

ANALISIS SENTIMEN KEPUASAN PENGGUNA APLIKASI FLO HEALTH PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES CLASSIFIER

Azzahra Shabbani Putri¹, Umi Hayati², Riri Narasati³, Ade Rizki Rinaldi⁴.

Program Studi Teknik Informatika¹²³
Program Studi Rekayasa Perangkat Lunak⁴

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

(*) Corresponding Author : azzahrashabbaniputri@gmail.com
Published : 30 Juni 2026

Abstract—The development of mobile health (mHealth) applications makes it easier for users to monitor their health independently, one of which is through the Flo Health application used to track menstrual cycles and women's reproductive health. User reviews on the Google Play Store are an important source of information for determining the level of satisfaction and user experience in using the application. However, the large number of reviews makes manual analysis difficult, requiring an automated machine learning-based approach to identify user sentiment more efficiently. This study aims to analyze the sentiment of Flo Health app user reviews and evaluate the performance of the Multinomial Naïve Bayes algorithm in classifying sentiment into positive, negative, and neutral categories. The research data was obtained through a web scraping process on Google Play Store, which generated 9,442 user reviews. The preprocessing stage was carried out through several processes, namely cleaning, case folding, normalization, tokenization, stopword removal, and stemming to improve the quality of the text data. Next, word weighting was performed using the Term Frequency-Inverse Document Frequency (TF-IDF) method to convert the text data into numerical representations. The dataset was then divided into 80% training data and 20% test data before applying the Multinomial Naïve Bayes algorithm for the classification process. The model was evaluated using accuracy, precision, recall, and F1-score metrics. The results showed that most user reviews had positive sentiments related to ease of use and menstrual cycle tracking features, while negative sentiments were generally related to technical issues with the application and data privacy issues. This study shows that the Multinomial Naïve Bayes algorithm is effective for sentiment analysis in digital health application reviews.

Keywords: sentiment analysis, Multinomial Naïve Bayes, TF-IDF, mobile health, user reviews.

Abstrak—Perkembangan aplikasi *mobile health* (mHealth) memberikan kemudahan bagi pengguna dalam memantau kondisi kesehatan secara mandiri, salah satunya melalui aplikasi Flo Health yang digunakan untuk melacak siklus menstruasi dan kesehatan reproduksi wanita. Ulasan pengguna pada *Google Play Store* menjadi sumber informasi penting untuk mengetahui tingkat kepuasan serta pengalaman pengguna dalam menggunakan aplikasi tersebut. Namun, jumlah ulasan yang sangat banyak menyulitkan proses analisis secara manual, sehingga diperlukan pendekatan otomatis berbasis machine learning untuk mengidentifikasi sentimen pengguna secara lebih efisien. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna aplikasi Flo Health serta mengevaluasi kinerja algoritma *Multinomial Naïve Bayes* dalam mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral. Data penelitian diperoleh melalui proses *web scraping* pada *Google Play Store* yang menghasilkan 9.442 ulasan pengguna. Tahap prapemrosesan dilakukan melalui beberapa proses, yaitu *cleaning*, *case folding*, *normalisasi*, *tokenisasi*, *stopword removal*, dan *stemming* untuk meningkatkan kualitas data teks. Selanjutnya dilakukan pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) untuk mengubah data teks menjadi representasi numerik. Dataset kemudian dibagi menjadi data latih sebesar 80% dan data uji sebesar 20% sebelum diterapkan algoritma *Multinomial Naïve Bayes* untuk proses klasifikasi. Evaluasi model dilakukan menggunakan *metrik akurasi*, *presisi*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa sebagian besar ulasan pengguna memiliki sentimen positif yang berkaitan dengan kemudahan penggunaan dan fitur pelacakan siklus menstruasi, sedangkan sentimen negatif umumnya berkaitan dengan kendala teknis aplikasi dan isu privasi data. Penelitian ini menunjukkan bahwa

algoritma *Multinomial Naïve Bayes* efektif digunakan untuk analisis sentimen pada ulasan aplikasi kesehatan digital.

Kata Kunci: analisis sentimen, Multinomial Naïve Bayes, TF-IDF, mobile health, ulasan pengguna.

INTRODUCTION

Perkembangan pesat di bidang Informatika dalam dua dekade terakhir telah merevolusi hampir setiap aspek kehidupan manusia. Transformasi digital yang ditopang oleh kemajuan komputasi, kecerdasan buatan (*AI*), dan pembelajaran mesin (*machine learning*) telah secara nyata mengubah cara manusia bekerja, belajar, dan berinteraksi di berbagai sektor. Dalam bidang bisnis, teknologi informasi telah menghadirkan otomatisasi proses, analitik prediktif, serta pengambilan keputusan berbasis data yang meningkatkan efisiensi operasional dan mempercepat inovasi. Di bidang pendidikan, teknologi digital telah mendorong lahirnya sistem pembelajaran daring interaktif dan personalisasi kurikulum berbasis kecerdasan buatan. Sementara itu, di sektor kesehatan, kemunculan aplikasi *mobile health* (*mHealth*) telah memperluas akses layanan medis, meningkatkan kesadaran masyarakat terhadap kesehatan, serta memperkuat kemampuan pemantauan mandiri pasien melalui perangkat seluler [1], [2], [3], [4], [5]. Inovasi-inovasi tersebut menegaskan bahwa Informatika menjadi fondasi utama transformasi sosial-ekonomi global, sekaligus membuka peluang baru bagi pengembangan riset analisis data dan pemodelan kecerdasan buatan dalam konteks aplikasi kesehatan digital.

Meskipun kemajuan Informatika telah menghasilkan berbagai inovasi signifikan, muncul pula beragam tantangan baru yang menuntut solusi ilmiah dan teknologis yang tepat. Salah satu isu utama adalah kualitas dan pengalaman pengguna terhadap aplikasi digital, khususnya pada aplikasi *mobile health* (*mHealth*) yang kini banyak digunakan untuk pemantauan kesehatan dan kesejahteraan individu. Studi terkini menunjukkan bahwa banyak aplikasi kesehatan masih menghadapi masalah seperti ketidakstabilan performa, ketidaksesuaian fitur dengan kebutuhan pengguna, serta tingkat kepuasan yang bervariasi antar platform [6], [7]. Selain itu, analisis ulasan pengguna yang tersebar luas di platform seperti *Google Play Store* belum dimanfaatkan secara optimal untuk menilai persepsi dan kepuasan pengguna terhadap aplikasi kesehatan. Hal ini menimbulkan kesenjangan riset antara potensi data yang

tersedia dan pemanfaatannya dalam pengembangan sistem berbasis pembelajaran mesin (*machine learning*) [8], [9]. Di sisi lain, meskipun metode analisis sentimen telah banyak diterapkan dalam bidang *e-commerce* dan media sosial, penerapannya pada ulasan aplikasi kesehatan wanita seperti *Flo Health* masih terbatas dan belum memiliki standar metodologis yang baku [10], [11]. Keterbatasan ini menimbulkan pertanyaan mengenai efektivitas algoritma klasik seperti *Naïve Bayes* dalam mengklasifikasi emosi dan kepuasan pengguna dibandingkan dengan model modern seperti *Support Vector Machine* (*SVM*) atau *transformer-based models* (misalnya *BERT*) [12], [13]. Selain itu, ketidakseimbangan data ulasan (antara positif, negatif, dan netral) serta keragaman bahasa dan konteks ekspresi pengguna sering kali mengurangi akurasi model analisis sentimen, terutama pada dataset yang besar dan tidak terstruktur [14]. Dengan demikian, terdapat kebutuhan mendesak untuk mengembangkan pendekatan analitik yang efisien, transparan, dan dapat diinterpretasikan guna mengevaluasi sentimen dan kepuasan pengguna aplikasi *mHealth* secara komprehensif. Permasalahan ini bukan hanya relevan dari sisi teknis, tetapi juga memiliki implikasi praktis bagi pengembang aplikasi dan peneliti kesehatan digital, karena hasil analisis sentimen dapat memberikan wawasan langsung terhadap kualitas fitur, penerimaan pengguna, serta potensi peningkatan layanan berbasis *AI* [15]. Oleh karena itu, riset dalam bidang ini bernilai strategis untuk menjembatani kesenjangan antara sains data dan desain pengalaman pengguna (*UX*) dalam konteks aplikasi kesehatan modern.

Penelitian mengenai analisis sentimen dan ulasan pengguna aplikasi kesehatan digital (*mHealth*) telah berkembang pesat dalam beberapa tahun terakhir, seiring meningkatnya ketergantungan masyarakat terhadap layanan kesehatan berbasis seluler. Salah satu studi oleh Aydin & Silahtaroglu (2021) melakukan analisis konten pasar aplikasi *mHealth* berskala besar dengan metode pembelajaran mesin untuk memahami pola kepuasan dan keluhan pengguna. Hasilnya menunjukkan bahwa pendekatan berbasis leksikon dan klasifikasi otomatis mampu mengidentifikasi tren kepuasan pengguna secara agregat, namun belum menggali dimensi emosional dan konteks spesifik dari ulasan individu [16]. Penelitian lain oleh Biswas et al. (2021) memperkenalkan *ACCU³RATE*, sebuah skala

evaluasi aplikasi mHealth berbasis ulasan pengguna, yang menggabungkan analisis kuantitatif dan kualitatif untuk menilai kualitas aplikasi. Meskipun metodologi tersebut inovatif, penilaian masih bergantung pada fitur leksikal dasar dan belum sepenuhnya memanfaatkan model klasifikasi probabilistik seperti *Naïve Bayes* untuk pemetaan sentimen yang lebih presisi [17]. Selanjutnya, Funnell et al. (2022) menerapkan analisis tematik dan sentimen terhadap aplikasi kesehatan mental untuk memahami persepsi pengguna terhadap fitur diagnosis berbasis *AI*. Studi tersebut menemukan bahwa sebagian besar pengguna menilai aplikasi dengan sentimen positif terkait aksesibilitas dan kemudahan penggunaan, namun kekhawatiran mengenai privasi dan akurasi masih dominan. Hasil ini menegaskan pentingnya penggunaan model analitik yang mampu menyeimbangkan sensitivitas linguistik dan klasifikasi sentimen secara efisien, terutama dalam konteks aplikasi yang melibatkan data pribadi [18]. Sementara itu, Malik et al. (2022) mengevaluasi umpan balik pengguna terhadap aplikasi terapi kognitif berbasis *AI* (*Wysa*) menggunakan pendekatan kualitatif tematik. Temuan mereka menyoroti bahwa pengklasifikasi klasik seperti *Naïve Bayes* atau *Support Vector Machine* (*SVM*) masih jarang digunakan dalam konteks aplikasi kesehatan mental, padahal metode tersebut memiliki potensi tinggi untuk mengotomatisasi penilaian kualitas aplikasi dari ribuan ulasan [19]. Di sisi lain, Palanivivayagam et al. (2023) melakukan tinjauan sistematis terhadap 224 studi klasifikasi teks dan menemukan bahwa *Naïve Bayes* tetap populer karena kesederhanaannya, tetapi performanya dapat menurun pada dataset dengan fitur yang saling bergantung tinggi [20]. Dari literatur tersebut dapat disimpulkan bahwa, meskipun berbagai pendekatan analisis sentimen telah diterapkan pada aplikasi kesehatan, masih terdapat kesenjangan riset dalam penerapan algoritma *Naïve Bayes* yang dioptimalkan untuk konteks ulasan pengguna aplikasi kesehatan wanita seperti *Flo Health*. Penelitian sebelumnya belum sepenuhnya mengeksplorasi kombinasi representasi fitur (seperti *TF-IDF*), evaluasi multi-metrik (*akurasi*, *presisi*, *recall*, *F1*), serta analisis komparatif dengan algoritma lain untuk menilai efektivitas model. Oleh karena itu, studi lanjutan dibutuhkan untuk mengisi celah tersebut dan memberikan kontribusi empiris terhadap penerapan pembelajaran mesin klasik dalam analisis sentimen mHealth yang terstandarisasi dan reproduktibel.

Penelitian ini bertujuan untuk menganalisis sentimen kepuasan pengguna terhadap aplikasi kesehatan wanita *Flo Health* berdasarkan ulasan pengguna di Google Play Store dengan menggunakan algoritma klasifikasi *Naïve Bayes*. Tujuan utama penelitian ini adalah untuk mengidentifikasi pola persepsi pengguna, menilai tingkat kepuasan terhadap fitur aplikasi, serta mengevaluasi efektivitas model *Naïve Bayes* dalam mengklasifikasi sentimen positif dan negatif. Dengan menggabungkan representasi fitur berbasis *TF-IDF* dan metrik evaluasi seperti *akurasi*, *presisi*, *recall*, dan skor *F1*, penelitian ini diharapkan mampu memberikan hasil yang lebih objektif dan terukur dalam menilai kinerja model klasifikasi teks [21]; [22]. Secara ilmiah, penelitian ini memiliki signifikansi penting dalam mengisi kesenjangan pengetahuan terkait keterbatasan riset terdahulu yang masih berfokus pada model analisis sentimen berbasis leksikon atau algoritma modern seperti *SVM* dan *BERT* tanpa mengevaluasi efisiensi algoritma klasik dalam konteks dataset ulasan aplikasi kesehatan [20]; [23]. Dengan demikian, studi ini memberikan kontribusi empiris terhadap bidang Informatika, khususnya dalam penerapan machine learning untuk analisis sentimen dan penambangan teks (*text mining*) pada data pengguna dunia nyata. Dari sisi praktis, hasil penelitian ini diharapkan dapat memberikan wawasan bagi pengembang aplikasi *mHealth*, khususnya terkait aspek yang paling memengaruhi kepuasan pengguna dan area yang perlu ditingkatkan untuk mendukung adopsi yang lebih luas. Selain itu, hasil klasifikasi dan analisis sentimen ini dapat menjadi dasar pengembangan sistem rekomendasi fitur atau peningkatan kualitas aplikasi berbasis data pengguna nyata, sebagaimana disarankan dalam studi oleh Li et al. (2024) dan Haoues et al. (2023) [7]; [6]. Dengan demikian, penelitian ini tidak hanya berkontribusi terhadap kemajuan metodologis dalam ranah Informatika, tetapi juga memberikan manfaat praktis dalam meningkatkan kualitas pengalaman pengguna dan efektivitas aplikasi kesehatan digital.

Penelitian ini menggunakan pendekatan kuantitatif berbasis pembelajaran mesin (*machine learning*) melalui penerapan algoritma *Naïve Bayes Classifier* untuk melakukan analisis sentimen pada ulasan pengguna aplikasi *Flo Health* di Google Play Store. Proses penelitian ini dimulai dengan pengumpulan data ulasan publik menggunakan metode *web scraping* yang etis dan terverifikasi, sebagaimana direkomendasikan oleh Dąbrowski et al. (2022) dan Yasin et al. (2024), yang menekankan pentingnya metadata dan anonimisasi pengguna untuk menjaga validitas serta kepatuhan etika penelitian [24]; [25]. Setelah data diperoleh,

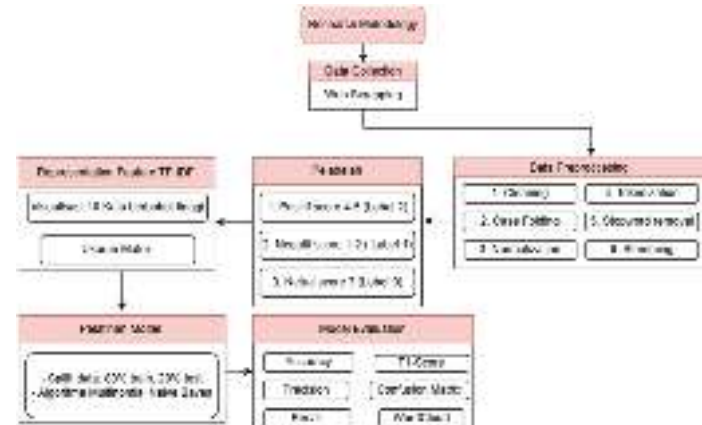
dilakukan tahapan prapemrosesan teks seperti pembersihan data, tokenisasi, penghapusan *stopwords*, dan *stemming* untuk meningkatkan akurasi klasifikasi [26], [27]. Selanjutnya, fitur teks dikonversi menjadi representasi numerik menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF)* guna memperkuat pembobotan kata dan meningkatkan diskriminasi antar kelas sentimen [28], [29]. Model *Multinomial Naïve Bayes* kemudian dilatih untuk mengklasifikasi ulasan ke dalam kategori sentimen positif, negatif, dan netral, diikuti dengan evaluasi kinerja menggunakan metrik akurasi, *presisi*, *recall*, dan skor *F1* guna mengukur reliabilitas model [30], [31]. Validasi model dilakukan dengan pendekatan *k-fold cross-validation* untuk memastikan generalisasi dan mengurangi risiko *overfitting* [32]. Pendekatan ini dipilih karena *Naïve Bayes* menawarkan kecepatan, efisiensi, dan interpretabilitas tinggi, yang menjadikannya ideal untuk analisis dataset besar dengan teks pendek seperti ulasan pengguna aplikasi seluler [20]. Dengan kombinasi metodologis ini, penelitian diharapkan mampu menghasilkan model analisis sentimen yang akurat, efisien, dan dapat direplikasi untuk mendukung pengembangan aplikasi kesehatan digital berbasis data pengguna nyata.

Apabila penelitian ini berhasil mencapai tujuannya, hasil yang diperoleh akan memberikan kontribusi penting terhadap pengembangan ilmu dan praktik di bidang Informatika, khususnya dalam ranah analisis sentimen dan pembelajaran mesin (*machine learning*). Secara teoretis, penelitian ini akan memperkuat pemahaman tentang bagaimana algoritma *Naïve Bayes* dapat diterapkan secara efektif dalam mengklasifikasi sentimen pengguna pada aplikasi *mobile health (mHealth)* dengan data ulasan yang tidak terstruktur [20]. Hasil penelitian ini juga diharapkan dapat menjadi acuan metodologis bagi penelitian selanjutnya yang berfokus pada analisis persepsi pengguna terhadap aplikasi kesehatan digital berbasis data besar dan representasi teks[6]. Dari sisi praktis, temuan penelitian ini dapat digunakan oleh pengembang aplikasi kesehatan seperti *Flo Health* untuk mengidentifikasi aspek fitur yang paling memengaruhi kepuasan pengguna, seperti akurasi prediksi siklus, kemudahan antarmuka, dan keandalan sistem. Dengan menerapkan analisis sentimen berbasis AI, pengembang dapat mengambil keputusan berbasis data (*data-driven decision making*) dalam

memperbaiki dan memperbarui aplikasi secara berkelanjutan [7]. Lebih lanjut, hasil penelitian ini berpotensi menjadi model referensi evaluasi kepuasan pengguna yang dapat diadaptasi untuk berbagai aplikasi *mHealth* lainnya, memperkuat ekosistem kesehatan digital yang berpusat pada pengguna. Bagi komunitas akademik, penelitian ini memberikan kontribusi empiris terhadap validasi model klasik *Naïve Bayes* di tengah dominasi model modern seperti *Support Vector Machine (SVM)* dan *BERT*, sekaligus memperluas wawasan tentang efektivitas model probabilistik sederhana pada data ulasan yang besar dan multibahasa [23]; [33]. Dengan demikian, hasil penelitian ini tidak hanya berimplikasi pada pengembangan teknologi kecerdasan buatan dalam bidang Informatika, tetapi juga mendorong integrasi antara analisis data, pengalaman pengguna (UX), dan inovasi digital dalam sektor kesehatan modern.

MATERIALS AND METHODS

Pada penelitian ini digunakan tahapan metodologi penelitian untuk menganalisis sentimen ulasan pengguna aplikasi Flo Health. Tahapan tersebut meliputi proses pengumpulan data, pelabelan data, prapemrosesan teks, representasi fitur menggunakan *TF-IDF*, pelatihan model menggunakan algoritma *Multinomial Naïve Bayes*, serta evaluasi model. Alur metodologi penelitian yang digunakan dalam penelitian ini dapat dilihat pada Gambar 1. Alur Penelitian.



Gambar 1. Alur Penelitian

Desain penelitian ini terdiri dari beberapa tahapan sistematis untuk menganalisis sentimen ulasan pengguna aplikasi Flo Health. Penelitian dimulai dengan pengumpulan data melalui teknik *web scraping* dari *Google Play Store*. Setelah data terkumpul, dilakukan pelabelan sentimen berdasarkan skor ulasan, yaitu positif, negatif, dan netral. Data kemudian diproses melalui tahap *preprocessing*, termasuk pembersihan teks, *case folding*, *normalisasi*, *tokenisasi*, penghapusan

stopword, dan *stemming* agar teks siap digunakan untuk analisis.

Tahap berikutnya adalah representasi fitur menggunakan *TF-IDF*, yang mengubah teks menjadi bentuk numerik serta menampilkan visualisasi kata berbobot tinggi dan ukuran matriks. Setelah itu, dilakukan pelatihan model dengan membagi dataset menjadi 80% data latih dan 20% data uji, menggunakan algoritma *Multinomial Naïve Bayes* untuk proses klasifikasi sentimen. Terakhir, performa model dievaluasi melalui metrik seperti *accuracy*, *precision*, *recall*, *F1-score*, serta visualisasi *wordcloud* dan *confusion matrix* untuk melihat hasil klasifikasi secara lebih detail.

RESULTS AND DISCUSSION

Tahap pembobotan fitur dengan metode *TF-IDF* selesai dilakukan, langkah berikutnya adalah melatih model klasifikasi sentimen menggunakan algoritma *Multinomial Naïve Bayes*. Algoritma ini dipilih karena memiliki kemampuan yang baik dalam mengolah data teks yang telah diubah ke dalam bentuk numerik, serta mampu bekerja secara efisien pada jumlah fitur yang besar.

Menurut [34] penelitian mengenai analisis sentimen pada aplikasi Kinemaster menggunakan metode *Naïve Bayes* menunjukkan bahwa algoritma tersebut mampu mengklasifikasikan ulasan pengguna ke dalam kategori sentimen positif dan negatif dengan cukup baik. Proses penelitian dilakukan melalui tahapan *preprocessing*, pembobotan *TF-IDF*, serta pembagian data menjadi data latih dan data uji dengan rasio 80:20 untuk menghasilkan model yang optimal.

Pada penelitian ini, pembagian data menggunakan metode *hold-out* dengan rasio 80:20, di mana 80% data digunakan sebagai data latih dan 20% sebagai data uji. Pemilihan metode ini didasarkan pada pertimbangan efisiensi komputasi dan kesederhanaan implementasi, terutama dengan jumlah dataset yang cukup besar sehingga sudah mampu merepresentasikan distribusi data secara baik.

Metode lain seperti *k-fold cross validation* sebenarnya dapat digunakan untuk menghasilkan evaluasi model yang lebih *robust* karena melakukan pelatihan dan pengujian secara berulang pada beberapa pembagian data. Namun, metode tersebut membutuhkan waktu komputasi yang lebih tinggi karena proses pelatihan dilakukan berkali-kali. Oleh karena itu, dalam penelitian ini dipilih metode *hold-out* 80:20 yang dinilai sudah cukup untuk

memberikan evaluasi performa model yang objektif dan efisien, serta telah banyak digunakan dalam penelitian analisis sentimen sebagai pendekatan standar.

Pembagian Data dibagi menjadi dua bagian dengan proporsi 80% sebagai data latih dan 20% sebagai data uji. Pada implementasinya, program menghasilkan pembagian sebesar 79,99% data latih dan 20,01% data uji.

Tujuan dari pembagian ini adalah agar model dapat mempelajari pola sentimen secara optimal dari sebagian besar data, sekaligus diuji performanya pada data yang tidak pernah dilihat sebelumnya. Dengan cara ini, kemampuan generalisasi model terhadap ulasan baru dapat dievaluasi secara lebih akurat. Dapat dilihat pada Tabel 1 Menunjukkan Hasil Pembagian Dataset

Tabel 1 Hasil Pembagian Dataset

No	Jenis Data	Jumlah	Persentase (%)
1	Data Latih	7.553	79.99%
2	Data Uji	1.889	20.01%
3	Total	9.442	100.00%

Berdasarkan hasil pembagian dataset penelitian, sebanyak 7.553 data digunakan sebagai data latih dengan persentase 79,99% dari keseluruhan dataset. Data ini berfungsi sebagai dasar bagi model untuk mempelajari pola sentimen yang terdapat dalam ulasan pengguna aplikasi Flo Health. Sementara itu, 1.889 data lainnya digunakan sebagai data uji, yang mencakup 20,01% dari total data. Data uji ini dimanfaatkan untuk mengevaluasi performa model terhadap ulasan yang tidak pernah dilihat sebelumnya, sehingga kemampuan generalisasi model dapat diukur secara objektif. Secara keseluruhan, jumlah data yang digunakan dalam penelitian ini mencapai 9.442 ulasan, yang kemudian diolah dan dianalisis untuk membangun model klasifikasi sentimen yang optimal.

Selanjutnya dilakukan proses pelatihan model menggunakan algoritma *Multinomial Naïve Bayes*. Proses ini melibatkan perhitungan probabilitas setiap kata dalam fitur *TF-IDF* terhadap masing-masing kelas sentimen (positif, negatif, netral). Dapat dilihat Pada Gambar 4.5

No	Parameter	Nilai
1	Jumlah Data Latih	7553
2	Jumlah Data Uji	1889
3	Jumlah Fitur TF-IDF	5658
4	Alpha (Laplace Smoothing)	1.0
5	Class Prior	[0.33, 0.34, 0.33]
6	Fit Prior	True
7	Force Alpha	True

Gambar 2 Hasil Pelatihan Model Naïve Bayes

Gambar 2 menunjukkan konfigurasi parameter yang digunakan pada proses pelatihan dan pengujian model *Naïve Bayes*. Dari total dataset yang tersedia, sebanyak 7.553 data dialokasikan sebagai data latih, sedangkan 1.889 data digunakan sebagai data uji. Proporsi ini menunjukkan pembagian sekitar 80:20 yang umum digunakan untuk menjaga keseimbangan antara kemampuan model untuk belajar dan kemampuan untuk dievaluasi secara objektif. Model dilatih menggunakan 5.658 fitur *TF-IDF*, yang merupakan representasi numerik dari kata-kata penting dalam seluruh korpus ulasan.

Selain itu, parameter *alpha* ditetapkan sebesar 1.0 sebagai bentuk *Laplace Smoothing*, yang bertujuan mencegah probabilitas bernilai nol pada kata-kata yang jarang muncul. Nilai *class prior* berada pada kisaran [0.33, 0.34, 0.33], menandakan distribusi kelas yang relatif seimbang antara sentimen negatif, netral, dan positif pada data latih. Opsi *Fit Prior* disetel ke *True*, model memperkirakan *prior* berdasarkan distribusi data. Sementara itu, *Force Alpha* juga bernilai *True*, memastikan penggunaan *smoothing* diterapkan secara konsisten. Secara keseluruhan, pengaturan parameter ini menunjukkan bahwa model *Naïve Bayes* dibangun dengan konfigurasi yang stabil dan sesuai untuk analisis sentimen berbasis teks.

Penggunaan *class prior* dalam skripsi ini dilakukan untuk mengatasi ketidakseimbangan distribusi data antar kelas sentimen tanpa melakukan manipulasi pada jumlah data. Pada dataset ulasan pengguna aplikasi *Flo Health*, kelas positif memiliki jumlah data yang jauh lebih dominan dibandingkan kelas negatif dan netral. Ketidakseimbangan ini berpotensi menyebabkan model lebih cenderung memprediksi kelas mayoritas. Dengan menerapkan *class prior*, probabilitas awal untuk setiap kelas dapat disesuaikan sehingga model tetap mempertimbangkan keberadaan kelas minoritas secara proporsional selama proses

prediksi. Pendekatan ini memungkinkan model bekerja lebih stabil tanpa harus menambah atau mengurangi jumlah data pada tiap kelas.

Class prior memiliki kelebihan karena dapat membantu menyeimbangkan kecenderungan prediksi model tanpa harus menambah atau mengubah jumlah data pada setiap kelas. Dengan mengatur probabilitas awal, model menjadi lebih peka terhadap kelas minoritas sehingga mengurangi bias terhadap kelas mayoritas. Selain itu, penggunaan *class prior* lebih sederhana dan efisien dibandingkan teknik *balancing* lain, serta sangat sesuai dengan sifat algoritma *Naïve Bayes* yang secara langsung memanfaatkan probabilitas awal dalam proses perhitungannya.

Kendati demikian, *class prior* memiliki kelemahan karena tidak dapat memperbaiki ketidakseimbangan pola fitur antar kelas. Jika jumlah data kelas minoritas sangat sedikit, model tetap kesulitan mengenali pola yang khas dari kelas tersebut. Efektivitasnya juga bergantung pada penentuan nilai *prior* jika tidak tepat, model dapat menjadi bias atau kurang stabil. Pada kondisi ketidakseimbangan yang sangat ekstrem, *class prior* sering kali kurang efektif dibandingkan metode *balancing* berbasis data.

Gambar berikut menampilkan hasil evaluasi model *Multinomial Naïve Bayes* yang telah diterapkan menggunakan pengaturan *class prior* untuk menyeimbangkan probabilitas awal setiap kelas sentimen. *Classification report* ini memuat nilai *precision*, *recall*, dan *F1-score* untuk kategori negatif, netral, dan positif, sehingga memberikan gambaran kinerja model dalam melakukan prediksi berdasarkan data uji. Dapat dilihat Pada Gambar 4.7 Akurasi Model: 87.08%

--- CLASSIFICATION REPORT ---				
	precision	recall	f1-score	support
Negatif	0.2987	0.4237	0.3448	59
Netral	0.0985	0.2321	0.1383	56
Positif	0.9617	0.9059	0.9329	1774
accuracy			0.8708	1889
macro avg	0.4503	0.5206	0.4720	1889
weighted avg	0.9152	0.8708	0.8910	1889

Gambar 3 Hasil Classification Report Model Naïve Bayes

Hasil evaluasi model *Multinomial Naïve Bayes* yang diterapkan menggunakan pengaturan *class prior* menunjukkan bahwa model memiliki performa yang cukup baik dalam mengklasifikasikan sentimen ulasan pengguna aplikasi *Flo Health*. Berdasarkan *classification report*, model memperoleh akurasi sebesar 87,08%, yang mengindikasikan bahwa sebagian besar prediksi sesuai dengan label sebenarnya. Pada kelas sentimen negatif, model menghasilkan nilai

Gambar berikut menampilkan confusion matrix hasil prediksi model *Multinomial Naïve Bayes*. Grafik ini digunakan untuk menunjukkan jumlah prediksi benar dan salah pada setiap kelas sentimen, sehingga memudahkan dalam mengevaluasi kemampuan model dalam membedakan sentimen positif, netral, dan negatif. Dapat dilihat Pada Gambar 4.8 Visualisasi Confusion Matrix Model Naïve Bayes



Word Cloud berikut menggambarkan kata-kata yang paling sering muncul pada ulasan dengan sentimen positif. Visualisasi ini membantu mengidentifikasi pola linguistik yang menunjukkan kepuasan atau pengalaman baik dari pengguna aplikasi *Flo Health*. Seperti pada gambar 4.9 wordcloud Sentimen Positif



7

Terlihat bahwa kata “sangat”, “bantu”, “bagus”, dan “aplikasi” mendominasi word cloud, menandakan bahwa pengguna banyak memberikan apresiasi terhadap manfaat serta kualitas aplikasi. Kemunculan kata-kata seperti “akurasi”, “tepat”, dan “prediksi” juga menunjukkan bahwa pengguna menilai aplikasi memberikan hasil yang akurat, terutama dalam konteks perhitungan siklus haid atau masa subur.

Selain itu, kata-kata seperti “haid”, “siklus”, “masa subur”, “menstruasi”, dan “hamil” mengindikasikan bahwa sebagian besar ulasan positif berkaitan langsung dengan fungsi utama aplikasi, yaitu membantu pengguna memantau siklus reproduksi. Ucapan syukur seperti “alhamdulillah” serta kata emosional seperti “keren”, “suka”, dan “manfaat” memperkuat interpretasi bahwa pengguna merasa terbantu dan puas setelah menggunakan aplikasi. Adapun kemunculan kata “flo” menunjukkan bahwa pengguna sering menyebut nama aplikasi secara langsung dalam ulasan positif mereka.

CONCLUSION

Berdasarkan hasil penelitian mengenai analisis sentimen kepuasan pengguna aplikasi Flo Health pada *Google Play Store* menggunakan algoritma *Naïve Bayes Classifier*, maka dapat disimpulkan beberapa hal sebagai berikut.

Proses pengumpulan dan pengolahan data ulasan pengguna aplikasi Flo Health berhasil dilakukan melalui metode *web scraping* dari *Google Play Store* sehingga diperoleh sebanyak 9.442 data ulasan yang kemudian digunakan sebagai dataset penelitian. Data tersebut selanjutnya melalui tahap *preprocessing* yang meliputi *cleaning*, *case folding*, *normalisasi*, *tokenisasi*, *stopword removal*, dan *stemming* untuk meningkatkan kualitas teks. Setelah melalui tahap pelabelan berdasarkan rating pengguna, diperoleh distribusi sentimen yaitu 8865 data sentimen positif (93,9%), 297 data sentimen negatif (3,1%), dan 280 data sentimen netral (3,0%), yang menunjukkan bahwa mayoritas pengguna memberikan ulasan positif terhadap aplikasi Flo Health.

Penerapan algoritma *Multinomial Naïve Bayes* dengan pembobotan fitur menggunakan metode *TF-IDF* berhasil digunakan untuk mengklasifikasikan sentimen ulasan pengguna ke dalam tiga kategori yaitu positif, negatif, dan netral. Proses pelatihan model dilakukan dengan pembagian dataset 80% data latih dan 20% data uji, sehingga model dapat mempelajari pola kata dari ulasan pengguna dan mengelompokkan

sentimen berdasarkan probabilitas kemunculan kata dalam setiap kelas.

Hasil evaluasi model menunjukkan bahwa algoritma *Naïve Bayes* mampu melakukan klasifikasi sentimen dengan performa yang cukup baik. Model menghasilkan akurasi sebesar 87,08%, dengan nilai *precision weighted* sebesar 0,9152, *recall* sebesar 0,8708, dan *F1-score* sebesar 0,8910. Hasil ini menunjukkan bahwa model memiliki kemampuan yang cukup efektif dalam mengidentifikasi sentimen ulasan pengguna aplikasi Flo Health, meskipun terdapat ketidakseimbangan jumlah data antara kelas positif, negatif, dan netral.

REFERENCE

- [1] E. K. Ameyaw, A. P. Adusei, and O. Ezezika, “Effectiveness of mHealth Apps for Maternal Health Care Delivery: Systematic Review of Systematic Reviews,” *Journal of Medical Internet Research*, vol. 26, p. e49510, 2024.
- [2] C. Lee and et al., “Mobile Health Apps for Breast Cancer: Content Analysis and Quality Assessment,” *JMIR mHealth and uHealth*, vol. 11, p. e43522, 2023.
- [3] M. Thanet, E. Wibowo, and P. Santosa, “Mobile health application for Thai women: investigation and model,” *BMC Medical Informatics and Decision Making*, vol. 22, p. 202, 2022.
- [4] M. Haoues, R. Mokni, and A. Sellami, “Machine learning for mHealth apps quality evaluation: An approach based on user feedback analysis,” *Software Quality Journal*, 2023.
- [5] J. Li et al., “Mining user reviews from hypertension management mobile health apps to explore factors influencing user satisfaction and their asymmetry: Comparative study,” *JMIR mHealth and uHealth*, vol. 12, p. e55199, 2024.
- [6] M. Haoues, R. Mokni, and A. Sellami, “Machine learning for mHealth apps quality evaluation: An approach based on user feedback analysis,” *Software Quality Journal*, 2023.
- [7] J. Li et al., “Mining user reviews from hypertension management mobile health apps to explore factors influencing user satisfaction and their asymmetry: Comparative study,” *JMIR*

- mHealth and uHealth*, vol. 12, p. e55199, 2024.
- [8] G. A. Id and G. Silahatoglu, "Insights into mobile health application market via a content analysis of marketplace data with machine learning," pp. 1–22, 2021, doi: 10.1371/journal.pone.0244302.
- [9] M. B. Id *et al.*, "ACCU 3 RATE: A mobile health application rating scale based on user reviews," pp. 1–24, 2021, doi: 10.1371/journal.pone.0258050.
- [10] E. L. Funnell, B. Spadaro, N. Martin-key, T. Metcalfe, S. Bahn, and S. Bahn, "mHealth Solutions for Mental Health Screening and Diagnosis: A Review of App User Perspectives Using Sentiment and Thematic Analysis," vol. 13, no. April, pp. 1–17, 2022, doi: 10.3389/fpsy.2022.857304.
- [11] T. Malik, A. J. Ambrose, C. Sinha, and T. Malik, "Evaluating User Feedback for an Artificial Intelligence – Enabled , Cognitive Behavioral Therapy – Based Mental Health App (Wysa): Qualitative Thematic Analysis Corresponding Author :," vol. 9, 2022, doi: 10.2196/35668.
- [12] A. S. Review, "Twenty Years of Machine-Learning-Based Text Classification : A Systematic Review," pp. 1–28, 2023.
- [13] A. Ulinuha, E. Majid, R. Nuari, U. T. Indonesia, and B. Lampung, "PERFORMANCE COMPARISON OF BERT METRICS AND CLASSICAL MACHINE LEARNING MODELS (SVM , NAIVE BAYES) FOR SENTIMENT ANALYSIS PERBANDINGAN KINERJA METRIK BERT DAN MODEL MACHINE LEARNING KLASIK (SVM , NAIVE BAYES) UNTUK," vol. 10, no. 2, pp. 741–752, 2025.
- [14] A. Bahar, T. Astuti, P. Arsi, and U. A. Purwokerto, "PERFORMANCE COMPARISON OF SVM , NAIVE BAYES , AND LOGISTIC REGRESSION CLASSIFICATION ALGORITHMS IN ANALYZING NOICE APP USER PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI SVM , NAIVE BAYES , DAN LOGISTIC REGRESSION DALAM ANALISIS ULASAN PENGGUNA," vol. 5, no. 4, pp. 469–477, 2024.
- [15] C. H. A. Vasconcelos and H. Wang, "A large scale analysis of mHealth app user reviews," *Empirical Software Engineering*, 2022.
- [16] G. A. Id and G. Silahatoglu, "Insights into mobile health application market via a content analysis of marketplace data with machine learning," pp. 1–22, 2021, doi: 10.1371/journal.pone.0244302.
- [17] M. B. Id *et al.*, "ACCU 3 RATE: A mobile health application rating scale based on user reviews," pp. 1–24, 2021, doi: 10.1371/journal.pone.0258050.
- [18] E. L. Funnell, B. Spadaro, N. Martin-key, T. Metcalfe, S. Bahn, and S. Bahn, "mHealth Solutions for Mental Health Screening and Diagnosis: A Review of App User Perspectives Using Sentiment and Thematic Analysis," vol. 13, no. April, pp. 1–17, 2022, doi: 10.3389/fpsy.2022.857304.
- [19] T. Malik, A. J. Ambrose, C. Sinha, and T. Malik, "Evaluating User Feedback for an Artificial Intelligence – Enabled , Cognitive Behavioral Therapy – Based Mental Health App (Wysa): Qualitative Thematic Analysis Corresponding Author :," vol. 9, 2022, doi: 10.2196/35668.
- [20] A. S. Review, "Twenty Years of Machine-Learning-Based Text Classification: A Systematic Review," pp. 1–28, 2023.
- [21] A. Rass, K. Tammi, and G. Demidova, "Mechatronics Technology and Transportation Sustainability," pp. 10–12, 2022.
- [22] M. Cristina, H. Lee, J. Braet, and J. Springael, "applied sciences Performance Metrics for Multilabel Emotion Classification: Comparing Micro , Macro , and Weighted F1-Scores," 2024.
- [23] A. Ulinuha, E. Majid, R. Nuari, U. T. Indonesia, and B. Lampung, "PERFORMANCE COMPARISON OF BERT METRICS AND CLASSICAL MACHINE LEARNING MODELS (SVM , NAIVE BAYES) FOR SENTIMENT

- ANALYSIS PERBANDINGAN KINERJA METRIK BERT DAN MODEL MACHINE LEARNING KLASIK (SVM , NAIVE BAYES) UNTUK,” vol. 10, no. 2, pp. 741–752, 2025.
- [24] J. Dąbrowski, E. Letier, A. Perini, and A. Susi, “Analysing app reviews for software engineering: a systematic literature review,” *Empirical Software Engineering*, vol. 27, no. 2, 2022, doi: 10.1007/s10664-021-10065-7.
- [25] A. Yasin, R. Fatima, A. Nauman, and Z. Wei, “Python data odyssey : Mining user feedback from google play store ☆,” *Data in Brief*, vol. 54, p. 110499, 2024, doi: 10.1016/j.dib.2024.110499.
- [26] M. A. Palomino, “applied sciences Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis,” 2022.
- [27] Rianto, A. B. Mutiara, E. P. Wibowo, and P. I. Santosa, “Improving the accuracy of text classification using stemming method, a case of non-formal Indonesian conversation,” *Journal of Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00413-1.
- [28] N. Umaira, C. Mohd, and N. A. Shafie, “Performance of TF-IDF for Text Classification Reviews on Google Play Store : Shopee,” vol. 9, no. 2, 2024, doi: 10.24191/jcrinn.v9i2.410.
- [29] M. I. Alfarizi, L. Syafaah, and M. Lestandy, “Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory),” vol. 10, no. 2, pp. 225–232, 2022.
- [30] M. Cristina, H. Lee, J. Braet, and J. Springael, “applied sciences Performance Metrics for Multilabel Emotion Classification : Comparing Micro , Macro , and Weighted F1-Scores,” 2024.
- [31] A. Rass, K. Tammi, and G. Demidova, “Mechatronics Technology and Transportation Sustainability,” pp. 10–12, 2022.
- [32] J. Opitz, “A Closer Look at Classification Evaluation Metrics and a Critical Reflection of Common Evaluation Practice,” vol. 12, no. 2018, pp. 820–836, 2024.
- [33] A. Bahar, T. Astuti, P. Arsi, and U. A. Purwokerto, “PERFORMANCE COMPARISON OF SVM , NAIVE BAYES , AND LOGISTIC REGRESSION CLASSIFICATION ALGORITHMS IN ANALYZING NOICE APP USER PERBANDINGAN KINERJA ALGORITMA KLASIFIKASI SVM , NAIVE BAYES , DAN LOGISTIC REGRESSION DALAM ANALISIS ULASAN PENGGUNA,” vol. 5, no. 4, pp. 469–477, 2024.
- [34] M. Enjeli and A. Wijaya, “Analisis Sentimen Penggunaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes,” vol. 2, pp. 89–98, 2024.