

# Analisis Sentimen Menggunakan Algoritma Naïve Bayes Pada Aplikasi m-BCA berdasarkan Ulasan Pengguna di Google Play Store

Irma Rahmawati<sup>1</sup>, Tiara Rika Fitriani<sup>2</sup>, Ajif Yunizar Pratama Yusuf<sup>3</sup>

Informatika, Ilmu Komputer, Universitas Bhayangkara Jakarta Raya

Email : [irmarahma119@gmail.com](mailto:irmarahma119@gmail.com) , [rikatiara56@gmail.com](mailto:rikatiara56@gmail.com) , [ajif.yunizar@dsn.ubharajaya.ac.id](mailto:ajif.yunizar@dsn.ubharajaya.ac.id)

## ABSTRAKSI

m-BCA (*Mobile Banking*) adalah layanan perbankan dari PT.Bank Central Asia,Tbk yang memungkinkan akses langsung melalui telepon seluler. Meskipun populer, pengguna sering mengalami pengalaman negatif, seperti error dalam mengakses aplikasi dan keluhan terkait transaksi. Meskipun terdapat opini negatif, opini positif di Play Store juga ditemukan. Penelitian ini menggunakan metode Naïve Bayes untuk menganalisis sentimen pengguna BCA Mobile berdasarkan ulasan di Google Play Store. Hasil analisis diharapkan dapat membantu BCA meningkatkan kualitas layanan mereka. Penelitian terkait menggunakan berbagai metode, dan penelitian ini memilih Naïve Bayes karena kecepatan dan efisiensinya. Tahap penelitian mencakup pengumpulan data dari Google Play Store, preprocessing data, dan penerapan algoritma Naïve Bayes, Random Forest, serta Logistic Regression. Penelitian ini menyimpulkan bahwa sebagian besar ulasan memiliki orientasi negatif dan model Naïve Bayes menonjol dengan tingkat akurasi yang paling tinggi, mencapai presisi 84%, recall 82%, F1-score 81%, dan tingkat akurasi sebesar 82%.

**Kata Kunci:** m-BCA, *Mobile Banking*, Analisis Sentimen, Naïve Bayes, Google Play Store, Kualitas Layanan.

## ABSTRACT

m-BCA (*Mobile Banking*) is a banking service provided by PT Bank Central Asia Tbk that enables direct access through mobile phones. Despite its popularity, users often encounter negative experiences, such as errors in accessing the application and complaints related to transactions. While negative opinions exist, positive reviews are also found on the Play Store. This research employs the Naïve Bayes method to analyze user sentiment regarding m-BCA based on reviews from the Google Play Store. The results of the analysis are expected to assist m-BCA in improving the quality of their services. This study uses the Naive Bayes method among a range of options in related research, mostly because of its rapidity and effectiveness. The research stages involve data collection from the Google Play Store, data preprocessing, and the implementation of Naïve Bayes, Random Forest, and Logistic Regression algorithms. The study concludes that the majority of reviews have a negative orientation, and the Naïve Bayes model stands out with the highest accuracy rates, achieving precision of 84%, recall of 82%, F1-score of 81%, and an accuracy rate of 82%.

**Keywords:** M-BCA, *Mobile Banking*, *Sentiment Analysis*, *Naïve Bayes*, *Google Play Store*, *Service Quality*.

## Penulis Korespondensi

Irma Rahmawati

Tanggal Submit : 01/12/2023

Tanggal Diterima : 29/03/2024

Tanggal Terbit : 30/03/2024

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license



**Copyright:** © 2023 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 (CC BY-NC-SA 4.0) International License (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

Publisher's Note: JPPM stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## I. PENDAHULUAN

m-BCA (*Mobile Banking*) merupakan layanan perbankan yang ditawarkan oleh PT.Bank Central Asia

Tbk ("BCA") yang memungkinkan Nasabah untuk mengaksesnya secara langsung melalui telepon seluler atau handphone. *Mobile Banking* menjadi salah satu

layanan populer diantara layanan internet lainnya[1]. Namun demikian aplikasi ini memiliki pengalaman negatif dari penggunaannya. Seperti halnya *error* dalam mengakses aplikasi dengan indikator tidak menyala hijau, tidak bisa di gunakan setiap log in selalu ada keterangan transaksi tidak bisa di gunakan silahkan coba beberapa saat padahal baru masukin kode akses, komplain setor tunai tidak masuk tapi penyelesaian lama.

Selain opini terkait buruknya layanan pada m-BCA, di play store juga banyak beredar opini yang bersifat positif, khususnya terkait layanan yang diberikan. Namun, seringkali tidak muncul dipermukaan karena hampir sebagian besar pengguna lebih fokus pada opini negatif dibandingkan opini positif. Padahal jika kedua jenis opini ini dianalisis bisa memungkinkan diperoleh informasi utuh terkait opini masyarakat pada pengguna m-BCA. Dengan demikian, hasil analisis opini dapat digunakan untuk meningkatkan pelayanan kepada nasabah dengan lebih baik.

Terdapat banyak teknik yang digunakan untuk analisis sentimen, salah satunya adalah Naïve Bayes. Algoritma ini merupakan salah satu algoritma yang termasuk kedalam klasifikasi. Metode ini merupakan pendekatan terbaik dalam klasifikasi yang telah diperkenalkan oleh Thomas Bayes, di mana kita dapat memproyeksikan kemungkinan kejadian di masa mendatang berdasarkan informasi yang ada dari masa sebelumnya. Naive Bayes menggunakan metode statistik sederhana yang mengacu pada teorema Bayes dengan mengasumsikan independensi antara fitur-fitur yang ada terhadap kelas tertentu, tanpa adanya keterkaitan antar fitur tersebut[2]. Teknik ini banyak digunakan untuk menyelesaikan banyak masalah, seperti prediksi ketidakhadiran[3], prediksi ketepatan waktu[4], mendeteksi virus komputer[5], dan penentuan kelayakan penerima kartu Jakarta pintar plus[6].

Penelitian ini dilakukan dengan tujuan memberikan masukan kepada pengguna aplikasi m-BCA agar mereka dapat menggali persepsi pelanggan melalui analisis pendapat yang terdapat di Google Play Store. Melalui proses ini, pihak m-BCA diharapkan dapat meningkatkan kualitas layanan mereka untuk calon pengguna atau nasabah, sesuai dengan kebutuhan yang diidentifikasi.

## II. PENELITIAN YANG TERKAIT

Pembahasan terkait analisis sentimen telah banyak dilakukan oleh peneliti dan menjadi topik yang sangat populer digunakan untuk penelitian. Beberapa diantaranya menggunakan Metode Support Vector Machine[7], latent dirichlet allocation[8], Sistematis Literature Review[9], Deep Neural Network[10], dan Long Short Term Memory[11]. Akan tetapi penggunaan pendekatan seperti yang dilakukan oleh [7 – 11] memiliki tingkat akurasi yang baik jika memiliki dataset

yang banyak. Oleh karena itu, penelitian ini menggunakan pendekatan lainnya yang dapat mengatasi kelemahan pada algoritma penelitian [7-11], yaitu menggunakan Naïve Bayes. Naïve Bayes merupakan suatu klasifikasi yang berdasarkan pada teorema Bayes dengan memiliki tujuan untuk menghitung peluang pada tiap kelas serta mempunyai asumsi bahwa hubungan antar kelas adalah independen[12].

Beberapa diantaranya digunakan untuk mengetahui polarity pada sebuah teks apakah *Extremly positive, positive, neutral, negative* atau *Extremly negative*[13]. Beberapa peneliti sebagian besar menggunakan algoritma Logistic Regression dengan memanfaatkan media sosial untuk mendapatkan ulasan yang nantinya di klasifikasikan. Pada penelitian sebelumnya dilakukan menggunakan algoritma Logistic Regression dengan jumlah 870 data *tweet*, namun mendapatkan nilai akurasi yang kurang baik karena pelabelan dilakukan secara manual. Proses klasifikasi pada penelitian ini menggunakan metode *Naïve Bayes*. *Naïve Bayes* adalah sebuah metode klasifikasi yang berasal dari dasar-dasar teorema Bayes. Pendekatan ini mengandalkan prinsip probabilitas dan statistik yang pertama kali dijabarkan oleh ilmuwan Inggris bernama Thomas Bayes. Metodenya berkaitan dengan memproyeksikan kemungkinan kejadian di masa depan berdasarkan pengalaman dari masa sebelumnya, oleh karena itu dikenal sebagai teorema Bayes. Namun metode ini rentan terhadap asumsi setiap variabel independen membuat berkurangnya akurasi karena ada korelasi antara variabel yang satu dengan variabel yang lain. Dalam hal tersebut untuk mencapai nilai akurasi yang terbesar dapat digunakan metode *confusion matrix*[1]. Tanthy Tawaqalia mengungkapkan bahwa data pengujian dengan metode Naïve Bayes menghasilkan akurasi yang tinggi yaitu sebesar 91.48%[14].

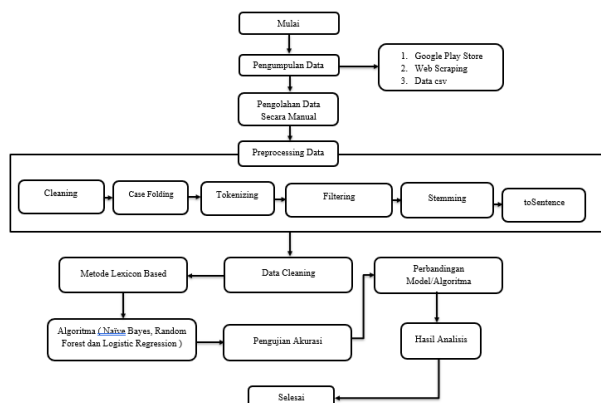
Berdasarkan dari masalah diatas dapat disimpulkan bahwa menggunakan algoritma Naïve Bayes dengan perhitungan cepat dan efisien. Algoritma ini sudah banyak dibuktikan dan dibandingkan dengan algoritma lain, sebab menghasilkan akurasi baik.

## III. METODE PENELITIAN

### A. Tahapan Penelitian

Penelitian ini diawali dengan tahap pengumpulan data dengan menggunakan data dari *review google playstore*, dimana data yang diperoleh sebanyak 1499 data ulasan dalam bentuk csv. Selanjutnya data diolah secara manual dengan dibagi menjadi dua class yaitu berdasarkan positif, dan negatif. Dilanjutkan dengan preprocessing data, data cleaning, metode lexicon based, pemodelan sistem yang mencakup implementasi algoritma Naïve Bayes, Random Forest dan Logistic Regression. Pada tahap terakhir dilakukan proses

pengujian akurasi, perbandingan dan hasil klasifikasi berdasarkan algoritma yang digunakan bisa dilihat pada gambar 1.



Gambar 1 Proses Pengujian Akurasi

### B. Preprocessing

Pada langkah ini, melanjutkan ke fase selanjutnya dilakukan dengan mengurangi fitur yang tidak memiliki dampak signifikan pada klasifikasi. Data yang digunakan dalam prosedur ini masih belum diolah dan akan diperbaiki menggunakan fungsi Python yang dikenal sebagai "*cleantext*" untuk menyaring data yang tidak relevan atau tidak akurat. Setelah itu, huruf kapital akan diubah menjadi huruf kecil melalui suatu proses yang disebut "*casefolding*". Selanjutnya, proses "*tokenization*" akan memecah kalimat menjadi kata-kata yang membentuknya, memisahkan kalimat dengan spasi atau tanda baca. Setelah itu, proses penyaringan digunakan untuk menghilangkan kata-kata yang tidak memberikan kontribusi dari data yang telah diproses sebelumnya. Langkah berikutnya adalah "*stemming*", di mana kata-kata diubah menjadi bentuk yang sesuai dengan aturan tertentu. Pada tahap terakhir, metode "*tosentences*" digunakan untuk mengubah daftar kata menjadi kalimat secara berurutan.

### C. Naïve Bayes

Algoritma Naïve Bayes adalah suatu metode klasifikasi yang memperhitungkan nilai probabilitas statistik. Dengan demikian, prediksi kelas keanggotaan dari suatu data baru dilakukan dengan memproyeksikan probabilitas keanggotaan data tersebut[15]. Persamaan 1 merupakan teorema bayes.

$$P(c | X) = \frac{P(X|c) \cdot P(c)}{P(X)} \quad (1)$$

Keterangan:

$P(c | X)$  = Probabilitas hipotesis berdasar kondisi (posteriori probability)

$P(c)$  = Probabilitas hipotesis (prior probability)

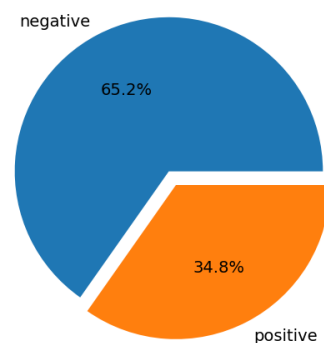
$P(X|c)$  = Probabilitas berdasarkan kondisi pada hipotesis

$P(X)$  = Probabilitas c

## IV. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan dengan menggunakan data dari review google playstore dan mendapatkan 1499 data. Dari data yang terkumpul dibagi menjadi dua kelas yaitu positif dan negatif. Data yang digunakan menggunakan ulasan berbahasa Indonesia mengenai ulasan pengguna aplikasi m-BCA. Setelah itu untuk melakukan membersihkan data yang tidak penting, simbol, angka, tanda baca digunakan proses *preprocessing* seperti *clenigtext*, *casefolding*, *tokenization*, *filteringtext*, *stemming* dan *tosentences*[16]. Proses Selanjutnya adalah data *cleaning*, proses ini data mentah yang di dapatkan dari Google Play Store akan di modifikasi, agar data yang dihasilkan lebih sistematis dan rapi. Kemudian dilakukannya proses metode *Lexicon Based* untuk analisis sentimen dari data tersebut, hasil yang didapatkan yaitu 978 kelas negatif, 521 diidentifikasi kelas positif.

Analisis Sentimen berdasarkan ulasan (total = 1499 ulasan)



Gambar 2 Visualisasi Data Ulasan

Pada gambar 2 data yang dihasilkan dari proses tersebut di visualisasikan dalam bentuk persentase, dari 1499 ulasan mendapatkan 65.2% yang mengatakan negatif, 34,8% mengatakan positif. Dapat disimpulkan bahwa setimen terhadap aplikasi m-BCA adalah ulasan negatif.

Setelah tahap visualisasi data, langkah berikutnya adalah membangun model klasifikasi menggunakan tiga algoritma, yaitu Naïve Bayes, Random Forest dan Logistic Regression. Sebelum algoritma belajar dari dataset, data dibagi ke dalam data training dan testing dengan proporsi 50:50% serta data di ekstraksi fitur menggunakan pendekatan *bag-of-words*.

Tabel 1 Perbandingan Model

Pengukuran	Naïve Bayes	Random Forest	Linear Regression
Precision	0.84	0.77	0.82
Recall	0.82	0.75	0.80
F1 - measure	0.81	0.73	0.78
Accuracy	0.82	0.75	0.80

Dari tabel 1 dapat di simpulkan bahwa menggunakan algoritma Naïve Bayes dapat dilakukan pemodelan secara baik dimana mendapatkan nilai akurasi 82%, nilai Precision sebesar 84% dibandingkan dengan penelitian sebelumnya yang menggunakan metode Linear Regression dan Random Forest hanya mendapatkan nilai rata – rata akurasi sebesar 75 – 80%.

## V. KESIMPULAN

Berdasarkan hasil penelitian analisis sentimen menggunakan algoritma Naïve Bayes pada aplikasi Mobile BCA berdasarkan ulasan pengguna play store dapat disimpulkan bahwa para pengguna aplikasi tersebut memberikan ulasan negatif terhadap Mobile Banking BCA. Dari data yang sudah dianalisis yang berjumlah 1499 hasil yang didapatkan yaitu 978 kelas negatif, 521 diidentifikasi kelas positif. Dengan menggunakan algoritma Linear Regression dapat dilakukan pemodelan secara baik dimana mendapatkan nilai akurasi 82%, nilai Precision sebesar 82% dibandingkan dengan penelitian sebelumnya yang menggunakan metode Naïve Bayes dan Random Forest hanya mendapatkan nilai rata – rata akurasi sebesar 46%. Hasil ini hanya mencakup positif dan negatif, namun pada penelitian ini tidak mengatakan bahwa ada pernyataan netral.

## DAFTAR PUSTAKA

- [1] A. Miftahusalam, H. Pratiwii, and I. Slamet, “Perbandingan Metode Random Forest dan Naïve Bayes pada Analisis Sentimen Review Aplikasi BCA Mobile,” pp. 1–8, 2023.
- [2] R. Sari, R. Yulia Hayuningtyas, and S. Nusa Mandiri, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Pada Wisata TMII Berbasis Website,” *IJSE-Indonesian Journal on Software Engineering*, vol. 5, no. 2, pp. 51–60, 2019.
- [3] H. Nalatissifa, W. Gata, S. Diantika, and K. Nisa, “Perbandingan Kinerja Algoritma Klasifikasi Naive Bayes, Support Vector Machine (SVM), dan Random Forest untuk Prediksi Ketidakhadiran di Tempat Kerja,” *Jurnal Informatika Universitas Pamulang*, vol. 5, no. 4, p. 578, Dec. 2021, doi: 10.32493/informatika.v5i4.7575.
- [4] P. C. Algoritma, D. Naïve Bayes Untuk Prediksi Ketepatan Waktu Studi Mahasiswa, J. Nata

- Permana, R. Goejantoro, and S. Prangga, “Comparison Of C4.5 Algorithm and Naïve Bayes for Prediction Of Student Study Timeliness (Case Study: Departement of Statistics Mulawarman University),” *EKSPONENSIAL*, vol. 13, no. 2, pp. 161–170, 2022.
- [5] J. Elektronik *et al.*, “Implementasi Algoritma Naive Bayes Classifier (NBC) Dan Information Gain Untuk Mendeteksi DDoS,” vol. 11, no. 2, pp. 273–282, 2022, [Online]. Available: <https://research.unsw.edu.au/projects/unsw-nb15-dataset>.
- [6] E. Widya Ningsih, “Penerapan Algoritma Naïve Bayes Dalam Penentuan Kelayakan Penerima Kartu Jakarta Pintar Plus,” 2019, doi: 10.31294/jtk.v4i2.
- [7] K. Ivana Ruslim and P. Pandu Adikara, “Analisis Sentimen Pada Ulasan Aplikasi Mobile Banking Menggunakan Metode Support Vector Machine dan Lexicon Based Features,” 2019. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [8] B. A. Tondang, Muhammad Rizqan Fadhil, Muhammad Nugraha Perdana, Akhmad Fauzi, and Ugra Syahda Janitra, “Analisis pemodelan topik ulasan aplikasi BNI, BCA, dan BRI menggunakan latent dirichlet allocation,” *INFOTECH: Jurnal Informatika & Teknologi*, vol. 4, no. 1, pp. 114–127, Jun. 2023, doi: 10.37373/infotech.v4i1.601.
- [9] Y. A. Singgalen, “Pemilihan Metode dan Algoritma dalam Analisis Sentimen di Media Sosial : Systematic Literature Review,” *Journal of Information Systems and Informatics*, vol. 3, no. 2, 2021, [Online]. Available: <http://journal-isi.org/index.php/isi>
- [10] E. Y. Hidayat, R. W. Hardiansyah, and A. Affandy, “Analisis Sentimen Twitter untuk Menilai Opini Terhadap Perusahaan Publik Menggunakan Algoritma Deep Neural Network,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 2, pp. 108–118, Sep. 2021, doi: 10.25077/teknosi.v7i2.2021.108-118.
- [11] J. Nurvania and K. Muslim Lhaksamana, “Analisis Sentimen Pada Ulasan di TripAdvisor Menggunakan Metode Long Short-Term Memory (LSTM),” 2021.
- [12] J. Sains and D. Seni Its, “Analisis Sentimen Nasabah Pada Layanan Perbankan Menggunakan Metode Regresi Logistik Biner, Naive Bayes Classifier(NBC), dan Support Vctor Machine (SVM),” vol. 8, no. 2, pp. 2337–3520, 2019.
- [13] Kelvin, J. Banjarnahor, E. Indra, and H. S. Sinurat, “ANALISIS PERBANDINGAN SENTIMEN CORONA VIRUS DISEASE- 2019 (COVID19) PADA TWITTER MENGGUNAKAN METODE LOGISTIC REGRESSION DAN SUPPORT VECTOR MACHINE (SVM),” *Jurnal Sistem Informasi dan Ilmu Komputer Prima*, vol. 5, no. 2, pp. 47–52, Feb. 2022.
- [14] T. T. Widowati and M. Sadikin, “ANALISIS SENTIMEN TWITTER TERHADAP TOKOH

PUBLIK DENGAN ALGORITMA NAIVE BAYES DAN SUPPORT VECTOR MACHINE,” *Jurnal SIMETRIS*, vol. 11, no. 2, 2020, [Online]. Available: <https://t.co/Xzf91zHK41>

- [15] V. A. Permadi, “Analisis Sentimen Menggunakan Algoritma Naïve Bayes Terhadap Review Restoran di Singapura 141,” *Jurnal Buana Informatika*, vol. 11, no. 2, pp. 141–151, 2020, [Online]. Available: <https://www.kaggle.com/hj5992/restaurantreviews>
- [16] H. Santoso and D. Desliani, “Analisis Sentimen Mahasiswa Terkait Pembelajaran Tatap Muka Menggunakan Metode Naïve Bayes Classifier,” *Agustus*, vol. 21, no. 3, pp. 644–654, Aug. 2022.