

ANALISIS SENTIMEN TERHADAP REVIEW APLIKASI BRIMO DI *GOOGLE PLAY STORE* MENGGUNAKAN ALGORITMA *NAÏVE BAYES*

Asyifa Nuraini, Ahmad Faqih, Gifthera Dwilestari, Nisa Dienwati Nuris, Riri Narasati

Teknik Informatika STMIK Ikmi Cirebon

JL. Perjuangan No 10B-Majasem, Kesambi Karya Mulya, Kota Cirebon, Jawa Barat Indonesia

Asyifanuraini730@gmail.com

ABSTRAK

Dalam era perkembangan teknologi informasi, khususnya dalam ranah analisis sentimen menggunakan algoritma *Naive Bayes*, dampak signifikan terlihat pada aplikasi Brimo, platform perbankan internet dan *mobile* yang dikembangkan oleh Bank BRI. Analisis sentimen terhadap ulasan pengguna di *Google Play Store* dengan *Naive Bayes* menghadapi tantangan, seperti tingkat ketidakpastian dalam hasil analisis. Algoritma ini cenderung memperlakukan kata-kata secara terpisah, mengabaikan konteks, dan berpotensi menghasilkan hasil yang kurang akurat, terutama pada sentimen kompleks seperti ironi atau sarkasme. Pentingnya text mining, penggunaan metode klasifikasi, dan proses *Knowledge Discovery in Database (KDD)* muncul dalam mendukung analisis sentimen. Penelitian ini menggunakan metode *KDD* dan *Naive Bayes* untuk menganalisis sentimen ulasan pengguna Brimo di *Google Play Store*. Tahapan melibatkan seleksi data, pembersihan data, transformasi dengan TF-IDF, data mining dengan *Naive Bayes*, dan evaluasi pola. Hasil menunjukkan tingkat akurasi tinggi sebesar 89.58%, dengan nilai presisi 85.94%, *Recall* 89.58%, dan F1-Score 86.90%. Meskipun demikian, perhatian khusus diperlukan untuk sentimen negatif guna meningkatkan performa *Recall*. Proses implementasi model *Naive Bayes* pada data baru menunjukkan efisiensi melalui penggunaan *joblib* dan visualisasi hasil prediksi. Penelitian ini memberikan kontribusi dalam meningkatkan pemahaman evaluasi sentimen pengguna, mendukung pengambilan keputusan, dan mengembangkan strategi bisnis yang efektif pada aplikasi Brimo.

Kata kunci: Analisis sentimen, *Google Play*, Brimo, *Naive bayes*

1. PENDAHULUAN

Kemajuan yang pesat dalam ranah informatika, khususnya dalam domain analisis sentimen dengan pemanfaatan algoritma *Naive Bayes*, telah mengalami perubahan signifikan dalam strategi operasional aplikasi brimo. Hal ini telah berdampak pada perbaikan pengalaman pengguna dan penyediaan informasi yang sangat bermanfaat dalam berbagai aspek kehidupan, termasuk dalam ranah teknologi, aspek bisnis, dan sektor pendidikan. Memahami dengan lebih mendalam mengenai evaluasi ulasan pengguna telah menjadi aset berharga dalam proses pengambilan keputusan serta dalam pengembangan strategi bisnis yang efektif. Dalam kerangka analisis sentimen terhadap evaluasi aplikasi brimo di *Google Play Store* dengan menggunakan pendekatan algoritma *Naive Bayes*, muncul berbagai permasalahan dan tantangan yang memiliki relevansi dan nilai untuk diselesaikan. Beberapa masalah signifikan meliputi tingkat ketidakpastian dalam hasil analisis. Permasalahan utama terletak pada tingkat ketidakpastian dalam hasil analisis sentimen tersebut.

Algoritma *Naive Bayes* memiliki kecenderungan untuk memperlakukan kata-kata secara terpisah tanpa mempertimbangkan konteks dengan kedalaman yang cukup, sehingga berpotensi menghasilkan hasil yang tidak selalu akurat ketika mengidentifikasi sentimen yang bersifat kompleks, seperti ironi, sarkasme, atau makna ganda yang sering muncul dalam ulasan. Dengan mengatasi tantangan-tantangan yang disebutkan di atas, analisis sentimen

ulasan aplikasi brimo di *Google Play Store* dengan menggunakan pendekatan algoritma *Naive Bayes* berpotensi meningkatkan tingkat akurasi, efisiensi, dan relevansi. Brimo adalah sebuah aplikasi perbankan berbasis internet dan *mobile* terbaru yang dikembangkan oleh bank bri. Aplikasi ini berfokus pada pengguna data internet dan menawarkan kemudahan dalam bertransaksi bagi nasabah serta individu yang bukan nasabah bank bri. Brimo menyediakan antarmuka pengguna dan pengalaman pengguna terbaru, serta sebagai fitur menarik.

Selain itu, aplikasi ini memberikan fleksibilitas dalam pemilihan sumber dana untuk setiap transaksi, yang dapat berasal dari rekening giro atau tabungan. Upaya ini berkontribusi pada kemampuan aplikasi brimo untuk mengambil keputusan berdasarkan evaluasi pengguna yang lebih tepat, merampingkan pengalaman pelanggan, serta memenuhi keperluan bisnis mereka secara lebih optimal. Selain itu, hal ini juga berperan dalam mengurangi risiko yang berkaitan dengan privasi dan keamanan data pengguna yang mungkin muncul akibat pemanfaatan data. Berdasarkan penelitian sebelumnya yang berjudul "Analisis sentimen aplikasi brimo pada ulasan pengguna di *Google Play Store* menggunakan algoritma *naive bayes*" menjelaskan tentang analisis sentimen pengguna aplikasi Brimo di *Google Play* menggunakan algoritma *Naive Bayes*. Penelitian ini menggunakan metode *Knowledge Discovery in Databases (KDD)* dengan hasil tingkat akurasi sebesar 84.52% [1].

Berdasarkan penelitian sebelumnya yang berjudul “Komparasi Algoritma *Support Vector Machine* dengan *naïve bayes* untuk analisis sentimen pada aplikasi brimo “Menjelaskan tentang analisis sentimen terhadap ulasan pengguna aplikasi mobile banking BRImo menggunakan algoritma *Support Vector Machine (SVM)* dan *Naive Bayes*. Penelitian ini menggunakan metode analisis sentimen dengan algoritma *Support Vector Machine (SVM)* dan *Naive Bayes*. Dengan hasil menunjukkan bahwa algoritma *Support Vector Machine (SVM)* memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma *Naive Bayes* dalam melakukan analisis sentimen terhadap ulasan pengguna aplikasi mobile banking BRImo. *SVM* mencapai tingkat akurasi sebesar 97.69%, sedangkan *Naive Bayes* mencapai tingkat akurasi sebesar 96.53%. [2].

Berdasarkan penelitian sebelumnya yang berjudul “Perbandingan metode *Random Forest* dan *naïve bayes* pada analisis sentimen review aplikasi bc mobile “menjelaskan tentang perbandingan antara metode *Naive Bayes* dan *Random Forest* dalam analisis sentimen terhadap ulasan pengguna aplikasi *BCA Mobile*. Penelitian ini menggunakan metode perbandingan antara *Naive Bayes* dan *Random Forest* dalam analisis sentimen terhadap ulasan pengguna aplikasi *BCA Mobile*. dengan hasil penelitian ini menunjukkan bahwa metode *Random Forest* menghasilkan prediksi yang lebih baik daripada metode *Naive Bayes* dalam analisis sentimen terhadap ulasan pengguna aplikasi *BCA Mobile*. Metode *Random Forest* memiliki akurasi, presisi, *Recall*, dan nilai *F1-score* yang lebih tinggi dibandingkan dengan metode *Naive Bayes*. Penelitian ini juga menekankan pentingnya menganalisis ulasan pengguna untuk meningkatkan kualitas aplikasi perbankan mobile. [3].

Tujuan utama dari penelitian ini adalah untuk melakukan analisis sentimen terhadap ulasan pengguna aplikasi brimo di *Google Play Store* dengan menggunakan algoritma *Naive Bayes*. Melalui penelitian ini, dapat diharapkan diketahui sejauh mana algoritma *naïve bayes* efektif dalam mengklasifikasi sentimen positif, negatif, atau netral dari ulasan pengguna. Selain itu, penelitian ini bertujuan untuk memahami hubungan antara analisis sentimen dengan ulasan yang diberikan oleh pengguna, dengan fokus pemahaman keseluruhan persepsi pengguna terhadap aplikasi Brimo. Dengan memanfaatkan metode komputasi yang melibatkan penerapan algoritma *naïve bayes*, akan menjalankan analisis sentimen terhadap ulasan yang diajukan oleh pengguna di *Google Play Store* yang berkaitan dengan aplikasi brimo. Melalui kerangka pendekatan ini, akan menghimpun ulasan tersebut dan mengelompokkannya ke dalam kategori – kategori yang mencakup aspek positif, negatif, atau netral. Sehingga memungkinkan untuk mengenali dan memahami sentimen yang tersirat dalam setiap ulasan

Dengan memanfaatkan metode komputasi yang melibatkan penerapan algoritma *naïve bayes*, akan menjalankan analisis sentimen terhadap ulasan yang diajukan oleh pengguna di *Google Play Store* yang berkaitan dengan aplikasi brimo. Melalui kerangka pendekatan ini, akan menghimpun ulasan tersebut dan mengelompokkannya ke dalam kategori – kategori yang mencakup aspek positif, negatif, atau netral. Sehingga memungkinkan untuk mengenali dan memahami sentimen yang tersirat dalam setiap ulasan. Hasil penelitian ini membawa implikasi yang signifikan dalam ranah informatika. Dengan berhasil mengidentifikasi serta mengelompokkan sentimen dalam ulasan pengguna aplikasi brimo, penelitian ini akan memberikan wawasan yang lebih mendalam mengenai penerimaan aplikasi perbankan berbasis internet dan mobile oleh pengguna. Hal ini akan berkontribusi dalam memahami aspek positif, dan negatif dari aplikasi brimo, serta menyediakan umpan balik yang berharga yang dapat dimanfaatkan oleh bank bri untuk meningkatkan kualitas layanan pengguna.

2. TINJAUAN PUSTAKA

2.1. Teks Mining

Teks mining adalah proses penambahan data (data mining) yang mengambil informasi dari data berupa teks. Teks mining mengekstrak data data berupa teks seperti email, dokumen teks, file HTML serta data dari sosial media.[4] Text Mining ialah langkah analisis dan ekstraksi pemahaman dari sejumlah besar data teks yang tidak terstruktur. Kategori data ini dapat berasal dari beragam sumber, seperti platform media.[5]

2.2. Klasifikasi

Metode klasifikasi telah dirancang untuk menganalisis sentimen dalam data berupa teks, salah satunya yaitu *naïve bayes*. Dalam pengolahan teks, terdapat dua jenis model klasifikasi *naïve bayes* yang dapat digunakan yaitu *naïve bayes* multinomial dan bernouli. Model *Naive Bayes* multinomial menghitung frekuensi kemunculan setiap kata dalam dokumen, sementara model *naïve bayes* Bernoulli hanya memperhitungkan apakah kata tersebut muncul pada dokumen atau tidak, tanpa memperhitungkan jumlah kemunculannya. [6] klasifikasi merujuk pada tugas memprediksi label kelas dari suatu sampel berdasarkan sekumpulan fitur atau karakteristik yang diberikan. Sebagai contoh, dengan data mengenai usia, jenis kelamin, dan pendapatan seseorang, tujuan dari tugas klasifikasi mungkin adalah untuk memprediksi apakah orang tersebut cenderung memiliki pekerjaan dengan gaji tinggi atau tidak. Terdapat berbagai algoritma yang dapat digunakan untuk klasifikasi, seperti algoritma *Naive Bayes (NB)* dan *K-Nearest Neighbors (KNN)*. Algoritma-algoritma ini menggunakan pendekatan yang berbeda untuk belajar dari data pelatihan dan melakukan prediksi terkait label kelas dari sampel yang baru.[7]

2.3. Naïve bayes

Naïve bayes adalah metode yang tidak memiliki aturan dan menggunakan cabang matematika yang disebut teori probabilitas untuk mendapatkan peluang setinggi mungkin dengan melihat frekuensi atau jumlah kemunculan setiap klasifikasi dalam data pelatihan.

Dalam pengembangan basis data, *naïve bayes* melibatkan pembelajaran yang diawasi jenis pembelajaran mesin yang membutuhkan sampel sebagai data pelatihan label, dikelompokkan menjadi dua bagian, yaitu. Klasifikasi dan regresi. Klasifikasi saat variabel menjadi kategori, panas dan dingin, sakit atau tidak sakit dll. Variabel berupa nilai riil seperti bobot, nilai uang dll contoh lainnya yaitu *Support Vector Machine (SVM)*, *K-Nearest Neighbor (KNN)*, *Artificial Neural Network (ANN)*. *Naïve bayes* juga merupakan metode yang menggunakan Teknik probabilistik, dimana satu fitur dan fitur yang lainnya dalam data yang sama tidak saling berhubungan. [1] Algoritma klasifikasi *Naïve bayes* bertujuan untuk mengelompokkan data ke dalam kelas tertentu. Kinerja pengklasifikasi diukur dengan nilai akurasi prediksi. Langkah-langkah dalam proses algoritma *Naïve bayes* adalah sebagai berikut: 1. Menghitung total kelas/label. 2. Menghitung jumlah kasus yang memiliki label yang sama. 3. Mengalikan semua variabel kelas. 4. Membandingkan hasil dari semua variabel. [8]

2.4. Knowledge Discovery in Database (KDD)

Tf-Idf (Term Frequency-Inverse Document Frequency) Merupakan suatu metode yang berguna untuk menghitung bobot dari setiap kata yang digunakan. Metode ini menghitung nilai TF dan IDF pada setiap kata. Semakin banyak frekuensi kemunculan dari kata tersebut, maka semakin besar nilai TF-IDF. IDF menunjukkan bagaimana suatu term di distribusikan secara luas dalam dokumen, Semakin besar nilai IDFnya [9] proses term weighting, di mana term akan diberi bobot (nilai) yang mencerminkan tingkat pentingnya terhadap dokumen. Selanjutnya, dilakukan perhitungan bobot pada term yang dicari dalam setiap dokumen dengan tujuan untuk mengetahui relevansi dan kesamaan term di dalam dokumen tersebut. Semakin sering sebuah kata muncul dalam koleksi dokumen, semakin tinggi nilai atau bobot kata tersebut. Setelah selesai tahap pembobotan, dilanjutkan dengan proses klasifikasi.[10]

3. METODE PENELITIAN

3.1. Selection

Di awal tahapan terhadap tahap selection, yang mana data selection adalah proses menganalisis data yang relevan dari database karena sering ditemukan bahwa tidak semua data dibutuhkan dalam proses data mining. Pemilihan data mendahului praktik pengumpulan data yang sebenarnya. Data diambil dari review aplikasi Brimo di *Google Play Store*

dengan cara scrapping data yang di peroleh sebanyak 1200 data.

3.2. Data Cleansing atau Pre-Processing

Pembersihan data untuk proses mengidentifikasi bagian data yang tidak lengkap, salah, benar, tidak akurat atau hilang dan mengoreksi, mengganti atau menghapusnya. Cleansing penelitian ini akan menghapus @, #, karakter URL serta atribut yang hanya teks. Operatore yang digunakan adalah substansi.

3.3. Transformation

Dalam proses pembobotan kata menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF), perhitungan bobot kata-kata pada setiap ulasan menjadi krusial dalam menganalisis sentimen. Pendekatan ini memainkan peran penting dalam menilai kebermaknaan kata-kata dalam konteks analisis sentimen. TF-IDF melakukan perhitungan seberapa sering suatu kata muncul dalam dokumen tertentu (Term Frequency) dan seberapa umum kata tersebut dalam seluruh dataset (Inverse Document Frequency). Hasilnya adalah bobot yang mencerminkan tingkat signifikansi kata dalam suatu teks.

3.4. Data Mining

Data mining proses mengubah teks yang tidak teratur menjadi format terstruktur untuk mengidentifikasi pola yang bermakna dan wawasan baru. Pada tahap riset ini menerapkan *naïve bayes*.

3.5. Pattern Evaluation

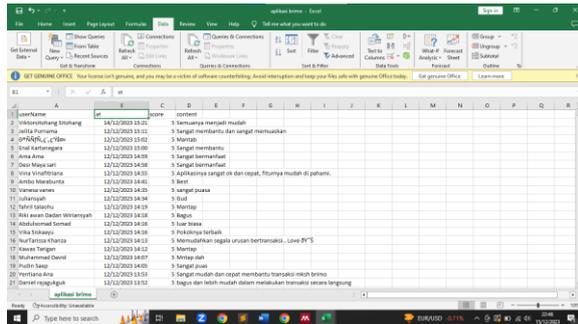
Pattern Evaluation didefinisikan sebagai mengidentifikasi model yang meningkatkan wawasan berdasarkan ukuran yang diberikan. Temukan tempat menarik untuk setiap model menggunakan ringkasan dan visualisasi agar informasi dapat dipahami oleh pengguna.

4. HASIL DAN PEMBAHASAN

4.1. HASIL

4.1.1. Data Selection

Di awal tahapan terhadap tahap selection, yang mana data selection adalah proses menganalisis data yang relevan dari database karena sering ditemukan bahwa tidak semua data dibutuhkan dalam proses data mining. Pemilihan data mendahului praktik pengumpulan data yang sebenarnya. Data diambil dari review aplikasi Brimo di *Google Play Store* dengan dengan cara scrapping data melalui tools website google colab yang di peroleh sebanyak 1200 data.



Gambar 1 Hasil dari Scrapping Data excel dari google colab

Tabel 1 Hasil Input output

Input Proses		Output Proses	
Semuanya	menjadi	semuanya	menjadi
berubah		mudah	
Pokonya	terbaik	Pokonya	terbaik

4.1.2. Data Cleansing

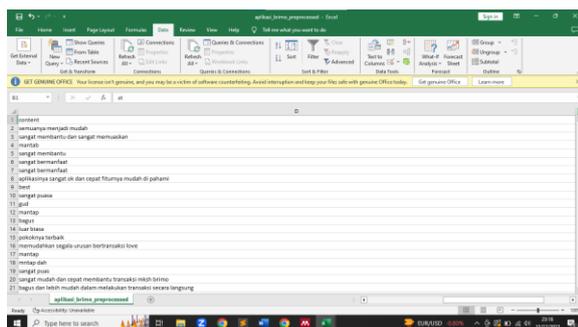
Pembersihan data untuk proses mengidentifikasi bagian data yang tidak lengkap, salah, benar, tidak akurat atau hilang dan mengoreksi, mengganti atau menghapusnya. Cleansing penelitian ini akan menghapus @, #, karakter URL serta atribut yang hanya teks. Operatore yang digunakan adalah substasi.

Dibawah ini adalah gambar proses cleansing data yang prosesnya memalui tools website google colab



Gambar 2 Proses cleansing data dari tools google colab

Dan adapun gambar hasil cleansing data di bawah ini Dan data hasil cleansing tersebut melalui pengolahan website google colab yang dimpan menjadi file csv di excel sebagai berikut :



Gambar 3 Hasil Proses Cleansing Data dari google colab

4.1.3. Case Folding

Case folding adalah proses normalisasi teks yang melibatkan pengubahan semua huruf dalam suatu teks menjadi huruf kecil atau huruf besar. Langkah ini sering digunakan untuk memastikan konsistensi dan mengurangi kompleksitas dalam pemrosesan teks

4.1.4. Tokenizing

Tokenizing adalah proses penting dalam pemrosesan teks yang melibatkan pembagian teks atau kalimat menjadi unit-unit lebih kecil yang disebut token. Jenis tokenisasi dapat mencakup pembagian menjadi kata-kata, kalimat, karakter, frasa, atau angka, tergantung pada tujuan analisis.

4.1.5. Stop Word Removal

Tahap proses stop word removal adalah langkah dalam pemrosesan teks yang bertujuan untuk menghilangkan kata-kata stop (stop words) dari teks yang sedang diolah. Stop words adalah kata-kata umum yang sering muncul namun memiliki sedikit nilai informatif dalam analisis teks, seperti "dan", "atau", "yang", dan sebagainya

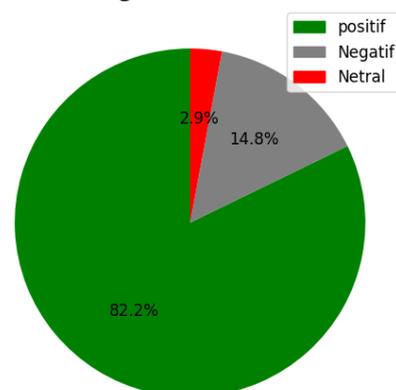
4.1.6. stemming

Stemming adalah proses dalam pemrosesan teks yang bertujuan untuk menghapus awalan atau akhiran kata sehingga hanya menyisakan akar kata atau bentuk dasar. Tujuannya adalah untuk mengurangi variasi kata ke dalam bentuk yang lebih sederhana, sehingga kata-kata dengan akar yang sama dapat dianggap identic

4.1.7. Labeling

Proses labeling dalam konteks pemrosesan data sering merujuk pada tugas memberikan label atau kategori tertentu kepada setiap observasi atau data dalam dataset. Dalam proses ini yaitu menentukan score, jika score 3, 4, dan score 5 bisa disebut sentimen positif negatif. Beriku adalah gambar setelah pelabelan. Proses pelabelan juga melalui tahap pemrosesan data yang di proses melalui website google colab proses pelabelan juga melalui tahap pemrosesan data yang di proses melalui tools website google colab

Diagram Sentimen



Gambar 4 Hasil visualisasi setelah pelabelan dari tools google colab

4.1.8. Splitting Data

Proses splitting data merupakan tahap penting dalam pengembangan analisis data, yang melibatkan pembagian dataset menjadi dua atau lebih subset yang berbeda. Dataset umumnya dibagi menjadi set pelatihan (*training set*) dan set pengujian (*test set*). Set pelatihan digunakan untuk melatih model, sementara set pengujian digunakan untuk menguji kinerja model yang telah dilatih pada data baru yang belum pernah dilihat sebelumnya. Proses selanjutnya adalah tahap instalasi library untuk splitting data. Di bawah ini adalah proses dari splitting data

```

Proses Split Data
# Membaca file CSV
df = pd.read_csv('sentiment_data.csv')

# Memisahkan kolom 'review_content' sebagai fitur dan kolom 'sentiment' sebagai target
X = df['review_content']
y = df['sentiment']

# Mengubah data menjadi data latih dan data uji dengan pembagian 80:20
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Mengubah data latih dan data uji menjadi DataFrame
train_df = pd.DataFrame(X_train, columns=['review_content'], index=range(X_train.shape[0]))
test_df = pd.DataFrame(X_test, columns=['review_content'], index=range(X_test.shape[0]))
    
```

Gambar 5 Proses Splitting data

4.1.9. Transformasi

Proses transformasi data dalam konteks analisis data melibatkan manipulasi struktur atau nilai data untuk memenuhi persyaratan tertentu atau meningkatkan kualitas data. Ini dapat mencakup berbagai tindakan seperti normalisasi, pengkodean variabel kategorikal, pemilihan fitur, atau perubahan skala nilai. Transformasi data bertujuan untuk memastikan bahwa data yang digunakan sesuai dengan kebutuhan analisis atau model yang akan dibangun.

4.1.10. Data Mining

Data mining proses mengubah teks yang tidak teratur menjadi format terstruktur untuk mengidentifikasi pola yang bermakna dan wawasan baru. Pada tahap riset ini menerapkan *naive bayes*

4.1.11. Evaluasi

Dari hasil evaluasi model klasifikasi *Naive Bayes* pada data uji, terlihat bahwa model ini mencapai tingkat akurasi sebesar 87.92%, yang menggambarkan seberapa baik model dapat memprediksi dengan tepat pada data uji. Presisi sebesar 83.65% menunjukkan seberapa akurat model dalam mengidentifikasi kasus positif, sementara *Recall* sebesar 87.92% menunjukkan kemampuan model untuk menemukan sebagian besar kasus positif dalam data uji. *F1 Score* sebesar 84.39% memberikan gambaran keseimbangan antara presisi dan *Recall*. Dengan hasil evaluasi ini, peneliti dapat melihat bahwa model *Naive Bayes* memiliki kinerja yang positif dalam mengklasifikasikan teks pada konteks dataset yang digunakan. Di bawah ini adalah table hasil model *Naive bayes*

Tabel 2 Hasil Model *naive bayes*

Akurasi	87,92%
Presisi	83,65%
Recall	87,92%
F1- Score	84,39%

4.2. PEMBAHASAN

4.2.1. Penerapan Algoritma *Naive bayes*

Penelitian ini difokuskan pada implementasi metode *Naive bayes* untuk menganalisis sentimen pengguna aplikasi Brimo. Proses dimulai dengan pengumpulan data menggunakan teknik scrapping, dengan 1200 data ulasan sebagai sampel. Preprocessing melibatkan labelisasi, pembersihan data, dan transformasi menggunakan TF-IDF. Model *Naive Bayes* diterapkan dengan evaluasi melibatkan matriks kebingungan, akurasi, presisi, *Recall*, dan *F1-Score*. Meskipun tingkat akurasi tinggi sebesar 87.92%, perhatian khusus diperlukan untuk meningkatkan kemampuan model mengenali sentimen negatif, ditunjukkan oleh nilai *Recall* yang relatif rendah pada kelas tersebut. Pengoptimalan lebih lanjut pada parameter atau pendekatan model mungkin diperlukan untuk meningkatkan kinerja pada sentimen negatif.

4.2.2. Implementasi Algoritma *Naive bayes*

Proses prediksi sentimen pada data baru melibatkan langkah-langkah penting, termasuk penyimpanan dan pemuat ulang model *Naive Bayes* menggunakan *joblib* untuk penggunaan masa mendatang tanpa pelatihan ulang. Pengolahan data uji baru dengan *pandas* memberikan gambaran struktur data sebelum prediksi sentimen. Model *Naive Bayes* digunakan untuk melakukan prediksi sentimen pada data uji baru, dan hasilnya ditambahkan sebagai kolom 'predicted_sentiment' dalam *DataFrame*. Proses ini mengilustrasikan kemampuan model dalam memberikan prediksi sentimen untuk data teks yang belum pernah dilihat sebelumnya, memperkuat keberlanjutan dan aplikabilitas algoritma *Naive Bayes* dalam analisis sentimen pada data teks.

5. KESIMPULAN DAN SARAN

Penerapan metode *Naive bayes* dalam analisis sentimen aplikasi Brimo melibatkan proses pengumpulan data melalui web scraping dengan menggunakan 1200 data ulasan sebagai sampel. Langkah awal ini penting untuk mendapatkan representasi yang cukup dari berbagai pendapat pengguna, yang selanjutnya dilabeli berdasarkan rating yang diberikan. Proses preprocessing data merupakan tahapan kritis dalam persiapan data, yang mencakup labelisasi, pembersihan data, perubahan huruf menjadi kecil, penghapusan kata pengisi, serta pemisahan dan pemangkasan kata.

Evaluasi kinerja algoritma *Naive bayes* pada analisis sentimen terhadap ulasan pengguna aplikasi Brimo menunjukkan tingkat akurasi yang tinggi sebesar 87.92%. Meskipun demikian, terdapat keterbatasan dalam mengidentifikasi sentimen negatif, sebagaimana tercermin dari nilai *Recall* yang relatif rendah pada kelas tersebut. Presisi dan *F1-Score* memberikan gambaran keseimbangan antara akurasi positif dan kemampuan mengenali seluruh sentimen positif. Keseluruhan, meskipun model

mampu mengklasifikasikan sentimen dengan baik secara keseluruhan, perhatian khusus perlu diberikan pada perbaikan performa dalam menghadapi sentimen negatif. Proses implementasi model *Naive Bayes* pada data baru menunjukkan efisiensi melalui penggunaan `joblib` untuk menyimpan dan memuat model, memungkinkan penggunaan kembali model tanpa pelatihan ulang. Penggunaan data uji baru ('data_uji3.csv') dan pengolahan awal dengan `pandas` memberikan pemahaman tentang struktur data sebelum melakukan prediksi sentimen. Model berhasil memprediksi sentimen pada data teks baru, dan hasilnya dapat divisualisasikan melalui kolom '`predicted_sentiment`' pada `DataFrame`.

Penambahan data latih, perluasan dataset dengan menambah jumlah data latih dapat membantu model untuk memahami variasi yang lebih besar dari sentimen pengguna. Dengan lebih banyak data latih, model dapat belajar pola yang lebih kompleks dan dapat meningkatkan kinerjanya. Penanganan Ketidakseimbangan Kelas, keterbatasan dalam mengenali sentimen negatif, seperti yang terlihat dari nilai *Recall* yang rendah, dapat diatasi dengan menyeimbangkan jumlah sampel untuk setiap kelas.

Teknik *oversampling* atau *undersampling* dapat digunakan untuk mencapai keseimbangan yang lebih baik antara sentimen positif dan negatif. Penanganan Kata-Kata Tidak Biasa, proses *preprocessing* dapat ditingkatkan dengan menangani kata-kata yang tidak umum atau slang yang mungkin tidak terpenuhi dengan baik oleh metode TF-IDF. Mungkin perlu adanya pengelompokan atau penanganan khusus untuk kata-kata ini agar model dapat lebih efektif

DAFTAR PUSTAKA

- [1] M. K. Khoirul Insan, U. Hayati, and O. Nurdiawan, "Analisis sentimen aplikasi brimo pada ulasan pengguna di *Google Play* menggunakan algoritma *Naive Bayes*," *JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 1, pp. 478–483, 2023, doi: 10.36040/jati.v7i1.6373.*
- [2] I. J. Anggi Puji Astuti, Syariful Alam, "Komparasi algoritma *Support Vector Machine* dengan *Naive Bayes* untuk analisis sentimen pada aplikasi BRImo," *J. Bangkit Indones., vol. 11, no. 2, pp. 1–6, 2022, doi: 10.52771/bangkitindonesia.v11i2.196.*
- [3] A. M. Miftahusalam, "Perbandingan metode *Random Forest* dan *Naive Bayes* pada analisis sentimen review aplikasi *BCA Mobile*," *SIPTEK Semin.* [Online]. Available: <https://proceeding.unesa.ac.id/index.php/siptek/article/view/184>
- [4] D. Sepri, "Penerapan algoritma *Naive Bayes* untuk analisis kepuasan penggunaan aplikasi bank," *J. Comput. Syst. Informatics (JoSYC, vol. 2, no. 1, pp. 135–139, 2020.*
- [5] M. Tegar Lazuardi, T. Suprapti, and Y. Arie Wijaya, "Perancangan Model Sentimen Tweet Terhadap Pilkada Dki Jakarta Tahun 2017 Menggunakan Algoritma *Naive Bayes*," *JATI (Jurnal Mhs. Tek. Inform., vol. 7, no. 1, pp. 308–312, 2023, doi: 10.36040/jati.v7i1.6328.*
- [6] M. Raffi, A. Suharso, and I. Maulana, "Analisis sentimen ulasan aplikasi binar pada *Google Play Store* menggunakan algoritma *naive bayes*," *INTECOMS J. Inf. Technol. Comput. Sci., vol. 6, no. 1, pp. 450–462, 2023, doi: 10.31539/intecom.v6i1.6117.*
- [7] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, and Fitri Nurapriani, "Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma *Naive Bayes* dan KNN," *J. KomtekInfo, vol. 10, pp. 1–7, 2023, doi: 10.35134/komtekinfo.v10i1.330.*
- [8] K. Anwar, "Analisa sentimen Pengguna Instagram Di Indonesia Pada Review Smartphone Menggunakan *Naive Bayes*," *KLIK Kaji. Ilm. Inform. dan Komput., vol. 2, no. 4, pp. 148–155, 2022, doi: 10.30865/klik.v2i4.315.*
- [9] M. Suhendra, W. Swastika, and M. Subianto, "Analisis sentimen pada ulasan aplikasi video conference menggunakan *Naive Bayes*," ... *Jurnal Ilmiah Sains & sainsbertek.machung.ac.id, 2021. [Online]. Available: http://sainsbertek.machung.ac.id/index.php/sbtek/article/download/145/89*
- [10] W. Wahyuni, "Analisis Sentimen terhadap Opini Feminisme Menggunakan Metode *Naive Bayes*," *J. Inform. Ekon. Bisnis, vol. 4, pp. 148–153, 2022, doi: 10.37034/infekon.v4i4.162.*