

# Analisis Sentimen Ulasan Mobile JKN Menggunakan TF-IDF dan Logistic Regression

Alders Paliling<sup>a,1,\*</sup> dan Muh. Nurtanzis Sutoyo<sup>b,2</sup>

<sup>a</sup>Program Studi Ilmu Komputer, Universitas Sembilanbelas November Kolaka, Kolaka, Indonesia

<sup>b</sup>Program Studi Sistem Informasi, Universitas Sembilanbelas November Kolaka, Kolaka, Indonesia

<sup>1</sup>palilingalders@gmail.com; <sup>2</sup>mns.usn21@gmail.com;

\*corresponding author

## Informasi Artikel

Diserahkan : 12 Februari 2026  
Diterima : 12 Maret 2026  
Direvisi : 17 Maret 2026  
Diterbitkan : 17 Maret 2026

### Kata Kunci:

Analisis sentimen  
Pembelajaran mesin  
Aplikasi kesehatan  
Penilaian pengguna  
Pelabelan Berdasarkan Peringkat

## ABSTRAK

Transformasi digital dalam layanan publik mendorong pemanfaatan aplikasi mobile sebagai sarana utama penyediaan layanan kesehatan, salah satunya melalui aplikasi Mobile JKN. Ulasan di *Google Play Store* menjadi sumber data penting untuk mengetahui persepsi dan pengalaman pengguna terhadap kualitas layanan aplikasi seiring dengan peningkatan pengguna. Tujuan penelitian ini adalah untuk mempelajari sentimen pengguna tentang aplikasi Mobile JKN dan menemukan elemen layanan yang paling sering menyebabkan ketidakpuasan pengguna. Data penelitian terdiri dari 5.000 ulasan pengguna yang dikumpulkan dari *Google Play Store*. *Rating based labeling* digunakan untuk melabelkan sentimen, dengan praproses teks meliputi pembersihan data dan *stemming*. Representasi fitur teks menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)*, sedangkan klasifikasi sentimen dilakukan menggunakan algoritma *Logistic Regression*. Menurut hasil penelitian, model dapat mencapai tingkat akurasi sebesar 92,15%, dengan *precision* sebesar 96,42%, *recall* sebesar 89,79%, dan *F1-score* sebesar 92,99%. Analisis sentimen berdasarkan kategori layanan menunjukkan bahwa sentimen negatif paling dominan terdapat pada aspek login dan autentikasi, khususnya terkait pengiriman kode *OTP*, serta pada aspek kinerja dan stabilitas sistem. Sebaliknya, kategori manfaat dan kemudahan layanan aplikasi menunjukkan tingkat sentimen negatif yang relatif rendah. Penelitian ini menunjukkan bahwa pemanfaatan analisis sentimen terhadap ulasan pengguna dapat berperan dalam mendukung evaluasi kualitas layanan pada aplikasi kesehatan digital.

## ABSTRACT

Digital transformation in public services has encouraged the use of mobile applications as a primary medium for delivering healthcare services, one of which is the Mobile JKN application. User reviews on *Google Play Store* serve as an important data source for understanding user perceptions and experiences regarding the quality of application services, especially as the number of users continues to grow. This study aims to analyze user sentiment toward the Mobile JKN application and identify service aspects that most frequently lead to user dissatisfaction. The dataset consists of 5,000 user reviews collected from *Google Play Store*. Sentiment labeling was performed using a rating-based labeling approach, while text preprocessing included data cleaning and stemming. Text features were represented using the *Term Frequency-Inverse Document Frequency (TF-IDF)* method, and sentiment classification was conducted using the *Logistic Regression* algorithm. The results show that the model achieved an accuracy of 92.15%, with a precision of 96.42%, recall of 89.79%, and an *F1-score* of 92.99%. Sentiment analysis based on service categories indicates that negative sentiment is most dominant in the login and authentication aspects, particularly related to *OTP* code delivery, as well as in system performance and stability. In contrast, the benefits and ease-of-use category shows a relatively low level of negative sentiment. These findings demonstrate that sentiment analysis of user reviews can support the evaluation of service quality in digital health applications.

### Keywords:

Sentiment analysis  
Machine Learning  
Health application  
User reviews  
Rating-Based Labeling

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



## I. Pendahuluan

Perubahan besar dalam penyediaan layanan publik dan kesehatan telah disebabkan oleh transformasi digital, terutama melalui penggunaan aplikasi berbasis perangkat bergerak. Aplikasi mobile memungkinkan institusi publik menyediakan layanan yang lebih cepat, mudah diakses, dan terintegrasi dengan kebutuhan

masyarakat. Platform digital dalam bidang kesehatan tidak hanya berfungsi sebagai sumber informasi, tetapi juga membantu pengguna membuat keputusan dan mengubah perilaku yang berkaitan dengan kesehatan pengguna. Berbagai studi menunjukkan bahwa tingkat penerimaan (*acceptability*) dan keterlibatan (*engagement*) pengguna terhadap *platform* digital menjadi faktor kunci keberhasilan implementasi layanan kesehatan berbasis teknologi [1]. Oleh karena itu, pemahaman terhadap pengalaman dan persepsi pengguna menjadi aspek penting dalam pengembangan dan evaluasi aplikasi layanan publik digital.

Seiring dengan meningkatnya jumlah pengguna aplikasi mobile, platform distribusi aplikasi seperti *Google Play Store* menyediakan sumber data yang kaya berupa ulasan pengguna. Ulasan tersebut bersifat spontan dan mencerminkan pengalaman nyata pengguna dalam menggunakan aplikasi, baik dalam bentuk kepuasan maupun keluhan. Sejumlah penelitian telah memanfaatkan ulasan pengguna sebagai sarana untuk mengevaluasi kualitas layanan dan persepsi pengguna pada berbagai jenis aplikasi [2]. Beberapa penelitian terdahulu melaporkan bahwa analisis sentimen terhadap ulasan aplikasi mampu memetakan persepsi pengguna terhadap kualitas layanan sekaligus mengidentifikasi area layanan yang memerlukan perbaikan [3], [4]. Beberapa penelitian menunjukkan bahwa rating numerik tidak selalu sejalan dengan polaritas sentimen teks ulasan, sehingga analisis sentimen berbasis teks tetap diperlukan untuk memperoleh pemahaman yang lebih akurat terhadap persepsi pengguna [5], [6]. Selain dimanfaatkan untuk evaluasi kualitas layanan, ulasan aplikasi juga banyak digunakan dalam konteks rekayasa perangkat lunak untuk mengidentifikasi kebutuhan pengguna, keluhan utama, serta prioritas pengembangan fitur melalui pendekatan *app review mining* [7], [8], [9]. Pendekatan berbasis ulasan pengguna ini memungkinkan pengembang dan pengelola sistem memperoleh umpan balik dalam skala besar secara efisien tanpa perlu melakukan survei konvensional yang memerlukan waktu dan biaya lebih besar.

Seiring dengan meningkatnya volume data ulasan pengguna, analisis manual menjadi tidak efektif lagi. Oleh sebab itu, teknik *sentiment analysis* berbasis *text mining* dan *machine learning* banyak digunakan untuk mengklasifikasikan opini pengguna secara otomatis. Berbagai penelitian menunjukkan bahwa metode pembelajaran mesin klasik, seperti *Naïve Bayes*, *Support Vector Machine*, dan *Logistic Regression* yang dikombinasikan dengan representasi fitur *TF-IDF*, mampu menghasilkan kinerja klasifikasi sentimen yang baik dan stabil [10], [11]. Meskipun pendekatan pembelajaran mendalam banyak digunakan pada penelitian terkini, metode pembelajaran mesin klasik tetap relevan karena stabilitas, interpretabilitas, serta efisiensi komputasi yang lebih baik pada data ulasan aplikasi berskala menengah [12].

Dalam domain layanan publik dan kesehatan, Analisis sentimen telah digunakan dalam beberapa studi kasus. Sentimen negatif pengguna umumnya berkaitan dengan permasalahan teknis, seperti kegagalan sistem, proses autentikasi, dan stabilitas aplikasi, sedangkan sentimen positif lebih banyak dikaitkan dengan kemudahan serta manfaat layanan yang diberikan [4], [13], [14], [15]. Penelitian pada aplikasi Mobile JKN menunjukkan bahwa permasalahan autentikasi, pengiriman kode *OTP*, serta kestabilan sistem merupakan sumber utama sentimen negatif pengguna dan berdampak langsung pada kepuasan layanan [16], [17]. Temuan serupa juga dilaporkan pada studi analisis sentimen aplikasi kesehatan, yang menekankan bahwa ulasan pengguna dapat digunakan untuk mengidentifikasi kebutuhan pengguna dan permasalahan fungsional aplikasi secara lebih mendalam [18]. Hal ini menunjukkan bahwa analisis sentimen memiliki peran strategis dalam mendukung peningkatan kualitas layanan aplikasi publik.

Meskipun demikian, penelitian terdahulu masih memiliki sejumlah keterbatasan. Beberapa penelitian hanya membahas kategori sentimen secara keseluruhan tanpa mengaitkannya dengan kategori layanan tertentu, sehingga sulit untuk mengidentifikasi aspek layanan yang perlu diprioritaskan untuk perbaikan. Selain itu, penelitian yang menggabungkan pendekatan *rating based labeling* dengan model *machine learning* untuk menganalisis sentimen ulasan aplikasi layanan publik secara terintegrasi masih terbatas [19], [20], [21]. Akibatnya, pengelola aplikasi layanan publik masih menghadapi kesulitan dalam mengidentifikasi secara sistematis aspek layanan yang paling banyak menimbulkan ketidakpuasan pengguna berdasarkan data ulasan yang tersedia. Namun demikian, masih terbatas penelitian yang mengintegrasikan analisis sentimen ulasan aplikasi dengan pemetaan kategori layanan secara spesifik. Keterbatasan ini menyebabkan rekomendasi perbaikan layanan yang dihasilkan belum bersifat operasional dan terprioritaskan [22], [23]. Mengacu pada kesenjangan penelitian yang ada, penelitian ini difokuskan pada analisis sentimen dan layanan aplikasi Mobile JKN menggunakan pendekatan *machine learning* dengan memanfaatkan *review* pengguna dari *Google Play Store*. Diharapkan penelitian ini dapat memberikan kontribusi berupa analisis sentimen berbasis teks yang dikaitkan dengan kategorisasi layanan, sehingga mampu memberikan rekomendasi yang lebih terarah bagi pengelola aplikasi layanan publik dalam meningkatkan kualitas sistem dan layanan.

## II. Metode

Penelitian ini menerapkan pendekatan kuantitatif dengan memanfaatkan metode analisis sentimen berbasis machine learning untuk mengkaji ulasan pengguna aplikasi Mobile JKN yang dikumpulkan dari platform Google Play Store. Rancangan metodologi penelitian disusun agar proses analisis dapat dilakukan secara terstruktur dan dapat direplikasi, dimulai dari tahap pengumpulan data, dilanjutkan dengan proses pengolahan dan klasifikasi sentimen, hingga tahap evaluasi hasil analisis serta pemetaan ulasan berdasarkan kategori layanan yang relevan.

### A. Pengumpulan Data

Sumber data penelitian ini adalah ulasan pengguna aplikasi Mobile JKN yang diperoleh dari platform Google Play Store melalui proses pengambilan data menggunakan metode web scraping. Informasi yang dikumpulkan meliputi teks ulasan pengguna serta rating bintang yang diberikan terhadap aplikasi tersebut. Pemanfaatan ulasan aplikasi sebagai sumber data penelitian telah banyak dilakukan dalam berbagai studi sebelumnya, karena data tersebut mampu mencerminkan pengalaman dan persepsi pengguna secara langsung, sekaligus menyediakan jumlah data yang besar untuk dianalisis [20], [21]. Pada tahap ini, ulasan duplikat dan ulasan tanpa teks dihapus untuk menjaga kualitas dataset.

### B. Pelabelan Sentimen

Pelabelan sentimen dalam penelitian ini dilakukan dengan menggunakan pendekatan *rating-based labeling*. Pada metode ini, ulasan yang pada penelitian ini, label sentimen ditentukan berdasarkan pendekatan *rating-based labeling* yang menggunakan nilai rating sebagai acuan dalam klasifikasi sentimen. Rating bintang empat dan lima diklasifikasikan sebagai sentimen positif, sedangkan ulasan dengan rating bintang satu hingga tiga dimasukkan ke dalam kategori sentimen negatif. Pendekatan ini digunakan karena rating yang diberikan pengguna dianggap dapat merepresentasikan kecenderungan sikap atau penilaian mereka terhadap aplikasi. Pendekatan ini dipilih karena efisien dan umum digunakan pada analisis sentimen ulasan aplikasi berskala besar [10]. Meskipun demikian, beberapa studi menunjukkan bahwa rating numerik tidak selalu sepenuhnya sejalan dengan polaritas sentimen teks, sehingga analisis sentimen berbasis teks tetap diperlukan untuk memperoleh pemahaman yang lebih akurat [5].

### C. Text Preprocessing

Untuk memastikan kualitas data yang lebih baik, dilakukan tahap pra-proses teks yang bertujuan mengurangi noise sekaligus meningkatkan representasi data sebelum digunakan dalam proses analisis lebih lanjut. Tahap ini terdiri atas sejumlah proses pembersihan teks, seperti *case folding*, penghapusan tanda baca, angka, *URL*, dan karakter non-alfabet, serta eliminasi kata-kata umum (*stopwords*) yang tidak berperan signifikan dalam pembentukan makna kalimat. Selanjutnya diterapkan proses *stemming* yang berfungsi untuk mengubah berbagai bentuk kata menjadi bentuk dasar yang seragam. Proses pra-proses teks merupakan tahapan yang krusial dalam analisis teks karena secara langsung memengaruhi kualitas fitur yang dihasilkan serta kinerja model klasifikasi yang digunakan [12], [24].

### D. Ekstraksi Fitur dan Klasifikasi Sentimen

Tahap ekstraksi fitur dalam penelitian ini dilakukan dengan menerapkan metode *TF-IDF*, yang berfungsi mengubah data ulasan berbasis teks menjadi representasi numerik sehingga dapat digunakan sebagai input bagi algoritma *machine learning*. Metode *TF-IDF* menghitung bobot kata dengan mempertimbangkan frekuensi kemunculannya dalam dokumen serta distribusinya pada seluruh korpus. Konsep ini menyebabkan kata yang sering muncul dalam satu dokumen namun jarang ditemukan pada dokumen lain memperoleh bobot yang lebih besar. Secara matematis, nilai bobot *TF-IDF* untuk suatu term  $t$  pada dokumen  $d$  dirumuskan sebagai berikut:

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t) \quad (1)$$

dengan *Term Frequency*  $\text{TF}(t, d)$  menyatakan frekuensi kemunculan term  $t$  dalam dokumen  $d$ , dan *Inverse Document Frequency*  $\text{IDF}(t)$  dirumuskan sebagai:

$$\text{IDF}(t) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

$N$  menyatakan jumlah keseluruhan dokumen yang terdapat dalam korpus dan  $df(t)$  menunjukkan jumlah dokumen yang memuat term  $t$ . Pendekatan ini dipilih karena mampu menyeimbangkan frekuensi kata yang sering muncul dengan tingkat kepentingannya dalam keseluruhan data teks [25].

Klasifikasi sentimen dilakukan dengan menerapkan algoritma *Logistic Regression*, yang bertujuan untuk mengestimasi probabilitas sebuah dokumen berada pada kelas sentimen tertentu. Algoritma ini menggunakan fungsi sigmoid untuk mentransformasikan nilai linier ke dalam rentang probabilitas antara 0 dan 1, yang dirumuskan sebagai:

$$P(y = 1 | x) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (3)$$

di mana  $x$  merupakan vektor fitur hasil ekstraksi TF-IDF,  $w$  adalah vektor bobot model, dan  $b$  adalah bias. Logistic Regression dipilih karena memiliki performa yang stabil, interpretabilitas yang baik, serta efisiensi komputasi pada klasifikasi teks berdimensi tinggi [18]. Pembagian dataset dilakukan dengan memisahkan 80% data sebagai data latih dan 20% sebagai data uji, yang digunakan untuk mengevaluasi performa model klasifikasi.

#### E. Evaluasi Model

Evaluasi kinerja model dilakukan dengan menggunakan beberapa metrik, yaitu *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur performa klasifikasi secara komprehensif. Selain itu, *confusion matrix* dimanfaatkan untuk mengidentifikasi dan menganalisis kesalahan klasifikasi pada setiap kelas sentimen secara lebih rinci. Penggunaan metrik evaluasi ini umum diterapkan dalam penelitian klasifikasi teks dan analisis sentimen.

#### F. Analisis Sentimen Berdasarkan Kategori Layanan

Untuk memperoleh pemahaman yang lebih komprehensif mengenai berbagai aspek layanan dalam aplikasi, ulasan pengguna diklasifikasikan ke dalam beberapa kategori layanan utama, seperti autentikasi dan login, kinerja dan stabilitas sistem, pendaftaran dan data peserta, layanan rujukan dan antrean faskes, serta manfaat dan kemudahan penggunaan aplikasi. Pengelompokan kategori dilakukan berdasarkan kemunculan kata kunci tertentu pada teks ulasan. Setiap ulasan kemudian dipetakan ke kategori layanan yang paling dominan. Hasil klasifikasi sentimen dianalisis berdasarkan kategori layanan untuk mengidentifikasi aspek layanan yang paling banyak menerima sentimen negatif maupun positif, sehingga memungkinkan penyusunan rekomendasi perbaikan layanan yang lebih terarah dan operasional [9].

#### G. Tahap Persiapan Data

Sebelum tahap analisis sentimen dilakukan, dataset dalam penelitian ini dipersiapkan melalui sejumlah proses pengolahan awal guna memastikan kualitas dan kesiapan data. Proses persiapan data dilakukan untuk memastikan keandalan serta konsistensi data teks, yang pada akhirnya dapat mendukung peningkatan performa model klasifikasi dalam proses analisis. Pada tahap awal, data ulasan pengguna yang diambil dari *Google Play Store* dibersihkan dengan menghapus data duplikat serta ulasan yang tidak memiliki konten teks. Dataset hasil pembersihan ini kemudian diberi label sentimen menggunakan pendekatan *rating based labeling*.

Selanjutnya, dilakukan tahap praproses teks (*text preprocessing*) yang mencakup beberapa langkah, yaitu *case folding*, penghapusan tanda baca, angka, serta karakter non-alfabet, dilanjutkan dengan penghapusan *stopwords*, yaitu kata-kata umum yang tidak memiliki kontribusi berarti terhadap proses analisis teks. Proses selanjutnya adalah *stemming*, yang bertujuan mengubah berbagai bentuk kata menjadi bentuk dasar yang seragam, sehingga representasi kata dengan makna serupa dapat diproses secara lebih konsisten dalam analisis. Tahap ini dilakukan untuk membersihkan data teks dari noise serta menyelaraskan representasi kata agar lebih konsisten dalam dokumen. Setelah proses praproses selesai, ulasan pengguna kemudian dikelompokkan ke dalam kategori layanan berdasarkan kemunculan kata kunci tertentu yang merepresentasikan aspek layanan aplikasi.

Setiap tahapan dalam proses persiapan data menghasilkan dataset yang semakin terstruktur dan relevan untuk dianalisis. Melalui proses tersebut, kualitas data menjadi lebih baik sehingga siap digunakan pada tahap selanjutnya, yaitu ekstraksi fitur dan proses klasifikasi sentimen. Ringkasan jumlah data pada setiap tahapan persiapan dataset disajikan pada Tabel 1, yang menunjukkan proses penyaringan dan transformasi data hingga didapatkan dataset akhir yang dipakai dalam analisis sentimen berbasis kategori layanan.

Tabel 1. Ringkasan Tahapan Persiapan Dataset Ulasan Mobile JKN

Tahapan Dataset	Deskripsi	Jumlah Ulasan
Data awal ( <i>raw</i> )	Ulasan hasil pengambilan langsung dari <i>Google Play Store</i>	5.000
Setelah <i>cleaning</i> dan seleksi data	Data duplikat, ulasan tanpa teks, dan data tidak relevan dihapus	3.630
Setelah <i>text preprocessing</i>	Data teks telah melalui <i>case folding</i> , penghapusan <i>stopwords</i> , dan <i>stemming</i>	3.630
Dataset akhir terkategori	Data dipetakan ke dalam kategori layanan aplikasi	3.630

## H. Validitas dan Keandalan Data

Validitas data dalam penelitian ini dijaga melalui serangkaian tahapan persiapan data yang dilakukan secara sistematis, yang mencakup proses pembersihan data (data cleaning), pelabelan sentimen, serta praproses teks. Untuk menjaga kualitas data, dilakukan pembersihan data dengan menghapus ulasan duplikat dan ulasan tanpa isi teks, sehingga data yang digunakan dapat merepresentasikan pendapat pengguna secara lebih akurat. Pendekatan rating-based labeling digunakan untuk memberikan label sentimen secara konsisten dan terukur pada dataset berskala besar, sehingga memungkinkan analisis sentimen dilakukan secara efisien.

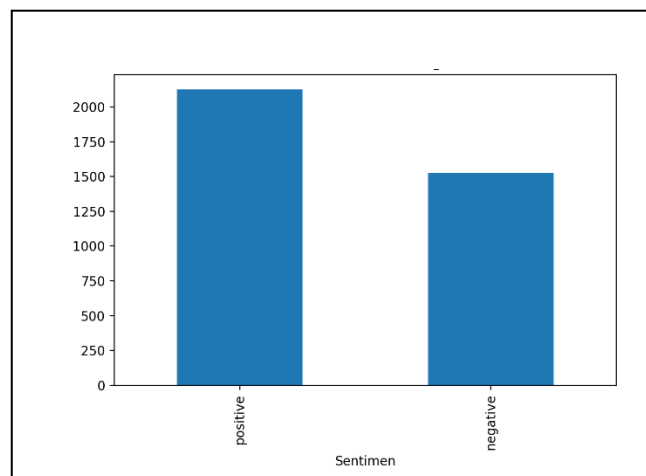
Keandalan (reliability) analisis didukung oleh penggunaan metode representasi teks dan algoritma klasifikasi yang sangat sering digunakan dalam penelitian analisis sentimen. *TF-IDF* diterapkan sebagai metode untuk merepresentasikan fitur teks dalam proses analisis, yang memungkinkan setiap kata diberi bobot berdasarkan tingkat kepentingannya dalam keseluruhan korpus ulasan. Sementara itu, algoritma *logistic regression* dipilih karena stabilitas, interpretabilitas, serta efisiensi komputasinya dalam mengolah data teks berukuran menengah. Performa model klasifikasi dalam penelitian ini dinilai menggunakan metrik accuracy dan didukung oleh analisis confusion matrix untuk mengevaluasi ketepatan hasil klasifikasi.

Penggunaan pelabelan sentimen berbasis *rating* memungkinkan adanya ketidaksesuaian antara *rating numerik* dan isi teks ulasan pengguna. Selain itu, pendekatan kategorisasi layanan berbasis kata kunci berpotensi belum sepenuhnya menangkap konteks semantik yang kompleks pada beberapa ulasan. Keterbatasan ini menjadi peluang untuk pengembangan penelitian selanjutnya dengan memanfaatkan pendekatan pelabelan manual atau metode deep learning yang lebih mampu menangkap konteks bahasa secara mendalam.

## III. Hasil dan Pembahasan

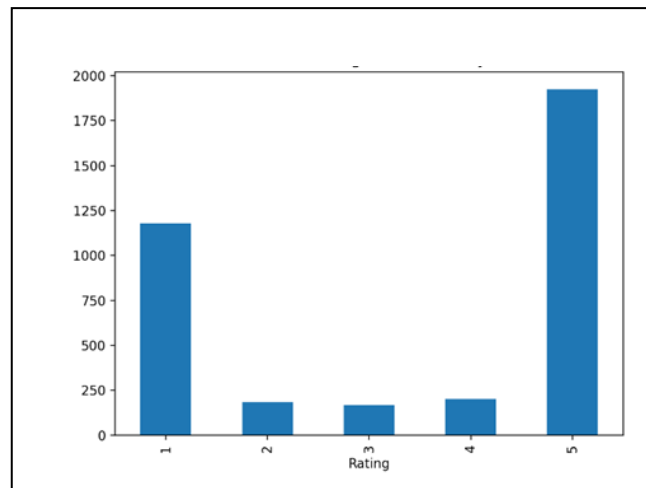
### A. Hasil Klasifikasi Sentimen

Penelitian ini menganalisis sebanyak 3.630 ulasan pengguna aplikasi Mobile JKN yang bersumber dari *Google Play Store*. Distribusi sentimen berdasarkan pelabelan berbasis rating menunjukkan bahwa sentimen positif lebih dominan dibandingkan sentimen negatif.



Gambar 1. Distribusi Sentimen Ulasan Aplikasi Mobile JKN

Distribusi rating ulasan pengguna menunjukkan bahwa sebagian besar ulasan memiliki rating ekstrem, yaitu satu bintang dan lima bintang, sebagaimana ditunjukkan pada Gambar 2. Pola ini mengindikasikan bahwa rating numerik belum tentu sepenuhnya merepresentasikan persepsi pengguna terhadap kualitas layanan aplikasi.



Gambar 2. Distribusi Rating Ulasan Pengguna Aplikasi Mobile JKN

Hasil evaluasi performa model klasifikasi sentimen berbasis fitur *TF-IDF* dengan penerapan algoritma *logistic regression* ditampilkan pada Tabel 2. Model mencapai nilai *accuracy* sebesar 92,15%, dengan *precision* 96,42%, *recall* 89,79%, dan *F1-score* 92,99%. Hasil ini menunjukkan bahwa pendekatan pembelajaran mesin klasik mampu memberikan performa yang stabil dan efektif dalam proses klasifikasi sentimen ulasan pengguna aplikasi Mobile JKN.

Tabel 2. Kinerja Model Klasifikasi Sentimen

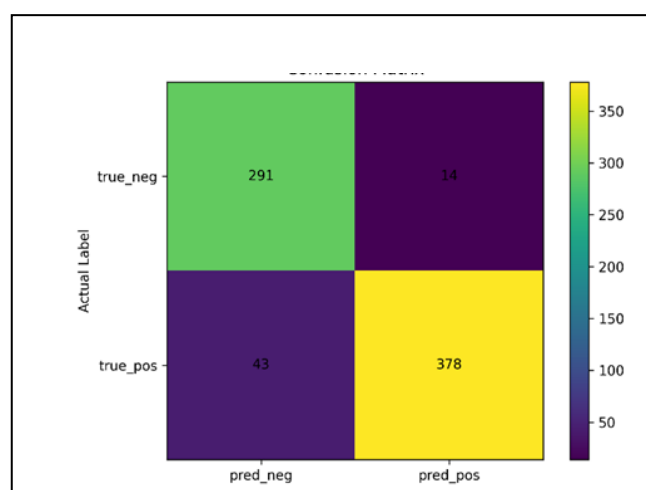
Model	Accuracy	Precision	Recall	F1-Score
TF-IDF + Logistic Regression	92,15%	96,42 %	89,79 %	92,99 %

### B. Analisis Confusion Matrix

Analisis lanjutan terhadap tingkat keberhasilan maupun kesalahan klasifikasi dilakukan dengan menggunakan confusion matrix. Hasil analisis tersebut disajikan dalam bentuk tabel dan visualisasi grafis, yang ditampilkan pada Tabel 3 dan Gambar 3, sehingga memudahkan dalam mengidentifikasi distribusi prediksi model pada setiap kelas sentimen.

Tabel 3. Confusion Matrix Klasifikasi Sentimen

Aktual \ Prediksi	Negatif	Positif
Negatif	291	14
Positif	43	378

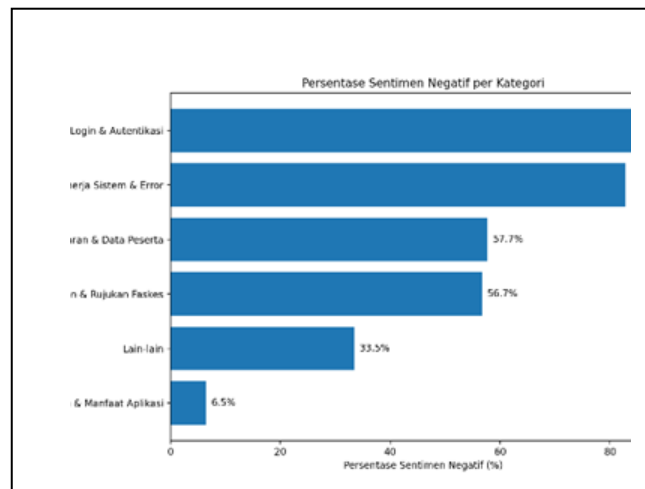


Gambar 3. Visualisasi Confusion Matrix Klasifikasi Sentimen

Hasil evaluasi menggunakan *confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan mayoritas ulasan dengan tingkat ketepatan yang baik. Namun demikian, masih ditemukan kesalahan klasifikasi yang sebagian besar disebabkan oleh ketidaksesuaian antara rating bintang dan isi teks ulasan.

### C. Analisis Sentimen Berdasarkan Kategori Layanan

Untuk mendapatkan pengetahuan yang lebih mendalam terkait aspek layanan aplikasi Mobile JKN, dilakukan analisis sentimen berdasarkan kategori layanan. Hasil analisis persentase sentimen negatif pada masing-masing kategori layanan ditunjukkan pada Gambar 4.



Gambar 4. Persentase Sentimen Negatif Berdasarkan Kategori Layanan Aplikasi Mobile JKN

Hasil analisis menunjukkan bahwa kategori *Login & Autentikasi* memiliki persentase sentimen negatif tertinggi sebesar 87,6%, diikuti oleh kategori *Kinerja Sistem & Error* sebesar 82,76%. Sebaliknya, kategori *Layanan & Manfaat Aplikasi* menunjukkan persentase sentimen negatif terendah, yaitu 6,46%, yang mengindikasikan tingkat kepuasan pengguna yang relatif tinggi pada aspek tersebut.

Temuan penelitian ini menunjukkan bahwa pemanfaatan ulasan pengguna dalam analisis sentimen mampu memberikan gambaran yang komprehensif terkait persepsi pengguna terhadap kualitas layanan pada aplikasi Mobile JKN. Dominasi sentimen negatif pada kategori autentikasi dan stabilitas sistem menandakan bahwa aspek keandalan teknis masih menjadi tantangan utama dalam implementasi aplikasi layanan publik berbasis mobile.

Penerapan pendekatan *rating-based labeling* yang dipadukan dengan klasifikasi sentimen berbasis teks terbukti efektif dalam mengolah data ulasan dalam jumlah besar dengan tingkat akurasi yang tinggi. Selain itu, integrasi antara hasil klasifikasi sentimen dan pemetaan kategori layanan memungkinkan penyusunan rekomendasi perbaikan layanan yang lebih terarah, spesifik, dan operasional, sehingga dapat dimanfaatkan sebagai dasar dalam meningkatkan kualitas layanan aplikasi.

## IV. Kesimpulan dan saran

Penelitian ini menerapkan pendekatan *machine learning* untuk melakukan analisis sentimen pada ulasan pengguna aplikasi Mobile JKN, dengan menggunakan *TF-IDF* sebagai metode representasi fitur dan *logistic regression* sebagai model klasifikasi. Berdasarkan analisis terhadap 3.630 ulasan pengguna yang didapatkan dari platform *Google Play Store*, hasil penelitian menunjukkan bahwa pendekatan yang digunakan dapat melakukan klasifikasi sentimen ulasan dengan akurasi yang tinggi, yaitu 92,15%, serta mendapatkan nilai *precision* sebesar 96,42%, *recall* sebesar 89,79%, dan *F1-score* sebesar 92,99%. Hal ini mengindikasikan bahwa metode pembelajaran mesin klasik masih efektif dan relevan untuk analisis sentimen teks pada data ulasan aplikasi berskala menengah, terutama dalam konteks evaluasi layanan publik berbasis aplikasi mobile. Hasil analisis sentimen berdasarkan kategori layanan menunjukkan bahwa permasalahan utama yang memicu sentimen negatif pengguna terletak pada aspek login dan autentikasi, khususnya terkait kegagalan proses masuk dan pengiriman kode *OTP*, serta pada aspek kinerja dan stabilitas sistem. Sebaliknya, kategori layanan yang berkaitan dengan manfaat dan kemudahan penggunaan aplikasi menunjukkan tingkat sentimen negatif yang relatif rendah, yang mengindikasikan bahwa pengguna pada umumnya mengapresiasi nilai guna layanan yang disediakan oleh aplikasi Mobile JKN. Temuan ini menegaskan bahwa ulasan pengguna dapat digunakan sebagai sumber informasi strategis dalam mengidentifikasi prioritas dalam meningkatkan kualitas layanan digital. Integrasi antara analisis sentimen berbasis teks dan pemetaan kategori layanan memungkinkan identifikasi aspek layanan yang perlu diprioritaskan untuk perbaikan secara lebih terarah dan operasional. Temuan penelitian ini dapat dijadikan sebagai referensi atau bahan pertimbangan bagi pengelola aplikasi Mobile JKN dalam meningkatkan kualitas layanan digital, khususnya pada aspek yang paling banyak dikeluhkan oleh pengguna. Pendekatan ini juga berpotensi diterapkan pada aplikasi layanan publik lainnya sebagai mekanisme evaluasi berbasis data yang efisien dan berkelanjutan. Sebagai rekomendasi untuk penelitian berikutnya, eksplorasi terhadap penerapan pendekatan *deep learning* perlu dipertimbangkan guna

meningkatkan kualitas analisis serta teknik pelabelan sentimen berbasis anotasi manual atau semi supervised learning guna meningkatkan ketepatan klasifikasi dan menangkap nuansa sentimen yang lebih kompleks. Selain itu, analisis temporal terhadap ulasan pengguna juga dapat dilakukan untuk mengamati perubahan persepsi pengguna seiring dengan pembaruan sistem dan kebijakan layanan. Pengembangan lebih lanjut juga dapat mempertimbangkan pendekatan aspect based *sentiment analysis* agar pemetaan opini pengguna terhadap fitur layanan dapat dilakukan secara lebih rinci

#### Daftar Pustaka

- [1] E. V Eppes, M. Augustyn, S. M. Gross, P. Vernon, L. E. Caulfield, and D. M. Paige, "Engagement With and Acceptability of Digital Media Platforms for Use in Improving Health Behaviors Among Vulnerable Families: Systematic Review," *J Med Internet Res*, vol. 25, p. e40934, 2023, doi: 10.2196/40934.
- [2] M. Ariyanti, S. Widiyanesti, and W. H. Aprillia, "Service Quality Analysis of Telkomsel Case Study Based on Online Customer Reviews in Google Play Store," in *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)*, 2024, pp. 1669–1674. doi: 10.1109/IC2PCT60090.2024.10486619.
- [3] L. Çallı, "Exploring mobile banking adoption and service quality features through user-generated content: the application of a topic modeling approach to Google Play Store reviews," *International Journal of Bank Marketing*, vol. 41, no. 2, pp. 428–454, Dec. 2022, doi: 10.1108/IJBM-08-2022-0351.
- [4] Y. Astuti, Y. Ruldeviyani, F. Salbari, and A. Prayogi, "Sentiment Analysis of Electricity Company Service Quality Using Naïve Bayes," *Jurnal RESTI*, vol. 7, no. 2, pp. 389–396, Apr. 2023, doi: 10.29207/resti.v7i2.4627.
- [5] S. Sadiq, M. Umer, S. Ullah, S. Mirjalili, V. Rupapara, and M. Nappi, "Discrepancy detection between actual user reviews and numeric ratings of Google App store using deep learning," *Expert Syst. Appl.*, vol. 181, p. 115111, 2021, doi: <https://doi.org/10.1016/j.eswa.2021.115111>.
- [6] P. H. C. Samanmali and R. A. H. M. Rupasingha, "Sentiment analysis on google play store app users' reviews based on deep learning approach," *Multimed. Tools Appl.*, vol. 83, no. 36, pp. 84425–84453, 2024, doi: 10.1007/s11042-024-19185-w.
- [7] L. Lu, P. Xu, Y.-Y. Wang, and Y. Wang, "Measuring service quality with text analytics: Considering both importance and performance of consumer opinions on social and non-social online platforms," *J. Bus. Res.*, vol. 169, p. 114298, 2023, doi: <https://doi.org/10.1016/j.jbusres.2023.114298>.
- [8] X. Wang, T. Zhang, Y. Tan, W. Shang, and Y. Li, "How to effectively mine app reviews concerning software ecosystem? A survey of review characteristics," *Journal of Systems and Software*, vol. 213, p. 112040, 2024, doi: <https://doi.org/10.1016/j.jss.2024.112040>.
- [9] N. Genc-Nayebi and A. Abran, "A systematic literature review: Opinion mining studies from mobile app store user reviews," *Journal of Systems and Software*, vol. 125, pp. 207–219, 2017, doi: <https://doi.org/10.1016/j.jss.2016.11.027>.
- [10] S. A. H. Bahtiar, C. K. Dewa, and A. Luthfi, "Comparison of Naïve Bayes and Logistic Regression in Sentiment Analysis on Marketplace Reviews Using Rating-Based Labeling," *Journal of Information Systems and Informatics*, vol. 5, no. 3, pp. 915–927, Aug. 2023, doi: 10.51519/journalisi.v5i3.539.
- [11] S. Al-Hagree and G. Al-Gaphari, "Arabic Sentiment Analysis Based Machine Learning for Measuring User Satisfaction with Banking Services' Mobile Applications: Comparative Study," in *2022 2nd International Conference on Emerging Smart Technologies and Applications (eSmarTA)*, 2022, pp. 1–4. doi: 10.1109/eSmarTA56775.2022.9935486.
- [12] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014, doi: <https://doi.org/10.1016/j.asej.2014.04.011>.
- [13] F. B. Wijaya, W. Budiaji, R. P. R. Bagaskara, Z. A. Tazkia, and D. D. A. Pertiwi, "The Digital Frontline: A Thematic Analysis of User Grievances and Satisfaction Drivers for Indonesian Public Service Apps," in *Proceedings of The International Conference on Data Science and Official Statistics*, 2025, pp. 587–612.
- [14] M. L. F. Martanto and W. Istiono, "Sentiment Analysis of M-Paspor App Reviews Using Multinomial Naive Bayes," *Journal of Logistics, Informatics and Service Science*, vol. 11, no. 10, pp. 311–326, 2024, doi: 10.33168/JLISS.2024.1017.
- [15] I. Arief, M. Farhandika, A. S. Indrapriyatna, A. A. Yulianto, and Y. Meuthia, "Enhancing User Interface and Experience of the Bukalapak Application: A Sentiment Analysis Approach for Improved Usability and User Satisfaction in Indonesia's E-Commerce Sector," *Jurnal RESTI*, vol. 7, no. 5, pp. 1192–1204, Oct. 2023, doi: 10.29207/resti.v7i5.5184.
- [16] G. Tamami, W. A. Triyanto, and S. Muzid, "Sentiment Analysis Mobile JKN Reviews Using SMOTE Based LSTM," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 19, no. 1, p. 13, Jan. 2025, doi: 10.22146/ijccs.101910.

- [17] M. Y. Al Qahar, Y. Ruldeviyani, U. N. Mukharomah, M. A. Fidyawan, and R. Putra, "Factor analysis influencing Mobile JKN user experience using sentiment analysis," *IAES International Journal of Artificial Intelligence*, vol. 13, no. 2, pp. 1782–1793, Jun. 2024, doi: 10.11591/ijai.v13.i2.pp1782-1793.
- [18] C. I. Ossai and N. Wickramasinghe, "Automatic user sentiments extraction from diabetes mobile apps – An evaluation of reviews with machine learning," *Inform. Health Soc. Care*, vol. 48, no. 3, pp. 211–230, Jul. 2023, doi: 10.1080/17538157.2022.2097083.
- [19] N. Haug, S. Dan, and I. Mergel, "Digitally-induced change in the public sector: a systematic review and research agenda," *Public Management Review*, vol. 26, no. 7, pp. 1963–1987, Jul. 2024, doi: 10.1080/14719037.2023.2234917.
- [20] R. Massenon *et al.*, "Mobile app review analysis for crowdsourcing of software requirements: a mapping study of automated and semi-automated tools," *PeerJ Comput. Sci.*, vol. 10, p. e2401, Nov. 2024, doi: 10.7717/peerj-cs.2401.
- [21] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, "Mining and searching app reviews for requirements engineering: Evaluation and replication studies," *Inf. Syst.*, vol. 114, p. 102181, 2023, doi: <https://doi.org/10.1016/j.is.2023.102181>.
- [22] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, "Analysing app reviews for software engineering: a systematic literature review," *Empir. Softw. Eng.*, vol. 27, no. 2, p. 43, 2022, doi: 10.1007/s10664-021-10065-7.
- [23] S. McIlroy, N. Ali, H. Khalid, and A. E. Hassan, "Analyzing and automatically labelling the types of user issues that are raised in mobile app reviews," *Empir. Softw. Eng.*, vol. 21, no. 3, pp. 1067–1106, 2016, doi: 10.1007/s10664-015-9375-7.
- [24] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc., 2009.
- [25] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513–523, 1988, doi: [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0).