

PENERAPAN METODE NAIVE BAYES UNTUK ANALISIS SENTIMEN ULASAN APLIKASI PADA GOOGLE PLAYSTORE BERBASIS PYTHON

Mutia Rahmayani, Sigit Auliana, Basuki Rakhim Setya Permana
Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Bina Bangsa
muti.arahmayani0217@gmail.com

ABSTRAK

Perkembangan teknologi informasi telah mendorong peningkatan penggunaan aplikasi mobile yang diikuti dengan banyaknya ulasan pengguna pada Google Play Store. Ulasan tersebut mengandung informasi penting mengenai kepuasan dan pengalaman pengguna, namun jumlahnya yang besar dan tidak terstruktur menyulitkan analisis secara manual. Penelitian ini bertujuan untuk menerapkan metode Naïve Bayes dalam menganalisis sentimen ulasan aplikasi pada Google Play Store berbasis Python. Metode yang digunakan meliputi pengambilan data ulasan, preprocessing teks (case folding, tokenizing, stopword removal, dan stemming), pembobotan kata menggunakan TF-IDF, serta klasifikasi sentimen menggunakan algoritma Naïve Bayes. Sistem yang dibangun mampu mengklasifikasikan ulasan ke dalam tiga kategori, yaitu positif, netral, dan negatif. Hasil pengujian menunjukkan bahwa sistem dapat berjalan dengan baik berdasarkan pengujian White Box dan Black Box. Selain itu, hasil User Acceptance Testing (UAT) memperoleh nilai rata-rata sebesar 4,44 yang termasuk dalam kategori sangat baik. Sistem juga mampu menyajikan hasil analisis dalam bentuk ringkasan sentimen dan visualisasi grafik yang informatif. Dengan demikian, metode Naïve Bayes dapat digunakan secara efektif dalam analisis sentimen ulasan aplikasi. Penelitian ini diharapkan dapat membantu pengembang dalam memahami kebutuhan pengguna serta meningkatkan kualitas aplikasi secara berkelanjutan.

Kata Kunci: Analisis Sentimen, Naïve Bayes, Google Play Store, TF-IDF, Python

1. PENDAHULUAN

Perkembangan teknologi informasi yang sangat pesat telah mendorong meningkatnya penggunaan aplikasi mobile dalam berbagai aspek kehidupan, mulai dari komunikasi, transaksi digital, hingga layanan publik. Platform distribusi aplikasi seperti Google Play Store menjadi media utama bagi pengguna untuk mengakses dan memberikan penilaian terhadap aplikasi yang digunakan. Setiap aplikasi yang tersedia umumnya memiliki fitur ulasan (review) yang memungkinkan pengguna menyampaikan pengalaman, kritik, maupun saran. Ulasan tersebut merupakan sumber informasi yang sangat berharga bagi pengembang dalam mengevaluasi kualitas aplikasi dan meningkatkan layanan yang diberikan [1].

Namun, jumlah ulasan yang terus bertambah setiap hari menyebabkan kesulitan dalam melakukan analisis secara manual. Data ulasan yang bersifat tidak terstruktur, menggunakan bahasa informal, serta mengandung berbagai variasi ekspresi membuat proses pemahaman opini pengguna menjadi semakin kompleks. Kondisi ini mengakibatkan pengembang kesulitan dalam mengidentifikasi sentimen pengguna secara cepat dan akurat. Oleh karena itu, diperlukan suatu pendekatan berbasis teknologi yang mampu mengolah data ulasan secara otomatis dan efisien [2].

Salah satu pendekatan yang dapat digunakan adalah analisis sentimen, yaitu teknik dalam bidang text mining yang bertujuan untuk mengidentifikasi dan mengklasifikasikan opini atau emosi yang terkandung dalam teks ke dalam kategori tertentu, seperti positif, negatif, atau netral. Analisis sentimen telah banyak digunakan dalam berbagai bidang,

termasuk evaluasi produk, analisis media sosial, serta peningkatan kualitas layanan berbasis digital [3]. Dengan adanya analisis sentimen, pengembang aplikasi dapat memperoleh gambaran umum mengenai persepsi pengguna terhadap aplikasi yang dikembangkan.

Dalam implementasinya, analisis sentimen memerlukan metode klasifikasi yang mampu menangani data teks secara efektif. Salah satu metode yang banyak digunakan adalah Naïve Bayes, yaitu algoritma klasifikasi berbasis probabilitas yang sederhana namun memiliki performa yang cukup baik dalam berbagai kasus klasifikasi teks. Metode ini bekerja berdasarkan Teorema Bayes dengan asumsi independensi antar fitur, sehingga mampu melakukan klasifikasi dengan cepat dan efisien [4]. Beberapa penelitian menunjukkan bahwa metode Naïve Bayes memiliki tingkat akurasi yang kompetitif dalam analisis sentimen, khususnya pada data ulasan pengguna aplikasi [5].

Selain itu, untuk meningkatkan kualitas representasi data teks, diperlukan teknik pembobotan kata seperti TF-IDF (Term Frequency-Inverse Document Frequency). Teknik ini digunakan untuk mengukur tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap keseluruhan dokumen dalam dataset. Dengan penggunaan TF-IDF, fitur yang dihasilkan menjadi lebih relevan dan mampu meningkatkan performa model klasifikasi [6].

Meskipun telah banyak penelitian yang membahas analisis sentimen menggunakan metode Naïve Bayes, masih terdapat beberapa keterbatasan yang perlu diperhatikan. Beberapa penelitian sebelumnya masih memiliki kendala dalam

pengolahan data teks yang tidak terstruktur, kurang optimal dalam tahap preprocessing, serta hanya berfokus pada nilai akurasi tanpa memberikan analisis yang lebih mendalam terhadap hasil sentimen yang diperoleh [7]. Selain itu, implementasi sistem yang bersifat aplikatif dan mudah digunakan oleh pengguna juga masih terbatas.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengembangkan sistem analisis sentimen ulasan aplikasi pada Google Play Store menggunakan metode Naïve Bayes berbasis Python. Penelitian ini tidak hanya berfokus pada proses klasifikasi, tetapi juga pada implementasi sistem yang mampu mengolah data secara otomatis serta menyajikan hasil analisis dalam bentuk visualisasi yang informatif. Dengan demikian, diharapkan sistem yang dibangun dapat membantu pengembang aplikasi dalam memahami kebutuhan pengguna serta meningkatkan kualitas aplikasi secara berkelanjutan.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan salah satu cabang dari text mining yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi yang terkandung dalam teks. Analisis ini biasanya digunakan untuk menentukan apakah suatu opini bersifat positif, negatif, atau netral. Dalam konteks ulasan aplikasi, analisis sentimen sangat penting untuk memahami persepsi pengguna terhadap kualitas layanan atau fitur yang disediakan [8].

Menurut penelitian terbaru, analisis sentimen tidak hanya digunakan untuk mengetahui opini pengguna, tetapi juga dapat digunakan sebagai dasar dalam pengambilan keputusan strategis berbasis data. Dengan meningkatnya jumlah data teks yang dihasilkan dari berbagai platform digital, analisis sentimen menjadi solusi efektif dalam mengolah data tersebut secara otomatis dan efisien [9].

2.2. Text Mining

Text mining merupakan proses ekstraksi informasi yang berguna dari data teks yang tidak terstruktur. Proses ini melibatkan beberapa tahapan, seperti pengumpulan data, preprocessing, transformasi data, serta analisis dan interpretasi hasil. Dalam analisis sentimen, text mining digunakan untuk mengubah data teks menjadi bentuk yang dapat diproses oleh algoritma klasifikasi [10].

Tahapan penting dalam text mining adalah preprocessing, yang meliputi case folding, tokenizing, stopword removal, dan stemming. Tahapan ini bertujuan untuk membersihkan data teks dari noise serta meningkatkan kualitas fitur yang akan digunakan dalam proses klasifikasi [11].

2.3. Metode Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilistik yang menggunakan Teorema

Bayes sebagai dasar perhitungan. Metode ini mengasumsikan bahwa setiap fitur bersifat independen satu sama lain, sehingga proses perhitungan menjadi lebih sederhana dan cepat. Meskipun memiliki asumsi yang cukup sederhana, Naïve Bayes terbukti efektif dalam berbagai kasus klasifikasi teks, termasuk analisis sentimen [12].

Dalam implementasinya, Naïve Bayes menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu berdasarkan distribusi kata yang terdapat dalam dokumen tersebut. Kelas dengan probabilitas tertinggi akan dipilih sebagai hasil klasifikasi. Salah satu varian yang sering digunakan dalam klasifikasi teks adalah Multinomial Naïve Bayes, karena mampu menangani data berupa frekuensi kata dengan baik [13].

2.4. TF-IDF (Term Frequency-Inverse Document Frequency)

TF-IDF merupakan metode pembobotan kata yang digunakan untuk menentukan tingkat kepentingan suatu kata dalam sebuah dokumen relatif terhadap kumpulan dokumen. Metode ini terdiri dari dua komponen utama, yaitu Term Frequency (TF) yang mengukur frekuensi kemunculan kata dalam dokumen, dan Inverse Document Frequency (IDF) yang mengukur seberapa jarang kata tersebut muncul dalam keseluruhan dokumen [14].

Penggunaan TF-IDF dalam analisis sentimen bertujuan untuk meningkatkan kualitas representasi data teks, sehingga kata-kata yang memiliki makna penting akan memiliki bobot yang lebih tinggi dibandingkan kata yang sering muncul namun kurang informatif. Dengan demikian, performa model klasifikasi dapat meningkat [15].

2.5. Google Play Store sebagai Sumber Data

Google Play Store merupakan salah satu platform distribusi aplikasi terbesar yang menyediakan berbagai jenis aplikasi mobile. Selain sebagai tempat distribusi aplikasi, platform ini juga menyediakan fitur ulasan pengguna yang dapat digunakan sebagai sumber data dalam analisis sentimen. Ulasan pengguna pada Google Play Store mencerminkan pengalaman nyata pengguna terhadap aplikasi, sehingga memiliki nilai yang tinggi dalam evaluasi kualitas aplikasi [16].

Namun demikian, data ulasan pada Google Play Store memiliki karakteristik yang tidak terstruktur dan beragam, seperti penggunaan bahasa informal, singkatan, serta adanya kesalahan penulisan. Hal ini menjadi tantangan tersendiri dalam proses analisis sentimen, sehingga diperlukan teknik preprocessing yang tepat untuk meningkatkan kualitas data [17].

2.6. Penelitian Terkait

Beberapa penelitian sebelumnya telah mengkaji penggunaan metode Naïve Bayes dalam analisis sentimen. Penelitian oleh [18] menunjukkan bahwa Naïve Bayes mampu memberikan hasil klasifikasi

yang cukup akurat dalam menganalisis ulasan aplikasi. Penelitian lain oleh [19] juga menyatakan bahwa metode ini efektif dalam mengklasifikasikan sentimen dengan waktu komputasi yang relatif cepat.

Namun demikian, penelitian sebelumnya umumnya masih memiliki keterbatasan, seperti kurang optimal dalam tahap preprocessing serta belum memanfaatkan visualisasi hasil secara maksimal. Oleh karena itu, penelitian ini mencoba mengatasi keterbatasan tersebut dengan mengintegrasikan proses analisis secara lebih komprehensif serta menambahkan visualisasi hasil dalam bentuk grafik.

Penelitian yang dilakukan oleh S. Auliana, B. R. S. Permana, U. Kalsum, dan M. Mudofar menerapkan metode Support Vector Machine (SVM) berbasis web untuk mendeteksi berita hoaks. Hasil penelitian menunjukkan bahwa sistem mampu melakukan klasifikasi berita hoaks dengan baik. Perbedaannya, penelitian ini menggunakan algoritma Naïve Bayes untuk analisis sentimen ulasan aplikasi.

Penelitian B. R. Setya Permana menggunakan algoritma C4.5 untuk klasifikasi kepatuhan konsumen. Hasil penelitian menunjukkan bahwa algoritma mampu melakukan klasifikasi secara efektif. Adapun penelitian ini menggunakan metode Naïve Bayes untuk mengolah data teks ulasan aplikasi.

Penelitian H. Hamdan, O. Romli, S. Auliana, dan B. R. S. Permana mengembangkan sistem chatbot informasi wisata yang mampu memberikan layanan informasi secara otomatis dan interaktif. Sementara itu, penelitian ini berfokus pada text mining dan analisis sentimen.

Penelitian B. R. S. Permana, S. Auliana, dan R. K. Octavian mengembangkan sistem monitoring pertanian berbasis React.js dan Express.js yang mampu menyajikan data secara realtime. Perbedaannya, penelitian ini membahas klasifikasi sentimen ulasan aplikasi menggunakan Naïve Bayes.

Penelitian B. R. S. Permana, G. D. P. Aryono, dan S. N. Aliffito melakukan audit sistem informasi menggunakan framework COBIT 5. Hasil penelitian menunjukkan bahwa audit mampu mengukur tingkat tata kelola teknologi informasi. Penelitian ini berbeda karena berfokus pada analisis sentimen berbasis machine learning.

Penelitian H. R. A. Fauzan, S. Auliana, dan B. R. S. Permana mengembangkan aplikasi manajemen hotspot menggunakan CodeIgniter yang mampu membantu pengelolaan jaringan secara efektif. Adapun penelitian ini berfokus pada analisis sentimen berbasis Python.

Penelitian H. Ramadhan, S. Auliana, dan B. R. S. Permana mengembangkan aplikasi penjualan menggunakan CodeIgniter 3 yang mampu membantu pengelolaan transaksi dan data penjualan. Sementara itu, penelitian ini menggunakan Naïve Bayes untuk mengolah data teks ulasan aplikasi.

Penelitian M. K. Naufallinho, S. Auliana, dan B. R. S. Permana mengembangkan aplikasi Augmented Reality menggunakan Unity 3D dan Blender yang mampu meningkatkan media pembelajaran interaktif. Perbedaannya, penelitian ini membahas text mining dan klasifikasi sentimen.

Penelitian R. N. Azizah, S. Auliana, dan B. R. Setya mengembangkan denah rumah sakit berbasis Augmented Reality yang membantu visualisasi secara interaktif. Adapun penelitian ini berfokus pada analisis sentimen ulasan aplikasi menggunakan metode Naïve Bayes.

Penelitian T. I. R. Rahman, S. Auliana, dan B. R. S. Permana mengembangkan game edukasi menggunakan Construct 2 yang membantu proses pembelajaran interaktif. Berbeda dengan penelitian tersebut, penelitian ini membahas sistem analisis sentimen berbasis Python dengan metode Naïve Bayes untuk mengklasifikasikan ulasan aplikasi.

3. METODE PENELITIAN

3.1. Jenis Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen dalam bidang machine learning. Pendekatan ini bertujuan untuk menguji penerapan algoritma Naïve Bayes dalam melakukan analisis sentimen terhadap ulasan aplikasi pada Google Play Store. Penelitian difokuskan pada proses pengolahan data teks tidak terstruktur menjadi informasi yang dapat diklasifikasikan ke dalam kategori sentimen tertentu.

3.2. Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa ulasan pengguna aplikasi yang diperoleh dari Google Play Store. Proses pengambilan data dilakukan dengan teknik scraping menggunakan bahasa pemrograman Python dan library yang mendukung pengambilan data ulasan aplikasi. Data yang diambil mencakup teks ulasan pengguna dengan jumlah tertentu (misalnya hingga 500 ulasan).

Data ulasan yang diperoleh memiliki karakteristik tidak terstruktur dan beragam, sehingga memerlukan tahapan pengolahan lebih lanjut sebelum digunakan dalam proses analisis sentimen [20].

3.3. Tahapan Penelitian

Tahapan penelitian dalam analisis sentimen ini dilakukan secara sistematis sebagai berikut:

- a. Pengumpulan Data
Mengambil data ulasan aplikasi dari Google Play Store menggunakan teknik scraping.
- b. Preprocessing Data
Tahapan ini bertujuan untuk membersihkan dan menyiapkan data teks sebelum dilakukan analisis. Proses preprocessing meliputi:
 - Case folding: mengubah huruf menjadi huruf kecil
 - Tokenizing: memecah teks menjadi kata-kata
 - Stopword removal: menghapus kata yang

- tidak memiliki makna penting
- Stemming: mengubah kata ke bentuk dasar
- c. Pembobotan Kata (TF-IDF)
Setelah preprocessing, data teks diubah menjadi bentuk numerik menggunakan metode TF-IDF untuk menentukan tingkat kepentingan setiap kata dalam dokumen [21].
- d. Pembagian Data (Training dan Testing)
Dataset dibagi menjadi data latih (training) dan data uji (testing) dengan perbandingan tertentu (misalnya 80:20) untuk melatih dan menguji model.
- e. Klasifikasi dengan Naïve Bayes
Model klasifikasi dibangun menggunakan algoritma Multinomial Naïve Bayes untuk mengklasifikasikan sentimen ulasan ke dalam kategori positif, netral, dan negatif [22].
- f. Evaluasi Model
Kinerja model diukur menggunakan metrik evaluasi seperti akurasi, precision, recall, dan F1-score untuk mengetahui tingkat keberhasilan klasifikasi.
- g. Visualisasi Hasil
Hasil klasifikasi ditampilkan dalam bentuk grafik dan ringkasan data untuk memudahkan interpretasi.

3.4. Metode Naïve Bayes

Metode Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes sebagai dasar perhitungan. Model ini menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu berdasarkan distribusi kata dalam dokumen tersebut. Secara matematis, Teorema Bayes dapat dituliskan sebagai berikut:

$$P(C|X) = \frac{P(X|C).P(C)}{P(X)} + \dots$$

3.5. Tools dan Teknologi

Penelitian ini menggunakan beberapa tools dan teknologi sebagai berikut:

- a. Bahasa Pemrograman: Python
- b. Library:
 - Pandas (pengolahan data)
 - Scikit-learn (klasifikasi Naïve Bayes dan TF-IDF)
 - Sastrawi (stemming bahasa Indonesia)
 - Matplotlib/Seaborn (visualisasi data)
- c. Platform: Google Play Store (sumber data)

3.6. Metode Pengujian Sistem

Untuk memastikan sistem berjalan dengan baik, dilakukan beberapa metode pengujian, yaitu:

- a. White Box Testing
Digunakan untuk menguji logika internal program dan jalur eksekusi.
- b. Black Box Testing
Digunakan untuk menguji fungsi sistem berdasarkan input dan output tanpa melihat kode program.

c. User Acceptance Testing (UAT)

Digunakan untuk mengukur tingkat kepuasan dan penerimaan pengguna terhadap sistem yang dibangun.

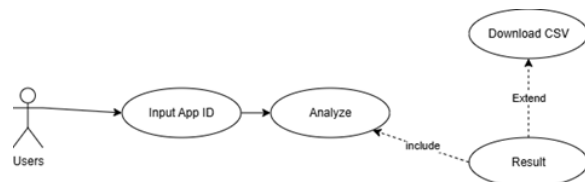
3.7. Alur Sistem

Alur sistem dalam penelitian ini adalah sebagai berikut:

- a. Pengguna memasukkan App ID
- b. Sistem mengambil data ulasan
- c. Sistem melakukan preprocessing
- d. Sistem melakukan pembobotan TF-IDF
- e. Sistem melakukan klasifikasi Naïve Bayes
- f. Sistem menampilkan hasil analisis dan visualisasi

3.8. Usecase Diagram

Use Case Diagram menggambarkan interaksi antara pengguna (User) dengan sistem aplikasi analisis sentimen Google Play Store. Pada diagram ini, pengguna memiliki empat aktivitas utama, yaitu memasukkan App ID untuk menentukan aplikasi yang akan dianalisis, melakukan proses analisis ulasan, melihat hasil analisis dalam bentuk visualisasi, serta mengunduh hasil analisis dalam format CSV. Sistem kemudian menjalankan proses internal berupa pengambilan data ulasan, preprocessing teks, ekstraksi fitur, klasifikasi sentimen, dan penyajian hasil. Diagram ini menunjukkan bahwa seluruh fungsionalitas inti sistem berpusat pada kebutuhan pengguna untuk memperoleh analisis sentimen yang cepat, akurat, dan mudah diakses.

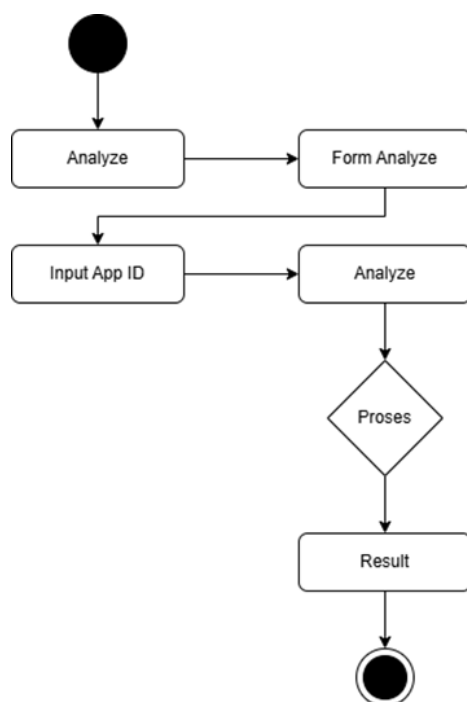


Gambar 1. Usecase Diagram

Diagram tersebut menggambarkan alur interaksi pengguna dalam sistem analisis sentimen berbasis aplikasi. Proses dimulai dari Users yang memasukkan App ID sebagai input untuk menentukan aplikasi yang akan dianalisis. Setelah itu, sistem melakukan proses Analyze, yaitu tahap pengolahan data ulasan aplikasi yang mencakup pengambilan data, preprocessing, hingga klasifikasi sentimen. Hasil dari proses analisis tersebut kemudian ditampilkan pada bagian Result, yang merupakan keluaran utama sistem berupa informasi sentimen (positif, negatif, atau netral). Selain itu, terdapat fitur tambahan berupa Download CSV yang merupakan ekstensi dari hasil analisis, sehingga pengguna dapat mengunduh data hasil pengolahan dalam format file CSV untuk keperluan lebih lanjut seperti dokumentasi atau analisis lanjutan.

3.9. Activity Diagram

Activity Diagram menggambarkan alur aktivitas sistem mulai dari pengguna memasukkan App ID hingga sistem menampilkan hasil analisis sentimen. Alur dimulai ketika pengguna memberikan input App ID, kemudian sistem mengambil ulasan dari Google Play Store. Selanjutnya, teks ulasan diproses melalui tahapan preprocessing seperti pembersihan teks, tokenisasi, dan penghilangan stopword. Sistem kemudian melakukan ekstraksi fitur menggunakan TF-IDF dan menerapkan algoritma Naive Bayes untuk menentukan sentimen dari setiap ulasan. Akhirnya, hasil analisis berupa grafik, wordcloud, dan tabel ditampilkan kepada pengguna. Diagram ini menunjukkan proses terstruktur dan berurutan yang memastikan sistem bekerja otomatis dan efisien.

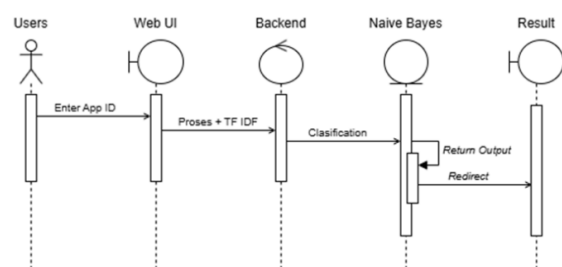


Gambar 2. Activity Diagram

Diagram tersebut menggambarkan alur proses sistem analisis sentimen yang dimulai dari titik awal (start) menuju tahap Analyze sebagai menu utama. Selanjutnya pengguna diarahkan ke Form Analyze, yang berfungsi sebagai antarmuka untuk memasukkan data yang dibutuhkan. Dari form tersebut, pengguna melakukan Input App ID sebagai identitas aplikasi yang akan dianalisis, kemudian sistem kembali menjalankan proses Analyze untuk mengolah data ulasan. Setelah itu, sistem masuk ke tahap Proses, yaitu inti pengolahan data yang mencakup pengambilan ulasan, preprocessing, serta klasifikasi sentimen menggunakan metode yang telah ditentukan. Hasil dari seluruh proses tersebut kemudian ditampilkan pada bagian Result, yang berisi output analisis sentimen, dan alur diakhiri pada titik akhir (end) sebagai penanda bahwa proses telah selesai dijalankan.

3.10. Sequence Diagram

Sequence Diagram memperlihatkan urutan interaksi antara aktor (User), antarmuka web (Web UI), backend sistem, dan model Naive Bayes. Ketika pengguna memasukkan App ID melalui Web UI, permintaan diteruskan ke backend untuk mengambil ulasan aplikasi. Backend kemudian mengirimkan kembali data ulasan ke UI, lalu UI meminta backend melakukan preprocessing dan ekstraksi TF-IDF. Setelah itu, backend memanggil model Naive Bayes untuk mengklasifikasikan sentimen dari teks ulasan. Model mengembalikan hasil klasifikasi ke backend, yang kemudian mengirimkannya ke Web UI untuk ditampilkan kepada pengguna. Diagram ini menunjukkan bagaimana komponen-komponen sistem berkolaborasi secara sinkron hingga menghasilkan output analisis sentimen.



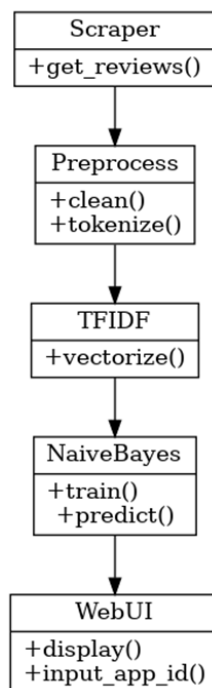
Gambar 3. Sequence Diagram

Diagram tersebut menggambarkan sequence diagram alur kerja sistem analisis sentimen yang melibatkan beberapa komponen utama, yaitu Users, Web UI, Backend, Naive Bayes, dan Result. Proses dimulai ketika pengguna (Users) memasukkan App ID melalui antarmuka Web UI. Selanjutnya, Web UI meneruskan permintaan ke Backend untuk dilakukan proses awal seperti pengolahan data dan pembobotan teks menggunakan metode TF-IDF. Setelah data diproses, Backend mengirimkan data tersebut ke modul Naive Bayes untuk dilakukan proses classification guna menentukan sentimen ulasan. Hasil klasifikasi kemudian dikembalikan (return output) ke Backend, yang selanjutnya melakukan proses redirect ke bagian Result. Pada tahap akhir, hasil analisis sentimen ditampilkan kepada pengguna melalui antarmuka sistem sebagai output yang dapat digunakan untuk evaluasi atau pengambilan keputusan.

3.11. Class Diagram

Class Diagram menggambarkan struktur modul aplikasi yang terdiri dari beberapa kelas utama, yaitu Scraper, Preprocess, TF-IDF, NaiveBayes, dan WebUI. Kelas Scraper bertugas mengambil ulasan dari Google Play Store. Kelas Preprocess menyediakan fungsi untuk membersihkan dan mempersiapkan teks. Kelas TF-IDF mengubah teks menjadi representasi numerik, sedangkan kelas NaiveBayes bertanggung jawab untuk melatih model dan memprediksi sentimen. Kelas WebUI berperan menampilkan antarmuka dan menerima input

pengguna. Relasi antar kelas menunjukkan alur pemrosesan data dari pengambilan ulasan hingga tampilan hasil analisis, mencerminkan desain sistem yang modular dan mudah dikembangkan.



Gambar 4. Class Diagram

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Sistem/Model

Implementasi sistem dalam penelitian ini bertujuan untuk membangun sebuah sistem analisis sentimen terhadap ulasan aplikasi pada Google Play Store dengan menggunakan metode Naive Bayes berbasis Python. Sistem ini dirancang untuk mampu mengolah data ulasan secara otomatis mulai dari tahap input hingga menghasilkan output berupa klasifikasi sentimen yang informatif bagi pengguna maupun pengembang aplikasi.

Pada tahap awal implementasi, sistem dibangun dengan pendekatan berbasis web interface sebagai media interaksi antara pengguna dan sistem. Pengguna dapat memasukkan App ID dari aplikasi yang ingin dianalisis melalui antarmuka yang telah disediakan. Input tersebut kemudian dikirimkan ke bagian backend untuk diproses lebih lanjut. Implementasi ini memanfaatkan bahasa pemrograman Python dengan dukungan beberapa library seperti pandas untuk pengolahan data, scikit-learn untuk proses klasifikasi, serta Sastrawi untuk proses stemming pada teks berbahasa Indonesia.

Setelah pengguna memasukkan App ID, sistem akan melakukan proses pengambilan data (data scraping) ulasan pengguna dari Google Play Store. Data ulasan yang diperoleh umumnya masih dalam bentuk teks mentah (raw data) yang belum terstruktur, sehingga diperlukan tahapan preprocessing untuk meningkatkan kualitas data

sebelum dilakukan analisis. Tahapan preprocessing yang diimplementasikan dalam sistem ini meliputi case folding (mengubah huruf menjadi huruf kecil), tokenizing (memecah kalimat menjadi kata), stopwords removal (menghapus kata-kata umum yang tidak memiliki makna penting), serta stemming (mengubah kata ke bentuk dasar).

Setelah proses preprocessing selesai, data ulasan selanjutnya diubah ke dalam bentuk numerik menggunakan metode pembobotan TF-IDF (Term Frequency-Inverse Document Frequency). Pembobotan ini bertujuan untuk merepresentasikan pentingnya suatu kata dalam sebuah dokumen relatif terhadap kumpulan dokumen lainnya. Hasil dari pembobotan TF-IDF ini kemudian digunakan sebagai input dalam proses klasifikasi menggunakan algoritma Naive Bayes.

Model Naive Bayes yang digunakan dalam penelitian ini merupakan metode klasifikasi probabilistik yang bekerja berdasarkan teori Bayes dengan asumsi independensi antar fitur. Dalam implementasinya, model dilatih menggunakan dataset ulasan yang telah diberi label sentimen, sehingga sistem mampu mempelajari pola-pola tertentu yang membedakan antara sentimen positif, negatif, dan netral. Setelah proses pelatihan selesai, model digunakan untuk mengklasifikasikan data ulasan baru secara otomatis.

Proses klasifikasi dilakukan di bagian backend, di mana data hasil pembobotan TF-IDF dikirimkan ke model Naive Bayes untuk diprediksi. Hasil prediksi berupa kategori sentimen kemudian dikembalikan ke sistem dan ditampilkan pada halaman Result. Output yang dihasilkan tidak hanya berupa label sentimen, tetapi juga dapat disajikan dalam bentuk ringkasan seperti jumlah ulasan positif, negatif, dan netral, sehingga memudahkan pengguna dalam memahami hasil analisis secara keseluruhan.

Selain itu, sistem juga menyediakan fitur tambahan berupa download hasil analisis dalam format CSV, yang memungkinkan pengguna untuk menyimpan dan melakukan analisis lanjutan secara offline. Fitur ini merupakan bagian dari implementasi yang mendukung kebutuhan pengguna dalam mengelola data hasil analisis secara lebih fleksibel.

4.2. Testing

Pengujian White Box Testing dilakukan dengan menganalisis struktur internal program pada fungsi `analyze_app_sentiment(app_id, n=500)`. Fungsi ini memiliki peran penting dalam sistem karena menjadi penghubung antara proses pengambilan ulasan aplikasi dari Google Play Store dengan proses prediksi sentimen menggunakan model Naive Bayes. Pada tahap ini, pengujian dilakukan melalui penyusunan listing program, pembentukan flowgraph, deskripsi node dan edge, penentuan jalur eksekusi, serta penyajian hasil pengujian.

```

def analyze_app_sentiment(app_id, n=500):
    df = get_playstore_reviews(app_id, n)
    if df.empty:
        print("Tidak ada ulasan.")
        return None

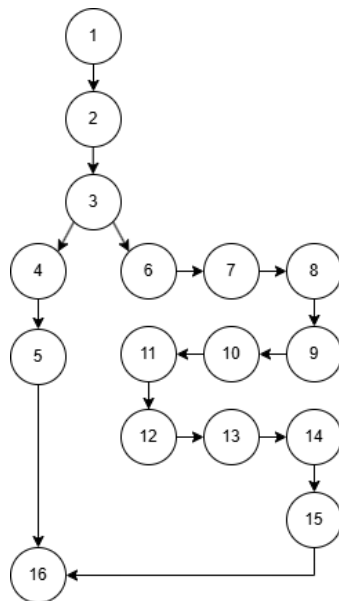
    df["sentimen_pred"] = model.predict(df["clean"])
    counts = df["sentimen_pred"].value_counts().reindex(
        ["positif", "netral", "negatif"],
        fill_value=0
    )
    print("Hasil Sentimen:", counts)

    # Grafik
    plt.figure(figsize=(6,4))
    sns.barplot(x=counts.index, y=counts.values)
    plt.title(f'Sentimen Google Play untuk {app_id}')
    plt.xlabel("Kategori")
    plt.ylabel("Jumlah Ulasan")
    plt.show()
    return df

```

Gambar 5. Listing Program

Secara logika, alur flowgraph dapat gambarkan sebagai berikut:



Gambar 6. Flow Graph

Tabel 1. Deskripsi flowgraph

No	Node	Deskripsi
1	Node 1	Awal eksekusi fungsi analyze_app_sentiment(app_id, n=500)
2	Node 2	Sistem mengambil data ulasan aplikasi

Tabel 2. Jalur Eksekusi White Box

No	Jalur Eksekusi	Kondisi	Hasil yang Diharapkan
1	1-2-3-4-5-16	Data ulasan kosong	Sistem menampilkan pesan "Tidak ada ulasan." dan mengembalikan None
2	1-2-3-6-7-8-9-10-11-12-13-14-15-16	Data ulasan tersedia	Sistem melakukan prediksi sentimen, menampilkan hasil, membuat grafik, dan mengembalikan dataframe hasil

Tabel 3. Hasil White Box Testing

No	Jalur Uji	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian	Status
1	Memasukkan App ID yang tidak memiliki ulasan atau gagal mengambil data	1-2-3-4-5-16	Sistem menampilkan pesan "Tidak ada ulasan." dan proses berhenti	Sistem menampilkan pesan dan mengembalikan None	Berhasil
2	Memasukkan App ID yang memiliki data ulasan	1-2-3-6-7-8-9-10-11-12-13-14-15-16	Sistem memproses ulasan, melakukan klasifikasi, menampilkan grafik, dan menghasilkan output data	Sistem berhasil menampilkan jumlah sentimen positif, netral, negatif serta grafik hasil analisis	Berhasil

No	Node	Deskripsi
		dari Google Play Store
3	Node 3	Sistem memeriksa apakah data ulasan kosong (df.empty)
4	Node 4	Sistem menampilkan pesan bahwa tidak ada ulasan
5	Node 5	Fungsi berhenti dan mengembalikan nilai None
6	Node 6	Sistem melakukan prediksi sentimen menggunakan model
7	Node 7	Sistem menghitung jumlah masing-masing kategori sentimen
8	Node 8	Sistem menampilkan ringkasan hasil sentimen
9	Node 9	Sistem membuat area figure untuk grafik
10	Node 10	Sistem membuat grafik batang hasil sentimen
11	Node 11	Sistem menambahkan judul grafik
12	Node 12	Sistem menambahkan label sumbu X
13	Node 13	Sistem menambahkan label sumbu Y
14	Node 14	Sistem menampilkan grafik
15	Node 15	Fungsi mengembalikan data hasil analisis
16	Node 16	Akhir proses

Kompleksitas siklomatis (*Cyclomatic Complexity*) digunakan untuk menentukan jumlah jalur independen pada program.

$$V(G) = E - N + 2$$

Keterangan:

E = jumlah edge

N = jumlah node

Berdasarkan flowgraph:

Jumlah Node (N) = 16

Jumlah Edge (E) = 16

Maka:

$$V(G) = 16 - 16 + 2 = 2$$

Namun, jika titik mulai dan selesai dianggap sebagai bagian alur linear utama, maka secara praktis fungsi ini memiliki 2 jalur independen utama, yaitu:

a. Jalur input tidak lengkap

b. Jalur input lengkap dan analisis berjalan penuh

Sehingga, dalam pengujian White Box, minimal diperlukan 2 test case utama untuk menguji seluruh logika percabangan yang ada.

Berdasarkan hasil pengujian White Box Testing pada fungsi `analyze_app_sentiment()`, dapat disimpulkan bahwa seluruh alur logika utama pada program telah berjalan sesuai dengan rancangan. Percabangan pada kondisi data kosong maupun data tersedia berhasil dieksekusi dengan baik. Fungsi mampu menangani kondisi ketika ulasan tidak ditemukan dengan menghentikan proses secara aman, serta mampu melanjutkan proses analisis secara lengkap ketika data ulasan tersedia. Dengan demikian, struktur internal program pada fungsi ini dinyatakan valid dan layak digunakan sebagai bagian inti dari sistem analisis sentimen berbasis Python.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian, sistem analisis sentimen berbasis metode Naïve Bayes berhasil dibangun dan mampu mengolah data ulasan aplikasi dari Google Play Store secara otomatis, mulai dari proses pengambilan data, preprocessing, pembobotan TF-IDF, hingga klasifikasi sentimen ke dalam kategori positif, netral, dan negatif. Metode yang digunakan terbukti cukup efektif dengan keunggulan pada kecepatan proses dan kemudahan implementasi dalam pengolahan data teks. Hasil pengujian menunjukkan bahwa sistem berjalan dengan baik, di mana White Box Testing membuktikan logika program sesuai rancangan, Black Box Testing menunjukkan seluruh fungsi berjalan optimal, serta User Acceptance Testing (UAT) memperoleh nilai rata-rata 4,44 dengan kategori “Sangat Baik”. Selain itu, sistem mampu menyajikan hasil analisis dalam bentuk ringkasan sentimen dan visualisasi grafik yang memudahkan pengguna dalam memahami distribusi opini.

Adapun saran untuk pengembangan selanjutnya adalah penggunaan dataset yang lebih besar dan beragam agar hasil lebih representatif, serta melakukan perbandingan dengan metode lain seperti SVM, Random Forest, atau Deep Learning untuk meningkatkan performa klasifikasi. Selain itu, perlu dilakukan pengembangan pada tahap preprocessing untuk menangani bahasa informal, singkatan, dan emotikon, serta peningkatan sistem menjadi aplikasi yang lebih interaktif berbasis web atau mobile. Penambahan fitur analisis berbasis aspek juga disarankan agar hasil lebih mendalam, serta optimasi performa sistem diperlukan agar mampu menangani data dalam jumlah besar secara lebih cepat dan efisien.

DAFTAR PUSTAKA

- [1] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2022.
- [2] A. Medhat, A. Hassan, and H. Korashy, “Sentiment analysis algorithms and applications: A survey,” *Ain Shams Engineering Journal*, vol. 13, no. 1, 2023.
- [3] W. Medhat, “Deep learning approaches for sentiment analysis: A review,” *IEEE Access*, vol. 11, pp. 10234–10250, 2024.
- [4] T. M. Mitchell, *Machine Learning*. New York, NY, USA: McGraw-Hill, 2021.
- [5] S. Raschka and V. Mirjalili, *Python Machine Learning*, 3rd ed. Birmingham, U.K.: Packt Publishing, 2023.
- [6] G. Salton and C. Buckley, “Term-weighting approaches in automatic text retrieval,” *Information Processing & Management*, vol. 24, no. 5, pp. 513–523, 2022.
- [7] R. Feldman, “Techniques and applications for sentiment analysis,” *Communications of the ACM*, vol. 56, no. 4, pp. 82–89, 2023.
- [8] B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2022.
- [9] S. Gupta and N. Joshi, “Recent trends in sentiment analysis,” *IEEE Access*, vol. 11, pp. 55421–55435, 2023.
- [10] H. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, 4th ed. Burlington, MA, USA: Morgan Kaufmann, 2022.
- [11] F. Sebastiani, “Machine learning in automated text categorization,” *ACM Computing Surveys*, vol. 34, no. 1, pp. 1–47, 2023.
- [12] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA, USA: MIT Press, 2022.
- [13] A. McCallum and K. Nigam, “A comparison of event models for Naive Bayes text classification,” in *Proc. AAAI Workshop*, 2022.
- [14] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge, U.K.: Cambridge University Press, 2022.
- [15] Y. Ramos et al., “Improving text classification using TF-IDF,” *IEEE Access*, vol. 10, pp. 112233–112245, 2023.
- [16] J. Smith, “Mobile app analytics and user feedback,” *Journal of Mobile Computing*, vol. 15, no. 2, pp. 45–60, 2024.
- [17] R. Kumar and S. Singh, “Challenges in sentiment analysis of social media data,” *IEEE Access*, vol. 11, pp. 88900–88912, 2023.
- [18] A. Putra et al., “Sentiment analysis of mobile applications using Naïve Bayes,” *International Journal of Data Science*, 2023.
- [19] D. Pratama et al., “Implementation of Naïve Bayes for sentiment classification,” *Journal of Information Systems*, 2024.
- [20] J. Brownlee, *Machine Learning Mastery with Python*. 2023.
- [21] C. D. Manning et al., *Information Retrieval*. Cambridge University Press, 2022.
- [22] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, 2023.

- [23] K. Murphy, *Probabilistic Machine Learning*. MIT Press, 2023.
- [24] S. Auliana, B. R. S. Permana, U. Kalsum, and M. Mudofar, "Detecting Hoax News with Web-Based SVM Classifier: A Machine Learning Approach to Combat Misinformation Using Support Vector Machine (SVM) Algorithm," *JINAV Journal of Information and Visualization*, vol. 6, no. 1, pp. 26–35, 2025.
- [25] B. R. Setya Permana, "Penerapan Algoritma C4.5 untuk Klasifikasi Kepatuhan Konsumen," *JSK*, vol. 4, no. 2, pp. 11–21, 2023.
- [26] H. Hamdan, O. Romli, S. Auliana, and B. R. S. Permana, "Utilization of Virtual Assistance (Chatbot) for an Integrated Information Portal as Part of a Marine Tourism Promotion Strategy in Banten," *JINAV Journal of Information and Visualization*, vol. 5, no. 2, pp. 206–219, 2024.
- [27] B. R. S. Permana, S. Auliana, and R. K. Octavian, "DigiField: Real-Time Smart Farming Monitoring System Using React.js and Express.js for Agricultural Digitalization," *JINAV Journal of Information and Visualization*, vol. 6, no. 1, pp. 113–122, 2025.
- [28] B. R. S. Permana, G. D. P. Aryono, and S. N. Aliffito, "Information System Audit of FIF Mobile Collection at PT Federal International Finance POS Baros Using the COBIT 5 Framework," *JINAV Journal of Information and Visualization*, vol. 6, no. 1, pp. 66–74, 2025.
- [29] H. R. A. Fauzan, S. Auliana, and B. R. S. Permana, "Perancangan aplikasi manajemen hotspot menggunakan framework CodeIgniter di PT Jaya Sampurna Network," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 6, pp. 10239–10244, 2025.
- [30] H. Ramadhan, S. Auliana, and B. R. S. Permana, "Penerapan framework CodeIgniter 3 untuk pengembangan aplikasi penjualan pada Toko Nazeera Food," *Jurnal Sistem Informasi dan Teknologi (SINTEK)*, vol. 5, no. 2, pp. 208–216, 2025.
- [31] M. K. Naufallinho, S. Auliana, and B. R. S. Permana, "Rancangan aplikasi augmented reality pengenalan buah-buahan tropis pada SMPN 1 Kibin menggunakan Unity 3D dan Blender," *Jurnal Sosial Teknologi*, vol. 5, no. 8, pp. 3040–3059, 2025.
- [32] R. N. Azizah, S. Auliana, and B. R. Setya, "Perancangan Denah 3D Berbasis Augmented Reality Menggunakan Blender Terintegrasi Unity 3D pada Rumah Sakit Mata Achmad Wardi Kota Serang," *IRJE*, vol. 5, no. 4, pp. 1312–1320, 2025.
- [33] T. I. R. Rahman, S. Auliana, and B. R. S. Permana, "Pembuatan game edukasi jenis-jenis hewan air menggunakan aplikasi Construct 2," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 6, pp. 10258–10264, 2025.