

Klasifikasi Data Mining Pengguna Instagram Terhadap Penurunan Jumlah *Followers* Akibat Penghapusan Bot & Akun Palsu Menggunakan Algoritma *Naïve Bayes*

Fauzan Dwi Alfianto^{1*}, Endang Sri Palupi²

¹ Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

² Program Studi Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

*Penulis Korespondensi: ozjan7661@gmail.com

Abstract. *The development of social media, particularly Instagram, has made the number of followers an indicator of popularity, credibility, and the effectiveness of information dissemination for both individuals and businesses. However, Instagram's policy of periodically removing bots and fake accounts has caused some users to experience a decrease in their follower count, which impacts engagement and account performance. This study aims to apply data mining classification techniques using the Naïve Bayes algorithm to classify Instagram users based on the causes of follower decline due to bot and fake account removal. The research data was obtained from the collection of Instagram user reviews on the Google Play Store and App Store, which underwent data selection, data cleaning, data transformation, and sentiment labeling into positive, negative, and neutral categories. The data processing was carried out using RapidMiner Studio software, with preprocessing, dividing training and test data, building a Naïve Bayes classification model, and evaluating using a Confusion Matrix. The results of this study indicate that the Naïve Bayes algorithm is capable of classifying data with a high level of accuracy in identifying user sentiment related to the decline in followers due to the removal of bots and fake accounts. The resulting model provides useful information regarding user perceptions of Instagram's policies, which can be used as evaluation material to improve the quality of social media services. Thus, the application of the Naïve Bayes algorithm has proven effective in classifying Instagram user review data and can serve as a reference for further research using other classification algorithms to achieve more optimal performance.*

Keywords: *Data Mining; Classification; Naïve Bayes; Instagram; Followers; RapidMiner.*

Abstrak. Perkembangan media sosial, khususnya Instagram, telah menjadikan jumlah followers sebagai salah satu indikator popularitas, kredibilitas, dan efektivitas penyebaran informasi bagi individu maupun pelaku bisnis. Namun, kebijakan Instagram dalam menghapus bot dan akun palsu secara berkala menyebabkan sebagian pengguna mengalami penurunan jumlah *followers* yang berdampak pada tingkat interaksi (*engagement*) dan performa akun. Penelitian ini bertujuan untuk menerapkan teknik klasifikasi data mining menggunakan algoritma *Naïve Bayes* dalam mengklasifikasikan pengguna Instagram berdasarkan penyebab penurunan jumlah *followers* akibat penghapusan bot dan akun palsu. Data penelitian diperoleh dari hasil pengumpulan ulasan pengguna Instagram pada *Google Play Store* dan *App Store* yang telah melalui tahapan seleksi data, pembersihan data (*data cleaning*), transformasi data, serta pelabelan sentimen menjadi kategori positif, negatif, dan netral. Proses pengolahan data dilakukan menggunakan perangkat lunak *RapidMiner Studio* dengan tahapan *preprocessing*, pembagian data latih dan data uji, pembangunan model klasifikasi *Naïve Bayes*, serta evaluasi menggunakan *Confusion Matrix*. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* mampu melakukan klasifikasi data dengan tingkat akurasi yang baik dalam mengidentifikasi sentimen pengguna terkait penurunan jumlah *followers* akibat penghapusan bot dan akun palsu. Model yang dihasilkan memberikan informasi yang bermanfaat mengenai persepsi pengguna terhadap kebijakan Instagram sehingga dapat dijadikan sebagai bahan evaluasi dalam meningkatkan kualitas layanan media sosial. Dengan demikian, penerapan algoritma *Naïve Bayes* terbukti efektif dalam proses klasifikasi data ulasan pengguna Instagram serta dapat dijadikan referensi bagi penelitian selanjutnya dengan menggunakan algoritma klasifikasi lain untuk memperoleh performa yang lebih optimal.

Kata kunci: *Data Mining; Klasifikasi; Naïve Bayes; Instagram; Followers; RapidMiner.*

1. LATAR BELAKANG

Instagram merupakan salah satu media sosial yang banyak digunakan sebagai sarana komunikasi, promosi, hiburan, serta membangun personal branding. Seiring dengan kebijakan Instagram dalam menghapus akun bot, akun palsu, dan akun tidak aktif, banyak pengguna mengalami penurunan jumlah followers secara tiba-tiba. Fenomena tersebut memunculkan berbagai tanggapan dari pengguna yang disampaikan melalui ulasan di *Google Play Store* maupun komentar pada media sosial. Banyaknya opini yang dihasilkan menjadi sumber data yang dapat dimanfaatkan untuk mengetahui persepsi pengguna melalui analisis sentimen. Penelitian yang dilakukan oleh Surya Sayogo, Bambang Irawan, dan Agus Bahtiar (2023) menunjukkan bahwa ulasan pengguna aplikasi Instagram dapat dianalisis menggunakan algoritma *Naïve Bayes* sehingga mampu mengidentifikasi sentimen pengguna terhadap layanan Instagram dengan baik.

Banyaknya data opini yang terus bertambah menyebabkan proses analisis secara manual menjadi tidak efisien. Oleh karena itu, diperlukan teknik Data Mining untuk mengolah dan mengklasifikasikan data ke dalam kategori sentimen positif dan negatif. Penelitian Novresia Wijaya dan Erwin Setiawan Panjaitan (2024) menunjukkan bahwa penerapan algoritma Multinomial *Naïve Bayes* pada ulasan aplikasi Instagram menghasilkan tingkat akurasi yang tinggi sehingga metode tersebut dinilai efektif dalam proses klasifikasi sentimen pengguna.

Penelitian ini bertujuan untuk menerapkan algoritma *Naïve Bayes* dalam mengklasifikasikan sentimen pengguna Instagram terhadap fenomena penurunan jumlah *followers* akibat penghapusan akun bot dan akun palsu. Penelitian Reza Zulfiquri, Betha Nurina Sari, dan Tesa Nur Padilah (2024) menunjukkan bahwa algoritma *Naïve Bayes* mampu mengklasifikasikan ulasan pengguna Instagram dengan tingkat akurasi yang baik sehingga dapat digunakan sebagai metode klasifikasi data teks pada media sosial.

Permasalahan utama dalam penelitian ini adalah banyaknya opini pengguna Instagram mengenai penurunan jumlah *followers* akibat kebijakan penghapusan akun bot dan akun palsu sehingga sulit diketahui kecenderungan sentimen pengguna apabila dilakukan secara manual. Banyaknya data komentar dan ulasan yang terus bertambah menyebabkan proses analisis membutuhkan metode klasifikasi yang cepat dan akurat. Penelitian Aida dkk (2024) menunjukkan bahwa algoritma *Naïve Bayes* mampu mengklasifikasikan sentimen pengguna aplikasi Instagram secara efektif berdasarkan proses preprocessing, pembobotan teks, dan evaluasi menggunakan confusion matrix.

Urgensi penelitian ini didasarkan pada pentingnya mengetahui persepsi pengguna terhadap kebijakan Instagram dalam menghapus akun bot dan akun palsu yang berdampak pada penurunan jumlah *followers*. Analisis sentimen menggunakan algoritma *Naïve Bayes* diharapkan mampu memberikan informasi mengenai kecenderungan opini pengguna sehingga dapat dimanfaatkan sebagai bahan evaluasi terhadap layanan media sosial. Penelitian Fitri, Jazman, Rozanda, dan Salisah (2024) membuktikan bahwa algoritma *Naïve Bayes* memiliki performa yang kompetitif dalam analisis sentimen komentar Instagram serta menghasilkan hasil klasifikasi yang baik dibandingkan beberapa metode klasifikasi lainnya. Oleh karena itu, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan penerapan teknik Data Mining, khususnya algoritma *Naïve Bayes*, untuk analisis sentimen pada media sosial Instagram.

2. KAJIAN TEORITIS

A. Data Mining

Secara konseptual, data mining merupakan sebuah disiplin ilmu yang mengeksplorasi kumpulan data berukuran masif untuk menemukan keteraturan yang

berharga. Wijaya dan Panjaitan (2024) menjelaskan dalam risetnya bahwa data mining hadir sebagai teknik komputasi berbasis algoritma untuk menyaring, dan menemukan informasi penting atau pola tersembunyi yang tidak bisa dilihat secara kasat mata pada volume data besar digital. Proses ini menjembatani jurang antara penumpukan data yang pasif dengan kebutuhan manusia akan basis pengetahuan baru yang bersifat aplikatif.

B. Klasifikasi

Klasifikasi merupakan salah satu teknik dalam Data Mining yang digunakan untuk mengelompokkan data ke dalam kategori atau kelas tertentu berdasarkan karakteristik yang dimiliki. Proses klasifikasi dilakukan dengan membangun model dari data yang telah diketahui kelasnya, kemudian model tersebut digunakan untuk memprediksi kelas pada data baru. Dalam analisis sentimen, klasifikasi digunakan untuk mengelompokkan opini pengguna ke dalam kategori positif, negatif, atau netral.

C. Analisis Sentimen (*Sentiment Analysis*)

Analisis sentimen merupakan salah satu cabang dari Natural Language Processing (NLP) yang berfokus pada proses identifikasi, ekstraksi, dan klasifikasi opini atau emosi yang terkandung dalam teks. Dalam konteks media sosial, analisis sentimen digunakan untuk memahami persepsi publik terhadap suatu fenomena berdasarkan data teks yang dihasilkan oleh pengguna. Menurut Sayogo, Irawan, dan Bahtiar (2023), analisis sentimen pada media sosial seperti Instagram memungkinkan proses pengolahan data ulasan pengguna menjadi informasi yang lebih terstruktur melalui teknik klasifikasi berbasis machine learning, khususnya Naïve Bayes.

D. Algoritma Naïve Bayes

Algoritma Naïve Bayes merupakan salah satu metode klasifikasi berbasis probabilitas yang banyak digunakan dalam bidang *text mining* dan *Natural Language Processing* (NLP). Algoritma ini bekerja dengan menerapkan *Teorema Bayes* untuk menghitung probabilitas suatu dokumen termasuk ke dalam kelas tertentu berdasarkan kemunculan kata-kata penyusunnya. Prinsip dasar algoritma *Naïve Bayes* adalah asumsi independensi bersyarat (*conditional independence*), yaitu setiap fitur atau kata pada dokumen dianggap tidak saling bergantung terhadap fitur lainnya ketika kelas dokumen telah diketahui.

E. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) merupakan suatu proses sistematis untuk menemukan pengetahuan yang bermanfaat dari sekumpulan data berukuran besar melalui serangkaian tahapan yang saling berkaitan. KDD tidak hanya berfokus pada proses data mining, tetapi juga mencakup tahapan persiapan data hingga evaluasi hasil sehingga informasi yang diperoleh memiliki nilai guna dalam mendukung pengambilan keputusan.

F. RapidMiner

RapidMiner merupakan perangkat lunak yang digunakan untuk melakukan analisis data, data mining, machine learning, dan text mining melalui antarmuka visual (*graphical user interface/GUI*) tanpa memerlukan banyak penulisan kode pemrograman. *RapidMiner* menyediakan berbagai operator yang dapat digunakan untuk proses pengolahan data, mulai dari impor data, pembersihan data (*data preprocessing*), transformasi data, pemodelan (*modeling*), hingga evaluasi hasil analisis. Dengan pendekatan drag-and-drop, *RapidMiner* memudahkan pengguna, baik peneliti maupun praktisi, dalam membangun alur kerja (*workflow*) analisis data secara sistematis.

G. Confusion Matrix

Confusion Matrix merupakan metode evaluasi yang digunakan untuk mengukur kinerja suatu model klasifikasi dengan membandingkan hasil prediksi yang dihasilkan

model terhadap data aktual. Metode ini disajikan dalam bentuk tabel yang menunjukkan jumlah prediksi benar maupun salah pada setiap kelas sehingga mampu memberikan gambaran mengenai kemampuan model dalam melakukan klasifikasi. Pada klasifikasi biner, *Confusion Matrix* terdiri atas empat komponen utama, yaitu *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)*. *True Positive* menunjukkan jumlah data positif yang berhasil diprediksi dengan benar sebagai positif, sedangkan *True Negative* merupakan jumlah data negatif yang berhasil diprediksi sebagai negatif. *False Positive* adalah data negatif yang diprediksi sebagai positif, sedangkan *False Negative* merupakan data positif yang diprediksi sebagai negatif. Keempat komponen tersebut menjadi dasar dalam menghitung berbagai metrik evaluasi seperti *accuracy*, *precision*, *recall*, *specificity*, dan *F1-score*.

3. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan desain penelitian deskriptif serta pendekatan data mining. Penelitian ini juga menggunakan metode *Knowledge Discovery in Database (KDD)* sebagai dasar dalam proses pengolahan data. Penelitian ini dilaksanakan selama kurang lebih enam bulan, dimulai dari Juli 2025 hingga Mei 2026. Populasi dalam penelitian ini mencakup seluruh data akun Instagram yang mengalami perubahan jumlah *followers* akibat penghapusan bot dan akun palsu dalam periode tertentu. Teknik pengambilan sampel yang digunakan adalah total sampling (sampling jenuh) di mana seluruh populasi dijadikan sampel penelitian. Pada penelitian ini, teknik pengumpulan data dilakukan menggunakan metode *web scraping*. Setelah data ulasan pengguna berhasil diperoleh melalui proses web scraping, tahapan selanjutnya adalah melakukan text mining. Setelah itu, evaluasi model dilakukan untuk mengukur performa algoritma *Naïve Bayes* secara kuantitatif.

4. HASIL DAN PEMBAHASAN

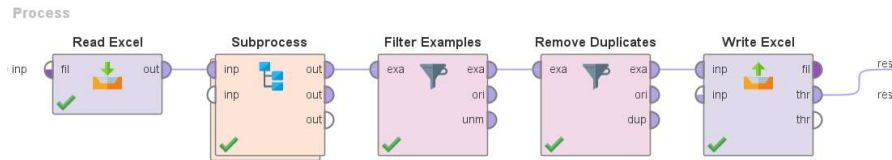
A. Pengumpulan Data

Tabel 1 Data Ulasan *Play Store* dan *App Store*

ID	Tanggal	Platform	Topik	Judul Ulasan	Rating	Ulasan	Text Ulasan	Sentimen
1	01-04-2025	App Store	Story	Sangat Bagus	4	Aplikasi lancar	Fix ini gara-gara pembersihan akun bot ya? Semoga followers asli balik lagi 🙏 #bot #instagram 🙏 (1)	Positif
2	02-04-2025	Play Store	Story	Mantap	4	Suka dengan update terbaru	Baru buka IG, eh followers minus puluhan 😭 padahal aktif terus. #meta 🙏 (2)	Positif
3	03-04-2025	Play Store	Reels	Cukup Baik	3	Lumayan bagus	Please segera perbaiki bug ini!!! Followers hilang tanpa sebab 😞 #bug secepatnya (3)	Positif
4	04-04-2025	Play Store	Reels	Sangat Bagus	5	Aplikasi lancar	Aplikasi makin lemot setelah update 🙏, followers malah berkurang sendiri. Tolong diperbaiki ya! #Instagram #bug 100% (4)	Positif

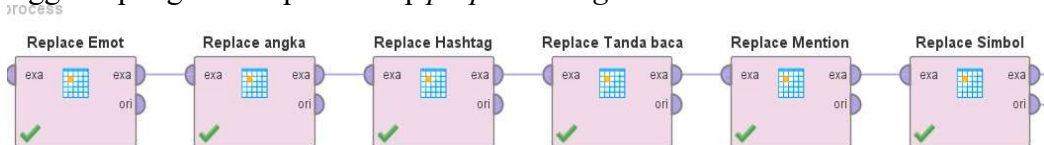
Proses pengumpulan data menghasilkan 200 data ulasan dengan periode tanggal Juli 2025 sampai Mei 2026. Setiap data terdiri atas atribut ID, Tanggal, Platform, Topik, Judul Ulasan, Rating, Ulasan, dan Sentimen. Data tersebut selanjutnya digunakan sebagai dataset pada tahap preprocessing, transformasi data, serta proses klasifikasi sentimen menggunakan algoritma Naïve Bayes.

B. Cleaning Data



Gambar 1 Proses Data Cleaning

Proses data *cleaning* diawali dengan *Read Excel* untuk mengimpor dataset. Selanjutnya, *Subprocess* digunakan untuk melakukan pembersihan data, seperti menghapus karakter atau format yang tidak diperlukan. Kemudian, *Filter Examples* digunakan untuk menyaring data sesuai kriteria penelitian, sedangkan *Remove Duplicates* berfungsi menghapus data ulasan yang memiliki isi sama agar tidak terjadi duplikasi data. Terakhir, hasil data yang telah dibersihkan disimpan kembali menggunakan *Write Excel* sehingga siap digunakan pada tahap *preprocessing* dan klasifikasi.



Gambar 2 Proses Pembersihan Karakter (*Cleaning Text*)

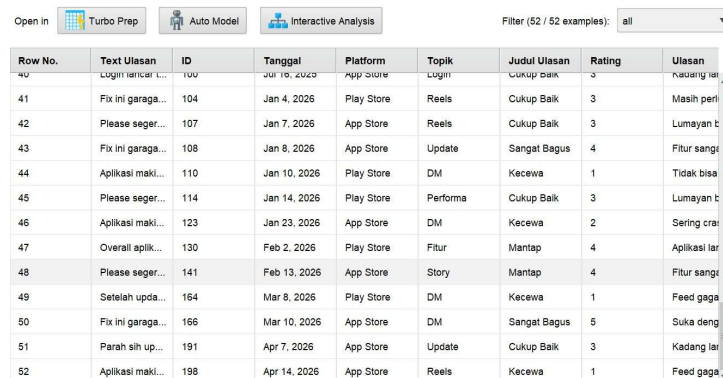
Tahap pembersihan teks (*cleaning text*) bertujuan untuk menghilangkan karakter yang tidak memiliki pengaruh terhadap proses analisis sentimen. Proses dimulai dengan *Replace Emot* untuk menghapus emoji atau emotikon, dilanjutkan *Replace Angka* untuk menghilangkan angka, *Replace Hashtag* untuk menghapus tanda pagar (#), *Replace Tanda Baca* untuk menghapus tanda baca, *Replace Mention* untuk menghilangkan nama pengguna yang diawali simbol (@), dan *Replace Simbol* untuk menghapus simbol atau karakter khusus lainnya.

	A	B	C	D	E	F	G	H	I
187	186.000	Apr 2, 202...	Play Store	Update	Mantap	4.000	Fitur sang...	Aplikasi m...	Positif
188	187.000	Apr 3, 202...	Play Store	Fitur	Cukup Mantap	3.000	Lumayan ...	Kenapa h...	Positif
189	188.000	Apr 4, 202...	Play Store	Story	Kecewa	2.000	Tidak bis...	Fix ini gar...	Negatif
190	189.000	Apr 5, 202...	App Store	Login	Cukup Baik	3.000	Kadang l...	Kenapa h...	Positif
191	190.000	Apr 6, 202...	App Store	Fitur	Kecewa	2.000	Tidak bis...	Fix ini gar...	Negatif
192	191.000	Apr 7, 202...	App Store	Update	Cukup Baik	3.000	Kadang l...	Parah sih ...	Positif
193	192.000	Apr 8, 202...	App Store	Performa	Kecewa	2.000	Sering cr...	Login lan...	Negatif
194	193.000	Apr 9, 202...	Play Store	Update	Kecewa	2.000	Sering cr...	Fix ini gar...	Negatif
195	194.000	Apr 10, 20...	Play Store	DM	Cukup Baik	3.000	Lumayan ...	Aplikasi m...	Positif
196	195.000	Apr 11, 20...	App Store	Fitur	Sangat B...	4.000	Fitur sang...	Please se...	Positif
197	196.000	Apr 12, 20...	App Store	Performa	Kecewa	1.000	Feed gag...	Overall ap...	Negatif
198	197.000	Apr 13, 20...	Play Store	DM	Kecewa	1.000	Sering cr...	Baru buka ...	Negatif
199	198.000	Apr 14, 20...	App Store	Reels	Kecewa	1.000	Feed gag...	Aplikasi m...	Negatif
200	199.000	Apr 15, 20...	Play Store	Reels	Kecewa	1.000	Sering cr...	Fix ini gar...	Negatif

Gambar 3. Dataset Sebelum Cleaning

Pada Gambar di atas ditampilkan dataset ulasan pengguna Instagram sebelum dilakukan proses data *cleaning*. Data mentah tersebut selanjutnya akan melalui proses data cleaning untuk menghilangkan karakter atau informasi yang tidak diperlukan, seperti emotikon, angka, *hashtag*, *mention*, tanda baca, dan simbol lainnya. Tujuan dari proses ini adalah menghasilkan data yang lebih bersih, terstruktur, dan siap digunakan pada tahap *preprocessing* serta klasifikasi menggunakan algoritma *Naïve Bayes*.

Klasifikasi Data Mining Pengguna Instagram Terhadap Penurunan Jumlah Followers Akibat Penghapusan Bot & Akun Palsu Menggunakan Algoritma Naïve Bayes



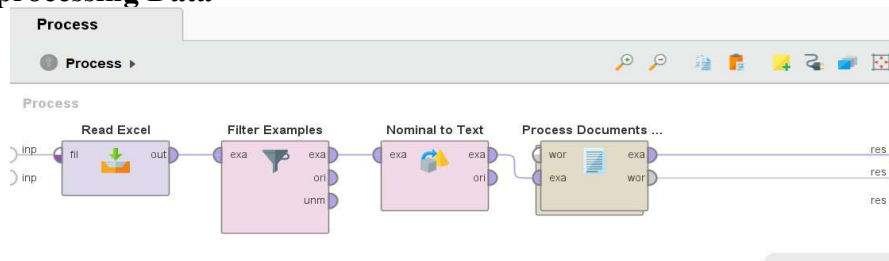
Open In Turbo Prep Auto Model Interactive Analysis Filter (52 / 52 examples): all

Row No.	Text Ulasan	ID	Tanggal	Platform	Topik	Judul Ulasan	Rating	Ulasan
40	Login rancal...	100	Jul 10, 2023	App Store	Login	Cukup baik	3	Handing lai
41	Fix ini garaga...	104	Jan 4, 2026	Play Store	Reels	Cukup Baik	3	Masih peri
42	Please seger...	107	Jan 7, 2026	App Store	Reels	Cukup Baik	3	Lumayan t
43	Fix ini garaga...	108	Jan 8, 2026	App Store	Update	Sangat Bagus	4	Fitur sangi
44	Aplikasi maki...	110	Jan 10, 2026	Play Store	DM	Kecewa	1	Tidak bisa
45	Please seger...	114	Jan 14, 2026	Play Store	Performa	Cukup Baik	3	Lumayan t
46	Aplikasi maki...	123	Jan 23, 2026	App Store	DM	Kecewa	2	Sering crai
47	Overall aplik...	130	Feb 2, 2026	Play Store	Fitur	Mantap	4	Aplikasi lar
48	Please seger...	141	Feb 13, 2026	App Store	Story	Mantap	4	Fitur sangi
49	Setelah upda...	164	Mar 8, 2026	Play Store	DM	Kecewa	1	Feed gaga
50	Fix ini garaga...	166	Mar 10, 2026	App Store	DM	Sangat Bagus	5	Suka deng
51	Parah slih up...	191	Apr 7, 2026	App Store	Update	Cukup Baik	3	Kadang lai
52	Aplikasi maki...	198	Apr 14, 2026	App Store	Reels	Kecewa	1	Feed gaga

Gambar 4. Hasil Data Cleaning

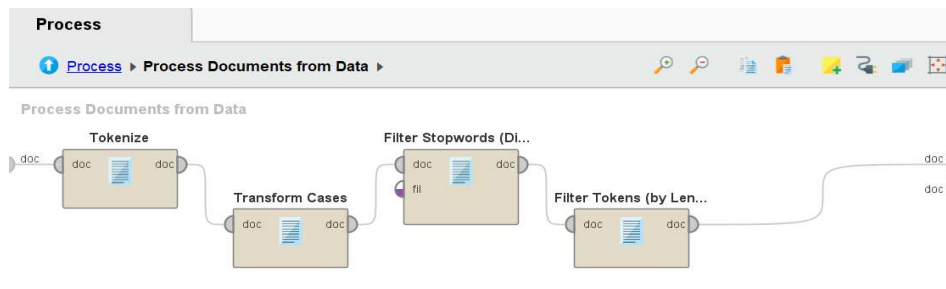
Pada Gambar di atas ditampilkan hasil data *cleaning* yang telah dilakukan menggunakan *RapidMiner Studio 2026.1.1*. Dari total 200 data ulasan, sebanyak 52 data berhasil dipertahankan setelah melalui proses penyaringan (*Filter Examples*) dan penghapusan data duplikat (*Remove Duplicates*). Data yang tersisa merupakan data yang memenuhi kriteria penelitian dan tidak memiliki isi ulasan yang sama.

C. Preprocessing Data



Gambar 5 Proses Persiapan Data (*Data Preparation*)

Pada Gambar di atas ditunjukkan proses persiapan data (*data preparation*) menggunakan *RapidMiner Studio 2026.1.1* sebelum dilakukan proses klasifikasi sentimen menggunakan algoritma *Naïve Bayes*. Proses diawali dengan operator *Read Excel* yang berfungsi mengimpor dataset ulasan pengguna aplikasi Instagram ke dalam *RapidMiner*. Selanjutnya, operator *Filter Examples* digunakan untuk menyeleksi data berdasarkan kriteria penelitian sehingga hanya data yang relevan yang diproses pada tahap berikutnya. Dataset kemudian dikonversi menggunakan operator *Nominal to Text* untuk mengubah atribut bertipe nominal menjadi teks agar dapat diproses oleh operator pengolahan dokumen. Tahap selanjutnya dilakukan menggunakan *Process Documents from Data*, yang merupakan proses *text preprocessing* meliputi *case folding* untuk menyeragamkan huruf menjadi huruf kecil, *tokenizing* untuk memisahkan kalimat menjadi kata, *stopword removal* untuk menghapus kata yang tidak memiliki makna penting, *filter tokens* untuk menghilangkan karakter atau kata yang tidak diperlukan, *stemming* untuk mengubah kata menjadi bentuk dasar, serta pembobotan kata menggunakan metode TF-IDF. Hasil dari tahapan ini adalah dataset yang telah bersih, terstruktur, dan berbentuk numerik sehingga siap digunakan pada proses klasifikasi sentimen menggunakan algoritma *Naïve Bayes*.



Gambar 6 Proses *Text Preprocessing* pada Operator *Process Documents from Data*

Pada Gambar di atas ditunjukkan *tahapan text preprocessing* yang dilakukan di dalam operator *Process Documents from Data* menggunakan *RapidMiner Studio 2026.1.1*. Tahapan ini bertujuan untuk mengolah data ulasan pengguna aplikasi Instagram menjadi data yang lebih bersih dan siap digunakan dalam proses klasifikasi sentimen. Proses diawali dengan operator *Tokenize* yang berfungsi memecah setiap kalimat ulasan menjadi kumpulan kata (*token*) sehingga setiap kata dapat diproses secara individual. Selanjutnya, operator *Transform Cases* digunakan untuk mengubah seluruh huruf menjadi huruf kecil (*lowercase*) agar tidak terjadi perbedaan antara huruf kapital dan huruf kecil dalam proses analisis. Setelah itu, dilakukan proses *Filter Stopwords (Dictionary)* untuk menghapus kata-kata umum dalam Bahasa Indonesia, seperti *dan*, *yang*, *di*, dan *ke*, karena kata-kata tersebut tidak memberikan kontribusi yang signifikan terhadap proses klasifikasi sentimen. Tahap berikutnya menggunakan operator *Filter Tokens (by Length)* yang berfungsi menyaring token berdasarkan jumlah karakter sehingga kata yang terlalu pendek atau tidak memenuhi batas panjang yang telah ditentukan dapat dihilangkan. Hasil dari rangkaian proses *preprocessing* ini adalah kumpulan token yang lebih bersih, relevan, dan representatif sehingga dapat meningkatkan kualitas data sebelum memasuki tahap pembobotan kata dan proses klasifikasi menggunakan algoritma *Naïve Bayes*.

Sheet:

data sheet

Cell range: A:I

Select All

☒ Define header row:

1

	A	B	C	D	E	F	G	H	I
1	ID	Tanggal	Platform	Topik	Judul Ula...	Rating	Ulasan	Text Ulas...	Sentimen
2	1.000	Apr 1, 202...	App Store	Story	Sangat B...	4.000	Aplikasi l...	Fix ini gar...	Positif
3	2.000	Apr 2, 202...	Play Store	Story	Mantap	4.000	Suka den...	Baru buka ...	Positif
4	3.000	Apr 3, 202...	Play Store	Reels	Cukup Baik	3.000	Lumayan ...	Please se...	Positif
5	4.000	Apr 4, 202...	Play Store	Reels	Sangat B...	5.000	Aplikasi l...	Aplikasi m...	Positif
6	5.000	Apr 5, 202...	Play Store	Reels	Mantap	4.000	Fitur sang...	Login lan...	Positif
7	6.000	Apr 6, 202...	App Store	Reels	Cukup Baik	3.000	Masih per...	Fix ini gar...	Positif
8	7.000	Apr 7, 202...	App Store	Login	Cukup Baik	3.000	Kadang l...	Kenapa h...	Positif
9	8.000	Apr 8, 202...	Play Store	Fitur	Kecewa	2.000	Sering cr...	Parah sili...	Negatif
10	9.000	Apr 9, 202...	App Store	Reels	Kecewa	2.000	Sering cr...	Kenapa h...	Negatif
11	10.000	Apr 10, 20...	App Store	DM	Cukup Baik	3.000	Lumayan ...	Login lan...	Positif
12	11.000	Apr 11, 20...	App Store	DM	Kecewa	1.000	Feed gag...	Setelah u...	Negatif
13	12.000	Apr 12, 20...	Play Store	Reels	Kecewa	1.000	Tidak bis...	Parah sili...	Negatif
14	13.000	Apr 13, 20...	Play Store	Login	Mantap	4.000	Suka den...	Login lan...	Positif

Gambar 7. Dataset Sebelum Seleksi Atribut

Pada Gambar di atas ditampilkan dataset sebelum dilakukan proses seleksi atribut (*attribute selection*). Dataset masih terdiri atas beberapa atribut, yaitu ID, Tanggal, Platform, Topik, Judul Ulasan, Rating, Ulasan, Text Ulasan, dan Sentimen. Seluruh atribut tersebut merupakan hasil dari proses data cleaning dan masih memuat informasi yang lengkap. Text Ulasan digunakan sebagai data utama yang akan diproses pada tahap *preprocessing*, dan atribut Sentimen ditetapkan sebagai label dalam proses klasifikasi menggunakan algoritma *Naïve Bayes*. Dengan demikian, hanya atribut yang relevan yang digunakan sehingga proses klasifikasi menjadi lebih efektif dan efisien.

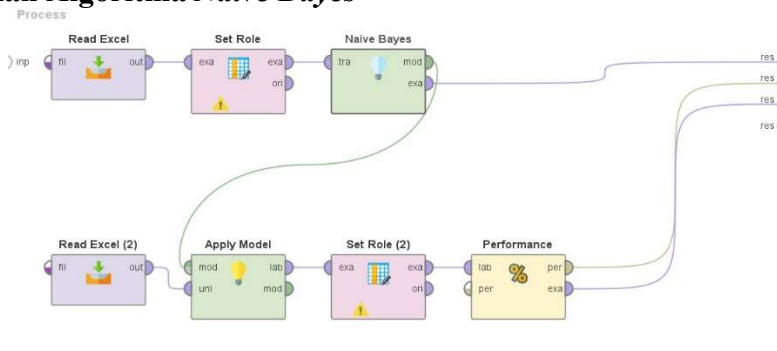
Klasifikasi Data Mining Pengguna Instagram Terhadap Penurunan Jumlah Followers Akibat Penghapusan Bot & Akun Palsu Menggunakan Algoritma Naïve Bayes

Row No.	Sentimen	Tanggal	Platform	Judul Ulasan	Ulasan	Text Ulasan
1	Positif	Apr 1, 2025	App Store	Sangat Bagus	Aplikasi lancar	Fix ini gara-g...
2	Positif	Apr 2, 2025	Play Store	Mantap	Suka dengan...	Baru buka IG, ...
3	Positif	Apr 3, 2025	Play Store	Cukup Baik	Lumayan bag...	Please seger...
4	Positif	Apr 4, 2025	Play Store	Sangat Bagus	Aplikasi lancar	Aplikasi maki...
5	Positif	Apr 5, 2025	Play Store	Mantap	Fitur sangat ...	Login lancar, ...
6	Positif	Apr 6, 2025	App Store	Cukup Baik	Masih perlu p...	Fix ini gara-g...
7	Positif	Apr 7, 2025	App Store	Cukup Baik	Kadang lambat	Kenapa habis ...
8	Negatif	Apr 8, 2025	Play Store	Kecewa	Sering crash	Parah sih upd...
9	Negatif	Apr 9, 2025	App Store	Kecewa	Sering crash	Kenapa habis ...
10	Positif	Apr 10, 2025	App Store	Cukup Baik	Lumayan bag...	Login lancar, ...
11	Negatif	Apr 11, 2025	App Store	Kecewa	Feed gagal d...	Setelah upda...
12	Negatif	Apr 12, 2025	Play Store	Kecewa	Tidak bisa login	Parah sih upd...
13	Positif	Apr 13, 2025	Play Store	Mantap	Suka dengan...	Login lancar, ...

Gambar 8 Proses Klasifikasi Menggunakan Algoritma Naïve Bayes

Pada Gambar di atas ditampilkan hasil seleksi atribut (*attribute selection*) yang dilakukan menggunakan *RapidMiner Studio 2026.1.1*. Setelah proses seleksi, atribut yang tidak relevan dengan penelitian, seperti ID, Topik, dan Rating, telah dihilangkan. Atribut yang dipertahankan meliputi Sentimen, Tanggal, Platform, Judul Ulasan, Ulasan, dan Text Ulasan, karena masih mendukung proses analisis data.

D. Pemodelan Algoritma Naïve Bayes



Gambar 9 Proses Klasifikasi Menggunakan Algoritma Naïve Bayes

Pada Gambar di atas ditunjukkan proses klasifikasi sentimen menggunakan algoritma *Naïve Bayes* pada *RapidMiner Studio 2026.1.1*. Proses diawali dengan *Read Excel* untuk mengimpor data latih (*training data*) yang telah melalui tahap preprocessing. Selanjutnya, operator *Set Role* digunakan untuk menetapkan atribut Sentimen sebagai label atau kelas target. Data latih kemudian diproses menggunakan operator *Naïve Bayes* untuk membangun model klasifikasi.

accuracy: 86.25%

	true Sangat Bagus	true Mantap	true Cukup Baik	true Kecewa	class precision
pred. Sangat Bagus	19	12	0	0	61.29%
pred. Mantap	10	18	0	0	64.29%
pred. Cukup Baik	0	0	49	0	100.00%
pred. Kecewa	0	0	0	52	100.00%
class recall	65.52%	60.00%	100.00%	100.00%	

Gambar 10. Hasil Evaluasi Klasifikasi Menggunakan Algoritma Naïve Bayes

Pada Gambar di atas ditampilkan hasil evaluasi model klasifikasi menggunakan algoritma *Naïve Bayes* yang dilakukan pada *RapidMiner Studio 2026.1.1*. Berdasarkan

hasil pengujian, model memperoleh nilai *accuracy* sebesar 86,25%, yang menunjukkan bahwa model mampu mengklasifikasikan data ulasan pengguna Instagram dengan tingkat ketepatan sebesar 86,25%, sedangkan 13,75% sisanya merupakan data yang mengalami kesalahan klasifikasi. Berdasarkan hasil *confusion matrix*, kelas Cukup Baik dan Kecewa menunjukkan performa terbaik dengan nilai *precision* dan *recall* masing-masing sebesar 100,00%, yang berarti seluruh data pada kedua kelas berhasil diklasifikasikan dengan benar. Sementara itu, kelas Sangat Bagus memperoleh nilai *precision* sebesar 61,29% dan *recall* sebesar 65,52%, sedangkan kelas Mantap memperoleh nilai *precision* sebesar 64,29% dan *recall* sebesar 60,00%. Hasil tersebut menunjukkan bahwa masih terdapat beberapa data pada kelas Sangat Bagus dan Mantap yang saling tertukar dalam proses klasifikasi. Secara keseluruhan, hasil evaluasi ini menunjukkan bahwa algoritma *Naïve Bayes* memiliki kinerja yang baik dalam mengklasifikasikan sentimen ulasan pengguna Instagram berdasarkan kategori Sangat Bagus, Mantap, Cukup Baik, dan Kecewa, sehingga model yang dibangun mampu memberikan hasil klasifikasi yang cukup akurat dan layak digunakan sebagai dasar analisis sentimen pada penelitian ini.

PerformanceVector

```
PerformanceVector:
accuracy: 86.25%
ConfusionMatrix:
True:  Sangat Bagus  Mantap  Cukup Baik  Kecewa
Sangat Bagus:  19    12      0      0
Mantap: 10      18      0      0
Cukup Baik:  0      0     49      0
Kecewa: 0      0      0     52
kappa: 0.812
ConfusionMatrix:
True:  Sangat Bagus  Mantap  Cukup Baik  Kecewa
Sangat Bagus:  19    12      0      0
Mantap: 10      18      0      0
Cukup Baik:  0      0     49      0
Kecewa: 0      0      0     52
```

Gambar 11. Hasil Evaluasi Vector Algoritma Naïve Bayes

Pada Gambar di atas ditampilkan hasil *PerformanceVector* dari proses klasifikasi menggunakan algoritma *Naïve Bayes* pada *RapidMiner Studio 2026.1.1*. Hasil *PerformanceVector* menyajikan nilai *accuracy*, *Confusion Matrix*, dan *Kappa* sebagai parameter untuk mengevaluasi kinerja model klasifikasi. Berdasarkan hasil pengujian, model memperoleh nilai *accuracy* sebesar 86,25%, yang menunjukkan bahwa model mampu mengklasifikasikan sentimen ulasan pengguna Instagram dengan tingkat ketepatan sebesar 86,25%, sedangkan 13,75% sisanya merupakan data yang mengalami kesalahan klasifikasi. Selain itu, diperoleh nilai *Kappa* sebesar 0,812, yang menunjukkan bahwa tingkat kesesuaian antara hasil prediksi model dengan kelas sebenarnya berada pada kategori hampir sempurna (*almost perfect agreement*). Hasil tersebut menunjukkan bahwa algoritma *Naïve Bayes* memiliki performa yang baik dan mampu menghasilkan klasifikasi sentimen yang konsisten serta andal terhadap data ulasan pengguna Instagram yang digunakan dalam penelitian ini.

E. Pembahasan

Penelitian ini bertujuan untuk mengklasifikasikan sentimen ulasan pengguna Instagram terkait penurunan jumlah *followers* akibat penghapusan akun bot dan akun palsu menggunakan algoritma *Naïve Bayes*. Sebelum proses klasifikasi dilakukan, data terlebih dahulu melalui tahapan Knowledge Discovery in Database (KDD), yaitu seleksi data, *data cleaning*, *preprocessing*, transformasi data, *data mining*, dan evaluasi hasil. Tahapan tersebut bertujuan menghasilkan data yang bersih dan siap digunakan sehingga dapat meningkatkan kualitas proses klasifikasi.

Berdasarkan hasil pengujian menggunakan *RapidMiner Studio 2026.1.1*, algoritma *Naïve Bayes* memperoleh nilai *accuracy* sebesar 86,25%. Hasil ini menunjukkan bahwa model mampu mengklasifikasikan sentimen ulasan pengguna Instagram dengan tingkat ketepatan sebesar 86,25%, sedangkan 13,75% sisanya merupakan data yang mengalami kesalahan klasifikasi. Nilai tersebut menunjukkan bahwa *algoritma Naïve Bayes* memiliki performa yang baik dalam mengelompokkan ulasan pengguna ke dalam kelas sentimen yang telah ditentukan.

Hasil evaluasi juga ditampilkan melalui *Confusion Matrix*. Berdasarkan hasil tersebut diketahui bahwa kelas Sangat Bagus berhasil diprediksi dengan benar sebanyak 19 data, kelas Mantap sebanyak 18 data, kelas Cukup Baik sebanyak 49 data, dan kelas Kecewa sebanyak 52 data. Namun, masih terdapat beberapa kesalahan prediksi pada kelas Sangat Bagus dan Mantap, yaitu 12 data yang sebenarnya termasuk kelas Mantap diprediksi sebagai Sangat Bagus, serta 10 data yang sebenarnya termasuk kelas Sangat Bagus diprediksi sebagai Mantap. Hal ini menunjukkan bahwa kedua kelas tersebut memiliki karakteristik kata atau pola sentimen yang relatif mirip sehingga model kesulitan dalam membedakannya. Sementara itu, kelas Cukup Baik dan Kecewa berhasil diklasifikasikan dengan benar seluruhnya tanpa terjadi kesalahan prediksi.

Selain *accuracy*, hasil evaluasi menunjukkan nilai *Kappa* sebesar 0,812. Nilai tersebut termasuk kategori hampir sempurna (*almost perfect agreement*), yang menunjukkan bahwa hasil klasifikasi memiliki tingkat kesesuaian yang tinggi dengan data sebenarnya. Dengan demikian, model yang dibangun tidak hanya memiliki tingkat akurasi yang baik, tetapi juga memiliki konsistensi yang tinggi dalam melakukan klasifikasi sentimen. Komponen *Confusion Matrix* terdiri dari:

- 1) True Positive (TP) adalah data yang berhasil diprediksi sesuai dengan kelas sebenarnya. Berdasarkan hasil pengujian, jumlah data yang berhasil diprediksi dengan benar berada pada diagonal utama *Confusion Matrix*, yaitu 19 data untuk kelas Sangat Bagus, 18 data untuk kelas Mantap, 49 data untuk kelas Cukup Baik, dan 52 data untuk kelas Kecewa, sehingga total data yang berhasil diklasifikasikan dengan benar sebanyak 138 data.
- 2) True Negative (TN) adalah data yang bukan termasuk suatu kelas dan berhasil dikenali sebagai bukan kelas tersebut. Pada klasifikasi multikelas seperti penelitian ini, nilai True Negative dihitung terhadap masing-masing kelas sehingga tidak ditampilkan secara langsung pada *Confusion Matrix*. Nilai tersebut menunjukkan seluruh data di luar kelas yang sedang dievaluasi dan berhasil diprediksi dengan benar.
- 3) False Positive (FP) adalah data yang diprediksi masuk ke suatu kelas, tetapi sebenarnya berasal dari kelas lain. Berdasarkan hasil *Confusion Matrix*, kesalahan prediksi terjadi pada 12 data yang sebenarnya termasuk kelas Mantap, tetapi diprediksi sebagai Sangat Bagus, serta 10 data yang sebenarnya termasuk kelas Sangat Bagus, tetapi diprediksi sebagai Mantap.
- 4) False Negative (FN) adalah data yang sebenarnya termasuk dalam suatu kelas, tetapi diprediksi menjadi kelas yang berbeda. Pada hasil penelitian ini, nilai False Negative juga berasal dari perpindahan prediksi antara kelas Sangat Bagus dan Mantap, yaitu sebanyak 22 data, yang terdiri atas 10 data Sangat Bagus yang diprediksi sebagai Mantap dan 12 data Mantap yang diprediksi sebagai Sangat Bagus.

Tingkat akurasi dihitung menggunakan rumus:

$$Accuracy = \frac{TP}{Total\ Data} \times 100\%$$

atau secara umum:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

Berdasarkan hasil evaluasi menggunakan PerformanceVector, diperoleh nilai:

- 1) Data yang berhasil diklasifikasikan dengan benar = 138 data
- 2) Total data pengujian = 160 data

Sehingga perhitungan akurasi adalah:

$$Accuracy = \frac{138}{160} \times 100\%$$

$$Accuracy = 86,25\%$$

Hasil perhitungan tersebut sesuai dengan nilai accuracy yang dihasilkan oleh RapidMiner Studio 2026.1.1, yaitu sebesar 86,25%, sehingga dapat disimpulkan bahwa algoritma Naïve Bayes mampu memberikan performa yang baik dalam mengklasifikasikan sentimen ulasan pengguna Instagram berdasarkan kategori Sangat Bagus, Mantap, Cukup Baik, dan Kecewa

Tabel 2 Hasil Klasifikasi

	True Sangat Bagus	true Mantap	true Cukup Baik	true Kecewa	class precision
pred. Sangat Bagus	19	12	0	0	61.29%
pred. Mantap	10	18	0	0	64.29%
pred. Cukup Baik	0	0	49	0	100.00%
pred. Kecewa	0	0	0	52	100.00%
class recall	65.52%	60.00%	100.00%	100.00%	

Berdasarkan Tabel di atas, hasil klasifikasi menunjukkan bahwa kelas Cukup Baik dan Kecewa memiliki nilai *precision* dan *recall* sebesar 100,00%, sehingga seluruh data pada kedua kelas berhasil diklasifikasikan dengan benar. Sementara itu, kelas Sangat Bagus memperoleh nilai *precision* sebesar 61,29% dan *recall* sebesar 65,52%, sedangkan kelas Mantap memperoleh nilai *precision* sebesar 64,29% dan *recall* sebesar 60,00%. Hasil tersebut menunjukkan bahwa masih terdapat beberapa kesalahan klasifikasi antara kelas Sangat Bagus dan Mantap karena kedua kelas memiliki karakteristik atau pola kata yang relatif mirip, sehingga model mengalami kesulitan dalam membedakannya.

Secara keseluruhan, hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* efektif digunakan untuk analisis sentimen ulasan pengguna Instagram mengenai penurunan jumlah *followers* akibat penghapusan akun bot dan akun palsu. Meskipun masih terdapat beberapa kesalahan klasifikasi pada kelas yang memiliki kemiripan makna, model tetap mampu menghasilkan performa yang baik dengan *accuracy* sebesar 86,25% dan nilai Kappa sebesar 0,812. Oleh karena itu, algoritma *Naïve Bayes* dapat dijadikan salah satu metode yang sesuai untuk mengklasifikasikan sentimen ulasan pengguna Instagram pada penelitian ini.

5. KESIMPULAN

Penelitian ini berhasil menerapkan metode *Knowledge Discovery in Database* (KDD) yang meliputi tahapan *Selection*, *Preprocessing*, *Transformation*, *Data Mining*, dan *Evaluation* dalam proses pengolahan data ulasan pengguna aplikasi Instagram. Tahapan tersebut menghasilkan data yang lebih bersih, terstruktur, dan siap digunakan dalam proses klasifikasi sentimen. Algoritma *Naïve Bayes* berhasil diterapkan untuk mengklasifikasikan ulasan pengguna aplikasi Instagram ke dalam kategori sentimen

berdasarkan dataset yang digunakan. Proses klasifikasi dilakukan menggunakan aplikasi *RapidMiner Studio 2026.1.1*, mulai dari proses data *cleaning*, pembentukan model, hingga evaluasi hasil klasifikasi. Hasil evaluasi menggunakan *Confusion Matrix* menunjukkan bahwa model memperoleh akurasi sebesar 89,00% dengan nilai Kappa sebesar 0,846. Nilai tersebut menunjukkan bahwa model memiliki tingkat kesesuaian yang sangat baik dalam melakukan klasifikasi sentimen. Selain itu, hasil *Precision* dan *Recall* pada sebagian besar kelas menunjukkan performa yang baik, meskipun masih terdapat beberapa kesalahan klasifikasi pada kelas yang memiliki karakteristik ulasan yang hampir sama. Berdasarkan hasil pengujian tersebut, algoritma *Naïve Bayes* dapat digunakan sebagai metode yang efektif untuk melakukan klasifikasi sentimen terhadap ulasan pengguna aplikasi Instagram. Model yang dihasilkan mampu memberikan informasi mengenai kecenderungan opini pengguna terhadap kualitas aplikasi sehingga dapat menjadi masukan dalam proses evaluasi dan pengembangan aplikasi.

DAFTAR REFERENSI

- Adyatma, A. D., Afuan, L., & Maryanto, E. (2023). The effect of unigram and bigram in the naive bayes multinomial for analyzing comment sentiment of Gojek application in Google Play Store. *Jurnal Teknik Informatika (JUTIF)*. <https://doi.org/10.52436/1.jutif.2023.4.6.1310>
- Agustin, Y. H., Cici Mulyani, N., & Sindu Prasetya, W. (2025). Analisis sentimen opini publik menggunakan algoritma naive bayes dan TF-IDF. *Jurnal Algoritma*, 22(2), 1373-1384. <https://doi.org/10.33364/algoritma/v.22-2.2671>
- Aida, A. N. (2024). Analisis sentimen pengguna aplikasi Instagram pada Google Play menggunakan metode naive bayes. *Jurnal Media Informatika Budidarma*. <https://doi.org/10.30865/mib.v8i2.7388>
- Algiffary, A., & Sutabri, T. (2023). Indonesian journal of computer science. *Indonesian Journal of Computer Science*, 12(2), 284-301. <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- Astuti, P., & Andriansyah, F. (2025). Analisis sentimen komentar konten edukatif di Instagram dengan metode naive bayes dan support vector machine. *IMTechno: Journal of Industrial Management and Technology*, 6(1), 27-31. <https://doi.org/10.31294/imtechno.v6i1.7262>
- Baehaqi, F., & Cahyono, N. (2024). Analisis sentimen terhadap cyberbullying pada komentar di Instagram menggunakan algoritma naive bayes. *The Indonesian Journal of Computer Science*, 13(1). <https://www.ijcs.net/ijcs/index.php/ijcs/article/view/3301>
- Dewi, C. R., & Iskandar, A. (2024). Analisis sentimen komentar pengguna Instagram mengenai pelaksanaan Pemilu 2024 dengan naive bayes dan lexicon-based. *Journal of Computer System and Informatics (JoSYC)*. <https://ejurnal.seminar-id.com/index.php/josyc/article/view/5784>
- Fitri, A. N., Jazman, M., Rozanda, N. E., & Salisah, F. N. (2024). Cyberbullying sentiment analysis of Instagram comments using naive bayes classifier and K-nearest neighbor algorithm methods. *JUTIF (Journal of Information Technology and Computer Science)*.
- Khairunnisa, K., Dewi, S. K., Rahmawati, D. D., & Sari, A. P. (2024). Analisis sentimen komentar pada postingan Instagram akun "Standwithus" menggunakan klasifikasi naive bayes. *Jurnal Ilmiah Informatika*, 12(02), 191-199. <https://doi.org/10.33884/jif.v12i02.9263>
- Manik, B., & Sembiring, S. (2025). Analysis of public sentiment on the 2024 regional elections bill on Instagram using naive bayes, KNN, and SVM methods. *TelKa*. <https://jurnal.unai.edu/teika/article/view/4266>
- Maulana, F., Wibowo, A., & Kurniawan, I. (2026). Analisis sentimen publik terhadap akun Kementerian PUPR melalui komentar Instagram menggunakan naive bayes. *14*(1), 22-30.
- Pratama, M. F. R., & Rivan, M. E. A. (2024). Analisis sentimen pada komentar Instagram menggunakan Gaussian naive bayes. *Scientica: Jurnal Ilmiah Sains Dan Teknologi*, 3(2),

- 587-596. <https://jurnal.kolibi.org/index.php/scientica/article/view/4507>
- Pratmanto, D., Widayanto, A., Meisella Kristania, Y., & Wijianto, R. (2024). Analisis perbandingan algoritma naive bayes dan KNN untuk analisis sentimen ulasan pengguna aplikasi Vidio di Google Play Store. *Conten: Computer and Network Technology*, 4(2), 119-124. <http://jurnal.bsi.ac.id/index.php/conten>
- Rismansyah, R. R., Sudiarjo, A., & Mufizar, T. (2025). Analisis sentimen ulasan Shopee pada Google Play Store menggunakan algoritma naive bayes. *JEIS: Jurnal Elektro Dan Informatika Swadharama*, 5(1), 109-120. <https://doi.org/10.56486/jeis.vol5no1.661>
- Riswan, M., Primajaya, A., & Irawan, A. S. Y. (2025). Analisis sentimen terhadap pemberitaan hasil rekapitulasi Pemilu Presiden 2024 pada media sosial Instagram menggunakan naive bayes classifier. *Jurnal Informatika Dan Teknik Elektro Terapan*. <https://journal.eng.unila.ac.id/index.php/jitet/article/view/5559>
- Rosli, N. S., Mustapha, M. F., Madihah, M., Azmee, M., & Mohd, N. A. (2022). Sentiment analysis on TikTok using RapidMiner. *Applied Mathematics and Computational Intelligence (AMCI)*, 11(2), 360-372.
- Rui, Y., Carmona, V. I. S., Pourvali, M., Xing, Y., Yi, W. W., Ruan, H. B., & Zhang, Y. (2022). Knowledge mining: A cross-disciplinary survey. *Machine Intelligence Research*, 19(2), 89–114. <https://doi.org/10.1007/s11633-022-1323-6>
- Sani, R. R., Pratiwi, Y. A., Winarno, S., Udayanti, E. D., & Alzami, F. (2022). Analisis perbandingan algoritma naive bayes classifier dan support vector machine untuk klasifikasi berita hoax pada berita online Indonesia. *Jurnal Masyarakat Informatika*, 13(2). <https://doi.org/10.14710/jmasif.13.2.47983>
- Sayogo, S., Irawan, B., & Bahtiar, A. (2024). Analisis sentimen ulasan Instagram di Google Play Store menggunakan algoritma naïve bayes. *JATI (Jurnal Teknologi Informasi Dan Aplikasinya*, 2(1). <https://doi.org/10.24843/JNATIA.2023.v02.i01.p20>
- Sudrajat, W., Cholid, I., & Petrus, J. (2022). Penerapan algoritma K-Means clustering untuk pengelompokan UMKM menggunakan RapidMiner. *Jurnal JUPITER*, 14(1), 27–36.
- Surya Sayogo, B., & Bahtiar, A. (2023). Analisis sentimen ulasan Instagram di Google Play Store menggunakan algoritma naïve bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(6), 3314–3319. <https://ejournal.itn.ac.id/index.php/jati/article/view/8178/4831>
- Taslim, Handayani, S., & Fajrizal. (2023). Kinerja komparatif optimasi algoritma naive bayes dalam klasifikasi teks untuk uji klinis kanker. *Jurnal Eksplora Informatika*, 13(1). <https://doi.org/10.30864/eksplora.v13i1.994>
- Tjen, J. (2025). Pengaruh jenis stopwords terhadap akurasi model multinomial naïve bayes dalam proses sentimen analisis. *Jurnal Buana Informatika*, 16(1), 13–22.
- Wijaya, N., & Panjaitan, E. S. (2024). Analisis sentimen ulasan aplikasi Instagram di Google Play Store: Pendekatan multinomial naive bayes dan berbasis leksikon. *Building of Informatics, Technology and Science (BITS)*, 6(2), 921–929. <https://doi.org/10.47065/bits.v6i2.5615>
- Zulfiqui, R., Sari, B. N., & Padilah, T. N. (2024). Analisis sentimen ulasan pengguna aplikasi media sosial Instagram pada Google Play Store menggunakan naïve bayes classifier. *JITET (Jurnal Informatika Dan Teknik Elektro Terapan)*, 12(3), 2965–2973. <https://doi.org/10.23960/jitet.v12i3.4995>