



## ANALISIS SENTIMEN TERHADAP TAGAR #KABURAJADULU DI MEDIA SOSIAL X MENGGUNAKAN METODE NAÏVE BAYES

Tsanil Hamim Aftoni<sup>1</sup>, Wibowo Harry Sugiharto<sup>2</sup>, Muhammad Imam Ghozali<sup>3</sup>

<sup>1,2,3</sup> Universitas Muria Kudus, Kudus (59327)

\* Email Korespondensi: 201851272@std.umk.ac.id

INFO ARTIKEL	ABSTRAK
<p><b>Sejarah Artikel:</b> Diterima Tgl. 4/07/2025 Diperbaiki Tgl. 18/07/2025 Disetujui Tgl.30/07/2025 Tersedia daring Tgl. 31/07/2025</p>	<p>Tingginya aktivitas masyarakat Indonesia di media sosial, khususnya platform X (dahulu Twitter), mencerminkan kuatnya opini publik terhadap berbagai isu sosial, termasuk kritik terhadap kebijakan pemerintah. Salah satu fenomena yang viral adalah tagar #kaburajadulu, yang merepresentasikan ekspresi kekecewaan terhadap kondisi sosial, ekonomi, dan politik. Penelitian ini bertujuan untuk mengklasifikasikan sentimen masyarakat terhadap tagar tersebut ke dalam tiga kategori: positif, negatif, dan netral, serta mengevaluasi performa metode Naïve Bayes Multinomial dalam klasifikasi sentimen. Data dikumpulkan melalui proses crawling dari platform X, kemudian diproses melalui tahap pra-pemrosesan teks, pelabelan menggunakan InSet Lexicon, dan pembobotan fitur menggunakan TF-IDF. Model Naïve Bayes dilatih dan diuji dengan pembagian data 80:20 serta dievaluasi menggunakan 10-Fold Cross Validation dengan metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa model berhasil mengklasifikasikan sentimen negatif dengan baik (akurasi 75%), namun belum optimal dalam mengenali sentimen positif dan netral. Penelitian ini mengonfirmasi dominasi opini negatif terhadap tagar #kaburajadulu dan membuka peluang pengembangan metode klasifikasi yang lebih kompleks pada isu-isu sosial digital.</p>
<p>e-ISSN 2961-9009 p-ISSN 2963-1289</p>	
<p><b>DOI:</b> <a href="https://doi.org/10.58290/jukomtek.v4i2.443">https://doi.org/10.58290/jukomtek.v4i2.443</a></p>	<p><b>Kata Kunci:</b> Analisis sentimen, Naïve Bayes, media sosial X, #kaburajadulu, opini publik.</p>
<p>©2022. Diterbitkan oleh Jurnal Komputer dan Teknologi (JUKOMTEK). Artikel ini memiliki akses terbuka di bawah lisensi CC BY (<a href="https://creativecommons.org/licenses/by/4.0/">https://creativecommons.org/licenses/by/4.0/</a>)</p>	

### PENDAHULUAN

Seiring dengan perkembangan zaman, media sosial berfungsi sebagai media utama bagi banyak orang untuk

mengekspresikan diri, berkomunikasi dan berbagi pengalaman (Aurellia *et al.*, 2024). Salah satu media yang paling populer untuk berkomentar pada saat ini yaitu Twitter. Twitter atau sekarang

disebut juga dengan X merupakan aplikasi berbasis teks yang memungkinkan pengguna untuk menulis, menyampaikan dan mengirim opini. Sosial media ini memungkinkan penggunanya untuk melakukan diskursus mengenai kejadian yang ada melalui unggahan diri sendiri maupun unggahan orang lain.

Menurut survei gabungan yang dilakukan oleh We Are Social dan Meltwater, Twitter menjadi salah satu sosial media populer yang menduduki peringkat 7 dalam kategori *Time Spent Using Social Media Apps* pada tahun 2024 dengan total waktu 4 jam 40 menit per-orang pada tahun 2024 (Thompson, 2024). Sedangkan untuk penggunanya, sosial media X di Indonesia mencapai angka 62% dari total jumlah penduduk di Indonesia. Hal ini menunjukkan bahwa lebih dari 50% penduduk Indonesia menggunakan aplikasi X untuk bermain sosial media. Hal ini menunjukkan bahwa penggunaan media sosial ini berpengaruh besar dalam pengaruh terhadap kebijakan pemerintah melalui gerakan opini digital (Warawilis, 2025). Banyaknya opini digital ini berkembang menjadi isu baik di dalam aplikasi maupun isu nasional.

Isu yang baru-baru ini menjadi *trending* yaitu isu Kabur Aja Dulu. Isu ini berkembang pesat di berbagai macam sosial media Indonesia, termasuk X. Awalnya isu ini hanya menjadi candaan dari pengguna sosial media X. Namun, hal ini menjadi fenomena serius untuk mengkritisi

kebijakan pemerintah dari mulai susah mencari pekerjaan, biaya hidup yang semakin mahal dan kebijakan yang nyeleneh. Fenomena ini menjadi kritik serius yang disuarakan melalui tagar #kaburajadulu. Hal ini dibenarkan oleh sejumlah ahli. Menurut berita yang dikutip dari Kompas.com, sejumlah ahli termasuk sosiolog UIN Walisongo Semarang, Nur Hasyim berpendapat bahwa tren Kabur Aja Dulu merupakan kalkulasi dari ekspresi kemarahan, keputusasaan, dan protes yang disampaikan masyarakat di sejumlah sosial media kepada Pemerintah (Zaenuddin, 2025). Isu ini menjadi isu serius yang menimbulkan berbagai macam sentimen dari masyarakat kepada pemerintah.

Seiring dengan perkembangan teknologi yang semakin berkembang, sentimen masyarakat terhadap suatu isu dapat dikategorikan dengan mudah. Perkembangan *Artificial Intelligence* (AI), dapat menjadi suatu alat untuk mengategorikan sentimen masyarakat tersebut menjadi beberapa kategori seperti sentimen positif, negatif atau netral. Dalam hal ini, AI yang digunakan yaitu *machine learning*. *Machine Learning* adalah sebuah teknik yang memungkinkan komputer untuk belajar dari data yang diberikan tanpa harus diprogram secara menyeluruh dan eksplisit untuk setiap masalah (Meng *et al.*, 2020). Penggunaan *Machine Learning* untuk menganalisis berbagai macam sentimen menjadi beberapa kategori disebut dengan

analisis sentimen.

Ada berbagai macam metode yang dapat digunakan dalam analisis sentimen seperti SVM, Naïve Bayes dan Random Forest. Salah satu penelitian yang menggunakan metode Naïve Bayes telah dilakukan oleh Haq et al., (2024) dengan judul “Algoritma Naïve Bayes untuk Mengidentifikasi Hoaks di Media Sosial”. Penelitian ini menggunakan dataset dari postingan media sosial yang dikategorikan sebagai hoaks atau bukan hoaks. Proses *preprocessing data* mencakup *tokenizing*, ekstraksi fitur, dan pembersihan teks menggunakan teknik TF-IDF. Model Naïve Bayes yang digunakan menghasilkan akurasi sebesar 88%, *Precision* 86%, *Recall* 89%, dan *F1-Score* sebesar 87%. Model yang digunakan berhasil mengidentifikasi 450 dari 500 hoaks dengan benar dan 470 dari 500 non-hoaks dengan benar.

Selanjutnya yaitu penelitian yang dilakukan oleh (Ahmed Khan et al., 2024) dengan judul “*Sentiment Analysis using Support Vector Machine and Random Forest*”. Penelitian diawali dengan melakukan tinjauan literatur yang komprehensif, serta menerapkan teknik *preprocessing*, ekstraksi fitur, pelatihan model, dan evaluasi. Hasil yang didapatkan dari penelitian ini menunjukkan bahwa metode SVM memiliki akurasi sebesar 80%, sedangkan Random Forest sebesar 78%. Hasil tersebut menunjukkan bahwa SVM mendapatkan akurasi yang lebih unggul daripada Random

Forest.

Berdasarkan perbandingan di atas, metode yang lebih baik dalam menganalisis sentimen masyarakat yaitu metode Naïve Bayes. Sehingga berdasarkan latar belakang tersebut penelitian ini bertujuan untuk menganalisis sentimen masyarakat di sosial media Twitter (X) dengan tagar atau kata kunci Kabur Aja Dulu menggunakan metode Naive Bayes Classifier.

## LANDASAN TEORI

Naive Bayes merupakan salah satu metode populer dalam *machine learning* yang menggunakan pengklasifikasian, menggunakan statistic dan probabilitas (Dwiramadhan, Wahyuddin and Hidayatullah, 2022). Naïve Bayes merupakan penerapan dari teorema Bayes yang diperkenalkan oleh ilmuwan asal Inggris, Thomas Bayes, pada abad ke-18. Naive Bayes digunakan untuk berbagai macam keperluan seperti klasifikasi, sistem pendukung keputusan, dan analisis sentimen (Isnain, Marga and Alita, 2021).

Salah satu penerapan umum yang menggunakan metode Naive Bayes adalah analisis sentimen. Analisis sentimen merupakan proses mengidentifikasi dan mengategorikan berbagai macam opini yang berbentuk teks untuk menentukan apakah sentimen tersebut positif, negatif atau netral (Syahril Dwi Prasetyo, Shofa Shofiah Hilabi and Fitri Nurapriani, 2023). Proses ini bertujuan untuk mengetahui apakah sentimen yang berasal dari masyarakat cenderung positif, negatif atau netral terutama

dalam tren Kabur Aja Dulu yang beberapa bulan lalu marak di aplikasi Twitter (X). Pengelompokan sentimen ini dapat dilakukan dengan menggunakan berbagai macam algoritma *machine learning*. Penelitian ini secara khusus berfokus pada penggunaan metode Naive Bayes untuk menentukan sentimen positif, negatif dan netral pada #KaburAjaDulu.

Beberapa penelitian terdahulu telah mengkaji metode ini. Pada penelitian yang dilakukan oleh Kahi, Talakua and Reynaldi, (2024) menggunakan metode Naive Bayes Classifier untuk mengklasifikasi sentimen menjadi positif dan negatif. Klasifikasi dibagi ke dalam dua sentimen kategori positif dan negatif pada kasus tanggapan masyarakat terhadap masa pemerintahan Anies Baswedan sebagai Gubernur Daerah Khusus Ibu Kota Jakarta, setelah proses *crawling* data, *preprocessing*, pelabelan data, maka dapat diketahui performa tingkat akurasi 76,04%, presisi 71,43, *recall* 81,63 % dan *f-measure* 76,19%. Data diperoleh dengan melakukan *crawling* data dari Twitter rentang waktu Januari 2019 – Oktober 2022 dengan jumlah data 1563 *Tweet*, dapat diketahui hasil tanggapan masyarakat kategori positif sebanyak 746 (47,7%). Akan tetapi respons masyarakat pada masa pemerintahan Anies Baswedan berkategori negatif.

Penelitian selanjutnya dilakukan oleh Silitonga and Sihotang, (2019) dengan judul "Analisis Sentimen Pemilihan Presiden Indonesia Tahun 2019 Di Twitter Berdasarkan Geolocation Menggunakan Metode Naive Bayesian Classifier", konfigurasi algoritma Naive Bayesian Classifier terhadap pra-pemilihan presiden tahun 2019 dimulai dari pengumpulan data manual

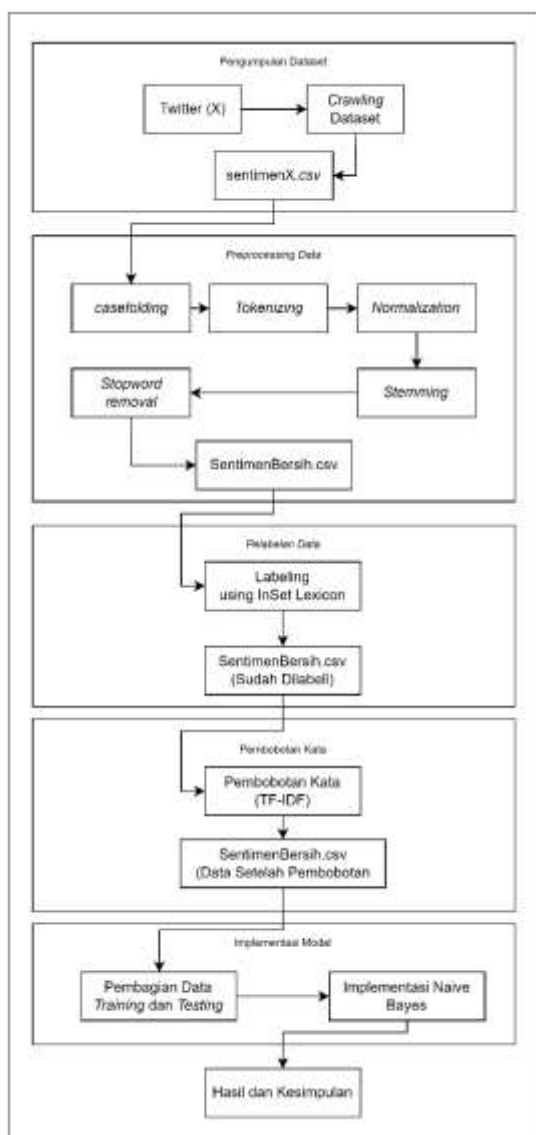
melalui Twitter. Lalu dilakukannya *pre-processing* terhadap data sampai diklasifikasinya pada aplikasi Weka dengan perhitungan Naive Bayesian Classifier terhadap probabilitas kata. Dengan memperoleh hasil probabilitas tertinggi, sehingga dapat mengategorikan sebuah kalimat menjadi sebuah sentimen yang sebenarnya. Perolehan sentimen *tweet* berdasarkan wilayah didapat melalui titik koordinat geografis bumi yang terhubung pada aplikasi Twitter (X) melalui API Geolocation yang terdiri dari garis lintang dan garis bujur pada google maps yang telah ditetapkan oleh pengguna. Perbandingan klasifikasi algoritma Naive Bayesian Classifier dengan algoritma Tree.J48 memiliki tingkat akurasi yang cukup berbeda. Di mana klasifikasi algoritma Naive Bayesian Classifier lebih tinggi dengan tingkat akurasi 77,62% dibandingkan dengan algoritma Tree.J48 yang hanya 70%.

Kedua penelitian di atas membuktikan keefektifitasan metode Naive Bayes untuk analisis sentimen, namun belum ada penelitian untuk mengkaji sentimen masyarakat terhadap fenomena sosial #kaburajadulu dengan menggunakan metode Naive Bayes Classifier. Sehingga penelitian ini memiliki urgensi dan kontribusi tersendiri dalam pengembangan kajian analisis opini publik berbasis media sosial.

## METODE PENELITIAN

Metode yang digunakan dalam penelitian ini yaitu metode Naive Bayes dengan fokus untuk mengklasifikasikan sentimen publik mengenai #KaburAjaDulu yang berada di aplikasi Twitter (X). Ada beberapa

tahapan penelitian yang dibagi menjadi pengumpulan dataset, *preprocessing data*, pelabelan data, pembobotan kata, implementasi model Naïve Bayes, lalu hasil dan Kesimpulan. Tahapan penelitian tersebut dapat dilihat pada Gambar 1.



Gambar 1. Desain Penelitian

### Pengumpulan Dataset

Penelitian ini dimulai dari pengumpulan dataset. Dataset yang dikumpulkan berasal dari aplikasi Twitter (X) dengan menggunakan Tweet Harvest. Tweet harvest digunakan untuk meng-crawling data Twitter dengan memasukkan kata

kunci tertentu. Pada penelitian ini, kata kunci yang digunakan yaitu Kabur Aja Dulu dan #KaburAjaDulu. Dengan menggunakan API Twitter (X), data dikumpulkan dari Februari 2023 hingga Maret 2025. Dikarenakan kebijakan Twitter (X) yang membatasi jumlah *scroll*, maka pada proses ini dilakukan hingga beberapa tahapan untuk menghindari diblokirnya akun, tahapan ini dilakukan dengan jeda waktu selama 5 menit untuk per-tahapan. Data yang berhasil dikumpulkan sebanyak 1510 cuitan yang mengandung kata kunci Kabur Aja Dulu atau #KaburAjaDulu.

### Preprocessing Data

Setelah mendapatkan dataset, proses selanjutnya yaitu *preprocessing data*. Data yang telah dikumpulkan akan melalui beberapa tahapan termasuk *casefolding*, *tokenizing*, *normalization*, *stemming*, dan *stopword removal*.

### Casefolding

Proses *casefolding* dimaksudkan untuk menyamaratakan huruf yang mengandung huruf kapital, kemudian diubah menjadi huruf kecil semua (Darmawan *et al.*, 2023).

### Tokenizing

Proses dalam *tokenizing* dilakukan untuk memecah kalimat menjadi satuan kata perkata. *Library* yang digunakan dalam proses ini yaitu NLTK, *library* ini memungkinkan proses *tokenizing* dilakukan melalui beberapa tahapan sehingga teks benar-benar tertoken secara sempurna,

### Normalization

*Normalization* bertujuan untuk mengubah kata yang tidak baku menjadi kata baku sesuai dengan

standar KBBI. Sebagai contoh, kata “gk” diubah menjadi “tidak” dan kata “bgt” diubah menjadi “banget”.

### Stemming

Selanjutnya yaitu *stemming*, proses ini dilakukan untuk mengubah kata variasi menjadi kata akarnya. Sebagai contoh, kata “memakan” diubah menjadi “makan”, dan kata “merubah” menjadi “ubah”.

### Stopword Removal

Tujuan dari proses *stopword removal* adalah untuk mengurangi *noise* pada dataset sehingga hanya kata yang benar-benar penting yang akan digunakan. Proses *stopword removal* pada penelitian ini menggunakan *library* Sastrawi. Penggunaan data “dan, di, dari, adalah” tidak memberikan makna yang signifikan. Contoh penggunaan *stopword removal* yaitu “saya sangat kecewa dengan kebijakan pemerintah” kemudian diubah menjadi “kecewa kebijakan pemerintah”.

### Pelabelan Data

Proses selanjutnya yaitu pelabelan data. Penelitian ini menggunakan *library* Inset Lexicon untuk melakukan pelabelan data. *Library* ini digunakan karena sudah teruji cukup baik untuk analisis sentimen berbahasa indonesia (Musfiroh *et al.*, 2021). *Library* ini terdiri dari 3,609 kata positif dan 6,609 kata negatif berbahasa indonesia. InSet Lexicon bekerja dengan cara menghitung bobot nilai atau *polarity score* pada setiap katanya yang berkisar antara -5 hingga +5. Kemudian, *polarity score* ini akan digunakan untuk klasifikasi jenis sentimen baik itu positif, negatif atau netral. Dataset yang telah melalui proses pelabelan

menggunakan InSet Lexicon mendapatkan hasil label negatif sebanyak 1180, label positif sebanyak 247 dan netral sebanyak 83.

### Pembobotan Kata

Setelah data berhasil dilabelkan menjadi kategori positif, negatif dan netral. Selanjutnya yaitu dilakukan pembobotan kata. Pembobotan kata dilakukan untuk mengubah data kata menjadi numerik. Hal ini dimaksudkan untuk memudahkan *machine learning* mempelajari setiap kata dengan baik. Penelitian ini menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk melakukan pembobotan pada dataset Kabur Aja Dulu. TF-IDF yaitu teknik yang digunakan untuk mengolah teks dengan memberikan bobot pada kata-kata dalam dokumen. Teknik ini mempunyai tujuan untuk mengidentifikasi kata-kata yang paling penting dalam suatu dokumen atau dataset (Wati, Ernawati and Rachmi, 2023). Cara kerja teknik TF-IDF adalah dengan menghitung nilai TF (*Term Frequency*), yaitu frekuensi kemunculan kata pada suatu dokumen. Adapun formulanya TF dijabarkan pada Formula 1:

$$TF(t, d) = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (1)$$

Keterangan:

- $TF(t, d)$  = *Term Frequency* kata  $t$  dalam dokumen  $d$
- $f_{t,d}$  = Jumlah kemunculan kata  $t$  dalam dokumen  $d$
- $\sum_k f_{k,d}$  = Total jumlah kata  $t$  dalam dokumen  $d$

Nilai TF tidak memberikan informasi tentang seberapa penting kata dalam dokumen tersebut. Akan banyak kata

yang muncul dalam sebuah dokumen sehingga memiliki nilai TF yang tinggi, namun tidak memberikan dampak yang penting dalam dokumen. Jadi, dibutuhkan teknik lain yang disebut dengan IDF (*Inverse Document Frequency*). IDF inilah yang kemudian memberikan bobot pada kata-kata yang muncul di dokumen. Adapun formula IDF dapat dilihat pada Formula 2.

$$= \log\left(\frac{N}{df_t}\right) \quad (2)$$

Keterangan:

- $IDF(t)$  = *Inverse Document Frequency* dari kata ( $t$ )
- $N$  = jumlah dokumen dari kumpulan teks sebagai data utama
- $df_t$  = jumlah dokumen yang mengandung kata ( $t$ )

Setelah mendapatkan masing-masing nilai TF dan IDF, selanjutnya yaitu dapat dihitung nilai TF-IDF-nya dengan mengalikan nilai TF terhadap nilai IDF. Kata yang memiliki nilai TF-IDF tinggi dianggap penting karena memberikan kontribusi lebih dalam penentuan topik dalam suatu dokumen.

### Implementasi Model

Setelah melakukan serangkaian proses, selanjutnya yaitu implementasi model. Sebelum diimplementasikan ke dalam model Naïve Bayes, dataset yang telah melalui proses pembobotan akan dibagi menjadi data *training* dan *testing*. Dataset yang digunakan akan dibagi dengan rasio 80:20. Di mana, 80% data akan dimasukkan ke dalam data *training* yang nantinya berfungsi untuk melatih model Naïve Bayes. Sedangkan 20% sisanya akan menjadi data *testing* untuk menguji model Naïve Bayes yang digunakan. Rasio ini umum

digunakan dalam *training* dan *testing* pada *machine learning*. Setelah dibagi, data kemudian akan diimplementasikan dengan model Naïve Bayes.

Naïve Bayes digunakan untuk memprediksi kejadian masa depan berdasarkan data dari masa lalu. Model ini menempuh dua tahap dalam proses klasifikasi teks. Proses pertama yaitu proses *training*, kemudian disusul dengan proses kedua yaitu klasifikasi. Adapun formula pada Naïve Bayes dijelaskan pada Formula 3.

$$\hat{y}_i = \underset{y}{\operatorname{arg\,max}} P(y) \prod_{j=1}^m P(X_{ij}|y) \quad (3)$$

Keterangan:

- $\hat{y}_i$  adalah hasil prediksi untuk data ke- $i$
- $P(y)$  adalah probabilitas awal dari kelas  $y$
- $x_{ij}$  adalah fitur ke- $j$  dari data ke- $i$
- $P(X_{ij}|y)$  adalah probabilitas fitur ke- $j$  dari data ke- $i$  diberikan kelas  $y$
- $m$  yaitu jumlah fitur dalam satu kata

Dalam model Naïve Bayes, setiap dokumen yang ada direpresentasikan dengan atribut  $x_1, x_2, x_3 \dots x_n$ .  $x_1$  adalah kata pertama, diikuti dengan  $x_2$  sebagai kata kedua sampai seterusnya. Sedangkan  $V$  merupakan himpunan kategori *Tweet* (Duei Putri, Nama and Sulistiono, 2022). Saat proses klasifikasi, model akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (VMAP). Adapun formulanya dapat Dilihat pada Formula 4.

$$V_{MAP} = \underset{V \in V}{\operatorname{arg\,max}} \frac{p(X_1, X_2, X_3, \dots, X_n | V_j P(V))}{P(X_1, X_2, X_3, \dots, X_n)} \quad (4)$$

Keterangan:

- $V_{MAP}$  = Kelas hasil prediksi berdasarkan MAP
- $V_j$  = salah satu kelas himpunan kelas  $j$
- $X_1, \dots, X_n$  = Fitur atau *evidence*

Karena nilai  $P(X_1, X_2, X_3, \dots, X_n)$  bernilai konstan untuk kategori ( $V_j$ ). Maka persamaan di atas dapat disederhanakan menjadi Formula 5:

$$V_{MAP} = \underset{V_j \in V}{\operatorname{argmax}} \prod_{i=1}^n P(x_i|V_j)P(V_j) \quad (5)$$

Keterangan:

- $V_j$  = Kategori dari twit  $j = 1, 2, 3, \dots, n$
- $J_1$  = Kategori twit sentimen negatif
- $J_2$  = Kategori twit sentimen positif
- $J_3$  = Kategori twit sentimen netral
- $P(x_i|V_j)$  = probabilitas  $x_i$  pada kategori
- $P(V_j)$  = Probabilitas dari  $V_j$

Pengujian metode Naïve Bayes dilakukan dengan menghitung nilai akurasi, presisi, *recall* dan *F1-score*. Perhitungan nilai akurasi, presisi, *recall* dan *F1-Score* dengan model Naïve Bayes secara sistematis dapat dilihat di formula 6-9.

$$\text{Akurasi} = \frac{\sum \text{data benar}}{n \text{ dokumen}} \times 100\% \quad (6)$$

$$\text{Presisi} = \frac{\sum \text{data positif atau negatif}}{n \text{ dokumen positif atau negatif}} \times 100\% \quad (7)$$

$$\text{Recall} = \frac{\sum \text{doc relevan dan terambil}}{\sum \text{seluruh dokumen relevan}} \times 100\% \quad (8)$$

$$F1 - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (9)$$

## HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dari setiap tahapan yang telah dijelaskan di Metode Penelitian. Adapun tahapan-tahapan tersebut dimulai dari pengumpulan dataset, *preprocessing data*, pelabelan data, pembobotan kata dan implementasi model Naïve Bayes terhadap dataset Kabur Aja Dulu.

### Hasil Pengumpulan Dataset

Dataset yang berhasil dikumpulkan menggunakan Tweet Harvest dan API Twitter berjumlah sebanyak 1510 cuitan. Adapun cuitan tersebut dapat dilihat pada Tabel 1.

Tabel 1. Dataset Terkumpul

Index	Text
0	putra putri terbaik bangsa lebih milih #KaburAjaDulu di prediksi indonesia bubar! a thread <a href="https://t.co/DQEIZ1gw9R">https://t.co/DQEIZ1gw9R</a>
1	Rame bgt #KaburAjaDulu beberapa waktu terakhir. Buat kalian yang mau kabur khususnya ke Jepang aku mau share ke kalian blog belajar bhs Jepang yang kubuat waktu masih sering ngajar Nihongo zaman kuliah. Memang masih sedikit tapi paling nggak ini gratis buat dipelajari mandiri
2	kabur aja dulu <a href="https://t.co/QxakVm4ZLc">https://t.co/QxakVm4ZLc</a>
3	Fenomena kabur aja dulu adalah indikasi negara gagal mewujudkan harapan banyak anak bangsa. Malah menterinya meragukan nasionalisme warganya. Apa enggak bercermin? itu artinya bapak-ibu yang enggak berhasil membuat

	mereka merasa punya harapan di sini.
4	Kabur aja dulu <a href="https://t.co/TUf2AHOnIo">https://t.co/TUf2AHOnIo</a>

### Hasil Preprocessing Data

Setelah dataset terkumpul, selanjutnya data melalui serangkaian tahap *preprocessing* dari mulai *casefolding*, *tokenizing*, *normalization*, *stemming* dan *stopword removal*. Adapun setiap tahap tersebut dapat dilihat dari Tabel 2-6.

Tabel 2. Hasil *Casefolding*

Index	Text
0	putra putri terbaik bangsa lebih milih #kabura...
1	rame bgt #kaburajadulu beberapa waktu terakhir...
2	kabur aja dulu <a href="https://t.co/qxakvm4zlc">https://t.co/qxakvm4zlc</a>
3	fenomena kabur aja dulu adalah indikasi negara...
4	kabur aja dulu <a href="https://t.co/tuf2ahonio">https://t.co/tuf2ahonio</a>

Setelah dilakukan tahap *casefolding*, kemudian data melakukan tahap *tokenizing*. Hasil dari tahap ini dapat dilihat di Tabel 3.

Tabel 3. Hasil *Tokenizing*

Index	Text
0	[putra, putri, terbaik, bangsa, lebih, milih, ...]
1	[Rame, bgt, beberapa, waktu, terakhir, Buat, k...]
2	[Kabur, aja, dulu]
3	[Fenomena, kabur, aja, dulu, adalah, indikasi,...]
4	[Kabur, aja, dulu]

Selanjutnya yaitu tahap *normalization*. Adapun hasil dari tahap ini dapat dilihat pada Tabel 4.

Tabel 4. Hasil *Normalization*

Index	Text
0	[putra, putri, terbaik, bangsa, lebih, memilih...]
1	[Rame, banget, beberapa, waktu, terakhir, Buat...]
2	[kabur, saja, dulu]
3	[Fenomena, kabur, saja, dulu, adalah, indikasi...]
4	[kabur, saja, dulu]

Kemudian masuk ke tahap *stemming*. Hasil dari tahap ini dapat dilihat pada Tabel 5.

Tabel 5. Hasil *Stemming*

Sebelum <i>Stemming</i>	Setelah <i>Stemming</i>
pinggiran	pinggir
konflik	konflik
kepentingan	penting
pertanianlahan	pertanianlahan
senjata	senjata

Tahap terakhir yaitu *stopword removal*, adapun kata-kata yang termasuk menjadi *stopword removal* yaitu "yg", "dg", "rt", "dgn", "ny", 'klo', 'kalo', 'amp', 'biar', 'bikin', 'bilang', 'gak', 'ga', 'krn', 'nya', 'nih', 'sih', 'si', 'tau', 'tdk', 'tuh', 'utk', 'ya', 'jd', 'jgn', 'sdh', 'aja', 'nyg', 'hehe', 'pen', 'u', 'nan', 'loh', 'rt', '&', 'yah', 'sdgkan', 'sdg', 'emg', 'sm', 'pls', 'mlu', 'ken', 'allah', 'brb', 'btw', 'b/c', 'cod', 'cmiiw', 'fyi', 'gg', 'ggwp', 'idk', 'ikr', 'lol', 'ootd', 'lmao', 'oot', 'pap', 'otw', 'tfl', 'vc', 'ygy'. Sehingga hasilnya dapat dilihat pada Tabel 6.

Tabel 6. Hasil *Stopword Removal*

Index	Text
0	[putra, putri, bangsa, pilih, prediksi, indone..
1	[rame, banget, kabur, khusus, jepang, share, b...
2	[kabur]
3	[fenomena, kabur, indikasi, negara, gagal, wuj...
4	[kabur]

### Hasil Pelabelan Data

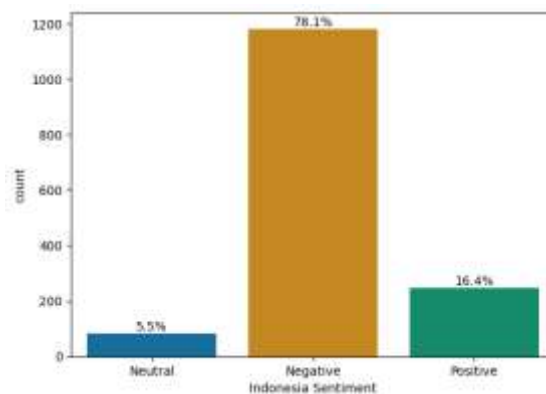
Dataset yang telah melalui proses *Preprocessing* kemudian dilakukan pelabelan data. Pada tahap ini, data yang dilabeli akan memunculkan *polarity score* dan juga label apakah data tersebut positif, negatif atau netral. Hasil dari pelabelan data ini dapat dilihat pada Tabel 7.

Tabel 7. Hasil Implementasi Pelabelan Data

Index	Text Filtering	Polarity Score	Sentiment
0	[putra, putri, bangsa, pilih, prediksi, indone..	0	Netral
1	[rame, banget, kabur, khusus, jepang, share,	-13	Negatif
2	[kabur]	-2	Negatif
3	[fenomena, kabur, indikasi, negara, gagal, wuj...	-3	Negatif

4	[kabur]	-2	Negatif
---	---------	----	---------

Setelah dilakukan pelabelan, dapat diketahui berapa jumlah data yang mempunyai sentimen positif, negatif dan netral. Perbandingan dari data tersebut dapat dilihat di Gambar 2 berikut ini.



Gambar 2. Perbandingan Data Sentimen

Dapat dilihat pada Gambar 2, sentimen negatif terhadap tagar atau kata kunci Kabur Aja Dulu di sosial media Twitter (X) didominasi oleh sentimen negatif dengan total sentimen negatif sebanyak 78,1%, disusul oleh sentimen positif sebanyak 16,4% dan sentimen netral sebanyak 5,5%. Dari pelabelan di atas menunjukkan bahwa tagar dan kata kunci Kabur Aja Dulu digunakan untuk mengekspresikan kekecewaan publik. Hal ini sejalan dengan pendapat dari sosiolog yang menganggap bahwa fenomena ini adalah bentuk protes digital terhadap kondisi sosial-ekonomi dan kebijakan pemerintah.

### Hasil dan Implementasi TF-IDF

Proses TF-IDF ini memungkinkan setiap kata mempunyai bobot tersendiri. Sehingga model Naïve Bayes yang digunakan bisa bekerja lebih maksimal. Adapun hasil dari pembobotan menggunakan TF-IDF ini dapat dilihat pada Tabel 8.

Tabel 8. Hasil Pembobotan TF-IDF

Kata	TF-IDF
'kabur'	0.06137412020713726
'tapi'	0.27418520464266216
'aja'	0.059910806471040134
'dulu'	0.12425009346248424
'sih'	0.18095123013653197

### Penerapan Model Naïve Bayes

Pada tahap ini, algoritma Naïve bayes digunakan untuk mengklasifikasikan data sentimen dengan kelas positif, negatif dan netral. Data yang digunakan yaitu data yang telah melalui serangkaian tahap *preprocessing* dan pembobotan TF-IDF. Tujuannya yaitu untuk mengidentifikasi kecenderungan sentimen masyarakat di sosial media Twitter (X) dengan tagar atau kata kunci Kabur Aja Dulu. Pada penelitian ini, model Naïve Bayes yang digunakan yaitu Naïve Bayes Multinomial. Naïve Bayes Multinomial dipilih karena cocok untuk data TF-IDF. Implementasi pada penelitian ini menggunakan *library* Scikit-learn.

Dataset yang digunakan kemudian dibagi menjadi data *training* dan data *testing*. Dataset dibagi dengan rasio 80:20, yaitu 80% untuk data *training* dan 20% untuk data *testing*. Pembagian data dilakukan dengan menggunakan fungsi `train_test_split` dari *library* `sklearn.model_selection`. Pembagian data ini juga menggunakan parameter `random_state = 0` untuk memastikan hasil pembagian data bersifat konsisten. Data *testing* digunakan untuk melatih model Naïve Bayes Multinomial supaya dapat mempelajari pola dari data. Sedangkan data *testing* digunakan untuk mengukur performa model yang telah dilatih terhadap data yang belum pernah diuji sebelumnya. Karena data positif dan netral tidak

seimbang dengan data negatif, maka dilakukan teknik SMOTE (*Synthetic Minority Over-sampling Technique*). Teknik ini digunakan untuk mengatasi masalah data yang tidak seimbang dengan cara membuat sampel data buatan untuk kelas minoritas. Dalam kasus ini, SMOTE digunakan untuk menambah kelas sentimen positif dan netral. Pembagian data setelah dilakukan teknik SMOTE dapat dilihat pada Tabel 9.

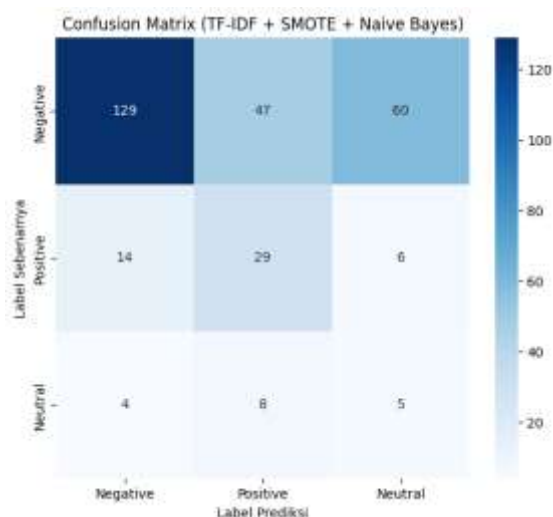
Tabel 9. Distribusi Data SMOTE

	Sebelum SMOTE	Setelah SMOTE
<b>Negatif</b>	944	944
<b>Positif</b>	198	944
<b>Netral</b>	66	944

Setelah dataset dibagi, data masuk ke dalam model Naïve Bayes Multinomial. Model diuji dengan data *testing* sebanyak 302 data cuitan. Evaluasi performanya dilakukan dengan menggunakan *confusion matrix* dan *classification report* untuk mengetahui nilai akurasi, presisi, *recall* dan *F1-score*. Adapun tabel hasil *confusion matrix* dapat dilihat pada Tabel 10 dan Gambar 3.

Tabel 10. Hasil *Confusion Matrix*

	Pred. Negatif	Pred. Positif	Pred. Netral
<b>Aktual Negatif</b>	129	47	60
<b>Aktual Positif</b>	14	29	6
<b>Aktual Netral</b>	4	8	5



Gambar 3. Confusion Matrix

Untuk hasil dari *classification report* dapat dilihat pada Tabel 11.

Tabel 11. Hasil *Classification Report*

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0.75	1.00	0.85	236
Positif	0.35	0.59	0.44	49
Netral	0.07	0.29	0.11	17
<b>Akurasi</b>			0.75	302
<i>Macro Avg</i>	0.43	0.48	0.41	302
<i>Weighted Avg</i>	0.75	0.54	0.60	302

Berdasarkan hasil dari tabel *confusion matrix* dan *classification report* di atas. Dapat disimpulkan bahwa model Naïve Bayes memiliki kemampuan yang baik dalam mengklasifikasikan sentimen negatif, namun tidak cukup kuat untuk mengenali sentimen positif dan netral. Hal ini dapat dilihat dari nilai *precision*, *recall*, dan *F1-Score* dengan nilai yang lebih rendah untuk kelas positif dan netral.

## KESIMPULAN

Penelitian ini mempunyai tujuan untuk mengklasifikasikan sentimen masyarakat di sosial media Twitter (X)

terhadap tagar dan kata kunci Kabur Aja Dulu menggunakan metode Naïve Bayes Multinomial dengan pembobotan menggunakan TF-IDF. Setelah dilakukan evaluasi model menggunakan data uji, didapatkan akurasi sebesar 75%. Model Naïve Bayes Multinomial yang digunakan berhasil mengklasifikasikan sentimen negatif dengan baik. Namun, tidak cukup kuat untuk mengenali sentimen positif dan netral. Model Naive Bayes yang digunakan masih belum seimbang untuk mengklasifikasikan sentimen terutama sentimen positif dan netral. Hal ini disebabkan oleh kurangnya data positif dan netral dari dataset Twitter (X) terkait tagar dan kata kunci Kabur Aja Dulu. Namun, ini justru memperkuat argumen bahwa Masyarakat tidak puas dengan kondisi sosial dan ekonomi saat ini sehingga menjadikan tren Kabur Aja Dulu sebagai pilihan untuk dilakukan. Temuan ini penting untuk memahami persepsi masyarakat terhadap isu sosial tertentu serta dapat digunakan sebagai landasan dalam pengambilan kebijakan publik berbasis opini daring. Penelitian selanjutnya disarankan untuk mengumpulkan data yang seimbang terkait tagar atau kata kunci Kabur Aja Dulu agar model Naïve Bayes Multinomial dapat lebih adil dalam mengenali semua kelas sentimen. Selain itu, penggunaan metode klasifikasi lain dapat dieksplorasi sehingga ada pembandingan untuk penelitian ini.

## DAFTAR PUSTAKA

Ahmed Khan, T. *et al.* (2024) 'Sentiment Analysis using Support Vector Machine and Random Forest', *Journal of Informatics and Web Engineering*,

- 3(1), pp. 67–75. Available at: <https://doi.org/10.33093/jiwe.2024.3.1.5>.
- Aurellia, Z. et al. (2024) *Pengalaman Pengguna X Sebagai Platform Curhat (Studi Fenomenologi)*, Seminar Nasional Universitas Negeri Surabaya.
- Darmawan, G. et al. (2023) 'Analisis Sentimen Berdasarkan Ulasan Pengguna Aplikasi My Pertamina Pada Google Playstore Menggunakan Metode Naive Bayes', *Jurnal Ilmiah Teknik dan Ilmu Komputer*, 2(3), pp. 100–108. Available at: <https://doi.org/10.55123>.
- Duei Putri, D., Nama, G.F. and Sulistiono, W.E. (2022) 'Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier', *Jurnal Informatika dan Teknik Elektro Terapan*, 10(1). Available at: <https://doi.org/10.23960/jitet.v10i1.2262>.
- Dwiramadhan, F., Wahyuddin, M.I. and Hidayatullah, D. (2022) 'Sistem Pakar Diagnosa Penyakit Kulit Kucing Menggunakan Metode Naive Bayes Berbasis Web', *Jurnal Teknologi Informasi dan Komunikasi*, 6(3), p. 2022. Available at: <https://doi.org/10.35870/jti>.
- Haq, M.Z. et al. (2024) 'Algoritma Naive Bayes untuk Mengidentifikasi Hoaks di Media Sosial', *Jurnal Minfo Polgan*, 13(1), pp. 1079–1084. Available at: <https://doi.org/10.33395/jmp.v13i1.13937>.
- Isnain, A.R., Marga, N.S. and Alita, D. (2021) 'Sentiment Analysis Of Government Policy On Corona Case Using Naive Bayes Algorithm', *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 15(1), p. 55. Available at: <https://doi.org/10.22146/ijccs.60718>.
- Kahi, F.R.B., Talakua, A. and Reynaldi, R. (2024) 'Analisis Sentimen Masyarakat Di Twitter Terhadap Pemerintahan Anies Baswedan Menggunakan Metode Naive Bayes Classifier', *Jurnal Minfo Polgan*, 13(1), pp. 324–336. Available at: <https://doi.org/10.33395/jmp.v13i1.13636>.
- Meng, T. et al. (2020) 'A survey on machine learning for data fusion', *Information Fusion*, 57, pp. 115–129. Available at: <https://doi.org/10.1016/j.inffus.2019.12.001>.
- Musfiroh, D. et al. (2021) 'Science Sentiment Analysis of Online Lectures in Indonesia from Twitter Dataset Using InSet Lexicon Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon', *MALCOM: Indonesian Journal of Machine Learning and Computer*, 1, pp. 24–33.
- Silitonga, W.H. and Sihotang, J.I. (2019) 'Analisis Sentimen Pemilihan Presiden Indonesia Tahun 2019 Di Twitter Berdasarkan Geolocation Menggunakan Metode Naive Bayesian Classification', *TelKa (Jurnal Teknonogi Informasi dan Komunikasi)*, 9(2).
- Syahril Dwi Prasetyo, Shofa Shofiah Hilabi and Fitri Nurapriani (2023) 'Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naive Bayes dan KNN', *Jurnal KomtekInfo*, pp. 1–7. Available at: <https://doi.org/10.35134/komtekinf.o.v10i1.330>.
- Thompson, A. (2024) *Digital 2024: 5 billion social media users - We Are Social Indonesia. We Are Social Indonesia*, <https://wearesocial.com/id/blog/2024/01/digital-2024-5-billion-social-media-users/>.
- Warawilis, D. (2025) 'Analisis Sentimen dan Opini Digital Media Sosial X Pada Isu #KABURAJADULU', *Universitas Ahmad Dahlan* [Preprint].
- Wati, R., Ernawati, S. and Rachmi, H. (2023) 'Pembobotan TF-IDF Menggunakan Naive Bayes pada Sentimen

Masyarakat Mengenai Isu Kenaikan BIPIH', *Jurnal Manajemen Informatika (JAMIKA)*, 13(1), pp. 84–93. Available at:

<https://doi.org/10.34010/jamika.v13i1.9424>.

Zaenuddin, M. (2025) *Awal Mula Tren Tagar Kabur Aja Dulu Ramai Digunakan, Mengapa?*,

<https://www.kompas.com/tren/read/2025/02/18/083000865/awal-mula-tren-tagar-kabur-aja-dulu-ramai-digunakan-mengapa-?page=all>.