
['hallo', 'indodax', 'pas', 'gua', 'butuh', 'apk', 'error', 'password', 'reset', 'tolong']	hallo indodax pas gua butuh apk error password reset tolong
['mengalami', 'kerugian', 'ambil', 'limit', 'order', 'aset', 'langsung', 'terjual', 'market', 'order', 'temukan', 'bug', 'transaksi', 'arah', 'btcidr', 'refresh', 'berkali', 'kali', 'aplikasi', 'ngehang', 'ngeframe']	alami rugi ambil limit order aset langsung jual market order temu bug transaksi arah btcidr refresh kali kali aplikasi ngehang ngeframe

3.3 Translate Data

Sebelum masuk dalam tahap pelabelan, semua data ulasan diterjemahkan kedalam bahasa Inggris . Tujuan untuk translate data ini adalah untuk mendapatkan nilai compound dari suatu teks ulasan. Pada tabel 7 berikut menunjukkan proses translate data dilakukan :

Tabel 7. Translate Data

Sebelum	Sesudah
lengkap coin coin usdtbnb contoh binance tokocrypt lengkap tambah filitur swab tolong tambahin updatenya	complete coin coin usdt bnb example complete binance tokocrypto plus swab feature please add the update
hallo indodax pas gua butuh apk error password reset tolong	hello Indodax when I need apk error password reset, please

Sebelum	Sesudah
alami rugi ambil limit order aset langsung jual	experience losses take limit orders on assets immediately
market order temu bug transaksi arah btcidr	sell market orders find transaction bugs btcidr directions
refresh kali kali aplikasi ngehang ngeframe	refresh times the application hangs framing

3.4 Pelabelan Data

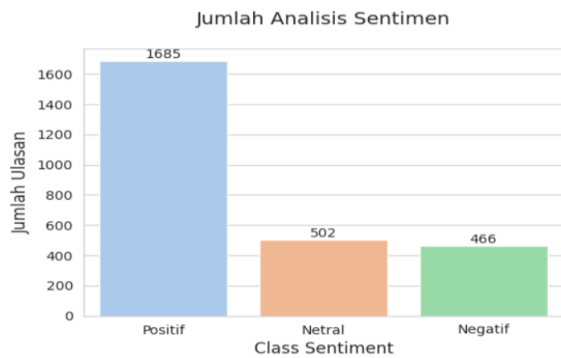
Setelah melalui proses terjemahan semua data ulasan, langkah selanjutnya setiap kalimat akan diberikan label dengan membaginya ke dalam tiga kategori sentimen, yaitu netral, positif, negatif. Penentuan label ini dilakukan berdasarkan pada nilai compound score dari setiap ulasan teks, dengan kondisi sebagai berikut:

- Dalam kasus nilai compound score yang lebih besar dari 0, maka ulasan akan diberi label "sentimen positif", menunjukkan adanya kecenderungan positif dalam ulasan tersebut.
- Ketika nilai compound score sama dengan 0, ulasan akan diberi label "sentimen netral", menunjukkan ketiadaan kecenderungan positif maupun negatif yang signifikan dalam ulasan tersebut.
- Jika nilai compound score kurang dari 0, maka ulasan akan diberi label "sentimen negatif", menunjukkan adanya kecenderungan negatif dalam ulasan tersebut.

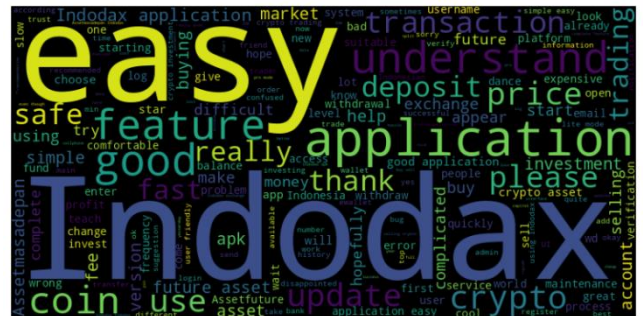
index	at	userName	steming_data	Compound_Score	Sentiments
0	2022-01-05 11:22:22	Good ones	complete coin coin usdt bnb example complete binance tokocrypto plus swab feature please add the update	0.3182	Positif
1	2022-01-16 03:14:21	Ridhwan Fathoni	hello Indodax when I need apk error password reset please	-0.1027	Negatif
2	2022-01-22 00:11:33	Indarto Brotherhood	experience losses take limit orders on assets immediately sell market orders find transaction bugs btcidr directions refresh times the application hangs framing	-0.25	Negatif

Gambar 4. Pelabelan Data

Pada gambar 4, menampilkan hasil pelabelan data seperti positif, netral serta negatif yang telah diproses menggunakan library vader yang disesuaikan dengan compoun score.



Gambar 5. Jumlah analisis sentimen

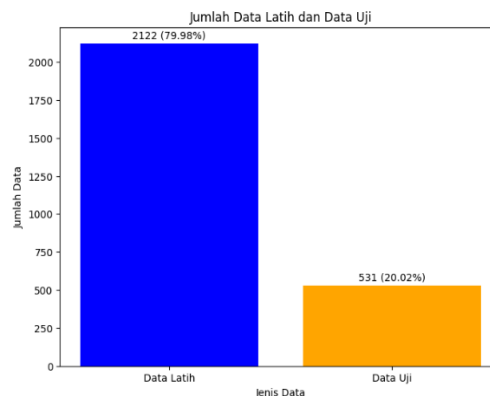


Gambar 6. Wordcloud Analisis

Berdasarkan pada gambar 5, sentimen positif memperoleh data paling terbanyak dengan jumlah 1685, disusul dengan sentimen netral sebanyak 466 dan negatif sebanyak 502. Dan pada gambar 6 menunjukkan wordcloud ulasan yang sering muncul dengan tampilan kata yang besar.

3.5 Klasifikasi

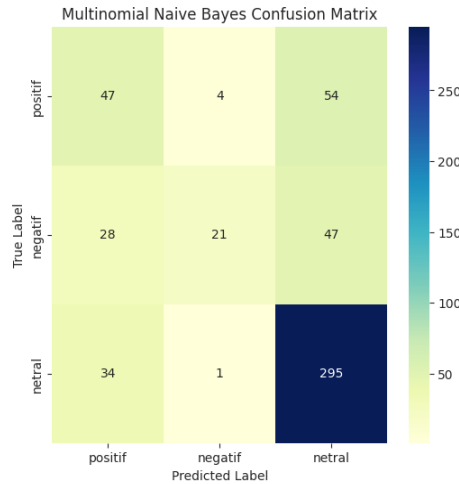
Sebelum membangun model, data akan dibagi dua yaitu data latih dan data uji dengan rasio 80:20, di mana data latih akan mencakup 80% dari jumlah total data, dan data uji akan mencakup 20% sisanya. Pada gambar 7 jumlah data latih akan menjadi sebanyak 2122, sementara jumlah data uji akan menjadi sebanyak 531.



Gambar 7. Pembagian Data

3.6 Pengujian

Kinerja model algoritma Naïve Bayes dievaluasi dengan Confusion matrix yang ditampilkan pada figure berikut.



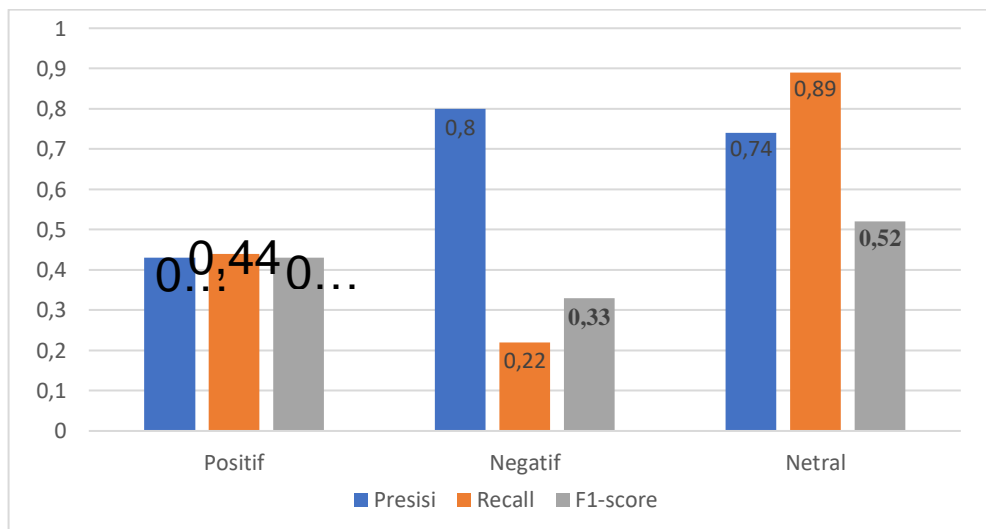
Gambar 8. Visualisasi Confusion Matrix

Pada gambar 8 menunjukkan figure Confusion Matrix yang dapat menghasilkan beberapa parameter seperti accuracy, precision, recall, dan f1-score diukur. Hasil perhitungan parameter tersebut dideskripsikan pada tabel berikut.

Tabel 8. Performace Matrix

Sentimen	Parameter		
	Presisi	Recall	F1-Score
Positif	0.43	0.44	0.43
Negatif	0.80	0.22	0.33
Netral	0.74	0.89	0.52
Average	0.65	0.51	0.52

Data dalam tabel 8. diatas diilustrasikan dalam grafik bar. Tampilan kinerja pada model Naïve Bayes dalam grafik dapat disajikan dalam gambar berikut.



Gambar 9. Grafik Bar Analisis

Mengacu pada tabel 8 dan gambar 9 diatas, kinerja model NB untuk setiap kategori sentimen dapat dijelaskan sebagai berikut.

- a. Positive: model memperoleh nilai rendah pada parameter Precison (0.43), Recall (0.44) dan F1-score (0.43). Hal ini menunjukkan bahwa model cenderung melewatkan banyak instance Positive. Selain itu, keseimbangan parameter Precision dan Recall juga rendah.



- b. Neutral: model mempergich nilai yang tinggi pada parameter Precision (0.74), recall (0.89) dan F1-score dengan masing-masing sebesar 0.81. Hasil ini menunjukkan bahwa model memiliki kinerja yang baik untuk mengklasifikasikan sentimen Neutral.
- c. Negative: model memperoleh nilai rendah pada parameter Precision (0.80), Recall (0.22) dan F1-score (0.33). Hal ini menunjukkan bahwa model cenderung melewati banyak instance Positive. Selain itu, keseimbangan parameter Precision dan Recall juga rendah.

4. KESIMPULAN

Penelitian ini dilakukan untuk memperoleh analisis sentimen pengguna terhadap aplikasi Indodax dengan mengadopsi teknik leksikon Vader dan membangun model klasifikasi berbasis metode Naïve Bayes. Hasil analisis dengan Vader menunjukkan bahwa aplikasi Indodax memiliki sentimen positif tertinggi dengan persentase sebesar 63,5%. Selanjutnya, hasil pengujian model dengan confusion matrix menunjukkan bahwa parameter presisi pada sentimen Negatif memperoleh nilai persentase tertinggi sebesar, sedangkan parameter recall tertinggi pada kategori Neutral. Sementara itu, nilai parameter f1-score agak rendah pada semua sentimen. Untuk penelitian selanjutnya, perlu diterapkan identifikasi kata tidak baku dan perbaikannya supaya kinerja analisis sentimen meningkat.

REFERENCES

- [1] A. Kusuma and H. Nurramdhani Irmanda, “Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine,” 2022.
- [2] Y. Afrillia, L. Rosnita, and D. Siska, “Analisis Sentimen Ciutan Twitter Terkait Penerapan Permendikbudristek Nomor 30 Tahun 2021 Menggunakan TextBlob dan Support Vector Machine,” *G-Tech: Jurnal Teknologi Terapan*, vol. 6, no. 2, pp. 387–394, Oct. 2022, doi: 10.33379/gtech.v6i2.1778.
- [3] R. Abdillah, E. Haerani, and R. M. Candra, “Analisis Sentimen Ulasan Aplikasi Wetv Untuk Peningkatan Layanan Menggunakan Metode Support Vector Machine,” *Journal of Information System Research (JOSH)*, vol. 4, no. 3, pp. 865–873, Apr. 2023, doi: 10.47065/josh.v4i3.3353.
- [4] M. Taufiq Anwar, D. Riandhita Arief Permana, P. STMI Jakarta, P. Sistem Informasi Industri Otomotif, J. Letjen Suprpto No, and J. Pusat, “Analisis Sentimen Masyarakat Indonesia Terhadap Produk Kendaraan Listrik Menggunakan VADER,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 10, no. 1, pp. 783–792, Mar. 2023, [Online]. Available: <http://jurnal.mdp.ac.id>
- [5] D. Aryanti, “Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naïve Bayes Classifier,” *Journal of Information System Research (JOSH)*, vol. 3, no. 4, pp. 524–531, Jul. 2022, doi: 10.47065/josh.v3i4.1944.
- [6] L. O. Sihombing, H. Hannie, and B. A. Dermawan, “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naïve Bayes Classifier,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 2, pp. 233–242, Dec. 2021, doi: 10.29408/edumatic.v5i2.4089.
- [7] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, “Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine,” *Jurnal Pengembangan IT (JPIT)*, vol. 6, no. 3, Jul. 2021, [Online]. Available: <http://situs.com>
- [8] D. Normawati and S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, Sep. 2021.
- [9] W. Sejati, A. Singh Bist, and A. Tambunan, “Karya ini berlisensi di bawah Creative Commons Attribution 4.0 (CC BY 4.0) Pengembangan Analisis Sentimen dalam Rekayasa Software Engineering menggunakan tinjauan literatur sistematis,” *Jurnal Manajemen Pendidikan dan Teknologi Informasi*, vol. 2, no. 1, pp. 95–103, Sep. 2023, [Online]. Available: <https://journal.pandawan.id/mentari/article/view/377>
- [10] A. Dwiki, A. Putra, and S. Juanita, “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Bibit Dan Bareksa Dengan Algoritma KNN,” *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 2, pp. 636–646, Jun. 2021, [Online]. Available: <http://jurnal.mdp.ac.id>
- [11] P. Arsi and R. Waluyo, “ANALISIS SENTIMEN WACANA PEMINDAHAN IBU KOTA INDONESIA MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [12] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, “Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine,” *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY)*, vol. 2, no. 1, Mar. 2022.
- [13] D. R. Alghifari, M. Edi, and L. Firmansyah, “Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia,” *Jurnal Manajemen Informatika (JAMIKA)*, vol. 12, no. 2, pp. 89–99, Sep. 2022, doi: 10.34010/jamika.v12i2.7764.
- [14] A. Safira, A. S. Masyarakat...v, and F. N. Hasan, “ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” *Jurnal Sistem Informasi*, vol. 5, no. 1, Jan. 2023.
- [15] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *Jurnal KomtekInfo*, vol. 10, no. 1, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [16] E. E. Amelia and I. Yustiana, “Analisis Sentimen Pada Ulasan Produk UNIQLO dengan Algoritma Naive Bayes,” *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 8, no. 1, pp. 141–148, Mar. 2024.
- [17] A. N. Alfarooby and H. Irawan, “Analisis Sentimen Kepuasan Konsumen Pengguna Transportasi Online Pada Ulasan Google Playstore Menggunakan Indobert Dan Topic Modeling (Studi kasus: Gojek dan Grab),” 2024.



- [18] D. Nugraha and D. Gustian, “Analisis Sentimen Penggunaan Aplikasi Transportasi Online Pada Ulasan Google Play Store dengan Metode Naive Bayes Classifier,” *Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 5, no. 1, pp. 326–335, Jan. 2024.
- [19] A. N. Hasanah and B. N. Sari, “ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI JASA OJEK ONLINE MAXIM PADA GOOGLE PLAY DENGAN METODE NAÏVE BAYES CLASSIFIER,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 1, Jan. 2024, doi: 10.23960/jitet.v12i1.3628.
- [20] A. Handayani and I. Zufria, “Analisis Sentimen Terhadap Bakal Capres RI 2024 di Twitter Menggunakan Algoritma SVM,” *Journal of Information System Research (JOSH)*, vol. 5, no. 1, pp. 53–63, Oct. 2023, doi: 10.47065/josh.v5i1.4379.
- [21] L. A. Pramesti and N. Pratiwi, “Analisis Sentimen Twitter Terhadap Program MBKM Menggunakan Decision Tree dan Support Vector Machine,” *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1145–1154, Jul. 2023, doi: 10.47065/josh.v4i4.3807.
- [22] N. Q. Rizkina and F. N. Hasan, “Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes,” *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1136–1144, Jul. 2023, doi: 10.47065/josh.v4i4.3803.
- [23] I. M. Karo Karo, J. A. Karo Karo, Y. Yunianto, H. Hariyanto, M. Falah, and M. Ginting, “Analisis Sentimen Ulasan Aplikasi Info BMKG di Google Play Menggunakan TF-IDF dan Support Vector Machine,” *Journal of Information System Research (JOSH)*, vol. 4, no. 4, pp. 1423–1430, Jul. 2023, doi: 10.47065/josh.v4i4.3943.
- [24] E. Putri Nirwandani and R. Cahya Wihandika, “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode Modified Term Frequency Scheme Dan Naïve Bayes,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 3, pp. 1039–1047, Mar. 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [25] A. Tazidan OctaN et al., “ALGORITMA DECISION TREE UNTUK ANALISIS SENTIMEN PUBLIC TERHADAP MARKETPLACE DI INDONESIA,” *Jurnal Ilmiah Nasional Riset Aplikasi dan Teknik Informatika*, vol. 05, no. 5, Jun. 2023.