PREDIKSI SENTIMEN PENGGUNA SPOTIFY MENGGUNAKAN METODE NAÏVE BAYES STUDI KASUS ULASAN PENGGUNA DI PLAY STORE

Aisyah Fatihaturrahmah, Dhita Amanda Ardhani, Musdalifa Putri Casanova, Syifa Cahya Aulia, Yesya Najwa Widasari, Ken Ditha Tania, Winda Kurnia Sari
Sistem Informasi, Universitas Negeri Sriwijaya
Jalan Palembang-Prabumulih, KM 32 Inderalaya, Kabupaten Ogan Ilir, Sumatera Selatan (30662)
aisyahfrahmah@gmail.com

ABSTRAK

Spotify merupakan salah satu *platform* layanan *streaming music* digital yang banyak digunakan, tetapi ulasan pengguna menunjukkan variasi tingkat kepuasan terhadap layanan yang tersedia. Penelitian ini berfokus pada analisis dan prediksi sentimen pengguna *Spotify* berdasarkan ulasan yang diperoleh dari *Google Play Store* dengan menggunakan metode Naïve Bayes. Penelitian ini mencakup beberapa proses, yaitu pengumpulan data, pra-pemrosesan teks (tokenisasi, normalisasi, penghapusan kata tidak bermakna, dan *stemming*) serta ekstraksi fitur menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). Sentimen pengguna diklasifikasikan ke dalam dua kategori, yaitu positif dan negative. Evaluasi model dilakukan dengan menggunakan metrik akurasi, presisi *recall* dan F1-*score*. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen pengguna *Spotify* dengan akurasi sebesar 88,42%, *precision* 0,89, *recall* 0,74, dan F1-*Score* 0,79 (*macro average*). Selain itu, nilai *Area Under Curve* (AUC) sebesar 0,94 yang mengindikasikan bahwa model memiliki kemampuan klasifikasi yang sangat baik. Sehingga hasil ini menunjukkan bahwa pendekatan yang digunakan sudah sangat efektif dalam menganalisis sentimen pengguna dan dapat menjadi acuan bagi pengembang untuk meningkatkan kualitas terhadap layanan *streaming music Spotify*.

Kata kunci : *Analisis sentiment, Naïve Bayes, Spotify, Ulasan Pengguna, Google Play Store*

1. PENDAHULUAN

Di era transformasi *digital*, jumlah data yang dihasilkan terus bertambah secara signifikan setiap saat. Data ini berasal dari berbagai sumber, termasuk transaksi daring, media sosial, perangkat *Internet of Things (IoT)*, serta interaksi pengguna dengan platform digital. Informasi yang melimpah ini menjadi aset berharga bagi berbagai industri, khususnya di sektor teknologi dan hiburan. Namun, tanpa teknik analisis yang tepat, data tersebut hanya akan menjadi kumpulan informasi yang tidak terstruktur dan sulit dimanfaatkan secara efektif [1].

Salah satu pendekatan yang digunakan untuk mengatasi tantangan ini adalah *data mining*, yaitu teknik analisis data yang berfungsi untuk mengekstrak pola atau informasi berharga dari kumpulan data yang besar. Seiring dengan perkembangan teknologi *digital* dan semakin meningkatnya *volume data, data mining* menjadi semakin esensial dalam berbagai bidang, termasuk bisnis, kesehatan, keuangan, dan media sosial. Teknik ini memungkinkan perusahaan dan peneliti untuk mengidentifikasi pola tersembunyi, membuat prediksi, serta mendukung pengambilan keputusan berbasis data [2].

Salah satu penerapan utama dari *data mining* adalah analisis sentimen, yaitu proses mengidentifikasi dan mengkategorikan opini atau emosi yang terkandung dalam teks. Dengan menggunakan algoritma klasifikasi seperti Naïve Bayes, analisis sentimen dapat memberikan wawasan berharga mengenai opini publik terhadap suatu produk, layanan, atau isu tertentu. Teknik ini menjadi sangat relevan dalam memahami persepsi pengguna terhadap layanan digital seperti *Spotify* [3].

Metode Naïve Bayes memiliki beberapa keunggulan yang menjadikannya lebih unggul dibandingkan dengan metode klasifikasi lainnya, seperti *Support Vector Machine* (SVM) atau *Random Forest*. Berdasarkan Penelitian yang dilakukan oleh Muhammad rifqi fauzi dan Kemas muslim laksmana, Naïve Bayes mampu memberikan akurasi yang cukup tinggi serta efisiensi pemrosesan yang baik dibandingkan dengan metode SVM dalam analisis sentimen *Spotify*, yang menjadikan metode Naïve Bayes sebagai pilihan tepat untuk penelitian dengan data besar dan kebutuhan prediksi yang cepat [4].

Selain itu, menurut Devi Nazhifa penerapan Naïve Bayes yang dikombinasikan dengan pembobotan TF-IDF dapat menghasilkan akurasi hingga 93,56%, yang berarti lebih unggul dari metode pembobotan lainnya dalam klasifikasi teks yang kompleks [5].

Layanan streaming musik kini telah menjadi bagian penting dalam kehidupan sehari-hari, dengan *Spotify* sebagai salah satu platform yang paling populer. *Spotify* menyediakan jutaan lagu serta fitur personalisasi untuk meningkatkan pengalaman mendengarkan musik. Seiring dengan meningkatnya jumlah pengguna, opini mereka terhadap layanan ini semakin banyak tersebar di berbagai platform, termasuk media sosial dan toko aplikasi seperti *Google Play Store*. Ulasan pengguna menjadi sumber informasi yang sangat berguna bagi pengembang *Spotify* untuk memahami pengalaman pengguna dan meningkatkan kualitas layanan mereka [6].

Analisis sentimen merupakan salah satu metode dalam data mining yang digunakan untuk mengklasifikasikan opini pengguna ke dalam kategori positif, negatif, atau netral [7]. Salah satu pendekatan yang umum digunakan dalam analisis sentimen adalah Naïve Bayes, sebuah algoritma berbasis probabilistik yang efisien dalam mengolah teks dalam jumlah besar. Metode ini telah terbukti efektif dalam berbagai penelitian analisis sentimen karena kemampuannya dalam menangani data dengan akurasi yang cukup baik serta kecepatan pemrosesannya yang tinggi [8].

Penelitian ini bertujuan untuk menganalisis dan memprediksi sentimen pengguna terhadap aplikasi *Spotify* dengan menerapkan metode Naïve Bayes berdasarkan ulasan yang dikumpulkan dari *Google Play Store*. Dengan menganalisis pola sentiment pengguna, diharapkan penelitian ini dapat memberikan wawasan yang bermanfaat bagi pihak *Spotify* dalam mengidentifikasi aspek layanan yang perlu diperbaiki serta meningkatkan pengalaman pengguna secara keseluruhan.

2. TINJAUAN PUSTAKA

Analisis sentimen terhadap ulasan pengguna di *Google Play Store* menjadi fokus penelitian yang semakin berkembang dalam beberapa tahun terakhir. Metode Naïve Bayes banyak digunakan karena kemampuannya dalam klasifikasi teks yang sederhana namun efektif. Berbagai penelitian menunjukkan penerapan metode ini dalam berbagai aplikasi, termasuk layanan perbankan, streaming musik, hingga transportasi.

Penelitian oleh Ginabila dan Ahmad Fauzi membahas penerapan algoritma Naïve Bayes dan *Support Vector Machine* untuk menganalisis dan memprediksi sentimen dari ulasan pengguna terhadap aplikasi *Spotify* di *platform Play Store*. Metode ini dipilih karena kesederhanaannya dalam implementasi dan efektivitas dalam mengklasifikasikan teks berbasis probabilitas. Penelitian ini mengungapkan bahwa algoritma Naïve Bayes dapat mencapai akurasi yang cukup tinggi dalam memprediksi sentiment pengguna, sehingga dapat memberikan informasi bagi pengembang untuk mengingkatkan kualitas layanan aplikasi berdasarkan tanggapan pengguna [8].

Rahmawati meneliti sentimen pengguna terhadap aplikasi *m-BCA* menggunakan algoritma Naïve Bayes. Hasil penelitian mereka menunjukkan bahwa sebagian besar ulasan yang dianalisis bersentimen negatif. Model yang diterapkan berhasil mencapai akurasi sebesar 82%, dengan presisi 84%, *recall* 82%, dan *F1-score* 81%. Temuan ini mengonfirmasi bahwa Naïve Bayes memiliki performa yang cukup baik dalam mengklasifikasikan sentimen ulasan pada aplikasi perbankan [9].

Penelitian lain oleh Fadillah membandingkan kinerja algoritma Naïve Bayes dengan *Random Forest* dalam menganalisis sentimen pengguna terhadap aplikasi *Spotify*. Dengan *dataset* berisi 1.500 ulasan berbahasa Indonesia, penelitian ini menemukan bahwa Naïve Bayes lebih unggul dibandingkan *Random Forest*. Dalam skenario pembagian data 70:30, model Naïve Bayes menunjukkan akurasi 96%, dengan presisi, *recall*, dan *F1-score* masing-masing sebesar 95%. Hal ini menegaskan efektivitas Naïve Bayes dalam mengkategorikan sentimen pengguna pada aplikasi *streaming* musik [10].

Khotimah mengaplikasikan Naïve Bayes dalam analisis sentimen ulasan pengguna aplikasi Pintu yang tersedia di *Google Play Store*. Dalam penelitian ini, dilakukan pra-pemrosesan data serta penerapan teknik *oversampling SMOTE* untuk mengatasi ketidakseimbangan data. Model yang dihasilkan mencapai tingkat akurasi sebesar 95,07%. Namun, hasil analisis menunjukkan bahwa performa model pada kelas sentimen negatif masih tergolong rendah, dengan presisi 62% dan *recall* 58%. Ini menunjukkan adanya tantangan dalam mendeteksi sentimen negatif pada dataset yang tidak seimbang [11].

Husnina melakukan penelitian terhadap sentimen pengguna aplikasi *RedBus* menggunakan metode Naïve Bayes. Penelitian ini membandingkan dua metode pembobotan kata, yaitu TF-RF dan TF-IDF. Pembobotan TF-IDF memberikan hasil paling optimal dengan pengujian menggunakan metode *cross-validation* dengan k=10, yang menghasilkan rata-rata akurasi 93,56%. Hal ini menunjukkan bahwa pemilihan teknik pembobotan kata yang tepat dapat meningkatkan kinerja model dalam analisis sentimen berbasis Naïve Bayes [12].

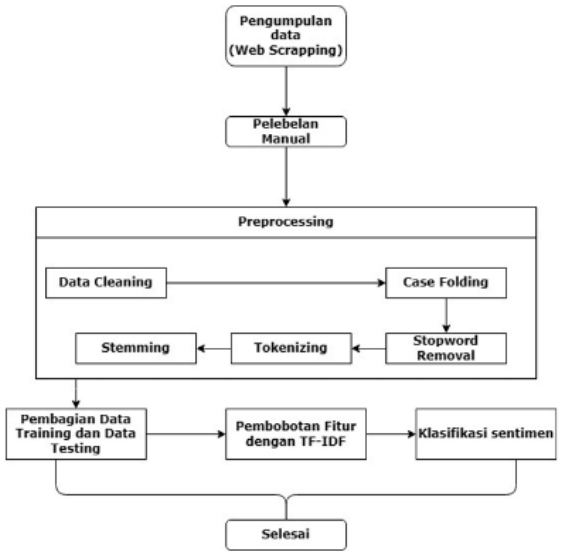
Selain pendekatan klasik seperti Naïve Bayes, berbagai penelitian terkini mulai mengeksplorasi model-model deep learning dalam analisis sentimen, khususnya pada ulasan aplikasi berbasis e-commerce dan media sosial. Novalia dalam penelitiannya yang dipublikasikan oleh IEEE, mengevaluasi efektivitas berbagai metode *word embedding* dalam meningkatkan performa model *sentiment analysis* terhadap ulasan pengguna TikTok Shop. Dengan menggabungkan arsitektur *Convolutional Neural Network (CNN)* dan *Long Short-Term Memory (LSTM)*, penelitian ini membandingkan embedding klasik seperti GloVe dengan embedding hasil pelatihan khusus menggunakan Word2Vec dan FastText pada data domain spesifik. Hasilnya menunjukkan bahwa Word2Vec yang dilatih secara kustom menghasilkan performa terbaik dalam hal akurasi, presisi, F1-score, dan AUC-ROC, diikuti oleh FastText, sementara GloVe menunjukkan performa yang relatif lebih rendah [13].

Berdasarkan studi literatur yang telah dibahas, terdapat beberapa gap penelitian yang mendasari urgensi penelitian ini. Sebagian besar penelitian terdahulu telah menggunakan metode Naïve Bayes untuk menganalisis sentimen ulasan pengguna aplikasi di berbagai bidang. Namun, kajian yang secara khusus membahas sentimen pengguna terhadap aplikasi *streaming* music terutama *Spotify* masih terbatas. Dengan demikian, penelitian ini bertujuan mengisi

kekosongan tersebut dengan menganalisis sentimen ulasan pengguna *Spotify* di *Google Play Store.*

Selain itu, meskipun Naïve Bayes telah terbukti efektif, beberapa penelitian mengindikasikan bahwa metode ini masih memiliki beberapa batasan dalam mengenali sentimen negatif secara optimal terutama pada dataset yang tidak seimbang. Seperti yang ditemukan oleh Khotimah, presisi dan *recall* untuk kelas negatif masih rendah sehingga penelitian ini akan mengeksplorasi strategi optimasi seperti teknik pembobotan kata TF-IDF dan seleksi fitur untuk meningkatkan akurasi model. Dengan demikian, penelitian ini berkontribusi pada pengembangan metode analisis sentimen yang lebih akurat dan relevan untuk aplikasi *streaming* musik yang selama ini masih jarang diteliti secara mendalam.

3. METODE PENELITIAN



Gambar 1. Tahapan Proses Penelitian

Proses penelitian dimulai dengan *data collection*, *pre-processing* data, *aspect-based sentiment analysis* (ABSA), data *labeling*, *feature extraction* (TF-IDF), split data, klasifikasi, dan evaluasi.

3.1. *Data Collection*

Penelitian ini menggunakan pendekatan data mining dalam analisis sentimen pengguna aplikasi *Spotify* berdasarkan ulasan yang diperoleh dari *Google Play Store.* Metode klasifikasi yang digunakan adalah Naïve Bayes, sebuah algoritma berbasis probabilistik yang telah terbukti efektif dalam pengolahan teks dan analisis sentimen.

3.2. *Aspect Based Sentiment Analysis (ABSA)*

Penelitian ini menggunakan data yang diperoleh dari ulasan pengguna *Spotify* di *Google Play Store.* Pengambilan data dilakukan dalam periode tertentu untuk mendapatkan informasi yang lebih relevan mengenai pengalaman pengguna.

3.3. *Data Labeling*

Pemberian label dalam analisis sentimen adalah proses mengklasifikasikan sentimen dalam teks berdasarkan kategori tertentu. Pada penelitian ini, proses pelabelan bertujuan menentukan apakah sebuah teks memiliki sentimen positif atau negatif.

3.4. *Preprocessing Data*

Preprocessing data merupakan langkah penting dalam *Natural Language Processing (NLP)* yang bertujuan merubah teks mentah ke dalam format yang lebih terorganisir dan mudah dianalisis. Pada tahap ini dilakukan beberapa proses seperti *cleaning, case folding, normalization, tokenization, stopword removal,* dan *stemming*.

a. *Cleaning*
   Proses pembersihan data dimulai dengan memilih kolom yang akan dianalisis, kemudian dilanjutkan dengan menghapus data duplikat, menangani nilai yang hilang, membersihkan simbol khusus, menghapus angka yang tidak relevan, merapikan spasi yang berlebih, serta menghilangkan baris yang tidak mengandung informasi yang diperlukan. Setelah dilakukan proses *cleaning*, data yang akan dianalisis menjadi 1.000 ulasan.
b. *Case folding*
   Case folding adalah proses standarisasi teks dengan mengubah semua huruf menjadi huruf kecil atau huruf besar. Pada penelitian ini, semua huruf dalam ulasan akan diubah menjadi huruf kecil.
c. *Normalization*
   Normalization adalah proses memperbaiki ejaan kata yang tidak baku atau disingkat agar sesuai dengan aturan bahasa standar. Salah satu contohnya yaitu mengubah kata “udh” menjadi “sudah”.
d. *Tokenization*
   Tokenization adalah proses membagi teks menjadi unit-unit kecil yang disebut token.
e. *Stopword Removal*
   Stopword removal adalah proses menghapus kata-kata yang tidak bermakna dalam teks. Contoh kata-katanya seperti, “atau”, “yang”, dan “di”.
f. *Stemming*
   Stemming adalah teknik dalam pemrosesan bahasa alami yang mengubah kata menjadi bentuk dasarnya dengan menghilangkan imbuhan seperti awalan dan akhiran.

3.5. *Feature Extraction (TF-IDF)*

Metode TF-IDF diterapkan pada penelitian ini bertujuan untuk mengevaluasi signifikansi kata-kata dalam kumpulan dokumen. Metode ini menghitung relevansi suatu kata dengan mempertimbangkan dua faktor, yaitu frekuensi kemunculan kata dalam dokumen tertentu (*Term Frequency atau TF*) dan frekuensi kemunculan kata tersebut di seluruh kumpulan dokumen (*Document Frequency* atau DF). Melalui perhitungan *Inverse Document Frequency* (IDF), kata-kata yang terlalu sering muncul di banyak

dokumen diberi bobot yang lebih rendah, sehingga fokus analisis dapat diarahkan pada kata-kata yang lebih spesifik dan informatif.

3.6. *Split Data*

Dalam upaya memvalidasi performa model pembelajaran mesin yang dikembangkan, dataset dibagi menjadi dua, yaitu data pelatihan dan data pengujian. Alokasi data sebesar 80% untukdata pelatihan dan 20% untuk data pengujian diterapkan mengacu pada praktik standar yang umumnya menghasilkan evaluasi yang lebih baik terhadap kemampuan generalisasi model.

3.7. Model Klasifikasi

a. Algoritma Naive Bayes

$$P(x \mid y) = \frac{P(y|x) \cdot P(x)}{P(y)} \quad (1)$$

y : Data yang belum diketahui.

x : Hipotesis data y merupakan suatu class spesifik.

P(x|y) : Probabilitas hipotesis y berdasarkan kondisi y (*Posterior Probability*).

P(H) : Probabilitas hipotesis x (*Prior Probability*).

P(y|x) : Probabilitas y berdasarkan Hipotesis x.

P(y) : Probabilitas y.

3.8. Model Evaluasi

a. Evaluasi Kinerja

Evaluasi kinerja digunakan Untuk mengetahui seberapa baik model mengklasifikasikan data. Akurasi, presisi, *recall*, dan F1-*score* adalah beberapa metrik evaluasi yang digunakan dalam penelitian ini.

b. Akurasi

Untuk menentukan proporsi jumlah prediksi yang benar terhadap keseluruhan data, akurasi digunakan. Akurasi dihitung menggunakan Persamaan berikut:

$$\text{Akurasi} = \frac{Jumlah\ prediksi\ benar}{Total\ jumlah\ data} \quad (2)$$

c. Presisi

Presisi adalah rasio antara seluruh data yang diprediksi sebagai positif dan prediksi yang benar. Presisi dihitung menggunakan Persamaan berikut:

$$\text{Presisi} = \frac{TP}{TP + FP} \quad (3)$$

d. *Recall*

Recall menunjukkan kemampuan model untuk menemukan data positif secara akurat. Metrik ini penting disaat biaya dari *false negative* lebih besar. *Recall* dihitung menggunakan Persamaan berikut:

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

e. F1-*Score*

F1-*Score* adalah rata-rata harmonis dari presisi dan *recall*, yang digunakan ketika data tidak seimbang (*imbalanced data*). F1-*Score* dihitung menggunakan Persamaan berikut:

$$\text{F1-}Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (5)$$

Keterangan:

- TP (*True Positive*): Data positif yang diprediksi positif oleh model.
- FP (*False Positive*): Data negatif yang diprediksi sebagai positif.
- FN (*False Negative*): Data positif yang diprediksi sebagai negatif.

4. HASIL DAN PEMBAHASAN

4.1. *Data Collection*

Proses pengumpulan data dilakukan melalui Teknik *web scrapping* menggunakan Bahasa pemrograman yaitu *Python* pada platform *Google Collaboratory*. Sebanyak 5.000 data ulasan berhasil dikumpulkan dari bulan Januari hingga Februari 2025. Berikut hasil web scraping menggunakan *google colabolatory*.

reviewId	userName	userImage	content	score
9419d358-	Pengguna	https://pl	maksa banget ke premium blm lagi iklan ny di setiap lagu baru.. stelah di pe	1
3a691651-	Pengguna	https://pl	Baik tampa biyaya yg besar	5
633efc17-	Pengguna	https://pl	Bagus sekali	5
c1deffba-9	Pengguna	https://pl	Lagu nya selalu mati kalo lagi buka aplikasi lain	2
00f34385-	Pengguna	https://pl	Kurang memuaskan !!!	4
2db9bc3a-	Pengguna	https://pl	Mantap	5
0d08d7e3-	Pengguna	https://pl	Maksain banget buat ke premium, iklannya banyak banget, mana iklannya 3	1
6a67e03c-	Pengguna	https://pl	Btw,,,,,udah itu aja	5
d6bb72f0-	Pengguna	https://pl	Bagus	5
e28eb8a7-	Pengguna	https://pl	,	1
99588175-	Pengguna	https://pl	Bagus	5
efdf736d-1	Pengguna	https://pl	Dulu Spotify family yang 84k ada 6 orang, sekarang ada 4. Pelit, jahat lu.	1
160ecf60-1	Pengguna	https://pl	Lancar untuk menemani kerjaku tiap hari Terima kasih Spotify	5
e4c90bcb-	Pengguna	https://pl	Sangat baik	1
3c366293-	Pengguna	https://pl	Bisa mendenar lagu bagus	5
3bcf98e3-	Pengguna	https://pl	nice	5
973339c1-	Pengguna	https://pl	Top	5
04bb7f61-	Pengguna	https://pl	Bagus	5
801f0f97-1	Pengguna	https://pl	Suka bgt	5
08e0ed69-	Pengguna	https://pl	Bagus	5

Gambar 2. Hasil Web Scrapping

4.2. *Aspect Based Sentiment Analysis (ABSA)*

Berdasarkan hasil analisis sentiment terhadap ulasan pengguna *Spotify* di *Google Play Store*, sentiment diklasifikasikan ke dalam dua kategori utama, yaitu positif dan nnegatif. Guna memperdalam analisis, penelitian ini juga menerapkan *Aspect-Based Sentiment Analysis* (ABSA), yang memungkinkan identifikasi sentimen berdasarkan aspek-aspek spesifik dari layanan *Spotify*.

Berikut adalah hasil analisis berdasarkan aspek yang diidentifikasi :

Tabel 1. Hasil Analisis Berdasarkan Aspek

Aspek	Sentimen Positif	Sentimen Netral	Sentimen Negatif
Kinerja Aplikasi	30%	20%	50%
Fitur	45%	30%	25%
Antarmuka (UI/UX)	50%	20%	30%
Stabilitas Aplikasi	25%	30%	45%
Komparasi dengan Kompetitor	20%	25%	55%
Kepuasan Pengguna	40%	30%	30%

Dari tabel di atas, dapat disimpulkan bahwa aspek fitur dan antarmuka mendapatkan lebih banyak sentimen positif, menunjukkan bahwa pengguna cukup puas dengan elemen-elemen ini. Sebaliknya, aspek stabilitas aplikasi dan kinerja cenderung mendapat sentimen negatif lebih tinggi, mengindikasikan bahwa pengguna sering mengalami masalah teknis saat menggunakan *Spotify*.

4.3. *Data Labelling*

Labelling dalam penelitian ini bertujuan untuk mengklasifikasikan sentimen pengguna *Spotify* berdasarkan aspek-aspek tertentu. Setiap ulasan yang dikumpulkan diberi label positif, netral, atau negatif, tergantung pada opini yang terkandung di dalamnya.

Tabel 2. Contoh Pelabelan

Ulasan					
banyak bug nya,play lagu kadang seret ga bisa dipencet,ditambah lagi buka *podcast* mau pencet tombol back malah gak bisa,mending *apple music* dah daripada *spotify* mah					
Aspek					
Kinerja	Fitur	Antar-muka	Stabilitas	Komparasi	Kepuasan
Negative	Negative	Negative	Negative	Negative	Negative

4.4. *Preprocessing Data*

Pre-processing merupakan tahap penting dalam *Natural Language Processing* (NLP) yang bertujuan untuk membersihkan dan menormalisasi teks sebelum dianalisis. Pada penelitian ini, *pre-processing* diterapkan untuk mengubah ulasan mentah dari *Play Store* ke dalam format yang lebih terstruktur sehingga dapat diolah oleh model Naïve Bayes.

Tabel 3. Hasil *Pre-Processing*

Proses	**Hasil**
Cleaning	Banyak bug, memutar lagu kadang lambat dan tidak bisa ditekan. Saat membuka podcast, tombol kembali tidak berfungsi.
Case folding	aplikasi ini memiliki banyak bug. pemutaran lagu kadang lambat dan tidak bisa ditekan. selain itu, saat membuka podcast, tombol kembali tidak berfungsi.
Normalization	aplikasi ini memiliki banyak kesalahan. pemutaran lagu kadang lambat dan tidak dapat ditekan. selain itu, saat membuka podcast, tombol kembali tidak berfungsi.
Tokenization	["aplikasi", "ini", "memiliki", "banyak", "kesalahan", ".", "pemutaran", "lagu", "kadang", "lambat", "dan", "tidak", "dapat", "ditekan", ".", "selain", "itu", ",", "saat", "membuka", "podcast", ",", "tombol", "kembali", "tidak", "berfungsi", "."]
Stopword Removal	["aplikasi", "memiliki", "banyak", "kesalahan", ".", "pemutaran", "lagu", "kadang", "lambat", "tidak", "dapat", "ditekan", ".", "membuka", "podcast", ",", "tombol", "kembali", "tidak", "berfungsi", "."]
Stemming	["aplikasi", "miliki", "banyak", "salah", ".", "putar", "lagu", "kadang", "lambat", "tidak", "dapat", "tekan", ".", "buka", "podcast", ",", "tombol", "kembali", "tidak", "fungsi", "."]

4.5. *Feature Extraction (TF-IDF)*

4.5.1 *Term Frequency (TF)*

Term Frequency (TF) menghitung seberapa sering suatu kata muncul dalam satu dokumen dibandingkan dengan total kata dalam dokumen tersebut.

$$tf(t,d) = \frac{ft,d}{Nd} \quad (6)$$

Penerapan *Term Frequency (TF)* dalam penelitian ini bertujuan untuk menghitung bobot setiap kata dalam dokumen berdasarkan jumlah kemunculannya relatif terhadap total kata dalam dokumen tersebut. Setelah dilakukan ekstraksi fitur menggunakan metode TF-IDF, model klasifikasi yang diterapkan menunjukkan akurasi sebesar 88,42%.

Berdasarkan laporan klasifikasi, model memiliki *precision* sebesar 0,91 untuk kelas 0 dan 0,88 untuk kelas 1. Namun, model mengalami kesulitan dalam mengidentifikasi kelas 0, sebagaimana ditunjukkan oleh nilai *recall* yang lebih rendah, yaitu 0,50, dibandingkan dengan kelas 1 yang memiliki *recall* 0,99. Hal ini menunjukkan bahwa model lebih sering salah mengklasifikasikan kelas 0 sebagai kelas 1, yang terlihat dari matriks kebingungan:

- 99 sampel kelas 0 diklasifikasikan dengan benar, tetapi 100 sampel kelas 0 salah diklasifikasikan sebagai kelas 1.
- 741 sampel kelas 1 diklasifikasikan dengan benar, sementara 10 sampel kelas 1 salah diklasifikasikan sebagai kelas 0.

Rata-rata makro (*macro average*) dari model ini menunjukkan *precision* sebesar 0,89, *recall* 0,74, dan *f1-score* 0,79, sedangkan rata-rata berbobot (*weighted average*) memiliki nilai *precision* 0,89, *recall* 0,88, dan *f1-score* 0,87. Perbedaan nilai *recall* yang cukup signifikan antara kelas 0 dan 1 menunjukkan bahwa distribusi data dalam kelas tidak seimbang, di mana kelas 1 memiliki jumlah sampel yang lebih banyak dibandingkan kelas 0.

Hasil ini menunjukan bahwa meskipun TF membantu dalam mengekstraksi fitur dari teks, metode ini masih memiliki keterbatasan dalam menangani ketidakseimbangan data. Oleh karena itu, pendekatan tambahan seperti penyeimbangan data melalui Teknik (*oversampling* atau *undersampling*), penerapkan teknik pembobotan lainnya, atau eksplorasi fitur tambahan dapat menjadi solusi untuk meningkatkan performa model dalam mengidentifikasi kelas dengan jumlah sampel yang lebih kecil.

4.5.2 *Inverse Document Frequency (IDF)*

Inverse Document Frequency (IDF) merupakan komponen dalam metode TF-IDF yang berfungsi untuk mengurangi bobot kata-kata yang sering muncul dalam banyak dokumen dan meningkatkan bobot kata-kata yang jarang muncul. IDF dihitung menggunakan rumus berikut:

$$tf - IDF(t,d) = TF(t,d)\ x\ IDF(t) \quad (7)$$

$$IDF(t) = log\ \frac{N}{DFt} \quad (8)$$

di mana:

- NNN adalah jumlah total dokumen dalam korpus,
- dftdf_tdft adalah jumlah dokumen yang mengandung kata ttt.

Dalam penelitian ini, penggunaan metode IDF bertujuan untuk memperkuat fitur kata yang lebih informatif dalam proses klasifikasi. Setelah diterapkan bersama dengan *Term Frequency* (TF), model menghasilkan akurasi 88,42%.

Berdasarkan laporan klasifikasi, kelas 0 memiliki nilai *precision* sebesar 0,91, namun *recall* yang lebih rendah (0,50), yang menunjukkan adanya kesulitan model mengalami kesulitan dalam mengenali kelas ini. Sebaliknya, kelas 1 menunjukkan *recall* yang sangat tinggi (0,99) dengan *precision* 0,88, mengindikasikan bahwa sebagian besar data kelas 1 diklasifikasikan dengan benar.

Matriks kebingungan menunjukkan bahwa dari 199 sampel kelas 0, sebanyak 99 diklasifikasikan dengan benar, sementara 100 diklasifikasikan sebagai kelas 1. Di sisi lain, dari 751 sampel kelas 1, sebanyak 741 diklasifikasikan dengan benar, dan 10 salah diklasifikasikan sebagai kelas 0.

Rata-rata makro dari model menunjukkan *precision* 0,89, *recall* 0,74, dan *f1-score* 0,79, sedangkan rata-rata berbobot memiliki nilai *precision* 0,89, *recall* 0,88, dan *f1-score* 0,87. Rendahnya nilai *recall* yang rendah pada kelas 0 mengindikasikan bahwa model cenderung keliru dalam mengklasifikasikan sampel dari kelas tersebut, yang bisa disebabkan oleh ketidakseimbangan distribusi data.

4.5.3 TF-IDF Calculation

Setelah menerapkan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), model klasifikasi yang digunakan dalam penelitian ini mencapai akurasi 88,42%. TF-IDF dihitung dengan mengombinasikan *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) menggunakan rumus berikut:

di mana:

- TF(t,d)TF(t,d)TF(t,d) adalah jumlah kemunculan kata ttt dalam dokumen ddd, dinormalisasi terhadap panjang dokumen.
- IDF(t)IDF(t)IDF(t) adalah logaritma dari rasio jumlah total dokumen dengan jumlah dokumen yang mengandung kata ttt, yang digunakan untuk mengurangi bobot kata yang sering muncul di banyak dokumen.

Setelah fitur diekstraksi menggunakan TF-IDF, model yang digunakan dalam klasifikasi menunjukkan performa Akurasi Model 88,42%

Tabel 4. *TF-IDF Calculation*

Kelas	Precision	Recall	F1-Score	Support
0	0.91	0.50	0.64	199
1	0.88	0.99	0.93	751

Berdasarkan hasil evaluasi model yang ditampilkan dalam Tabel 4, untuk kelas 0 (sentimen negatif), model memiliki nilai *precision* sebesar 0,91, yang berarti sebagian besar prediksi sentimen negatif adalah benar. Namun, nilai *recall* untuk kelas ini hanya sebesar 0,50, menunjukkan bahwa model hanya mampu mendeteksi setengah dari seluruh ulasan negatif yang ada. Hal ini menandakan bahwa model masih belum cukup sensitif dalam mengidentifikasi keluhan atau kritik dari pengguna. Nilai *F1-score* untuk kelas ini sebesar 0,64, memperkuat indikasi bahwa performa model terhadap ulasan negatif masih perlu ditingkatkan.

Sebaliknya, untuk kelas 1 (sentimen positif), nilai *recall* sangat tinggi, yaitu 0,99, menunjukkan bahwa hampir semua ulasan positif berhasil diklasifikasikan dengan benar. Nilai *precision* untuk kelas ini adalah 0,88, sedangkan *F1-score*-nya adalah 0,93. Kombinasi nilai-nilai ini mengindikasikan bahwa model sangat baik dalam mendeteksi sentimen positif dan hanya melakukan sedikit kesalahan prediksi terhadap ulasan berlabel positif.

4.6. *Split Data*

Pada tahap awal, sebanyak 5.000 data ulasan pengguna *Spotify* dari *Google Play Store* berhasil dikumpulkan melalui proses *Web Scrapping.* Namun, setelah dilakukan proses pembersihan dan pemrosesan data, hanya 1.000 data yang berhasil diolah dan digunakan dalam penelitian ini. Setelah dilakukan *split* data dengan rasio 80:20, hasilnya adalah sebagai berikut:

Dataset yang digunakan dalam penelitian ini terdiri dari 1.000 ulasan pengguna *Spotify* dari *Google Play Store*. Setelah dilakukan split data dengan rasio 80:20, hasilnya sebagai berikut :

Tabel 5. *Split Data*

Kategori	Jumlah Data	Persentase (%)
Training set	800	80%
Testing set	200	20%
Total	1000	100%

Setelah didapatkan model dari data training selanjutnya dilakukan pengklasifikasian pada data testing dengan algoritma Naïve Bayes untuk mengetahui hasil probabilitas sentimen.

4.7. Model Klasifikasi

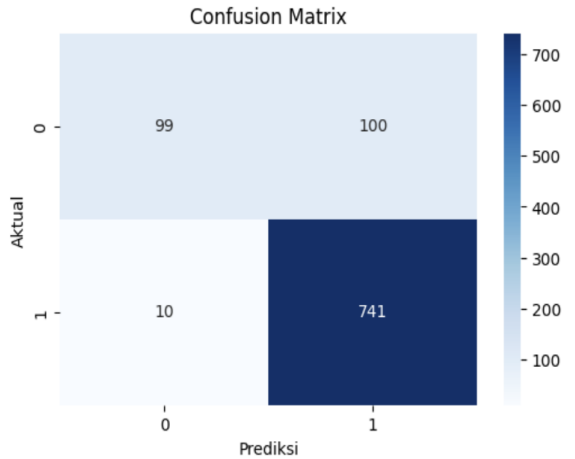
Model klasifikasi yang diterapkan dalam penelitian ini adalah Naïve Bayes, sebuah algoritma berbasis probabilistik yang umum digunakan dalam analisis sentimen. Algoritma ini bekerja dengan menghitung probabilitas suatu ulasan termasuk dalam kategori sentimen tertentu, berdasarkan asumsi bahwa setiap kata dalam teks bersifat independen. Naïve Bayes menghitung probabilitas *posterior* menggunakan rumus:

$$P(x|y) = \frac{P(x).P(x)}{P(y)} \quad (9)$$

Proses klasifikasi dimulai dengan *pre-processing* data, yang mencakup pembersihan teks, penghapusan kata-kata yang tidak relevan (*stopword)*, dan stemming untuk memastikan bahwa data dalam bentuk yang optimal sebelum diproses lebih lanjut. Setelah itu, dilakukan ekstraksi fitur menggunakan metode *Term Frequency - Inverse Document Frequency* (TF-IDF) untuk mengubah teks menjadi representasi numerik. Dataset yang telah diproses kemudian dibagi menjadi dua bagian, yaitu 80% data untuk pelatihan dan 20% data untuk pengujian, menggunakan metode *train-test split*. Model Naïve Bayes dilatih menggunakan data pelatihan dan kemudian diuji dengan data pengujian untuk mengukur performanya dalam mengklasifikasikan sentimen pengguna.

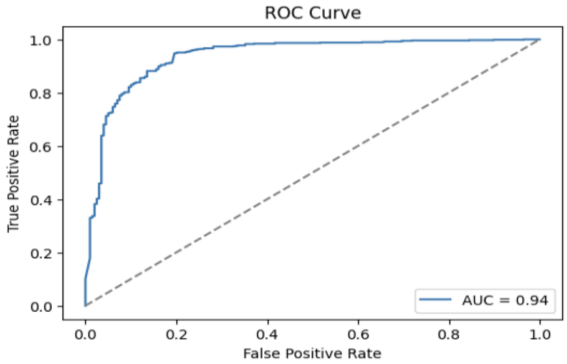
Model klasifikasi yang digunakan dalam penelitian ini adalah algoritma Naïve Bayes. Berdasarkan hasil pengujian terhadap data testing yang telah dipisahkan sebelumnya (rasio 80:20), model menunjukkan performa yang cukup baik dengan tingkat akurasi sebesar 88,42%. Nilai ini menunjukkan bahwa secara umum model mampu mengklasifikasikan data ulasan pengguna Spotify dengan baik ke dalam kategori sentimen positif dan negatif. Namun demikian, mengingat distribusi data yang tidak seimbang (mayoritas sentimen positif), akurasi saja belum cukup mewakili kualitas model secara keseluruhan. Oleh karena itu, dilakukan analisis lanjutan menggunakan metrik evaluasi lainnya, yaitu *precision*, *recall*, dan *F1-score*.

4.8. Model Evaluasi



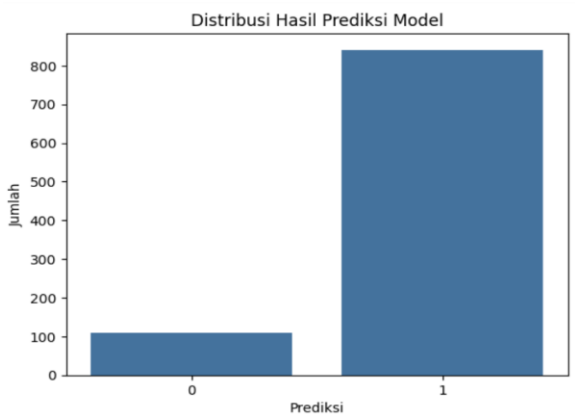
Gambar 3. *Confusion Matrix*

Gambar *confusion matrix* di atas menunjukkan performa model klasifikasi dalam memprediksi dua kelas, yaitu kelas 0 dan kelas 1. Dari hasil evaluasi, model berhasil mengklasifikasikan 99 sampel kelas 0 dengan benar (*True Negative*) dan 741 sampel kelas 1 dengan benar (*True Positive*). Namun, terdapat 100 sampel kelas 0 yang salah diprediksi sebagai kelas 1 (*False Positive*) dan 10 sampel kelas 1 yang salah diklasifikasikan sebagai kelas 0 (*False Negative*). Nilai *True Positive* yang tinggi (741) menunjukkan bahwa model sangat baik dalam mengenali kelas 1, namun nilai *False Positive* yang cukup besar (100) menunjukkan bahwa masih terdapat banyak sampel kelas 0 yang salah diklasifikasikan. Hal ini dapat mengindikasikan adanya ketidakseimbangan data atau kemiripan fitur antar kelas yang menyebabkan kesulitan dalam membedakan kedua kelas secara optimal. Untuk meningkatkan akurasi klasifikasi, beberapa pendekatan dapat diterapkan, seperti penyesuaian *threshold* keputusan, penanganan ketidakseimbangan data, atau penggunaan model yang lebih kompleks guna meningkatkan performa dalam mengenali kelas 0 dengan lebih baik.



Gambar 4. Kurva ROC

Dari grafik tersebut, terlihat bahwa kurva ROC berada jauh di atas garis diagonal, yang menunjukkan bahwa model memiliki performa yang baik dalam membedakan kelas positif dan negatif. Nilai *Area Under the Curve* (AUC) sebesar 0.94 mengindikasikan bahwa model memiliki tingkat diskriminasi yang sangat baik. Semakin mendekati nilai 1, semakin baik model dalam membedakan antara kelas positif dan negatif. Dengan AUC sebesar 0.94, model dapat dikatakan memiliki kemampuan klasifikasi yang sangat baik dan dapat diandalkan dalam pengambilan keputusan berdasarkan prediksi yang dihasilkan. Dengan demikian, meskipun model masih memiliki kelemahan dalam mendeteksi sentimen negatif akibat ketidakseimbangan data, secara keseluruhan performanya tergolong sangat baik dan dapat diandalkan.



Gambar 5. Prediksi Model Klasifikasi

Gambar di atas menunjukkan distribusi hasil prediksi model klasifikasi. Grafik ini menggambarkan jumlah sampel yang diklasifikasikan ke dalam masing-masing kelas, yaitu kelas 0 dan kelas 1. Dari hasil visualisasi, terlihat bahwa model lebih banyak memprediksi data sebagai kelas 1, dengan jumlah lebih dari 800 sampel, sedangkan kelas 0 hanya sekitar 100 sampel.

Distribusi hasil prediksi menunjukkan bahwa model cenderung lebih sering mengklasifikasikan ulasan sebagai sentimen positif. Hal ini juga sejalan dengan analisis sebelumnya yang menunjukkan bahwa sebagian besar data ulasan memang berasal dari pengguna yang memberikan tanggapan positif terhadap Spotify. Berdasarkan pola ini, dapat disimpulkan bahwa secara umum, sentimen pengguna terhadap aplikasi Spotify cenderung positif. Fitur-fitur aplikasi dan antarmuka pengguna (UI/UX) mendapatkan apresiasi dari banyak pengguna. Namun demikian, aspek seperti stabilitas aplikasi dan performa teknis masih menjadi titik keluhan, sebagaimana tercermin dalam ulasan dengan sentimen negatif yang terdeteksi, walaupun jumlahnya lebih sedikit.

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan untuk menganalisis sentimen pengguna terhadap aplikasi *Spotify* berdasarkan ulasan yang diperoleh dari *Google Play Store* menggunakan metode *Naïve Bayes.* Tahapan-tahapan penelitian terdiri dari pengumpulan data, *pre-processing*, seperti pembersihan teks, tokenisasi, penghapusan *stopward*, dan *stemming*, serta fitur yang diekstraksi dengan metode *Term Frequency-inverse Document* (TF-IDF). Kemudian, data dibagi menjadi dua bagian, yaitu 80% untuk pelatihan dan 20% untuk pengujian. Hasil dari evaluasi tersebut menunjukkan bahwa, dengan model klasifikasi Naïve Bayes mampu mengklasifikasikan sentimen yang cukup baik, dengan nilai akurasi sebesar 88,42%, *precision* rata-rata 0,89, *recall* rata-rata 0,74, F1-*score* rata-rata 0,79, dan nilai AUC sebesar 0,94. Mayoritas ulasan pengguna tergolong dalam kategori sentimen positif, khususnya pada aspek fitur dan antarmuka pada aplikasi, sedangkan sentimen negatif ditemukan pada aspek stabilitas dan kinerja teknis. Dengan demikian, algoritma Naïve Bayes terbukti efektif dalam melakukan prediksi sentimen pengguna terhadap *Spotify*. Penelitian ini diharapkan dapat menjadi referensi bagi pengembang aplikasi untuk meningkatkan kualitas layanan, dan disarankan untuk penelitian selanjutnya agar menggunakan dataset yang lebih besar, serta menerapkan teknik lanjutan seperti *word embedding* atau *transformer-based models* untuk meningkatkan sensitivitas terhadap sentimen negatif dan performa klasifikasi secara keseluruhan.

DAFTAR PUSTAKA

[1] Ardhani. R and et al, "Analisis Sentimen Terhadap Layanan Aplikasi Grab Indonesia Menggunakan Metode Naïve Bayes," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 8, no. 1, pp. 303–309, 2024.

[2] M. K. K. Insan, Hayati U, and Nurdiawan, "Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informasi)*, vol. 7, no. 1, pp. 478–483, 2023.

[3] Ramdhani M and Lhaksmana K.M, "Analisis Sentimen Menggunakan Metode Naïve Bayes dan Support Vector Machine Pada Ulasan Aplikasi Spotify," *Universitas Telkom*, 2023.

[4] M. Rifqi and F. Ramdhani, "Analisis Sentimen Menggunakan Metode Naive Bayes dan Support Vector Machine pada Ulasan Aplikasi Spotify," 2023.

[5] D. Nazhifa, N. Husnina, D. E. Ratnawati, and B. Rahayudi, "Analisis Sentimen Pengguna Aplikasi RedBus berdasarkan Ulasan di Google Play Store menggunakan Metode Naïve Bayes," 2023. [Online]. Available: http://j-ptiik.ub.ac.id

[6] Nugraha M.T, Sulistyowati N.M, and Enri U.U, " Analisis Sentimen Ulasan Aplikasi Satu Sehat Pada Google Play Store Menggunakan Naïve Bayes Classifier," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 5, pp. 3593–3601, 2023.

[7] Widodo B.K and Prasvita D.S, "Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Penggunaan Aplikasi Jobstreet," *Techno*, vol. 21, no. 3, 2022.

[8] Ginabila and Ahmad Fauzi, "Analisis Sentimen Terhadap Pemutar Musik Online Spotify Dengan Algoritma Naive Bayes dan Support Vector Machine," *Jurnal Ilmu Komputer dan Informatika*, vol. 6, no. 2, pp. 111–122, Jul. 2023.

[9] Rahmawati A, Sari D.P, and Nugroho R, "Analisis sentimen pengguna aplikasi m-BCA menggunakan algoritma Naïve Bayes," *Jurnal Riset Teknologi Informasi*, vol. 10, no. 2, pp. 45–56, 2023.

[10] Fadhilah R, "Perbandingan algoritma Random Forest dan Naïve Bayes dalam analisis sentimen

ulasan Spotify," *Jurnal Sistem dan Informatika*, vol. 15, no. 1, pp. 78–90, 2024.

[11] Khotimah S, Putra H, and Yulianti, " Penerapan Naïve Bayes untuk analisis sentimen ulasan aplikasi Pintu di Google Play Store," *Jurnal Teknologi Informasi dan Elektronika*, vol. 9, no. 3, pp. 112–125, 2023.

[12] R. Husnina and et al, "Analisis sentimen pengguna aplikasi RedBus dengan metode Naïve Bayes dan pembobotan TF-IDF," *Jurnal Pengolahan Data dan Kecerdasan Buatan*, vol. 11, no. 4, pp. 98–110, 2023.

[13] Vanessa Novalia, Ken Ditha Tania, Allsela Meiriza, and Ari Wedhasmara, "Knowledge Discovery of Application Review Using Word Embedding's Comparison with CNN-LSTM Model on Sentiment Analysis," *Conference: 2024 International Conference on Electrical Engineering and Computer Science (ICECOS)*, 2024.