Journal of Science and Social Research ISSN 2615 – 4307 (Print)
August 2026, IX (4): 6829 – 6836 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

IDENTIFIKASI SENTIMEN PENGGUNA APLIKASI JKN BERBASIS TERM FREQUENCY-DOCUMENT FREQUENCY DAN SUPPORT VECTOR MACHINE

Alicia Mulya Maharani[1], Musli Yanto[2], Billy Hendrik[3]
Universitas Putra Indonesia YPTK, Padang
e-mail: [1]aliciamulyaamhrni@gmail.com, [2]musli_yanto@upiyptk.ac.id,
[3]billy_hendrik@upiyptk.ac.id

Abstract: *User reviews of the Jaminan Kesehatan Nasional (JKN) mobile application on app distribution platforms continue to increase in line with the growing use of digital health services in Indonesia. These reviews contain sentiments related to the application, yet they have not been systematically analyzed to reveal user perception and satisfaction toward the JKN application. This study aims to analyze sentiment in JKN application reviews by employing the Term Frequency–Inverse Document Frequency (TF-IDF) method combined with the Support Vector Machine (SVM) algorithm. The review data processing stage began with preprocessing, comprising cleaning, case folding, tokenization, stopword removal, and stemming, followed by weighting using TF-IDF to transform textual data into numerical data in accordance with word occurrence frequency. The weighting results then served as input for the SVM algorithm to construct a separating hyperplane between classes, thereby classifying reviews into positive or negative categories. The research dataset consisted of 10,000 JKN application user reviews obtained through scraping from the Google Play Store platform. The testing results showed an accuracy of 99.83%, precision of 99.92%, recall of 99.83%, and an F1-score of 99.87%. These results demonstrate that the combination of TF-IDF and SVM can produce highly accurate sentiment classification of JKN application user reviews. User perception and satisfaction obtained from this sentiment identification serve as an evaluative basis for JKN application administrators to encourage improvements in the quality of digital health services in Indonesia.*

Keyword: *Sentiment Analysis, TF-IDF, JKN, Support Vector Machine, User Reviews.*

Abstrak: Ulasan pengguna aplikasi Jaminan Kesehatan Nasional (JKN) pada platform distribusi aplikasi terus bertambah seiring meningkatnya penggunaan layanan kesehatan digital di Indonesia. Ulasan tersebut memuat sentimen terkait kualitas layanan, kemudahan penggunaan, dan kinerja fitur, namun belum dianalisis secara sistematis untuk mengetahui persepsi dan kepuasan pengguna terhadap aplikasi JKN. Sentimen dalam ulasan pengguna aplikasi JKN dianalisis menggunakan metode Term Frequency Inverse Document Frequency (TF-IDF) dan Support Vector Machine (SVM) sebagai tujuan dari penelitian ini. Tahapan untuk mengolah data ulasan pada penelitian ini merupakan preprocessing (cleaning, case folding, tokenization, stopword removal, dan stemming), disusul pembobotan memakai TF-IDF agar data teks bertransformasi menjadi data numerik selaras dengan frekuensi kemunculan kata. Hasil pembobotan menjadi input bagi algoritma SVM untuk membentuk hyperplane pemisah antar kelas, sehingga ulasan diklasifikasikan ke dalam kategori positif atau negatif. Dataset penelitian berjumlah 10.000 ulasan pengguna aplikasi JKN yang diperoleh melalui scraping dari platform Google Play Store. Hasil pengujian dari penelitian yaitu akurasi sebesar 99,83%, presisi 99,92%, recall 99,83%, dan F1-score 99,87%. Hasil dari pengujian tersebut membuktikan kombinasi TF-IDF dan SVM dapat menghasilkan klasifikasi sentimen ulasan pengguna aplikasi JKN dengan ketepatan tinggi. Gambaran mengenai persepsi dan kepuasan pengguna diperoleh dari hasil identifikasi sentimen ini untuk bahan evaluasi bagi

pengelola aplikasi JKN guna mendorong peningkatan kualitas layanan kesehatan digital di Indonesia.

Kata kunci: Analisis Sentimen, TF-IDF, JKN, Support Vector Machine, Ulasan Pengguna.

PENDAHULUAN

Ulasan pengguna pada aplikasi digital menjadi sumber data penting karena memuat pendapat dan pengalaman mereka terhadap kinerja layanan yang disediakan (Ulya et al., 2022). Ulasan tersebut mencakup pengalaman, penilaian, dan pendapat mengenai kinerja serta kualitas layanan aplikasi yang digunakan (Helmayanti et al., 2023). Kendala teknis dan tingkat kemudahan penggunaan aplikasi turut membentuk persepsi masyarakat yang tercermin dalam ulasan pengguna. Ulasan data biasanya berupa teks bebas yang tidak terstruktur sehingga saat diolah sulit secara manual, terutama dalam jumlah besar (Afdal & Elita, 2022). Ulasan pengguna aplikasi mencerminkan berbagai persepsi yang penting terhadap umpan balik secara keseluruhan terhadap berbagai komentar seperti kinerja aplikasi, kemudahan penggunaan, ataupun kualitan layanan. (Ulya et al., 2022). Hal itu menjadi urgensi terhadap ulasan pengguna aplikasi dalam analisis sentimen secara sistematis pada aplikasi Jaminan Kesehatan Nasional (JKN), terutama karna

Studi mengenai ulasan aplikasi Access by KAI mengindikasikan bahwa sebagian besar sentimen termasuk dalam kategori positif, sementara sentimen negatif yang muncul akibat kendala teknis tetap dapat dijadikan bahan evaluasi layanan (Damanhuri & Husein, 2024). Penelitian pada ulasan aplikasi TikTok di Google Play Store juga menemukan sentimen negatif lebih dominan pada aspek performa dan kebijakan aplikasi, sedangkan sentimen positif banyak muncul pada fitur hiburan yang disediakan (Maulana et al., 2025). Penelitian mengenai analisis sentimen terhadap aplikasi Jamsostek Mobile (JMO) menunjukkan bahwa analisis sentimen diperlukan untuk mengelompokkan opini pengguna di Google Play Store, yang menghasilkan beragam sentimen mulai dari kepuasan terhadap kemudahan layanan hingga keluhan terkait kendala teknis (Butsianto & Muhammad, 2025).

Pendekatan text mining yang dimanfaatkan guna mengenali arah opini sekaligus menggolongkan data teks ke dalam kategori sentimen tertentu disebut analisis sentimen. (Husen et al., 2023). Ketidakmampuan algoritma klasifikasi memproses data teks pada ulasan secara langsung melatarbelakangi kebutuhan akan teknik pembobotan kata, seperti Term Frequency Inverse Document Frequency (TF-IDF). Bentuk numerik pun diperoleh teks berdasarkan frekuensi kemunculan kata melalui teknik tersebut. (Emarapenta et al., 2024). Kata kunci ditemukan TF-IDF dengan melihat seberapa sering kata muncul dalam teks. Kemungkinan menjadi kata kunci semakin besar seiring semakin seringnya kata muncul (Wibowo & Hidayat, 2024). Bobot pada setiap kata ditetapkan TF-IDF, tingkat kepentingan suatu kata dalam setiap dokumen pun ditunjukkan, untuk mengubah data teks menjadi data numerik (Fahmi et al., 2025).

Teknik ini sering dikombinasikan dengan algoritma klasifikasi yang lebih canggih karena masih memiliki keterbatasan, seperti tidak mempertimbangkan urutan kata dalam kalimat (Emarapenta et al., 2024). Masukan bagi algoritma Support Vector Machine (SVM) diperoleh dari hasil pembobotan TF-IDF tersebut. Algoritma ini bekerja dengan menelusuri hyperplane paling ideal guna memilah data ke dalam kelas sentimen yang berlainan disertai

jarak pemisah semaksimal mungkin (Azzahra Livia, 2025). Keefektifan SVM dalam konteks analisis sentimen ulasan aplikasi tergambar dari kesanggupannya menangani data teks berdimensi tinggi serta bersifat jarang hasil TF-IDF. (Pertiwi & Utomo, 2023). SVM memanfaatkan prinsip optimasi matematis untuk membangun model berbasis metode linier, dan tujuan SVM adalah menentukan fungsi optimal yang membentuk hyperplane sebagai batas pemisah untuk mengklasifikasikan data masukan ke berbagai kategori (Nugraha & Hardjianto, 2024).

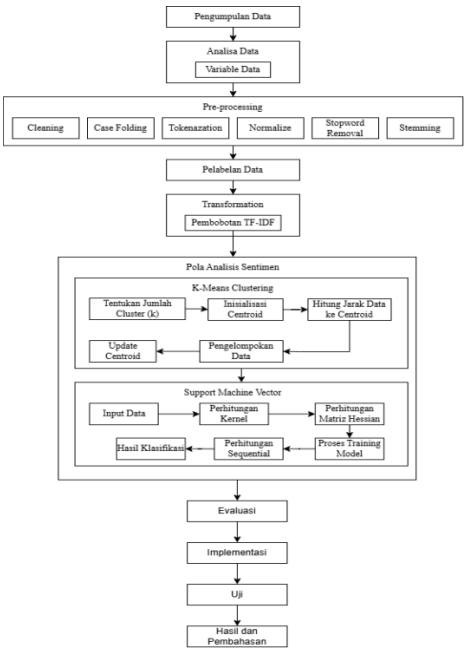
Penelitian terdahulu menunjukkan bahwa kombinasi TF-IDF dan SVM efektif digunakan dalam analisis sentimen karena TF-IDF mampu menghasilkan fitur teks yang informatif, sementara SVM berperan membangun model klasifikasi dengan margin pemisah yang optimal. Karakteristik data teks yang tidak tertata mampu ditangani lewat penerapan TF-IDF dan SVM pada ulasan aplikasi digital lain, disertai capaian akurasi yang konsisten (Husain & Syam, 2024). Penelitian terdahulu mengenai analisis sentimen ulasan aplikasi digital umumnya menerapkan TF-IDF dan SVM dengan label sentimen yang diambil langsung dari rating bintang atau pelabelan manual, seperti pada ulasan aplikasi Access by KAI (Damanhuri & Husein, 2024), Jamsostek Mobile (Butsianto & Muhammad, 2025), dan Mobile JKN (Sugihartono et al., 2024). Pendekatan ini berisiko menimbulkan bias label karena rating bintang tidak selalu sejalan dengan isi ulasan, sedangkan pelabelan manual kurang efisien untuk data berskala besar. Penerapan metode tersebut masih terbatas pada domain transportasi, hiburan, dan perbankan digital, sementara layanan kesehatan digital publik seperti aplikasi JKN Mobile dengan karakteristik data dan sensitivitas berbeda belum banyak disentuh. Penelitian ini menawarkan pendekatan berbeda dengan menggabungkan TF-IDF dan K-Means Clustering sebagai tahap pelabelan sentimen otomatis berdasarkan kemiripan bobot kata pada tiga variabel ulasan (kinerja aplikasi, kemudahan penggunaan, dan kualitas layanan), sebelum diklasifikasikan menggunakan SVM. Kombinasi tiga tahap ini menjadi kebaruan penelitian karena belum diterapkan pada analisis sentimen layanan publik digital domain kesehatan digital nasional yang masih jarang dieksplorasi.

Kecenderungan sentimen pada ulasan pengguna aplikasi Jaminan Kesehatan Nasional (JKN) hendak diidentifikasi lewat pendekatan text mining pada penelitian ini, bertumpu pada uraian yang telah dipaparkan sebelumnya. Metode TF-IDF diterapkan untuk memberi bobot pada kata-kata dalam teks ulasan, sedangkan algoritma SVM dipakai untuk menggolongkan hasil bobot tersebut ke dalam kategori sentimen positif dan negatif. Sebanyak 10.000 ulasan pengguna aplikasi JKN yang dihimpun lewat scraping dari platform Google Play Store dipakai pada penelitian ini. Tahap preprocessing pun dilalui data tersebut guna memperoleh teks yang bersih dan konsisten, sebelum ditindaklanjuti pada tahap berikutnya. Metrik akurasi, presisi, recall, dan F1-score dipakai untuk menakar kinerja model yang dibangun, guna memastikan kelayakan hasil klasifikasi yang diperoleh. Hasil identifikasi sentimen dalam penelitian ini melihat gambaran yang jelas mengenai persepsi dan tingkat kepuasan pengguna terhadap aplikasi JKN, serta bahan pertimbangan bagi pengelola aplikasi dalam menyusun langkah pembenahan guna mendongkrak mutu layanan kesehatan digital di Indonesia.

METODE

Bagian ini menguraikan tahapan penelitian secara sistematis, mulai dari kerangka kerja yang menjadi acuan penelitian hingga dasar teori metode yang diterapkan. Urutan tertentu diikuti dalam penyusunan setiap tahapan, supaya proses analisis sentimen ulasan pengguna

aplikasi Jaminan Kesehatan Nasional (JKN) berlangsung runtut, terhitung dari pengumpulan data, pra-pemrosesan (pre-processing), pembobotan memakai Term Frequency–Inverse Document Frequency (TF-IDF), sampai algoritma Support Vector Machine (SVM) untuk penggolongan sentimen. Kerangka penelitian tersebut disajikan pada Gambar 1.



Gambar 1 Kerangka Penelitian

Tahapan penelitian yang ditampilkan pada Gambar 1 diawali dengan proses pengumpulan data, yaitu ulasan pengguna aplikasi JKN Mobile yang bersumber dari platform Google Play Store. Data yang berhasil dikumpulkan selanjutnya dianalisis untuk mengidentifikasi karakteristik serta menentukan variabel yang menjadi fokus dalam penelitian. Tahapan pre-processing yang mencakup cleaning, case folding, tokenization, normalization, stopword removal, serta stemming dilalui oleh data tersebut, supaya data teks yang dihasilkan lebih tertata dan siap dimanfaatkan pada tahap analisis. Tahap berikutnya dilakukan transformasi menggunakan metode TF-IDF yang dapat ubah data teks menjadi numerik. Analisis sentimen memanfaatkan proses tersebut, diawali dengan pengelompokan data memakai algoritma K-Means Clustering. Proses klasifikasi memakai algoritma SVM pun ditempuh sebagai tahap lanjutan. Hasil dari model yang didapatkan kemudian dievaluasi guna mengetahui performa klasifikasi, sebelum diterapkan dan diuji pada tahap implementasi. Seluruh rangkaian proses tersebut membentuk alur penelitian yang sistematis dalam menghasilkan analisis sentimen pada ulasan pengguna aplikasi JKN Mobile

Text Mining

Teknik pemrosesan bahasa alami yang dimanfaatkan guna menggali informasi dari data teks tidak tertata disebut text mining, sehingga pola dan pengetahuan yang tersembunyi di dalamnya bisa dikenali secara runtut (Husen et al., 2023).

Term Frequency-Document Frequency

Metode pembobotan kata yang mengukur tingkat kepentingan suatu kata berdasarkan frekuensi kemunculannya dalam sebuah dokumen pada seluruh kumpulan dokumen disebut Term Frequency-Inverse Document Frequency (TF-IDF). (Emarapenta et al., 2024) Nilai TF diperoleh dari tingkat frekuensi kemunculan sebuah kata di dalam dokumen, secara matematis pada persamaan 1:

$$TF_{d,t} = \frac{t,d}{d} \quad (1)$$

frekuensi kemunculan disimbolkan dengan f, term (t) berupa kata atau frasa, dan simbol d merupakan dokumen. Selanjutnya tahapan perhitungan IDF. Berikut untuk memperoleh nilai IDF pada Persamaan 2.

$$IDF_t = log\ \frac{N}{df(t)} \quad (2)$$

Simbol N menyatakan jumlah seluruh dokumen, sedangkan df(t) merupakan jumlah dokumen yang memuat istilah t. Bobot TF-IDF didapatkan dari perkalian kedua nilai tersebut, seperti pada Persamaan 3.

$$W_{t,d} = TF_{d,t} \times IDF \quad (3)$$

Simbol W menyatakan bobot suatu kata pada dokumen ke-n, d menunjukkan

dokumen, sedangkan simbol t merupakan kata atau term. TF digunakan untuk mengukur frekuensi kemunculan kata dalam sebuah ulasan, IDF menunjukkan tingkat kepentingan kata tersebut berdasarkan seluruh kumpulan dokumen.

Support Vector Machine

Algoritma machine learning dalam kategori supervised learning yang dipakai untuk menyelesaikan persoalan klasifikasi maupun regresi dengan mengandalkan pola-pola yang terbentuk pada data disebut Support Vector Machine (SVM) (Nugraha & Hardjianto, 2024). Support vector adalah titik-titik data yang letaknya paling dekat dengan hyperplane tersebut dan paling berpengaruh dalam menentukan posisinya. Dasar penentuan hyperplane terbaik ditentukan oleh prinsip margin maksimum. Ketepatan model dalam menggeneralisasi data yang belum pernah dijumpai sebelumnya turut meningkat seiring membesarnya jarak pemisah antar kelas (Fahmi et al., 2025). Untuk kasus SVM linier, hubungan tersebut dirumuskan pada Persamaan 4:

$$f(x) = w \cdot x_i + b = 0 \quad (4)$$

$f(x)$ adalah fungsi prediksi, yang berperan sebagai vektor normal *hyperplane* disimbolkan *w* (*weight vector*), x_i adalah vektor fitur dari data input, dan *b* adalah bias atau *intercept*. Kelas bernilai -1 dinyatakan pada Persamaan 5:

$$(w \cdot x_i + b) \leq y_i = -1 \quad (5)$$

Kelas bernilai +1 dinyatakan pada Persamaan 6:

$$(w \cdot x_i + b) \geq y_i = 1 \quad (6)$$

SVM pada dasarnya bekerja sebagai pengklasifikasi linier, tetapi penerapan fungsi kernel memungkinkan data yang tidak dapat dipisahkan secara linier ditangani oleh metode ini. (Husain & Syam, 2024). Terdapat tiga jenis kernel yang umum digunakan. Kernel linear digunakan untuk data dua kelas yang sudah terpisah secara linier, dirumuskan pada Persamaan 7:

$$K(x_i, x) = x_i^t \quad (7)$$

$K(x_i, x)$ sebagai fungsi kernel yang mengukur kedekatan antara dua data, x_i^t sebagai transpose dari data latih, dan x sebagai data uji. Kernel RBF (*Radial Basis Function*) digunakan untuk data yang tidak terpisah secara linier, dirumuskan pada Persamaan 8:

$$K(x_i, x) = \exp(-\frac{||x_i - x||^2}{2\delta^2}) \quad (8)$$

$||x_i, x||^2$ menyatakan jarak Euclidean kuadrat antar dua data, sedangkan δ adalah parameter yang mengatur luas pengaruh suatu data terhadap data lain dalam proses klasifikasi. Kernel polynomial digunakan pula untuk data yang tidak terpisah secara linier, dirumuskan pada Persamaan 9:

$$K(x_i, x) = (x \cdot x_i + c)^p \quad (9)$$

c adalah konstanta yang menyesuaikan hasil transformasi agar model lebih fleksibel dalam memisahkan data, sedangkan *p* adalah derajat polinomial (*degree*) yang menentukan kompleksitas pola klasifikasi, semakin besar nilai *p*, semakin kompleks pola yang dapat dibentuk kernel tersebut (Azzahra Livia, 2025).

HASIL DAN PEMBAHASAN

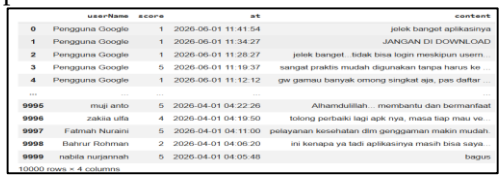
Pembahasan diawali dengan penjelasan mengenai dataset yang digunakan dalam penelitian, dilanjutkan dengan analisis variabel yang mencakup hasil preprocessing, pembobotan kata menggunakan TF-IDF, serta pengelompokan data menggunakan K-Means Clustering, terakhir hasil klasifikasi sentimen menggunakan algoritma SVM beserta evaluasi kinerja model yang dihasilkan.

Dataset

Data teks ulasan pengguna aplikasi Jaminan Kesehatan Nasional (JKN) menjadi dataset yang dipakai pada penelitian ini yang diperoleh dari *platform*

Journal of Science and Social Research ISSN 2615 – 4307 (Print)
August 2026, IX (4): 1 – 6 ISSN 2615 – 3262 (Online)
Available online at http://jurnal.goretanpena.com/index.php/JSSR

Google Play Store. Pengambilan data dilakukan melalui teknik *scraping* menggunakan *script Python* dengan pustaka *google_play_scraper* yang dijalankan pada *Google Colab*, sehingga proses pengumpulan ulasan dapat berjalan secara otomatis dan lebih efisien. Proses tersebut menghasilkan sebanyak 10.000 ulasan pengguna aplikasi JKN, yang selanjutnya disimpan dalam format *file* Excel untuk digunakan pada tahap pengolahan data berikutnya. Hasil pengumpulan data tersebut ditampilkan pada Gambar 2.



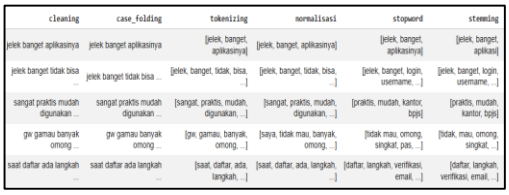
	userName	score	at	content
0	Pengguna Google	1	2026-06-01 11:41:54	jelek banget aplikasinya
1	Pengguna Google	1	2026-06-01 11:34:27	JANGAN DI DOWNLOAD
2	Pengguna Google	1	2026-06-01 11:28:27	jelek banget...tidak bisa login meskipun usern...
3	Pengguna Google	5	2026-06-01 11:19:37	sangat praktis mudah digunakan tanpa harus ke...
4	Pengguna Google	1	2026-06-01 11:12:12	gw gamau banyak omong singkat aja, pas daftar ...
...	...	...	...	...
9995	muji anto	5	2026-04-01 04:22:26	Alhamdulillah... membantu dan bermanfaat
9996	zakiia ulfa	4	2026-04-01 04:19:50	tolong perbaiki lagi apk nya, masa tiap mau ve...
9997	Fatmah Nuraini	5	2026-04-01 04:11:00	pelayanan kesehatan dim genggaman makin mudah
9998	Bahrur Rohman	2	2026-04-01 04:06:20	ini kenapa ya tadi aplikasinya masih bisa saya...
9999	nabila nurjannah	5	2026-04-01 04:05:48	bagus

10000 rows × 4 columns

Gambar 2 Dataset JKN

Analisis Variabel

Data ulasan yang telah dikumpulkan diolah melalui serangkaian tahapan untuk menghasilkan informasi yang dapat digunakan pada proses klasifikasi sentimen, meliputi preprocessing, pembobotan TF-IDF, pengelompokan menggunakan K-Means Clustering, hingga klasifikasi menggunakan Support Vector Machine (SVM). Tahap preprocessing meliputi cleaning, case folding, tokenizing, normalisasi, stopword removal, hingga stemming untuk mengurangi ketidakkonsistenan penulisan dan menghasilkan bentuk teks yang lebih seragam, sebagaimana ditampilkan pada Gambar 3



cleaning	case_folding	tokenizing	normalisasi	stopword	stemming
jelek banget aplikasinya	jelek banget aplikasinya	[jelek, banget, aplikasinya]	[jelek, banget, aplikasinya]	[jelek, banget, aplikasinya]	[jelek, banget, aplikasi]
jelek banget tidak bisa ...	jelek banget tidak bisa ...	[jelek, banget, tidak, bisa, ...]	[jelek, banget, tidak, bisa, ...]	[jelek, banget, login, username, ...]	[jelek, banget, login, username, ...]
sangat praktis mudah digunakan ...	sangat praktis mudah digunakan ...	[sangat, praktis, mudah, digunakan, ...]	[sangat, praktis, mudah, digunakan, ...]	[praktis, mudah, kantor, bpjs]	[praktis, mudah, kantor, bpjs]
gw gamau banyak omong ...	gw gamau banyak omong ...	[gw, gamau, banyak, omong, ...]	[saya, tidak mau, banyak, omong, ...]	[tidak mau, omong, singkat, pas, ...]	[tidak, mau, omong, singkat, ...]
saat daftar ada langkah ...	saat daftar ada langkah ...	[saat, daftar, ada, langkah, ...]	[saat, daftar, ada, langkah, ...]	[daftar, langkah, verifikasi, email, ...]	[daftar, langkah, verifikasi, email, ...]

Gambar 3 Hasil Data Preprocessing

Pembobotan kata pada setiap dokumen ulasan dilakukan menggunakan TF-IDF berdasarkan variabel penelitian (X1, X2, X3). Nilai TF diperoleh dari kemunculan kata pada dokumen (bernilai 1 jika muncul, 0 jika tidak), nilai DF dihitung dari jumlah dokumen yang memuat kata tersebut; nilai TF kemudian dinormalisasi untuk mengatasi perbedaan panjang dokumen. Perkalian antara TF dan IDF menghasilkan nilai TF-IDF.

Tabel 1 Pola Analisis

Dokumen	Kinerja Aplikasi (X1)	Kemudahan Penggunaan (X2)	Kualitas Layanan (X3)
D1	0.3983	0.0000	0.0000
D2	1.4432	0.0000	0.0000
…	…	…	…
D6031	1.3368	0.0000	0.0000
D6032	2.0565	0.0000	0.0000

Data hasil pola analisis yang telah diperoleh selanjutnya digunakan sebagai masukan pada proses pengelompokan menggunakan algoritma K-Means Clustering. Proses ini bertujuan mengelompokkan data ulasan ke dalam cluster berdasarkan kemiripan karakteristik pada nilai bobot yang telah dihitung. Hasil pengelompokan tersebut ditampilkan pada Tabel 2.

Tabel 2 K-Means Clustering

X1	X2	X3	Cluster	Sentimen
0.3983	0.0000	0.0000	1	Negatif
1.4432	0.0000	0.0000	1	Negatif
…	…	…	…	…

1.3368	0.0000	0.0000	1	Negatif
2.0565	0.0000	0.0000	1	Negatif

Proses iterasi berhenti pada iterasi ke-12 karena tidak lagi terjadi perpindahan data antar cluster, menandakan posisi centroid telah stabil dan algoritma K-Means Clustering mencapai konvergensi, sehingga hasil klasterisasi bersifat optimal dan dapat dijadikan dasar analisis lanjutan menggunakan SVM. Hasil klasterisasi ini juga memberikan gambaran jelas mengenai distribusi opini pengguna pada aplikasi JKN.

Hasil Support Vector Machine

Data ulasan yang telah digolongkan menggunakan K-Means Clustering menjadi proses penerapan tahap klasifikasi sentimen dengan SVM. Keunggulan dalam memisahkan data berdasarkan kelas sentimennya melatarbelakangi pemilihan SVM.

Hasil Klasifikasi SVM

	No	Label Aktual	Label Prediksi	Keterangan
0	1	Positif	Positif	Benar
1	2	Negatif	Negatif	Benar
2	3	Positif	Positif	Benar
3	4	Positif	Positif	Benar
4	5	Positif	Positif	Benar
...	...	...	...	...
1807	1808	Negatif	Negatif	Benar
1808	1809	Negatif	Negatif	Benar
1809	1810	Negatif	Negatif	Benar
1810	1811	Negatif	Negatif	Benar
1811	1812	Negatif	Negatif	Benar

Gambar 4 Klasifikasi SVM

Evaluasi

Kinerja model Support Vector Machine (SVM) dalam menggolongkan sentimen ulasan pengguna aplikasi JKN diukur melalui tahap evaluasi dalam perhitungan nilai accuracy, precision, recall, serta F1-score pada data uji.

```
Akurasi: 0.9983443708609272

Confusion Matrix:
 [[ 628    1]
 [   2 1181]]

Laporan Klasifikasi:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00       629
           1       1.00      1.00      1.00      1183

    accuracy                           1.00      1812
   macro avg       1.00      1.00      1.00      1812
weighted avg       1.00      1.00      1.00      1812
```

Gambar 5 Performa Akurasi Model SVM

99,83% ditunjukkan oleh tingkat akurasi model SVM yang tertera pada Gambar 5. Hal ini menunjukkan bahwa model mampu menggolongkan data uji dengan ketepatan yang tinggi. Kemampuan klasifikasi yang sangat baik pada model tergambar dari nilai precision, recall, serta F1-score yang mendekati 1,00. Hasil pendekatan TF-IDF dan SVM memiliki tingkat keandalan tinggi dalam mengidentifikasi kecenderungan opini pengguna pada aplikasi JKN.

SIMPULAN

Penerapan pembobotan Term Frequency–Inverse Document Frequency (TF-IDF) dan klasifikasi Support Vector Machine (SVM) pada 6.037 data ulasan pengguna aplikasi JKN yang telah melalui proses penyaringan dari 10.000 data awal memberikan hasil evaluasi yang sangat baik. Nilai accuracy yang diperoleh mencapai 99,83%, disertai precision sebesar 99,92%, recall sebesar 99,83%, dan F1-score sebesar 99,87%. Nilai keempat metrik tersebut mebuktikan bahwa kombinasi kedua metode mampu memisahkan sentimen positif dan negatif secara konsisten. Hasil ini menunjukkan bahwa pembobotan kata pada TF-IDF cukup kuat untuk membentuk batas pemisah kelas yang baik bagi SVM. tersebut ke layanan kesehatan digital publik, yang karakteristik topik dan sensitivitas datanya berbeda dari kedua domain tersebut. Hasil ini menegaskan bahwa pendekatan serupa dapat diterapkan untuk memantau sentimen pengguna pada platform layanan publik digital lain secara otomatis dan pada skala data besar.

DAFTAR PUSTAKA

Afdal, M., & Elita, L. R. (2022). *Penerapan Text Mining Pada*

Aplikasi Tokopedia. *8*(1), 78–87.

Anas, A., & Zebua, A. J. (2025). *Clustering Nilai Mahasiswa Menggunakan Algoritma K-Means*. *19*(1), 7–14.

Azzahra Livia, M. E. (2025). *Analisis Sentimen Terhadap Rsud Salatiga Menggunakan Abstrak*. *6*(1), 478–489.

Barokah, L., Lorenza, F. O., & Tyas, F. A. (2025). *Jurnal Processor Penerapan Clustering*. *20*(1), 44–57.

Butsianto, S., & Muhammad, A. (2025). *Analisis Sentimen Ulasan Aplikasi Jamsostek Dengan Svm , Random Forest , Dan Logistic Regression*. *7*, 700–706.

Damanhuri, R., & Husein, V. A. (2024). *Analisis Sentimen Pada Ulasan Aplikasi Access By Kai Berbahasa Indonesia Menggunakan Word-Embedding Dan Classical Machine Learning*. *15*.

Dewi, M. S., & Hasugian, A. H. (2025). *Mutiara Sintia Dewi, 2) Abdul Halim Hasugian*. *10*(2), 911–922.

Emarapenta, J., Sinulingga, B., Cesar, H., & Sitorus, K. (2024). *Analisis Sentimen Masyarakat Terhadap Film Horor Indonesia Menggunakan Metode Svm Dan Tf-Idf*. *14*, 42–53.

Fahmi, R. U., Arifiyanti, A. A., Luhur, T., & Sugata, I. (2025). *Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi Midi Kriing Menggunakan Support Vector Machine (Svm)*. *9*(3), 4831–4839.

Fatah, Z. F., & Maghfiroh, S. (2025). *Analisis Data Mining Dengan Algoritma K-Means Clustering Untuk Menentukan Siswa Berprestasi Di Mts Miftahul Ulum Bengkak*. *4*.

Helmayanti, S. A., Hamami, F., & Fa'rifah, R. Y. (2023). *Jurnal Indonesia : Manajemen Informatika Dan Komunikasi Penerapan Algoritma Tf-Idf Dan Naïve Bayes Jurnal Indonesia 4*(3), 1822–1834.

Husain, P., & Syam, A. F. (2024). *Analisis Sentimen Ulasan Pengguna Tiktok Pada Google Play Store Berbasis Tf-Idf Dan Support Vector Machine*. *5*(1), 91–102.

Husen, R. A., Astuti, R., Marlia, L., & Efrizoni, L. (2023). *Analisis Sentimen Opini Publik Pada Twitter Terhadap Bank Bsi Menggunakan Algoritma Machine Learning*. *3*, 211–218.

Maulana, M. T. F., Nugroho, B. I., Unggul, E., & Utami, S. (2025). *Analisis Sentimen Ulasan Aplikasi Tiktok Di Google Playstore Menggunakan Algoritma Naive Bayes*. *4*(3), 6627–6635.

Nugraha, W., & Hardjianto, M. (2024). *Optimasi Support Vector Machines (Svm) Menggunakan Particle Swarm Optimization (Pso) Pada Analisis Sentimen Ulasan Pengguna Layanan Bukalapak Di Twitter*. *4*(3), 1082–1100.

Nurian, A., Sari, B. N., Karawang, U. S., & Timur, T. (2023). *Analisis Sentimen Ulasan Pengguna Aplikasi Google Play Menggunakan Naïve Bayes*. *11*(3), 829–835.

Pertiwi, A. A., & Utomo, A. N. (2023). *Implementasi Metode Support Vector Machine(Svm) Pada Citra Fundus Retina Mata Untuk Deteksi Penyakit Glaukoma*. *12*(1), 19–27.

Sugihartono, T., Rian, R., Putra, C., Informasi, F. T., Informatika, T., Machine, S. V., Mining, T., & Pengguna, U. (2024). *Penerapan Metode Support Vector Machine Dalam Classifikasi Ulasan Pengguna Aplikasi Mobile Jkn*. *7*, 144–153.

Ulya, S., Ridwan, A., Wahyudin, W. C., & Hana, F. M. (2022). *Text Mining Sentimen A Nalisis P Engguna Aplikasi Marketplace Tokopedia Berdasar Rating An Komentar*. 33–40.

Wibowo, A. P., & Hidayat, N. (2024). *Eksplorasi Linguistik Komputasional Dalam Analisis Bahasa Alami Untuk Mengungkap Evolusi Dialek Digital Di Era Media Sosial Global*.