

ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI MENGGUNAKAN PENDEKATAN MACHINE LEARNING BERBASIS TEXT MINING

Mutiara Fatmiah¹, Rudi Kurniawan², Bani Nurhakim³, Aris Pratama Putra⁴, Tati Suprapti⁵

Program Studi Teknik Informatika^{1,2,5}
Program Studi Manajemen Informatika³
Program Studi Sistem Informasi⁴

STMIK IKMI Cirebon
<https://ikmi.ac.id/page/18/?lang=de>
mutiarafatmiah9@gmail.com

(*) Corresponding Author : mutiarafatmiah9@gmail.com
Published : 30 Maret 2026

Abstract— The rapid growth of mobile applications has led to an increasing volume of user reviews that can be utilized as a valuable source of information for systematically evaluating application quality. This study aims to analyze the sentiment of user reviews of the Kopi Kenangan application on the Google Play Store using a machine learning approach based on text mining. Review data were collected through web scraping and classified into three sentiment classes: positive, neutral, and negative. Initial labeling was performed based on star ratings and subsequently manually verified to improve label consistency. The research stages include text cleaning, case folding, tokenization, stopword removal, and stemming using Sastrawi. Class imbalance was addressed using Random Oversampling, while feature extraction was carried out using Term Frequency–Inverse Document Frequency (TF–IDF). The Multinomial Naïve Bayes classifier was employed and evaluated using accuracy, precision, recall, F1-score, and 5-fold cross-validation. Experimental results show that the model achieved an accuracy of 98.98%, with macro-averaged precision, recall, and F1-score of 99.00%, and demonstrated consistent performance across different data subsets. These findings indicate that a machine learning approach based on text mining is effective for sentiment analysis of Indonesian-language application reviews and can provide relevant insights for application developers in improving service quality.

Keywords: Sentiment analysis, Machine learning, Text mining, Multinomial Naïve Bayes, TF–IDF.

Abstrak- Pertumbuhan aplikasi mobile mendorong peningkatan jumlah ulasan pengguna yang dapat dimanfaatkan sebagai sumber informasi untuk mengevaluasi kualitas aplikasi secara sistematis. Penelitian ini bertujuan menganalisis sentimen ulasan pengguna aplikasi Kopi Kenangan pada platform Google Play Store menggunakan pendekatan machine learning berbasis text mining. Data dikumpulkan melalui web scraping dan diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif. Pelabelan awal dilakukan berdasarkan rating bintang dan diverifikasi secara manual untuk meningkatkan konsistensi label. Tahapan penelitian meliputi pembersihan teks, case folding, tokenisasi, penghapusan stopword, dan stemming menggunakan Sastrawi. Ketidakseimbangan kelas ditangani menggunakan Random Oversampling, sedangkan ekstraksi fitur dilakukan dengan Term Frequency–Inverse Document Frequency (TF–IDF). Model klasifikasi yang digunakan adalah Multinomial Naïve Bayes dan dievaluasi menggunakan metrik accuracy, precision, recall, F1-score, serta validasi silang 5-fold. Hasil eksperimen menunjukkan bahwa model mencapai akurasi sebesar 98,98% dengan nilai precision, recall, dan F1-score makro masing-masing sebesar 99,00%, serta menunjukkan kinerja yang konsisten pada berbagai subset data. Temuan ini menunjukkan bahwa pendekatan machine learning berbasis text mining efektif untuk menganalisis sentimen ulasan aplikasi berbahasa Indonesia dan dapat memberikan masukan yang relevan bagi pengembang aplikasi dalam meningkatkan kualitas layanan.

Kata kunci: Analisis sentimen, Machine learning, Text mining, Multinomial Naïve Bayes, TF–IDF.

INTRODUCTION

Pertumbuhan aplikasi mobile yang pesat telah mengubah pola interaksi antara pengguna dan penyedia layanan digital. Salah satu implikasi utama

dari fenomena ini adalah meningkatnya jumlah ulasan pengguna pada platform distribusi aplikasi seperti Google Play Store, yang berisi opini, pengalaman, serta evaluasi spontan terhadap

kualitas aplikasi. Ulasan pengguna tersebut merupakan sumber data tidak terstruktur yang bernilai tinggi karena merefleksikan persepsi pengguna secara langsung dan real time. Namun, volume ulasan yang terus bertambah menjadikan proses analisis manual tidak lagi efisien dan sulit dilakukan secara konsisten, sehingga diperlukan pendekatan otomatis berbasis komputasi untuk mengekstraksi informasi yang relevan secara sistematis [1]

Analisis sentimen menjadi salah satu pendekatan utama dalam text mining yang bertujuan mengidentifikasi polaritas opini pengguna, seperti sentimen positif, netral, dan negatif. Pendekatan ini banyak dimanfaatkan dalam berbagai domain, termasuk evaluasi layanan digital dan aplikasi mobile, karena mampu memberikan gambaran kuantitatif mengenai tingkat kepuasan dan permasalahan pengguna. Penelitian terkini menunjukkan bahwa analisis sentimen berbasis machine learning masih menjadi pendekatan yang dominan karena mampu memberikan kinerja yang stabil dan interpretatif, terutama pada data ulasan aplikasi yang bersifat pendek dan kontekstual [2]. Pendekatan supervised machine learning banyak digunakan dalam analisis sentimen karena kemampuannya mempelajari pola linguistik dari data berlabel dan menghasilkan prediksi yang konsisten. Meskipun model deep learning seperti LSTM dan transformer menunjukkan performa yang unggul pada dataset berskala besar, beberapa penelitian terbaru menegaskan bahwa algoritma machine learning klasik tetap relevan dan kompetitif pada dataset berukuran kecil hingga menengah, terutama ketika dikombinasikan dengan teknik ekstraksi fitur yang tepat [3]. Selain itu, model klasik cenderung memiliki kompleksitas komputasi yang lebih rendah dan interpretabilitas yang lebih baik, sehingga lebih sesuai untuk penelitian terapan dan studi kasus spesifik.

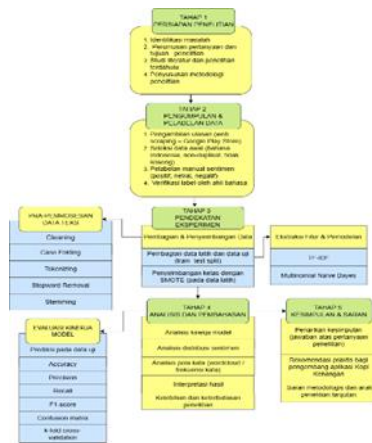
Pada konteks ulasan aplikasi berbahasa Indonesia, tantangan yang dihadapi menjadi lebih kompleks. Ulasan pengguna umumnya ditulis dalam bahasa informal, mengandung singkatan, variasi ejaan, serta campuran bahasa, yang dapat menurunkan kinerja model klasifikasi jika tidak ditangani dengan baik. Penelitian terbaru menekankan bahwa tahapan pra-pemrosesan teks, seperti pembersihan data, normalisasi kata, dan stemming, memiliki pengaruh signifikan terhadap kualitas representasi fitur dan performa klasifikasi sentimen [4]. Selain itu, distribusi kelas sentimen yang tidak seimbang juga menjadi permasalahan umum pada data ulasan aplikasi, di mana ulasan positif cenderung mendominasi, sehingga berpotensi menimbulkan bias prediksi terhadap kelas mayoritas [5].

Algoritma Multinomial Naïve Bayes merupakan salah satu metode probabilistik yang banyak digunakan dalam klasifikasi teks karena kesederhanaan, efisiensi komputasi, dan kemampuannya menangani data berdimensi tinggi. Studi terbaru menunjukkan bahwa kombinasi Multinomial Naïve Bayes dengan representasi fitur berbasis Term Frequency-Inverse Document Frequency (TF-IDF) masih mampu menghasilkan kinerja yang kompetitif dalam analisis sentimen ulasan aplikasi, khususnya pada bahasa dengan struktur morfologi yang kompleks seperti bahasa Indonesia [6]. Pendekatan ini juga dinilai lebih stabil ketika diterapkan pada dataset dengan ukuran terbatas dibandingkan model yang lebih kompleks.

Berdasarkan latar belakang tersebut, penelitian ini memfokuskan pada analisis sentimen ulasan pengguna aplikasi Kopi Kenangan menggunakan pendekatan machine learning berbasis text mining. Penelitian ini berkontribusi dengan menyajikan pipeline pemrosesan teks yang sistematis, mencakup pra-pemrosesan data, penanganan ketidakseimbangan kelas, serta pemodelan menggunakan Multinomial Naïve Bayes dengan fitur TF-IDF. Evaluasi kinerja model dilakukan secara komprehensif untuk memastikan keandalan dan kemampuan generalisasi. Diharapkan hasil penelitian ini tidak hanya memberikan gambaran empiris mengenai sentimen pengguna terhadap aplikasi, tetapi juga menjadi referensi metodologis bagi penelitian analisis sentimen ulasan aplikasi berbahasa Indonesia pada konteks dan domain serupa.

MATERIALS AND METHODS

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional berbasis supervised machine learning. Pendekatan ini bertujuan untuk membangun model klasifikasi sentimen dengan mempelajari pola dari data teks yang telah diberi label, serta mengevaluasi kinerjanya menggunakan metrik kuantitatif yang terukur dan dapat direproduksi. Desain penelitian ini memungkinkan pengujian performa model secara objektif melalui tahapan eksperimen yang sistematis dan terstruktur.



Gambar 1 Tahapan Penelitian

Pengumpulan Data

Data penelitian berupa ulasan pengguna aplikasi Kopi Kenangan yang diperoleh dari platform Google Play Store. Pengumpulan data dilakukan menggunakan teknik web scraping untuk memperoleh ulasan teks beserta informasi rating bintang. Data yang diperoleh kemudian diseleksi untuk memastikan hanya ulasan berbahasa Indonesia yang digunakan, tidak bersifat duplikat, dan tidak mengandung teks kosong, sehingga kualitas data tetap terjaga sebelum masuk ke tahap pemrosesan berikutnya.

Pelabelan Data

Pelabelan sentimen dilakukan melalui dua tahap. Tahap pertama adalah pelabelan awal berdasarkan rating bintang, di mana ulasan dengan rating tinggi dikategorikan sebagai sentimen positif, rating rendah sebagai sentimen negatif, dan rating menengah sebagai sentimen netral. Tahap kedua dilakukan melalui verifikasi manual untuk memastikan kesesuaian isi ulasan dengan label sentimen yang diberikan, khususnya pada ulasan yang bersifat ambigu atau mengandung lebih dari satu kecenderungan sentimen.

Pra-pemrosesan Teks

Pra-pemrosesan teks dilakukan untuk menormalkan data dan mengurangi noise yang dapat memengaruhi kinerja model. Tahapan pra-pemrosesan meliputi pembersihan teks dari karakter non-alfabet dan simbol yang tidak relevan, case folding untuk menyeragamkan bentuk huruf, tokenisasi untuk memecah teks menjadi unit kata, penghapusan stopword, serta stemming untuk mengembalikan kata ke bentuk dasarnya. Tahapan ini bertujuan menghasilkan representasi teks yang lebih ringkas dan informatif.

Penanganan Ketidakseimbangan Data

Distribusi kelas sentimen pada data ulasan pengguna cenderung tidak seimbang, yang berpotensi menyebabkan bias model terhadap kelas mayoritas. Untuk mengatasi permasalahan tersebut, penelitian ini menerapkan teknik penyeimbangan data pada data latih dengan menambah jumlah sampel pada kelas minoritas. Pendekatan ini dilakukan untuk meningkatkan kemampuan model dalam mengenali seluruh kelas sentimen secara lebih proporsional.

Model Klasifikasi

Model klasifikasi yang digunakan dalam penelitian ini adalah Multinomial Naïve Bayes. Algoritma ini bekerja berdasarkan pendekatan probabilistik untuk memprediksi kelas sentimen suatu ulasan berdasarkan distribusi kata yang muncul dalam teks. Model ini dipilih karena memiliki efisiensi komputasi yang baik dan mampu menangani data teks berdimensi tinggi.

Evaluasi Model

Evaluasi kinerja model dilakukan menggunakan data uji yang terpisah dari data latih. Kinerja model diukur menggunakan beberapa metrik evaluasi, yaitu accuracy, precision, recall, dan F1-score. Metrik accuracy digunakan untuk menggambarkan tingkat ketepatan prediksi secara keseluruhan, sedangkan precision dan recall digunakan untuk menilai kemampuan model dalam mengklasifikasikan setiap kelas sentimen secara spesifik. F1-score digunakan sebagai ukuran keseimbangan antara precision dan recall, khususnya pada kondisi distribusi kelas yang tidak seimbang. Selain itu, confusion matrix digunakan untuk menganalisis pola kesalahan klasifikasi pada setiap kelas sentimen.

Validasi Silang

Untuk menguji konsistensi dan kemampuan generalisasi model, penelitian ini menerapkan teknik validasi silang k-fold dengan nilai k sebesar 5. Dataset dibagi menjadi lima sub.

RESULTS AND DISCUSSION

Dataset

Dataset penelitian terdiri dari ulasan pengguna aplikasi yang diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif, netral, dan negatif. Data telah melalui proses pelabelan sehingga setiap teks memiliki kategori sentimen yang jelas. Dataset kemudian dibagi menjadi data pelatihan dan data pengujian dengan skema stratified split untuk menjaga keseimbangan distribusi kelas.

Hasil pembagian menunjukkan bahwa data pelatihan memiliki distribusi yang seimbang dengan masing-masing 651 data per kelas, sedangkan data pengujian memiliki 163 data per kelas. Keseimbangan ini penting untuk mencegah bias model terhadap kelas mayoritas dan memastikan evaluasi performa yang adil.

```
Distribusi y_train:
Label
positif    651
netral     651
negatif    651
Name: count, dtype: int64

Distribusi y_test:
Label
netral     163
positif    163
negatif    163
Name: count, dtype: int64
```

Gambar 2. Data Train dan Test

Text Preprocessing

Tahapan pra-pemrosesan teks dilakukan untuk meningkatkan kualitas data sebelum proses ekstraksi fitur. Tahapan ini meliputi:

1. konversi huruf menjadi lowercase
2. penghapusan tanda baca dan karakter non-alfabet
3. penghapusan angka dan symbol
4. pembersihan spasi berlebih
5. normalisasi teks

Proses ini bertujuan untuk mengurangi noise dan memastikan konsistensi representasi kata sehingga proses ekstraksi fitur menjadi lebih efektif.

Feature Extraction Menggunakan TF-IDF

Representasi numerik teks dilakukan menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). Metode ini mengubah dokumen teks menjadi vektor berbobot berdasarkan frekuensi kata dalam dokumen dan tingkat kelangkaannya di seluruh korpus.

TF-IDF dipilih karena mampu menekan pengaruh kata umum yang sering muncul dan memperkuat kontribusi kata yang lebih informatif dalam membedakan sentimen.

Model Klasifikasi: Multinomial Naïve Bayes

Model klasifikasi yang digunakan adalah Multinomial Naïve Bayes, yang sesuai untuk data berbasis frekuensi kata. Model dilatih menggunakan fitur TF-IDF dari data pelatihan. Algoritma ini bekerja dengan pendekatan probabilistik untuk menghitung kemungkinan

suatu dokumen termasuk dalam kelas sentimen tertentu berdasarkan distribusi kata.

Model dilatih menggunakan data latih yang telah diproses, kemudian digunakan untuk memprediksi label sentimen pada data uji.

Experimental Setup

Eksperimen dilakukan dengan alur berikut:

1. Pembagian data train-test secara stratified
2. Transformasi teks menggunakan TF-IDF vectorizer
3. Pelatihan model Multinomial Naïve Bayes
4. Prediksi pada data uji
5. Evaluasi performa menggunakan metrik klasifikasi

Seluruh proses dilakukan menggunakan pipeline pemrosesan teks dan klasifikasi terintegrasi untuk menjaga konsistensi eksperimen.

Evaluation Metrics

Evaluasi model dilakukan menggunakan beberapa metrik kinerja klasifikasi, yaitu:

1. Accuracy
2. Precision
3. Recall
4. F1-score
5. Confusion Matrix

Classification report digunakan untuk menampilkan performa model pada setiap kelas sentimen secara rinci. Confusion matrix digunakan untuk menganalisis distribusi prediksi benar dan salah antar kelas.

Laporan klasifikasi menampilkan nilai precision, recall, dan F1-score untuk setiap kelas sentimen (negatif, netral, dan positif) pada data uji. Hasil menunjukkan bahwa model memiliki kinerja yang sangat tinggi dan seimbang antar kelas, dengan nilai rata-rata metrik mendekati 1,00 serta akurasi keseluruhan sebesar 99%, yang mengindikasikan kemampuan klasifikasi yang sangat baik dimuat pada gambar 3.

```
=== Classification Report (TEST) ===
              precision    recall  f1-score   support

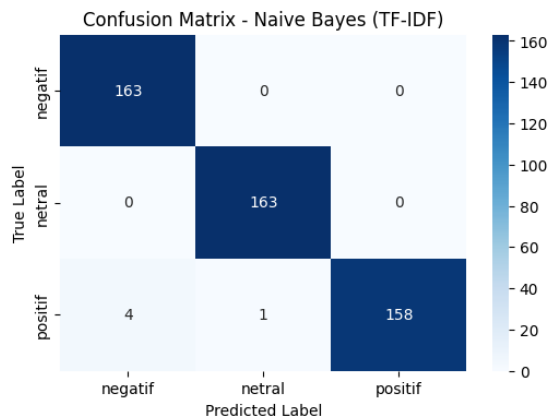
negatif      0.98         1.00         0.99         163
netral       0.99         1.00         1.00         163
positif      1.00         0.97         0.98         163

accuracy                0.99         489
macro avg              0.99         0.99         0.99         489
weighted avg          0.99         0.99         0.99         489

=== Confusion Matrix (Raw) ===
[[163  0  0]
 [  0 163  0]
 [  4  1 158]]
```

Gambar 3. Classification Report Model Naïve Bayes

Selain itu, Confusion matrix memperlihatkan distribusi prediksi model terhadap label aktual pada tiga kelas sentimen. Sebagian besar data berhasil diklasifikasikan dengan benar pada diagonal matriks, dengan jumlah kesalahan prediksi yang sangat kecil, terutama pada kelas positif. Visualisasi ini menunjukkan bahwa model mampu membedakan setiap kelas sentimen secara efektif dan konsisten. dimuat pada Gambar 4.



Gambar 4. Confusion Matrix

Visualisasi ini membantu menunjukkan bahwa model mampu mengenali seluruh kelas dengan akurasi tinggi.

CONCLUSION

Penelitian ini berhasil menerapkan pendekatan text mining dan machine learning untuk melakukan klasifikasi sentimen ulasan pengguna aplikasi menggunakan representasi fitur TF-IDF dan algoritma Multinomial Naïve Bayes. Tahapan pemrosesan yang sistematis, mulai dari pembersihan teks, ekstraksi fitur, hingga pembagian data yang seimbang antar kelas, berkontribusi terhadap kestabilan dan akurasi model dalam proses klasifikasi.

Hasil pengujian menunjukkan bahwa model mampu mencapai tingkat akurasi yang sangat tinggi, dengan nilai precision, recall, dan F1-score yang konsisten pada setiap kelas sentimen. Analisis confusion matrix memperlihatkan bahwa sebagian besar data berhasil diklasifikasikan dengan benar, dengan jumlah kesalahan prediksi yang relatif kecil. Selain itu, hasil validasi silang 5-fold menunjukkan performa yang stabil di setiap

fold, yang mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik terhadap variasi data.

Secara keseluruhan, kombinasi TF-IDF dan Multinomial Naïve Bayes terbukti efektif dan efisien untuk analisis sentimen ulasan aplikasi berbahasa Indonesia. Pendekatan ini dapat dijadikan dasar untuk pengembangan sistem pemantauan opini pengguna secara otomatis, serta dapat diperluas pada penelitian selanjutnya dengan membandingkan algoritma lain atau menggunakan representasi fitur berbasis embedding untuk meningkatkan pemahaman konteks bahasa.

REFERENCE

- [1] A. Kastrati, A. Imran, and A. Kurti, "Sentiment analysis of user-generated content using machine learning algorithms," *Appl. Sci.*, vol. 11, no. 2, pp. 1–22, 2021, doi: 10.3390/app11020814.
- [2] K. Sawicki and R. Wójcik, "Evaluation of Naïve Bayes variants for sentiment classification of short texts," *Procedia Comput. Sci.*, vol. 207, pp. 2392–2401, 2023, doi: 10.1016/j.procs.2022.12.302.
- [3] L. Ashbaugh and Y. Zhang, "A comparative study of sentiment analysis on customer reviews using machine learning and deep learning," *Computers*, vol. 13, no. 12, p. 340, 2024, doi: 10.3390/computers13120340.
- [4] A. Hanić, "Challenges of text classification in informal and user-generated content," *Inf. Process. Manag.*, vol. 61, no. 1, p. 103122, 2024, doi: 10.1016/j.ipm.2023.103122.
- [5] S. Ashbaugh and T. Zhang, "Performance of ensemble methods in user review sentiment classification," *Appl. Sci.*, vol. 14, no. 3, p. 1125, 2024, doi: 10.3390/app14031125.
- [6] K. Sawicki and R. Wójcik, "Evaluation of naïve Bayes variants for sentiment classification of short texts," *Procedia Comput. Sci.*, vol. 192, pp. 299–308, 2021, doi: 10.1016/j.procs.2021.08.031.