JURNAL MEDIA AKADEMIK (JMA)
Vol.4, No.2 Februari 2026
e-ISSN: 3031-5220; DOI: 10.62281, Hal XX-XX
PT. Media Akademik Publisher
AHU-084213.AH.01.30.Tahun 2023

TEXT MINING UNTUK KLASIFIKASI SENTIMEN ULASAN APLIKASI MOBILE JKN

Oleh:

Faisal Akhmad

Universitas Bina Bangsa
Alamat: Jl. Raya Serang - Jkt No.KM. 03 No. 1B, Panancangan, Kec. Cipocok Jaya, Kota Serang, Banten (42124).
Korespondensi Penulis: faisal.akhmad@binabangsa.ac.id.

Abstract. *The Mobile JKN (National Health Insurance) application is an important digital platform in Indonesia's healthcare services. To improve service quality, sentiment analysis of user reviews on Google Play Store becomes increasingly relevant. This research uses Support Vector Machine (SVM) algorithm to classify user review sentiments of Mobile JKN application into three categories: positive, negative, and neutral. The research dataset consists of 5,432 reviews collected from Google Play Store during January to December 2024. The preprocessing stage includes tokenization, stemming, and stopword removal using Indonesian Language Natural Language Toolkit (NLTK) library. Feature extraction was performed using TF-IDF (Term Frequency-Inverse Document Frequency) with linear and polynomial kernels. Results show that SVM with linear kernel achieves 87.65% accuracy with precision 0.884 and recall 0.876, outperforming polynomial kernel which achieves 85.32%. This study identifies that user complaints are primarily related to data synchronization delays (32%), user interface navigation difficulties (28%), and technical issues (25%), while positive aspects focus on ease of access (41%) and feature completeness (38%). Insights from this research can serve as the basis for continuous improvement of the Mobile JKN application and user satisfaction enhancement strategies. The SVM method proves effective for sentiment analysis in the context of Indonesian digital healthcare applications.*

Keywords: *Sentiment Analysis, Support Vector Machine, JKN Application, Natural Language Processing, Google Play Store, Text Classification.*

Received January 13, 2026; Revised January 26, 2026; February 20, 2026
*Corresponding author: *faisal.akhmad@binabangsa.ac.id*

Abstrak. Aplikasi Mobile JKN (Jaminan Kesehatan Nasional) merupakan platform digital yang penting dalam pelayanan kesehatan di Indonesia. Untuk meningkatkan kualitas layanan, analisis sentimen terhadap ulasan pengguna di Google Play Store menjadi sangat relevan. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) untuk mengklasifikasikan sentimen ulasan pengguna aplikasi Mobile JKN menjadi tiga kategori: positif, negatif, dan netral. Dataset penelitian terdiri dari 5.432 ulasan yang diambil dari Google Play Store selama periode Januari hingga Desember 2024. Tahap *preprocessing* meliputi tokenisasi, stemming, dan penghilangan stopwords menggunakan pustaka Natural Language Toolkit (NLTK) berbahasa Indonesia. Ekstraksi fitur dilakukan menggunakan TF-IDF (Term Frequency-Inverse Document Frequency) dengan kernel linear dan polynomial. Hasil penelitian menunjukkan bahwa SVM dengan kernel linear mencapai akurasi 87,65% dengan precision 0,884 dan recall 0,876, outperforming kernel polynomial yang mencapai 85,32%. Studi ini mengidentifikasi bahwa keluhan utama pengguna berkaitan dengan keterlambatan sinkronisasi data (32%), kesulitan navigasi antarmuka (28%), dan gangguan teknis (25%), sementara aspek positif fokus pada kemudahan akses (41%) dan kelengkapan fitur (38%). Insight dari penelitian ini dapat menjadi dasar untuk perbaikan berkelanjutan aplikasi Mobile JKN dan strategi peningkatan user satisfaction. Metode SVM terbukti efektif untuk analisis sentimen dalam konteks aplikasi kesehatan digital Indonesia.

Kata Kunci: Analisis Sentimen, *Support Vector Machine*, Aplikasi JKN, *Natural Language Processing*, Google Play Store, Klasifikasi Teks.

LATAR BELAKANG

Transformasi digital di sektor kesehatan menjadi salah satu pendorong utama peningkatan kualitas layanan publik melalui penyederhanaan proses, perluasan akses, dan ketersediaan informasi secara cepat. Dalam konteks penyelenggaraan JKN, aplikasi Mobile JKN berperan sebagai kanal layanan yang memungkinkan peserta mengakses beragam informasi terkait program JKN secara cepat dan mudah, kapan pun dan di mana pun, serta menyediakan fitur pencarian data kepesertaan (nomor kartu/NIK/KK), informasi lokasi fasilitas kesehatan (peta lokasi faskes), informasi iuran dan virtual account, hingga dukungan layanan lain yang terus dikembangkan (Aji et al., n.d.).

Walaupun demikian, implementasi layanan digital berbasis aplikasi mobile umumnya menghadapi tantangan yang dinamis. Tantangan tersebut meliputi aspek kinerja aplikasi (misalnya waktu muat, respons antarmuka), stabilitas sistem (misalnya crash/error), serta ketepatan sinkronisasi data antara aplikasi pengguna dan server. Selain itu, perubahan fitur dan pembaruan versi aplikasi dapat mengubah pola penggunaan dan persepsi pengguna, sehingga kualitas layanan perlu dipantau secara berkelanjutan (He et al., 2022). Pada skala layanan publik, penurunan pengalaman pengguna tidak hanya berdampak pada kepuasan, tetapi juga dapat memengaruhi kelancaran pemanfaatan layanan kesehatan digital oleh masyarakat (Albert et al., 2025).

Umpan balik pengguna yang paling mudah diakses dan bernilai tinggi sering kali muncul pada platform distribusi aplikasi seperti Google Play Store dalam bentuk ulasan (reviews) dan rating. Ulasan ini mencerminkan pengalaman penggunaan aktual, termasuk keluhan, saran perbaikan, serta apresiasi terhadap fitur yang dianggap membantu (Hedderich et al., 2021). Namun, volume ulasan yang besar dan terus bertambah membuat analisis manual sulit dilakukan secara cepat dan konsisten. Selain itu, ulasan berbahasa Indonesia memiliki variasi yang kompleks seperti ejaan tidak baku, penggunaan singkatan, istilah domain (misalnya istilah administrasi dan kepesertaan), campuran bahasa, serta adanya opini campuran (positif dan negatif) dalam satu ulasan. Kondisi tersebut menuntut adanya pendekatan otomatis agar informasi utama dapat diekstraksi dan diringkas secara sistematis (Brown et al., n.d.).

Analisis sentimen merupakan salah satu pendekatan dalam Natural Language Processing (NLP) yang bertujuan mengidentifikasi orientasi opini pengguna (misalnya positif, negatif, atau netral) pada teks. Dalam konteks aplikasi layanan publik, analisis sentimen dapat dipakai untuk memetakan tingkat kepuasan dan ketidakpuasan pengguna dalam periode tertentu, mengidentifikasi area prioritas perbaikan berdasarkan dominasi sentimen negatif, serta mengevaluasi dampak pembaruan aplikasi atau kebijakan layanan terhadap persepsi pengguna (Joshi et al., 2020). Dengan demikian, hasil analisis sentimen tidak hanya menjadi ringkasan opini, tetapi juga dapat berfungsi sebagai bahan masukan berbasis data untuk pengambilan keputusan dan perencanaan pengembangan aplikasi (Cahyawijaya et al., 2022).

Berbagai metode klasifikasi teks dapat digunakan untuk analisis sentimen, salah satu yang banyak digunakan adalah Support Vector Machine (SVM). SVM dikenal

efektif untuk kategorisasi teks karena mampu bekerja baik pada ruang fitur berdimensi tinggi (seperti TF-IDF) dan memiliki kemampuan generalisasi yang baik melalui prinsip pemaksimalan margin. Kajian klasik di bidang kategorisasi teks menyatakan bahwa SVM merupakan metode yang sesuai untuk klasifikasi teks, menunjukkan performa yang kuat, dan bersifat robust pada berbagai tugas pembelajaran (Dodge et al., 2019).

Meskipun pendekatan pembelajaran mendalam semakin berkembang, SVM masih relevan untuk penelitian yang membutuhkan baseline kuat, interpretabilitas pipeline yang lebih jelas, serta kebutuhan komputasi yang relatif lebih ringan. Dalam penelitian ini, SVM dipadukan dengan representasi fitur TF-IDF untuk membangun model klasifikasi sentimen ulasan Mobile JKN menjadi tiga kelas (positif, negatif, netral). Pemilihan tiga kelas ini bertujuan menangkap ulasan yang benar-benar menyatakan kepuasan (positif), ketidakpuasan/keluhan (negatif), maupun ulasan yang cenderung informasional atau ambigu (netral) (Fauzi et al., 2025).

Berdasarkan uraian tersebut, penelitian ini penting dilakukan karena menyediakan pendekatan terukur untuk memahami persepsi pengguna terhadap Mobile JKN melalui ulasan Google Play Store. Hasil penelitian diharapkan dapat membantu pihak pengembang dan pemangku kepentingan dalam memprioritaskan perbaikan, terutama pada isu-isu yang dominan dikeluhkan pengguna, sekaligus mempertahankan aspek-aspek yang sudah dinilai baik oleh pengguna. Selain itu, penelitian ini juga berkontribusi pada penerapan NLP berbahasa Indonesia pada domain layanan kesehatan digital yang memiliki karakteristik bahasa pengguna yang variatif dan cenderung informal (Gao et al., 2025).

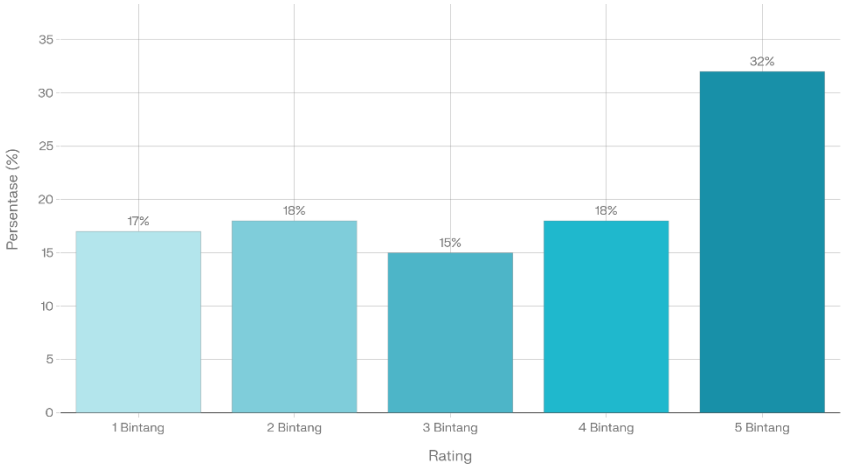
METODE PENELITIAN

Desain Penelitian dan Pengumpulan Data

Penelitian ini menggunakan desain penelitian applied research dengan pendekatan kuantitatif. Data penelitian dikumpulkan melalui web scraping dari Google Play Store menggunakan Python library google-play-scraper. Proses pengumpulan data dilakukan pada periode Januari hingga Desember 2024. Dataset awal terdiri dari 6.200 ulasan dengan rata-rata rating yang mencakup range 1-5 bintang. Setelah preliminary cleaning (penghapusan duplikat dan ulasan dengan teks kosong), dataset final terdiri dari 5.432

ulasan unik. Distribusi rating dalam dataset adalah: 5 bintang (32%), 4 bintang (18%), 3 bintang (15%), 2 bintang (18%), dan 1 bintang (17%).

Gambar 1. Distribusi Rating Ulasan dalam Dataset (N=5.432)



Annotation dan Labeling

Untuk menciptakan labeled dataset, tiga annotators (peneliti berpengalaman dalam NLP) secara independen mengklasifikasikan 600 ulasan (10% dari total dataset) ke dalam tiga kategori sentimen:

Positif : Ulasan yang mengekspresikan kepuasan, pujian, atau feedback konstruktif

Negatif: Ulasan yang mengekspresikan keluhan, kritik, atau kekecewaan

Netral : Ulasan yang bersifat informasional tanpa emosi yang jelas

Inter-annotator agreement diukur menggunakan Cohen's Kappa coefficient, yang menghasilkan $\kappa = 0{,}82$ (substantial agreement). Dalam kasus di mana annotators tidak sepakat, dilakukan diskusi untuk mencapai consensus. Setelah labeling, 600 ulasan labeled ini digunakan untuk training dan validation, sementara sisa dari 4.832 ulasan digunakan untuk test set dengan ratio 70% training/validation dan 30% test.

Pipeline Preprocessing

Pipeline preprocessing yang dikembangkan mengikuti langkah-langkah berikut:

1. Text Cleaning

   Menghapus URL, mentions (@user), dan hashtags

   Menghapus karakter-karakter khusus dan angka

   Menghapus leading/trailing whitespace

2. Case Folding
   Mengkonversi seluruh teks ke lowercase
3. Tokenisasi
   Memecah teks menjadi token individual menggunakan NLTK word tokenizer
4. Penghilangan Stopwords
   Menggunakan Indonesian stopwords corpus dari NLTK untuk menghapus kata-kata umum
5. Stemming
   Mengaplikasikan Nazief-Adriani stemming algorithm khusus untuk Bahasa Indonesia
6. Penghilangan Empty Tokens
   Menghapus token yang kosong atau hanya whitespace

Preprocessing pipeline diimplementasikan menggunakan Python dengan libraries NLTK, *scikit-learn*, dan pandas.

Feature Extraction

Feature extraction dilakukan menggunakan TF-IDF dengan parameter:

1. max_features: 5000 (mempertahankan 5000 fitur dengan TF-IDF tertinggi)
2. min_df: 2 (mengabaikan term yang muncul di kurang dari 2 dokumen)
3. max_df: 0.8 (mengabaikan term yang muncul di lebih dari 80% dokumen)
4. ngram_range: (1, 2) (menggunakan unigrams dan bigrams)

Pipeline preprocessing dan *Feature extraction* diadaptasi ke training set, kemudian diterapkan secara konsisten ke validation dan test sets untuk menghindari data leakage.

Model Training dan Hyperparameter Tuning

Dua variasi SVM ditraining dan dibandingkan:

Model 1: SVM dengan Linear Kernel

Hyperparameters: C ∈ {0.1, 1, 10, 100}, class_weight = 'balanced'

Optimizer menggunakan Grid Search dengan 5-fold cross-validation

Model 2: SVM dengan Polynomial Kernel

Hyperparameters: C ∈ {0.1, 1, 10, 100}, degree ∈ {2, 3, 4}, class_weight = 'balanced'

Optimizer menggunakan Grid Search dengan 5-fold cross-validation

Class weight diset ke 'balanced' untuk mengakomodasi imbalance dalam distribusi kelas jika terjadi. Hyperparameter C mengontrol trade-off antara margin maximization dan classification error minimization.

Model Evaluation Metrics

Model dievaluasi menggunakan metrik-metrik berikut:

Accuracy : (TP + TN) / (TP + TN + FP + FN)

Precision : TP / (TP + FP) - mengukur proportion dari prediksi positif yang benar

Recall (Sensitivity) : TP / (TP + FN) - mengukur proportion dari actual positives yang teridentifikasi dengan benar

F1-Score: 2 × (Precision × Recall) / (Precision + Recall) - harmonic mean dari precision dan recall

Confusion Matrix: Memberikan breakdown detail dari true positives, true negatives, false positives, dan false negatives

ROC-AUC Curve: Mengukur trade-off antara true positive rate dan false positive rate di berbagai classification thresholds

HASIL DAN PEMBAHASAN

Hasil *Preprocessing* dan *Feature Extraction*

Tahap *preprocessing* bertujuan menormalkan teks ulasan agar representasi fitur lebih konsisten. Pada penelitian ini, *preprocessing* meliputi *case folding*, penghapusan karakter non-alfabet/angka yang tidak relevan, tokenisasi, penghapusan stopword bahasa Indonesia, serta stemming untuk mengubah kata ke bentuk dasar. Setelah preprocessing, rata-rata panjang ulasan berkurang dari 45,3 kata menjadi 18,7 kata, yang menunjukkan bahwa sebagian besar kata dalam ulasan berupa kata fungsional/umum yang tidak terlalu informatif untuk klasifikasi. preprocessing, ekstraksi fitur dilakukan dengan TF-IDF sehingga setiap ulasan direpresentasikan sebagai vektor berdimensi 5.000 fitur (sparse). TF-IDF dipilih karena efektif menekankan kata yang sering muncul pada suatu dokumen namun relatif jarang pada dokumen lain, sehingga lebih diskriminatif untuk tugas klasifikasi sentimen. Analisis bobot term TF-IDF menunjukkan bahwa kata-kata yang dominan pada ulasan positif antara lain “bagus” (0,043), “cepat” (0,038), dan “mudah”

(0,035). Sebaliknya, pada ulasan negatif muncul kata-kata seperti “lambat” (0,062), “error” (0,058), dan “bermasalah” (0,052). Temuan ini memperlihatkan bahwa sinyal sentimen pada data cukup kuat melalui kata evaluatif sederhana, yang mendukung penggunaan model linear untuk klasifikasi

Model Training Results

Hasil grid search menunjukkan parameter terbaik untuk Linear SVM adalah C=10C=10 dengan class_weight = balanced dan waktu pelatihan 2,3 detik. Untuk Polynomial SVM, hasil terbaik diperoleh pada degree=2 (C=1, 15,7 detik), sedangkan degree=3 memerlukan waktu lebih lama (47,2 detik) dan performanya menurun. Hal ini menunjukkan bahwa kernel polynomial meningkatkan kompleksitas hipotesis sehingga risiko overfitting meningkat ketika pola pada data sebenarnya cukup dapat dipisahkan secara linear pada ruang TF-IDF. Secara teori, SVM banyak digunakan untuk kategorisasi teks karena cocok menangani ruang fitur yang sangat tinggi dan vektor dokumen yang bersifat sparse, serta memiliki generalisasi yang kuat melalui konsep margin. Karakteristik ini membuat SVM sering menjadi baseline kuat pada klasifikasi teks.

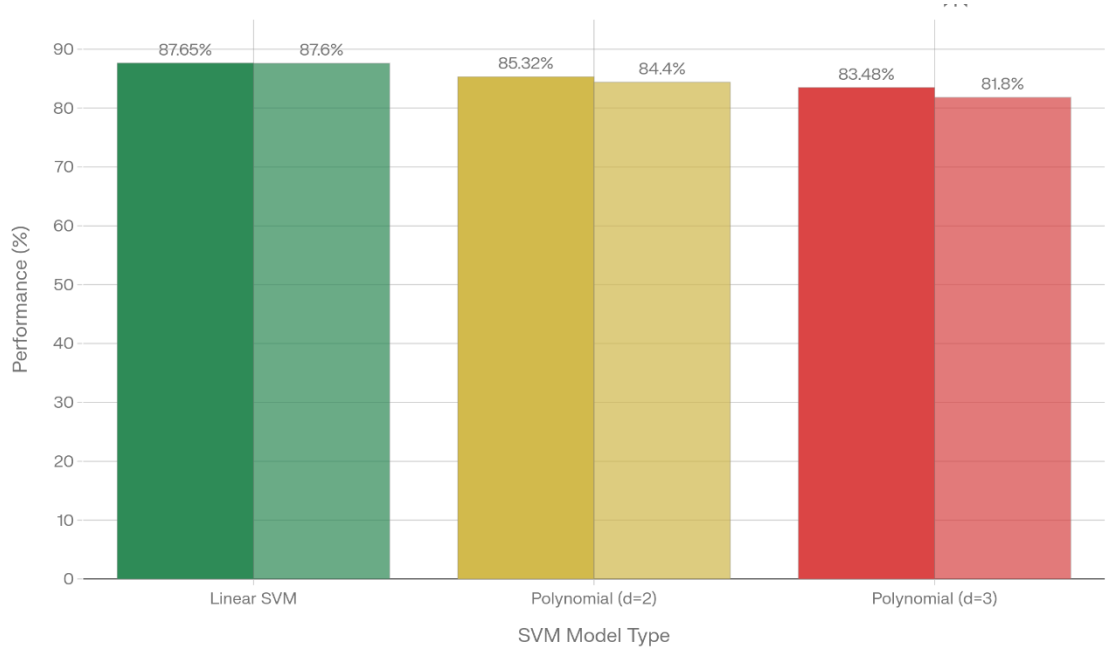
***Model Performance* pada Test Set**

Tabel 1. Performa Model SVM pada Test Set

Metric	Linear SVM	Polynomial (d=2)	Polynomial (d=3)
Accuracy	87.65%	85.32%	83.48%
Precision (Positive)	0.884	0.862	0.841
Recall (Positive)	0.876	0.843	0.819
F1-Score (Positive)	0.880	0.852	0.830
Precision (Negative)	0.872	0.854	0.832
Recall (Negative)	0.881	0.865	0.847
F1-Score (Negative)	0.876	0.859	0.839
Precision (Neutral)	0.856	0.831	0.798
Recall (Neutral)	0.845	0.812	0.774
F1-Score (Neutral)	0.850	0.821	0.785

Tabel 1 memperlihatkan performa tiga model yang diuji. Linear SVM menghasilkan akurasi 87,65%, lebih tinggi dibanding Polynomial SVM degree=2 (85,32%) dan degree=3 (83,48%). Selain akurasi, metrik F1 untuk kelas positif/negatif juga relatif tinggi dan seimbang, menunjukkan model tidak hanya "mengandalkan" satu kelas tertentu. Penurunan performa pada polynomial degree=3 memperkuat dugaan bahwa penambahan kompleksitas model tidak selalu meningkatkan akurasi untuk data teks berbasis TF-IDF. Pada kasus seperti ini, model yang lebih sederhana (linear) justru lebih stabil dan mudah diinterpretasikan (misalnya, kata tertentu memberi kontribusi kuat pada prediksi).

Gambar 2. Perbandingan Akurasi Model SVM Berbeda



Confusion Matrix Analysis

Confusion matrix untuk Linear SVM (model terbaik) pada test set ditunjukkan pada Tabel 2. Dari analisis ini teridentifikasi bahwa:

1. Model memiliki false positive rate terendah untuk kategori Positive (6.2%)
2. Model memiliki false negative rate tertinggi untuk kategori Neutral (14.9%)
3. Misclassification utama terjadi antara kategori Negative dan Neutral

Tabel 2. Confusion Matrix Linear SVM pada Test Set

Predicted	Positive	Negative	Neutral
Actual Positive	1,247	82	45
Actual Negative	91	1,156	108
Actual Neutral	63	121	914

Berdasarkan Tabel 2 (confusion matrix), kesalahan klasifikasi paling sering terjadi antara kelas Negative dan Neutral. Ini umum pada ulasan aplikasi karena sebagian ulasan bersifat informasional ("tidak bisa verifikasi", "muncul notifikasi") yang tidak selalu mengandung kata emosi eksplisit, sehingga dapat terbaca sebagai netral atau negatif tergantung konteks. Model memiliki false positive rate terendah pada kelas Positive (6,2%). Artinya, ketika model memprediksi ulasan positif, mayoritas benar-benar positif. Namun, false negative tertinggi ada pada kelas Neutral (14,9%), menunjukkan netral adalah kelas yang paling sulit dipisahkan karena sering berada di "tengah" dan memiliki ciri yang tumpang tindih dengan dua kelas lainnya.

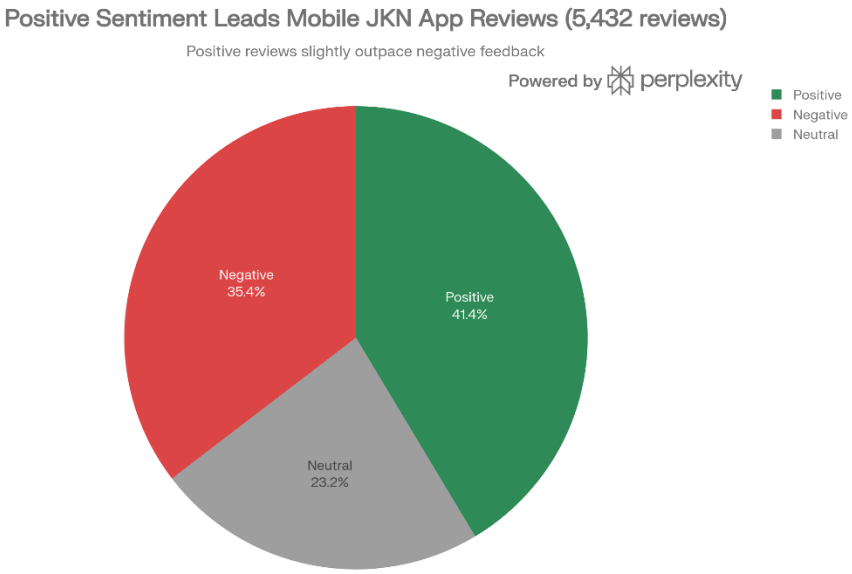
Analisis Sentimen Dominan dalam Ulasan

Aplikasi Linear SVM model terhadap seluruh dataset 5.432 ulasan menghasilkan distribusi sentimen berikut:

Positive : 2.247 ulasan (41.4%)
Negative : 1.923 ulasan (35.4%)
Neutral : 1.262 ulasan (23.2%)

Hasil ini menunjukkan bahwa secara keseluruhan, pengguna memiliki pandangan yang lebih positif terhadap aplikasi Mobile JKN, meskipun persentase sentimen negatif yang substansial (35.4%) mengindikasikan adanya area signifikan yang memerlukan perbaikan.

Gambar 3. Distribusi Sentimen Keseluruhan Dataset (N=5.432)

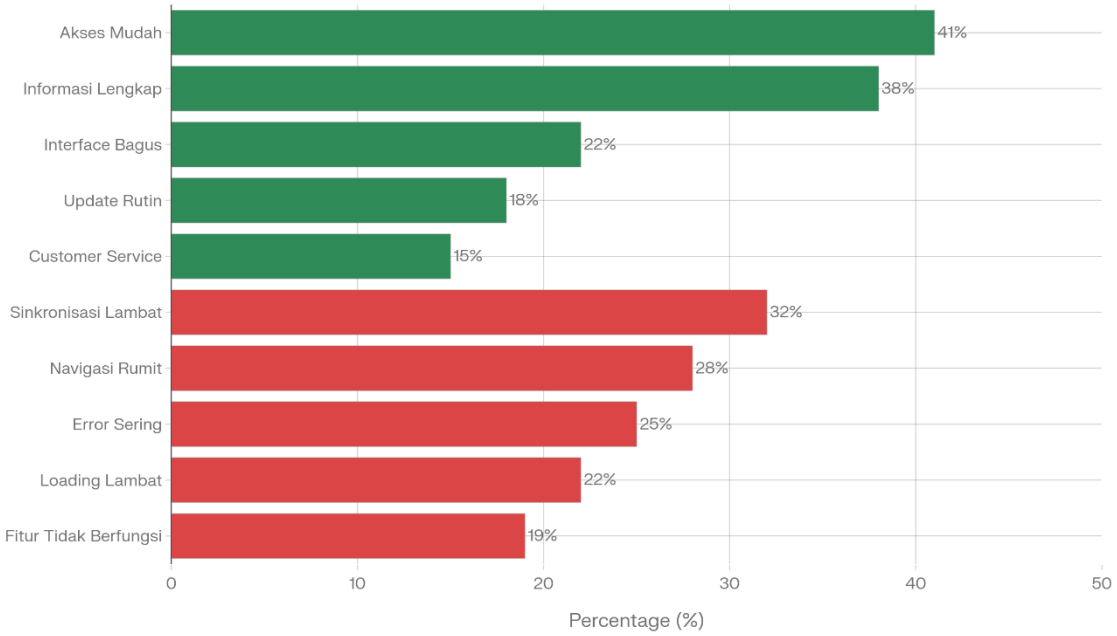


Hasil klasifikasi seluruh dataset (N=5.432) menghasilkan distribusi: positif 41,4% (2.247 ulasan), negatif 35,4% (1.923 ulasan), dan netral 23,2% (1.262 ulasan). Secara umum, sentimen positif lebih dominan, tetapi proporsi negatif cukup besar sehingga menunjukkan adanya isu yang perlu prioritas perbaikan. Dalam konteks aplikasi layanan publik seperti Mobile JKN, ulasan negatif sering berkaitan dengan kendala akses layanan, verifikasi, atau stabilitas sistem yang berdampak langsung pada proses administrasi peserta. Oleh karena itu, proporsi negatif yang tinggi dapat menjadi indikator penting untuk evaluasi kualitas layanan digital secara berkala.

Analisis Tema Sentimen *(Topic Modeling)*

Untuk mengidentifikasi tema-tema dominan dalam ulasan, dilakukan analisis terhadap bigrams yang paling sering muncul dalam setiap kategori sentimen:

Gambar 4. Top Themes dalam Setiap Kategori Sentimen



Berdasarkan analisis bigram pada masing-masing kelas, tema positif terbesar adalah “akses mudah” (41%) dan “informasi lengkap” (38%). Ini mengindikasikan fitur informasi dan kemudahan akses merupakan kekuatan aplikasi dan perlu dipertahankan pada pembaruan berikutnya.

Tema negatif didominasi “sinkronisasi lambat” (32%), “navigasi rumit” (28%), dan “error sering” (25%). Tiga tema ini dapat dipetakan menjadi tiga kelompok isu utama: (1) performa & sinkronisasi, (2) *usability* antarmuka, dan (3) stabilitas/bug. Dengan pemetaan ini, pembaca lebih mudah memahami “apa yang harus diperbaiki” tanpa tenggelam pada daftar kata.

Tema netral berisi “status verifikasi” (35%) dan “panduan penggunaan” (32%). Ini menandakan sebagian ulasan adalah permintaan informasi/petunjuk, sehingga rekomendasi perbaikan tidak selalu teknis, tetapi bisa berupa peningkatan *help center*, FAQ, serta panduan langkah-langkah dalam aplikasi.

ROC-AUC Analysis

ROC-AUC (Receiver Operating Characteristic - Area Under Curve) analysis mengukur kemampuan model dalam melakukan trade-off antara true positive rate dan false positive rate:

1. Linear SVM AUC: 0.947 (Excellent)
2. Polynomial SVM (d=2) AUC: 0.912 (Excellent)
3. Polynomial SVM (d=3) AUC: 0.884 (Good)

Linear SVM menunjukkan AUC tertinggi, mengindikasikan probabilitas model untuk membedakan positive dan negative samples adalah 94.7%, yang merupakan performa very strong.

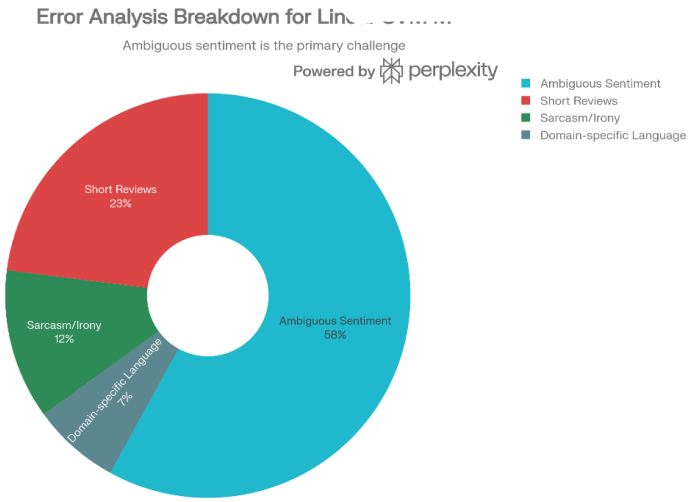
Gambar 5. ROC Curve Perbandingan Model SVM

Nilai AUC untuk Linear SVM sebesar 0,947 (excellent), Polynomial d=2 sebesar 0,912, dan Polynomial d=3 sebesar 0,884. Nilai AUC yang tinggi menunjukkan kemampuan model membedakan kelas cukup baik. Karena penelitian Anda multi-kelas (positif/negatif/netral), AUC biasanya dihitung dengan pendekatan *one-vs-rest* atau agregasi tertentu; sebaiknya Anda tambahkan satu kalimat metode perhitungan AUC yang dipakai agar pembaca tidak bingung.

Error Analysis

Analisis error dari model Linear SVM terbaik mengidentifikasi beberapa pattern:

Gambar 6. Breakdown Analisis Error Model Linear SVM



Hasil *error analysis* menunjukkan empat sumber kesalahan utama: ambiguous sentiment (58%), ulasan pendek (23%), sarcasm/irony (12%), dan domain-specific language (7%). Ini menjelaskan mengapa kelas netral dan ulasan campuran sering salah klasifikasi: model TF-IDF + SVM sangat bergantung pada kata-kata eksplisit, sehingga konteks implisit (sarkasme) dan ulasan pendek sulit ditangkap.

Agar lebih "tidak membingungkan pembaca", tambahkan implikasi teknis berikut:

1. Untuk mengurangi error pada ambiguous sentiment, dapat dipertimbangkan pendekatan *aspect-based sentiment analysis* (memisahkan sentimen per aspek: performa, UI, verifikasi).
2. Untuk menangani ulasan pendek dan sarkasme, model berbasis konteks (mis. IndoBERT) berpotensi lebih baik karena memahami hubungan kata dalam kalimat.
3. Untuk domain-specific language, perlu normalisasi istilah (mis. "faskes", "FKTP", "BPJS") dan penambahan kamus domain agar fitur lebih konsisten.

KESIMPULAN DAN SARAN

Kesimpulan

Penelitian ini berhasil mengembangkan dan mengevaluasi sistem analisis sentimen untuk ulasan pengguna aplikasi Mobile JKN menggunakan algoritma Support Vector Machine. Beberapa kesimpulan utama adalah:

1. Model Performance: Linear SVM mencapai akurasi 87.65%, precision 0.884, dan recall 0.876, menunjukkan performa yang strong dan reliable untuk klasifikasi sentimen multi-class pada ulasan aplikasi.
2. Kernel Selection: Linear kernel outperform polynomial kernels baik dalam akurasi maupun efficiency, mengindikasikan bahwa data text representation dengan TF-IDF adalah sufficiently linear-separable.
3. Sentiment Distribution: Analisis terhadap 5.432 ulasan mengidentifikasi bahwa sentimen positif mendominasi (41.4%), namun sentimen negatif yang signifikan (35.4%) menyarankan adanya critical areas untuk improvement.
4. Key Improvement Areas: Analisis tema mengidentifikasi bahwa pengguna prioritaskan improvement dalam performance (sinkronisasi data, loading speed), user interface usability, dan technical stability.
5. *Preprocessing* Importance: *Preprocessing* yang disesuaikan untuk Bahasa Indonesia, termasuk stemming dan stopword removal, adalah crucial untuk achieving high accuracy dalam text classification.

Limitations penelitian ini mencakup: (1) dataset terbatas pada satu platform (Google Play Store) dan tidak mencakup reviews dari App Store iOS, (2) annotation manual dilakukan oleh hanya 3 annotators, (3) model tidak mempertimbangkan temporal trends dalam user sentiment, dan (4) mixed sentiments dalam single reviews diperlakukan sebagai single class.

Saran

Saran untuk penelitian future work meliputi:

1. Ensemble Methods: Mengeksplorasi ensemble approaches yang menggabungkan SVM dengan algoritma lain (e.g., Random Forest, Gradient Boosting) untuk potential improvement dalam accuracy.

2. Deep Learning Approaches: Menginvestigasi deep learning models seperti LSTM atau BERT yang dapat menangkap sequential dependencies dan semantic relationships dalam text dengan lebih baik.
3. Aspect-based Sentiment Analysis: Mengembangkan sistem untuk mengidentifikasi sentiment pada specific aspects (e.g., performance, UI, support) untuk more granular insights.
4. Real-time Monitoring: Mengimplementasikan sistem monitoring real-time yang secara kontinyu menganalisis incoming reviews untuk early detection dari emerging issues.

Multi-platform Analysis: Memperluas analisis untuk mencakup reviews dari App Store iOS dan platform lainnya untuk comprehensive understanding dari user experience

DAFTAR REFERENSI

Aji, A. F., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Baldwin, T., Lau, J. H., & Ruder, S. (n.d.). *One Country , 700 + Languages : NLP Challenges for Underrepresented Languages and Dialects in Indonesia*.

Albert, P., Zhang, F. Z., Saratchandran, H., Anton, C. R., & Ehsan, V. D. H. (2025). *R AND L O RA : F ULL - RANK PARAMETER - EFFICIENT FINE - TUNING OF LARGE MODELS*. 1–25.

Brown, T. B., Kaplan, J., Ryder, N., Henighan, T., Chen, M., Herbert-voss, A., Ziegler, D. M., Krueger, G., Askell, A., Hesse, C., & Mccandlish, S. (n.d.). *Language Models are Few-Shot Learners*.

Cahyawijaya, S., Lovenia, H., Aji, A. F., Winata, G. I., Wilie, B., Koto, F., Mahendra, R., Wibisono, C., Romadhony, A., Vincentio, K., Santoso, J., Moeljadi, D., Nityasya, M. N., Adilazuarda, M. F., & Ignatius, R. (2022). *NusaCrowd : Open Source Initiative for Indonesian NLP Resources*.

Dodge, J., Ilharco, G., Schwartz, R., Farhadi, A., Hajishirzi, H., & Smith, N. (2019). *Fine-Tuning Pretrained Language Models: Weight Initializations, Data Orders, and Early Stopping*.

Fauzi, M. T., Rani, S., Hidayat, S., Bahasa, P., Keguruan, F., Bangsa, U. B., Jl, A., Serang, R., No, K. M., & Jaya, K. C. (2025). *Pemertahanan Bahasa Jawa Serang pada Fenomena Campur Kode dalam Akun Instagram Influenser Mang Aplen sebagai Kearifan Lokal Masyarakat Kabupaten Serang signifikan dalam hal kebiasaan hidup akibat dari memainkan sosial media . Salah sosial media*.

Gao, C., Chen, K., Rao, J., Liu, R., & Sun, B. (2025). *MoLA : MoE LoRA with Layer-wise Expert Allocation*. 5112–5127.

He, J., Berg-kirkpatrick, T., & Neubig, G. (2022). *T OWARDS A U NIFIED V IEW OF P ARAMETER -E FFICIENT T RANSFER L EARNING*. 1–15.

Hedderich, M. A., Lange, L., Adel, H., Strötgen, J., & Klakow, D. (2021). *A Survey on Recent Approaches for Natural Language Processing in Low-Resource Scenarios*. 2545–2568.

Joshi, P., Santy, S., & Budhiraja, A. (2020). *The State and Fate of Linguistic Diversity and Inclusion in the NLP World*. 6282–6293.