JURNAL TEKNOLOGI TERPADU VOL.13 NO. 2 OKTOBER 2025 ISSN 2338 - 6649

Received: Agustus 2024	Accepted: Oktober 2025	Published: Oktober 2025

Analisis Sentimen Terhadap Ulasan Aplikasi MyPertamina pada Google PlayStore dengan Menggunakan Algoritma *K-Nearest Neighbor* dan *Support Vector Machine*

Deota Abie Pramudya[1*], Nengah Widya Utami[2], Putri Anugrah Cahya Dewi[3]

[1,2,3] Fakultas Teknologi dan Design, Universitas Primakara*

[*]Email: deotaabie12@gmail.com

Abstract

In the era of digital transformation, technology-based applications are key to improving the efficiency and transparency of public services. MyPertamina, an application from PT Pertamina (Persero), is designed to support fuel subsidies and facilitate digital transactions. However, the app has faced obstacles, such as complaints related to performance, user interface, and technical issues. This research aims to analyze the sentiment of MyPertamina user reviews on Google Playstore using K-Nearest Neighbor and Support Vector Machine algorithms. Review data is collected through web scraping techniques and classified into positive and negative sentiments. The research methodology used is Knowledge Discovery in Database. Of the 24,777 review data about the MyPertamina application, it resulted in 21,391 clean data after preprocessing, then the data labeling process using IndoBert. Based on the labeling, there are 8,849 positive class reviews, and 12,542 negative class reviews. This research shows that the MyPertamina application gets more negative classes than positive classes. The results of applying the K-Nearest Neighbor algorithm resulted in an accuracy of 85% and the results of applying the Support Vector Machine algorithm resulted in an accuracy of 91%.

Keywords: Sentiment analysis, k-nearest neighbor, support vector machine, mypertamina, knowledge discovery in database

Abstrak

Di era transformasi digital, aplikasi berbasis teknologi menjadi kunci untuk meningkatkan efisiensi dan transparansi pelayanan publik. MyPertamina, aplikasi dari PT Pertamina (Persero), dirancang untuk mendukung subsidi BBM dan mempermudah transaksi digital. Namun, aplikasi ini menghadapi kendala, seperti keluhan terkait performa, antarmuka pengguna, dan masalah teknis. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna MyPertamina di Google Playstore menggunakan *algoritma K-Nearest Neighbor dan Support Vector Machine*. Data ulasan dikumpulkan melalui teknik *web scraping* dan diklasifikasikan menjadi sentimen positif dan negatif. Metodologi penelitian yang digunakan adalah *Knowledge Discovery In Database.* Dari 24.777 data ulasan mengenai aplikasi MyPertamina, menghasilkan 21.391 data yang bersih setelah dilakukan *preprocessing*, selanjutnya proses pelabelan data menggunakan IndoBert. Berdasarkan pelabelan tersebut mendapatkan ulasan kelas positif sebanyak 8.849 data, dan ulasan kelas negatif sebanyak 12.542 data. Penelitian ini menunjukkan bahwa aplikasi MyPertamina mendapatkan kelas negatif lebih banyak dibandingkan dengan kelas positif. Hasil penerapan algoritma *K-Nearest Neighbor* menghasilkan akurasi sebesar 85% dan hasil penerapan algoritma *Support Vector Machine* menghasilkan akurasi sebesar 91%.

Kata kunci: Analisis sentiment, *k-nearest neighbor*, *support vector machine,* mypertamina, *knowledge discovery in database*

1. Pendahuluan

Era digital ini, teknologi informasi menjadi elemen penting dalam kehidupan. Inovasi terus berkembang untuk mendukung berbagai sektor, termasuk sektor layanan publik [1]. Pemerintah dan perusahaan di berbagai negara, termasuk Indonesia, telah memanfaatkan aplikasi berbasis digital untuk memberikan layanan yang lebih efisien dan mudah diakses oleh masyarakat. Bentuk nyata dari transformasi ini adalah peluncuran aplikasi MyPertamina oleh PT Pertamina (Persero), yang dirancang untuk mendukung program subsidi bahan bakar minyak (BBM) dan memberikan pengalaman transaksi digital yang modern.

Aplikasi Mypertamina menawarkan kemudahan yang memungkinkan pengguna untuk melakukan pembayaran dalam pembelian bahan bakar secara non-tunai, melacak transaksi, mengakses berbagai fitur lainnya seperti pendaftaran program subsidi BBM, dan memberikan umpan balik atau ulasan dari pengguna untuk aplikasi Mypertamina. Dapat disimpulkan bahwa aplikasi ini memudahkan konsumen dalam melakukan transaksi. Hingga saat ini, Lebih dari 5 juta pengguna dengan *rating* telah mengunduh MyPertamina dan lebih dari 200 ribu ulasan di Google Play Store [2]. Aplikasi ini dirilis pertama kali pada 21 Desember 2016, sebagai bagian dari strategi pemerintah untuk mendigitalisasi layanan publik, dan diharapkan mampu memberikan manfaat yang signifikan, baik bagi konsumen maupun pemerintah, dalam hal pengelolaan data dan transparansi program subsidi.

Namun, meskipun MyPertamina membawa banyak potensi manfaat, implementasinya di masyarakat tidak terlepas dari berbagai tantangan. Ulasan pengguna untuk aplikasi di *platform* seperti Google PlayStore dalam beberapa tahun terakhir menunjukkan berbagai sentimen, termasuk ulasan positif dan keluhan tentang performa aplikasi, antarmuka pengguna (*user interface*), pengalaman pengguna (*user experience*), dan kendala teknis. Ada beberapa masalah yang dihadapi oleh pengguna aplikasi MyPertamina, salah satunya adalah seringnya aplikasi mengalami *error* dan munculnya *bug*, yang menyebabkan pengguna menjadi enggan untuk menggunakan aplikasi tersebut. Selain itu, salah satu kendala lainnya adalah minimnya pemahaman aplikasi oleh petugas di lapangan [3].

Sebagian pengguna merasa aplikasi ini kurang responsif, sulit digunakan, atau memiliki fitur yang belum sesuai dengan kebutuhan mereka. Kompas.com melaporkan bahwa banyak kendala yang ditemukan pengguna saat menggunakan aplikasi MyPertamina, yang berujung pada ulasan buruk. Dari 158 ribu ulasan yang ditinggalkan, banyak pengguna menyoroti masalah yang muncul saat instalasi aplikasi tersebut. Beberapa di antaranya bahkan tidak ragu memberikan *rating* bintang satu untuk aplikasi MyPertamina. Objek penelitian ini memilih MyPertamina karena aplikasi ini merupakan bagian dari strategi pemerintah dalam mendigitalisasi layanan publik dan pengelolaan subsidi bahan bakar, sehingga permasalahannya relevan untuk dianalisis guna mendukung kebijakan nasional dan meningkatkan kepuasan pengguna. Mengingat banyaknya keluhan masyarakat dari ulasan aplikasi MyPertamina, diperlukan analisis sentimen terhadap aplikasi ini untuk mengetahui aspek apa saja yang perlu diperbaiki.

Salah satu langkah penting dalam memahami persepsi masyarakat terhadap sebuah aplikasi yaitu dengan menganalisis ulasan yang diberikan oleh pengguna. Pendekatan ini dapat memberikan perspektif apa yang perlu diperbaiki, baik dari sisi teknis maupun fungsionalitas aplikasi. Ulasan dari pengguna aplikasi seperti MyPertamina biasanya mencakup saran positif maupun keluhan negatif, yang terkadang dapat berupa sedikit atau banyak. Mengingat jumlah ulasan yang sangat banyak di media sosial, proses pengolahan secara manual menjadi tidak efisien. Oleh karena itu, metode diperlukan atau teknik khusus yang dapat dengan cepat dan otomatis menyortir dan memantau ulasan

tersebut, mengkategorikan ulasan sebagai positif atau negatif. Data untuk penelitian ini diambil dari Google PlayStore karena data dari Cloudfire 2024 pengguna Android lebih didominasi sekitar 91,3% pengguna, sedangkan IOS sebesar 8,7% pengguna di Indonesia [4].

Metode yang digunakan untuk mengumpulkan ulasan pada penelitian ini adalah dengan teknik *web scraping*, yaitu proses membaca, mengambil, dan mengumpulkan data atau dokumen dari situs *web* yang ditargetkan [5]. Setelah data dikumpulkan menggunakan teknik *web scraping*, langkah selanjutnya adalah menganalisis sentimen dari pengguna aplikasi MyPertamina. Analisis sentimen sendiri merupakan suatu proses untuk memahami kesan atau pandangan seseorang terhadap suatu isu, peristiwa, atau layanan dari sebuah objek [6]. Analisis sentimen bertujuan untuk menggali pola emosi dan pendapat dalam teks yang dibuat oleh pengguna. Cara untuk menganalisis ulasan atau pendapat yaitu dengan cara analisis sentimen. Selain itu, analisis sentimen juga membantu mengubah data yang tidak teratur menjadi lebih terstruktur. Tugas analisis sentimen adalah mengklasifikasikan polaritas teks pada tingkat aspek, dokumen, atau kalimat, untuk menentukan teks tersebut bersifat positif dan negatif [7].

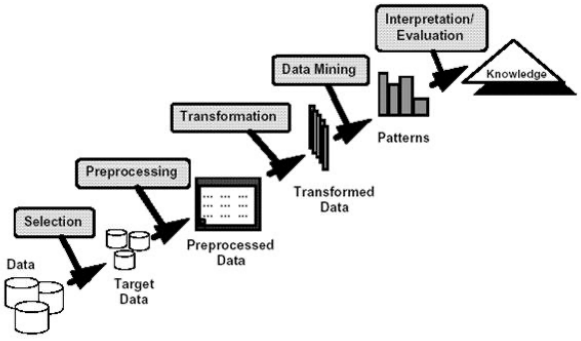
Untuk menganalisis sentimen, algoritma yang cocok digunakan adalah algoritma *K-Nearest Neighbor* karena algoritma ini berbasis jarak yang mengklasifikasikan data baru dengan membandingkannya dengan data sebelumnya yang sudah diberi label. Algoritma ini memiliki tujuan untuk mengklasifikasikan objek berdasarkan atributnya dan data latih. Cara kerja sederhana *K-Nearest Neighbor* mengukur jarak terdekat antara sampel uji dan sampel pelatihan memberinya keunggulan [8]. Karena sifatnya yang sederhana dan efektif dalam menangani dataset kecil hingga menengah, untuk analisis sentimen pada ulasan MyPertamina, metode *K-Nearest Neighbor* cocok digunakan.

Selain itu, algoritma *Support Vector Machine* juga sering digunakan dalam analisis sentimen karena kemampuannya dalam menemukan hyperplane terbaik untuk memisahkan data pada kelas yang berbeda. *Support Vector Machine* bekerja dengan baik dalam menangani data berdimensi tinggi dan sering digunakan untuk tugas klasifikasi teks yang kompleks. Maka dari itu penelitian ini menggunakan dua algoritma untuk membandingkan tingkat akurasi dari kedua algoritma tersebut.

Dari penelitian sebelumnya mengenai analisis sentimen terhadap review *Fintech* menggunakan algoritma *K-Nearest Neighbor* pada tahun 2020. Hasil dari penelitian ini mendapatkan nilai tingkat akurasi 84.76% [9]. Selanjutnya yaitu analisis sentimen terhadap Pandemi Covid-19 pada media sosial twitter. Penelitian ini menggunakan algoritma SVM, *Naive Bayes* dan *K-Nearest Neighbor* pada tahun 2021. Memiliki hasil tingkat akurasi algoritma *K-Nearest Neighbor* lebih tinggi dibanding algoritma SVM. Penelitian ini menggunakan K=20 dan mendapatkan tingkat akurasi 90.01% [10]. Berdasarkan penjelasan di atas dan didukung oleh berbagai penelitian sebelumnya, penelitian ini akan menjadikan algoritma *K-Nearest Neighbor* untuk menganalisis data.

Berdasarkan hal diatas, penelitian ini memiliki tujuan untuk melakukan analisis sentimen pada ulasan pengguna aplikasi MyPertamina menggunakan algoritma *K-Nearest Neighbor* dan *Support Vector Machine*. Diharapkan hasil penelitian ini memberikan rekomendasi bermanfaat bagi pengembang aplikasi dalam meningkatkan kualitas layanan, sekaligus membantu pemerintah dalam mengoptimalkan implementasi program digitalisasi layanan publik.

2. Metoda Penelitian



Gambar 1. *Knowledge Discovery in Database*

Knowledge Discovery in Database (KDD) merupakan proses menganalisis, mencari, serta mengekstraksi informasi yang dilakukan oleh komputer. Penulis menggunakan metode ini dengan alur penelitian sebagai berikut.

1) Identifikasi Masalah

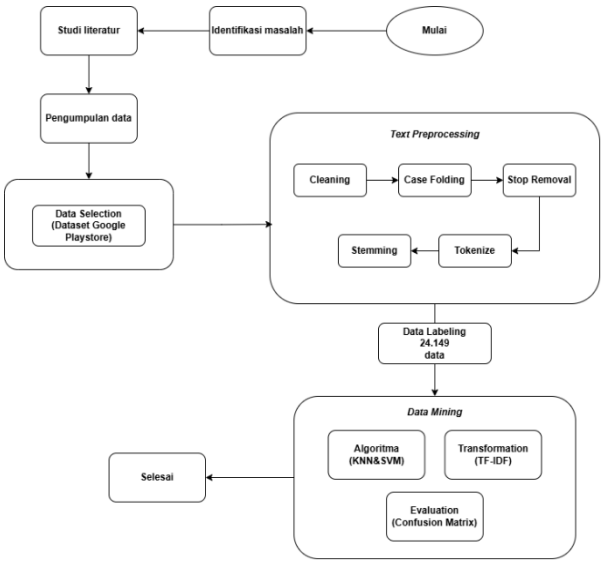
Peneliti menggali permasalahan yang terkait dengan objek penelitian, yaitu aplikasi MyPertamina. Salah satu masalah yang ditemukan adalah banyaknya keluhan terhadap fitur-fitur dalam aplikasi.

2) Studi Literatur

Peneliti melanjutkan dengan meninjau teori-teori yang relevan dari berbagai jurnal berkaitan dengan penelitian yang akan dilakukan. Studi literatur ini mencakup pembahasan mengenai data mining dan algoritma *K-Nearest Neighbor* dan *Support Vector Machine* sebagai dasar teori.

Penelitian ini memiliki sejumlah perbedaan dengan penelitian terdahulu yang membahas topik serupa. Penelitian ini menggunakan algoritma K-Nearest Neighbor, sedangkan penelitian sebelumnya menggunakan algoritma NBC [11]. Penggunaan algoritma *K-Nearest Neighbor* dilakukan karena algoritma ini memiliki tingkat akurasi terbaik secara berturut turut [12].

Dengan tingginya tingkat akurasi pada penggunaan algoritma *K-Nearest Neighbor*, diharapkan penelitian ini dapat memberikan hasil maksimal dan dapat berguna untuk pembelajaran kedepannya. Selain itu, perbedaan lainnya pada penelitian ini menggunakan algoritma *K-Nearest Neighbor* dan *Support Vector Machine* untuk dibandingkan, sedangkan beberapa penelitian membandingkan algoritma *K-Nearest Neighbor* dan *Naive Bayes* [12] [13].



Gambar 2. Alur Penelitian

3) Pengumpulan Data

Data dikumpulkan dengan mengambil ulasan dari pengguna aplikasi MyPertamina melalui teknik *web scraping*. *Web scraping* dilakukan secara otomatis menggunakan program *Python*, rentang waktu pengambilan data dari Januari 2023 hingga November 2024.

4) *Data Selection*

Data diperoleh dari Google PlayStore menggunakan teknik *Web Scraping*. Data yang diambil berupa teks, yaitu ulasan pengguna aplikasi MyPertamina. Setelah data ulasan dikumpulkan, data tersebut akan diproses untuk melakukan klasifikasi awal.

5) *Preprocessing*

Proses mengolah data menjadi data yang berkualitas sehingga bisa digunakan untuk keperluan analisis atau pemrosesan. Alur *preprocessing* yang diterapkan pada penelitian ini sebagai berikut:

a. *Cleaning Data*
Cleaning data adalah proses untuk menghapus *noise* dalam dokumen, seperti kata atau karakter yang tidak relevan.
b. *Case Folding*
Case Folding bertujuan untuk mengubah semua huruf menjadi huruf kecil.
c. *Stopword*
Stopword bertujuan untuk menghapus kata-kata yang tidak relevan.
d. *Tokenization*
Tokenization bertujuan untuk membagi teks menjadi bagian-bagian kecil berdasarkan kata.
e. *Stemming*
Stemming bertujuan untuk menghapus imbuhan atau akhiran sebuah kata.

6) *Transformation*

Pada tahap ini, digunakan algoritma pembobotan kata TF-IDF yang berfungsi merubah teks menjadi vektor. Pembobotan TF-IDF adalah teknik yang digunakan dalam pemrosesan bahasa alami untuk mengukur pentingnya suatu kata dalam sebuah dokumen.

Kata-kata yang memiliki bobot tinggi berarti kata-kata tersebut penting dalam dokumen tersebut dan memberikan kontribusi signifikan terhadap pemahaman dan representasi dokumen[11].

$$Wtf_{t,d}\ x\ idf_t \quad (1)$$

Keterangan:
$W_{t,d}$: TF-IDF pada term ke t, dokumen ke d
$Wtf_{t,d}$: Log frequency weight pada term ke t, dokumen ke d
idf_t: inverse document frequency pada term ke t

7) Data Mining

Pada penelitian ini digunakan 2 algoritma sebagai klasifikasi perbandingan tingkat akurasi. Algoritma yang pertama digunakan adalah algoritma KNN. Selanjutnya menggunakan algoritma SVM.

Algoritma *K-Nearest Neighbor* digunakan untuk mengklasifikasikan data dengan mempertimbangkan kedekatan atau jarak antara satu data dengan data lainnya [14].

Tujuan dari algoritma *K-Nearest Neighbor* yaitu mengklasifikasikan objek berdasarkan atribut dan sampel dari data latih. *K-Nearest Neighbor* memiliki kelebihan dalam kestabilannya, meskipun terdapat variasi pada nilai k. Langkah-langkah menghitung menggunakan metode *K-Nearest Neighbor* berikut:

1. Menentukan K
2. Menghitung jarak antara data training dan data testing

Rumus *Euclidean* sering digunakan untuk menghitung algoritma *K-Nearest Neighbor*.

$$Euc = \sqrt{\sum_{i=o}^{n} (x_i - x_2)^2} \quad (2)$$

Keterangan:

i : index dari atribut
n: Jumlah data
x_i : atribut dari data ke-i, (i=1, 2, 3,..,..., n)
x_2 : atribut dari data ke-i, (i=1, 2, 3,..,..., n)

Algoritma selanjutnya adalah *Support Vector Machine*. *Support Vector Machine* adalah metode yang baik untuk melakukan prediksi dalam kasus klasifikasi. SVM beroperasi dengan mengandalkan prinsip *machine learning*, statistik, dan optimasi, kinerja algoritma ini ditentukan sejauh mana model sesuai dengan data latih yang diukur melalui kesalahan empiris dan ruang hipotesis [15].

8) *Evaluation*

Pada tahap ini dilakukan perhitungan akurasi, presisi, dan recall menggunakan *confusion matrix*. Alat evaluasi dalam data mining digunakan menilai kinerja model atau algoritma yang telah diterapkan [16].

Proses evaluasi ini penting memastikan model atau algoritma tersebut dapat menghasilkan prediksi yang tepat. Untuk analisis sentimen, salah satu metode evaluasi yang paling tepat digunakan adalah confusion matrix.

$$AKURASI = \frac{TP+TN}{TP+TN+FP+FN} * 100 \quad (3)$$

$$PRESISI = \frac{TP}{FP+TP} * 100 \quad (4)$$

$$RECALL = \frac{TP}{FN+TP} * 100 \quad (5)$$

Keterangan:

1. Akurasi adalah untuk mengukur seberapa sering model membuat prediksi yang benar.
2. Presisi adalah untuk mengukur seberapa banyak prediksi positif tanpa memperhatikan prediksi negatif.
3. Recall adalah untuk mengukur seberapa banyak dari semua data positif yang berhasil diprediksi sebagai positif.

3. Hasil Penelitian

Terdapat beberapa tahapan yang dilakukan proses KDD dalam penelitian ini. Tahapan-tahapan yang dilakukan adalah sebagai berikut:

3.1. Selection

Data yang digunakan pada penelitian ini adalah data ulasan aplikasi MyPertamina yang diambil melalui Google PlayStore. Data tersebut dipilih dari bulan Januari 2023 hingga November 2024 dan mendapatkan data sebanyak 24.777.

3.2. Preprocessing

Setelah data terkumpul selanjutnya akan masuk kedalam tahap *preprocessing* meliputi tahap *tokenizing, case folding, stopword, dan stemming*. Berikut adalah contoh hasil dari proses preprocessing data.

Tabel 1. *Preprocessing*

Tahapan	Ulasan
Cleaning Data	Sangat menyenangkan sekali
	Ribet Terlalu banyak input data
	Aplikasi cukup mudah digunakan
Case Folding	sangat menyenangkan sekali
	ribet terlalu banyak input data
	aplikasi cukup mudah digunakan
Normalization	sangat menyenangkan sekali
	ribet terlalu banyak input data
	aplikasi cukup mudah digunakan
Stop Removal	Menyenangkan
	ribet input data
	aplikasi mudah
Tokenize	['menyenangkan']
	['ribet', 'input', 'data']
	['aplikasi', 'mudah']
Stemming	['senang']
	['ribet', 'input', 'data']
	['aplikasi', 'mudah']

3.3. Data Mining

Pada proses data mining akan membandingkan penggunaan algoritma untuk menganalisis data. Pada penelitian ini menggunakan algoritma *K-Nearest Neighbor* dan *Support Vector Machine* menggunakan tools Google Colab untuk melakukan analisis data ini.

3.3.1 IndoBert

Dataset yang telah melalui *preprocessing*, selanjutnya dilakukan proses labeling untuk memberi label pada ulasan.

Tabel 2. *Labeling*

Ulasan	Label
sangat menyenangkan sekali	Positif
ribet terlalu banyak input data	Negatif
aplikasi cukup mudah digunakan	Positif

3.3.2 TF-IDF

TF – IDF adalah metode statistik yang mengukur pentingnya kata dalam sebuah dokumen.

Tabel 3. TF-IDF

Term	TF	IDF	TF-IDF
senang	1,0	0,477	0,4777
ribet	0,333	0,477	0,159
input	0,333	0,477	0,159
data	0,333	0,477	0,159
aplikasi	0,5	0,477	0,239
mudah	0,5	0,477	0,239

3.3.3 Klasifikasi KNN & SVM

Dataset yang digunakan dalam penelitian ini berasal dari hasil pengumpulan ulasan pengguna pada aplikasi, yang telah melalui tahap *preprocessing* seperti pembersihan teks dan labeling sentimen menggunakan IndoBert. Setelah dilakukan pelabelan data maka didapatkan ulasan yang termasuk ke dalam kelas negatif sebanyak 12.542 data, ulasan kelas positif sebanyak 8.849 data. Hasil penelitian menggunakan algoritma *K-Nearest Neighbor* mendapatkan akurasi 85%, sedangkan algoritma *Support Vector Machine* mendapatkan akurasi 91%. Dapat disimpulkan bahwa algoritma *Support Vector Machine* mendapatkan hasil akurasi yang lebih tinggi dibandingkan *K-Nearest Neighbor*.

Tabel 4. Hasil KNN

KNN	Precision	Recall	F1-Score	Support
Negatif	0.82	0.95	0.88	2530
Positif	0.91	0.70	0.79	1749
Accuracy			0.85	4279
Macro Avg	0.86	0.82	0.84	4279
Weighted avg	0.86	0.85	0.84	4279

Tabel 5 Hasil SVM

SVM	Precision	Recall	F1-Score	Support
Negatif	0.89	0.97	0.93	2530
Positif	0.96	0.83	0.89	1749
Accuracy			0.91	4279
Macro Avg	0.92	0.90	0.91	4279
Weighted avg	0.92	0.91	0.91	4279

3.4. Visualisasi

Dalam penelitian ini, visualisasi dilakukan untuk membantu memahami persebaran data sentimen dan performa model klasifikasi. Pada penelitian ini visualisasi dilakukan dengan cara wordcloud, confusion matrix, perbandingan algoritma KNN dan SVM. Untuk mendapatkan gambaran awal mengenai kata-kata yang paling sering muncul dalam ulasan pengguna, dilakukan visualisasi dalam bentuk *Wordcloud*. *Wordcloud* ini dihasilkan dari pemisahan berdasarkan label sentimen yaitu positif, dan negatif.

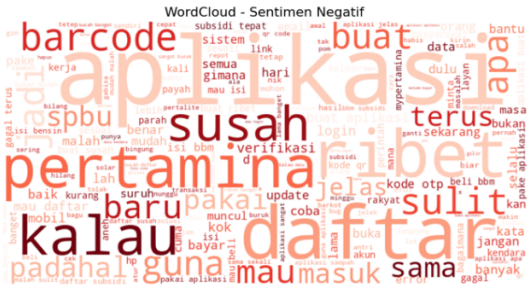
3.4.1 *Wordcloud* Positif

Data sentimen positif merupakan hasil dari pelabelan yang diklasifikasi ke dalam kelas positif.

Gambar 3. *Wordcloud* Positif

3.4.2 *Wordcloud* Negatif

Ulasan yang memiliki sentimen negatif adalah hasil pelabelan data yang termasuk dalam kelas negatif dalam analisis sentimen.



Gambar 4. *Wordcloud* Negatif

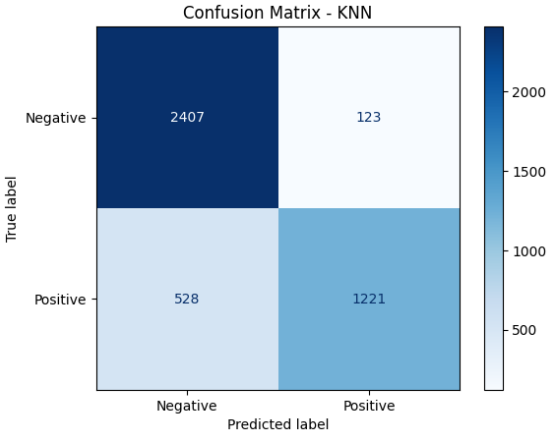
Pada *Wordcloud* sentimen negatif, didominasi kata seperti "tidak bisa", "susah", "tidak bisa", “ribet”, “daftar”, “barcode”, dan "aplikasi" menunjukkan kecenderungan keluhan dari pengguna.

Wordcloud ini memberikan gambaran visual yang membantu dalam memahami persepsi dan fokus perhatian pengguna berdasarkan kategorisasi sentimen.

3.4.3 *Confusion Matrix* KNN

Gambar di bawah menunjukkan hasil visualisasi confusion matrix dari algoritma KNN. Berdasarkan confusion matrix tersebut, diperoleh hasil sebagai berikut:

a. KNN berhasil memprediksi dengan benar 2407 data ulasan yang memiliki sentimen negatif dari total keseluruhan data berlabel negatif, sementara 123 ulasan negatif salah diklasifikasikan menjadi positif.
b. Di sisi lain, ulasan positif algoritma KNN berhasil mengklasifikasikan 1.221 ulasan dengan benar sebagai positif, namun masih terdapat 528 ulasan positif salah diklasifikasikan menjadi negatif.



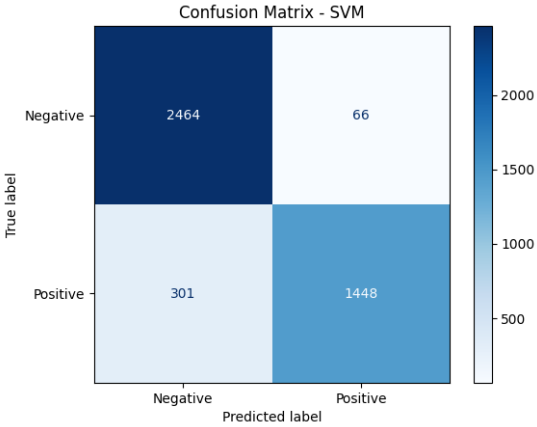
Gambar 5. *Confusion Matrix* KNN

Hasil menunjukkan bahwa algoritma KNN mengklasifikasikan baik dalam mendeteksi ulasan dengan sentimen negatif dibandingkan ulasan positif.

3.4.4 *Confusion Matrix* SVM

Berdasarkan gambar di atas, model SVM menunjukkan performa yang baik dalam mengklasifikasikan data sentimen.

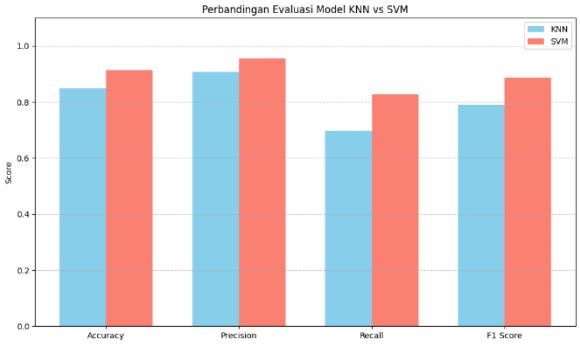
a. Sebanyak 3.267 data ulasan yang berlabel negatif berhasil diklasifikasikan dengan benar oleh model sebagai negatif. Namun, terdapat 31 data negatif yang salah diklasifikasikan sebagai positif.
b. Model berhasil mengklasifikasikan 1.254 data positif dengan benar. Namun terdapat kesalahan, di mana 243 data positif salah diklasifikasikan sebagai negatif.



Gambar 6. *Confusion Matrix* SVM

Dari hasil ini, dapat disimpulkan bahwa model SVM memiliki tingkat akurasi yang tinggi, khususnya dalam mengurangi kesalahan klasifikasi baik untuk kelas negatif maupun positif.

3.4.5 Perbandingan algoritma KNN & SVM



Gambar 7. Perbandingan KNN & SVM

Evaluasi performa model *K-Nearest Neighbor* dan *Support Vector Machine* dilakukan untuk klasifikasi sentimen ulasan pengguna aplikasi MyPertamina, dengan hasil metrik *precision*, *recall*, *f1-score*, dan *accuracy*. Model SVM menunjukkan performa unggul dengan akurasi 91%, *precision* 92%, *recall* 91%, dan *f1-score* 91% (*weighted average*), dibandingkan KNN dengan akurasi 85%, *precision* 86%, *recall* 85%, dan *f1-score* 84%.

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan maka diperoleh beberapa kesimpulan sebagai berikut:

a. Dari 24.777 data ulasan mengenai aplikasi MyPertamina, menghasilkan 21.391 data yang bersih setelah dilakukan *preprocessing*, selanjutnya proses pelabelan data menggunakan IndoBert. Berdasarkan pelabelan tersebut mendapatkan ulasan kelas positif sebanyak 8.849 data, dan ulasan kelas negatif sebanyak 12.542 data. Penelitian ini menunjukkan bahwa aplikasi MyPertamina mendapatkan kelas negatif lebih banyak dibandingkan dengan kelas positif.
b. Hasil penerapan algoritma *K-Nearest Neighbor* menghasilkan akurasi sebesar 85% dan hasil penerapan algoritma *Support Vector Machine* menghasilkan akurasi sebesar 91%. Algoritma *Support Vector Machine* lebih unggul dibandingkan dengan *K-Nearest Neighbor.*

5. Daftar Pustaka

[1] R. Mutaqin, G. Mutaqin, and D. S. Dharmopadni, "Dampak Perkembangan Teknologi Informasi Dan Komunikasi Terhadap Dinas Militer," *J. Ilm. Multidisiplin*, vol. 2, no. 3, 2024, doi: 10.59000/jim.v2i3.213.

[2] C. G. Indrayanto, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Data Ulasan Pengguna Aplikasi MyPertamina di Indonesia pada Google Play Store menggunakan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 3, pp. 1131–1139, 2023, [Online]. Available: http://j-ptiik.ub.ac.id

[3] A. Amelia, L. Nur, and H. Darwis, "Analisis Sentimen Masyarakat Terhadap Sistem Pembayaran Mypertamina dengan Metode Random Forest , SVM , dan Naïve Bayes," vol. 1, no. 1, pp. 30–46, 2023.

[4] Cloudflare, "Year in Review 2024 – Indonesia: iOS vs Android." Cloudflare, Inc., 2024. [Online]. Available: https://radar.cloudflare.com/year-in-review/2024/id#ios-vs-android

[5] F. Djiwadikusumah, G. H. Irawan, and R. Haekal Al-Fadilah, "Web scraping situs e-commerce menggunakan teknik parsing dom," *J. Siliwangi*, vol. 7, no. 2, pp. 52–57, 2021, [Online]. Available: https://jurnal.unsil.ac.id/index.php/jssainstek/article/view/4223/1958

[6] R. A. Saputra *et al.*, "Analisis Sentimen Aplikasi Tokocrypto Berdasarkan Ulasan Pada Google Play Store Menggunakan Metode Naïve Bayes," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 4, no. 4, pp. 2028–2036, 2024, doi: 10.30865/klik.v4i4.1707.

[7] L. Purnama and T. Wahyudi, "Analisa Sentimen Tentang Piala Dunia u-20 Indonesia Menggunakan," vol. 6, no. 2, pp. 217–222, 2024.

[8] S. Surohman, S. Aji, R. Rousyati, and F. F. Wati, "Analisa Sentimen Terhadap Review Fintech Dengan Metode Naive Bayes Classifier Dan K- Nearest Neighbor," *EVOLUSI J. Sains dan Manaj.*, vol. 8, no. 1, pp. 93–105, 2020, doi: 10.31294/evolusi.v8i1.7535.

[9] F. S. Pamungkas and I. Kharisudin, "Analisis Sentimen dengan SVM, NAIVE BAYES dan KNNuntuk Studi Tanggapan Masyarakat Indonesia TerhadapPandemi Covid-19 pada Media Sosial Twitter," *Pros. Semin. Nas. Mat.*, vol. 4, pp. 1–7, 2021, [Online]. Available: https://journal.unnes.ac.id/sju/prisma/article/view/45038

[10] A. Apriani, H. Zakiyudin, and K.

Marzuki, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF System Penerimaan Mahasiswa Baru pada Kampus Swasta," *J. Bumigora Inf. Technol.*, vol. 3, no. 1, pp. 19–27, 2021, doi: 10.30812/bite.v3i1.1110.

[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, doi: 10.54914/jtt.v9i1.609.

[12] S. M. Siroj, I. Arwani, and D. E. Ratnawati, "Analisis Sentimen Opini Publik pada Twitter terhadap Efek Pembelajaran Daring di Universitas Brawijaya menggunakan Metode K-Nearest Neighbor," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 7, pp. 3131–3140, 2021, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/9494

[13] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN," *J. KomtekInfo*, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.

[14] I. Widaningrum, D. Mustikasari, R. Arifin, S. L. Tsaqila, and D. Fatmawati, "Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) dan K-Means Clustering Untuk Menentukan Kategori Dokumen," *Pros. Semin. Nas. Sist. Inf. dan Teknol.*, pp. 145–149, 2022.

[15] S. K. P. Loka and A. Marsal, "Perbandingan Algoritma K-Nearest Neighbor dan Naïve Bayes Classifier untuk Klasifikasi Status Gizi Pada Balita," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 1, pp. 8–14, 2023, doi: 10.57152/malcom.v3i1.474.

[16] I. T. Akinola, Y. Sun, I. G. Adebayo, and Z. Wang, "Daily peak demand forecasting using Pelican Algorithm optimised Support Vector Machine (POA-SVM)," *Energy Reports*, vol. 12, pp. 4438–4448, Dec. 2024, doi: 10.1016/J.EGYR.2024.10.017.

[17] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI*, vol. 5, no. 2, pp. 697–711, 2021.