

PENERAPAN DAN PERBANDINGAN ALGORITMA NAÏVE BAYES DAN K-NEAREST NEIGHBOR DALAM ANALISIS SENTIMEN TERHADAP KEPUASAN PENGGUNA APLIKASI FLO

Basmallah Ramadhani Aisyah Putri¹⁾, Fitrianti²⁾

^{1), 2)} Teknik Informatika, Universitas Indraprasta PGRI

Jl. Raya Tengah No.80, RT.6/RW.1, Gedong, Ps. Rebo, Kota Jakarta Timur, DKI Jakarta 13760

Email : basmallahramadhani08@gmail.com¹⁾, fitriantif11@gmail.com²⁾

ABSTRACT

Digital technology is developing rapidly and has a wide and significant impact on the health sector, including through the presence of health monitoring applications such as the Flo application. This application is designed to help women track their menstrual cycles, fertile periods, and pregnancy. As an application that is personal and used routinely, user satisfaction is an important factor that determines the quality and sustainability of services. Sentiment analysis is needed to explore user views and preferences for this application. This study aims to analyze 18215 user reviews of the Flo application from the Google Play Store to classify sentiment, using web scraping techniques as a data retrieval method. Naïve Bayes and K-Nearest Neighbor are used as classification algorithms in data analysis. Data are analyzed through several stages, namely preprocessing, sentiment classification, model evaluation, and interpretation of results. The results showed that 93.1% of reviews were positive and 6.9% of reviews were negative. In terms of performance, the Naïve Bayes algorithm showed the best results with an accuracy value of 99%, precision 100%, recall 98%, and f-measure 99%, and without False Positive errors. Meanwhile, the K-Nearest Neighbor algorithm obtained an accuracy of 95%, precision of 97%, recall of 90%, and f-measure of 93%. The results of the study showed that the Naïve Bayes algorithm was more effective in analyzing the sentiment of Flo application user reviews.

Keywords : Sentiment Analysis, Flo app, Naïve Bayes, K-Nearest Neighbor, User Reviews.

ABSTRAK

Teknologi digital berkembang pesat dan berdampak luas dan signifikan dalam sektor kesehatan, termasuk melalui kehadiran aplikasi pemantau kesehatan seperti aplikasi Flo. Aplikasi ini dirancang untuk membantu perempuan dalam melacak siklus menstruasi, masa subur, dan juga kehamilan. Sebagai aplikasi yang bersifat personal dan digunakan secara rutin, kepuasan pengguna menjadi faktor penting yang menentukan kualitas dan keberlanjutan layanan. Analisis sentimen diperlukan untuk menggali pandangan dan preferensi pengguna terhadap aplikasi ini. Penelitian ini bertujuan untuk menganalisis 18215 ulasan pengguna aplikasi Flo dari Google Play Store untuk mengklasifikasikan sentimen, menggunakan teknik web scraping sebagai metode pengambilan data. Naïve Bayes dan K-Nearest Neighbor digunakan sebagai algoritma klasifikasi dalam analisis data. Data dianalisis melalui beberapa tahapan, yaitu preprocessing, klasifikasi sentimen, evaluasi model, dan interpretasi hasil. Hasil penelitian menunjukkan bahwa 93,1% ulasan bersifat positif dan 6,9% ulasan bersifat negatif. Dari segi performa, algoritma Naïve Bayes menunjukkan hasil terbaik dengan nilai accuracy 99%, precision 100%, recall 98%, dan f - measure 99%, serta tanpa kesalahan False Positive. Sementara itu, algoritma K-Nearest Neighbor memperoleh accuracy sebesar 95%, precision 97%, recall 90%, dan f - measure 93%. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes lebih efektif dalam menganalisis sentimen ulasan pengguna aplikasi Flo.

Kata Kunci : Analisis Sentimen, Aplikasi Flo, Naïve Bayes, K-Nearest Neighbor, Ulasan Pengguna.

1. Pendahuluan

Kemajuan teknologi di Indonesia membawa manfaat dan mempermudah berbagai aspek kehidupan, terutama dalam sektor kesehatan (Asrin et al., 2024). Salah satu penerapannya adalah kehadiran aplikasi yang berfungsi sebagai asisten pribadi untuk memantau kondisi kesehatan secara mandiri. Aplikasi Flo merupakan salah satu contoh aplikasi populer yang dirancang khusus untuk membantu perempuan dalam melacak siklus menstruasi, masa subur, serta memantau kehamilan (Al-Frida et al., 2025).

Sebagai aplikasi yang personal dan digunakan secara rutin, tingkat kepuasan pengguna menjadi faktor penting yang menentukan keberhasilan dan keberlanjutan aplikasi ini. Pengguna sering kali menyampaikan pengalaman dan pendapat mereka melalui ulasan di aplikasi. Ulasan pengguna berisi informasi berharga terkait kelebihan, kekurangan, serta kebutuhan pengguna (Fauzan Abdillah & Pramesti, 2024). Untuk mendapatkan insight dari data ulasan, analisis sentimen menjadi salah satu metode yang efektif digunakan (Mursyid & Dwi Indriyanti, 2024).

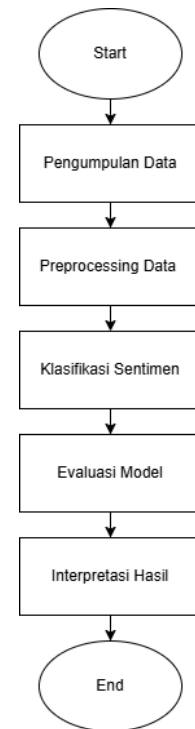
Dalam text mining dan *Natural Language Processing* (NLP), analisis sentimen digunakan untuk mengidentifikasi opini atau perasaan pengguna terhadap suatu produk atau layanan, dengan mengklasifikasikannya ke dalam kategori *positive*, *negative*, atau *neutral* (Muliani et al., 2023). Dalam konteks aplikasi Flo, analisis sentimen terhadap ulasan pengguna dapat memberikan informasi yang berguna bagi pengembang untuk meningkatkan kualitas layanan serta memahami kebutuhan pengguna secara lebih mendalam (Ramdan Adi Surya & Hayati, 2024).

Untuk melakukan analisis sentimen secara efektif, pemilihan algoritma klasifikasi yang tepat sangat diperlukan. *Naïve Bayes* dan *K-Nearest Neighbor* (K-NN) adalah dua algoritma yang sering dipakai untuk mengelompokkan data dengan cara yang sederhana dan efektif (Zahri et al., 2025). *Naïve Bayes* dikenal sebagai algoritma yang sederhana namun efisien dalam menangani data teks berskala besar (Syafa'aturrohmah et al., 2024), sementara *K-Nearest Neighbor* bekerja berdasarkan kedekatan antar data dan dikenal fleksibel terhadap distribusi data (Salmon et al., 2024). Setiap algoritma memiliki karakteristik dan kinerja yang berbeda saat digunakan dalam analisis sentimen.

Penelitian ini bertujuan menerapkan dan membandingkan kinerja algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dalam analisis sentimen ulasan pengguna aplikasi Flo. Hasil dari penelitian ini diharapkan dapat memberikan informasi yang relevan mengenai efektivitas kedua algoritma tersebut serta mendukung pengambilan keputusan dalam pengembangan aplikasi.

2. Metodologi Penelitian

Gambar 1 merupakan tahapan penelitian. Tahapan penelitian pada penelitian ini dimulai pada tahap pengumpulan data, *preprocessing* data, klasifikasi sentimen, evaluasi model, dan interpretasi hasil.



Gambar 1. Tahapan Penelitian

2.1 Pengumpulan Data

Pada tahap ini, dilakukan pengumpulan data menggunakan *web scraping*. *Web scraping* adalah suatu teknik untuk mendapatkan informasi dari situs tertentu yang bertujuan untuk dapat dilakukan pengambilan data baik secara manual maupun otomatis (Djiwadikusumah et al., 2021). *Web scraping* pada penelitian ini didapat dari ulasan pengguna pada aplikasi Flo di *Google Play Store*.

2.2 Preprocessing Data

Tahapan *preprocessing* data merupakan tahapan dalam pemrosesan data, baik itu untuk data teks, gambar, suara, maupun numerik, yang bertujuan untuk mempersiapkan data agar lebih bersih dan sesuai untuk dianalisis (Agung et al., 2023). Tahapan *preprocessing* data terdiri dari Integrasi Data, Pelabelan Data, Transformasi Data, dan Pembersihan Data (Gori et al., 2024).

2.3 Klasifikasi Sentimen

Pada tahap ini, dilakukan penerapan algoritma *Naïve Bayes* dan *K-Nearest Neighbor* untuk dilakukan analisis sentimen pada data yang telah diperoleh sebelumnya dan telah dilakukan tahapan *preprocessing* data. Hasil analisis tersebut berupa perbandingan hasil dari algoritma *Naïve Bayes* dan juga *K-Nearest Neighbor*. Masing-masing algoritma memiliki kelebihannya masing-masing. *Naïve Bayes* memiliki kelebihan, yaitu cepat dan efisien, mudah diimplementasikan, dan performa yang baik (Samosir & Wati, 2024). Dan *K-Nearest Neighbor* memiliki kelebihan, yaitu dapat digunakan untuk masalah klasifikasi dan regresi, adaptif

terhadap data yang baru, dan mudah dipahami dan diimplementasikan (Daulay, 2024).

2.4 Evaluasi Model

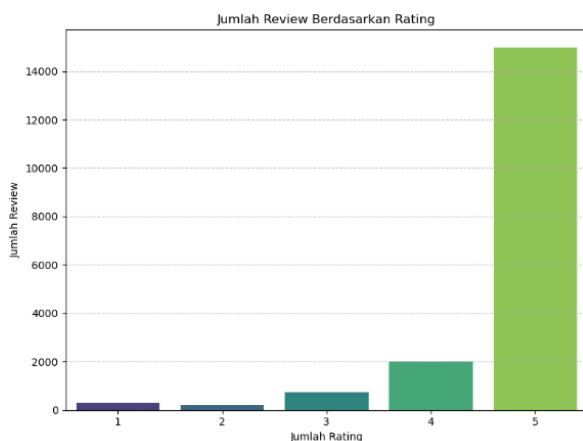
Tahapan evaluasi model adalah tahapan dalam menilai kinerja suatu model setelah dilakukannya proses suatu data testing selesai (Nugraha Poernamawan et al., 2025). Tujuan utamanya adalah untuk mengetahui seberapa baik model tersebut bekerja dalam memprediksi data (Mega Antika & Yusuf, 2025).

2.5 Interpretasi Hasil

Pada tahap ini, dilakukan interpretasi hasil dari evaluasi model dengan menggunakan *Confusion Matrix*. *Confusion Matrix* merupakan tabel yang digunakan untuk mengevaluasi performa model klasifikasi, dengan menunjukkan jumlah prediksi yang benar dan salah yang dilakukan oleh model dibandingkan dengan label *actual* (Aizah & Fatah, 2024).

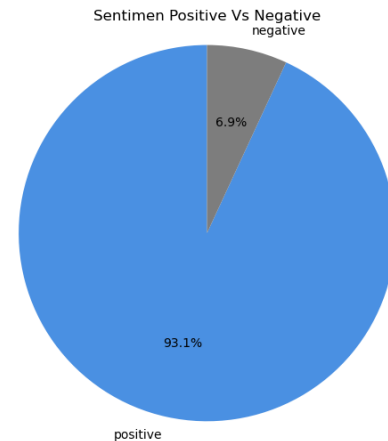
3. Hasil dan Pembahasan

Penelitian ini memperoleh data berjumlah 18215 ulasan pengguna, terdiri dari bintang 1 berjumlah 315 *review*, bintang 2 berjumlah 220 *review*, bintang 3 berjumlah 722 *review*, bintang 4 berjumlah 1993 *review*, dan bintang 5 berjumlah 14966 *review* yang ditunjukkan pada Gambar 2.



Gambar 2. Jumlah Review Berdasarkan Bintang

Ulasan pengguna aplikasi Flo menunjukkan hasil sebanyak 93.1% ulasan *positive* dan 6.9% ulasan *negative* yang ditunjukkan pada Gambar 3 yang berupa *pie chart*.



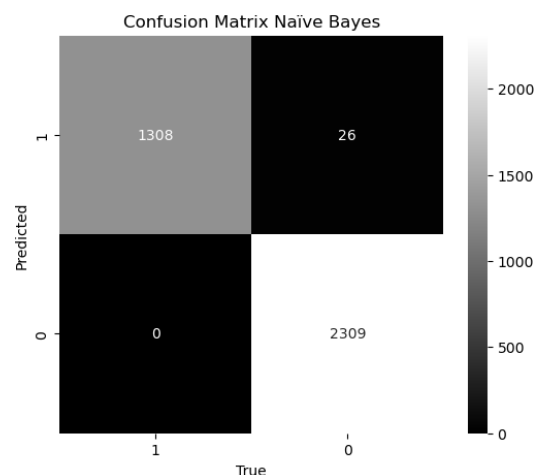
Gambar 3. Distribusi Kelas Sentimen

Hasil evaluasi kinerja algoritma *Naïve Bayes* dan *K-Nearest Neighbor* ditunjukkan pada Tabel 1. Hasil tersebut menunjukkan bahwa hasil algoritma *Naïve Bayes* lebih unggul dibandingkan algoritma *K-Nearest Neighbor*. Algoritma *Naïve Bayes* mendapatkan hasil *precision* yang sempurna yaitu 100%.

Tabel 1. Evaluasi Hasil Kinerja Algoritma

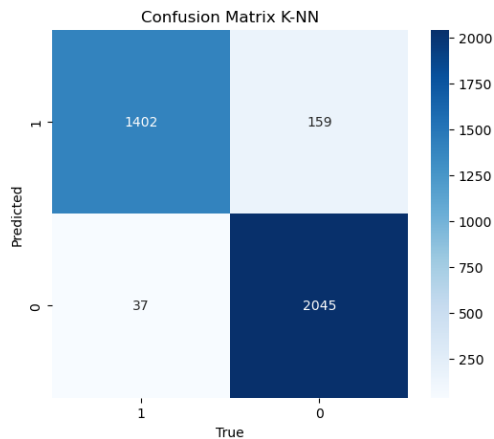
	Accuracy	Precision	Recall	f-measure
Naïve Bayes	0.99	1.00	0.98	0.99
K-Nearest Neighbor	0.95	0.97	0.90	0.93

Hasil *Confusion Matrix Naïve Bayes* ditunjukkan pada Gambar 4. Hasil tersebut menunjukkan bahwa nilai *True Positive* mendapatkan hasil 1308, *False Negative* mendapatkan hasil 26, *False Positive* mendapatkan hasil 0, dan *True Negative* sebanyak 2309.



Gambar 4. Confusion Matrix Naïve Bayes

Hasil *Confusion Matrix K-Nearest Neighbor* menunjukkan bahwa nilai *True Positive* mendapatkan hasil 1402, *False Negative* mendapatkan hasil 159, *False Positive* mendapatkan hasil 37, dan *True Negative* mendapatkan hasil 2045 yang ditunjukkan pada Gambar 5.



Gambar 5. Confusion Matrix K-Nearest Neighbor

4. Kesimpulan

Penelitian ini menunjukkan bahwa mayoritas pengguna aplikasi Flo memberikan ulasan *positive*, dengan persentase sebesar 93,1% dari total 18.215 ulasan. Dalam klasifikasi sentimen, algoritma *Naïve Bayes* terbukti lebih efektif dibandingkan *K-Nearest Neighbor*, dengan nilai *accuracy* sebesar 99%, *precision* 100%, *recall* 98%, dan *f - measure* 99%. Hasil *Confusion Matrix* juga menunjukkan performa klasifikasi yang sangat baik pada algoritma *Naïve Bayes*, tanpa kesalahan *False Positive*. Dengan demikian, algoritma *Naïve Bayes* merupakan algoritma yang paling efektif dalam mengklasifikasikan sentimen ulasan pengguna pada studi kasus ini.

Daftar Pustaka

- Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *JINACS: Journal of Informatics and Computer Science*, 05(01).
- Aizah, N., & Fatah, Z. (2024). Klasifikasi Kualitas Kopi Robusta dengan Algoritma K-Nearest Neighbor di PT. Indokom Citra Persada Situbondo. *Jurnal Kecerdasan Buatan, Komputasi Dan Teknologi Informasi*, 5(2), 109–118. <https://ejournal.unuja.ac.id/index.php/core>
- Al-Frida, M., Harjanti, A. I., & Oktiningrum, M. (2025). Pengaruh Pemberian Informasi Masa Subur dan “Flo Health” Terhadap Tingkat Pengetahuan Calon Pengantin di Semarang. *Termometer: Jurnal Ilmiah Ilmu Kesehatan Dan Kedokteran*, 3(2), 168–175.
- Asrin, F., Anra, H., Irwansyah, M. A., S, A. S., & Pratama, E. E. (2024). Pemahaman Dampak Positif dan Negatif Perkembangan Komputer Di Bidang Kesehatan. *Jurnal Abdimas Mandiri*, 8(2), 159–168. <https://doi.org/10.36982/jam.v8i2.4310>
- Daulay, R. S. (2024). Analisis Kritis dan Pengembangan Algoritma K-Nearest Neighbor (KNN): Sebuah Tinjauan Literatur. *Jurnal Pendidikan Sains Dan Komputer*, 4(02), 131–141. <https://doi.org/10.47709/jpsk.v4i02.5055>
- Djiwadikusumah, F., Hayindra Irawan, G., & Haekal Al-Fadilah, R. (2021). Web Scraping Situs E-Commerce Menggunakan Teknik Parsing DOM. *Saintek: Jurnal Siliwangi Seri Sains Dan Teknologi*, 7(2), 52–57.
- Fauzan Abdillah, R., & Pramesti, A. N. (2024). Dampak Rating dan Ulasan Konsumen terhadap Keputusan Pembelian di E-Commerce. *Prosiding Seminar Nasional Amikom Surakarta*, 2, 1480–1494.
- Gori, T., Sunyoto, A., & Al Fatta, H. (2024). Preprocessing Data dan Klasifikasi untuk Prediksi Kinerja Akademik Siswa. *JTIK: Jurnal Teknologi Informasi Dan Ilmu Komputer*, 11(1), 215–224. <https://doi.org/10.25126/jtiik.20241118074>
- Mega Antika, B., & Yusuf, R. (2025). Analisis Proyek Layanan Pengadaan Secara Elektronik (LPSE) Provinsi Lampung. *JINTEKS: Jurnal Informatika Teknologi Dan Sains*, 7(1), 213–222.
- Muliani, R., Solehudin, A., & Jamaludin, A. (2023). Analisis Sentimen Terhadap Provider Xyz di Twitter Menggunakan Algoritma Naïve Bayes Classifier. *JATI: Jurnal Mahasiswa Teknik Informatika*, 7(4).
- Mursyid, R., & Dwi Indriyanti, A. (2024). Perbandingan Akurasi Metode Analisis Sentimen Untuk Evaluasi Opini Pengguna Pada Platform Media Sosial (Studi Kasus: Twitter). *JINACS: Journal of Informatics and Computer Science*, 06(02), 371–383.
- Nugraha Poernamawan, A., Azhima Yoga Siswa, T., & Rudiman. (2025). Penerapan Metode GA-CBU Pada Algoritma Logistic Regression Untuk Mengatasi Class Imbalance Data Beasiswa KIP-Kuliah. *BITS: Building of Informatics, Technology and Science*, 6(4), 2322–2335. <https://doi.org/10.47065/bits.v6i4.6747>
- Ramdan Adi Surya, M., & Hayati, U. (2024). Analisis Sentimen Ulasan Pengguna OVO Menggunakan Algoritma Naïve Bayes pada Google Play Store. *JATI: Jurnal Mahasiswa Teknik Informatika*, 8(3), 2780–2786.
- Salmon, S., Azahari, A., & Ekawati, H. (2024). Perbandingan Kinerja Algoritma K-Nearest Neighbor dan Algoritma Random Forest Untuk Klasifikasi Data Mining Pada Penyakit Gagal Ginjal. *BITS: Building of Informatics, Technology and Science (BITS)*, 6(3), 1943–1953. <https://doi.org/10.47065/bits.v6i3.6476>
- Samosir, M. S., & Wati, L. (2024). Penerapan Naive Bayes Untuk Memprediksi Kelulusan Mahasiswa Rekayasa Perangkat Lunak Politeknik Negeri Bengkalis. *Remik: Riset Dan E-Jurnal Manajemen Informatika Komputer*, 8(3), 838–848. <https://doi.org/10.33395/remik.v8i3.13964>
- Syafa'aturrohman, A., Nurdian, O., Muhammad Basysyar, F., & Sulaeman, M. (2024). Naïve Bayes Meningkatkan Model Analisis Sentimen Pada Ulasan Aplikasi DANA Di Playstore Indonesia. *Information Management for Educators and Professionals*, 9(2), 171–180.
- Zahri, F., Insani, F., & Oktavia, L. (2025). Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor

(K-NN) Untuk Klasifikasi Penyakit Gagal Jantung.
JOSH: Journal of Information System Research,
6(2), 997–1005.
<https://doi.org/10.47065/josh.v6i2.6480>