

# Sentiment Analysis about Anti-LGBT Campaign using the Naïve Bayes Classifier

Rios Dacosta, Syafriandi\*, Dony Permana, Dina Fitria

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: [syafriandi\\_math@fmipa.unp.ac.id](mailto:syafriandi_math@fmipa.unp.ac.id)

Submitted : 06 Januari 2024

Revised : 06 Februari 2024

Accepted : 16 Februari 2024

## ABSTRACT

*Social media is growing so that the news that is discussed is also very fast to be known by everyone. The news or topic that is being discussed on social media is the anti-LGBT campaign. The conversation about the anti-LGBT campaign is expressed in the form of opinions that contain positive and negative feelings. The opinion is conveyed through Twitter. Twitter is a microblogging social media site that allows users to create short messages and share them easily and quickly. Opinions on Twitter are used to see whether the opinion rejects or supports the anti-LGBT campaign. The use of sentiment analysis helps to see the opinion supports or rejects the anti-LGBT campaign. The algorithm used to perform sentiment analysis is the Naïve Bayes Classifier. The purpose of this study is to determine the sentiment analysis of anti-LGBT campaign tweets on Twitter. This study using Python as the tools. The dataset used is 3103 tweets with 80% training data and 20% test data. The sentiment analysis results obtained in this study show that Twitter users in Indonesia have more positive opinions. The use of the Naïve Bayes Classifier algorithm produces an accuracy of 68,75%, precision of 99,6%, and recall of 92,8%.*

**Keywords:** *anti-lgbt campaign, naïve bayes classifier, sentiment analysis, social media*



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Dalam era globalisasi, perkembangan media sosial sangat terkait dengan adanya internet. Media sosial merupakan alat komunikasi yang memanfaatkan internet, sehingga tidak ada batasan jarak dan waktu dalam interaksinya. Twitter merupakan suatu platform sosial media yang populer. Twitter merupakan situs microblogging yang penggunanya dapat membuat pesan singkat yang biasa disebut "tweet". Tweet mempunyai limit maksimal 280 karakter. Twitter memiliki jutaan pengguna aktif di seluruh dunia, sehingga berita yang diunggah melalui platform ini dapat dengan cepat menyebar ke berbagai belahan dunia. Menurut Mailoa (2021), penggunaan Twitter sering meningkat saat terjadi peristiwa penting dan menarik seperti Piala Dunia, pemilihan presiden, kematian selebritis, dan sebagainya.

Salah satu topik yang sedang diperbincangkan di Twitter hingga akhir tahun 2022 adalah gerakan LGBT (*Lesbian, Gay, Biseksual, dan Transgender*). Istilah LGBT telah dipakai semenjak tahun 1990-an untuk merubah istilah komunitas *gay*, sebab lebih memadai dalam menggambarkan kelompok tersebut (Anjani dkk, 2022). Isu LGBT di Indonesia masih belum tuntas karena dianggap sebagai topik yang sensitif dan masyarakat Indonesia belum memahami isu tersebut (Hamzah & Dwiputri Maharani, 2021). Saat ini, masyarakat Indonesia sedang menghadapi peralihan budaya yang mempunyai konsekuensi negatif dan positif. Dalam konteks ini, terjadi penurunan nilai-nilai, etika, dan keyakinan agama di kalangan masyarakat. Salah satu contoh pergeseran budaya yang menarik perhatian masyarakat adalah keberadaan kaum LGBT di Indonesia. Twitter sebagai platform media sosial memberikan kebebasan bagi individu untuk berbicara, dan dampaknya bervariasi bagi masyarakat. Lewat media sosial ini, interaksi dan komunikasi kelompok LGBT sangat mudah. Menurut Mukhid (2019), kehadiran LGBT bukanlah dikarenakan oleh faktor genetik, ketidakseimbangan hormon, masalah mental, maupun perilaku jahat, tetapi lebih dipengaruhi oleh faktor lingkungan dan pengalaman pribadi.

Banyak kasus LGBT yang terjadi di Indonesia, kelompok LGBT berencana mengadakan pertemuan se-ASEAN (*Association of Southeast Asian Nations*) di Jakarta tapi gagal dilakukan karena banyak pihak yang menolak khususnya Majelis Ulama Indonesia (MUI). Tidak hanya itu, Pemerintah Daerah (Pemda) DKI Jakarta mengeluarkan pernyataan bahwa hutan kota di kawasan Cawang, Jakarta Timur diduga menjadi tempat berkumpulnya komunitas LGBT. Oleh

karena itu, Satuan Polisi Pamong Praja (Satpol PP) Jakarta Timur ditempatkan di hutan kota Cawang, Jakarta Timur untuk menjaga keamanan.

Banyak orang di Twitter menyatakan sikap anti-LGBT dan tidak mendukung interaksi LGBT. Untuk memahami pandangan anti-LGBT dalam masyarakat Indonesia, dapat digunakan analisis sentimen. Analisis sentimen ialah suatu metode untuk mengidentifikasi sikap emosional dalam teks, seperti ulasan, komentar, atau posting media sosial. Tujuannya adalah untuk menentukan apakah sentimen dalam teks tersebut positif, netral, atau negatif. Pendapat yang dipakai pada analisis ini diperoleh dari Twitter, di mana pengguna Indonesia mengungkapkan pandangan mereka tentang isu LGBT. Menurut Rozi dkk (2012), analisis sentimen berguna untuk mengklasifikasikan data dengan cepat, sehingga kita dapat memahami opini yang terkandung dalam kalimat-kalimat tersebut. Salah satu metode yang dapat dipakai dalam analisis sentimen ini ialah *Naïve Bayes Classifier* (NBC). NBC dianggap mudah namun mempunyai tingkat akurasi yang tinggi untuk melakukan klasifikasi sentimen. Tingkat akurasi ini dipengaruhi oleh jumlah data uji yang digunakan dalam proses analisis.

Penelitian yang berkaitan dengan analisis sentimen telah dilakukan Fitri dkk., (2019). Pada penelitian tersebut dilakukan perbandingan antara algoritma NBC, *Decision Tree*, dan *Random Forest* dalam mengklasifikasikan sentimen masyarakat Indonesia terhadap kasus kampanye anti-LGBT di Indonesia. Hasil penelitian menunjukkan NBC lebih baik dibandingkan *Decision Tree* dan *Random Forest*. Penelitian lain juga dilakukan oleh Daffa dan Apriade (2023) mengenai pengklasifikasian sentimen masyarakat Indonesia mengenai anti-LGBT melalui sosial media Twitter. Pada penelitian ini NBC juga baik dalam mengklasifikasikan sentimen masyarakat Indonesia terkait isu anti-LGBT.

Dalam penelitian terdahulu, metode ekstraksi fitur TF-IDF digunakan dalam NBC untuk menaikkan performanya. Penelitian sebelumnya telah mencoba menggunakan ekstraksi fitur *unigram* dan *negasi*, tetapi hasil yang diperoleh belum memuaskan. Oleh sebab itu, digunakan metode ekstraksi fitur lain, yaitu TF-IDF. Salah satu kelebihan TF-IDF adalah kemampuannya dalam menunjukkan pentingnya kata-kata dalam dokumen. TF-IDF memberi bobot terhadap kata-kata yang didasarkan pada jumlah kemunculannya. Dengan begitu, kata yang banyak timbul pada sebuah dokumen tetapi sedikit muncul dalam dokumen lain mempunyai bobot yang lebih besar. Berdasarkan latar belakang tersebut, penelitian dilakukan untuk melihat pandangan masyarakat Indonesia terhadap kampanye anti-LGBT, apakah mereka cenderung menentang atau mendukung kampanye tersebut.

## II. METODE PENELITIAN

### A. Sumber Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder yang didapatkan melalui proses *crawling* atau pengumpulan data dari Twitter dengan *Google Colaboratory*. Data yang digunakan adalah kampanye anti-LGBT yang terjadi di masyarakat Indonesia pada periode Januari-Maret 2023. Variabel penelitian terdiri dari variabel respon (Y) dan variabel prediktor (X), ditampilkan pada Tabel 1.

**Tabel 1.** Variabel Penelitian

Variabel	Keterangan
Y	Kelas data menggunakan sentimen positif dan sentimen negatif.
X	Frekuensi kemunculan kata dalam dokumen.

### B. Tahapan Analisis Data

Berikut ini merupakan tahap-tahap analisis sentimen pada NBC.

#### 1. Memberi label pada data

Pada tahap pelabelan data dengan mengumpulkan *tweet* dari Twitter berdasarkan opini positif dan negatif. Ketentuan opini positif adalah opini yang mendukung gerakan kampanye anti-LGBT dan opini negatif adalah opini yang menentang gerakan kampanye anti-LGBT. Proses pemberian label pada data terjadi secara otomatis dengan melibatkan metode berbasis leksikon (*lexicon-based*).

#### 2. *Preprocessing Data*

Tahapan ini ialah tahap awal, yang bertujuan untuk membersihkan data dengan tujuan mengatasi masalah pada dataset dan mempertahankan konsistensi data yang digunakan. Berikut merupakan langkah-langkah dalam *preprocessing data*.

- a. *Cleaning* merupakan tahap di mana karakter dan tanda baca yang tidak diperlukan harus dihapus dari ulasan untuk membersihkan data. Contoh karakter yang dihapuskan seperti URL, tag (#), titik (.), koma (,) maupun tanda baca lainnya.
- b. *Case Folding* merupakan proses yang mengubah huruf besar menjadi huruf kecil.
- c. *Stopword* adalah menghilangkan kata imbuhan, kata hubung, atau kata depan.
- d. *Stemming* adalah mengganti kata ke kata dasar yang sesuai kamus KBBI.
- e. *Tokenisasi* adalah proses membagi teks menjadi unit kata yang biasa disebut sebagai token. Pada penelitian ini, kata-kata akan dipisahkan berdasarkan *unigram*.

### 3. Memberi bobot kata

Tahapan ini, menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk memproses teks. Metode ini memberikan bobot pada kata berdasarkan seberapa sering kata tersebut muncul dalam dokumen dan seberapa umum kata tersebut dalam korpus (seluruh koleksi dokumen). Tujuannya adalah mengubah kata ke dalam data numerik. TF-IDF menggabungkan *Term Frequency* (TF), yang mengukur frekuensi kata dalam dokumen, dengan *Inverse Document Frequency* (IDF), yang mengukur keumuman atau kelangkaan kata dalam keseluruhan koleksi dokumen (Widyasanti dkk, 2018). Persamaan TF, IDF, dan TF-IDF disajikan oleh persamaan (1), (2), dan (3).

$$tf_{ij} = \frac{freq_i(d_j)}{\sum_{i=1}^k freq_i(d_j)} \quad (1)$$

Persamaan (1) memberikan gambaran tentang seberapa penting sebuah kata dalam sebuah dokumen.

$$idf_i = \log \frac{N}{df_i} \quad (2)$$

Nilai IDF menghasilkan bobot yang lebih besar dalam kata yang jarang muncul pada korpus dan bobot yang lebih kecil daripada kata umum.

$$(tfidf)_{ij} = tf_{ij} \times idf_i \quad (3)$$

TF-IDF menggabungkan konsep *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) untuk mendapatkan bobot kata pada sebuah dokumen. Dengan metode TF-IDF, kata yang sering muncul pada dokumen tertentu namun sedikit muncul pada dokumen lain di korpus memberikan bobot yang lebih tinggi.

Keterangan:

$freq_i(d_j)$  : jumlah kemunculan kata ke-i pada dokumen ke-j

$tf_{ij}$  : jumlah kemunculan kata ( $t_i$ ) pada dokumen ( $d_j$ )

$idf_i$  : logaritma dari perbandingan jumlah total dokumen dengan jumlah dokumen yang mengandung kata

$(tfidf)_{ij}$  : bobot kata ke-i yang terkait pada dokumen ke-j

$N$  : jumlah total dokumen

$df_i$  : jumlah dokumen yang mengandung kata

$i$  : 1,2,3,...,k ; k merupakan banyak kata yang dihasilkan

$j$  : 1,2,3,...,n ; n merupakan banyak dokumen.

### 4. Pemisahan Data *Training* dan Data *Testing*.

Data dipisah menjadi data *training* dan data *testing*, dengan rasio 80:20. Rasio tersebut, seperti yang disebutkan oleh Hastie (2009), merupakan rasio umum yang digunakan dalam analisis sentimen. Pembagian tersebut bertujuan untuk memanfaatkan 80% merupakan data *training* untuk melatih model dan 20% merupakan data *testing* untuk menguji kinerja model. Penggunaan rasio tersebut diharapkan dapat menghasilkan hasil yang optimal pada analisis sentimen.

5. Klasifikasi Sentimen dengan *Naïve Bayes Classifier*

*Naïve Bayes Classifier* (NBC) adalah suatu algoritma yang memanfaatkan konsep probabilitas dan sering dipakai untuk melakukan klasifikasi sentimen (Indarbensyah dan Rochmawati, 2021). NBC merupakan model algoritma yang sederhana dan mudah digunakan, namun tetap efektif dalam memprediksi kasus-kasus berdasarkan hasil klasifikasi teks.

Menurut Hamzah (2012), metode NBC melibatkan dua tahap penting pada proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. dalam tahap pelatihan, digunakan analisis pada sampel dokumen untuk memilih kosa kata, yaitu kata yang bakal muncul pada koleksi dokumen sampel dan mewakili keseluruhan dokumen. Selain itu, probabilitas prior pada sebuah kategori didasari pada sampel dokumen yang ada. Dalam tahap klasifikasi, nilai kategori sebuah dokumen didasari pada kemunculan kata-kata dalam dokumen tersebut yang sedang diproses klasifikasi. Berikut persamaan yang digunakan.

$$P(V_j|a_i) = \frac{P(a_i|V_j)P(V_j)}{P(a_i)} \tag{4}$$

- Dimana,
- $P(V_j|a_i)$  : peluang kategori-j pada saat kemunculan *term-i*
- $P(a_i|V_j)$  : peluang kata-*i* masuk pada kategori j
- $P(V_j)$  : peluang kemunculan sebuah kategori j
- $P(a_i)$  : peluang kemunculan sebuah kata ke-*i*
- i* : indeks kata, dari 1 sampai ke-*k*
- j* : indeks kategori, dari 1 sampai ke-*n*

Persamaan (4) merupakan rumus *Teorema Bayes* yang berfungsi untuk menghitung peluang kelas sentimen didasarkan pada kata dalam dokumen. Pada algoritma NBC sebuah dokumen direpresentasikan terhadap pasangan atribut  $(a_1, a_2, \dots, a_k)$  dengan  $a_1$  merupakan kata pertama,  $a_2$  merupakan kata kedua dan berikutnya. Pada tahap klasifikasi, NBC mendapatkan kategori/kelas yang paling besar peluangnya dengan memberikan atribut  $(a_1, a_2, \dots, a_k)$  dengan *V* merupakan himpunan kelas.  $V_{map}$  merupakan kategori dengan probabilitas terbesar. Berikut persamaan yang dipakai.

$$V_{map} = \arg \max_{V_j \in V} P(V_j) \prod_{i=1} P(a_i|V_j) \tag{5}$$

Nilai  $P(V_j)$  dan probabilitas  $a_i$  untuk sebuah kategori  $P(a_i|V_j)$  dihitung pada data training, rumus yang digunakan sebagai berikut.

$$P(a_i|V_j) = \frac{n_i+1}{n+|vocabulary|} \tag{6}$$

Persamaan (6) dipakai untuk mengkomputasi peluang kemunculan kata pada kelas sentimen yang berbeda. Hasil peluang ini akan dipakai bersama dengan probabilitas kelas sentimen lainnya untuk menghitung probabilitas keseluruhan. Probabilitas keseluruhan tersebut berfungsi mengelompokkan dokumen baru ke dalam kelas sentimen yang telah ditetapkan.  $n_i$  merupakan frekuensi kata  $a_i$  pada kategori  $V_j$ . *n* merupakan jumlah kosakata yang ada pada kategori  $V_j$  dan *vocabulary* merupakan jumlah kata dalam data secara keseluruhan.

6. Evaluasi Model

Evaluasi klasifikasi menggunakan *confusion matrix* berupa tabel yang untuk mengevaluasi kinerja model klasifikasi. *confusion matrix* ditampilkan dalam Tabel 2.

**Tabel 2.** *Confusion Matrix*

	Actual	
	Yes	No
Prediction	Yes TP (True Positive)	No FP (False Positive)
	No FN (False Negative)	TN (True Negative)

Dengan menggunakan informasi dalam *confusion matrix*, beberapa evaluasi kinerja model dapat dihitung, sebagai berikut:

$$a. \text{ Akurasi (acc)} = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

$$b. \text{ Presisi (pre)} = \frac{TP}{TP+FP} \tag{8}$$

$$c. \text{ Recall (re)} = \frac{TN}{TN+FP} \tag{9}$$

Ketiga ukuran evaluasi yang digunakan memiliki fungsi yang berbeda. Akurasi mengukur kemampuan model dalam mengklasifikasikan seluruh data secara benar. Presisi melihat sejauh mana prediksi positif yang dibuat pada model adalah benar atau relevan. Recall melihat sejauh mana model dapat menemukan atau mengklasifikasikan dengan benar semua sampel positif. Secara keseluruhan, akurasi, presisi, dan recall memberikan informasi yang komprehensif tentang kinerja model klasifikasi.

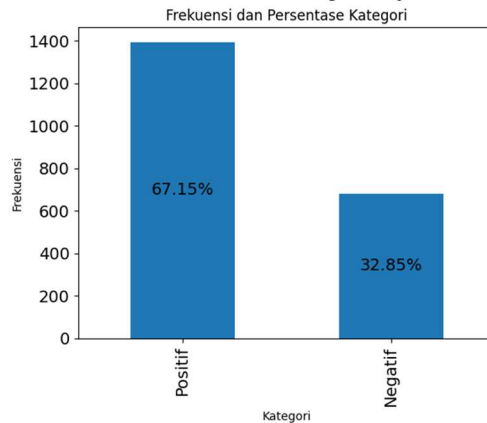
7. Visualisasi kata dengan *word cloud*

*Word cloud* adalah visualisasi dari kata-kata yang banyak muncul dalam data teks. Pada analisis sentimen, *word cloud* berfungsi untuk membedakan kata-kata yang paling banyak muncul dalam sebuah kategori sentimen. Melalui *word cloud*, kata-kata bisa jelas dilihat mana yang paling relevan pada setiap kategori sentimen secara visual.

### III. HASIL DAN PEMBAHASAN

#### A. Pelabelan Data

Pelabelan sentimen didapatkan menggunakan metode berbasis leksikon. Visualisasi sebaran hasil pelabelan dapat dilihat pada Gambar 1, yang memperlihatkan analisis sentimen terbagi menjadi kategori positif dan negatif.



Gambar 1. Sebaran Kelas Sentimen

Berdasarkan Gambar 1, mayoritas masyarakat Indonesia memberikan pandangan positif terhadap kampanye anti LGBT. Persentase menunjukkan bahwa sentimen positif mendominasi dalam data yang dianalisis, sementara sentimen negatif tetap ada meskipun dengan persentase yang lebih rendah. Hal ini menunjukkan ketidakseimbangan dalam distribusi sentimen, dengan mayoritas masyarakat memiliki pandangan yang positif terhadap kampanye tersebut.

#### B. Preprocessing Data

Tahapan ini, *tweet* mentah diproses dulu pada tahap *preprocessing* (*case folding, tokenizing, stopword removal, dan stemming*). Tahapan *preprocessing* ditampilkan pada Tabel 3.

Tabel 3. Contoh *Tweet Preprocessing*

Nama Proses	Hasil Proses
<b>Tweet</b>	@bensimmoon Mereka kampanye anti kriminalisasi bukan ngajak org jd lgbt kocak!
<b>Cleaning</b>	Mereka kampanye anti kriminalisasi bukan ngajak org jd lgbt kocak
<b>Case Folding</b>	mereka kampanye anti kriminalisasi bukan ngajak org jd lgbt kocak

Nama Proses	Hasil Proses
<b>Tokenizing</b>	['kampanye', 'anti', 'kriminalisasi', 'ngajak', 'org', 'lgbt', 'kocak']
<b>Stopword</b>	Kampanye anti kriminalisasi ngajak org lgbt kocak
<b>Normalization</b>	Kampanye anti kriminalisasi ajak orang lgbt lucu
<b>Stemming</b>	kampanye : kampanye anti : anti kriminalisasi : kriminalisasi ngajak : ajak lgbt : lgbt kocak : lucu

Tabel 3 menunjukkan tweet yang belum diolah terlihat tidak terstruktur dan banyak kata-kata yang tidak bermakna. Oleh karena itu, perlu dilakukan pembersihan data dengan menghapus tanda baca, simbol, dan kata-kata tak bermakna. Selanjutnya, teks diubah menjadi huruf kecil, dipisahkan menjadi kata-kata individu (*tokenizing*), dan kata-kata tak bermakna dihapus (*stopword removal*). Kata-kata yang tersisa dinormalisasi sesuai KBBI dan kemudian dilakukan *stemming* untuk mengubahnya ke bentuk dasar.

### C. Pembobotan

Pada tahap pembobotan menggunakan metode TF-IDF. TF-IDF bertujuan memberi bobot pada kata dengan cara berapa sering kata tersebut muncul pada dokumen dengan kata yang sedikit muncul dalam dokumen lain diberi dengan bobot yang tinggi dan sebaliknya. hasil pembobotan kata disajikan dalam Tabel 4.

**Tabel 4.** Pembobotan Kata dengan TF-IDF

No.Dok	Kata	aamiin	abai	abaikan	....	ada
1.		0	0	0	....	0
2.		0	0	0	....	0
3.		0	0	0	....	0
:	:	:	:	:	....	....
9.		0	0	0	....	0.2233507

Berdasarkan Tabel 4 diperoleh hasil pada kata “aamiin”, “abai”, “abaikan” diperoleh 0, yang artinya tidak ada kata tersebut pada dokumen. Sementara itu, kata “ada” pada dokumen 9 memiliki nilai sebesar 0,2233507 sebab kata tersebut berada pada dokumen 9. Pada TF-IDF bobot kata di pengaruhi oleh semakin jarang kata tersebut dibicarakan pada dokumen lain semakin tinggi bobot kata tersebut.

### D. Naïve Bayes Classifier

Penerapan NBC pada *Python*, menggunakan library *sklearn.naive\_bayes* dengan mengimpor modul *MultinomialNB*. Algoritma NBC menghasilkan dua klasifikasi yaitu positif dan negatif. Contoh hasil pengujian menggunakan NBC ditampilkan pada Tabel 5.

**Tabel 5.** Contoh Probabilitas Klasifikasi pada Kampanye Anti-LGBT

Dok	Tweet	Negatif	Positif	Sentimen
1	tidak ada calon presiden kampanye janji undang-undang anti-lgbt isu bisa terima	0,000108	0,0006655	Negatif
2	Hanya kampanye politik gila yang kambing hitam kaum anti-lgbt populer harap bisa menang	3,3275e-5	7,3205e-5	Positif

Dalam Tabel 5, hasil klasifikasi menunjukkan bahwa dokumen 1 diklasifikasikan sebagai sentimen negatif karena probabilitas negatif lebih besar daripada probabilitas positif. Sementara itu, dokumen 2 diklasifikasikan sebagai sentimen positif karena probabilitas positif lebih besar daripada probabilitas negatif.



**E. Evaluasi Performa Model**

Metode yang digunakan dalam mengevaluasi model klasifikasi analisis sentimen adalah *confusion matrix*. Tabel *confusion matrix* berfungsi membagi data *training* dan data *testing* 80%:20% ditampilkan dalam Tabel 6.

**Tabel 6. Confusion Matrix**

Prediksi	Aktual	
	Negatif	Positif
Negatif	13	1
Positif	129	273

Berdasarkan Tabel 6, terlihat bahwa model berhasil mengklasifikasikan dengan benar 273 data merupakan kelas positif dan 13 data merupakan kelas negatif. Namun, terdapat 129 data kelas negatif yang salah diprediksi sebagai kelas positif, dan 1 data kelas positif yang salah diprediksi sebagai kelas negatif. Untuk menghitung akurasi, presisi, dan recall, digunakan persamaan (7), (8), dan (9). Hasil akurasi, presisi, dan recall dapat ditemukan di Tabel 7.

**Tabel 7. Matriks Evaluasi**

Matriks Evaluasi	Hasil Evaluasi
Akurasi	0,6875
Presisi	0,996
Recall	0,928

Berdasarkan Tabel 7, algoritma NBC mampu mengklasifikasikan model dengan benar sebesar 68,75%, NBC juga mampu memprediksi sentimen positif dengan tepat sebesar 99,6%, dan mampu mengklasifikasikan sentimen positif dengan baik sebesar 92,8%. Hasil *confusion matrix* memiliki ketepatan klasifikasi pada data kampanye anti LGBT secara keseluruhan baik.

**F. Word Cloud**

Visualisasi tentang ulasan kampanye anti-LGBT di Twitter disajikan dalam bentuk *word cloud*. *Word cloud* berfungsi untuk menggambarkan kata-kata yang banyak muncul dalam data dengan ukuran kata yang berbeda-beda tergantung pada seberapa banyak kata tersebut muncul pada keseluruhan data. Hasil visualisasi dapat dilihat pada Gambar 2.



**Gambar 2.** (a) *Word cloud* Komentar Positif dan (b) *Word cloud* Komentar Negatif

Terlihat pada Gambar 2 (a) adalah *word cloud* dari keseluruhan ulasan positif masyarakat Indonesia mengenai kampanye anti-LGBT melalui media sosial Twitter, sedangkan (b) adalah *word cloud* dari ulasan negatif masyarakat Indonesia melalui media sosial Twitter. Berdasarkan *word cloud* tersebut dapat diketahui ulasan positif masyarakat Indonesia mengenai kampanye anti-LGBT dikaitkan dengan kata "LGBT", "kampanye", "anti", "tidak", "yang", "dukung". Hal ini menunjukkan adanya pemberitahuan bahwa seseorang tersebut tidak LGBT dan mendukung kampanye anti LGBT. Sedangkan pada tanggapan negatif menunjukkan perbincangan sering dikaitkan dengan kata "LGBT", "anti lgbt", "kampanye", "yang", "menolak". Hal ini menunjukkan penolakan terhadap kampanye anti-LGBT.

#### IV. KESIMPULAN

Analisis menggunakan dataset ulasan dari Twitter dan algoritma NBC menunjukkan bahwa sentimen masyarakat Indonesia terhadap kampanye anti-LGBT memiliki tingkat kecocokan sebesar 68,75%, dengan recall 92,8% dan presisi 99,6%. Secara umum, ulasan positif dari masyarakat Indonesia mendukung kampanye tersebut dan menyatakan bahwa mereka bukan LGBT. Di sisi lain, ulasan negatif mencerminkan penolakan terhadap kampanye tersebut. Persentase sentimen positif adalah 67,15% sedangkan sentimen negatif adalah 32,85%. Dengan demikian, mayoritas masyarakat Indonesia mendukung kampanye anti-LGBT. Saran untuk penelitian selanjutnya adalah meningkatkan referensi yang digunakan dalam tahap preprocessing teks untuk mengidentifikasi kata-kata yang kurang relevan dalam bahasa Indonesia, sehingga model dapat lebih memahami konteks teks dan memberikan hasil yang lebih optimal.

#### DAFTAR PUSTAKA

- Anjani, A., Chamid, A., & Murti, A. (2022). Analisis Sentimen Kaum LGBT pada Media Sosial Twitter Menggunakan Algoritma Naïve Bayes. *JTINFO : Jurnal Teknik Informatika*, 1(2), 1–8.
- Ardras, D. W., & Voutama, A. (2023). Analisis Sentimen Anti Lgbt Di Indonesia Melalui Media Sosial Twitter. *Jurnal Teknik*, 15(1), 23–28. <https://doi.org/10.30736/jt.v15i1.926>
- Chandra, M., Utomo, C., Taukhid, M., & Mujahidin, S. (2023). Analisis Sentimen Media Sosial Twitter pada Kasus Pemberlakuan Pembatasan Kegiatan Masyarakat dengan menggunakan Metode Naïve Bayes Classifier. 01, 41–48.
- Fitri, V. A., Andreswari, R., & Hasibuan, M. A. (2019). Sentiment analysis of social media Twitter with case of Anti-LGBT campaign in Indonesia using Naïve Bayes, decision tree, and random forest algorithm. *Procedia Computer Science*, 161, 765–772. <https://doi.org/10.1016/j.procs.2019.11.181>
- Hamzah, A. (2012). Klasifikasi Teks Dengan Naïve Bayes Classifier (NBC) Untuk Pengelompokan Teks Berita Dan Abstract Akademis. *Prosiding Seminar Nasional Aplikasi Sains & Teknologi (SNAST) Periode III, 2011*, 269–277. <https://doi.org/1979-911X>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd ed.)*. Springer, Bab 7.10 "Model Assessment and Selection".
- Imam Fahrur Rozi, Sholeh Hadi Pramono, & Erfan Achmad Dahlan. (2012). Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi. *Jurnal EECCIS*, 6(1), 37–43.
- Indarbensyah, P. P. E., & Rochmawati, N. (2021). Penerapan N-Gram menggunakan Algoritma Random Forest dan Naïve Bayes Classifier pada Analisis Sentimen Kebijakan PPKM 2021. *Journal of Informatics and Computer Science (JINACS)*, 2(04), 235–244. <https://doi.org/10.26740/jinacs.v2n04.p235-244>
- Liu, Bing. (2012) *Sentiment Analysis and Opinion Mining*, Chicago, Morgan & Claypool.
- Mailoa, F. F. (2021). Analisis sentimen data twitter menggunakan metode text mining tentang masalah obesitas di indonesia. *Journal of Information Systems for Public Health*, 6(1), 44. <https://doi.org/10.22146/jisph.44455>
- Mukhid, A. (2019). Kajian Teoritis Tentang Perilaku Lesbian, Gay, Biseksual, Transgender (LGBT) Dalam Perspektif Psikologis dan Teologis. *Sophist : Jurnal Sosial Politik, Kajian Islam Dan Tafsir*, 1(1), 53-75. <https://doi.org/10.20414/sophist.v1i1.756>
- Sari, F. V., & Wibowo, A. (2019). Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon Emosi. *Jurnal SIMETRIS*, 10(2), 681–686. <https://jurnal.umk.ac.id/index.php/simet/article/view/3487/1883>
- Widyasanti, N. K., Putra, I.K., & Rusjyanthi, N. K. (2018). Seleksi Fiur Bobot Kata dengan Metode TF-IDF untuk Ringkasan Bahasa Indonesia. *Jurnal Evolusi*, 4(1), 25-31.