Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473

ANALISIS PERBANDINGAN SENTIMEN ULASAN PENGGUNA APLIKASI MYBCA DAN BCA MOBILE MENGGUNAKAN NAÏVE BAYES

Fatra Nadia Rahmawati[1], Belsana Butar Butar[2], Kartika Mariskhana[3]
[1]fatranadia60@gmail.com, [2]belsana.bbb@bsi.ac.id, dan [3]kartika.kma@bsi.ac.id
[1,2,3]Sistem Informasi, Universitas Bina Sarana Informatika

Abstrak

Perkembangan teknologi informasi mendorong Bank Central Asia (BCA) mengoperasikan dua platform *mobile banking* secara bersamaan, yakni MyBCA dan BCA Mobile. Tingginya volume ulasan pengguna di *Google Play Store* memerlukan pendekatan komputasional untuk memahami persepsi pengguna secara objektif. Penelitian ini bertujuan mengklasifikasikan, membandingkan distribusi, dan mengukur kinerja model *Naïve Bayes* pada kedua aplikasi. Data dikumpulkan melalui *web scraping*, dengan total 8.000 ulasan. Tahapan penelitian meliputi *preprocessing* teks, pelabelan otomatis menggunakan metode *Lexicon-Based* dengan kamus InSet, pembobotan TF-IDF, dan klasifikasi *Naïve Bayes* dengan rasio 80:20. Hasil menunjukkan MyBCA memperoleh sentimen positif 54,31%, negatif 34,66%, dan netral 11,03%, sedangkan BCA Mobile memperoleh positif 63,70%, negatif 25,10%, dan netral 11,19%. Model MyBCA menghasilkan *accuracy* 63,70% dan BCA Mobile 73,62%. Penelitian ini menyimpulkan bahwa BCA Mobile memperoleh persepsi pengguna yang lebih positif dibandingkan MyBCA.

Kata kunci: Analisis Sentimen, *Naïve Bayes*, *Mobile Banking*, *Lexicon-Based*, TF-IDF

Abstract

Advances in information technology have led Bank Central Asia (BCA) to operate two mobile banking platforms simultaneously MyBCA and BCA Mobile. The high volume of user reviews on the Google Play Store necessitates a computational approach to objectively understand user perceptions. This study aims to classify, compare the distributions of, and measure the performance of the Naïve Bayes model on both applications. Data were collected via web scraping, totaling 8,000 reviews. The research stages included text preprocessing, automatic labelling using the Lexicon-Based method with the InSet dictionary, TF-IDF weighting, and Naïve Bayes classification with an 80:20 split. The result show that MyBCA received 54.31% positive sentiment, 34.66% negative sentiment, and 11,03% neutral sentiment, while BCA Mobile received 63.70% positive sentiment, 25.10% negative sentiment, and 11.19% neutral sentiment. The MyBCA model achieved an accuracy of 63.70%, and the BCA Mobile model achieved 73.62%. this study concludes that BCA Mobile receives a more positive user perception than MyBCA.

Keywords: *Sentiment Analysis, Naïve Bayes, Mobile Banking, Lexicon-Based, TF-IDF*

1. Pendahuluan

Perkembangan teknologi informasi mendorong transformasi layanan perbankan menuju sistem digital berbasis *mobile banking.* Sejak pandemi COVID-19, transaksi perbankan digital mengalami peningkatan signifikan, dengan pertumbuhan mencapai 61,80% pada Agustus 2021[1]. Sebagai salah satu bank terbesar di Indonesia, Bank Central Asia (BCA) mengembangkan dua aplikasi *mobile banking*, yakni MyBCA dan BCA Mobile, untuk memenuhi kebutuhan transaksi digital masyarakat. Hingga Desember 2023, jumlah pengguna *mobile banking* BCA mencapai 30,3 juta dengan total transaksi digital sebesar Rp24.825 triliun[2]. Tingginya jumlah pengguna menghasilkan ribuan ulasan pada *Google Play Store* yang mencerminkan pengalaman dan persepsi pengguna terhadap kualitas layanan aplikasi.

Ulasan pengguna merupakan sumber informasi yang bernilai dalam mengevaluasi kualitas layanan aplikasi. Namun, jumlah ulasan yang sangat besar menyebabkan proses analisis secara manual menjadi tidak efektif. Selain itu, isi ulasan tidak selalu sejalan dengan *rating* bintang yang diberikan, sehingga penentuan sentimen tidak dapat hanya didasarkan pada nilai *rating*[3]. Analisis sentimen merupakan salah satu teknik *text mining* yang digunakan untuk mengidentifikasi opini atau persepsi pengguna terhadap suatu objek berdasarkan data teks. Penerapan teknik ini memungkinkan informasi dari ulasan pengguna dianalisis

Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473

secara sistematis dan efisien menggunakan algoritma klasifikasi[4]. Oleh karena itu, diperlukan pendekatan analisis sentimen berbasis *machine learning* untuk mengidentifikasi opini pengguna secara lebih objektif dan sistematis.

Di antara berbagai algoritma klasifikasi yang tersedia, *Naïve Bayes* dipilih dalam penelitian ini karena sifatnya yang probabilistik, sederhana, dan memiliki efisiensi komputasi yang tinggi. Algoritma ini bekerja dengan baik pada data berdimensi tinggi seperti data teks, karena setiap kata direpresentasikan sebagai fitur independen sehingga proses pelatihan model dapat dilakukan dengan cepat meskipun jumlah data ulasan besar[5]. Selain itu, *Naïve Bayes* tidak memerlukan sumber daya komputasi maupun volume data pelatihan sebesar model *deep learning*, sehingga lebih mudah diimplementasikan dan diinterpretasikan hasilnya, karakteristik yang relevan untuk kebutuhan analisis sentimen pada ulasan aplikasi *mobile banking* dalam penelitian ini.

Penelitian oleh Yoman et al.[6] telah membandingkan sentimen pengguna aplikasi BCA Mobile dan MyBCA menggunakan metode *lexicon-based* dan menunjukkan bahwa mayoritas ulasan bersifat netral. Namun, penelitian tersebut belum menerapkan algoritma klasifikasi sehingga belum dapat mengevaluasi performa model dalam mengklasifikasikan sentimen. Sementara itu, Putri at al.[7] menerapkan metode *deep learning* berbasis *Gated Recurrent Unit* (GRU) dengan pelabelan otomatis menggunakan IndoBERT dan memperoleh akurasi yang tinggi. Meskipun demikian, pendekatan tersebut memiliki kompleksitas komputasi yang lebih tinggi, dan memerlukan sumber daya yang lebih besar. Hal ini membuat implementasinya relatif lebih kompleks dan kurang transparan dibandingkan algoritma klasifikasi konvensional.

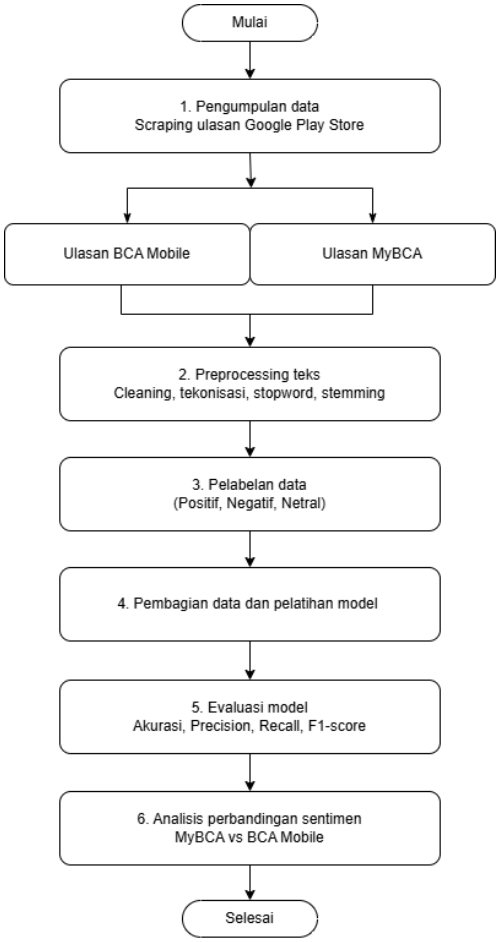
Berdasarkan kedua penelitian tersebut, masih terdapat kesenjangan penelitian dalam penerapan algoritma klasifikasi yang sederhana namun efektif untuk membandingkan sentimen pengguna aplikasi MyBCA dan BCA Mobile, disertai proses pelabelan otomatis yang transparan dan dapat ditelusuri dasar pengambilan keputusannya. Untuk mengatasi keterbatasan tersebut, penelitian ini menawarkan kebaruan berupa penerapan pelabelan otomatis berbasis *Lexicon-Based* menggunakan kamus InSet *(Indonesian Sentiment Lexicon)* yang dikombinasikan dengan algoritma *Naïve Bayes* untuk membandingkan sentimen pengguna kedua aplikasi. Kombinasi ini memungkinkan proses pelabelan yang transparan sekaligus proses klasifikasi yang ringan secara komputasi.

Penelitian ini bertujuan untuk membandingkan sentimen pengguna aplikasi MyBCA dan BCA Mobile berdasarkan ulasan pada *Google Play Store* menggunakan pelabelan otomatis berbasis *Lexicon-Based* InSet dan algoritma Naïve Bayes, serta mengevaluasi kinerja model klasifikasi yang dihasilkan menggunakan metrik *accuracy*, *precision*, *recall*, dan F1-*score*. Hasil penelitian diharapkan dapat memberikan gambaran yang lebih objektif mengenai persepsi pengguna terhadap kedua aplikasi serta menjadi masukan bagi BCA dalam meningkatkan kualitas layanan perbankan digital.

2. Metode

Penelitian ini menggunakan pendekatan kuantitatif berbasis komputasi dengan menerapkan teknik *text mining* dan *machine learning* untuk menganalisis serta membandingkan sentimen pengguna aplikasi MyBCA dan BCA Mobile. Data bersumber dari ulasan pengguna di *Google Play Store* yang dikumpulkan melalui teknik *web scraping*, dengan total 8.000 ulasan yang terdiri dari 4.000 ulasan MyBCA dan 4.000 ulasan BCA Mobile. Seluruh proses komputasi dilaksanakan menggunakan bahasa pemrograman Python di platform *Google Colaboratory*. Alur penelitian terdiri dari enam tahapan sistematis, sebagaimana disajikan pada Gambar 1.

Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473



Gambar 1. Flowchart Penelitian

2.1. Pengumpulan Data

Pengumpulan data dilakukan menggunakan teknik *web scraping* melalui *library google-play-scraper* pada Python. Data yang dikumpulkan mencakup atribut teks ulasan *(content)*, skor rating *(score)*, nama pengguna *(username)*, dan tanggal ulasan, dengan filter bahasa Indonesia untuk menjaga konsistensi analisis. Pengambilan data dilakukan secara bertahap *(batch)* dengan jeda satu detik antarbatch guna menghindari pembatasan akses dari server *Google Play Store*.

2.2. *Preprocessing* Teks

Tahap *preprocessing* bertujuan mengubah data teks mentah menjadi format yang siap diproses oleh algoritma klasifikasi. Tahapan ini penting karena algoritma *machine learning* tidak dapat mengolah teks secara langsung sehingga diperlukan proses pembersihan dan standarisasi data sebelum dilakukan pembobotan dan klasifikasi[8].Tahapan ini terdiri dari lima sub-proses yang dilaksanakan secara berurutan, sebagaimana dijelaskan pada Tabel 1.

Tabel 1. Tahapan *Preprocessing* Teks

No	Tahapan	Fungsi
1	*Case Folding*	Mengubah seluruh teks menjadi huruf kecil
2	*Cleaning*	Menghapus URL, emoji, angka, simbol, dan tanda baca
3	Tokenisasi	Memecah teks menjadi satuan kata *(token)*
4	*Stopword Removal*	Menghapus kata umum yang tidak mengandung makna sentimen
5	*Stemming*	Mengubah kata berimbuhan ke bentuk dasar menggunakan algoritma Sastrawi

Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473

Proses *stemming* menggunakan algoritma Sastrawi yang dirancang khusus untuk menangani morfologi bahasa Indonesia, sehingga variasi kata berimbuhan dapat dinormalisasi ke bentuk dasarnya secara tepat.

2.3. Pelabelan Sentimen *(Lexicon-Based)*

Pelabelan sentiment dilakukan secara otomatis menggunakan metode *lexicon-based* dengan kamus InSet *(Indonesian Sentiment Lexicon).* Metode ini bekerja dengan menghitung akumulasi skor kata positif dan skor kata negatif pada setiap ulasan. Penentuan label mengikuti aturan berikut: apabila total skor positif lebih besar dari skor negatif, ulasan diberi label **positif** (1); dan apabila skor negatif lebih besar, ulasan diberi label **negatif** (-1); dan apabila kedua skor sama atau tidak ditemukan kata sentimen yang relevan, ulasan diberi label **netral** (0). Pendekatan *lexicon-based* dipilih karena bersifat *unsupervised*, serta kamus InSet dirancang khusus untuk menangani kosakata bahasa Indonesia termasuk kata informal[9].

2.4. Pembobotan TF-IDF

Sebelum data teks dapat diproses oleh algoritma *Naïve Bayes*, setiap ulasan perlu direpresentasikan dalam bentuk numerik menggunakan metode TF-IDF *(Term Frequency-Inverse Document Frequency).* TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen secara relatif terhadap seluruh koleksi dokumen, sehingga kata yang lebih diskriminatif mendapatkan bobot lebih tinggi[10]. Nilai TF-IDF dihitung menggunakan persamaan berikut:

$$W_{t,d} = TF(t,d) \times IDF(t) \quad (1)$$

Di mana:

$$TF(t,d) = \frac{f_{t,d}}{\sum_{t' \in d} f_{t',d}} \ , IDF(t) = \log\left(\frac{N}{df_t}\right) \quad (2)$$

Keterangan: $f_{t,d}$ adalah frekuensi kata t dalam dokumen d, N adalah jumlah total dokumen, dan df_t adalah jumlah dokumen yang mengandung kata t. Proses pembobotan dilakukan menggunakan TfidfVectorizer dari *library* scikit-learn yang hanya dilatih pada data *training* untuk menghindari kebocoran data *(data leakage).*

2.5. Algoritma *Naïve Bayes*

Naïve Bayes merupakan algoritma klasifikasi berbasis probabilitas yang bekerja berdasarkan *Teorema Bayes* dengan asumsi independensi kondisional antar setiap fitur bersifat independen terhadap fitur lainnya. Meskipun menggunakan asumsi yang sederhana, algoritma ini memiliki proses komputasi yang efisien dan mampu menghasilkan performa yang baik pada berbagai permasalahan klasifikasi, terutama pada data teks. Oleh karena itu, *Naïve Bayes* banyak diterapkan pada klasifikasi dokumen, penyaringan spam, maupun analisis sentimen[11]. Persamaan dasar *Teorema Bayes* yang digunakan adalah:

$$P(C_k \mid X) = \frac{P(X \mid C_k) \cdot P(C_k)}{P(X)} \quad (3)$$

Keterangan: $P(C_k \mid X)$ adalah probabilitas kelas C_k diberikan fitur X *(posterior)*, $P(X \mid C_k)$ adalah *likelihood*, $P(C_k)$ adalah probabilitas kelas sebelum pengamatan (*pior*), dan $P(X)$ adalah probabilitas fitur sebagai konstanta normalisasi. Dalam penelitian ini, klasifikasi dilakukan dengan memilih kelas yang memiliki nilai probabilitas tertinggi.

Varian yang digunakan adalah Multinomial *Naïve Bayes* karena sesuai untuk data teks yang direpresentasikan menggunakan pembobotan TF-IDF. Data dibagi menjadi data *training* (80%) dan data *testing* (20%) menggunakan fungsi *train_test_split* dari *scikit-learn* dengan parameter *random_state*=42 untuk menjamin replikabilitas hasil, serta parameter *stratify=y* untuk mempertahankan proporsi kelas pada kedua subset data.

Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473

2.6. Evaluasi Model

Kinerja model dievaluasi berdasarkan *confusion matrix* menggunakan empat metrik evaluasi, yaitu *accuracy*, *precision*, *recall*, dan F1-*score*. Sebagaimana disajikan pada Tabel 2.

Tabel 2. Metrik Evaluasi Model

Metrik	Rumus	Keterangan
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Proprosi prediksi benar secara keseluruhan
Precison	$\frac{TP}{TP + FP}$	Ketepatan prediksi per kelas sentimen
Recall	$\frac{TP}{TP + FN}$	Kemampuan model mendeteksi kelas yang benar
F1-*Score*	$\frac{2 \times Precision \times Recall}{Precision + Recall}$	Rata-rata harmonik *precision* dan *recall*

Keterangan: TP = *True Positive*, TN = *True Negative*, FP = *False Positive*, FN = *False Negative*. Evaluasi dilakukan secara terpisah untuk model BCA Mobile dan MyBCA agar perbandingan kinerja bersifat adil dan proporsional.

3. Hasil dan Pembahasan

Tahap *preprocessing* dilakukan terhadap 8.000 ulasan mentah hasil *scraping* yang terdiri dari 4.000 ulasan MyBCA dan 4.000 ulasan BCA Mobile. Setelah melalui lima tahapan *preprocessing* secara berurutan, sejumlah ulasan tereliminasi karena menjadi kosong setelah proses *cleaning*, yakni ulasan yang hanya berisi emoji, simbol, atau karakter tunggal tanpa makna linguistik. Data bersih yang dihasilkan berjumlah 3.924 ulasan MyBCA dan 3.904 ulasan BCA Mobile. Contoh transformasi teks pada setiap tahapan *preprocessing* disajikan pada Tabel 3.

Tabel 3. Contoh Hasil *Preprocessing* Teks

Tahapan	Contoh Teks MyBCA	Contoh Teks BCA Mobile
Teks Asli	"MyBCA tampilannya bagus, fitur lengkap banget. Login juga mudah 👍 "	"TIBA² APK BCA MOBILE HILANG DARI HP SAYA... TRUS BAGAIMANA UNTUK TRANSAKSI?"
Case Folding	"mybca tampilannya bagus, fitur lengkap banget. login juga mudah 👍 "	"tiba² apk bca mobile hilang dari hp saya... trus bagaimana untuk transaksi?"
Cleaning	“mybca tampilannya bagus fitur lengkap banget login juga mudah”	"tiba apk bca mobile hilang dari hp saya trus bagaimana untuk transaksi"
Tokenisasi	[“bagus”,”fitur”,”lengkap”,”login ”,”mudah”]	[“tiba”,”apk”,”hilang”,”hp”,”trus”,”transaksi”]
Stopword Removal	[“bagus”,”fitur”,”lengkap”,”login ”,”mudah”]	[“apk”,”hilang”,”hp”,”trus”,”transaksi”]
Hasil Akhir	“bagus fitur lengkap login mudah”	“apk hilang hp trus transaksi”

Hasil *preprocessing* menunjukkan bahwa proses *stopword removal* secara signifikan mereduksi Panjang teks dengan menghapus kata-kata umum yang tidak mengandung informasi sentimen, sementara *stemming* menggunakan Sastrawi berhasil menormalisasi variasi morfologis bahasa Indonesia ke bentuk dasar yang lebih konsisten.

Pelabelan sentimen dilakukan secara otomatis menggunakan metode *lexicon-based* dengan kamus InSet terhadap seluruh data hasil *preprocessing*. Distribusi hasil pelabelan untuk masing-masing aplikasi disajikan pada Tabel 4.

Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473

Tabel 4. Distribusi Hasil Pelabelan Sentimen

Kelas Sentimen	MyBCA	%	BCA Mobile	%
Positif (1)	2.131	54,31%	2.487	63,70%
Negatif (-1)	1.360	34,66%	980	25,10%
Netral (0)	433	11,03%	437	11,19%
Total	**3.924**	**100%**	**3.904**	**100%**

Berdasarkan Tabel 4, kelas positif mendominasi pada kedua dataset, namun dengan proporsi yang berbeda. BCA Mobile memiliki proporsi positif lebih tinggi (63,70%) dibandingan MyBCA (54,31%), sementara MyBCA memiliki proporsi negatif yang lebih besar (34,66%) dibandingkan BCA Mobile (25,10%). Kelas netral pada kedua aplikasi berada pada kisaran yang relatif setara, yakni 11,03% untuk MyBCA dan 11,19% untuk BCA Mobile. Ketidakseimbangan distribusi antarkelas, khususnya minimnya data berlabel netral dibandingan dua kelas lainnya, berpotensi memengaruhi kinerja model pada kelas tersebut.

Setelah pelabelan selesai, data direpresentasikan secara numerik menggunakan TF-IDF dengan parameter *max_features*=5000 yang membatasi jumlah fitur pada 5.000 kata dengan bobot tertinggi. Data kemudian dibagi dengan rasio 80:20 menggunakan *train_test_split* dengan *random_state*=42. Rincian pembagian data dan hasil pelatihan model disajikan pada Tabel 5.

Tabel 5. Rincian Pembagian Data dan Pelatihan Model

Keterangan	MyBCA	BCA Mobile
Total data berlabel	3.924	3.904
Data *Training* (80%)	3.139	3.123
Data *Testing* (20%)	785	781
Jumlah fitur *(vocabulary)*	3.678 kata	3.684 kata
Algoritma	*Multinomial Naïve Bayes*	*Multinomial Naïve Bayes*

Model *Multinomial Naïve Bayes* dilatih pada data *training* masing-masing aplikasi secara terpisah. Perbedaan jumlah fitur antara kedua model (3.678 vs 3.684) mencerminkan perbedaan kekayaan kosakata pada masing-masing dataset yang dipengaruhi oleh variasi gaya penulisan pengguna kedua aplikasi.

Kinerja model *Naïve Bayes* dievaluasi menggunakan data *testing* yang tidak digunakan selama proses pelatihan. Hasil evaluasi menggunakan empat metrik standar disajikan pada Tabel 6, sedangkan kinerja model per kelas sentimen disajikan pada Tabel 7.

Tabel 6. Hasil Evaluasi Model *Naive Bayes*

Metrik	MyBCA	BCA Mobile
Accuracy	76,56%	73,62%
Precision	79,23%	78,02%
Recall	76,56%	73,62%
F1-*Score*	72,74%	67,97%

Tabel 7. *Classification Report* per Kelas Sentimen

Kelas	MyBCA *Precision*	MyBCA *Recall*	MyBCA F1	BCA Mobile *Precision*	BCA Mobile *Recall*	BCA Mobile F1
Positif	0,75	0,92	0,83	0,71	0,99	0,83
Negatif	0,79	0,75	0,77	0,92	0,39	0,54
Netral	1,00	0,06	0,11	0,86	0,07	0,13

Berdasarkan Tabel 6, model MyBCA menghasilkan kinerja yang lebih baik dibandingkan model BCA Mobile pada seluruh metrik evaluasi, dengan *accuracy* 76,56 berbanding 73,62%.

Perbedaan *accuracy* ini pada awalnya tampak berlawanan dengan intuisi, mengingat BCA Mobile memiliki proporsi kelas mayoritas (positif) yang lebih dominan (63,70%) dibandingkan MyBCA (54,31%), yang secara teoretis seharusnya memudahkan model mencapai akurasi tinggi karena sebagian besar data

Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473

terkonsentrasi pada satu kelas. Namun dominasi kelas mayoritas yang terlalu besar justru berdampak sebaliknya terhadap performa keseluruhan. Pada model BCA Mobile, kelas positif memang dikenali hampir sempurna (*recall* 0,99), tetapi kelas negatif yang menempati 25,10% data mengalami penurunan *recall* yang tajam menjadi hanya 0,39, artinya sebagain besar ulasan negatif justru diklasifikasikan sebagai positif. Karena kelas negatif tetap menyumbang seperempat dari keseluruhan data uji, kegagalan mengenali kelas ini secara signifikan menurunkan *accuracy* total, meskipun kelas mayoritas dikenali dengan sangat baik.

Sebaliknya, distribusi kelas pada MyBCA relatif lebih seimbang antara positif (54,31%) dan negatif (34,66%), sehingga proses estimasi probabilitas prior pada algoritma *Naïve Bayes* tidak terlalu timpang antara kedua kelas tersebut. Kondisi ini memungkinkan model mempelajari pola kosakata kedua kelas secara lebih proporsional, yang tercermin dari *recall* kelas negatif MyBCA yang jauh lebih tinggi (0,75) dibandingkan BCA Mobile (0,39). Dengan demikian, perbedaan *accuracy* antara kedua model bukan disebabkan oleh perbedaan jumlah fitur atau kompleksitas data melainkan lebih dipengaruhi oleh tingkat keseimbangan distribusi kelas pada masing-masing dataset.

Berdasarkan Tabel 7, terdapat beberapa temuan penting yang perlu diperhatikan. **Pertama**, kelas positif memperoleh F1-*score* tertinggi pada kedua model yakni 0,83, yang menunjukkan bahwa model paling andal dalam mengenali ulasan bersentimen positif. Hal ini sejalan dengan dominasi kelas positif pada dataset yang menyediakan cukup sampel pelatihan bagi model. **Kedua**, kelas negatif pada model MyBCA menunjukkan keseimbangan yang baik antara *precision* (0,79) dan *recall* (0,75), dengan F1-*score* 0,77, sedangkan pada model BCA Mobile terjadi kesenjangan signifikan antara *precision* (0,92) dan *recall* (0,39) yang menghasilkan F1-*score* hanya 0,54. Rendahnya *recall* negatif pada BCA Mobile mengindikasikan model cenderung memprediksi ulasan negatif sebagai positif, yang dipengaruhi oleh dominasi kelas positif pada dataset BCA Mobile.

Fenomena ini dapat dijelaskan lebih lanjut melalui mekanisme kerja *Multinomial Naïve Bayes*. Algoritma ini mengestimasi probabilitas posterior suatu kelas berdasarkan probabilitas prior (proporsi kelas pada data *training*) dikalikan dengan probabilitas *likelihood* kemunculan setiap kata pada kelas tersebut. Ketika proporsi data *training* kelas positif jauh lebih besar dibandingkan kelas negatif (data *training* BCA Mobile: kelas negatif hanya berkontribusi sekitar 784 dari 3.123 sampel, dibandingkan MyBCA yang memiliki sekitar 1.088 dari 3.139 sampel untuk kelas yang sama), probabilitas prior kelas positif menjadi jauh lebih besar sejak awal. Akibatnya, untuk ulasan yang mengandung campuran kata netral-negatif atau kata negatif dengan intensitas lemah, probabilitas posterior kelas positif tetap lebih tinggi karena pengaruh prior yang dominan, sehingga ulasan tersbeut salah diklasifikasikan sebagai positif.

Selain itu, asumsi independensi antar-fitur pada *Naïve Bayes* membuat model rentan terhadap ulasan negatif yang tetap mengandung sejumlah kata bernuansa positif (misalnya keluhan yang diselipi pujian pada fitur tertentu). Pada kondisi prior yang sudah timpang, keberadaan kata-kata positif tersbeut cukup untuk menggeser keputusan klasifikasi ke arah kelas mayoritas.

Ketiga, kelas netral menunjukkan kinerja terendah pada kedua model dengan F1-*score* 0,11 (MyBCA) dan 0,13 (BCA Mobile). Mesikipun nilai *precision* kelas netral tinggi pada MyBCA 1,00 dan 0,86 pada BCA Mobile, nilai *recall* yang sangat rendah (0,06 dan 0,07) mengindikasikan model hampir tidak mampu mendeteksi ulasan netral secara tepat. Kondisi ini merupakan dampak langsung dari ketidakseimbangan kelas *(class imbalance)*, di mana jumlah data netral (~11%) jauh lebih sedikit dibandingkan kelas positif dan negatif sehingga model tidak memiliki cukup sampel untuk mempelajari pola ulasan netral secara memadai.

Pola *precision* tinggi namun *recall* sangat rendah pada kelas netral mengindikasikan karakteristik prediksi model yang sangat konservatif terhadap kelas ini: model jarang sekali memutuskan suatu ulasan sebagai netral, tetapi pada saat model melakukannya, prediksinya cenderung tepat. Hal ini dapat dijelaskan melalui dua faktor yang saling berkaitan. Pertama, dari sisi jumlah data, kelas netral pada data *training* hanya berkisar 346 sampel untuk MyBCA (80% dari 433) dan sekitar 350 sampel untuk BCA Mobile (80% dari 437), jauh lebih sedikit dibandingkan kelas positif maupun negatif. Dengan jumlah sampel yang terbatas ini, estimasi probabilitas kemunculan kata pada kelas netral menjadi kurang stabil, dan *smoothing* yang digunakan *Naïve Bayes* untuk menangani kata yang tidak muncul pada suatu kelas (umumnya *Laplace/add-one smoothing)* justru mendominasi estimasi probabilitas untuk kelas dengan sampel sedikit ini, sehingga probabilitas posterior kelas netral jarang menjadi yang tertinggi dibandingkan dua kelas lainnya.

Kedua, karakteristik data ulasan yang dilabeli netral melalui pendekatan *lexicon-based* umumnya merupakan ulasan dengan skor sentimen mendekati nol, baik karena teksnya pendek dan minim kata

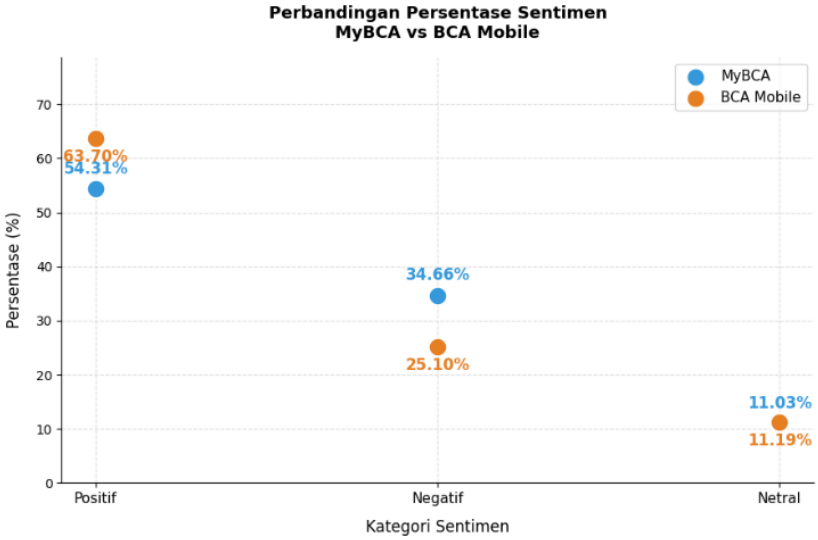
Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473

bermuatan sentimen (seperti keluhan teknis singkat: “aplikasi *force close*”, “tidak bisa login”), maupun karena mengandung kombinasi kata positif dan negatif yang saling meniadakan skornya. Pada kedua kondisi tersebut, kosakata yang muncul pada ulasan netral pada dasarnya juga muncul pada ulasan positif maupun negatif, sehingga tidak terdapat kosakata yang secara khas membedakan kelas netral dari kedua kelas lainnya. Akibatnya, model yang mengandalkan frekuensi kemunculan kata sebagai dasar klasifikasi kesulitan menemukan sinyal pembeda yang cukup kuat untuk mengenali kelas ini, dan pada situasi ambigu tersebut probabilitas posterior kelas netral secara sistematis kalah dibandingkan kelas positif atau negatif yang memiliki dukungan sampel dan sinyal kosakata jauh lebih besar.

Untuk mengevaluasi bagaimana model memang distribusi sentimen dibandingkan label aktualnya, dilakukan analisis terhadap hasil prediksi model *Naïve Bayes* pada data *testing* masing-masing aplikasi. Perbandingan distribusi sentimen antara MyBCA dan BCA Mobile disajikan pada Tabel 8 dan Gambar 1.

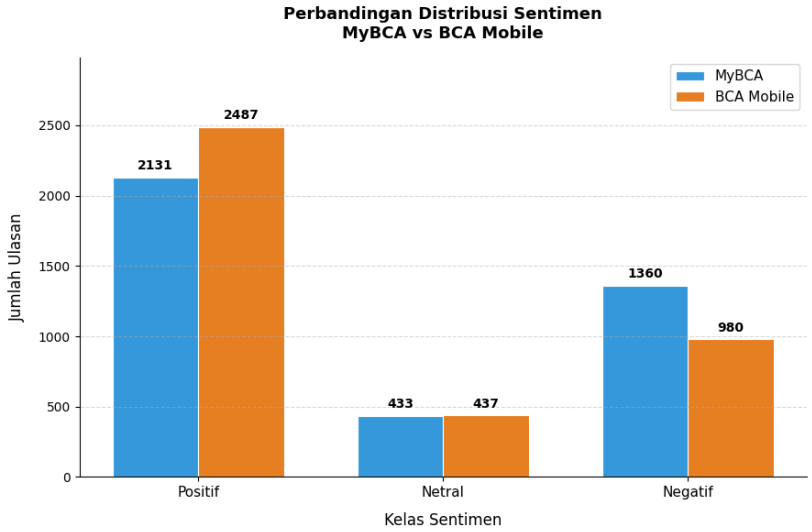
Tabel 8. Perbandingan Distribusi Hasil Prediksi Model Naïve Bayes pada Data *Testing* MyBCA dan BCA Mobile

Kelas Sentimen	MyBCA (Prediksi)	%	BCA Mobile (Prediksi)	%	Selisih (%)
Positif	523	66,62%	691	88,48%	+21,86 (BCA Mobile)
Negatif	5	0,64%	7	0,90%	+0,26 (BCA Mobile)
Netral	257	32,74%	83	10,63%	+22,11 (MyBCA)
Total	**785**	**100%**	**781**	**100%**	



Gambar 2. Scatter Plot Perbandingan Distribusi Sentimen MyBCA vs BCA Mobile

Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473



Gambar 3. Diagram Batang Distribusi Sentimen

Berdasarkan Tabel 8, distribusi hasil prediksi model pada data *testing* menunjukkan pola yang berbeda dari distribusi label aktual pada Tabel 4. Model MyBCA memprediksi 66,62% ulasan sebagai positif, meningkat dari proporsi label aktualnya sebesar 54,31%, sementara model BCA Mobile memprediksi 88,48% ulasan sebagai positif, jauh lebih tinggi dibandingkan proporsi label aktualnya sebesar 63,70%. Peningkatan proporsi prediksi positif pada kedua model ini konsisten dengan temuan Tabel 7, di mana *recall* kelas negatif dan netral tergolong rendah yang artinya sebagian ulasan yang sebenarnya berlabel negatif atau netral ikut terklasifikasi sebagai positif oleh model. Pola ini paling terlihat pada model BCA Mobile: proporsi prediksi negatif turun tajam dari 25,10% (label aktual) menjadi hanya 10,63% (hasil prediksi), sejalan dengan *recall* kelas negatif BCA Mobile yang hanya 0,39 pada Tabel 7. Sebaliknya, penurunan proporsi negatif pada model MyBCA lebih moderat, dari 34,66% (label aktual) menjad 32,74% (hasil prediksi), sejalan dengan *recall* kelas negatif MyBCA yang jauh lebih baik (0,75). Perbedaan besaran pergeseran ini mengkonfirmasi bahwa tingkat keseimbangan distribusi kelas pada data *training* berdampak langsung pada seberapa jauh model "menggeser" prediksi ke kelas mayoritas.

Sementara itu, kelas netral pada hasil prediksi kedua model tereduksi drastis dari sekitar 11% pada data aktual menjadi kurang dari 1% pada hasil prediksi (0,64% untuk MyBCA dan 0,90% untuk BCA Mobile), yang menegaskan temuan sebelumnya bahwa model hampir sepenuhnya gagal mengenali kelas ini akibat kombinasi antara jumlah sampel *training* yang minim dan tumpang tindih kosakata dengan kelas positif maupun negatif.

Temuan ini konsisten dengan hasil penelitian Yoman[6] yang juga menemukan perbedaan proporsi sentimen antara BCA Mobile dan MyBCA, meskipun dengan metode *lexicon-based* tanpa evaluasi model terstruktur. Penelitian ini memperkuat temuan tersebut dengan pendekatan yang lebih komprehensif melalui penggunaan algoritma Naïve Bayes yang terukur kinerjanya. Perbedan distribusi sentimen antara kedua aplikasi dapat diinterpretasikan dalam konteks karakteristik masing-masing platform: BCA Mobile sebagai aplikasi yang lebih matang secara teknis dengan basis pengguna yang lebih terbiasa, sementara MyBCA sebagai platform yang lebih baru masih dalam fase penyempurnaan fitur dan stabilitas layanan yang mencerminkan lebih tingginya proporsi keluhan pengguna pada periode pengambilan data.

Secara metodologis, hasil penelitian ini menunjukkan bahwa kombinasi metode *lexicon-based* untuk pelabelan otomatis dan algoritma Naïve Bayes untuk klasifikasi merupakan pendekatan yang efektif dan efisien untuk analisis sentimen ulasan aplikasi *mobile banking* berbahasa Indonesia, dengan tingkat akurasi yang kompetitif dibandingkan penelitian terdahulu menggunakan objek yang serupa, Al-Hakim[12] dan Marpaung[13]. Keterbatasan utama penelitian ini terletak pada rendahnya kinerja model pada kelas netral akibat *class imbalance*, serta minimnya kosakata khas yang membedakan ulasan netral dari ulasan berpolaritas, sebagaimana diuraikan sebelumnya, yang dapat diatasi pada penelitian selanjutnya melalui teknik *oversampling* seperti SMOTE atau penerapan bobot kelas *(class weighting)* dalam proses pelatihan model.

Jurnal Sistem Informasi dan Sistem Komputer, Vol. 11, No. 2, Juli 2026
ISSN: 2715-906X (Online)
10.51717/simkom.v11i2.1473

4. Kesimpulan dan Saran

4.1. Kesimpulan

Penelitian ini menunjukkan bahwa kombinasi metode *Lexicon-Based* untuk pelabelan otomatis dan algoritma Naive Bayes untuk klasifikasi dapat digunakan secara efektif dan efisien untuk menganalisis serta membandingkan sentimen pengguna aplikasi *mobile banking*, tanpa memerlukan proses pelabelan manual maupun sumber daya komputasi besar seperti pada pendekatan *deep leaning*. Pendekatan ini menjawab kesenjangan yang teridentifikasi pada penelitian sebelumnya, yaitu ketiadaan metode klasifikasi yang sederhana namun tetap transparan dalam proses pelabelannya untuk membandingkan sentimen pengguna MyBCA dan BCA Mobile.

Secara substantif, hasil penelitian mengonfirmasi bahwa kedua aplikasi memiliki persepsi pengguna yang secara konsisten berbeda yakni BCA Mobile menunjukkan tingkat kepuasan pengguna yang lebih tinggi dibandingkan MyBCA, tercermin dari proporsi sentimen positif yang lebih dominan dan proporsi keluhan yang lebih rendah. Temuan ini konsisten dengan karakteristik kedua platform, BCA Mobile sebagai aplikasi yang lebih matang dengan basis pengguna yang lebih terbiasa, sedangkan MyBCA sebagai platform yang relatif baru dan masih dalam tahap penyempurnaan fitur serta stabilitas layanan. Bagi BCA, temuan ini dapat menjadi masukan untuk memprioritaskan perbaikan pada aspek-aspek yang paling banyak dikeluhkan pengguna MyBCA, guna mempercepat kesetaraan kualitas layanan antara kedua aplikasi.

Dari sisi evaluasi model, penelitian ini juga mengungkap bahwa performa klasifikasi *Naïve Bayes* sangat dipengaruhi oleh tingkat keseimbangan distribusi kelas pada data *training*, bukan semata oleh jumlah maupun kekayaan kosakata *dataset*. Model dengan distribusi kelas yang lebih seimbang terbukti mampu mengenali kelas minoritas (negatif) jauh lebih baik dibandingkan model dengan distribusi yang lebih timpang, sementara kelas netral secara konsisten menjadi kelas yang paling sulit dikenali pada kedua model akibat keterbatasan jumlah sampel dan tumpang tindih karakteristik kosakata dengan kelas lain. temuan ini memberikan konstribusi metodologis bagi penelitian analisis sentimen berbasis algoritma klasifikasi konvensional, khususnya terkait pentingnya mempertimbangkan strategi penanganan *class imbalance* sejak tahap perancangan penelitian, bukan hanya sebagai catatan keterbatasan di akhir.

4.2. Saran

Berdasarkan keterbatasan yang ditemukan, penelitian ini memiliki keterbatasan pada rendahnya kemampuan model mengenali kelas netral, yang berpotensi memengaruhi akurasi interpretasi pada ulasan dengan sentimen ambigu. Penelitian selanjutnya disarankan untuk menerapkan teknik penanganan *class imbalance* seperti *oversampling* (misalnya SMOTE) atau pembobotan kelas *(class weighting)*, serta mempertimbangkan perluasan data pelabelan pada kelas netral agar model memiliki representasi yang lebih memadai. Selain itu, penelitian lanjutan juga dapat mengeksplorasi kombinasi pendekatan *lexicon-based* dengan algoritma klasifikasi lain untuk mengevaluasi sejauh mana temuan mengenai pengaruh distribusi kelas ini berlaku secara umum, tidak terbatas pada *Naïve Bayes*.

Daftar Pustaka

[1] M. Walfajri, “Digital banking tumbuh di tengah pandemi, masyarakat kian sering bertransaksi online,” Kontan.Co.Id-Jakarta. [Online]. Available: https://newssetup.kontan.co.id/news/digital-banking-tumbuh-di-tengah-pandemi-masyarakat-kian-sering-bertransaksi-online

[2] G. Buana, “Lebih Baik myBCA atau BCA Mobile Kenali Perbedaannya untuk Pilihan Terbaik,” *Media Indonesia*, 2025. [Online]. Available: https://mediaindonesia.com/ekonomi/804680/lebih-baik-mybca-atau-bca-mobile-kenali-perbedaannya-untuk-pilihan-terbaik#goog_rewarded

[3] A. Nadira, N. Y. Setiawan, and W. Purnomo, “Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Naïve Bayes Dengan Kamus Inset,” *Indexia*, vol. 5, no. 01, p. 35, 2023, doi: 10.30587/indexia.v5i01.5138.

[4] S. A. S. Mola, S. N. R. Djawa, and A. Y. Mauko, *Text Mining: Analisis Sentimen dengan Naïve Bayes*. Bandung, 2025.

[5] A. Wibowo, Firman Noor Hasan, L. A. Ramadhan, Rika Nurhayati, and Arief Wibowo, “Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier,” *J. Asiimetrik J. Ilm. Rekayasa Inov.*, pp. 239–248, 2022, doi: 10.35814/asiimetrik.v4i1.3577.

[6] J. P. Yoman, C. Cok, K. Laigusten, and G. Zovintho, "Analisis Sentimen Ulasan Perbedaan Aplikasi BCA Mobile dengan MYBCA di Playstore Menggunakan Metode Lexicon," *JDMIS J. Data Min. Inf. Syst.*, vol. 3, no. 2, pp. 64–76, 2025, doi: 10.54259/jdmis.v3i2.4236.

[7] S. A. Putri, K. Ditha Tania, N. Kawadha, and P. Gumay, "Knowledge Discovery Through Sentiment Analysis and Topic Modeling of BCA Mobile and MyBCA," *J. Math. Comput. Stat.*, vol. 8, no. 2, pp. 669–682, 2025, [Online]. Available: https://journal.unm.ac.id/index.php/JMATHCOS

[8] Y. Findawati and M. A. Rosid, *Buku Ajar Text Mining*. 2020.

[9] D. Musfiroh, U. Khaira, P. E. P. Utomo, and T. Suratno, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 24–33, 2021, doi: 10.57152/malcom.v1i1.20.

[10] V. Meida Hersianty, E. Larasati Amalia, D. Puspitasari, and D. Wahyu Wibowo, "Penerapan Algoritma Tf-Idf Dan Cosine Similarity Dalam Sistem Rekomendasi Lowongan Pekerjaan," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 1619–1625, 2025, doi: 10.36040/jati.v9i1.12406.

[11] Z. Setiawan *et al.*, *Buku Ajar Data Mining*, vol. 4, no. 1. 2023.

[12] M. G. Al Hakim and F. Irwiensyah, "Analisis Sentimen Terhadap Ulasan Pengguna Pada Aplikasi BCA Mobile Menggunakan Metode Naïve Bayes," *J. Inf. Syst. Res.*, vol. 5, no. 4, pp. 911–921, 2024, doi: 10.47065/josh.v5i4.5343.

[13] J. A. Marpaung, M. Devega, and Yuhelmi, "Analisis Sentimen Kepuasan Pengguna Aplikasi BCA Mobile menggunakan Metode Naïve Bayes dan Support Vector Machine (SVM)," *Prosiding-Seminar Nas. Teknol. Inf. Ilmu Komput.*, vol. 3, no. 1, pp. 249–261, 2024, [Online]. Available: https://journal.unilak.ac.id/index.php/Semaster/article/view/24124