

Evaluasi Metode Naive Bayes dan K-Nearest Neighbors untuk Analisis Sentimen pada Review Aplikasi Duolingo

Joko Dwi Mulyanto ^{1*}, Dany Pratmanto ², Aprih Widayanto ³, Pijar Sukma Prayogo ⁴,
Andi Yoko Satrio ⁵

¹ Sistem Informasi, Universitas Bina Sarana Informatika
^{2,3,4,5} Teknologi Komputer, Universitas Bina Sarana Informatika

* Corresponding Author. E-mail: joko.jdm@bsi.ac.id

Abstract

Sentiment analysis is a valuable method for understanding user opinions on digital applications. This study evaluates the performance of the Naive Bayes and K-Nearest Neighbors (KNN) algorithms in classifying sentiments from user reviews of the Duolingo application obtained from the Google Play Store. The dataset consists of 2,000 reviews, comprising 1,000 negative (1-star) and 1,000 positive (5-star) reviews. Preprocessing was conducted in RapidMiner through several stages, including case transformation, tokenization, stopword removal, and stemming, with features represented using TF-IDF. The experimental results show that Naive Bayes achieved an accuracy of 79.96%, recall of 87.76%, precision of 77.21%, and an AUC of 96.40%. Meanwhile, KNN achieved an accuracy of 78.34%, recall of 75.32%, precision of 81.06%, and an AUC of 92.20%. These findings suggest that Naive Bayes outperforms KNN overall, particularly in sensitivity and class separation, while KNN produces more precise positive predictions. Therefore, the choice of algorithm should depend on analysis objectives, whether emphasizing broader sentiment detection or higher precision in positive sentiment classification.

Keywords : Sentiment Analysis, Naive Bayes, K-Nearest Neighbors, RapidMiner, Duolingo

Abstrak

Analisis sentimen merupakan pendekatan penting untuk memahami opini pengguna terhadap aplikasi digital. Penelitian ini bertujuan untuk mengevaluasi kinerja algoritma Naive Bayes dan K-Nearest Neighbors (KNN) dalam mengklasifikasikan sentimen pada ulasan pengguna aplikasi Duolingo yang diperoleh dari Google Play Store. Dataset yang digunakan terdiri dari 2.000 ulasan, mencakup 1.000 ulasan negatif (rating 1 bintang) dan 1.000 ulasan positif (rating 5 bintang). Proses pra-pemrosesan dilakukan menggunakan RapidMiner melalui beberapa tahapan, yaitu case transformation, tokenisasi, stopword removal, dan stemming, dengan representasi fitur menggunakan TF-IDF. Hasil eksperimen menunjukkan bahwa algoritma Naive Bayes mencapai akurasi sebesar 79,96%, recall 87,76%, presisi 77,21%, dan AUC 96,40%. Sementara itu, KNN memperoleh akurasi 78,34%, recall 75,32%, presisi 81,06%, dan AUC 92,20%. Temuan ini mengindikasikan bahwa Naive Bayes memiliki performa yang lebih baik secara keseluruhan, terutama dalam hal sensitivitas dan kemampuan memisahkan kelas, sedangkan KNN memberikan prediksi positif yang lebih tepat. Dengan demikian, pemilihan algoritma sebaiknya disesuaikan dengan tujuan analisis, apakah menekankan pada deteksi sentimen yang lebih luas atau pada ketepatan prediksi positif.

Kata kunci : Analisis Sentimen, Naive Bayes, K-Nearest Neighbors, RapidMiner, Duolingo

1. Pendahuluan

Kemajuan teknologi digital telah membawa perubahan signifikan dalam dunia pendidikan, khususnya dalam hal pembelajaran bahasa asing. Salah satu aplikasi pembelajaran bahasa yang populer dan banyak digunakan secara global adalah Duolingo. Aplikasi ini menawarkan metode pembelajaran yang interaktif, menyenangkan, dan berbasis gamifikasi, sehingga mampu menarik perhatian jutaan pengguna dari berbagai latar belakang.(Sasmitha & Thamrin, 2022) Popularitas Duolingo dapat dilihat dari banyaknya unduhan dan ulasan yang diberikan oleh pengguna melalui platform distribusi aplikasi seperti Google Play Store.(Hardiyanto dkk., 2023)

Ulasan atau review yang diberikan oleh pengguna di Play Store mencerminkan pengalaman dan persepsi mereka terhadap aplikasi.(Zaky & Luthfi, 2021) Informasi tersebut dapat menjadi sumber data yang sangat berharga untuk mengetahui sejauh mana kepuasan pengguna serta aspek-aspek apa saja yang perlu ditingkatkan oleh pengembang aplikasi. Salah satu pendekatan yang dapat digunakan untuk memahami opini pengguna secara sistematis adalah analisis sentimen.(Hamka & Ratna Sari, 2022) Analisis sentimen merupakan proses pengolahan teks yang bertujuan untuk mengidentifikasi dan mengelompokkan

opini ke dalam kategori tertentu, seperti positif, negatif, atau netral.

Dalam konteks penelitian ini, dilakukan analisis sentimen terhadap ulasan pengguna aplikasi Duolingo di Google Play Store. Penelitian ini menggunakan dua metode klasifikasi yang umum digunakan dalam bidang pengolahan bahasa alami (Natural Language Processing), yaitu Naive Bayes dan K-Nearest Neighbors (KNN).(Anwar, 2022) Metode Naive Bayes merupakan algoritma klasifikasi probabilistik yang sederhana namun cukup efektif, sedangkan K-Nearest Neighbors merupakan metode berbasis kedekatan atau kemiripan antar data.(Wulandari & Anwar US, 2021) Kedua metode ini dipilih karena karakteristiknya yang berbeda dan kerap digunakan dalam berbagai studi analisis teks.(Zaenal & Astutik, 2023)

Tujuan utama dari penelitian ini adalah untuk membandingkan performa kedua metode tersebut dalam mengklasifikasikan sentimen ulasan pengguna, serta mengevaluasi akurasi dan efisiensi masing-masing metode.(Khoirul Insan dkk., 2023) Selain itu, penelitian ini juga bertujuan untuk menggambarkan persepsi umum pengguna terhadap aplikasi Duolingo berdasarkan analisis data yang dilakukan.(Yuliska & Syaliman, 2020)

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam

pengembangan sistem klasifikasi sentimen yang lebih akurat serta memberikan masukan yang konstruktif bagi pengembang aplikasi dalam meningkatkan kualitas layanan dan pengalaman pengguna.

2. Materials and Methods

Penelitian ini menggunakan kerangka CRISP-DM (Cross-Industry Standard Process for Data Mining) sebagai metodologi untuk proses analisis data.(Wahyuningsih & Hendry, 2023) CRISP-DM dipilih karena memberikan alur sistematis dalam mengolah data hingga memperoleh model yang dapat dievaluasi. Tahapan CRISP-DM dalam penelitian ini dijelaskan sebagai berikut.

1. Business Understanding

Tahap awal penelitian ini bertujuan memahami kebutuhan dan tujuan analisis, yaitu mengevaluasi performa metode Naive Bayes dan K-Nearest Neighbors dalam klasifikasi sentimen ulasan pengguna aplikasi Duolingo di Play Store. Fokus utama adalah membandingkan akurasi, presisi, recall, dan f1-score kedua metode untuk menentukan algoritma yang lebih sesuai digunakan pada kasus analisis sentimen aplikasi pembelajaran bahasa.

2. Data Understanding

Data penelitian berupa ulasan pengguna aplikasi Duolingo yang diambil dari platform Google Play Store. Ulasan tersebut terdiri

dari teks review dan rating bintang sebagai acuan label sentimen (positif, negatif, netral).

Proses pemahaman data meliputi:

Eksplorasi distribusi jumlah review berdasarkan kategori sentimen.

Analisis karakteristik teks, seperti panjang rata-rata review, jumlah kata unik, serta keberadaan kata-kata dominan.

Identifikasi potensi masalah pada data, seperti duplikasi, noise, dan ketidakseimbangan kelas.

3. Data Preparation

Tahap ini dilakukan untuk menyiapkan data sehingga dapat digunakan dalam proses pemodelan. Beberapa langkah yang diterapkan adalah:

Cleaning: menghapus duplikasi, tanda baca, angka, serta karakter khusus yang tidak relevan.

Case folding: mengubah semua huruf menjadi huruf kecil.(Kosasih & Alberto, 2021a)

Tokenizing: memecah kalimat review menjadi kata-kata.(Kosasih & Alberto, 2021b)

Stopword removal: menghapus kata umum yang tidak memiliki makna sentimen.(Rinandyaswara dkk., 2022)

Stemming/Lemmatization: mengembalikan kata ke bentuk dasar.(Sinaga & Nainggolan, 2023)

Labeling sentimen: berdasarkan rating,

misalnya bintang 5 sebagai positif, bintang ≤ 1 sebagai negatif.(Darwis dkk., 2020)

Feature extraction: menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) atau Bag of Words untuk mengubah teks menjadi vektor numerik.

4. Modeling

Pada tahap ini, dua algoritma machine learning diterapkan:

Naive Bayes (NB): digunakan karena sederhana, efisien, dan umum dipakai untuk klasifikasi teks.

K-Nearest Neighbors (KNN): digunakan untuk membandingkan performa NB, dengan prinsip kedekatan antar data berdasarkan jarak (misalnya Euclidean).

Proses pemodelan dilakukan dengan membagi data menjadi training set dan testing set menggunakan metode train-test split atau k-fold cross validation.

5. Evaluation

Model yang telah dibangun dievaluasi menggunakan metrik performa, antara lain:

Accuracy untuk mengukur ketepatan prediksi secara keseluruhan.

Precision untuk mengukur proporsi prediksi positif yang benar.

Recall untuk mengukur keberhasilan model dalam menemukan data positif.

F1-score sebagai harmonisasi antara precision dan recall.

Perbandingan hasil evaluasi dari kedua

algoritma digunakan untuk menentukan metode yang lebih efektif dalam analisis sentimen review aplikasi Duolingo.

3. Hasil dan Pembahasan

3.1. Hasil Pengumpulan Data

Data penelitian diperoleh dari Google Play Store dengan memanfaatkan Python Google Review Scraper. Play Store dipilih karena merupakan platform utama bagi pengguna untuk memberikan ulasan yang dapat merepresentasikan pengalaman mereka dalam menggunakan aplikasi Duolingo. Proses pengumpulan data dimulai dengan mengidentifikasi aplikasi target, yaitu Duolingo: Learn Languages Free, kemudian scraper digunakan untuk mengekstrak komentar, rating bintang, nama pengguna, serta tanggal review.

Agar penelitian lebih terfokus, hanya ulasan dengan rating bintang 1 dan bintang 5 yang diambil. Ulasan dengan bintang 1 diperlakukan sebagai representasi sentimen negatif, sedangkan ulasan dengan bintang 5 dianggap mewakili sentimen positif. Dari proses seleksi ini diperoleh masing-masing seribu ulasan negatif dan seribu ulasan positif sehingga total dataset berjumlah dua ribu ulasan. Jumlah yang seimbang antara kedua kategori dipilih untuk menghindari masalah ketidakseimbangan kelas yang dapat memengaruhi performa algoritma klasifikasi.

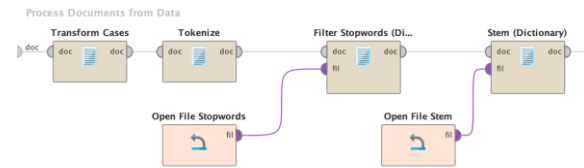
Karakteristik ulasan yang diperoleh

menunjukkan bahwa review positif banyak berisi ungkapan kepuasan pengguna, misalnya terkait kemudahan penggunaan, pengalaman belajar yang menyenangkan, hingga manfaat aplikasi dalam meningkatkan kemampuan bahasa. Sebaliknya, review negatif cenderung berisi keluhan mengenai banyaknya iklan, adanya bug atau error, keterbatasan fitur bagi pengguna gratis, dan kesulitan dalam mengikuti proses pembelajaran. Dari segi panjang teks, sebagian ulasan hanya terdiri dari satu hingga dua kata seperti “bagus” atau “buruk”, sementara yang lain berbentuk paragraf panjang dengan penjelasan detail.

Dengan dataset yang seimbang, representatif, dan mencerminkan opini nyata pengguna, penelitian ini memiliki dasar yang kuat untuk melakukan evaluasi terhadap kinerja metode Naive Bayes dan K-Nearest Neighbors dalam analisis sentimen.

3.2. Tahap Preprocessing Data

Sebelum dilakukan pemodelan, data ulasan yang telah dikumpulkan terlebih dahulu melalui tahap preprocessing menggunakan RapidMiner. Proses ini bertujuan untuk membersihkan dan menyiapkan data agar lebih terstruktur serta dapat diolah secara optimal oleh algoritma klasifikasi.



Gambar 1 Data Preparation

Pengolahan data dilakukan menggunakan RapidMiner, dengan tahapan sebagai berikut:

1. Transform Cases: seluruh teks ulasan diubah menjadi huruf kecil (lowercase) agar konsisten.
2. Tokenize: setiap kalimat dipecah menjadi kata-kata (token) untuk memudahkan analisis teks.
3. Stopword Removal: kata-kata umum yang tidak berkontribusi terhadap sentimen (misalnya dan, yang, untuk) dihapus.
4. Stemming: kata dikembalikan ke bentuk dasarnya, misalnya belajarnya → belajar.

Setelah preprocessing, data kemudian dikonversi menjadi bentuk numerik menggunakan metode TF-IDF (Term Frequency – Inverse Document Frequency) sebelum dimasukkan ke tahap pemodelan.

3.3. Hasil Pemodelan

Dua algoritma machine learning digunakan, yaitu Naive Bayes (NB) dan K-Nearest Neighbors (KNN). Dataset dibagi dengan skema train-test split menggunakan RapidMiner.

Hasil evaluasi model ditunjukkan pada Tabel

1.

Tabel 1 Hasil Evaluasi Model

Algoritma	Accuracy	Recall	Precision	AUC
Naive Bayes (NB)	74.74%	76.08%	74.59%	93.60%
K-Nearest Neighbors (KNN)	80.16%	79.90%	80.79%	93.00%

3.4. Diskusi Hasil

Berdasarkan hasil perbandingan, algoritma K-Nearest Neighbors (KNN) menunjukkan performa yang lebih baik secara keseluruhan dibandingkan Naive Bayes (NB). KNN memiliki akurasi tertinggi yaitu 80,16%, dengan nilai presisi 80,79% dan recall 79,90%. Hal ini menunjukkan bahwa KNN mampu mengklasifikasikan sentimen ulasan dengan lebih tepat dan konsisten. Sementara itu, Naive Bayes memperoleh akurasi 74,74% dengan presisi 74,59% dan recall 76,08%. Walaupun hasil ini lebih rendah dibandingkan KNN, Naive Bayes justru unggul pada metrik AUC dengan nilai 93,60%, sedikit lebih tinggi dibandingkan KNN yang memperoleh 93,00%. Temuan ini mengindikasikan bahwa Naive Bayes masih memiliki kemampuan yang baik dalam membedakan kelas positif dan negatif secara keseluruhan, meskipun tingkat ketepatan klasifikasinya lebih rendah. Dengan demikian, dapat disimpulkan bahwa KNN lebih direkomendasikan untuk digunakan pada analisis sentimen ulasan aplikasi Duolingo karena memberikan hasil klasifikasi yang

lebih akurat, sementara Naive Bayes tetap dapat menjadi alternatif ketika fokus penelitian adalah pada kemampuan pemisahan antar kelas.

3.8. Kesimpulan

Berdasarkan hasil penelitian analisis sentimen terhadap ulasan aplikasi Duolingo di Google Play Store dengan menggunakan algoritma Naive Bayes dan K-Nearest Neighbors (KNN), dapat disimpulkan bahwa kedua metode memiliki keunggulan masing-masing. Algoritma KNN menghasilkan akurasi yang lebih tinggi (80,16%) dibandingkan dengan Naive Bayes (74,74%), serta menunjukkan nilai presisi dan recall yang lebih seimbang. Hal ini menandakan bahwa KNN lebih efektif dalam memberikan klasifikasi sentimen yang tepat pada data ulasan pengguna.

Namun demikian, Naive Bayes tetap menunjukkan keunggulan pada metrik AUC dengan nilai 93,60%, sedikit lebih tinggi dibandingkan KNN yang memperoleh 93,00%. Temuan ini mengindikasikan bahwa meskipun akurasinya lebih rendah, Naive Bayes memiliki kemampuan yang baik dalam memisahkan kelas positif dan negatif secara konsisten.

Secara keseluruhan, KNN lebih direkomendasikan untuk digunakan dalam analisis sentimen ulasan aplikasi Duolingo

karena menghasilkan klasifikasi yang lebih akurat dan presisi. Akan tetapi, Naive Bayes tetap dapat menjadi pilihan alternatif terutama ketika fokus penelitian adalah pada kemampuan generalisasi dan pemisahan kelas. Hasil penelitian ini diharapkan dapat menjadi dasar bagi pengembangan sistem analisis sentimen yang lebih baik, serta memberikan masukan yang bermanfaat bagi pengembang aplikasi dalam memahami persepsi dan kebutuhan pengguna.

3.9. Acknowledgment

Penelitian ini didanai oleh Basis Informasi Penelitian & Pengabdian Masyarakat (BIPEMAS) LPPM Universitas Bina Sarana Informatika (UBSI). Penulis menyampaikan terima kasih atas dukungan yang telah diberikan sehingga penelitian ini dapat terlaksana dengan baik.

3.10. References

- Anwar, K. (2022). Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan Naive Bayes. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 2(4).
<https://doi.org/10.30865/klik.v2i4.315>
- Darwis, D., Pratiwi, E. S., & Pasaribu, A. F. O. (2020). PENERAPAN ALGORITMA SVM UNTUK ANALISIS SENTIMEN PADA DATA TWITTER KOMISI PEMBERANTASAN KORUPSI REPUBLIK INDONESIA. *Eductic - Scientific Journal of Informatics*, 7(1).
<https://doi.org/10.21107/edutic.v7i1.8779>
- Hamka, M., & Ratna Sari, D. (2022). ANALISIS SENTIMEN DAN INFORMATION EXTRACTION PEMBELAJARAN DARING MENGGUNAKAN PENDEKATAN LEXICON. *Djtechno: Jurnal Teknologi Informasi*, 3(1).
<https://doi.org/10.46576/djtechno.v3i1.2194>
- Hardiyanto, M. R. P., Pahlevi, G., & Nugroho, M. F. (2023). Pengaruh Fitur-Fitur Aplikasi Duolingo Terhadap Popularitasnya. *Jurnal Sains, Nalar, dan Aplikasi Teknologi Informasi*, 3(1).
<https://doi.org/10.20885/snati.v3i1.28>
- Khoirul Insan, M. K., Hayati, U., & Nurdiawan, O. (2023). ANALISIS SENTIMEN APLIKASI BRIMO PADA ULASAN PENGGUNA DI GOOGLE PLAY MENGGUNAKAN ALGORITMA NAIVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1).
<https://doi.org/10.36040/jati.v7i1.6373>
- Kosasih, R., & Alberto, A. (2021a). Analisis Sentimen Produk Permainan Menggunakan Metode TF-IDF Dan Algoritma K-Nearest Neighbor. *InfoTekJar : Jurnal Nasional Informatika dan Teknologi Jaringan*, 6(1).
- Kosasih, R., & Alberto, A. (2021b). Sentiment analysis of game product on shopee using the TF-IDF method and naive bayes classifier. *ILKOM Jurnal Ilmiah*, 13(2).

- <https://doi.org/10.33096/ilkom.v13i2.721.101-109>
- Rinandyaswara, R., Sari, Y. A., & Furqon, M. T. (2022). Pembentukan Daftar Stopword Menggunakan Term Based Random Sampling Pada Analisis Sentimen Dengan Metode Naïve Bayes (Studi Kasus: Kuliah Daring Di Masa Pandemi). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 9(4).
<https://doi.org/10.25126/jtiik.2022934707>
- Sasmitha, I., & Thamrin, H. (2022). Meningkatkan Minat Belajar Bahasa Inggris Dengan Aplikasi Duolingo Sebagai Media Interaktif di Rumah Pintar YAFSI. *Literasi Jurnal Pengabdian Masyarakat dan Inovasi*, 2(2).
<https://doi.org/10.58466/literasi.v2i2.1414>
- Sinaga, A., & Nainggolan, S. P. (2023). ANALISIS PERBANDINGAN AKURASI DAN WAKTU PROSES ALGORITMA STEMMING ARIFIN-SETIONO DAN NAZIEF-ADRIANI PADA DOKUMEN TEKS BAHASA INDONESIA. *Sebatik*, 27(1).
<https://doi.org/10.46984/sebatik.v27i1.2072>
- Wahyuningsih, N., & Hendry, H. (2023). PERBANDINGAN METODE KLASIFIKASI DALAM ANALISIS SENTIMEN MASYARAKAT TERHADAP IDENTITAS KEPENDUDUKAN DIGITAL (IKD). *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 8(4).
<https://doi.org/10.29100/jupi.v8i4.4155>
- Wulandari, A., & Anwar US, K. (2021). Analisis Manajemen Pemasaran Pada Aplikasi Shopee Dan Tokopedia. *Transekonomika: Akuntansi, Bisnis dan Keuangan*, 1(3).
<https://doi.org/10.55047/transekonomika.v1i3.49>
- Yuliska, Y., & Syaliman, K. U. (2020). Peningkatan Akurasi K-Nearest Neighbor Pada Data Index Standar Pencemaran Udara Kota Pekanbaru. *IT Journal Research and Development*, 5(1).
[https://doi.org/10.25299/itjrd.2020.vol5\(1\).4680](https://doi.org/10.25299/itjrd.2020.vol5(1).4680)
- Zaenal, Z., & Astutik, I. R. I. (2023). Sentiment Analysis of OYO App Reviews Using the Support Vector Machine Algorithm. *Procedia of Engineering and Life Science*, 3.
<https://doi.org/10.21070/pels.v3i0.1338>
- Zaky, A., & Luthfi, A. (2021). Analisa Pengukuran Usability Sistem Menggunakan Metode USE Questionary pada Aplikasi Duolingo. *Jurnal Teknik Informatika Kaputama (JTIK)*, 5(2).