



Implementasi Library Textblob dan Metode Support Vector Machine Pada Analisis Sentimen Pelanggan Terhadap Jasa Transportasi Online

Yardiana Laia*, Sunneng Sandino Berutu, Yo'el Pieter Sumihar, Haeni Budiati

Fakultas Ilmu Sains dan Komputer, Program Studi Informatika, Universitas Kristen Immanuel, Yogyakarta, Indonesia

Email: ^{1,*}Yardiana.L2042@student.ukrimuniversity.ac.id, ²Sandinoberutu@ukrimuniversity.ac.id,

³Pieter.haro@ukrimuniversity.ac.id, ⁴Heni@ukrimuniversity.ac.id

Email Penulis Korespondensi: Yardiana.L2042@student.ukrimuniversity.ac.id

Submitted: 18/04/2024; Accepted: 15/05/2024; Published: 23/06/2024

Abstrak—Jasa transportasi online telah menjadi bagian yang tidak dapat dipisahkan dari kehidupan manusia saat ini. Penelitian ini bertujuan untuk mengembangkan metode analisis sentimen yang efektif guna mengukur opini masyarakat tentang kualitas layanan jasa transportasi online yang memiliki dampak signifikan terhadap reputasi perusahaan dan penerimaan masyarakat terhadap layanan tersebut. Dalam penelitian ini, mengusulkan penggunaan library TextBlob untuk melakukan analisis sentimen terhadap opini masyarakat terhadap jasa transportasi online. Library ini memungkinkan untuk mengukur polaritas positif, negatif dan netral serta subjektivitas dari teks opini yang dikumpulkan dari ulasan aplikasi *Gojek*, *Maxim* dan *Grab* melalui *Google Play Store*. Langkah-langkah analisis sentimen dilakukan mulai dari persiapan data, pra pemrosesan data, pelabelan data menggunakan library *Text Blob*. Selanjutnya, membangun model klasifikasi sentimen berbasis algoritma Support Vector Machine (SVM) melalui tahap pelatihan dan pengujian. Hasil pengujian model dievaluasi dengan *confusion matrix*. Hasil analisis dengan textblob menunjukkan bahwa transportasi online memperoleh sentimen positif tertinggi sebesar 40,1%, disusul sentimen netral sebesar 26,7% dan negatif sebesar 25,2%. Sementara itu, hasil pengukuran kinerja model menunjukkan bahwa precision memperoleh nilai tertinggi pada sentimen positif sebesar 0,93. Parameter recall mencapai nilai tertinggi pada sentimen negatif sebesar 0,95 dan f1-score pada sentimen netral dan positif sebesar 0,92. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi dalam pengembangan klasifikasi analisis sentimen, tetapi juga memberikan dampak praktis yang signifikan dalam meningkatkan layanan jasa transportasi online dan memberikan informasi yang berguna bagi masyarakat, sehingga mendorong inovasi dan perbaikan berkelanjutan dalam layanan transportasi online.

Kata Kunci: Klasifikasi; Sentimen; Text Blob; Transportasi Online; Support Vector Machine

Abstract—Online transportation services have become an inseparable part of human life today. This research aims to develop an effective sentiment analysis method to measure public opinion about the quality of online transportation services, which has a significant impact on company reputation and public acceptance of these services. In this research, we propose the use of TextBlob library to perform sentiment analysis of public opinion on online transportation services. This library allows to measure the positive, negative and neutral polarity and subjectivity of opinion text collected from *Gojek*, *Maxim* and *Grab* application reviews through *Google Play Store*. Sentiment analysis steps are carried out starting from data preparation, data pre-processing, data labeling using the *Text Blob* library. Furthermore, building a sentiment classification model based on the Support Vector Machine (SVM) algorithm through training and testing stages. Model testing results are evaluated with *confusion matrix*. The results of the analysis with textblob showed that online transportation received the highest positive sentiment of 40.1%, followed by neutral sentiment of 26.7% and negative sentiment of 25.2%. Meanwhile, the model performance measurement results show that the precision obtained the highest value in positive sentiment of 0.93. The recall parameter reaches the highest value in negative sentiment of 0.95 and f1-score in neutral and positive sentiment of 0.92. Thus, this research not only contributes to the development of sentiment analysis classification, but also has a significant practical impact in improving online transportation services and providing useful information to the public, thus encouraging innovation and continuous improvement in online transportation services.

Keywords: Classification; Sentiment; Text Blob; Online; Transportation; Support Vector Machine

1. PENDAHULUAN

Dalam era digital yang semakin berkembang saat ini, opini dan ulasan masyarakat tentang berbagai produk dan layanan, termasuk jasa transportasi online, dapat sangat memengaruhi persepsi publik, keputusan konsumen, dan citra merek perusahaan. Oleh karena itu, ini topik yang sangat penting dalam memahami pandangan masyarakat terhadap layanan seperti ini. Salah satunya adalah bidang jasa transportasi. Transportasi merupakan salah satu penunjang kegiatan manusia yang tidak dapat dipisahkan, transportasi tidak hanya berfungsi menghubungkan berbagai lokasi, tetapi berperan dalam memfasilitasi interaksi sosial dan ekonomi. Dalam perkembangannya, sektor transportasi kini memasuki era baru dengan kehadiran layanan transportasi online yang menawarkan solusi transportasi yang lebih fleksibel dan terjangkau dibandingkan dengan model konvensional yang lebih kompleks dan tentu saja dapat diakses melalui smartphone sehingga lebih flexibel bagi para penggunanya. Manfaat lain yang didapat adalah pengguna tidak perlu lagi harus mencari pangkalan atau tempat singgah penyedia jasa transportasi (driver) begitu juga untuk pengemudi (driver) tidak perlu stand-by di tempat pangkalan atau mencari pelanggan secara manual. Semua cukup dihubungkan dengan aplikasi pemesanan transportasi online yang ada, seperti *Grab*, *Gojek*, *Maxim*, dll.[1]

Media transportasi online telah melampaui fungsi awalnya sebagai alat untuk memesan transportasi, dan saat ini juga berperan sebagai platform untuk pengiriman barang dan layanan pengantaran makanan. Kepopuleran aplikasi transportasi online ini dipicu oleh terobosan-terobosan inovatif dalam industri transportasi yang menggabungkan teknologi komunikasi online, memberikan kemudahan bagi masyarakat untuk memesan layanan transportasi di mana pun dan kapan pun. Menurut survei yang dilakukan oleh *We Are Social (CNBC Indonesia)* pada Januari 2020, terdapat



sekitar 21,7 juta orang di Indonesia yang menggunakan layanan ride-hailing atau berbagi tumpangan seperti Gojek, Maxim, dan lainnya. Selain itu, survei tersebut juga mengungkapkan bahwa sebanyak 75% pengguna internet di Indonesia menggunakan aplikasi mobile yang berkaitan dengan pemetaan.. Menurut survey Indef,(DataIndonesia.id) Gojek, Grab dan Maxim menjadi 3 layanan ojek online yang paling banyak digunakan dengan presentase 82,6%, 57% dan 19,6% dari responden.[2]

Hal tersebut menyebabkan banyak pengguna yang memberikan pendapat mereka melalui Google Play Store mengenai layanan transportasi online di Indonesia. Platform tersebut merupakan layanan android market dalam bentuk software sebagai tempat bagi pengguna (user) untuk men-download aplikasi, game, dan sebagainya untuk sistem operasi Android. Play store menyediakan fitur ulasan berupa teks sebagai tempat bagi para pengguna aplikasi untuk memberikan tanggapan atau opini mereka terhadap kinerja aplikasi yang digunakan. Nilai ulasan yang diberikan oleh pengguna kepada aplikasi yang digunakan sangat berpengaruh terhadap aplikasi yang tersedia di play store tersebut.[3]

Dari perkembangan platform tersebut, banyak masyarakat yang mulai memilihnya sebagai opsi utama dalam kegiatan sehari-hari. Ini mendapatkan berbagai pihak dalam mengelola informasi dan data untuk menganalisa kemajuan yang sedang kekinian, terutama dari ulasan yang ditemukan di Google Play Store. Beberapa orang memberikan pendapat positif tentang efisiensi dan kemudahan yang ditawarkan oleh layanan tersebut, sementara yang lain mengungkapkan kekurangan berdasarkan masalah kualitas dan keamanan layanan. Dalam rangka itu, analisis sentimen yang melibatkan proses text mining dan pengelompokan berdasarkan sentimen menjadi penting untuk menilai keakuratan informasi yang diperoleh dari ulasan-ulasan tersebut.

Pada tahun 2023 Aziz dkk. melakukan penelitian tentang perbandingan metode dalam menganalisis review film[4]. Metode yang digunakan dalam penelitian ini adalah Naïve Bayes, Regresi Logistik, dan SVM. Dataset review film yang digunakan merupakan kumpulan review film yang diambil dari website Rotten Tomatoes. Metode yang mempunyai performa terbaik adalah SVM dengan hasil skor akurasi 99,4% dan skor F1 93,5%. Dan untuk referensi di masa mendatang, penulis berharap bagi peneliti selanjutnya adalah menggunakan dataset yang lebih sederhana dan seimbang dari segi sentimennya, sebaiknya sentimen dari dataset tersebut hanya mempunyai Positif dan Negatif serta proporsinya seimbang.

Pada tahun 2023 Karo dkk. melakukan penelitian analisis sentimen pada aplikasi info BMKG[5]. Dalam penelitian ini model prediksi yang dilakukan oleh peneliti dengan menggunakan metode Support Vector Machine, diperoleh yaitu data split 75:25 memiliki tingkat akurasi tertinggi sebanyak 79% dilanjut dengan data split 80:20 dengan tingkat akurasi sebanyak 78% dan data split 70:30 memiliki jumlah benar terendah yaitu sebanyak 77%. Untuk pengembangan sistem kedepan penulis berharap diperlukannya kamus khusus untuk menangani kasus penggunaan data ulasan Google Playstore yang seringkali menggunakan bahasa yang tidak baku, dapat menggunakan pendekatan dengan deteksi emoticon untuk mendeteksi emoticon serta dapat membandingkan dengan metode klasifikasi lainnya.

Pada tahun 2023 Singgalen melakukan penelitian ekstrak sentimen pada klasifikasi ulasan tamu hotel. Penelitian ini menggunakan metode SVM[6]. Penelitiannya bertujuan untuk menganalisis sentimen tamu hotel melalui pendekatan pembelajaran mesin. Kinerja SVM menghasilkan nilai akurasi sebesar 92,05%, nilai presisi sebesar 100%, nilai recall sebesar 84,08%. Penulis memberikan rekomendasi untuk penelitian lebih lanjut melibatkan melakukan analisis komparatif algoritma klasifikasi sentimen dalam konteks ulasan tamu hotel untuk menilai kesesuaian kinerja di berbagai kumpulan data.

Pada tahun 2021 Arsi dan Waluyo melakukan penelitian analisis sentimen pada wacana pemindahan ibu kota Indonesia[7]. Sumber datanya diambil dari twitter dan metode yang digunakan Support Vector Machine. Teknis klasifikasi dilakukan dengan cara mengklasifikasikan menjadi 2 kelas yakni positif dan negatif. Hasil pengujian yang dilakukan terhadap tweetsentimen pemindahan ibu kota dari media sosial twitter sebanyak 1.236 tweets (404 positif dan 832 negatif) menggunakan SVM diperoleh akurasi = 96,68%, precision =95.82%, recall=94.04%. Sebagai perbaikan dari penelitian selanjutnya perlu dikaji lebih mendalam perbandingan hasil analisis sentimen dengan menerapkan penanganan negasi dengan yang tidak mengalami penanganan negasi dalam bahasa Indonesia.

Pada tahun 2023 Vamilina dan Novita melakukan penelitian analisis sentiemen pada aplikasi e-wallet[8]. Metode yang digunakan merupakan svm berbasis Particle Swarm Optimization. Fokus pada penelitian ini yakni membandingkan komentar pengguna aplikasi e-wallet. Penelitian ini menyatakan bahwa SVM -PSO memiliki persentase akurasi yang paling tinggi dibandingkan algoritma yang lain.

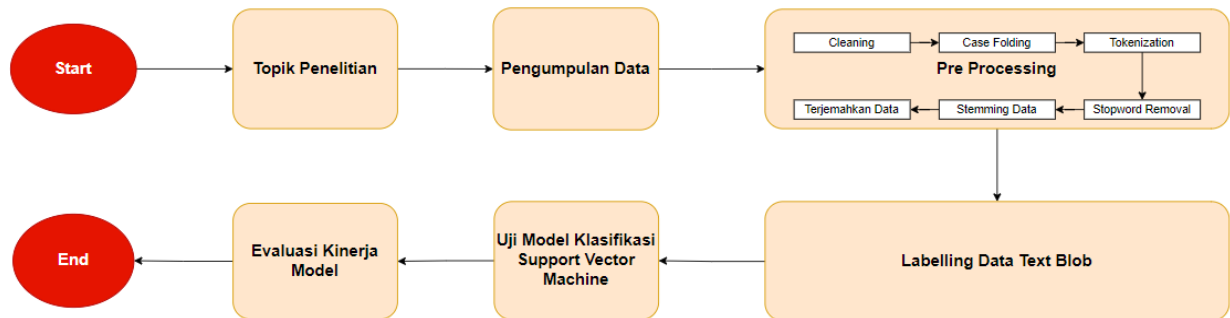
Penelitian ini bertujuan untuk mengklasifikasikan opini masyarakat Indonesia terhadap jasa transportasi online ke dalam sentimen positif, sentimen negatif, sentimen netral yang efektif untuk mengukur opini dan persepsi masyarakat Indonesia terhadap layanan jasa transportasi online melalui ulasan pengguna di Google Play Store. Memberikan pemahaman yang lebih mendalam tentang pandangan masyarakat terhadap layanan transportasi online di Indonesia, serta memberikan rekomendasi bagi penyedia layanan untuk meningkatkan kualitas layanan mereka berdasarkan hasil analisis sentimen yang diperoleh.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi yang signifikan bagi perkembangan industri transportasi online di Indonesia, sehingga layanan tersebut dapat terus meningkatkan kualitasnya dan mendapatkan lebih banyak kepercayaan dari masyarakat pengguna.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Tahapan ini dilakukan untuk pembuatan rancangan sistem dan alur pembuatan program yang akan dijadikan sebagai pedoman atau panduan dalam melakukan penelitian. Perancangan dapat berupa alur program, langkah-langkah atau flowchart[9]. Gambar 1 berikut menunjukkan gambaran lebih rinci mengenai alur tahapan metodologi penelitian.



Gambar 1. Tahapan Penelitian

2.2 Tahap Pengumpulan Data

Proses ini melakukan pengambilan ulasan dari tiga platform transportasi online yang tersedia di google play store (Gojek, Grab, Maxim) dengan menggunakan teknik crawling data *library* python yang telah diinstall. Google Play Store adalah platform distribusi digital yang berperan sebagai toko resmi untuk aplikasi dan konten digital pada perangkat berbasis Android. Semua data ulasan, rating dan informasi mengenai aplikasi android lainnya terdapat dalam Google Play Store. Proses crawling data dapat dilakukan secara manual dan otomatis menggunakan tools yang ada.

2.3 Tahap Pre-Processing Data

Tahapan sebelum pemrosesan data merupakan proses pengolahan data yang sudah didapat untuk meningkatkan kualitas sumber data untuk dimiliki lebih bersih dan terstruktur sehingga menambah keakuratan data yang akan diolah selama penelitian. Data yang sudah melalui tahap pre-processing akan digunakan dalam penelitian dan pengujian.

- Data Cleaning: Membersihkan ulasan yang tidak relevan, hapus dan hilangkan data yang tidak lengkap, tidak relevan, dan tidak akurat.
- Case folding adalah tahapan dalam preprocessing data yang berisi teks, dimana proses ini bertujuan untuk mengubah semua teks dalam dataset menjadi huruf kecil dan untuk memperbaiki konsistensi teks serta mempermudah proses pengolahan data yang akan dilakukan selanjutnya.
- Tahap tokenization adalah tahapan pemecahan seluruh isi teks ulasan menjadi potongan-potongan kata (termed word) yang berdiri secara tunggal.
- Stopword removal adalah tahapan penghapusan ulasan yang tidak memiliki makna atau kata yang kurang memiliki arti yang akan mempengaruhi proses sentimen.
- Tahap stemming data adalah tahapan mengubah seluruh kata-kata menjadi bentuk kata dasar atau mengembalikan kata ke bahasa alami tanpa imbuhan. Tujuan dari stemming data adalah mengurangi kata-kata menjadi bentuk yang sama atau setidaknya bentuk yang serupa.
- Terjemahkan data adalah seluruh data ulasan akan diterjemahkan kedalam bahasa Inggris untuk mendapatkan nilai polaritas.

2.4 Pelabelan Data

TextBlob adalah library yang digunakan untuk analisis sentimen dalam bahasa Inggris. TextBlob merupakan salah satu library yang canggih dalam analisis data teks yang dapat melakukan operasi yang kompleks. Salah satu fitur utamanya adalah kemampuannya untuk menilai sentimen individual dari setiap kata dalam sebuah kalimat. Hal ini dilakukan dengan menggunakan informasi tentang makna dan intensitas setiap kata, yang diperoleh dari kamus yang telah ditentukan sebelumnya untuk mengidentifikasi apakah kata tersebut memiliki makna positif atau negatif. Dengan demikian, TextBlob dapat memberikan informasi yang mendalam tentang polaritas dan subjektivitas dari kalimat tersebut, dengan memberikan skor yang mencakup rentang dari positif, netral, hingga negatif.

Dengan fitur-fitur tersebut, TextBlob menjadi salah satu pilihan utama dalam pemrosesan bahasa alami (Natural Language Processing/NLP) dalam lingkungan Python. Keunggulan TextBlob tidak hanya terletak pada kemampuannya dalam analisis sentimen, tetapi juga kemudahan penggunaannya dan fleksibilitasnya dalam melakukan berbagai tugas pemrosesan teks seperti tokenisasi, pemisahan kata, pembuatan frasa, analisis frasa, penerjemahan teks, dan lain sebagainya. Oleh karena itu, TextBlob sangat membantu dalam menganalisis data teks, terutama dalam konteks analisis sentimen untuk memahami pandangan dan respons pengguna terhadap produk atau layanan tertentu.



2.5 Pengembangan Model

Tahapan ini merupakan tahapan pembuatan program yang telah dirancangan dan mengimplementasikan semua metode yang digunakan untuk membuat program sesuai alur rancangan yang ada. Penelitian ini mengadopsi algoritma Support Vector Machine (SVM), sebuah pendekatan dalam pembelajaran mesin yang memanfaatkan konsep fungsi hyperplane untuk mengklasifikasikan data ke dalam berbagai daerah yang mewakili tiap kelas. Hyperplane, sebagai fungsi matematis, berperan sebagai pembagi antara kelas-kelas yang ada dalam data. Saat melakukan prediksi terhadap data baru, SVM akan menentukan label kelas berdasarkan letak data tersebut dalam ruang fitur yang dipisahkan oleh hyperplane dan memperhitungkan margin atau jarak antara data dan hyperplane tersebut.[10]

2.6 Pengujian

Tahap berikutnya dalam penelitian ini melibatkan pengujian program yang telah dikembangkan untuk mengevaluasi sejauh mana kinerja sistem dan akurasi metode yang diterapkan.[11] Tujuan utamanya adalah untuk menghasilkan informasi yang dapat diandalkan. Dalam pengujian ini, metode yang digunakan adalah Confusion Matrix. Confusion Matrix merupakan alat yang sangat berguna dalam melakukan perhitungan akurasi pada konsep data mining. Sering kali disebut juga sebagai error matrix, Confusion Matrix memberikan detail informasi mengenai hasil klasifikasi aktual dan prediksi yang dihasilkan oleh sistem klasifikasi. Performa sistem kemudian dievaluasi menggunakan informasi yang disajikan dalam bentuk matriks, memungkinkan analisis yang lebih mendalam tentang kualitas klasifikasi yang dilakukan.

Dalam pengujian, kelas terklasifikasi positif mencakup *True Positive* (TP) dan *False Negative* (FN), sementara kelas terklasifikasi negatif melibatkan *False Positive* (FP) dan *True Negative* (TN). Berdasarkan nilai-nilai ini, beberapa metrik kinerja yang umum digunakan adalah *f1-score*, *accuracy*, *recall* dan *precision*.

Untuk memperoleh nilai akurasi maka dilakukan pengukuran proporsi total data yang diprediksi dengan benar oleh model, sedangkan presisi mengukur proporsi dari data positif yang diprediksi benar dari total data yang diprediksi positif. Recall, di sisi lain, mengukur seberapa banyak dari semua kasus positif yang diidentifikasi oleh model. Sementara itu, F1-score diukur untuk memperoleh nilai perbandingan harmonik dan keseimbangan antara recall dan presisi. Rumus-rumusnya adalah sebagai berikut:

- Presisi (Precision) mengukur tingkat ketepatan model dalam mengklasifikasikan data positif, baik yang sebenarnya positif maupun yang sebenarnya negatif. Presisi dihitung dengan membagi jumlah data yang diklasifikasikan dengan benar (True Positive dan True Negative) dengan jumlah keseluruhan data yang diklasifikasikan (True Positive, True Negative, False Positive, dan False Negative), kemudian hasilnya dikalikan dengan 100%.

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

- Recall mengukur kemampuan model untuk menemukan kembali semua data positif. Recall dihitung dengan membagi jumlah data positif yang diklasifikasikan dengan benar (True Positive) dengan jumlah keseluruhan data yang sebenarnya positif (True Positive dan False Negative), kemudian hasilnya dikalikan dengan 100%.

$$\text{Rumus} = \frac{TP}{TP + FN} \times 100\% \quad (2)$$

- F1-score adalah perbandingan rata-rata harmonis antara Presisi dan Recall. F1-score mengukur keseimbangan antara Presisi dan Recall. F1-score yang tinggi menunjukkan bahwa model memiliki Presisi dan Recall yang baik secara bersamaan. F1-score dihitung dengan rumus:

$$\text{Rumus} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \times 100\% \quad (3)$$

3. HASIL DAN PEMBAHASAN

3.1 Koleksi Data

Data diperoleh menggunakan teknik web scraping di *Google Play Store* dengan bahasa pemrograman python. Data yang dikumpulkan meliputi review pada tanggal 1 Januari 2023 hingga 28 Februari 2024. Data yang diambil berupa review atau komentar yang diberikan oleh pengguna aplikasi transportasi online (Grab, Maxim, dan Gojek) dengan jumlah sebanyak 6503 ulasan. Berikut contoh data review aplikasi transportasi online. Didalam tabel 1 berikut terdapat semua ulasan-ulasan yang masih belum diproses atau belum di kelola.

Tabel 1. Contoh koleksi beberapa data ulasan

	at	userName	content
0	2023-09-19 19:52:56	Balfir YT	tolong dong gojek kenapa ya saya mau verifikasi email ribet amat katanya udah dikirim link verifikasinya tapi blm msk juga ke akun email saya padahal akun email saya fine fine aja ga ada problem udah dicek



1	2023-09-22 05:35:14	Jhonson M Tampubolon RM	saya kesulitan memasukkan alamat tepat penerima paket di gosend instan saya mesti jelasin lagi sama drivernya alamat penerima paket mohon perbaikan aplnya memang sampai saat ini kiriman2 paket saya masih sampai ketujuan alias masih oke
2	2023-09-23 03:42:20	Khasmita Farkhatunnaja h	setelah diupdate kok sering banget ngelag ya padahal aku pakai koneksi wifi pribadi dan memory ram 12gb gedhe tolong perbaiki aplikasinya kak error mulu baru beberapa menit buka aplikasi langsung error lagi

3.2 Pre – Processing

Data yang telah dikumpulkan adalah data mentah yang berisi tanda baca, karakter, emoticon, penulisan kata yang kurang lengkap sehingga perlu dilakukan proses untuk menyiapkan data. Data proses yang dimaksud adalah tahap Pre-Processing, dimana pada tahap ini menghapus beberapa atribut review yang tidak digunakan sehingga memastikan bahwa data dapat diproses secara efektif. Tabel 2 berikut adalah data yang menunjukkan yang melalui proses tahapan pre-processing terdiri dari cleaning, kemudian case folding, setelah itu melalui tokenizing, stopword removal, dan stemming.

Tabel 2. Tahap pre-Processing pada data ulasan

at	userName	content	Content_token	Stemmed	Text_string	
0	2023-09-19 19:52:56	Balfir YT	tolong dong gojek kenapa ya saya mau verifikasi email ribet amat katanya udah dikirim link verifikasinya tapi blm msk juga ke akun email saya padahal akun email saya fine fine aja ga ada problem udah dicek	['tolong', 'gojek', 'ya', 'verifikasi', 'email', 'ribet', 'udah', 'dikirim', 'link', 'verifikasinya', 'blm', 'msk', 'akun', 'email', 'fine', 'fine', 'aja', 'ga', 'problem', 'udah', 'dicek']	['tolong', 'gojek', 'ya', 'verifikasi', 'ribet', 'email', 'link', 'verifikasi', 'link', 'email', 'fine', 'fine', 'problem', 'udah', 'aja', 'ga', 'problem', 'udah', 'cek']	tolong gojek verifikasi email ribet udah kirim link verifikasi akun email fine fine problem udah
1	2023-09-22 05:35:14	Jhonson M Tampubolon RM	saya kesulitan memasukkan alamat tepat penerima paket di gosend instan saya mesti jelasin lagi sama drivernya alamat penerima paket mohon perbaikan aplnya memang sampai saat ini kiriman2 paket saya masih sampai ketujuan alias masih oke	['kesulitan', 'memasukkan', 'alamat', 'penerima', 'paket', 'gosend', 'instan', 'mesti', 'jelasin', 'drivernya', 'alamat', 'penerima', 'paket', 'mohon', 'perbaikan', 'aplnya', 'kiriman2', 'paket', 'tujuan', 'alias', 'oke']	['sulit', 'masuk', 'alamat', 'terima', 'paket', 'gosend', 'instan', 'mesti', 'jelasin', 'drivernya', 'alamat', 'paket', 'mohon', 'baik', 'aplnya', 'kiriman2', 'paket', 'aplnya', 'paket', 'tuju', 'alias', 'oke']	sulit masuk alamat terima paket gosend instan mesti jelasin drivernya alamat terima paket mohon baik aplnya kiriman2 paket tuju alias
2	2023-09-23 03:42:20	Khasmita Farkhatunnajah	setelah diupdate kok sering banget ngelag ya padahal aku pakai koneksi wifi pribadi dan memory ram 12gb gedhe tolong perbaiki aplikasinya kak error mulu baru beberapa menit buka aplikasi langsung error lagi	['diupdate', 'banget', 'ngelag', 'ya', 'pakai', 'koneksi', 'wifi', 'pribadi', 'memory', 'ram', '12gb', 'gedhe', 'tolong', 'perbaiki', 'aplikasinya', 'kak', 'error', 'mulu', 'menit', 'buka', 'aplikasi', 'langsung', 'error']	['diupdate', 'banget', 'ngelag', 'ya', 'pakai', 'koneksi', 'wifi', 'pribadi', 'memory', 'ram', '12gb', 'gedhe', 'tolong', 'perbaiki', 'aplikasi', 'kak', 'error', 'mulu', 'menit', 'buka', 'aplikasi']	diupdate banget ngelag pakai koneksi wifi pribadi memory 12gb gedhe tolong baik aplikasi error mulu menit buka aplikasi langsung error



'langsung',
'error']

Kemudian data pre-processing yang telah jadi dibuat kedalam data frame dimana kolom yang diambil adalah 'at', 'userName', 'text_string'. Tabel 3 berikut menunjukkan proses dataframe dilakukan.

Tabel 3. Dataframe

	At	userName	Text_string
0	2023-09-19 19:52:56	Balfir YT	tolong gojek verifikasi email ribet udah kirim link verifikasi akun email akun email fine fine problem udah
1	2023-09-22 05:35:14	Jhonsoon M Tampubolon RM	sulit masuk alamat terima paket gosend instan mesti jelasin drivernya alamat terima paket mohon baik aplnya kiriman2 paket tuju alias
2	2023-09-23 03:42:20	Khasmita Farkhatunnajah	diupdate banget ngelag pakai koneksi wifi pribadi memory 12gb gedhe tolong baik aplikasi error mulu menit buka aplikasi langsung error

3.3 Data Labelling

Setelah melalui tahap pre-processing, proses berikutnya meliputi tahap labelling, dimana data yang digunakan adalah kolom review atau text_string sebagai penentuan sentimen netral, positif, negatif. TextBlob adalah library yang memiliki fungsi analisis sentimen dalam bahasa Inggris. Sesuai dengan perancangan model, labelling data yang digunakan adalah labelling vader, dimana semua data terlebih dahulu diterjemahkan kedalam bahasa Inggris, kemudian masuk pada tahap penentuan sentimen positif, negatif dan netral. Tabel 4 berikut menunjukkan proses *translate* data dilakukan.

Tabel 4. Translate data

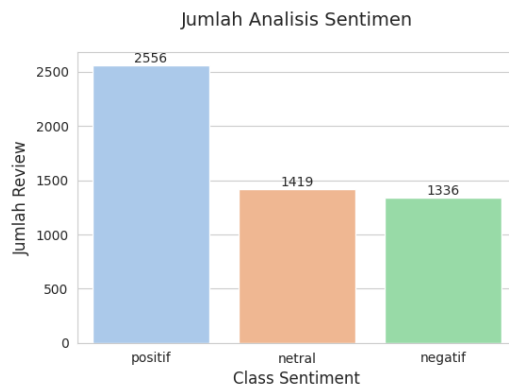
Sebelum terjemahan	Sesudah Terjemahan
tolong gojek verifikasi email ribet udah kirim link verifikasi akun email akun email fine fine problem udah	please Gojek email verification is complicated, I've sent a link to verify my email account, my email account is fine, the problem is fine
sulit masuk alamat terima paket gosend instan mesti jelasin drivernya alamat terima paket mohon baik aplnya kiriman2 paket tuju alias	it's difficult to enter the address for receiving instant Gosend packages, you have to explain the driver, address for receiving packages, please use the app to send packages to destination alias
diupdate banget ngelag pakai koneksi wifi pribadi memory 12gb gedhe tolong baik aplikasi error mulu menit buka aplikasi langsung error	Updates are really lagging using a private WiFi connection, big 12GB memory, please be good, the application error takes a few minutes to open the application, the error immediately occurs

Kemudian setelah semua data ulasan atau text_string telah diterjemahkan, maka akan dilabeli melalui library Text Blob sesuai dengan ketentuannya. Polarity dalam TextBlob adalah nilai yang menunjukkan seberapa positif, negatif, atau netral sebuah teks. Nilai polaritas berkisar dari -1 hingga 1, di mana nilai -1 menunjukkan sentimen yang sangat negatif, nilai 1 menunjukkan sentimen yang sangat positif, dan nilai 0 menunjukkan sentimen yang netral. Dengan kata lain, polarity mengukur tingkat kepositifan atau kenegatifan suatu teks. Semakin tinggi nilai polaritas, semakin positif sentimen teks tersebut, sementara semakin rendah nilai polaritas, semakin negatif sentimen teks tersebut. Jika nilai polaritas mendekati 0, itu menunjukkan bahwa teks memiliki sentimen yang netral. Tabel 5 berikut menunjukkan proses labelling data dilakukan.

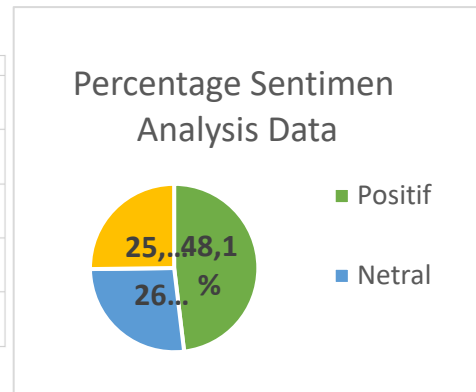
Tabel 5. Labelling Data

	Text_string	Sentiment_polarity	sentiment
0	please Gojek email verification is complicated, I've sent a link to verify my email account, my email account is fine, the problem is fine	0,111111111111111	Positif
1	it's difficult to enter the address for receiving instant Gosend packages, you have to explain the driver, address for receiving packages, please use the app to send packages to destination alias	-0,25	Negatif
2	Updates are really lagging using a private WiFi connection, big 12GB memory, please be good, the application error takes a few minutes to open the application, the error immediately occurs	0,116666666666667	Positif

Setelah melewati proses pelabelan dan semua data sudah diberi label, selanjutnya adalah melakukan distribusi sentimen. Sentimen yang akan di distribusikan dalam bentuk sns plot dan pie chart sehingga bisa terlihat visualisasi perbedaan jumlah data berdasarkan sentimen yang ada.



Gambar 2. Bar Analisis



Gambar 3. Pie chart Analisis

Didalam visualisasi gambar 2 dan gambar 3 diatas menunjukkan terdapat sentimen yang paling unggul adalah positif sebanyak 2556 (48,1%), kemudian sentimen netral sebanyak 1419 (26,7%), dan sentimen negatif 1336 (25,2%).

3.4 WordCloud

Data penelitian sekarang sudah memiliki label dan sudah dikelompokkan segmen, yaitu : netral, positif, negatif. Oleh karena itu, data akan melalui proses pembuatan word cloud untuk menghasilkan representasi visual dari teks yang terdapat dalam data. Wordcloud digunakan sebagai alat bantu dalam menganalisis teks, di mana kata-kata dalam teks digambarkan dalam visualisasi dimana ukuran huruf bergantung pada frekuensi kemunculan kata tersebut dalam teks. Dengan kata lain, kata-kata yang paling sering muncul dalam data akan ditampilkan dalam ukuran huruf yang lebih besar dalam word cloud. Fungsi wordcloud adalah untuk membantu dalam melihat atau menggambarkan kata-kata dari masing-masing segmen sehingga kata yang paling sering muncul akan ditampilkan.

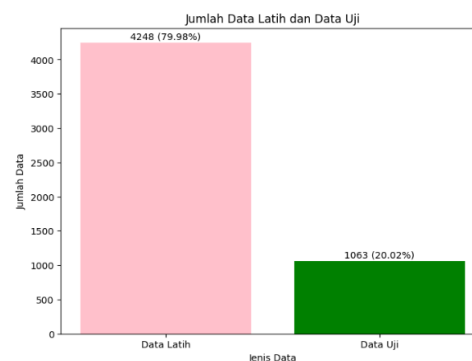


Gambar 4. wordcloud sentimen

Berdasarkan grafik diatas maka diketahui kata-kata yang memiliki frekuensi paling sering muncul terbanyak yaitu driver, application, dan order. Gambar 4 menunjukkan wordcloud ulasan yang sering muncul dengan tampilan kata yang besar.

3.5 Dataset

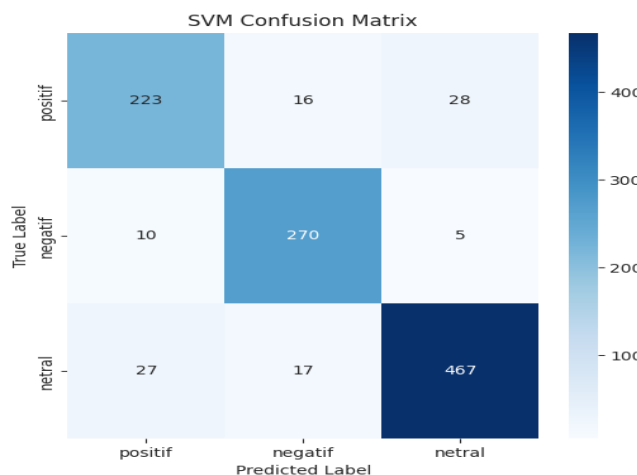
Data review aplikasi transportasi online dipecah menjadi bagian, training data (data pelatihan) dan testing data (data uji). Pada gambar 3.4 berikut adalah pembagian data testing dan data training. Pembagian dengan rasio 8:2 dimana data testing 80% sebanyak 4248 data dan data training 20% sebanyak 1063 data. Gambar 5 berikut adalah jumlah pemecahan data menjadi data latih dan data uji.



Gambar 5. Data latih dan uji

3.6 Evaluasi kinerja model

Setelah tahap pelatihan, kinerja model dievaluasi berdasarkan tahap pengujian. Confusion matrix menjadi instrumen kunci untuk memeriksa dan menganalisis hasil prediksi yang dihasilkan oleh algoritma Support Vector Machine. Melalui penggunaan confusion matrix, dapat mengevaluasi nilai-nilai seperti akurasi *precision*, *accuracy*, *f1-score*, dan *recall*. Nilai-nilai ini memberikan wawasan yang mendalam tentang kinerja sistem klasifikasi dalam memproses data dan memprediksi kelasnya dengan tepat.

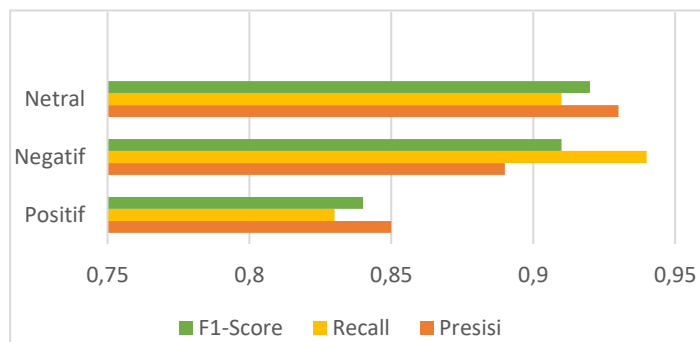


Gambar 6. Visualisasi Confusion Matrix

Sesuai dengan gambar 6 setiap kelas sudah didapat matrixnya, kemudian menentukan performance metrics dari setiap kelas dengan cara mengimplementasikan nilai TP, TN, FP, dan FN dari masing-masing kelas kedalam rumus. Berikut tabel 6 menunjukkan hasil performance matrix dan gambar 7 menunjukkan hasil dalam bentuk bar:

Tabel 6. Hasil Performance Matrix

Sentimen	Parameter		
	Presisi	Recall	F1-Score
Positif	0,85	0,83	0,84
Negatif	0,89	0,94	0,91
Netral	0,93	0,91	0,92
Average	0,89	0,89	0,89



Gambar 7. Bar Performance

Berikut hasil dari *Performance Metrics* untuk metode klasifikasi *Support Vector Machine*.

a. Sentimen Positif:

1. Presisi yang mencapai 0,85 menunjukkan bahwa sebagian besar teks yang diklasifikasikan sebagai sentimen positif oleh model memang benar-benar memiliki sentimen positif. Namun, masih ada sebagian kecil teks yang salah diklasifikasikan sebagai positif.
2. Recall sebesar 0,83 menunjukkan bahwa model berhasil mengidentifikasi sebagian besar teks yang sebenarnya memiliki sentimen positif. Namun, ada beberapa teks yang sebenarnya positif namun tidak terdeteksi oleh model.
3. F1-Score yang seimbang sebesar 0,84 menunjukkan keseimbangan yang baik antara presisi dan recall dalam mengklasifikasikan sentimen positif.

b. Sentimen Negatif:

1. Dengan presisi tinggi sebesar 0,89, model jarang melakukan kesalahan dalam mengklasifikasikan teks sebagai sentimen negatif. Namun, masih ada sejumlah kecil teks yang salah diklasifikasikan sebagai negatif.
 2. Recall yang sangat baik sebesar 0,94 menunjukkan bahwa model berhasil mengidentifikasi sebagian besar teks yang sebenarnya memiliki sentimen negatif.
 3. F1-Score yang tinggi sebesar 0,91 menunjukkan kinerja yang kuat dalam mengklasifikasikan sentimen negatif.
- c. Sentimen Netral:
1. Presisi yang tinggi sebesar 0,93 menunjukkan bahwa sebagian besar teks yang diklasifikasikan sebagai netral oleh model memang tidak memiliki sentimen yang jelas.
 2. Recall yang baik sebesar 0,91 menunjukkan bahwa model dapat mengidentifikasi sebagian besar teks yang sebenarnya memiliki sentimen netral.
 3. F1-Score yang tinggi sebesar 0,92 mencerminkan kinerja yang baik dalam mengklasifikasikan sentimen netral.
- Rata-rata presisi, recall, dan F1-Score yang tinggi sebesar 0,89 menunjukkan bahwa model memiliki kinerja yang konsisten dan seimbang dalam mengklasifikasikan sentimen untuk semua kategori.

4. KESIMPULAN

Berdasarkan hasil perancangan, implementasi dan analisis yang telah dilakukan dengan textblob dan metode klasifikasi SVM dan proses perbandingan pengukurannya pada studi kasus opini masyarakat Indonesia tentang Jasa Transportasi *Online*, dapat disimpulkan bahwa opini masyarakat Indonesia tentang Jasa Transportasi *Online* berdasarkan data yang diambil dari *google play store* menunjukkan bahwa sentimen Positif memperoleh persentase tertinggi sebesar 48,1%, kemudian disusul sentimen Netral sebesar 26,7% dan Negatif sebesar 25,2%. Kemudian, hasil evaluasi kinerja model SVM menunjukkan parameter *recall* memiliki nilai akurasi yang paling tinggi sebesar 0,94 pada sentimen Negatif. Sementara parameter *precision* dan *f1-score* memperoleh hasil tertinggi pada sentimen Netral dengan nilai masing-masing sebesar 0,93 dan 0,92.

REFERENCES

- [1] 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.
- [2] A. N. Alfarobby and H. Irawan, "Analisis Sentimen Kepuasan Konsumen Pengguna Transportasi Online Pada Ulasan Google Playstore Menggunakan Indobert Dan Topic Modeling (Studi kasus: Gojek dan Grab)," *e-Proceeding of Management*, vol. 11, no. 1, p. 72, Feb. 2024.
- [3] 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.
- [4] M. M. Aziz, M. D. Purbalaksone, and A. Adiwijaya, "Method comparison of Naïve Bayes, Logistic Regression, and SVM for Analyzing Movie Reviews," *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 4, pp. 1714-1720, Mar. 2023, doi: 10.47065/bits.v4i4.2644.
- [5] I. M. Karo Karo, J. A. Karo Karo, Y. Yuniarto, 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.
- [6] Y. A. Singgalen, "Extract Sentiment and Support Vector Machine (SVM) Performance of Hotel Guest Review Classification," *Technology and Science (BITS)*, vol. 5, no. 3, pp. 627-635, Dec. 2023, doi: 10.47065/bits.v5i3.4737.
- [7] 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.
- [8] V. Vamilina and R. Novita, "Analisis Sentimen E-Wallet Menggunakan Support Vector Machine Berbasis Particle Swarm Optimization," *Building of Informatics, Technology and Science (BITS)*, vol. 5, no. 1, pp. 40-48, Jun. 2023, doi: 10.47065/bits.v5i1.3526.
- [9] Z. Alhaq, A. Mustopa, and J. D. Santoso, "PENERAPAN METODE SUPPORT VECTOR MACHINE UNTUK ANALISIS SENTIMEN PENGGUNA TWITTER," *JOISM : JURNAL OF INFORMATION SYSTEM MANAGEMENT*, vol. 3, no. 1, 2021.
- [10] N. Tri Romadloni, I. Santoso, S. Budilaksone, and M. Ilmu Komputer STMIK Nusa Mandiri Jakarta, "PERBANDINGAN METODE NAIVE BAYES, KNN DAN DECISION TREE TERHADAP ANALISIS SENTIMEN TRANSPORTASI KRL COMMUTER LINE," *Jurnal IKRA-ITH Informatika*, vol. 3, no. 2, Jul. 2019.
- [11] 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.
- [12] 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>
- [13] 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.
- [14] 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>
- [15] 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>
- [16] 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.
- [17] 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.
- [18] A. Safira, A. S. Masyarakat...□, and F. N. Hasan, “ANALISIS SENTIMEN MASYARAKAT TERHADAP PAYLATER MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER,” *Jurnal Sistem Informasi*, vol. 5, no. 1, Jan. 2023.
- [19] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN,” *Jurnal KomtekInfo*, vol. 10, no. 1, pp. 1–7, Jan. 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [20] 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.
- [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] D. Aryanti, “Analisis Sentimen Ibukota Negara Baru Menggunakan Metode Naive Bayes Classifier,” *Journal of Information System Research (JOSH)*, vol. 3, no. 4, pp. 524–531, Jul. 2022, doi: 10.47065/josh.v3i4.1944.
- [23] 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.
- [24] A. Kusuma and H. Nurramdhani Irmanda, “Analisis Sentimen Pada Ulasan Aplikasi Indodax di Google Play Store Menggunakan Metode Support Vector Machine,” Aug. 2022.
- [25] 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.
- [26] E. Putri Nirwandani and R. Cahya Wihandika, “Analisis Sentimen Pada Ulasan Pengguna Aplikasi Mandiri Online Menggunakan Metode Modified Term Frequency Scheme Dan Naive 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>
- [27] 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>
- [28] 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.
- [29] L. O. Sihombing, H. Hannie, and B. A. Dermawan, “Sentimen Analisis Customer Review Produk Shopee Indonesia Menggunakan Algoritma Naive Bayes Classifier,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 5, no. 2, pp. 233–242, Dec. 2021, doi: 10.29408/edumatic.v5i2.4089.
- [30] 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.