ANALISIS SENTIMEN PENGGUNA TERHADAP APLIKASI MYPERTAMINA MENGGUNAKAN METODE SVM (STUDI KASUS: BBM SOLAR)

Novita Ramadhini, Nabila Ramadhani, Adiaz Salsabilah Putri, Fathoni, Ali Ibrahim
Sistem Informasi, Universitas Sriwijaya
Jl. Raya Palembang-Prabumulih, KM 32 Indralaya,
Kabupaten Ogan Ilir, Sumatera Selatan, Indonesia
novitaramadhini0@gmail.com

ABSTRAK

Implementasi distribusi BBM Solar bersubsidi melalui aplikasi MyPertamina masih menghadapi berbagai tantangan, seperti rendahnya literasi digital masyarakat serta ketidaksesuaian antara regulasi dan praktik di lapangan. Permasalahan ini tercermin dalam ulasan pengguna yang menunjukkan adanya kesenjangan antara tujuan kebijakan dan pengalaman nyata. Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi MyPertamina dalam konteks distribusi BBM Solar bersubsidi. Metode yang digunakan adalah klasifikasi sentimen berbasis machine learning dengan algoritma Support Vector Machine (SVM). Data penelitian berupa 1.103 ulasan pengguna yang diperoleh melalui teknik scraping dari Google Play Store. Data kemudian diproses melalui tahapan preprocessing yang meliputi cleansing, case folding, tokenization, stopword removal, normalization, dan stemming, serta direpresentasikan menggunakan metode TF-IDF. Hasil penelitian menunjukkan bahwa model SVM menghasilkan akurasi sebesar 83,71%. Model memiliki kinerja sangat baik dalam mengklasifikasikan sentimen negatif dengan nilai precision, recall, dan F1-score sebesar 0,91, namun kurang optimal dalam mengklasifikasikan sentimen positif dengan nilai sebesar 0,40. Temuan ini mengindikasikan dominasi sentimen negatif serta adanya ketidakseimbangan data yang memengaruhi performa model.

Kata kunci : *Analisis Sentimen, Support Vector Machine, MyPertamina, BBM Solar*

1. PENDAHULUAN

Distribusi BBM bersubsidi sering tidak tepat sasaran karena bergantung pada komoditas, sehingga pihak yang tidak memenuhi kriteria sering menikmati manfaatnya. Ini menunjukkan bahwa kebijakan perlu direformasi dan sistem pengawasan yang lebih baik dibangun [1]. Pemerintah mulai mendorong penggunaan teknologi digital dalam distribusi BBM bersubsidi sebagai cara untuk meningkatkan transparansi, keakuratan data subsidi, dan kesadaran mengenai penggunaan bakar-bakar bersubsidi. Salah satu metode penerapan kebijakan tersebut adalah implementasi aplikasi MyPertamina, yang digunakan sebagai sistem pendaftaran kendaraan dan alat untuk mengontrol pembelian BBM bersubsidi, termasuk jenis BBM Solar, yang penggunanya telah disesuaikan menggunakan mekanisme ini sehingga distribusinya dapat lebih terkontrol dan tepat sasaran [2].

Aplikasi MyPertamina dikembangkan sebagai salah satu alat digital untuk meningkatkan transparansi serta menekan potensi penyalahgunaan BBM bersubsidi. Namun, implementasinya di lapangan masih menghadapi berbagai tantangan. Beberapa studi menunjukkan bahwa sistem distribusi BBM tipe Solar masih belum beroperasi secara optimal, yang berarti potensi penggunaan distribusi masih mungkin terjadi pada tingkat operasional di lapangan [3]. Dalam konteks lain, penggunaan aplikasi MyPertamina dapat membantu meningkatkan transparansi dalam subsidi energi. Dalam praktiknya, masih ada masalah seperti kurangnya literasi digital di kalangan masyarakat umum dan penolakan untuk menggunakan sistem digital dalam proses pembelian BBM bersubsidi [2]. Situasi yang disebutkan di atas menyoroti kemungkinan ketidaksesuaian antara mekanisme pembatasan pembelian BBM yang telah diterapkan dengan praktik isi ulang di SPBU, khususnya pada BBM tipe Solar yang telah digunakan melalui sistem digital. Akibatnya, perlu dilakukan penelitian lanjutan untuk menentukan apakah implementasi aplikasi MyPertamina telah memenuhi syarat untuk mendistribusikan BBM Solar sesuai dengan peraturan yang ditetapkan. Salah satu cara yang dapat dilakukan, misalnya, dengan menganalisis tanggapan dan pengalaman pengguna yang menggunakan aplikasi MyPertamina.

Metode *machine learning* telah banyak digunakan dalam penelitian sebelumnya untuk menganalisis sentimen ulasan aplikasi di Google Play Store. Salah satu penelitian menggunakan algoritma Naïve Bayes Classifier (NBC) dan menunjukkan bahwa ulasan pengguna aplikasi MyPertamina cenderung bersentimen negatif dengan akurasi sebesar 87% [4], sementara penelitian lain menggunakan Word Embedding FastText dan algoritma K-Nearest Neighbor (K-NN) yang juga menunjukkan dominasi sentimen negatif dengan akurasi sebesar 73% [5]. Meskipun demikian, penelitian tersebut masih memiliki keterbatasan, sehingga diperlukan pengembangan metode lain seperti SVM atau Random Forest serta optimasi parameter untuk meningkatkan performa klasifikasi sentimen [5]. Kondisi ini menunjukkan adanya *research gap*, karena sebagian besar penelitian sebelumnya hanya berfokus pada

pengklasifikasian sentimen pengguna secara umum terhadap aplikasi MyPertamina dan belum memfokuskan pada aspek yang berkaitan dengan fungsi utama aplikasi. Padahal, aplikasi MyPertamina memiliki peran penting sebagai sarana pembatasan pembelian BBM bersubsidi, khususnya BBM jenis Solar, sehingga diperlukan analisis sentimen yang lebih spesifik pada ulasan pengguna terkait pembatasan BBM Solar untuk mengetahui apakah implementasi aplikasi telah berjalan sesuai dengan tujuan pembatasan yang telah ditetapkan.

Analisis sentimen pada ulasan aplikasi dapat digunakan untuk mengetahui persepsi serta kritik pengguna terhadap kualitas layanan suatu aplikasi. Untuk menilai pengalaman pengguna, ulasan pengguna di Google Play Store sering digunakan sebagai sumber data untuk menilai pengalaman pengguna, karena ulasan tersebut mencerminkan penilaian langsung terhadap fungsi dan kinerja aplikasi yang digunakan [6]. Dalam konteks aplikasi MyPertamina, ulasan pengguna tidak hanya memuat penilaian terhadap performa aplikasi, tetapi juga mengandung ulasan terkait kebijakan pembatasan pembelian BBM Solar yang diimplementasikan melalui aplikasi tersebut. Salah satu metode text mining yang dimanfaatkan untuk menganalisis pendapat pengguna dalam bentuk teks serta mengklasifikasikannya ke dalam kategori sentimen positif dan negatif [7]. Oleh karena itu, penelitian ini memanfaatkan metode *Support Vector Machine* karena memiliki keunggulan dalam melakukan klasifikasi teks berskala tinggi dan telah banyak digunakan dalam penelitian yang melibatkan analisis sentimen terhadap ulasan pengguna aplikasi [8].

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi MyPertamina dengan fokus pada ulasan yang berkaitan dengan penggunaan BBM Solar. Peneliti mengambil data ulasan dari Google Play Store, kemudian memfilternya untuk memperoleh komentar pengguna yang secara spesifik membahas pembatasan serta penggunaan BBM Solar melalui aplikasi tersebut. Penelitian ini melakukan analisis melalui beberapa tahapan, yaitu pengumpulan data ulasan, proses *preprocessing* teks, pembobotan kata, juga proses klasifikasi sentimen menggunakan metode SVM [9]. Hasil klasifikasi sentimen diharapkan dapat menunjukkan seberapa efektif aplikasi MyPertamina menurut pandangan pengguna dalam mengatur pembatasan BBM Solar, termasuk kritik pengguna terhadap ketidaksesuaian antara aturan yang diterapkan dalam aplikasi dengan praktik yang terjadi di lapangan. Temuan dari penelitian ini diharapkan dapat membantu pengelola layanan mengevaluasi cara meningkatkan kualitas sistem serta memperkuat pengelolaan hubungan dengan pengguna yang sejalan dengan konsep *Customer Relationship Management* (CRM) [10].

2. TINJAUAN PUSTAKA

Aplikasi MyPertamina merupakan salah satu inisiatif digitalisasi yang dilakukan pemerintah Indonesia untuk meningkatkan transparansi dan akuntabilitas distribusi BBM bersubsidi kepada masyarakat. Penelitian [1] mengkaji efektivitas penggunaan aplikasi MyPertamina dalam menekan penyalahgunaan distribusi BBM bersubsidi melalui metode *literature review* terhadap berbagai studi dan laporan implementasi MyPertamina pada periode 2021–2025. Penelitian sebelumnya yang secara khusus mengkaji penggunaan Pertalite menunjukkan bahwa pada tahun 2024 sekitar 93,9% transaksi telah dilakukan secara digital, dengan sekitar 5,5 juta transaksi tercatat dalam program Subsidi Tepat. Temuan ini membuktikan bahwa digitalisasi melalui aplikasi MyPertamina mampu meningkatkan transparansi distribusi BBM bersubsidi, meskipun banyak tantangan masih menghalangi pelaksanaannya seperti literasi digital masyarakat yang rendah, keterbatasan infrastruktur teknologi, serta potensi penyalahgunaan mekanisme kode QR. Namun, penelitian tersebut hanya berfokus pada Pertalite, sehingga masih diperlukan kajian lebih lanjut pada jenis bahan bakar lain, seperti solar.

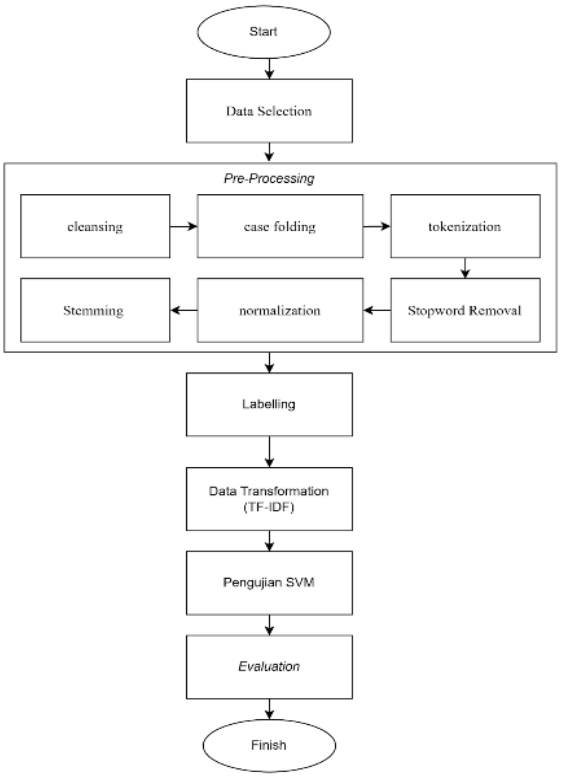
Penelitian [11] menerapkan metode *Naïve Bayes Classifier* (NBC) dengan TF-IDF pada 1.000 ulasan Google Play Store dan menghasilkan akurasi 82%, dengan temuan bahwa sentimen negatif mendominasi sebesar 821 ulasan, mencerminkan besarnya ketidakpuasan pengguna terhadap kebijakan BBM bersubsidi. Penelitian [12] kemudian menerapkan algoritma Bi-LSTM pada 6.000 ulasan MyPertamina dan menghasilkan akurasi lebih tinggi sebesar 90%, membuktikan bahwa pendekatan *Deep Learning* mampu menangkap konteks ulasan pengguna secara lebih akurat. Penelitian [13] secara khusus juga menerapkan metode SVM dengan TF-IDF pada ulasan MyPertamina dan berhasil memperoleh akurasi sentimen sebesar 92% serta rata-rata akurasi berbasis aspek sebesar 96%, menunjukkan bahwa metode SVM lebih unggul dalam proses klasifikasi sentimen pada ulasan aplikasi MyPertamina dibandingkan NBC.

Support Vector Machine (SVM) merupakan algoritma *machine learning* yang berfungsi untuk menemukan garis batas terbaik (*hyperplane*) sebagai pemisah antara kelas positif (+1) dan negatif (−1) dalam ruang fitur berdimensi tinggi [13]. Penelitian oleh [14] membuktikan bahwa SVM menghasilkan akurasi 83% dalam klasifikasi sentimen, juga mencapai hasil yang lebih baik dibandingkan *Naïve Bayes* yang hanya memperoleh akurasi sebesar 75,5%. Penelitian [15] juga membuktikan bahwa SVM yang dikombinasikan dengan *Recursive Feature Elimination* (RFE) mampu meningkatkan akurasi klasifikasi sentimen secara signifikan dengan mereduksi fitur yang tidak relevan pada data teks.

Berdasarkan analisis penelitian sebelumnya, masih memiliki celah penelitian yang belum terjawab dalam literatur yang ada. Seluruh penelitian yang

mengkaji aplikasi MyPertamina, baik oleh [11] maupun [12] menganalisis ulasan pengguna secara umum tanpa memfokuskan kajian pada jenis BBM tertentu, khususnya BBM Solar yang memiliki karakteristik distribusi dan segmentasi pengguna yang berbeda dibandingkan BBM lainnya. Di samping itu, belum ada penelitian yang secara khusus menggunakan metode SVM untuk menganalisis sentimen ulasan pengguna MyPertamina dalam konteks BBM Solar. Oleh karena itu, fokus penelitian ini adalah melakukan analisis ulasan pengguna terhadap aplikasi MyPertamina yang tersedia di Google Play Store terkait distribusi BBM Solar, dengan tujuan untuk mengetahui persepsi pengguna terhadap penggunaan aplikasi tersebut. Data ulasan kemudian diolah menggunakan metode SVM, dengan pembobotan fitur menggunakan TF-IDF.

3. METODE PENELITIAN



Gambar 1. Alur Penelitian

Gambar 1 menggambarkan alur penelitian yang diawali dari proses pengumpulan data, dilanjutkan dengan proses *pre-processing* yang meliputi *cleansing*, *case folding*, *tokenization*, *normalization*, *stemming*, dan *stopword removal* untuk membersihkan serta menyiapkan data agar siap dianalisis. Setelah itu, data diberi label (*labelling*), kemudian ditransformasikan menggunakan metode TF-IDF, dilanjutkan dengan pengujian menggunakan algoritma SVM, hingga tahap akhir berupa evaluasi untuk mengetahui hasil kinerja model yang digunakan. Penjelasan lebih lanjut dari setiap tahapan tersebut adalah sebagai berikut.

3.1. *Data Selection*

Penelitian ini menggunakan data sekunder berupa ulasan pengguna aplikasi MyPertamina yang diperoleh dari Google Play Store [11]. Data ulasan dikumpulkan secara otomatis dari halaman aplikasi melalui teknik web scraping [5]. Ulasan tersebut dipilih dengan mempertimbangkan relevansi isi ulasan dengan subjek penelitian serta penggunaan seluruh layanan aplikasi, hal ini dilakukan untuk membuat data yang dikumpulkan lebih konsisten dan representatif [11]. Data yang digunakan dalam penelitian ini berjumlah 1.103 ulasan.

3.2. *Data Preprocessing*

Preprocessing pada data ulasan akan menjadi langkah berikutnya setelah melakukan tahapan *data selection*. Tujuan dari tahap ini adalah untuk membersihkan sekaligus mempersiapkan data untuk digunakan dalam analisis sentimen, sehingga kualitas data meningkat dan noise yang mempengaruhi hasil klasifikasi dapat diminimalkan. Proses *preprocessing* umumnya meliputi beberapa tahapan yaitu *cleansing*, *case folding*, *tokenization*, *stopword removal*, *normalization*, dan *stemming* [13].

a. *Cleansing*: proses awal dalam *preprocessing* dataset ini bertujuan untuk menghilangkan noise (karakter yang tidak relevan) dari data ulasan atau tweet. Hal ini dilakukan dengan menghapus tanda baca, simbol, tautan URL, hashtag (#), dan mention (@) untuk mengurangi gangguan dalam data.
b. *Case Folding*: langkah ini dilakukan untuk menghindari duplikasi data dan mempermudah proses analisis pada tahap berikutnya dengan menyeragamkan bentuk kata dengan mengubah semua huruf menjadi huruf kecil.
c. *Tokenization*: proses memecah teks menjadi kata-kata berdasarkan spasi sehingga data siap digunakan pada tahap analisis berikutnya.
d. *Stopword Removal*: tujuan dari tahap ini adalah membuat data lebih relevan untuk dianalisis dengan menghilangkan kata-kata seperti "dan", "yang", dan "ke".
e. *Normalization*: tahap ini digunakan untuk memperbaiki kata tidak baku pada ulasan MyPertamina (solar) di Play Store agar sesuai dengan bahasa yang benar, sehingga memudahkan proses analisis data.
f. *Stemming*: kata yang memiliki imbuhan diubah menjadi bentuk aslinya (*stemming*) sehingga analisis dapat lebih fokus pada makna dari setiap kata.

3.3. *Labelling*

Pada tahap ini, label akan diberikan kepada data yang telah dibersihkan dan diproses sebelumnya. Data teks secara keseluruhan (*full text*) akan dikategorikan ke dalam sentimen positif apabila mengandung makna positif, sedangkan teks yang mengandung makna negatif akan dikategorikan sebagai sentimen negatif.

Dalam proses pelabelan data *full text*, digunakan kamus leksikon (*lexicon dictionary*) sebagai acuan untuk meningkatkan akurasi pelabelan [16]. Kamus leksikon atau leksikon dapat diartikan sebagai kumpulan kosakata, kamus sederhana, atau daftar istilah yang disusun secara alfabetis beserta maknanya. Selain itu, leksikon juga mencakup satuan bahasa yang memiliki makna dan penggunaannya, serta kata kunci tertentu [17].

3.4. *Data Transformation*

Pada tahap transformasi, teks diubah menjadi bentuk yang siap diprose. Metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk merepresentasikan teks sebagai vektor berbobot [10]. TF menunjukkan frekuensi kata yang muncul dalam dokumen, sehingga mencerminkan tingkat relevansinya, sedangkan DF menghitung jumlah dokumen yang mengandung kata tersebut, yang menunjukkan seberapa sering digunakan [18].

3.5. Pengujian SVM

Penelitian ini melakukan tahap pengujian model klasifikasi menggunakan algoritma SVM berdasarkan data yang telah direpresentasikan dalam bentuk vektor fitur [19]. Data yang diperoleh dari proses pembobotan fitur dengan metode TF-IDF digunakan sebagai input untuk proses pelatihan dan pengujian model [20]. Pada proses pengujian, model yang telah dilatih mengklasifikasikan data uji sehingga menghasilkan prediksi kelas berdasarkan pola yang telah dipelajari. Tujuan dari proses ini adalah untuk mengetahui kemampuan model untuk melakukan klasifikasi data baru dengan benar.

3.6. *Evaluation*

Penelitian ini melakukan tahap evaluasi untuk menilai kinerja model klasifikasi. Untuk mengevaluasi model menggunakan *confusion matrix* serta beberapa metrik, yaitu *accuracy* untuk menilai ketepatan prediksi pada kelas positif, *recall* untuk melihat kemampuan model dalam mendeteksi data relevan, serta F1-*score* untuk mengukur keseimbangan antara *precision* dan *recall* [21]. Dengan membandingkan hasil prediksi model dengan label sebenarnya, *confusion matrix* digunakan untuk menghasilkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) sebagai dasar perhitungan [22].

4. HASIL DAN PEMBAHASAN

Penelitian ini memanfaatkan data ulasan pengguna aplikasi MyPertamina yang diperoleh dari Google Play Store melalui teknik *scraping*. Jumlah data yang akan digunakan pada penelitian ini akan difokuskan pada pengambilan ulasan yang membahas terkait distribusi BBM Solar. Data tersebut kemudian diseleksi dan disaring untuk memastikan relevansi dengan topik penelitian

4.1. *Data Selection*

Proses pemilihan data menjadi tahap pertama yang digunakan dalam penelitian, dengan tujuan untuk menyeleksi data yang relevan. Dataset yang digunakan berasal dari ulasan pengguna untuk aplikasi MyPertamina di Google Play Store yang dikumpulkan melalui proses *scraping*. Data yang dikumpulkan berjumlah 1.103 ulasan, dengan kriteria hanya mengambil ulasan yang mengandung pembahasan terkait BBM Solar. Tabel 1 di bawah ini menampilkan dataset yang diperoleh melalui teknik *scraping* tersebut.

Tabel 1. Dataset awal hasil *scraping data*

Rating	Ulasan
5	mypertamina sangat membantu supir di lintasan Sumatra mengisi BBM solar
1	sudah memperbaiki data kendaraan 2 kali tetap saja di tolak .. mau nya apa? beli solar saja di persulit
1	tolong saya pa Bu barcode saya di blokir terus. padahal udah ada BPKB di rumah. saya 3 bulan ga kerja ga bisa beli solar

4.2. *Data Preprocessing*

Tahap *preprocessing* digunakan untuk memproses data yang telah dikumpulkan guna meningkatkan kualitas data sebelum tahap analisis. Tahapan ini mencakup proses pembersihan dan persiapan data agar lebih terorganisir dan siap untuk diolah. Segala sesuatu dilakukan melalui Google Colab dan bahasa pemrograman Python.

a. *Cleansing*

Sebelum tahap *cleansing*, data masih mengandung sejumlah gangguan, seperti kata-kata yang tidak relevan, kesalahan penulisan, serta karakter yang tidak diperlukan sehingga dapat mempengaruhi hasil analisis. Proses cleansing dilakukan untuk menghilangkan gangguan tersebut dan memperbaiki struktur data agar menjadi lebih bersih serta siap digunakan pada tahap selanjutnya. Tabel 2 menampilkan hasil proses *cleansing*.

Tabel 2. Hasil *cleansing*

Sebelum *Cleansing*	Setelah *Cleansing*
sudah memperbaiki data kendaraan 2 kali tetap saja di tolak .. mau nya apa? beli solar saja di persulit	sudah memperbaiki data kendaraan kali tetap saja di tolak mau nya apa beli solar saja di persulit

b. *Case Folding*

Setiap kata dalam dataset diubah menjadi uruf kecil (*lowercase*) untuk menyeragamkan bentuk teks. Hasil dari tahap ini ditampilkan pada tabel 3.

Tabel 3. Hasil *case folding*

Sebelum *Case Folding*	Setelah *Case Folding*
sudah memperbaiki data kendaraan kali tetap saja di tolak mau nya apa beli solar saja di persulit apa? beli solar saja di persulit	sudah memperbaiki data kendaraan kali tetap saja di tolak mau nya apa beli solar saja di persulit

c. *Tokenization*
Pada tahap ini, teks dipisahkan menjadi bagian-bagian dasar berupa token, seperti kata, berdasarkan spasi maupun tanda baca, sehingga memudahkan proses analisis selanjutnya. Tabel 4 menunjukkan teks yang sudah terbagi menjadi potongan kata.

Tabel 4. Hasil *tokenization*

Sebelum *Tokenization*	**Setelah *Tokenization***
sudah memperbaiki data kendaraan kali tetap saja di tolak mau nya apa beli solar saja di persulit	['sudah', 'memperbaiki', 'data', 'kendaraan', 'kali', 'tetap', 'saja', 'di', 'tolak', 'mau', 'ya', 'apa', 'beli', 'solar', 'saja', 'di', 'persulit']

d. *Stopword Removal*
Pada tahapan ini, teks dilakukan pembersihan dengan menghapus kata-kata yang sering muncul tetapi tidak memiliki makna, sehingga hasil analisis menjadi lebih relevan. Tabel 5 menunjukkan hasil dari tahapan pembersihan teks dari kata-kata yang tidak memiliki makna.

Tabel 5. Hasil *stopword removal*

Sebelum *Stopword Removal*	**Setelah *Stopword Removal***
['sudah', 'memperbaiki', 'data', 'kendaraan', 'kali', 'tetap', 'saja', 'di', 'tolak', 'mau', 'ya', 'apa', 'beli', 'solar', 'saja', 'di', 'persulit']	['memperbaiki', 'data', 'kendaraan', 'kali', 'tolak', 'ya', 'beli', 'solar', 'persulit']

e. *Normalization*
Pada tahap ini, kata-kata yang tidak baku dalam teks diubah menjadi bentuk yang baku atau standar untuk meningkatkan konsistensi. Tabel 6 menunjukkan hasil dari tahapan ini.

Tabel 6. Hasil *normalization*

Sebelum *Normalization*	**Setelah *Normalization***
['memperbaiki', 'data', 'kendaraan', 'kali', 'tolak', 'ya', 'beli', 'solar', 'persulit']	['memperbaiki', 'data', 'kendaraan', 'kali', 'tolak', 'ya', 'beli', 'solar', 'persulit']

f. *Stemming*
Pada tahapan ini, kata-kata dalam teks yang memiliki imbuhan diubah menjadi bentuk dasarnya melalui proses *stemming*. Tabel 7 menunjukkan hasil dari tahapan ini.

Tabel 7. Hasil *stemming*

Sebelum *Stemming*	**Setelah *Stemming***
['memperbaiki', 'data', 'kendaraan', 'kali', 'tolak', 'ya', 'beli', 'solar', 'persulit']	'baik', 'data', 'kendara', 'kali', 'tolak', 'ya', 'beli', 'solar', 'sulit']

4.3. *Data Transformation*

Tahap transformasi data memasukkan data yang telah melalui tahap *preprocessing*. Tujuan transformasi ini adalah untuk mengubah format, struktur, atau nilai data agar lebih sesuai dengan metode yang digunakan [23].

a. *Labelling*
Pada tahap ini, seluruh data yang telah diproses sebelumnya dilabelkan untuk mengelompokkan ulasan ke dalam kategori sentimen positif dan negatif. Proses pelabelan dilakukan berdasarkan kata-kata yang mengandung kecenderungan sentimen tertentu, di mana kata dengan muatan positif dan negatif digunakan sebagai acuan. Hasil pelabelan dapat dilihat pada tabel 8.

Tabel 8. *Labelling*

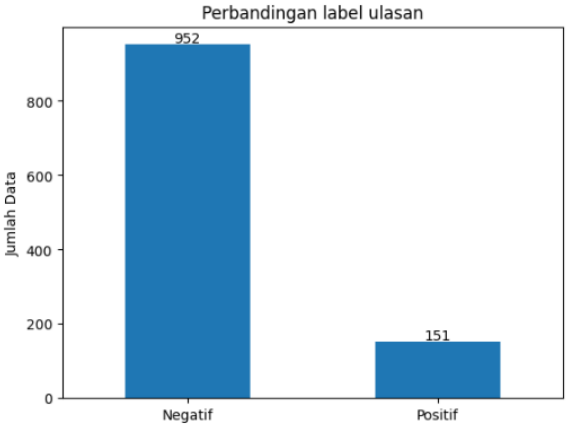
Positive Word	***Negative Word***
mypertamina sangat membantu supir di lintasan Sumatra mengisi BBM solar	tambah susah beli solar Dunia tambah ruwet

b. *Label Encoding*
Tahap selanjutnya adalah proses *label encoding*, yaitu mengubah label kategorikal hasil pelabelan ke dalam bentuk numerik agar sesuai dengan kebutuhan pemrosesan oleh algoritma *machine learning*. Setiap kategori sentimen direpresentasikan dalam bentuk nilai angka tertentu. Hasil tahap ini ditampilkan pada tabel 9.

Tabel 9. *Label encoding*

Label	***Label Encoding***
Positive	1
Negatif	0

Hasil dari proses *label encoding* menunjukkan bahwa terdapat 952 data memiliki sentimen negatif dan 151 data memiliki sentimen positif. Selanjutnya, distribusi sentimen tersebut disajikan dalam bentuk grafik, seperti yang ditunjukkan pada gambar 2.



Gambar 2. Grafik distribusi sentimen

4.4. Pengujian SVM

Pada tahap pengujian algoritma *Support Vector Machine* (SVM) diterapkan pada data ulasan pengguna yang telah melalui tahapan *preprocessing*. Data yang digunakan merupakan hasil representasi fitur

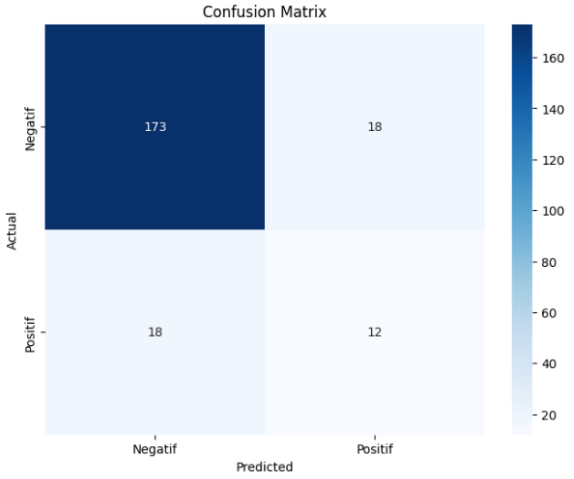
menggunakan metode TF-IDF, sedangkan label sentimen yang sebelumnya berbentuk kategori telah dikonversi ke dalam bentuk numerik melalui proses label encoding.

Selanjutnya, data digunakan untuk melatih model SVM untuk mengidentifikasi pola yang ditemukan di setiap ulasan. Data uji kemudian dikategorikan ke dalam dua kategori emosi, yaitu positif dan negatif, menggunakan model yang telah dilatih dalam dua kategori sentimen, yaitu positif dan negatif. Selanjutnya, performa model dievaluasi menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *recall*, dan F1-*score* guna menilai kemampuan model dalam melakukan klasifikasi.

```
Confusion Matrix:
[[173  18]
 [ 18  12]]

Classification Report:
              precision    recall  f1-score   support

     Negatif       0.91      0.91      0.91       191
     Positif       0.40      0.40      0.40        30

    accuracy                           0.84       221
   macro avg       0.65      0.65      0.65       221
weighted avg       0.84      0.84      0.84       221


Accuracy: 0.8371040723981901
```

Gambar 3. Hasil program klasifikasi SVM

Gambar tersebut menampilkan hasil evaluasi model berupa *classification report*, nilai akurasi, serta representasi *confusion matrix*. Berdasarkan hasil pengujian model SVM yang digunakan memperoleh tingkat akurasi sebesar 83,71%. Meskipun model memiliki tingkat akurasi yang cukup baik, terdapat perbedaan kinerja antar kelas. Model menunjukkan bahwa sentimen negatif lebih unggul dalam mengklasifikasikan sentimen negatif dibandingkan sentimen positif. Kondisi ini diduga disebabkan oleh adanya ketidakseimbangan dalam jumlah data di masing-masing kelas.



Gambar 4. *Confusion matrix*

4.5. *Evaluation*

Evaluasi dilakukan untuk mengetahui seberapa baik kinerja model SVM dalam mengkategorikan sentimen pengguna aplikasi MyPertamina tentang distribusi BBM Solar. Proses evaluasi dilakukan menggunakan *confusion matrix* serta metrik evaluasi berupa *accuracy*, *precision*, *recall*, dan F1-*score*. Untuk membandingkan hasil prediksi model dengan label aktual yaitu dengan menggunakan *confusion matrix*, sehingga menghasilkan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) sebagai dasar perhitungan. *Accuracy* digunakan untuk mengukur ketepatan secara keseluruhan model dalam mengklasifikasikan data.

Hasil evaluasi SVM menunjukkan nilai akurasi sebesar 0,84, yang menunjukkan bahwa model secara umum dapat mengklasifikasikan data dengan baik. Namun, jika dianalisis lebih lanjut melalui metrik *precision*, *recall*, dan F1-*score*, tampak adanya perbedaan performa antar kelas yang cukup signifikan.

Model menunjukkan kinerja yang sangat baik pada kelas negatif dengan nilai *precision*, *recall*, dan F1-*score* masing-masing sebesar 0,91. Ini menunjukkan bahwa model dapat secara konsisten dan akurat mengidentifikasi dan mengklasifikasikan data sentimen negatif. Sebaliknya, pada kelas positif, menunjukkan bahwa model masih mengalami kesulitan dalam mengidentifikasi dan memprediksi data sentimen positif yang tepat, seperti yang ditunjukkan oleh nilai *precision*, *recall*, dan F1-*score* yang hanya 0,40 pada kelas positif. Rendahnya nilai *recall* pada kelas ini mengindikasikan bahwa sebagian besar data positif tidak berhasil terdeteksi oleh model (*false negative* yang tinggi).

Ketimpangan performa ini mengindikasikan adanya potensi bias model terhadap kelas tertentu, yang mungkin terjadi karena *imbalanced dataset* atau data distribusi yang tidak seimbang, jumlah data pada kelas negatif lebih banyak daripada kelas positif. Kondisi ini menghasilkan model yang sudah terbiasa mengenali pola pada kelas yang jumlah datanya lebih dominan, sehingga kemampuan dalam mendeteksi kelas dengan jumlah data yang lebih sedikit menjadi kurang optimal.

5. KESIMPULAN DAN SARAN

Hasil penelitian ini menunjukkan bahwa klasifikasi sentimen pada ulasan pengguna aplikasi MyPertamina yang berfokus pada BBM Solar dapat dilakukan secara cukup efektif dengan menggunakan metode *Support Vector Machine* (SVM). Berdasarkan pengolahan terhadap 1.103 data ulasan yang diperoleh dari Google Play Store, model yang dikembangkan mampu mencapai akurasi sebesar 83,71%. Menurut evaluasi lanjutan *precision*, *recall*, dan F1-*score* masing-masing sebesar 0,91 menunjukkan kemampuan model yang luar biasa untuk mengklasifikasikan sentimen negatif. Sebaliknya, pada kelas sentimen positif, kinerja model masih relatif rendah dengan nilai *precision*, *recall*, dan F1-

score sebesar 0,40. Hal ini menunjukkan bahwa model menemukan ulasan dengan sentimen negatif lebih baik daripada ulasan dengan sentimen positif. Perbedaan kinerja tersebut mengindikasikan bahwa kemampuan model dalam menangkap pola pada setiap kelas belum merata, sehingga masih terdapat kesalahan klasifikasi, terutama pada data dengan sentimen positif. Untuk penelitian selanjutnya, disarankan untuk melakukan penyeimbangan jumlah data pada masing-masing kelas, melakukan optimasi parameter model secara lebih mendalam, serta mempertimbangkan penggunaan metode atau fitur tambahan yang dapat meningkatkan kemampuan model dalam memahami konteks teks, seperti penggunaan teknik representasi kata yang lebih kompleks. Dengan demikian, diharapkan performa model dapat meningkat dan memberikan hasil klasifikasi yang lebih optimal.

DAFTAR PUSTAKA

[1] R. D. Triwibowo and S. Pramono, "Kajian implementasi pengawasan subsidi bahan bakar minyak (BBM) di Indonesia: Perspektif teori Daniel A. Mazmanian dan Paul A. Sabatier," *J. Nas. Pengelolaan Energi MigasZoom*, vol. 7, no. 1, pp. 61–72, 2025.

[2] J. M. Putra and M. Menorizah, "An Analysis of the Use of the MyPertamina Application in Reducing the Misuse of Government-Subsidised Pertalite (A Literature Review)," *JERIT J. Educ. Res. Innov. Technol.*, vol. 2, no. 2, pp. 62–72, 2025.

[3] P. Hariqah and R. Yusran, "Supervision of the Distribution of Subsidized Fuel Oil (BBM) Type Solar (Biosolar) in Padang City," *Santhet (Jurnal Sej. Pendidik. Dan Humaniora)*, vol. 8, no. 1, pp. 711–720, 2024.

[4] B. Z. Ramadhan, R. I. Adam, and I. Maulana, "Analisis sentimen ulasan pada aplikasi e-commerce dengan menggunakan algoritma naïve bayes," *J. Appl. Informatics Comput.*, vol. 6, no. 2, pp. 220–225, 2022.

[5] N. Sepriadi, E. Budianita, and M. Fikry, "Analisis Sentimen Review Aplikasi Mypertamina Menggunakan Word Embedding Fasttext Dan Algoritma K-Nearest Neighbor," *Inf. (Jurnal Inform. dan Sist. Informasi)*, vol. 15, no. 1, pp. 91–109, 2023.

[6] M. F. Rahmatullah, P. L. L. Belluano, and H. Darwis, "Analisis Sentimen Review Aplikasi di Google Play Store Menggunakan Random Forest," *LINIER Lit. Inform. dan Komput.*, vol. 2, no. 3, pp. 380–389, 2025.

[7] R. A. Rahman, V. H. Pranatawijaya, and N. N. K. Sari, "Analisis Sentimen Berbasis Aspek pada Ulasan Aplikasi Gojek," *KONSTELASI Konvergensi Teknol. dan Sist. Inf.*, vol. 4, no. 1, pp. 70–82, 2024.

[8] S. D. Fitri, D. Lestari, R. R. Bintana, R. Aryani, M. Ilhami, and Y. Noverina, "Implementasi Model Support Vector Machine Dalam Analisa Sentimen Masyarakat Mengenai Kebijakan Penerapan Aplikasi Mypertamina," *Bridg. J. Publ. Sist. Inf. dan Telekomun.*, vol. 2, no. 2, pp. 176–193, 2024.

[9] E. Hokijuliandy, H. Napitupulu, and F. Firdaniza, "Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square," *SisInfo J. Sist. Inf. dan Inform.*, vol. 5, no. 2, pp. 40–49, 2023.

[10] A. W. S. Angga, H. S. Hamzah, and R. D. Rohman, "Analisis Sentimen Layanan Perwalian Mahasiswa UMSIDA Menggunakan Metode Support Vector Machine (SVM)," *JSAI (Journal Sci. Appl. Informatics)*, vol. 8, no. 1, pp. 93–99, 2025.

[11] R. Maulana, A. Voutama, and T. Ridwan, "Analisis Sentimen Ulasan Aplikasi MyPertamina pada Google Play Store menggunakan Algoritma NBC," *J. Teknol. Terpadu*, vol. 9, no. 1, pp. 42–48, 2023.

[12] A. Saputra, R. C. S. Hariyono, and N. M. Saraswati, "Analisis Sentimen Pengguna Aplikasi MyPertamina Menggunakan Algoritma Bidirectional Long Short Term Memory," *J. Eksplora Inform.*, vol. 13, no. 2, pp. 156–163, 2024.

[13] I. Maulana, W. Apriandari, and A. Pambudi, "Analisis Sentimen Berbasis Aspek Terhadap Ulasan Aplikasi Mypertamina Menggunakan Support Vector Machine," *Idealis Indones. J. Inf. Syst.*, vol. 6, no. 2, pp. 172–181, 2023.

[14] M. Safrudin, M. Martanto, and U. Hayati, "Perbandingan Kinerja Naïve Bayes Dan Support Vector Machine Untuk Klasifikasi Sentimen Ulasan Game Genshin Impact," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3182–3188, 2024.

[15] W. Romadhona and A. R. Isnain, "Analisis Sentimen Pengguna Media Sosial Terhadap Kebijakan Kenaikan Pajak Hiburan Menggunakan Metode Svm (Support Vector Machine)," *Jipi (Jurnal Ilm. Penelit. Dan Pembelajaran Inform.*, vol. 9, no. 4, pp. 2185–2195, 2024.

[16] N. Azza and N. Hadian, "Tweet Sentiment Analysis with Support Vector Machine (SVM) Algorithm for PT. XYZ Digital Strategy," *G-Tech J. Teknol. Terap.*, vol. 9, no. 4, pp. 1868–1877, 2025.

[17] A. Irawan and D. H. U. Ningsih, "IMPLEMENTASI ALGORITMA DECISION TREE DALAM SISTEM PENDUKUNG KEPUTUSAN SELEKSI CALON KARYAWAN PADA PT G-TEXTILE INDONESIA MENGGUNAKAN FRAMEWORK DJANGO PYTHON," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 10, no. 1, pp. 1302–1308, 2026.

[18] T. Wahyudi, R. Rudiman, and N. A. Verdikha, "Klasifikasi Sentimen X-Twitter Perihal

Pemindahan Ibu Kota Indonesia Menggunakan Ekstraksi Fitur TF-IDF dan Metode Support Vector Machine (SVM)," 2024, *vol*.

[19] A. A. Munandar, F. Farikhin, and C. E. Widodo, "Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM. JOINTECS (Journal of Information Technology and Computer Science), 8 (2), 77," 2023.

[20] A. Fauzi and A. H. Yunial, "Analisis sentimen pada media sosial menggunakan perbandingan algoritma data mining," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 10, no. 2, pp. 277–286, 2024.

[21] K. M. Sujon, R. Hassan, K. Choi, and M. A. Samad, "Accuracy, precision, recall, f1-score, or MCC? empirical evidence from advanced statistics, ML, and XAI for evaluating business predictive models," *J. Big Data*, vol. 12, no. 1, p. 268, 2025.

[22] S. Sathyanarayanan and B. R. Tantri, "Confusion matrix-based performance evaluation metrics," *African J. Biomed. Res.*, vol. 27, no. 4S, pp. 4023–4031, 2024.

[23] K. Paraskevas and T. Christos, "Data preprocessing and feature engineering for data mining: Techniques, tools, and best practices," *AI*, vol. 6, no. 10, p. 257, 2025.