ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI KASIR WIRAUSAHA MAJOO MENGGUNAKAN METODE SVM

Diana Masruroh, Andriyan Rizki Jatmiko, Listanto Tri Utomo
Sistem Informasi, Universitas Merdeka Malang
Jalan Terusan Dieng, Malang,Indonesia
andriyan.jatmiko@unmer.ac.id

ABSTRAK

Peningkatan adopsi teknologi aplikasi kasir digital seperti Aplikasi Kasir Wirausaha Majoo oleh pelaku usaha mikro, kecil, dan menengah (UMKM) mempengaruhi pentingnya pemahaman mengenai pandangan dan pengalaman pengguna terhadap aplikasi. Penelitian ini bertujuan untuk mengevaluasi perasaan dari ulasan pengguna Aplikasi Kasir Wirausaha Majoo dengan menerapkan metode *Support Vector Machine* (SVM) serta langkah-langkah dari alur kerja SEMMA (*Sample, Explore, Modify, Model, Assess).* Pada tahap *Sample* dilakukan pengumpulan data dari ulasan yang terdapat di *Google Play Store*. Selanjutnya tahap *Explore*, dilakukan analisis untuk memahami distribusi penilaian dari pengguna. Tahap *Modify* terdiri dari proses pembersihan teks. Pada tahapan *Model* dilakukan pelabelan data berbasis *lexicon* dan pembobotan kata dengan TF-IDF, kemudian model SVM dilatih dengan beberapa proporsi pembagian data (60:40, 70:30, 80:20, dan 90:10) untuk mengevaluasi kinerja klasifikasi. Proses evaluasi pada tahapan *Assess* dilakukan menggunakan metrik akurasi, *precision, recall,* dan *F1-score*. Hasil dari penelitian menunjukkan bahwa rasio 90:10 memberikan hasil terbaik dengan akurasi mencapai 88,12%, *precision* 88,51%, *recall* 88,12%, dan *F1-score* 88,26%. Klasifikasi sentimen positif dan negatif dilakukan dengan baik, namun performa untuk kelas netral terlihat kurang memuaskan. Temuan ini mengindikasikan bahwa SVM efektif dalam menganalisis opini pengguna dan dapat dijadikan sebagai dasar untuk meningkatkan layanan aplikasi kasir digital Majoo.

Kata kunci : *Analisis Sentimen, SVM, Aplikasi Kasir Wirausaha Majoo, SEMMA*

1. PENDAHULUAN

Berdasarkan kemajuan pesat di era digital, penggunaan teknologi di berbagai sektor, termasuk bisnis, menjadi hal yang sangat penting. Maka dari itu, pelaku Usaha Mikro Kecil dan Menengah (UMKM) mulai secara luas mengintegrasikan teknologi ke dalam kegiatan usaha mereka. Menurut majalah Program Adaptasi dan Transformasi Ekonomi Nasional yang diterbitkan oleh Kementerian Koperasi dan UKM, tercatat pada tahun 2022 terdapat 64,2 juta unit UMKM di Indonesia. Dari data tersebut, sekitar 30,4% atau 19,5 juta unit usaha telah melakukan digitalisasi dengan menerapkan teknologi dalam operasional mereka [1]. Hingga bulan Juli 2024, Kementerian Koperasi dan UKM melaporkan bahwa jumlah pelaku UMKM yang menjalani transformasi digital semakin meningkat, mencapai 25,5 juta. Menurut Siti Azizah, Deputi Kewirausahaan di Kementerian Koperasi dan UKM, digitalisasi dapat mencakup pemasaran secara digital, produksi, pengelolaan sumber daya manusia, serta sistem pembayaran [2].

Salah satu bentuk perkembangan teknologi untuk menunjang bisnis UMKM adalah hadirnya sistem *Point Of Sales* (*POS)*. Teknologi ini memudahkan pengguna untuk proses pengolahan data pelanggan, stok, transaksi, dan laporan penjualan berjalan secara efisien, cepat, akurat, dan aman [3]. Aplikasi Kasir Wirausaha Majoo merupakan salah satu platform kasir digital yang memiliki banyak pengguna di Indonesia. Aplikasi ini dirilis secara komersial pada tahun 2019 dengan tujuan awal untuk mendukung pelaku UMKM dari skala kecil hingga besar dalam menghadapi kesulitan dalam mengelola proses digitalisasi bisnis mereka. Pengembang dari aplikasi ini menyediakan berbagai fitur sebagai solusi menyeluruh yang mendukung perkembangan bisnis UMKM dengan biaya yang terjangkau. Berdasarkan data terbaru dari *Google Play Store* per tanggal 1 April 2025, aplikasi ini telah diunduh lebih dari 100 ribu kali. Dari total unduhan tersebut, terdapat lebih dari 5000 ulasan yang diberikan pengguna berkaitan dengan kinerja aplikasi. Dengan *rating* total 3,7, aplikasi ini dapat dikategorikan sebagai memiliki *rating* yang relatif rendah, dan banyak ulasan dari pengguna menunjukkan bahwa pengembang masih belum memenuhi harapan banyak pengguna. Melalui pengamatan singkat, ditemukan beberapa ulasan yang menyebutkan bahwa aplikasi sering mengalami kesalahan, dengan banyak *bug* yang muncul serta pelayanan pelanggan yang dianggap kurang responsif. Namun, beberapa pengguna juga memberikan tanggapan positif mengenai layanan yang tersedia dan bagaimana aplikasi ini membantu operasional UMKM. Banyaknya ulasan pengguna menyulitkan pengembang dalam memperoleh gambaran umum secara akurat. Oleh karena itu, dibutuhkan metode yang efisien untuk mengekstraksi informasi seperti keluhan, tanggapan, dan sikap pelanggan guna mendukung evaluasi dan pengambilan keputusan bisnis. Salah satu metode yang efektif adalah analisis sentimen melalui pengolahan data teks.

Analisis sentimen merupakan bagian dari NLP yang digunakan untuk mengidentifikasi sikap positif,

negatif, atau netral dalam ulasan pengguna terhadap suatu produk atau layanan [4]. Karena volume data yang besar, analisis manual menjadi tidak efisien, sehingga pendekatan *machine learning* diperlukan untuk mengklasifikasikan sentimen secara otomatis dan akurat [5]. Salah satu metode *machine learning* dalam proses analisis sentimen adalah *Support Vector Machine*. SVM adalah metode *machine learning* yang efektif untuk klasifikasi data linear dan non-linear dengan akurasi tinggi. Metode ini memisahkan kelas menggunakan *hyperplane* optimal dan memanfaatkan fungsi kernel untuk data non-linear [6].

Dalam penelitian oleh Marpaung dan kawan-kawan pada tahun 2024 menunjukkan bahwa *Support Vector Machine* menghasilkan akurasi 83,2% dalam analisis sentimen pengguna BCA Mobile, lebih tinggi dibanding Naïve Bayes yang hanya mendapatkan akurasi 71% [7]. Sementara itu, studi oleh Putra & Nababan di tahun 2025 pada ulasan Alfagift dan Midi Kriing juga menunjukkan performa SVM yang unggul dengan akurasi masing-masing 85% dan 87% [8]. Kedua studi ini menegaskan bahwa metode *Support Vector Machine* memiliki keunggulan dalam klasifikasi sentimen.

Berdasarkan latar belakang tersebut, penelitian ini fokus pada analisis sentimen menggunakan metode SVM terhadap ulasan Aplikasi Kasir Wirausaha Majoo. Minimnya studi pada aplikasi digital pendukung UMKM menjadi alasan utama, sehingga penelitian ini diharapkan mampu mengisi celah tersebut dan memberikan informasi akurat terkait evaluasi kinerja dan layanan Aplikasi Kasir Wirausaha Majoo.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen adalah salah satu teknik *Natural Language Processing* yang mengevaluasi opini, emosi, dan sikap terhadap berbagai entitas seperti produk atau layanan. Sentimen diklasifikasikan menjadi positif, negatif, atau netral, dan dapat disampaikan secara eksplisit maupun implisit [9].

2.2. SEMMA

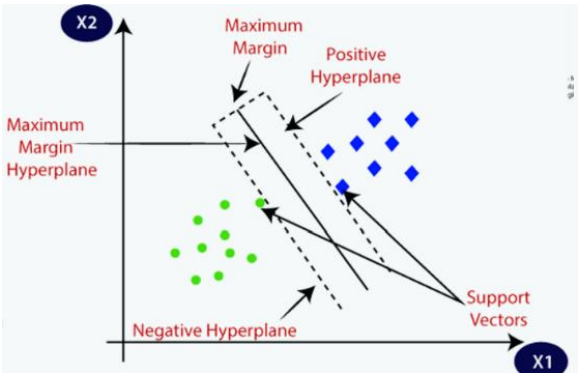
Data mining merupakan proses menemukan pola menarik dan berguna dalam kumpulan data besar. *Data Mining* sering disebut sebagai penambangan pengetahuan dari data yang terdiri dari berbagai teknik analisis seperti klasifikasi, klasterisasi, asosiatif, dan analisis regresi [10]. Salah satu metode yang ada pada proses *data mining* adalah metode SEMMA. Metode ini adalah pendekatan yang dikembangkan oleh SAS Institute. SEMMA merupakan akronim dari *Sample* (Ekstraksi data), *Explore* (Eksplorasi Data), *Modify* (Modifikasi Data), *Model* (Pemodelan Data) dan *Assess* (Evaluasi Model) yang mencerminkan tahapan utama dalam eksplorasi data secara sistematis untuk menghasilkan informasi [11].

2.3. *Text Preprocessing*

Text preprocessing adalah tahap awal dalam pengolahan data teks untuk membersihkan dan memperbaiki struktur teks agar siap dianalisis [12]. Dalam tahapan ini meliputi proses *cleaning, case folding, normalization, tokenizing, stopwords removal,* dan *stemming.* Tahap ini penting karena dapat meningkatkan akurasi model dengan menyusun kata menjadi lebih relevan dan mengurangi kata-kata yang tidak diperlukan. Hasil yang optimal diperoleh melalui normalisasi kata, koreksi *typo*, serta penghapusan kata-kata tidak penting [13].

2.4. *Support Vector Machine*

Support Vector Machine adalah salah satu algoritma *machine learning* yang diperkenalkan oleh Cortes & Vapnik pada tahun 1995 melalui publikasi dengan judul Support Vector Networks. Metode ini didasari oleh teori pembelajaran statistik yang efektif digunakan untuk menyelesaikan masalah klasifikasi dan regresi. Pada prosesnya metode SVM bekerja dengan menerapkan pencarian *hyperplane* secara optimal yang memiliki margin maksimum untuk memisahkan kelas-kelas dalam suatu data [14].



Gambar 1. *Support Vector Machine*

Pada gambar 1, SVM bekerja dengan mencari *hyperplane* optimal yang memisahkan dua kelas data dengan margin terbesar, yaitu jarak terdekat antar titik data ke *hyperplane*. Tujuannya adalah meminimalkan kesalahan klasifikasi dan meningkatkan generalisasi model. Proses dimulai dari pengumpulan data dengan fitur dan label yang relevan. *Hyperplane* dibentuk berdasarkan titik-titik data terdekat, yang disebut *support vectors*, karena berperan penting dalam menentukan posisi batas pemisah [15]. Persamaan umum dari SVM digambarkan sebagai berikut:

$$f(x) = w \cdot x + b$$

Dimana:

w = vektor bobot penentu arah dan orientasi *hyperplane*

x = vektor fitur dari titik data

b = bias, penentu posisi *hyperplane*

2.5. *Confussion Matrix*

Confusion matrix merupakan tabel yang menggambarkan jumlah kelas prediksi dan kelas aktual. Gambaran ini memberikan pemetaan hasil klasifikasi, termasuk jumlah prediksi yang benar dan kesalahan yang dibuat model dalam membedakan kelas [16]. Komponen utama dalam matriks ini diantaranya adalah *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Dari komponen tersebut selanjutnya dapat menghitung metrik evaluasi turunan yaitu dengan rumus manual sebagai berikut:

a. *Accuracy* (akurasi), mengukur hasil prediksi benar dibandingkan dengan total prediksi yang dibuat. Diperoleh dari perrsamaan:

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \quad (1)$$

b. *Precision* (Presisi), mengukur prediksi positif benar dibanding semua prediksi positif model. Hasil presisi didapatkan dari persamaan:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

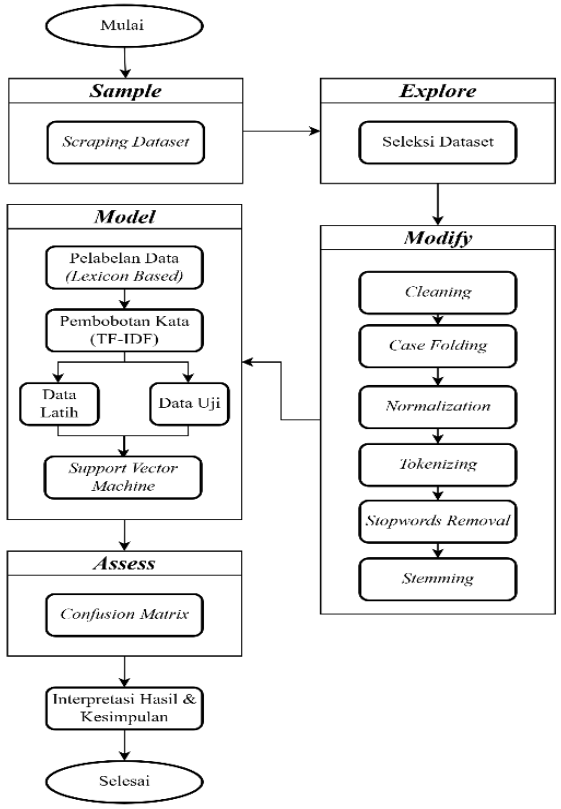
c. *Recall* (Sensitivitas), menunjukkan hasil prediksi positif dibanding nilai aktual positif diklasifikasikan benar oleh model. Diperoleh melalui persamaan:

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

d. *F1-Score (F-Measure*), digunakan untuk menyeimbangkan nilai *precision* dan *recall* melalui persamaan:

$$f1 - score = 2\ x\ \frac{Precision\ x\ Recall}{Precision + Recall} \quad (4)$$

3. METODE PENELITIAN



Gambar 2. Metode penelitian

Penelitian yang dilakukan menggunakan ulasan Aplikasi Kasir Wirausaha Majoo menjadi objek penelitian. Bertujuan untuk mengetahui pandangan pengguna mengenai layanan yang disediakan dapat berupa penilaian, pengalaman, kepuasan, atau bahkan keluhan.

Penelitian ini menjadikan metode SEMMA *(Sample, Explore, Modify, Model, Assess)* sebagai kerangka kerja utama. Pengunaan pendekatan metode ini dipilih agar dapat mengelolah dan melakukan analisis data sentimen secara runtut dan sistematis.

3.1. *Sample*

Tahap awal ini dilakukan pengumpulan data dengan teknik *scraping dataset* untuk bahan utama analisis data. Data yang dikumpulkan merupakan data ulasan pengguna Aplikasi Kasir Wirausaha Majoo pada *Google Play Store*.

3.2. *Explore*

Pada tahapan ini melakukan eksplorasi data dengan mencari tahu penjelasan mengenai ringkasan data. yang memberikan deskripsi mengenai informasi dari data yang digunakan. Tahapan ini juga mencakup proses seleksi fitur *dataset* dan memastikan penggunaan Bahasa Indonesia yang dibutuhkan dalam pengolahan data. Proses penghapusan duplikasi data dan pembersihan nilai kosong juga dilakukan pada tahapan ini. Selanjutnya melakukan visualisasi data untuk menampilkan distribusi *rating* dalam ulasan pengguna yang diberikan.

3.3. *Modify*

Pada tahapan *modify* melibatkan proses *text preprosesing* dengan tujuan untuk pembersihan, transformasi, dan persiapan lanjutan untuk pemodelan data. Proses diawali dengan *cleaning* untuk pembersihan karakter yang tidak dibutuhkan dalam analisis data sentimen. Karakter yang dihilangkan diantaranya adalah *mentions*, *hashtag*, tautan, angka, tanda baca, spasi berlebih, dan emoji. Dilanjutkan dengan proses *case folding* untuk menyama ratakan bentuk mengubah seluruh bentuk karakter menjadi huruf kecil dengan tujuan untuk memperkecil variasi kata karena bentuk kapitalisasi huruf. Proses berikutnya adalah *normalization* yang dilakukan untuk mengubah bentuk kata tidak baku atau kesalahan penulisan kata atau slang ke bentuk baku yang sesuai. Selanjutnya proses *tokenizing* dengan tujuan mengubah bentuk kalimat menjadi bentuk pecahan kata atau token. Proses selanjutnya yaitu *stopword removal* dilakukan untuk menghapus kata yang tidak relevan dalam proses analisis sentimen. Penghapusan dapat berupa kata yang tidak memiliki arti pada kalimat, kata depan, atau kata penghubung. Proses terakhir dalam *text preprocessing* adalah *stemming* dengan tujuan menemukan kata dasar dari *dataset* dengan menghilangkan awalan, sisipan, dan akhiran dalam sebuah kata.

3.4. *Model*

Pada tahapan model data hasil modifikasi akan dilakukan pelabelan menggunakan teknik *lexicon-based*, bertujuan untuk mengidentifikasi label sentimen pada setiap data sebelum melanjutkan ke tahap pemodelan menggunakan algoritma *machine learning*. Langkah selanjutnya adalah mengubah data dengan menggunakan teknik TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk mengubah teks ke dalam bentuk vektor yang bisa dimengerti oleh algoritma *Support Vector Machine*. Metode ini memberikan nilai pada kata-kata sesuai dengan seberapa sering mereka muncul dalam dokumen serta korpus, sehingga kata-kata yang lebih signifikan akan memiliki nilai yang lebih besar. Setelah itu, kumpulan data akan dipecah menjadi data latih dan data uji. Data latih dipergunakan untuk melatih model terkait data yang ada. Sedangkan data uji nantinya akan digunakan untuk menilai seberapa efektif model itu dalam mengategorikan sentimen dari ulasan. Pada penelitian ini menggunakan rasio pembagian data 90:10, 80:20, 70:30, dan 60:40. Setelah proses pengolahan data, langkah selanjutnya adalah pengembangan model analisis sentimen dengan menggunakan algoritma *Support Vector Machine*, yang sangat berguna untuk mengklasifikasikan data dengan dimensi tinggi, seperti halnya teks. Di tahap ini, model SVM dipersiapkan untuk mengidentifikasi sentimen dengan cara berdasarkan fitur numerik yang dihasilkan dari metode TF-IDF. Data pelatihan yang sudah melalui proses tersebut digunakan untuk melatih model supaya mampu mendeteksi pola-pola sentimen seperti positif, negatif, atau netral.

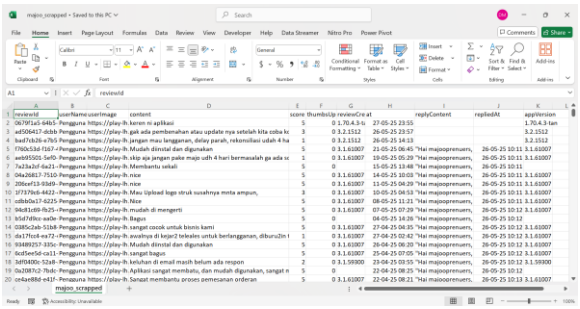
3.5. *Assess*

Tahap *Assess* merupakan langkah penutup, pada fase ini hasil dari evaluasi model atau pengujian dinamakan sebagai prediksi sentimen yang disajikan dalam bentuk *confusion matrix*. Kinerja suatu model dapat diukur dengan menunjukkan jumlah prediksi yang benar serta yang tidak tepat. Melalui *confusion matrix*, beberapa parameter turunan ditampilkan dalam *classification report* berisi *accuracy*, *precision*, *recall*, dan *F1-Score* dapat diketahui.

4. HASIL DAN PEMBAHASAN

4.1. *Sample*

Tahapan untuk teknik pengumpulan data yang diterapkan oleh peneliti adalah dengan menggunakan metode ekstraksi data dari dataset yang berasal dari ulasan pengguna Aplikasi Kasir Wirausaha Majoo di *Google Play Store*. Metode ini menggunakan bahasa pemrograman *Python* serta *library google-play-scraper* yang dijalankan di *Google Colaboratory* sebagai alat dalam penelitian. Penarikan data dibatasi waktu spesifik yaitu data ulasan dengan batas akhir 31 Mei 2025. Total tarikan data yang berhasil di ambil dari data ulasan sebanyak 5568 data ulasan, dengan rentang waktu 15 Juli 2018 hingga 26 Mei 2025. Selanjutnya data akan disimpan ke dalam format csv (*Comma Separated Values*).



Gambar 3. File csv hasil scraping data

4.2. *Explore*

Eksplorasi data perlu dilakukan untuk menggambarkan data yang berhasil ditarik pada tahapan *Sample* sebanyak 5568 data ulasan pengguna Aplikasi Kasir Wirausaha Majoo. Pada tahap ini beberapa hal yang dilakukan antara lain:

a. Memastikan kolom ulasan benar berbahasa Indonesia dengan *library langdetect*, menghasilkan 2933 data.
b. Seleksi fitur yang dibutuhkan, dan mengganti nama fitur sesuai kebutuhan. Fitur terpilih diantaranya "*user", "rating", "tanggal",* dan *"raw_ulasan".*
c. Penghapusan duplikasi data dan data dengan nilai kosong, kemudian mengambil data ulasan dengan panjang minimal karakter sebanyak tiga kata.

Dari eksplorasi data yang dilakukan didapatkan data bersih yang siap diolah sebanyak 2021 data. Data ulasan pertama terdapat pada tanggal 21 Juni 2019 hingga data terbaru pada 26 Mei 2025.

4.3. *Modify*

Pada tahap ini, serangkaian langkah dilakukan untuk merapikan dan menyesuaikan data demi meningkatkan kualitasnya untuk optimalisasi proses pemodelan. Melalui tahapan *text preprocessing* dengan langkah-langkah dalam mencakup *cleaning, case folding, normalization, tokenizing, stopwords removal,* dan *stemming.* Tabel 1 di bawah ini menunjukkan hasil dari proses *preprocessing* data.

Tabel 1. *Text preprocessing* pada tahapan *modify*

Proses	Hasil
raw_ulasan	Untuk pelayann training, oke dan sangt jelas Sangat membantu untuk yg masih awam
Cleaning	Untuk pelayann training oke dan sangt jelas Sangat membantu untuk yg masih awam
Case Folding	untuk pelayann training oke dan sangt jelas sangat membantu untuk yg masih awam
Normalization	untuk pelayan training ok dan sangat jelas sangat membantu untuk yang masih awam hilang atau terselesaikan

Proses	Hasil
Tokenizing	['untuk', 'pelayan', 'training', 'ok', 'dan', 'sangat', 'jelas', 'sangat', 'membantu', 'untuk', 'yang', 'masih', 'awam']
Stopword Removal	['pelayan', 'training', 'membantu', 'awam']
Stemming	['layan', 'training', 'bantu', 'awam']

4.4. *Model*

Pemodelan dimulai dengan melabeli data menggunakan pendekatan *lexicon-based*. Pada proses ini membutuhkan kamus *lexicon* positif dengan 1182 positif dan 2402 negatif. Kamus ini nantinya akan dibandingkan dengan data yang ada, jika kata yang muncul dalam data terdapat dalam kamus InSet positif nilai polaritas akan bertambah +1. Sedangkan perhitungan akan berkurang -1 jika kata terdapat pada kamus InSet negatif. Akumulasi jumlah perhitungan dengan hasil lebih dari 0 (nol) akan dikelompokkan menjadi label positif. Kemudian untuk jumlah perhitungan dengan hasil kurang dari 0 (nol) akan dikelompokkan menjadi label negatif. Sedangkan jumlah perhitungan sama dengan 0 (nol) akan masuk kelompok netral.

Tabel 2. Distribusi sentimen ulasan

Label Kelas	Distribusi	Jumlah Data
Positif	52,9%	1064
Negatif	27,8%	560
Netral	19,3%	387

Pada penelitian ini, seperti hasil dari pelabelan data berbasis *lexicon* dapat dilihat pada Tabel 2. Dari hasil menunjukkan mayoritas ulasan terklasifikasi menjadi ulasan positif. Kehadiran kelas netral dalam pengelompokan informasi dapat disebabkan oleh jumlah penilaian negatif yang sama dengan jumlah penilaian positif, sehingga total nilainya menjadi 0. Di samping itu, kelas netral juga dapat muncul karena kata-kata yang tidak dikenali dalam kamus InSet *lexicon*.

Gambar 4. *World cloud* label kelas data ulasan

Visualisasi *word cloud* pada Gambar 4 menunjukkan kata yang sering muncul menjadi topik utama pada setiap label kelas sentimen. Pada ulasan positif menunjukkan topik utama dengan kata “terima kasih”, “mudah”, “bagus”, “bantu”, “usaha”, “fitur”, “sederhana”, “trainer”, “paham”, “ramah”, dan “sesuai” menunjukkan tanggapan pelanggan dalam hal kepuasan. Pada kata yang sering muncul dalam visualisasi *word cloud* kelas negatif diantaranya “ribet”, “error”, “kecewa”, “lambat”, “susah”, “bug”, “stok”, “daftar”, “langgan”, dan “harga”. Munculnya istilah yang memiliki makna negatif ini dikenali sebagai bentuk ungkapan dari persoalan yang dicemooh dan dikeluhkan terkait pemakaian aplikasi. Kata-kata pada label kelas netral memberikan informasi dengan penilaian yang objektif. Mereka tidak menunjukkan pendapat yang sangat positif atau negatif karena memiliki nilai polaritas 0.

Proses penentuan bobot istilah menggunakan metode TF-IDF, di mana setiap kata diberikan nilai untuk mengubah data menjadi bentuk vektor. Ini membantu menentukan pentingnya suatu kata dalam dokumen berdasarkan frekuensinya. Kombinasi TF dan IDF menghasilkan hasil terbaik. Dalam *Python*, proses ini dapat dilakukan dengan pustaka *scikit-learn*, yang menyediakan fungsi *CountVectorizer* untuk menghitung Frekuensi Istilah dan *TfidfTransformer* untuk menghitung Frekuensi Dokumen Invers.

Pemanfaatan SVM dalam analisis sentimen untuk menetapkan garis pemisah yang paling efektif antara kategori ulasan yang memiliki sentimen positif, negatif, dan netral. Data yang sudah di ekstraksi fitur menggunakan TF-IDF akan dibagi menjadi beberapa rasio data latih dan data uji. Terdapat empat rasio pembagian, yaitu 90:10, 80:20, 70:30, dan 60:40 seperti pada Tabel 3. Kode untuk memulai kernel SVM adalah model = SVC(C=1.0, kernel='linear', class_weight=’balanced’, random_state=42). Penggunaan kernel linier dipandang tepat untuk data teks yang sering kali sudah memiliki banyak dimensi setelah proses vektorisasi TF-IDF. Kernel ini berusaha menemukan satu *hyperplane* linier untuk memisahkan kategori kelas. Selanjutnya penggunaan *class_weight=’balanced’* digunakan untuk menangani data teks yang sering kali memiliki distribusi sentimen yang tidak seimbang. Data latih dan uji yang digunakan dalam pemodelan diambil dari kolom *preprocessed_text* dan *sentiment*.

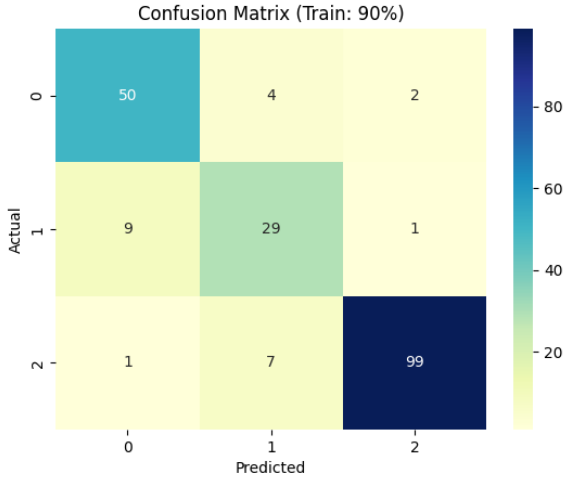
4.5. *Assess*

Pada tahap *assess* melakukan evaluasi kinerja model SVM menggunakan *confusion matrix* melalui metrik *accuracy, precision, recall,* dan *F1-Score* untuk mengukur efektivitas model dalam melakukan klasifikasi data.

Tabel 3. Perbandingan hasil akurasi rasio model

Rasio Data	Jumlah Data Latih	Jumlah Data Uji	*Accuracy*
90 : 10	1809	202	88,1%
80 : 20	1608	403	86,6%
70 : 30	1407	604	86,0%
60 : 40	1206	805	85,7%

Pada tabel 3 menunjukkan rasio pembagian data yang digunakan dan hasil *accuracy* dari klasifikasi yang sudah dilakukan oleh model SVM. Hasil pemodelan menunjukkan hasil *accuracy* paling tinggi pada rasio pembagian data 90:10. Pada rasio data latih ini menunjukkan performa yang baik dalam mengklasifikasikan sentimen pengguna.



Gambar 5. *Confusion Matrix* rasio 90:10

```
=== 📊 Evaluasi Model - Rasio Train 90% ===
✅ Akurasi: 0.8812
📄 Classification Report:
              precision    recall  f1-score   support

    negative       0.83      0.89      0.86        56
     neutral       0.72      0.74      0.73        39
    positive       0.97      0.93      0.95       107

    accuracy                           0.88       202
   macro avg       0.84      0.85      0.85       202
weighted avg       0.89      0.88      0.88       202
```

Gambar 6. *Classification Report* rasio 90:10

Dari gambar 5 menunjukkan tampilan *confusion matrix* dari rasio dengan hasil akurasi terbaik. Pada kelas negatif, model berhasil mengenali 50 dari 56 data dengan *recall* sebesar 89% dan *F1-score* 86%, menandakan kemampuannya dalam mendeteksi keluhan secara efektif. Sementara itu, kelas netral memiliki akurasi yang lebih rendah, dengan 29 dari 39 ulasan berhasil diklasifikasikan dengan benar. *F1-score* sebesar 73% mengindikasikan bahwa model masih kesulitan membedakan ulasan netral, kemungkinan akibat jumlah data yang lebih sedikit dan gaya bahasa yang ambigu. Di sisi lain, model sangat unggul dalam mengenali ulasan positif, dengan 99 dari 107 data diklasifikasikan secara tepat, dengan hasil *precision* 97% dan *F1-score* 95%. Secara keseluruhan, *accuracy* model mencapai 88,1%, yang menunjukkan performa tinggi dalam klasifikasi tiga kelas sentimen secara seimbang.

Dari tabel 3 juga didapatkan hasil bahwa semakin besar proporsi data latih, semakin baik performa model. Hal ini dikarenakan performa menurun secara bertahap pada rasio 80:20, 70:30, hingga 60:40. Penurunan ini menandakan berkurangnya kemampuan model dalam mengklasifikasikan ketiga kelas sentimen secara seimbang, terutama sentimen netral yang cenderung lebih sulit dikenali dengan jumlah data latih yang terbatas. Dari sisi *weighted average*, rasio 90:10 kembali unggul dengan *F1-score* 88,2%, menunjukkan performa stabil terhadap distribusi data yang tidak seimbang. Temuan ini mengindikasikan bahwa rasio 90:10 merupakan konfigurasi paling optimal dalam penelitian, karena mampu menjaga akurasi dan keseimbangan klasifikasi sentimen secara menyeluruh.

5. KESIMPULAN DAN SARAN

Berdasarkan analisis sentimen terhadap ulasan pengguna Aplikasi Kasir Wirausaha Majoo menggunakan metode *Support Vector Machine* (SVM) dengan pendekatan SEMMA (*Sample*, *Explore*, *Modify*, *Model*, *Assess*), penelitian ini berhasil mengumpulkan 5.568 data ulasan, yang setelah melalui proses pembersihan dan penghapusan duplikasi menghasilkan 2.021 data valid, dan 2.011 data bersih setelah tahap *text preprocessing*. Proses pelabelan sentimen berbasis *lexicon* menghasilkan 1.064 sentimen positif, 560 negatif, dan 387 netral. Visualisasi *word cloud* menunjukkan kata dominan pada sentimen positif seperti terima kasih, mudah, bantu, dan fitur, sementara sentimen negatif didominasi kata error, ribet, kecewa, dan stok. Model SVM menunjukkan performa terbaik pada rasio data latih 90:10 dengan akurasi 88,1%, *precision* 88,5%, *recall* 88,1%, dan *F1-score* 88,2%. Sentimen positif menjadi kelas yang paling mudah dikenali dengan *F1-score* tertinggi mencapai 95%. Hasil ini menegaskan bahwa model SVM bekerja secara presisi dan stabil, serta performanya meningkat seiring bertambahnya data pelatihan, menunjukkan pentingnya jumlah data latih dalam meningkatkan kemampuan klasifikasi sentimen secara merata. Dari keterbatasan penelitian, disarankan untuk mengumpulkan data ulasan dalam jumlah yang lebih besar dan jangka waktu yang lebih panjang guna meningkatkan akurasi dan generalisasi model. Selain itu, penelitian selanjutnya dapat mencoba algoritma lain seperti *Naive Bayes*, *Logistic Regression*, atau model *deep learning* lainnya untuk membandingkan performa dengan SVM. Hal ini dapat

memberikan gambaran yang lebih komprehensif terhadap pemilihan model terbaik dalam analisis sentimen.

DAFTAR PUSTAKA

[1] B. Mustopo, "Transformasi Digital Jadi Kunci Resiliensi UMKM," KEMENKOPUKM, hal. 1–28, 2022. [Daring]. Tersedia pada: https://jdih.kominfo.go.id/monografi_hukum/monografi/t/majalah/34

[2] S. Ayudiana, "Kemenkop UKM: 25,5 juta UMKM telah 'go digital,'" ANTARA, Jakarta, 14 Oktober 2024. [Daring]. Tersedia pada: https://www.antaranews.com/berita/4397157/kemenkop-ukm-255-juta-umkm-telah-go-digital

[3] S. N. Okofu, C. Asuai, O. Okumoku-evroro, dan A. Maureen, "Development of an Enhanced Point of Sales System for Retail Business in Developing Countries," J. Behav. Informatics, Digit. Humanit. Dev., vol. 11, no. 5, hal. 1–24, 2025, doi: 10.22624/AIMS/BHI/V11N1P1.

[4] D. Purnamasari et al., Pengantar Metode Analisis Sentimen. Depok: Penerbit Gunadarma, 2023.

[5] K. F. Ardhi, "Sentiment Analysis of Smartphone Accounting Application Users," J. Appl. Account. Tax., vol. 6, no. 2, hal. 161–174, 2021, doi: 10.30871/jaat.v6i2.3227.

[6] P. Sudhakaran dan M. Jaiganesh, "Sentiment Analysis Based Product Selection for Enhancing E-Commerce," Int. J. Innov. Technol. Explor. Eng., vol. 9, no. 3S, hal. 389–394, 2020, doi: 10.35940/ijitee.C1083.0193S20.

[7] J. A. Marpaung, M. Devega, dan Yuhelmi, "Analisis Sentimen Kepuasan Pengguna Aplikasi BCA Mobile Menggunakan Metode Naïve Bayes Dan Support Vector Machine (SVM)," Prosiding-Seminar Nas. Teknol. Inf. Ilmu Komput., vol. 3, no. 1, hal. 460–472, 2024.

[8] D. A. Putra dan E. R. Nababan, "Comparison of Sentiment for Midi Kriing and Alfagift Apps Using SVM with TF-IDF Weighting Diva," Data Sci. Inf. Technol. Data Anal., vol. 5, no. 1, hal. 84–90, 2025.

[9] B. Liu, Sentiment Analysis: Mining Opinions, Sentiments, and Emotions, Second Edition, no. May. 2012. doi: 10.1017/9781108639286.

[10] J. Han, M. Kamber, dan J. Pei, Data Mining: Concepts and Techniques, Third. Morgan Kaufmann Publishers, 2012.

[11] A. A. Baskara, N. M. Piranti, dan M. F. Romdendine, "FRAMEWORK DATA MINING : SEBUAH SURVEI," JATI (Jurnal Mhs. Tek. Inform., vol. 9, no. 3, hal. 4886–4895, 2025.

[12] W. S. Hoar, A. Zubair, dan L. Muflikhah, "Analisis sentimen kebijakan masuk sekolah pukul lima pagi menggunakan algoritma Naïve Bayes," J. Inf. Syst. Appl. Dev., vol. 2, no. 1, hal. 20–30, 2024, doi: 10.26905/jisad.v2i1.10935.

[13] R. B. Hadiprakoso, H. Setiawan, R. N. Yasa, dan Girinoto, "Text Preprocessing for Optimal Accuracy in Indonesian Sentiment Analysis Using a Deep Learning Model with Word Embedding," AIP Conf. Proc., vol. 2680, no. 1–8, 2022, doi: 10.1063/5.0126116.

[14] J. Sun, "Support Vector Machines and Kernel Methods," Suppl. Notes CSCI5525 Mach. Learn. Anal. Methods Support, hal. 1–22, 2024.

[15] M. Junus dan D. F. Abdillah, Penerapan Metode Support Vector Machine Untuk Klasifikasi Kualitas Air Minum Berbasis IOT, 1 ed., no. March. Malang: PT Penerbit Naga Pustaka Redaksi, 2025.

[16] S. Sathyanarayanan dan B. R. Tantri, "Confusion Matrix-Based Performance Evaluation Metrics," African J. Biomed. Res., vol. 27, no. 4, hal. 4023–4031, 2024, doi: 10.53555/AJBR.v27i4S.4345.