KLASIFIKASI SENTIMEN ULASAN APLIKASI MOTORKU X DI GOOGLE PLAYSTORE DENGAN METODE NAIVE BAYES DAN TEKNIK NLP

Lukas Febriyanto Chandra [1], Anugrah Prasetya [2], Novita Rahmayuna [3], Nurtriana Hidayati [4]
[1,2,4] Sistem Informasi, Universitas Semarang
[3] Sistem Informasi, Universitas Bina Nusantara
Jalan Soekarno-Hatta, Tlogosari, Semarang, Indonesia
febriyantolukas@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan sentimen dari ulasan pengguna aplikasi Motorku X yang tersedia pada platform Google Play Store, dengan menerapkan metode *Naïve Bayes* serta teknik Pengolahan Bahasa Alami (*Natural Language Processing/NLP*). Ulasan yang diberikan oleh pengguna merupakan sumber informasi yang bernilai, karena mencerminkan pengalaman serta persepsi mereka terhadap kualitas layanan aplikasi. Oleh sebab itu, analisis dan klasifikasi sentimen menjadi esensial dalam memperoleh pemahaman yang lebih mendalam mengenai tingkat kepuasan dan permasalahan yang dihadapi oleh pengguna. Permasalahan utama yang melatarbelakangi penelitian ini adalah rendahnya responsivitas pengembang dalam mengidentifikasi dan menangani keluhan pengguna secara sistematis. Ketidaktahuan terhadap pola sentimen dalam ulasan pengguna dapat menyebabkan keterlambatan dalam perbaikan layanan, berkurangnya kepercayaan pengguna, serta penurunan kepuasan yang berimplikasi pada tingkat retensi pengguna. Dalam konteks pengembangan aplikasi yang kompetitif, ketidakmampuan dalam memanfaatkan masukan pengguna secara efektif dapat berujung pada menurunnya daya saing dan reputasi aplikasi di pasar digital. Untuk menjawab permasalahan tersebut, data ulasan dikumpulkan dari Google Play Store dan selanjutnya diproses melalui beberapa tahap praproses teks, seperti tokenisasi, penghapusan stopwords, serta transformasi *TF-IDF* (*Term Frequency-Inverse Document Frequency*) untuk memberikan bobot pada kata-kata yang dianggap penting. Metode *Naïve Bayes* dipilih sebagai algoritma klasifikasi karena kemampuannya dalam melakukan prediksi dengan cepat dan akurat. Hasil penelitian menunjukkan bahwa model klasifikasi sentimen yang dibangun dengan *Naïve Bayes* berhasil mencapai akurasi sebesar 83,5% dengan rasio data latih-uji 80:20, dengan nilai presisi 84%, *recall* 99%, dan *F1-Score* 90%.

Kata kunci : *Klasifikasi Sentimen, Naive Bayes, NLP, Motorku X, Preprocessing, Google Playstore*

1. PENDAHULUAN

Klasifikasi sentimen merupakan salah satu bagian penting dalam analisis data, terutama dalam konteks ulasan aplikasi mobile yang semakin banyak digunakan di era digital saat ini. Ulasan pengguna di *platform* seperti Google Play Store memberikan wawasan berharga mengenai persepsi dan pengalaman pengguna terhadap suatu aplikasi. Di antara berbagai pendekatan yang ada, penggunaan metode klasifikasi berbasis *machine learning*, seperti *Naive Bayes*, telah menunjukkan efektivitas yang signifikan dalam mengidentifikasi sentimen positif dan negatif dalam teks [1], [2].

Aplikasi Motorku X, yang merupakan salah satu platform yang menyediakan layanan terkait kendaraan dan informasi untuk pengguna motor di Indonesia, memerlukan pemahaman yang mendalam tentang *feedback* yang diberikan oleh penggunanya. Ulasan pengguna pada platform distribusi aplikasi seperti Google Play Store tidak hanya mencerminkan kepuasan atau ketidakpuasan terhadap fitur tertentu, tetapi juga mengandung informasi yang berharga terkait pengalaman nyata pengguna dalam menggunakan aplikasi. Data ini bersifat dinamis, aktual, dan terus bertambah seiring meningkatnya jumlah pengguna. Dalam konteks Motorku X, ulasan-ulasan tersebut dapat mencerminkan isu-isu teknis, antarmuka pengguna, keandalan layanan, serta efisiensi fitur yang disediakan. Oleh karena itu, jika dianalisis secara sistematis, ulasan ini dapat berfungsi sebagai sumber masukan yang kaya untuk memperbaiki kualitas aplikasi secara berkesinambungan.. Selain itu, dengan meningkatnya jumlah pengguna dan ulasan yang terus bertambah, metode otomatis seperti *Naive Bayes* menawarkan solusi yang efisien dalam pengolahan data untuk menentukan klasifikasi sentimen secara cepat dan akurat [3], [4].

Naive Bayes sendiri merupakan algoritma yang berdasarkan pada teorema *Bayes* yang menerapkan asumsi independensi antar fitur. Meskipun sederhana, metode ini telah terbukti efektif dalam banyak aplikasi, termasuk di bidang analisis sentimen. Sebagai contoh, dalam studi yang dilakukan oleh Syaripah et al., penggunaan *Naive Bayes* untuk klasifikasi ulasan aplikasi lainnya menunjukkan capaian akurasi yang memuaskan, menggarisbawahi kemampuannya dalam menangani data teks yang besar dan beragam [5]. Selain itu, konsep teknik pengolahan bahasa alami (*Natural Language Processing* atau *NLP*) yang mendasari pemrosesan data teks juga sangat relevan untuk fitur-fitur seperti tokenisasi dan penghapusan *stop words*, yang esensial dalam persiapan dataset untuk analisis sentimen [6].

Dalam konteks ini, penelitian ini bertujuan untuk mengeksplorasi dan menerapkan metode *Naive Bayes*

dalam klasifikasi sentimen terhadap ulasan aplikasi Motorku X di Google Play Store. Dengan memanfaatkan teknik *NLP* untuk pengolahan awal data, diharapkan hasil analisis ini dapat memberikan gambaran yang jelas dan informatif mengenai respon pengguna aplikasi. Penelitian ini diharapkan dapat menjadi kontribusi bagi pengembang aplikasi dalam upaya meningkatkan pengalaman pengguna dan performa aplikasi di pasaran.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terdahulu

Penelitian mengenai klasifikasi sentimen pada ulasan pengguna aplikasi telah banyak dilakukan, khususnya dengan menggunakan metode *Naïve Bayes* sebagai salah satu pendekatan yang populer dalam *text classification*. Beberapa studi terdahulu dalam rentang waktu 2022 hingga 2024 telah mengkaji efektivitas metode ini pada berbagai *domain* aplikasi.

Septiani dan Budi (2022) melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi Ipusnas milik Perpustakaan Nasional Republik Indonesia. Penelitian ini membandingkan tiga algoritma, yaitu *Naïve Bayes*, *Logistic Regression*, dan *Support Vector Machine (SVM)*. Hasil evaluasi menunjukkan bahwa model *Naïve Bayes* memberikan kinerja yang kompetitif dengan akurasi sebesar 85%, presisi 84%, *recall* 82%, dan *F1 Score* 83%, menggunakan pendekatan TF-IDF untuk vektorisasi data ulasan [7].

Khoirunnisaa et al. (2024) mengkaji klasifikasi teks ulasan aplikasi Netflix pada platform Google Play Store menggunakan metode *Naïve Bayes* dan *SVM*. Dalam penelitian tersebut, model *Naïve Bayes* mencapai akurasi sebesar 78%, presisi 76%, *recall* 75%, dan *F1 Score* 75,5% [8]. Hasil ini menunjukkan bahwa *Naïve Bayes* tetap mampu memberikan hasil yang cukup baik meskipun dibandingkan dengan metode yang lebih kompleks seperti *SVM*.

Gilbert (2023) menerapkan metode *Naïve Bayes* untuk mengklasifikasikan sentimen dari ulasan aplikasi MyPertamina. Hasil penelitian menunjukkan bahwa metode ini memperoleh akurasi sebesar 72%, presisi 70%, *recall* 71%, dan *F1 Score* 71% [3], [4]. Meskipun hasilnya relatif lebih rendah dibandingkan beberapa penelitian lainnya, studi ini tetap menunjukkan potensi metode *Naïve Bayes* dalam konteks aplikasi layanan publik.

Tanggraeni dan Sitokdana (2022) menganalisis sentimen pengguna terhadap aplikasi *E-Government* yang tersedia di Google Play Store. Metode *Naïve Baye*s yang digunakan dalam studi ini berhasil mencapai akurasi sebesar 80%, presisi 77%, *recall* 78%, dan *F1 Score* 77,5% [3]. Temuan ini mengindikasikan bahwa metode *Naïve Bayes* cukup efektif dalam menangani data teks dari layanan pemerintahan digital.

Herjanto dan Carudin (2024) dalam penelitiannya terhadap aplikasi SIREKAP membandingkan algoritma *Random Forest* dengan *Naïve Bayes* dalam klasifikasi sentimen ulasan pengguna. Meskipun fokus utama penelitian pada *Random Forest*, hasil perbandingan menunjukkan bahwa *Naïve Bayes* tetap mampu memberikan akurasi sebesar 76%, dengan presisi 75%, *recall* 73%, dan F1 *Score* 74% [9].

Andi et al. (2023) melakukan penelitian berjudul "Analisis Sentimen Mahasiswa terhadap Layanan Bimbingan Tugas Akhir", yang menunjukkan bahwa algoritma *Naïve Bayes* mampu menghasilkan akurasi sebesar 85%, presisi 84%, *recall* 82%, dan *F1-Score* 83% [10]. Dalam konteks serupa, Meiyanti et al. (2023) menganalisis sentimen mahasiswa terhadap dosen melalui kuesioner dan mencatat hasil akurasi 88%, presisi 87%, *recall* 86%, serta *F1-Score* 86,5% [11].

Secara umum, ketujuh studi tersebut menunjukkan bahwa algoritma *Naïve Bayes* memiliki kinerja yang layak digunakan dalam konteks klasifikasi sentimen pada berbagai jenis aplikasi. Meskipun kinerjanya dapat bervariasi tergantung pada kualitas data, teknik praproses, dan pendekatan vektorisasi yang digunakan, *Naïve Bayes* tetap menjadi salah satu pilihan metode yang efisien dan dapat diandalkan dalam tugas klasifikasi teks.

2.2. Analisis Sentimen

Analisis sentimen merupakan pendekatan dalam bidang pengolahan data tekstual yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini atau emosi yang terkandung dalam suatu teks. Klasifikasi ini umumnya dibagi menjadi tiga kategori, yaitu sentimen positif, negatif, dan netral. Dalam konteks aplikasi digital, analisis sentimen sering digunakan untuk mengevaluasi persepsi pengguna terhadap produk atau layanan tertentu.

2.3. NLP

Pengolahan Bahasa Alami (*NLP*) adalah cabang dari kecerdasan buatan (*Artificial Intelligence*) yang memungkinkan komputer untuk memahami, menginterpretasikan, dan memproses bahasa manusia. Dalam konteks klasifikasi teks, proses *NLP* melibatkan beberapa tahapan seperti *case folding*, tokenisasi, penghapusan kata umum (*stopword removal*), dan *stemming*. Pratmanto et al. dalam penelitiannya menerapkan tahapan-tahapan preprocessing tersebut sebelum menganalisis data ulasan pengguna, dengan tujuan meningkatkan akurasi klasifikasi sentimen yang dilakukan [12]. Tahapan ini penting untuk menghasilkan representasi data yang bersih dan konsisten sebagai input bagi model klasifikasi.

2.4. Naïve Bayes

Naïve Bayes merupakan algoritma klasifikasi berbasis Teorema Bayes dengan asumsi independensi antar fitur. Algoritma ini dikenal sederhana, efisien dalam menangani data berskala besar, serta mampu memberikan hasil yang kompetitif dalam klasifikasi

teks, termasuk analisis sentimen. Widiastuti et al. dalam penelitiannya membuktikan bahwa penerapan *Naïve Bayes* dapat secara efektif mengklasifikasikan konten teks dengan performa yang kompetitif dibandingkan dengan algoritma lain seperti *SVM* atau *KNN* [3]. Keunggulan utamanya terletak pada kecepatan pelatihan dan rendahnya kebutuhan komputasi, menjadikannya pilihan yang relevan dalam berbagai penelitian terkini.. Menggunakan rumus:

$$P(C|X) = \frac{(P(X|C)xP(C)}{P(X)} \quad (1)$$

Keterangan:
P(C|X): Probabilitas hipotesis C diberikan data X
P(X|C): Probabilitas data X terjadi bila diketahui kelas C
P(C): Probabilitas awal dari kelas C
P(X): Probabilitas data (*evidence*)

2.5. TF-IDF

TF-IDF merupakan teknik pembobotan dalam representasi data teks yang mengukur seberapa penting suatu kata dalam dokumen tertentu relatif terhadap kumpulan dokumen lainnya. Pembobotan TF-IDF membantu dalam mengurangi dominasi kata-kata yang terlalu sering muncul namun tidak membawa makna spesifik (seperti kata hubung), sehingga dapat meningkatkan kualitas representasi fitur. Dalam penelitian Pratmanto et al., TF-IDF digunakan sebagai metode ekstraksi fitur utama sebelum proses klasifikasi sentimen dilakukan menggunakan algoritma *Naïve Bayes* dan *KNN* [2]. Penggunaan *TF-IDF* terbukti mampu meningkatkan kinerja model dengan merepresentasikan fitur secara lebih informatif.

2.6. Confusion Matrix

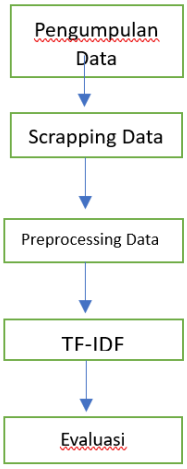
Confusion matrix adalah alat evaluasi yang digunakan untuk mengukur performa model klasifikasi dengan cara membandingkan hasil prediksi dengan label aktual. Matriks ini menghasilkan empat nilai utama, yaitu *True Positive (TP*), *True Negative (TN)*, *False Positive (FP*), dan *False Negative (FN*), yang kemudian digunakan untuk menghitung metrik evaluasi seperti akurasi, presisi, *recall*, dan *F1 Score*. *Confusion matrix* membantu dalam mengevaluasi efektivitas algoritma *machine learning* dengan lebih komprehensif [13].

2.7. Word Cloud

Wordcloud adalah representasi *visual* dari kumpulan kata dalam dokumen teks, di mana ukuran dari masing-masing kata mencerminkan frekuensinya dalam korpus tersebut. Visualisasi ini memberikan gambaran eksploratif terhadap topik atau tema dominan dalam data dan sangat efektif dalam konteks analisis sentimen karena dapat memperjelas opini atau isu yang paling sering dibicarakan oleh pengguna [14]. Teknik ini sering digunakan dalam tahapan eksplorasi data untuk mengidentifikasi kata-kata yang paling sering muncul dalam ulasan pengguna. Visualisasi *wordcloud* dapat memberikan gambaran awal mengenai topik-topik dominan atau fitur yang relevan sebelum dilakukan analisis yang lebih mendalam. Dalam konteks analisis sentimen, *wordcloud* berfungsi sebagai alat bantu untuk memahami persepsi umum pengguna terhadap suatu aplikasi.

3. METODE PENELITIAN

Penelitian ini bertujuan untuk melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi Motorku X yang tersedia pada platform Google Play Store. Untuk mencapai tujuan tersebut, penelitian ini menggunakan pendekatan kuantitatif dengan metode klasifikasi teks berbasis algoritma *Naïve Bayes*. Adapun tahapan yang dilakukan dalam proses penelitian ini meliputi pengumpulan data, prapemrosesan data teks, ekstraksi fitur menggunakan teknik *Term Frequency-Inverse Document Frequency (TF-IDF)*, proses klasifikasi, dan evaluasi performa model.



Gambar 1. Metodelogi penelitian

3.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan menggunakan teknik *web scraping* dengan memanfaatkan pustaka *google-play-scraper* berbasis *Python*. Teknik ini digunakan untuk mengakses dan mengambil data ulasan pengguna dari aplikasi Motorku X yang tersedia di Google Play Store. Data yang berhasil dikumpulkan berjumlah sebanyak 1.000 ulasan, mencakup informasi penting seperti nama pengguna, tanggal ulasan, rating yang diberikan, serta isi atau konten ulasan. Data ini selanjutnya digunakan sebagai dasar dalam proses prapemrosesan, pelabelan sentimen, dan klasifikasi.

3.2. Preprocessing

Setelah data berhasil dikumpulkan, tahap selanjutnya adalah melakukan pelabelan (*labeling*) terhadap data ulasan. Pelabelan dilakukan dengan mengklasifikasikan setiap ulasan ke dalam dua kategori sentimen, yaitu positif dan negatif,

berdasarkan nilai rating yang diberikan oleh pengguna. Ulasan dengan rating tinggi (misalnya 4 dan 5) dikategorikan sebagai sentimen positif, sementara ulasan dengan rating rendah (1 dan 2) dikategorikan sebagai sentimen negatif. Ulasan dengan rating netral (nilai 3) dikecualikan dari analisis untuk menyederhanakan proses klasifikasi biner. Setelah proses pelabelan, data ulasan kemudian melalui tahap prapemrosesan teks untuk meningkatkan kualitas representasi data dan mempersiapkannya sebelum dimasukkan ke dalam model klasifikasi. Prapemrosesan ini mencakup beberapa tahapan sebagai berikut:

a. *Case Folding*
   Tahap ini bertujuan untuk menyeragamkan bentuk huruf dalam teks dengan cara mengubah seluruh karakter menjadi huruf kecil (*lowercase*). Proses ini penting untuk menghindari perbedaan makna akibat variasi kapitalisasi, seperti kata "Aplikasi" dan "aplikasi" yang secara semantik merujuk pada entitas yang sama.
b. Tokenisasi
   Tokenisasi dilakukan setelah *case folding*, dengan memecah kalimat atau teks menjadi unit-unit kata (*token*). Proses ini bertujuan untuk mengidentifikasi elemen-elemen dasar dalam dokumen teks yang dapat digunakan sebagai fitur untuk analisis selanjutnya.
c. *Stopword Removal*
   Stopword removal adalah proses menghapus kata-kata umum yang tidak memberikan kontribusi signifikan terhadap pemaknaan sentimen, seperti kata "dan", "yang", "di", serta kata hubung dan kata bantu lainnya. Penghapusan kata-kata ini bertujuan untuk mengurangi *noise* dalam data dan meningkatkan efisiensi model.
d. *Stemming*
   Stemming merupakan proses mengubah kata turunan menjadi bentuk dasarnya (kata dasar). Dalam penelitian ini, proses *stemming* dilakukan menggunakan *library NLP* Bahasa Indonesia seperti Sastrawi. Tujuan dari *stemming* adalah untuk mengurangi kompleksitas fitur yang disebabkan oleh variasi morfologis kata.
e. Pembersihan Simbol dan Karakter non-teks
   Tahap ini mencakup penghapusan simbol-simbol khusus, angka, serta tanda baca yang tidak relevan dalam proses analisis. Pembersihan ini penting agar model hanya menerima data teks yang bersih dan bermakna.

3.3. TF-IDF

Tahapan ekstraksi fitur dalam penelitian ini dilakukan menggunakan metode *Term Frequency-Inverse Document Frequency (TF-IDF),* yang bertujuan mengubah data teks menjadi representasi vektor numerik sebagai input bagi model machine learning. *TF* mencerminkan frekuensi kemunculan kata dalam dokumen, sedangkan *IDF* mengukur kelangkaan kata di seluruh korpus. Kombinasi keduanya menghasilkan bobot yang mencerminkan pentingnya kata dalam konteks dokumen. Bobot tersebut digunakan untuk membentuk matriks fitur *TF-IDF*, yang merepresentasikan setiap dokumen dalam bentuk numerik. Matriks ini kemudian digunakan sebagai input bagi algoritma *Naïve Bayes* dalam proses klasifikasi sentiment.

3.4. Evaluasi

Tahap akhir dari metode penelitian ini adalah evaluasi performa model klasifikasi yang telah dibangun. Evaluasi dilakukan untuk mengukur sejauh mana model mampu mengklasifikasikan sentimen dengan akurasi dan konsistensi yang baik. Dalam penelitian ini, digunakan beberapa metrik evaluasi yang umum diterapkan dalam klasifikasi biner, yaitu akurasi, presisi, *recall*, dan *F1-score*. Sebagai pelengkap, evaluasi juga menggunakan confusion matrix, yaitu matriks yang menyajikan jumlah prediksi yang benar dan salah untuk masing-masing kelas (positif dan negatif). *Confusion matrix* memberikan informasi rinci mengenai distribusi kesalahan klasifikasi, seperti jumlah *True Positive*, *True Negative, False Positive,* dan *False Negative*. Seluruh proses evaluasi ini diimplementasikan menggunakan pustaka *Scikit-learn*, yang menyediakan berbagai fungsi dan modul untuk perhitungan metrik klasifikasi serta visualisasi hasil evaluasi. Dengan demikian, performa model dapat dianalisis secara menyeluruh untuk menilai efektivitas metode klasifikasi sentimen yang diterapkan.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dalam penelitian ini, teknik pengumpulan data dilakukan melalui metode *web scraping* menggunakan platform *Google Colab* dan bahasa pemrograman *Python*, dengan bantuan pustaka *google_play_scraper*. Pustaka ini digunakan untuk mengambil data ulasan pengguna aplikasi Motorku X yang tersedia di Google Play Store secara otomatis dan terstruktur. Data yang diperoleh mencakup berbagai informasi penting seperti nama pengguna, tanggal ulasan, *rating* (skor) yang diberikan, dan isi ulasan. Seluruh hasil pengambilan data kemudian disimpan dalam struktur *DataFrame* dan diekspor dalam format *CSV* (*Comma-Separated Values*) untuk keperluan analisis lebih lanjut

Tabel 1. Hasil *scrapping data*

score	***at***	***content***	***clean_content***
1	10/06/2025 08:33:00	Lelet	Lelet
2	07/03/2025 18:13:00	Bikin ribet	Bikin ribet
3	23/11/2024 13:28:00	Lumayan	Lumayan
4	16/02/2025 10:56:00	*Good*	*Good*
5	05/06/2025 03:30:00	Keren dan puas	Keren puas

4.2. Preprocessing

Tahap awal dalam proses analisis adalah pelabelan data yang dilakukan secara otomatis menggunakan skrip *Python* di Google Colab. Pelabelan didasarkan pada nilai rating yang diberikan pengguna, di mana ulasan dengan rating ≤ 2 diklasifikasikan sebagai sentimen negatif, sedangkan rating ≥ 3 diklasifikasikan sebagai sentimen positif.

Tabel 2. Hasil labeling data

score	***at***	***content***	***sentiment***	***clean_content***
1	10/06/2025 08:33:00	Lelet	*negative*	Lelet
2	07/03/2025 18:13:00	Bikin ribet	*negative*	Bikin ribet
3	23/11/2024 13:28:00	Lumayan	*positive*	Lumayan
4	16/02/2025 10:56:00	*Good*	*positive*	*Good*
5	05/06/2025 03:30:00	Keren dan puas	*positive*	Keren puas dan puas

Selanjutnya melakukan perbaikan deskripsi dan penjelasan untuk setiap proses dalam *text processing*:

a. *Data Cleaning*

Data cleaning merupakan proses pembersihan data teks dengan menghapus karakter yang tidak relevan seperti tanda baca, angka, emoji, dan simbol khusus. Tujuan dari tahap ini adalah untuk mengurangi *noise* dan memastikan data yang digunakan dalam analisis bersih serta konsisten..

Tabel 3. Hasil *data cleaning*

score	**at**	**content**	**sentiment**	**clean_content**
1	10/06/2025 08:33:00	Lelet	*negative*	Lelet
2	07/03/2025 18:13:00	Bikin ribet	*negative*	Bikin ribet
3	23/11/2024 13:28:00	Lumayan	*positive*	Lumayan
4	16/02/2025 10:56:00	*Good*	*positive*	*Good*
5	05/06/2025 03:30:00	Keren dan puas	*positive*	Keren puas dan puas

b. *Case Folding*

Case folding adalah proses mengubah seluruh teks menjadi huruf kecil guna memastikan konsistensi dalam pemrosesan kata. Misalnya, frasa “Layanan Bagus” diubah menjadi “layanan bagus”.

Tabel 4. Hasil *case folding*

score	***at***	***content***	**clean_*content***
1	10/06/2025 08:33:00	Lelet	lelet
2	07/03/2025 18:13:00	Bikin ribet	bikin ribet
3	23/11/2024 13:28:00	Lumayan	lumayan
4	16/02/2025 10:56:00	*Good*	*good*
5	05/06/2025 03:30:00	Keren dan puas	keren puas

c. *Stopword Removal*

Stopword removal adalah proses menghapus kata-kata umum yang sering muncul namun tidak memiliki makna signifikan dalam analisis, seperti “itu”, “yang”, “di”, dan “ke” dan lain sebagainya.

Tabel 5. Hasil *stopword removal*

Sebelum	**Sesudah**
sungguh memuaskan *service* di astra honda	sungguh memuaskan *service* astra honda

d. *Tokenizing*

Tokenization adalah proses memecah teks menjadi unit-unit kata atau token. Sebagai contoh, frasa “layanan bagus” diubah menjadi [“layanan”, “bagus”].

Tabel 6. Hasil *tokenizing*

Sebelum	**Sesudah**
sungguh memuaskan *service* astra honda	[sungguh, memuaskan, *service*, astra, honda]

e. *Stemming*

Stemming adalah proses mengubah kata ke bentuk dasarnya dengan menghapus imbuhan di awal atau akhir kata. Contohnya, kata “kemudahan” diubah menjadi “mudah”.

Tabel 7. Hasil *stemming*

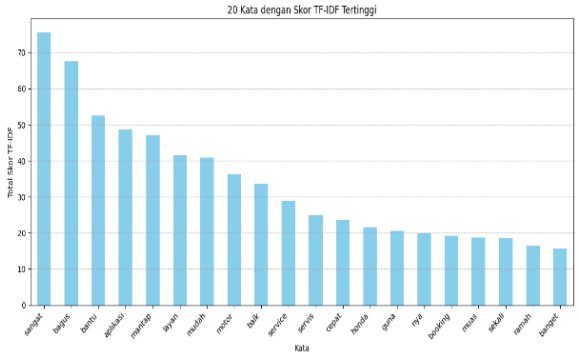
Sebelum	**Sesudah**
[sungguh, memuaskan, *service*, astra, honda]	[sungguh, puas, *service*, astra, honda]

f. Pembagian Data

Setelah melalui tahapan prapemrosesan, dataset dibagi menjadi dua subset utama, yaitu data latih (*training data*) dan data uji (*testing data*). Pembagian ini dilakukan untuk mengevaluasi kinerja model klasifikasi dalam berbagai skenario proporsi data. Terdapat tiga skenario pembagian data yang digunakan, yaitu: skenario pertama dengan 80% data latih dan 20% data uji (800 data latih, 200 data uji), skenario kedua dengan 70% data latih dan 30% data uji (700 data latih, 300 data uji), serta skenario ketiga dengan 60% data latih dan 40% data uji (600 data latih, 400 data uji). Pembagian ini bertujuan untuk menilai stabilitas dan kemampuan generalisasi model terhadap data yang belum pernah dilihat. Dataset terdiri dari dua kelas sentimen, yaitu positif dan negatif, yang digunakan sebagai label. Strategi ini memungkinkan evaluasi yang objektif terhadap efektivitas model dalam mengklasifikasikan sentimen ulasan pengguna.

4.3. TF-IDF

Metode Term Frequency-Inverse Document Frequency (TF-IDF) digunakan untuk mengubah data teks menjadi representasi numerik yang dapat diolah oleh model klasifikasi. TF-IDF menghitung bobot setiap kata berdasarkan frekuensinya dalam dokumen dan kelangkaannya di seluruh korpus, sehingga memungkinkan identifikasi fitur penting dalam teks. Proses ini menghasilkan kosakata yang digunakan untuk membentuk vektor fitur pada data latih dan data uji secara konsisten. Penerapan TF-IDF dalam penelitian ini membantu mengungkap kata-kata dengan pengaruh signifikan terhadap klasifikasi sentimen, di antaranya "sangat", "bagus", "bantu", "aplikasi", dan "mantap" yang dominan pada sentimen positif.



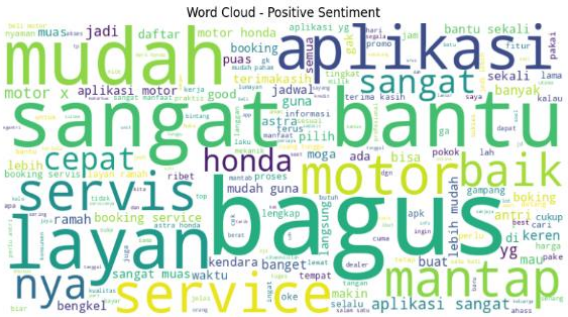
Gambar 2. Hasil *tf-idf*

4.4. Model Naïve Bayes

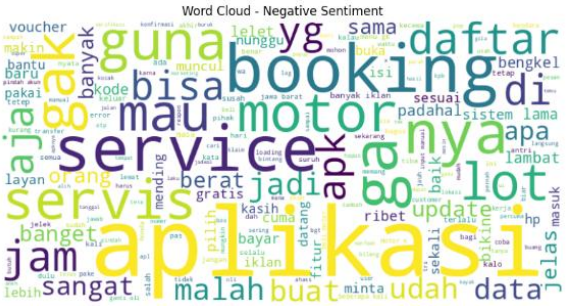
Proses klasifikasi pada penelitian ini menggunakan algoritma Multinomial Naïve Bayes, yang dikenal efektif untuk tugas klasifikasi teks. Algoritma ini mengandalkan prinsip probabilistik melalui penerapan *Teorema Bayes* dengan asumsi conditional independence antar fitur. Data ulasan yang digunakan telah melalui proses *preprocessing*, meliputi pembersihan teks, tokenisasi, penghapusan stopwords, normalisasi kata, serta transformasi menggunakan metode *TF-IDF* untuk merepresentasikan kata-kata penting dalam bentuk vektor numerik. Vektor ini kemudian digunakan sebagai input dalam pelatihan model. Selama pelatihan, algoritma mempelajari distribusi kata pada masing-masing kelas sentimen untuk membentuk model probabilistik. Model tersebut digunakan untuk memprediksi sentimen data baru dengan memilih kelas yang memiliki probabilitas posterior tertinggi. Evaluasi kinerja dilakukan menggunakan metrik akurasi, *precision*, *recall*, dan *F1-score*. Hasil evaluasi menunjukkan bahwa model memiliki performa yang baik dalam mengklasifikasikan sentimen, sehingga dapat digunakan sebagai alat bantu dalam menganalisis opini pengguna terhadap aplikasi secara efisien dan akurat.

4.5. Word Cloud

Word Cloud merupakan teknik visualisasi teks yang menampilkan kata-kata paling sering muncul dalam korpus secara grafis, dengan ukuran font yang proporsional terhadap frekuensinya. Dibangun berdasarkan *Term Document Matrix (TDM*), visualisasi ini mempermudah identifikasi kata-kata dominan dan memberikan gambaran awal mengenai konten dataset sebelum dilakukan analisis kuantitatif lanjutan.



Gambar 3. Hasil *sentiment* positif



Gambar 4. Hasil *sentiment* negatif

4.6. Evaluasi

Tahap akhir dalam penelitian ini adalah evaluasi performa model klasifikasi untuk menilai akurasi dan konsistensi dalam mengklasifikasikan sentimen. Evaluasi dilakukan menggunakan *confusion matrix*, yang menunjukkan perbandingan antara prediksi model dan label aktual melalui empat komponen: *True Positive (TP), False Positive (FP), False Negative (FN),* dan *True Negative (TN).* Matriks ini memungkinkan identifikasi tingkat kesalahan dan keberhasilan prediksi pada tiap kelas sentimen. Proses evaluasi dilakukan dengan bantuan pustaka Scikit-learn, yang menyediakan fungsi perhitungan metrik evaluasi dan visualisasi confusion matrix sebagai dasar penilaian kinerja model.

Tabel 8. Hasil *confusion matrix*

	Actual negative	***Actual positive***
Pred negative	*TN (8)*	*FP (31)*
Pred positive	*FN (2)*	*TP (159)*

Berdasarkan hasil pengujian model klasifikasi terhadap data uji, diperoleh nilai-nilai pada *confusion matrix* sebagai berikut: *True Negative (TN*): sebanyak 8 data dengan label negatif berhasil diklasifikasikan

secara tepat sebagai sentimen negatif. *True Positive (TP*): Sebanyak 159 data dengan label positif berhasil diprediksi dengan benar sebagai sentimen positif. *False negative (FN):* Terdapat 2 data dengan label positif yang salah diklasifikasikan sebagai negatif. *False Positive (FP):* Terdapat 31 data dengan label negatif yang keliru diprediksi sebagai positif.

Selanjutnya, nilai-nilai dari *confusion matrix* akan digunakan untuk menghitung kinerja model, yang mencakup akurasi, presisi, *recall*, dan skor F1. Proses perhitugannya adalah sebagai berikut:

a. $Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)}$ x100%

$Accuracy = \frac{(159+8)}{(159+8+31+2)}$ x100%

$Accuracy = \frac{167}{200}$ x 100% = 83.5%

b. $Precession = \frac{TP}{(TP+FP)}$

$Precession = \frac{159}{(159+31)} = 0.84$

c. $Recall = \frac{TP}{(TP+FN)}$

$Recall = \frac{159}{(159+2)} = 0.99$

d. $F1\text{-}Score = 2X \frac{(\text{Recall xPrecession})}{(\text{Recall+Precession})}$

$F1\text{-}Score = 2X \frac{(0.99x0.84)}{(0.99+0.84)} = 0.904$ (2)

Perhitungan nilai-nilai ini bertujuan untuk memberikan gambaran kuantitatif terhadap kemampuan model dalam mengklasifikasikan data secara tepat, khususnya dalam konteks analisis sentimen biner (positif dan negatif)

Tabel 9. Hasil *f1-score*

	precision	***recall***	***f1*-score**	***support***
support	0.80	0.21	0.33	39
positive	0.84	0.99	0.91	161
accuracy			0.83	200
macro avg	0.82	0.60	0.62	200
weighted avg	0.83	0.83	0.79	200

Berdasarkan Tabel 9, model Multinomial *Naïve Bayes* menunjukkan kinerja yang tinggi dalam mengklasifikasikan sentimen positif, dengan nilai *precision* sebesar 0.84, *recall* 0.99, dan *F1-score* 0.91. Namun, performa untuk kelas support jauh lebih rendah, dengan *recall* hanya 0.21 dan *F1-score* 0.33, yang mengindikasikan adanya ketidakseimbangan distribusi data antar kelas. Akurasi keseluruhan model mencapai 83%, namun nilai *macro average F1-score* sebesar 0.62 menunjukkan bahwa performa tidak merata di semua kelas. Untuk mengukur konsistensi performa model, evaluasi dilakukan pada tiga skenario pembagian data sebagai berikut:

Tabel 10. Hasil perbandingan data latih

Proporsi data latih	Akurasi	Presisi	*Recall*	*F1-Score*
80:20	83.5%	0.84	0.99	0.90
70:30	82.3%	0.82	0.97	0.89
60:40	80.8%	0.80	0.95	0.87

Tabel 10 menunjukkan perbandingan performa model pada tiga proporsi data latih. Hasil terbaik diperoleh pada proporsi 80:20 dengan akurasi 83,5% dan *F1-score* 0.90. Ketika jumlah data latih dikurangi menjadi 70:30 dan 60:40, terjadi penurunan akurasi masing-masing menjadi 82,3% dan 80,8%, serta penurunan *F1-score* menjadi 0.89 dan 0.87. Hal ini menunjukkan bahwa peningkatan jumlah data latih berdampak positif terhadap kinerja model dalam melakukan klasifikasi sentimen.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang dilakukan, diketahui bahwa mayoritas pengguna aplikasi Motorku X memberikan ulasan yang bersifat positif. Ulasan positif tersebut umumnya berkaitan dengan kemudahan penggunaan, pemahaman antarmuka, kualitas layanan, serta kenyamanan dalam pengoperasian aplikasi. Namun demikian, terdapat pula sejumlah ulasan negatif yang menyoroti kelemahan aplikasi, seperti kurangnya responsivitas sistem dan ketidakpuasan terhadap beberapa fitur layanan. Analisis sentimen yang diterapkan menggunakan algoritma *Multinomial Naïve Bayes* menunjukkan performa klasifikasi yang sangat baik. Model mencapai akurasi sebesar 83,5% pada proporsi pembagian data 80:20, yang juga menghasilkan nilai presisi 84%, *recall* 99%, dan *F1-score* 90%. Berdasarkan perbandingan terhadap tiga skenario pembagian data (80:20, 70:30, dan 60:40), diketahui bahwa proporsi 80:20 memberikan hasil evaluasi yang paling optimal. Hal ini mengindikasikan bahwa model mampu mengidentifikasi sentimen positif secara efektif, sekaligus menjaga keseimbangan antara ketepatan dan sensitivitas prediksi. Meskipun aplikasi ini masih memiliki beberapa kekurangan, dominasi ulasan positif menunjukkan bahwa Motorku X telah memberikan pengalaman pengguna yang cukup memuaskan. Oleh karena itu, disarankan agar pengembang atau pihak pengelola aplikasi terus meningkatkan kualitas sistem serta memberikan respons yang cepat dan efisien terhadap keluhan pengguna, guna mempertahankan dan meningkatkan kepuasan pelanggan di masa mendatang.

DAFTAR PUSTAKA

[1] E. B. Susanto, P. A. Christianto, M. R. Maulana, and S. W. Binabar, “Analisis Kinerja Algoritma Naïve Bayes Pada Dataset Sentimen Masyarakat

Aplikasi NEWSAKPOLE Samsat Jawa Tengah," 2022. doi: 10.37859/coscitech.v3i3.4343.

[2] A. R. Harungguan, H. Napitupulu, and F. Firdaniza, "Analisis Sentimen Dengan Metode Klasifikasi Naïve Bayes Dan Seleksi Fitur Chi-Square," 2023. doi: 10.37278/insearch.v22i2.762.

[3] A. I. Tanggraeni and M. N. Ngalumsine Sitokdana, "Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes," *Jatisi (Jurnal Teknik Informatika Dan Sistem Informasi)*, 2022, doi: 10.35957/jatisi.v9i2.1835.

[4] Gilbert, S. Alam, and M. I. Sulistyo, "Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi Mypertamina Pada Google Playstore Menggunakan Metode Naïve Bayes," *Storage Jurnal Ilmiah Teknik Dan Ilmu Komputer*, 2023, doi: 10.55123/storage.v2i3.2333.

[5] I. Syaripah, M. Martanto, and T. Suprapti, "Analisis Sentimen Pengguna Terhadap Aplikasi Binance Pada Ulasan Google Play Store Menggunakan Algoritma Naïve Bayes," *Jati (Jurnal Mahasiswa Teknik Informatika)*, 2024, doi: 10.36040/jati.v7i6.8289.

[6] D. Pratmanto, F. F. Dwi Imaniawan, and V. Maarif, "Analisis Sentimen Pada Ulasan Pengguna Aplikasi Identitas Kependudukan Digital Dengan Metode Naive Bayes Dan K-Nearest," *Computatio Journal of Computer Science and Information Systems*, 2023, doi: 10.24912/computatio.v7i2.26322.

[7] A. Septiani and I. Budi, "Klasifikasi Ulasan Pengguna Aplikasi: Studi Kasus Aplikasi Ipusnas Perpustakaan Nasional Republik Indonesia (PNRI)," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 7, no. 4, 2022, doi: 10.29100/jipi.v7i4.3216.

[8] N. Khoirunnisaa, K. Nabila Nastiti Kesuma, S. Setiawan, and A. Yunizar Pratama Yusuf, "KLASIFIKASI TEKS ULASAN APLIKASI NETFLIX PADA GOOGLE PLAY STORE MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN SVM," *SKANIKA: Sistem Komputer dan Teknik Informatika*, vol. 7, no. 1, pp. 64–73, Jan. 2024, doi: 10.36080/skanika.v7i1.3138.

[9] M. F. Y. Herjanto and C. Carudin, "ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI SIREKAP PADA PLAY STORE MENGGUNAKAN ALGORITMA RANDOM FOREST CLASSIFER," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 2, Apr. 2024, doi: 10.23960/jitet.v12i2.4192.

[10] A. Wi. Andi, Rohman Mujahidin, and Sasmito, "ANALISIS SENTIMEN MAHASISWA TERHADAP LAYANAN BIMBINGAN TUGAS AKHIR," *Jurnal Rekayasa Sistem Informasi dan Teknologi*, vol. 1, no. 2, pp. 216–222, Nov. 2023, doi: 10.59407/jrsit.v1i2.295.

[11] T. V. Meiyanti, M. Hatta, and A. Sevtiana, "Analisis Sentimen Mahasiswa Dengan Dosen Menggunakan Metode Naïve Bayes Classifier Pada Kuesioner Dosen," 2023. doi: 10.51920/jurminsi.v1i2.143.

[12] D. Pratmanto, R. Rousyati, F. F. Wati, A. E. Widodo, S. Suleman, and R. Wijianto, "App Review Sentiment Analysis Shopee Application in Google Play Store Using Naive Bayes Algorithm," in *Journal of Physics: Conference Series*, 2020. doi: 10.1088/1742-6596/1641/1/012043.

[13] E. Fauziningrum, M. Pd and M. P. Encis Indah Suryaningsih, S.T., "Evaluasi Dan Prediksi Penguasaan Bahasa Inggris Maritim Menggunakan Metode Decision Tree Dan Confusion Matrix (Studi Kasus Di Universitas Maritim Amni)," *Angewandte Chemie International Edition, 6(11), 951–952.*, 2021.

[14] S. Syafrizal, M. Afdal, and R. Novita, "Analisis Sentimen Ulasan Aplikasi PLN Mobile Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest Neighbor," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 1, pp. 10–19, Dec. 2023, doi: 10.57152/malcom.v4i1.983.