

ANALISIS SENTIMEN ULASAN APLIKASI SIAPKERJA MENGUNAKAN K-MEANS CLUSTERING, SUPPORT VECTOR MACHINE DAN RANDOM FOREST

Tio Febri¹, Yuhandri², Larissa Navia Rani³

¹Dinas Komunikasi dan Informatika, Pemerintah Kota Payakumbuh, Payakumbuh

^{2,3}Universitas Putra Indonesia YPTK Padang, Padang

e-mail: ¹tiofebrii@gmail.com, ²yuyu@upiyptk.ac.id, ³larissa_navia_rani@upiyptk.ac.id

Abstract: *SIAPkerja is a digital employment service platform organized by the Ministry of Manpower of the Republic of Indonesia to facilitate interactions between job seekers and employers. The increasing number of application users has resulted in a higher volume of reviews on the Google Play Store platform, reflecting diverse user experiences and satisfaction levels with the provided services. This study analyzes the sentiment of SIAPkerja user reviews, with 2,000 review data points collected through web scraping from the Google Play Store during the period from October 2022 to June 2026. The data were processed through cleaning, case folding, tokenization, normalization, stopword removal, and stemming, and then represented using TF-IDF weighting. K-Means was applied to cluster public opinions based on analytical patterns and sentiment as the basis for initial label determination, while SVM and Random Forest were used to classify these opinions into positive and negative sentiment categories. SVM achieved an accuracy of 99.63%, which was slightly higher than Random Forest. These results indicate that sentiment analysis based on text mining can be used to evaluate the quality of SIAPkerja services from the perspective of users and can serve as a consideration in formulating digital employment service policies that are more responsive to user needs.*

Keywords: *SIAPkerja, Google Play Store, Sentiment Analysis, Support Vector Machine, Random Forest.*

Abstrak: SIAPkerja merupakan platform layanan ketenagakerjaan digital yang diselenggarakan oleh Kementerian Ketenagakerjaan Republik Indonesia untuk memfasilitasi interaksi antara pencari kerja dan pemberi kerja. Peningkatan jumlah pengguna aplikasi tersebut berimplikasi pada bertambahnya volume ulasan pada platform Google Play Store yang merefleksikan variasi pengalaman serta tingkat kepuasan pengguna terhadap layanan yang disediakan. Penelitian ini menganalisis sentimen ulasan pengguna SIAPkerja sebanyak 2.000 data ulasan dikumpulkan melalui web scraping dari Google Play Store pada periode Oktober 2022 hingga Juni 2026. Data diproses melalui tahap cleaning, case folding, tokenisasi, normalisasi, stopword removal, dan stemming, kemudian direpresentasikan dengan pembobotan TF-IDF. K-Means diterapkan untuk mengelompokkan opini masyarakat berdasarkan pola analisis dan sentimen sebagai dasar penentuan label awal, sementara SVM dan Random Forest digunakan untuk mengklasifikasikan opini tersebut ke dalam kategori sentimen positif dan negatif. SVM menghasilkan akurasi 99,63%, sedikit lebih tinggi dibanding Random Forest. Hasil ini menunjukkan bahwa analisis sentimen berbasis text mining dapat digunakan untuk mengevaluasi kualitas layanan SIAPkerja berdasarkan persepsi pengguna, dan dapat menjadi bahan pertimbangan dalam penyusunan kebijakan layanan ketenagakerjaan digital yang lebih responsif terhadap kebutuhan pengguna.

Kata kunci: *SIAPkerja, Google Play Store, Analisis Sentimen, Support Vector Machine, Random Forest.*

PENDAHULUAN

Perkembangan teknologi informasi (TI) saat ini berlangsung sangat pesat dan menjadi pendorong utama transformasi digital di berbagai sektor. Kemajuan teknologi digital memungkinkan organisasi meningkatkan efisiensi proses, kualitas layanan, dan kinerja secara lebih efektif, serta mendorong inovasi teknologi yang berkelanjutan (Fang & Liu, 2024). Perkembangan ini telah menghasilkan perubahan mendasar dalam pengelolaan informasi, komunikasi, dan pelaksanaan berbagai kegiatan pelayanan oleh individu, organisasi, dan pemerintah secara lebih cepat dan terintegrasi (Mutmainna et al., 2025).

Kementerian Ketenagakerjaan Republik Indonesia (Kemnaker) telah menginisiasi transformasi digital layanan ketenagakerjaan melalui Ekosistem Digital SIAPkerja (Sistem Informasi dan Aplikasi Pelayanan Ketenagakerjaan), sebuah sistem yang mengintegrasikan layanan ketenagakerjaan secara nasional (Kemnaker, 2024). Ulasan pengguna di Google Play Store merupakan bentuk *user-generated content* (UGC) yang secara substansial mencerminkan persepsi dan penilaian pengguna terhadap kualitas layanan aplikasi (Suryaningsih et al., 2025). Pengelolaan digital terhadap ulasan pengguna di Google Play Store dapat dimanfaatkan sebagai indikator evaluasi kinerja layanan secara aktual, sehingga dapat merekam persepsi, keluhan, dan tingkat penerimaan pengguna terhadap kualitas layanan digital yang disediakan oleh pemerintah (Kelvin et al., 2025).

Berdasarkan kondisi tersebut, diperlukan pendekatan sistematis untuk mengekstraksi informasi dari data teks yang tidak terstruktur agar ulasan pengguna dapat dianalisis secara lebih objektif dan terarah (Purbaratri et al., 2024). Ulasan pengguna memiliki karakteristik bahasa informal, seperti singkatan, emotikon, dan variasi penulisan, sehingga sulit dianalisis secara manual dalam jumlah besar (Suhdi &

Siswoyo, 2026). Data tersebut dapat dimanfaatkan sebagai sumber informasi untuk mengevaluasi kualitas layanan dan mendukung perbaikan sistem berbasis bukti (Memon et al., 2025).

Natural Language Processing (NLP) berperan dalam mengubah data teks mentah menjadi informasi yang terstruktur melalui tahapan preprocessing, tokenisasi, normalisasi, ekstraksi fitur, dan pembobotan kata (Bimantara & Zufria, 2024). Pendekatan NLP dan machine learning terbukti relevan untuk mengklasifikasikan ulasan pengguna Google Play Store sehingga hasilnya dapat digunakan sebagai dasar evaluasi kualitas layanan digital pemerintah maupun aplikasi publik lainnya (Rizki et al., 2025).

Analisis sentimen terhadap ulasan pengguna Google Play Store umumnya mengikuti alur baku: data ulasan dikumpulkan melalui scraping, diproses melalui tahap preprocessing teks, dibobotkan menggunakan metode seperti TF-IDF, kemudian diklasifikasikan sentimennya dengan algoritma machine learning (Ikhsan et al., 2025). Metode Term Frequency-Inverse Document Frequency (TF-IDF) dapat digunakan untuk mengubah kata-kata dalam ulasan menjadi bobot numerik yang merepresentasikan tingkat kepentingan suatu kata dalam dokumen (Rizkiyanto et al., 2024). Representasi numerik pada teks ulasan yang sebelumnya bersifat bebas dapat diproses oleh algoritma machine learning untuk menemukan pola sentimen secara lebih sistematis (Barus et al., 2025). Setelah fitur teks terbentuk, algoritma klasifikasi seperti Support Vector Machine, Random Forest, Naïve Bayes, atau K-Nearest Neighbors dapat digunakan untuk mengelompokkan ulasan ke dalam kategori sentimen positif atau negatif (Madyatmadja et al., 2025).

Penelitian terdahulu menggunakan K-Means Clustering untuk pelabelan data dan Support Vector Machine (SVM) untuk klasifikasi sentimen. Dataset yang digunakan terdiri dari 647 ulasan Twitter terkait Karapan Sapi. Hasil penelitian

menunjukkan bahwa metode tersebut memperoleh akurasi sebesar 92%, meskipun masih terdapat keterbatasan dalam mengenali kelas sentimen minoritas berdasarkan nilai *recall* yang rendah (Rani & Khotimah, 2025).

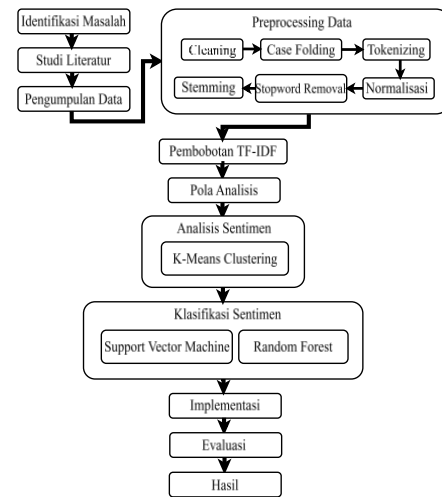
Berdasarkan uraian tersebut, penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi SIAPkerja pada Google Play Store menggunakan algoritma SVM dan Random Forest, melalui tahapan preprocessing teks dan pembobotan fitur dengan TF-IDF, untuk memperoleh gambaran objektif mengenai persepsi pengguna terhadap kualitas layanan aplikasi. Selain melakukan klasifikasi sentimen, penelitian ini juga membandingkan kinerja SVM dan Random Forest berdasarkan metrik evaluasi yang sama, sehingga dapat diketahui algoritma yang paling optimal dalam mengidentifikasi sentimen pengguna terhadap aplikasi SIAPkerja.

Hasil penelitian ini diharapkan berkontribusi bagi pengembangan studi analisis sentimen pada layanan publik digital, sekaligus menjadi bahan evaluasi bagi pengelola aplikasi dalam meningkatkan kualitas layanan ketenagakerjaan yang lebih responsif dan sesuai kebutuhan pengguna.

METODE

Bagian ini menjelaskan tahapan penelitian meliputi dengan pengumpulan data dan prapemrosesan teks melalui cleaning, case folding, tokenizing, normalisasi, stopword removal, dan stemming. Hasil prapemrosesan tersebut diekstraksi menggunakan TF-IDF, kemudian dianalisis polanya dengan K-Means Clustering sebelum diklasifikasikan menggunakan SVM dan Random Forest. Performa kedua algoritma klasifikasi diukur menggunakan empat metrik evaluasi, yaitu akurasi, presisi, recall, dan F1-score, untuk mengetahui seberapa tepat masing-masing model memprediksi kelas sentimen.

Keseluruhan tahapan penelitian, mulai dari identifikasi masalah hingga implementasi model, disajikan secara sistematis pada Gambar 1.



Gambar 1. Kerangka Penelitian

Text Mining

Text mining merupakan teknik analisis data yang digunakan untuk mengekstraksi informasi penting dari kumpulan data teks yang tidak terstruktur, seperti komentar, ulasan pengguna, dan dokumen digital (Rhomaningtias et al., 2025). Teknik ini memanfaatkan kombinasi Natural Language Processing (NLP) dan machine learning untuk mengolah serta mengklasifikasikan teks secara otomatis sesuai kebutuhan analisis (Bifari et al., 2024).

Selain sebagai metode pengolahan teks, *text mining* juga berperan dalam mengubah data tidak terstruktur menjadi informasi terorganisasi yang dapat mendukung pengambilan keputusan berbasis data (Bae et al., 2025).

Pembobotan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan salah satu metode ekstraksi fitur yang paling banyak digunakan dalam bidang Natural Language Processing (NLP) dan Information Retrieval untuk merepresentasikan dokumen teks ke dalam bentuk numerik (Agustin et al., 2026).

Berikut impementasi TF-IDF dapat digunakan pada persamaan berikut

$$W_{dt} = TF_{dt} \times IDF_{ft} \quad (1)$$

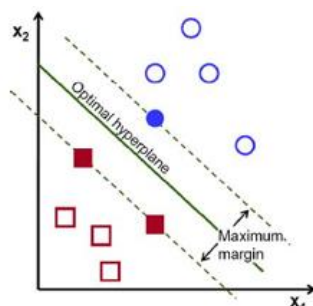
$$TF_{Normalisasi} = TF / d \quad (2)$$

$$idf_{ft} = \log (N / df_t) \quad (3)$$

W_{dt} menyatakan bobot dokumen ke-d terhadap kata ke-t, yaitu nilai TF-IDF akhir yang dihasilkan. TF_{dt} adalah Term Frequency, yaitu jumlah kemunculan suatu kata dalam dokumen tertentu. IDF_{ft} adalah Inverse Document Frequency, yaitu nilai kebalikan dari frekuensi dokumen yang mengukur seberapa penting suatu kata di seluruh kumpulan dokumen. DF_t adalah Document Frequency, yaitu jumlah dokumen yang mengandung kata tertentu. N adalah jumlah total dokumen dalam korpus penelitian. $TF_{Normalisasi}$ merupakan hasil pembagian nilai TF dengan jumlah total kata dalam dokumen, digunakan untuk menormalkan bobot kata agar tidak bias terhadap panjang dokumen. Adapun d menunjukkan dokumen tempat kata tersebut berada.

Metode Support Vector Machine

Support Vector Machine (SVM) merupakan algoritma supervised learning yang digunakan untuk klasifikasi dan regresi dengan membentuk hyperplane optimal sebagai batas pemisah antar kelas data (Rianti & Andarsyah, 2024). Ilustrasi konseptual mengenai pembentukan hyperplane pada SVM disajikan pada Gambar 2.



Gambar 2. Ilustrasi Hyperplane

Margin didefinisikan sebagai jarak terpendek antara hyperplane dengan titik data terdekat dari masing-masing kelas.

Jarak dari hyperplane menuju batas margin pada kelas positif sama dengan jarak menuju batas margin pada kelas negatif, sehingga kedua batas margin bersifat sejajar terhadap hyperplane. Titik-titik data yang berada tepat pada batas margin disebut sebagai support vector, yaitu data yang memiliki pengaruh paling besar dalam menentukan posisi hyperplane optimal.

Random Forest

Algoritma ini membangun sejumlah pohon keputusan (*decision tree*) menggunakan sampel data yang diperoleh melalui *bootstrap sampling* (Firnanda et al., 2025). Seluruh hasil prediksi pohon kemudian digabungkan melalui mekanisme *majority voting*. Pendekatan tersebut memungkinkan model menghasilkan kinerja klasifikasi yang lebih stabil dibandingkan dengan penggunaan satu pohon keputusan (Hilmi et al., 2025). Prediksi akhir *Random Forest* dihitung menggunakan Persamaan (4).

$$\hat{y} = \text{mode}(T_1(x), T_2(x), \dots, T_n(x)) \quad (4)$$

Pada Persamaan (4), $\hat{y}$ merupakan hasil prediksi akhir model *Random Forest*; $T_n(x)$ merupakan prediksi yang dihasilkan oleh pohon keputusan ke- n terhadap data masukan x ; sedangkan n menunjukkan jumlah pohon keputusan yang digunakan dalam model.

HASIL DAN PEMBAHASAN

Dataset

Data penelitian berupa ulasan pengguna aplikasi SIAPkerja yang diperoleh melalui teknik web scraping pada Google Play Store. Pengumpulan data dilakukan secara otomatis menggunakan skrip Python di Google Colab untuk memperoleh data secara efisien dan sistematis. Data yang dikumpulkan berjumlah 2.000 ulasan pengguna yang diambil dalam rentang waktu Oktober 2022 hingga Juni 2026. Penyaringan data berdasarkan data dilakukan berdasarkan tiga kategori yaitu

Performa Aplikasi (X1), Kemudahan Penggunaan (X2), dan Kualitas Layanan (X3). Proses ini, setelah melalui tahap prapemrosesan dan penyaringan sesuai kriteria penelitian, menyisakan 1.816 ulasan yang layak dianalisis dari total data awal. Dataset hasil seleksi ini disimpan dalam format CSV untuk digunakan pada tahap analisis sentimen berikutnya.

Preprocessing

Preprocessing atau prapemrosesan data merupakan tahapan penting dalam

analisis sentimen karena kualitas teks yang digunakan akan memengaruhi pembentukan fitur dan hasil pemodelan. Tahap ini bertujuan untuk mengubah ulasan pengguna yang masih tidak terstruktur menjadi data teks yang lebih bersih, seragam, dan siap dianalisis. Prapemrosesan pada penelitian ini meliputi *cleaning*, *case folding*, *tokenizing*, *normalisasi*, *stopword*, dan *stemming*. Berikut hasil

cleaning	case_folding	tokenizing	normalisasi	stopword	stemming
Setiap aku mau klik daftar pelatihan pasti di oper ke chrome di chrome juga sama aja	setiap aku mau klik daftar pelatihan pasti di oper ke chrome di chrome juga sama aja	['setiap', 'aku', 'mau', 'klik', 'daftar', 'pelatihan', 'pasti', 'di', 'oper', 'ke', 'chrome', 'di', 'chrome', 'juga', 'sama', 'aja']	['setiap', 'aku', 'mau', 'klik', 'daftar', 'pelatihan', 'pasti', 'di', 'oper', 'ke', 'chrome', 'di', 'chrome', 'juga', 'sama', 'aja']	['klik', 'daftar', 'pelatihan', 'oper', 'chrome', 'chrome']	['klik', 'daftar', 'latih', 'oper', 'chrome', 'chrome']
banyak banget error nya nih aplikasi Katanya aplikasi dari pemerintah tapi Banyak errornya Bukannya mempermudah malah mempersulit orang Bikin AK di suruh kesini sana katanya lebih mudah buat online tapi nyatanya nihil	banyak banget error nya nih aplikasi Katanya aplikasi dari pemerintah tapi Banyak errornya Bukannya mempermudah malah mempersulit orang bikin ak di suruh kesini sana katanya lebih mudah buat online tapi nyatanya nihil	['banyak', 'banget', 'error', 'nya', 'nih', 'aplikasi', 'katanya', 'aplikasi', 'dari', 'pemerintah', 'tapi', 'banyak', 'errornya', 'bukannya', 'mempermudah', 'malah', 'mempersulit', 'orang', 'bikin', 'ak', 'di', 'suruh', 'kesini', 'sana', 'katanya', 'lebih', 'mudah', 'buat', 'online', 'tapi', 'nyatanya', 'nihil']	['banyak', 'banget', 'error', 'nya', 'nih', 'aplikasi', 'katanya', 'aplikasi', 'dari', 'pemerintah', 'tapi', 'banyak', 'erromya', 'bukannya', 'mempermudah', 'bahkan', 'mempersulit', 'orang', 'bikin', 'aku', 'di', 'suruh', 'kesini', 'sana', 'katanya', 'lebih', 'mudah', 'buat', 'online', 'tapi', 'nyatanya', 'nihil']	['banget', 'error', 'aplikasi', 'aplikasi', 'pemerintah', 'erromya', 'mempermudah', 'mempersulit', 'orang', 'bikin', 'suruh', 'kesini', 'mudah', 'online', 'nihil']	['banget', 'error', 'aplikasi', 'aplikasi', 'perintah', 'erromya', 'mudah', 'sulit', 'orang', 'bikin', 'suruh', 'kesini', 'mudah', 'online', 'nihil']
Terimakasih berkat aplikasi ini sdh membantu apa yg saya cari smg lebih baik lagi kedepannya	terimakasih berkat aplikasi ini sdh membantu apa yg saya cari smg lebih baik lagi kedepannya	['terimakasih', 'berkat', 'aplikasi', 'ini', 'sdh', 'membantu', 'apa', 'yg', 'saya', 'cari', 'smg', 'lebih', 'baik', 'lagi', 'ke depannya']	['terima', 'kasih', 'berkat', 'aplikasi', 'ini', 'sudah', 'membantu', 'apa', 'yang', 'saya', 'cari', 'semoga', 'lebih', 'baik', 'lagi', 'ke depannya']	['terima', 'kasih', 'berkat', 'aplikasi', 'membantu', 'cari', 'semoga', 'ke depannya']	['terima', 'kasih', 'berkat', 'aplikasi', 'bantu', 'cari', 'moga', 'depan']

Gambar 4. Hasil Data Preprocessing

Gambar 4 memperlihatkan tahapan prapemrosesan ulasan pengguna aplikasi SIAPKerja yang diperoleh dari Google Play Store.

Pola Analisis

Tahap selanjutnya dilakukan ekstraksi fitur dengan memberikan bobot pada setiap kata dalam dokumen ulasan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). Proses pembobotan ini didasarkan pada tiga variabel penelitian, yaitu performa aplikasi (X1), kemudahan penggunaan (X2), dan kualitas layanan (X3).

Hasil perhitungan TF-IDF menghasilkan nilai bobot yang

merepresentasikan tingkat kepentingan suatu kata dalam dokumen. Semakin tinggi nilai TF-IDF, semakin besar kontribusi kata tersebut dalam menggambarkan karakteristik ulasan, sedangkan nilai nol menunjukkan bahwa fitur tidak terdapat pada dokumen yang dianalisis. Matriks fitur hasil pembobotan TF-IDF kemudian digunakan sebagai representasi numerik data teks untuk mengidentifikasi pola dan mendukung proses analisis sentimen pada tahap berikutnya. Contoh hasil pembobotan TF-IDF untuk setiap dokumen disajikan pada Tabel 1.

Tabel 1. Pola Analisis

Dokumen	Performa Aplikasi (X1)	Kemudahan Pengguna (X2)	Kualitas Layanan (X3)
D1	0,7952	0,0000	0,0000
D2	0,7238	1,0567	1,8789
D3	0,2173	0,0000	0,6956
...	...	...	...
D1814	1,3400	0,0000	0,6956
D1815	0,6026	2,2534	0,0000

Dokumen	Performa Aplikasi (X1)	Kemudahan Pengguna (X2)	Kualitas Layanan (X3)
D1816	1,4687	0,0000	2,0072

Analisis Sentimen

Pada tahap awal, data ulasan dikelompokkan menggunakan K-Means Clustering berdasarkan tingkat kemiripan fitur untuk membentuk kelompok sentimen sekaligus menghasilkan pelabelan awal.

Tahap pertama analisis dilakukan dengan menerapkan algoritma K-Means Clustering pada data pola analisis yang telah direpresentasikan dalam bentuk fitur numerik. Jumlah kluster ditetapkan sebanyak dua ($k = 2$), yang merepresentasikan kelompok sentimen positif dan negatif.

Tabel 2. Hasil K-Means Clustering

Dok.	X1	X2	X3	Cluster	Sentimen
D1	0,795223	0,000000	0,000000	1	Negatif
D2	0,723768	1,056653	1,878905	0	Positif
D3	0,217329	0,000000	0,695635	0	Positif
...	...	...	...	...	...
D1815	0,602591	2,253438	0,000000	1	Negatif
D1816	1,468724	0,000000	2,007221	0	Positif
D1812	1,502277	1,913133	0,000000	1	Negatif

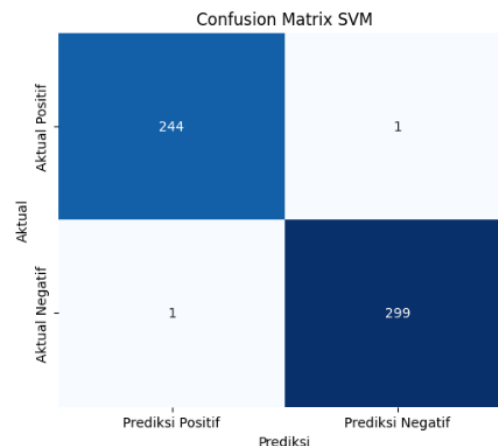
Tabel 2 menyajikan nilai centroid hasil penerapan K-Means Clustering pada data yang telah distandardisasi, mencakup tiga variabel penelitian: performa aplikasi (X1), kemudahan penggunaan (X2), dan kualitas layanan (X3). Pada iterasi ke-15, sebanyak 1.816 ulasan pengguna SIAPkerja terbagi ke dalam dua kluster dengan karakteristik skor yang berbeda. Kluster 0 berisi 818 ulasan dan diinterpretasikan sebagai sentimen positif, sedangkan Kluster 1 berisi 998 ulasan dan dipetakan sebagai sentimen negatif.

Hasil pengelompokan ini menjadi acuan label dalam klasifikasi sentimen menggunakan SVM dan Random Forest. Proporsi kelas positif dan negatif pada data latih maupun data uji tetap seimbang. Pembagian data dengan rasio 70:30 menghasilkan 1.271 ulasan sebagai data latih dan 545 ulasan sebagai data uji. Kedua model menggunakan skema pembagian data yang identik, sehingga dilatih dan diuji pada himpunan data yang setara dan hasil kinerjanya dapat dibandingkan secara adil.

Support Vector Machine

Model Support Vector Machine (SVM) diuji menggunakan 545 data uji untuk mengukur kinerja klasifikasi

sentimen secara kuantitatif. Confusion matrix hasil pengujian disajikan pada Gambar 5.



Gambar 5. Confusion Matrix SVM

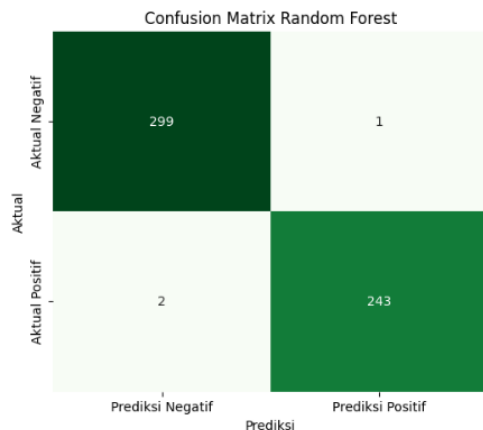
Pengujian terhadap 545 data uji menghasilkan akurasi 99,63%, dengan SVM hanya keliru mengklasifikasikan dua data dari keseluruhan data uji. Hasil ini menunjukkan fitur hasil pembobotan TF-IDF cukup representatif terhadap pola kata pada ulasan pengguna SIAPkerja.

Kelas Positif memperoleh precision, recall, dan F1-score sebesar 0,996, sementara kelas Negatif sedikit lebih tinggi dengan 0,997 pada ketiga metrik tersebut. Macro average dan weighted average yang sama-sama berada

di 0,996 menegaskan performa model tetap konsisten di kedua kelas, meski data uji tidak seimbang.

Random Forest

Model Random Forest yang telah dibangun selanjutnya dievaluasi menggunakan data uji untuk mengukur kemampuannya dalam mengklasifikasikan sentimen positif dan negatif. Visualisasi hasil evaluasi model Random Forest disajikan pada gambar berikut.



Gambar 6. Confusion Matrix Random Forest

Berdasarkan Gambar 6, dari 545 data uji, Random Forest berhasil mengklasifikasikan 542 data dengan benar dan menghasilkan tiga kesalahan klasifikasi. Kelas Positif memperoleh *precision* 0,996, *recall* 0,992, dan *F1-score* 0,994, sedangkan kelas Negatif memperoleh *precision* 0,993, *recall* 0,997, dan *F1-score* 0,995. Secara keseluruhan, model mencapai akurasi 99,45%, dengan performa yang relatif seimbang pada kedua kelas, meskipun masih sedikit di bawah SVM yang mencapai 99,63%. Random Forest terbukti mampu mempertahankan konsistensi klasifikasi meskipun kinerjanya sedikit di bawah model SVM yang mencapai akurasi 99,63% pada data uji yang sama.

SIMPULAN

Hasil penelitian menunjukkan bahwa analisis sentimen terhadap ulasan

pengguna aplikasi SIAPkerja dapat dilakukan melalui integrasi metode K-Means Clustering serta algoritma Support Vector Machine (SVM) dan Random Forest sebagai model klasifikasi. Hasil evaluasi menunjukkan bahwa model SVM mencapai akurasi sebesar 99,63%, sedangkan Random Forest memperoleh akurasi sebesar 99,45%. Temuan tersebut mengindikasikan bahwa kedua algoritma memiliki kemampuan yang sangat baik dalam mereproduksi label sentimen yang dihasilkan melalui proses klusterisasi.

Penerapan K-Means Clustering pada tahap awal, yang dilanjutkan dengan penggunaan SVM dan Random Forest sebagai model pembandingan, memberikan pendekatan alternatif dalam mengolah data ulasan tidak terstruktur dalam jumlah besar. Pendekatan ini juga memungkinkan dilakukannya perbandingan performa antarmodel pada kondisi data dan label yang sama. Secara praktis, hasil penelitian mampu memberikan gambaran awal mengenai pola apresiasi dan keluhan pengguna berdasarkan tiga aspek utama, yaitu performa aplikasi, kemudahan penggunaan, dan kualitas layanan. Hasil tersebut dapat dimanfaatkan sebagai bahan evaluasi bagi Kementerian Ketenagakerjaan Republik Indonesia dalam meningkatkan mutu layanan aplikasi SIAPkerja.

DAFTAR PUSTAKA

- Agustin, S., Syaifullah, W., & Rizaldy Pratama, A. (2026). *PERBANDINGAN TF-IDF DAN BM25 UNTUK FEATURE EXTRACTION PADA DATA ULASAN BUKU*. 3(1).
- Ayu Firnanda, P., Shofwatillah, L., Rahma, F., Fauzi, F., Studi Statistika, P., Muhammadiyah Semarang, U., Kedungmundu No, J., Tembalang, K., & Tengah, J. (2025). Analisis Perbandingan Decision Tree dan Random Forest dalam Klasifikasi Penjualan Produk pada Supermarket.

- Emerging Statistics and Data Science Journal*, 3(1).
- Bae, S., Jeon, M., & Moon, H. R. (2025). Text mining in MOF research: from manual curation to large language model-based automation. In *Chemical Communications* (Vol. 61, Number 60, pp. 11083–11094). Royal Society of Chemistry.
- Barus, H., Fajri, I. N., & Pristyanto, Y. (2025). Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Number 5). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Bifari, E., Basbrain, A., Mirza, R., Bafail, A., Albaeadie, S., & Alhalabi, W. (2024). Text mining and machine learning for crime classification: using unstructured narrative court documents in police academic. *Cogent Engineering*, 11(1). <https://doi.org/10.1080/23311916.2024.2359850>
- Bimantara, M. D., & Zufria, I. (2024). Text Mining Sentiment Analysis on Mobile Banking Application Reviews using TF-IDF Method with Natural Language Processing Approach. *JINAV: Journal of Information and Visualization*, 5(1), 115–123. <https://doi.org/10.35877/454ri.jinav2772>
- Fang, X., & Liu, M. (2024). How does the digital transformation drive digital technology innovation of enterprises? Evidence from enterprise's digital patents. *Technological Forecasting and Social Change*, 204, 123428. <https://doi.org/10.1016/J.TECHFOR.2024.123428>
- Hilmi, F., Taqiyassar, K., Romero, N., Pratama, P., Kusuma, S. C., Nurwachid, H. R., Fatyanosa, T. N., & Korespondensi, P. (2025). ANALISIS PERBANDINGAN MODEL MACHINE LEARNING TREE-BASED DAN NON-TREE-BASED UNTUK TUGAS KLASIFIKASI COMPARATIVE ANALYSIS OF TREE-BASED AND NON-TRE-BASED MACHINE LEARNING MODELS FOR CLASSIFICATION TASKS. 12(4), 875–884.
- Ikhsan, M. D., Huda, B., Hananto, A., & Nurapriani, F. (2025). Analisis Sentimen Ulasan Pengguna Alikasi Traveloka Pada Google Play Store Menggunakan Algoritma Naive Bayes. *Infotek: Jurnal Informatika Dan Teknologi*, 8(2), 367–377. <https://doi.org/10.29408/jit.v8i2.30444>
- Kelvin H., Erlin, Yenny Desnelita, & Dwi Oktarina. (2025). Sentiment Analysis of IKD Application Reviews on Play Store Using Random Forest. *Jurnal Nasional Teknik Elektro Dan Teknologi Informasi*, 14(3), 171–180. <https://doi.org/10.22146/jnteti.v14i3.20473>
- Kementerian Ketenagakerjaan Republik Indonesia. (2024). *Peraturan Menteri Ketenagakerjaan Republik Indonesia Nomor 17 Tahun 2024 tentang Sistem Informasi Ketenagakerjaan*. https://jdih.kemnaker.go.id/asset/data_puu/2024pmnaker017.pdf
- Madyatmadja, E. D., Winata, B. C. S., Pradhan, E., Yasmina, F. P., Adrian, F. A., Mahardhika, R., Christian, & Sembiring, D. J. M. (2025). Sentiment Analysis on User Reviews of Snapchat in Indonesia. *Journal of Computer Science*, 21(1), 158–167. <https://doi.org/10.3844/jcssp.2025.158.167>
- Memon, S. A., Mahar, M. A., Kehar, A., Oad, S., & Ahmed, I. (2025). User Experience Enhancement Through Sentiment Analysis: A Machine Learning Approach to App Reviews. *International Journal of Information Systems and Computer Technologies (IJISCT)*, 4(2), 2791–3643.

- <https://doi.org/https://doi.org/10.58325/ijisct.004.02.00131>
- Mutmainna, N., Studi, P., Keuangan, M., & Publik, S. (2025). *Inovasi Pelayanan Publik Berbasis Digital dalam Meningkatkan Efektivitas Pelayanan Pemerintah kepada Masyarakat Article History* (Vol. 1, Number 1). <https://journal.uwais.ac.id/index.php/amanah/article/view/11>
- Purbaratri, W., Purnomo, H. D., Manongga, D., Setyawan, I., & Hendry, H. (2024). Sentiment Analysis of e-Government Service Using the Naive Bayes Algorithm. *MATRIK: Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 23(2), 441–452. <https://doi.org/10.30812/matrik.v23i2.3272>
- Rani, Z., & Khotimah, B. K. (2025). ANALISIS SENTIMEN TERHADAP KARAPAN SAPI DI TWITTER MENGGUNAKAN METODE K-MEANS DAN SUPPORT VECTOR MACHINE (SVM). *Jurnal Informatika Dan Teknik Elektro Terapan*, 13(1). <https://doi.org/10.23960/jitet.v13i1.5685>
- Rhomaningtias, L., Khairunisa, A., Shella, S., Wara, M., & Hindrayani, K. M. (2025). *ANALISIS SENTIMEN ULASAN APLIKASI SMILE INDONESIA MENGGUNAKAN METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE (SVM)*. [https://doi.org/10.52972/hoaq.vol16n](https://doi.org/10.52972/hoaq.vol16n25/ijisct.004.02.00131)
- Rianti, R., & Andarsyah, R. (2024). Memprediksi Tingkat Atrisi Karyawan Menggunakan Machine Learning. *Jurnal Tekno Insentif*, 18(1), 39–52. <https://doi.org/10.36787/jti.v18i1.126>
- Rizki, A., Shafira, D., Syaiful Khabib, M., Rahmayuna, N., & Gayuh Utomo, V. (2025). *Klasifikasi Sentimen Ulasan Pengguna Aplikasi Layanan Publik Google Play Store Menggunakan NLP dan ML*. 20(1), 51–64. <https://doi.org/https://doi.org/10.33365/jtk.v20i1.586>
- Rizkiyanto, M. D., Purbolaksono, M. D., & Astuti, W. (2024). Sentiment Analysis Classification on PLN Mobile Application Reviews using Random Forest Method and TF-IDF Feature Extraction. *INTEK: Jurnal Penelitian*, 11(1), 37–43. <https://doi.org/10.31963/intek.v11i1.4774>
- Suhdi, & Siswoyo, A. D. (2026). Analisis Sentimen Berbasis Aspek dan Visualisasi Business Intelligence Pada Ulasan Pelanggan Produk Minuman Industri F&B Dari Instagram dan Google Review. *Karapan Network Journal*, 2(01). <https://doi.org/10.20473/KNJ.2.01.54-64>
- Suryaningsih, S. D., Andrinaldo, A., Hartati, D., & Kalsum, U. (2025). *Pengalaman Pengguna dan Praktik Pengelolaan Keuangan Digital Pada Aplikasi Seabank: Studi Kualitatif Terhadap Ulasan di Google Play Store*. 02(04).