

**ANALISIS SENTIMEN KAMPANYE ANTI-LGBT PADA TWITTER  
MENGUNAKAN METODE *NAÏVE BAYES CLASSIFIER*  
DENGAN EKSTRAKSI FITUR *N-GRAM***

**SKRIPSI**

*Diajukan sebagai salah satu persyaratan untuk memperoleh gelar  
Sarjana Statistika*



**Oleh  
RIOS DACOSTA  
NIM 19337013**

**PROGRAM STUDI SARJANA STATISTIKA  
DEPARTEMEN STATISTIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI PADANG  
2024**

## PERSETUJUAN SKRIPSI

### ANALISIS KAMPANYE ANTI-LGBT PADA TWITTER MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER* DENGAN EKSTRAKSI FITUR N-GRAM

Nama : Rios Dacosta  
NIM : 19337013  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

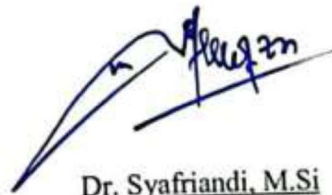
Padang, 12 Februari 2024

Mengetahui:  
Kepala Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si  
NIP. 1984022320101220005

Disetujui Oleh:  
Pembimbing



Dr. Syafrandi, M.Si  
NIP. 196609081991031003

## PENGESAHAN LULUS UJIAN SKRIPSI


Nama : Rios Dacosta  
NIM : 19337013  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

### ANALISIS KAMPANYE ANTI-LGBT PADA TWITTER MENGGUNAKAN METODE *NAIVE BAYES CLASSIFIER* DENGAN EKSTRAKSI FITUR N-GRAM

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi  
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam  
Universitas Negeri Padang

Padang, 23 Februari 2024

#### Tim Penguji

	Nama	Tanda Tangan
Ketua	: Dr. Syafriandi, M.Si	
Anggota	: Dr. Dony Permana, M.Si	
Anggota	: Dina Fitria, M.Si	

## SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertandatangan di bawah ini:

Nama : Rios Dacosta  
NIM : 19337013  
Program Studi : S1 Statistika  
Departemen : Statistika  
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul "**Analisis Sentimen Kampanye Anti-LGBT Pada Twitter Menggunakan Metode *Naïve Bayes Classifier* Dengan Ekstraksi Fitur *N-Gram***" adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan dengan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan.

Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran rasa tanggung jawab sebagai anggota masyarakat ilmiah.

Diketahui oleh,  
Kepala Departemen Statistika,



Dr. Yenni Kurniawati, S.Si., M.Si  
NIP. 198402232010122005

Saya yang menyatakan,



Rios Dacosta  
NIM. 19337013

# Analisis Sentimen Kampanye Anti-LGBT pada Twitter Menggunakan Metode *Naïve Bayes Classifier* dengan Ekstraksi Fitur N-Gram

Rios Dacosta

## ABSTRAK

Kampanye anti-LGBT (*Lesbian, Gay, Biseksual, dan Transgender*) merujuk pada upaya yang dilakukan oleh sekelompok atau individu untuk mempromosikan pandangan yang menentang hak-hak, kesetaraan, dan pengakuan komunitas LGBT. Kampanye anti-LGBT sering kali menjadi isu yang kontroversial dan memicu perdebatan yang hangat dalam masyarakat. Masyarakat yang memiliki pandangan berbeda tentang hak-hak LGBT mencerminkan adanya perbedaan dalam nilai-nilai, keyakinan agama, dan pemahaman sosial yang beragam. Pandangan yang berbeda pada kampanye anti-LGBT diungkapkan ke dalam sebuah opini yang mengandung sentimen positif atau negatif. Opini tersebut disampaikan melalui platform Twitter yang disebut *tweet*. Analisis yang digunakan untuk mengolah *tweet* menjadi sebuah informasi adalah analisis sentimen, yang dapat memberikan pemahaman yang lebih mendalam tentang bagaimana masyarakat menanggapi kampanye tersebut.

Metode yang digunakan dalam pengklasifikasian sentimen yaitu *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur N-Gram. Metode ini telah terbukti efektif dalam menganalisis sentimen karena kemampuannya dalam mengklasifikasikan *tweet* berdasarkan kemungkinan probabilitasnya. Penggunaan ekstraksi fitur N-Gram dilakukan untuk mengekstrak fitur-fitur penting dari data. N-Gram mengacu pada urutan  $n$  kata yang berurutan dalam sebuah *tweet*.

Sebelum memasuki tahap pemodelan, *tweet* dikumpulkan menggunakan teknik *crawling* dan melalui tahapan seperti *text preprocessing*, pelabelan berbasis leksikon, serta pembobotan data. Model yang dihasilkan dinilai berdasarkan evaluasi menggunakan perhitungan *confusion matrix* dari hasil prediksi. Hasil penelitian menunjukkan bahwa dari 1.315 *tweet* yang dianalisis menggunakan metode *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur N-Gram, 56% diklasifikasikan sebagai *tweet* yang mengandung sentimen positif dan 44% diklasifikasikan sebagai *tweet* yang mengandung sentimen negatif terkait pandangan masyarakat terhadap kampanye anti-LGBT. Selanjutnya, setelah pembagian data *training* dan data *testing* dengan rasio 90%:10% serta perhitungan *confusion matrix*, diperoleh hasil evaluasi dengan akurasi sebesar 76,19%, presisi sebesar 80%, recall sebesar 90,09%, dan F1-score sebesar 85,10%. Hasil ini menunjukkan bahwa metode NBC dengan ekstraksi fitur N-Gram memberikan hasil klasifikasi yang baik.

**Kata Kunci:** Analisis Sentimen, Kampanye Anti-LGBT, *Naïve Bayes Classifier*, N-Gram

# Sentiment Analysis of Anti-LGBT Campaign on Twitter Using *Naïve Bayes Classifier* with N-Gram Feature Extraction

Rios Dacosta

## ABSTRACT

Anti-LGBT (Lesbian, Gay, Bisexual, and Transgender) campaigns refer to efforts made by groups or individuals to promote views that oppose the rights, equality, and recognition of the LGBT community. These campaigns often become controversial issues that trigger heated debates within society. The diverse perspectives on LGBT rights reflect differences in values, religious beliefs, and social understanding. The varying viewpoints on anti-LGBT campaigns are expressed through opinions conveying positive or negative sentiments, which are shared on the Twitter platform as tweets. Analyzing these tweets using sentiment analysis can provide deeper insights into how the public responds to such campaigns.

The method used for sentiment classification is the Naïve Bayes Classifier (NBC) with N-Gram feature extraction. This method has proven to be effective in analyzing sentiment due to its ability to classify tweets based on probability. The use of N-Gram feature extraction is done to extract important features from the data. N-Gram refers to sequences of  $n$  consecutive words in a tweet.

Before entering the modeling stage, tweets were collected using the crawling technique and went through stages such as text preprocessing, lexicon-based labeling, and data weighting. The resulting model was evaluated based on the calculation of the confusion matrix from the prediction results. The research findings show that from the 1,315 tweets analyzed using the Naïve Bayes Classifier (NBC) method with N-Gram feature extraction, 56% were classified as tweets containing positive sentiment and 44% were classified as tweets containing negative sentiment regarding the public's perception of anti-LGBT campaigns. Furthermore, after dividing the data into a 90%:10% training and testing ratio and calculating the confusion matrix, the evaluation results were as follows: accuracy of 76.19%, precision of 80%, recall of 90.09%, and an F1-score of 85.10%. These findings indicate that the NBC method with N-Gram feature extraction provides good classification results.

**Keywords:** Anti-LGBT Campaign, Naïve Bayes Classifier, N-Gram, Sentiment Analysis

## KATA PENGANTAR

Puji syukur penulis ucapkan kepada Allah SWT, atas rahmat dan karunianya yang telah dilimpahkannya. Penulis dapat menyelesaikan penulisan skripsi ini yang berjudul “Analisis Sentimen Kampanye Anri-LGBT Menggunakan *Naive Bayes Classifier* Dengan Ekstraksi Fitur N-Gram”.

Penulisan skripsi ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan pada Program Studi Statistika (S1) Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang. Dalam menyelesaikan skripsi ini tidak lepas dari bimbingan dan bantuan dari berbagai pihak. Penulis menyadari bahwa skripsi ini mempunyai keterbatasan sesuai dengan ilmu yang dimiliki, oleh karena itu penulis menyambut baik saran dan kritik dari para pembaca demi perbaikan isi skripsi ini dan penulis mengucapkan terima kasih kepada:

1. Ibu Dr. Yenni Kurniawati., M.Si., Kepala Departemen Statistika dan Koordinator Program Studi S1 Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.
2. Bapak Dr. Syafriandi., M.Si., Dosen Pembimbing Skripsi dan Penasehat Akademik.
3. Bapak Dr. Dony Permana., M.Si., dan Ibu Dina Fitria., M.Si., Dosen Penguji Skripsi.
4. Bapak dan Ibu Dosen serta Tenaga Kependidikan Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.

5. Kedua orang tua dan saudara/i yang telah memberikan semangat, nasehat, dukungan dan doa selama masa perkuliahan dan dalam menyelesaikan skripsi ini.
6. Semua teman-teman yang selalu memberikan semangat serta dukungan kepada penulis selama menempuh perkuliahan di Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang.

Semoga semua kebaikan dan ketulusan dibalas oleh Allah SWT sebagai amal ibadah. Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan masih terdapat kekurangan di dalamnya. Oleh karena itu, penulis mengharapkan masukan dan saran yang membangun dari pembaca. Semoga penulisan skripsi ini dapat bermanfaat bagi seluruh pihak yang memerlukannya.

Padang, Februari 2024  
Penulis

Rios Dacosta



## DAFTAR ISI

<b>ABSTRAK .....</b>	<b>i</b>
<b>KATA PENGANTAR.....</b>	<b>iii</b>
<b>DAFTAR ISI.....</b>	<b>v</b>
<b>DAFTAR TABEL .....</b>	<b>vii</b>
<b>DAFTAR GAMBAR.....</b>	<b>ix</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>x</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
A. Latar Belakang Masalah.....	1
B. Batasan Masalah .....	6
C. Rumusan Masalah.....	6
D. Tujuan Penelitian .....	7
E. Manfaat Penelitian .....	7
<b>BAB II KAJIAN TEORITIS.....</b>	<b>8</b>
A. Lesbian, Guy, Biseksual, Transgender.....	8
B. Twitter.....	14
C. Text Mining .....	21
1. Preprocessing Data.....	21
2. Ekstraksi Fitur N-Gram.....	27
D. Analisis Sentimen .....	31
E. Machine Learning .....	35
1. <i>Supervised Learning</i> .....	36
2. <i>Unsupervised Learning</i> .....	37
F. Klasifikasi .....	37
G. Data Training dan Data Testing.....	42
1. <i>Data Training</i> .....	42
2. <i>Data Testing</i> .....	42
H. Confusion Matrix.....	43
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>47</b>

A. Jenis Penelitian.....	47
B. Sumber Data Penelitian.....	47
C. Metode Pengumpulan Data.....	49
D. Teknik Analisis Data.....	49
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>54</b>
A. Hasil Penelitian .....	54
1. Preprocessing .....	54
2. Pelabelan data.....	60
3. Ekstraksi menggunakan N-Gram ( $n = 2$ ) .....	62
4. Klasifikasi.....	64
B. Pembahasan.....	71
<b>BAB V PENUTUP.....</b>	<b>74</b>
A. Kesimpulan .....	74
B. Saran .....	75
<b>DAFTAR PUSTAKA.....</b>	<b>76</b>
<b>LAMPIRAN.....</b>	<b>79</b>

## DAFTAR TABEL

<b>Tabel</b>	<b>Halaman</b>
1. Ketentuan Pelabelan VADER.....	35
2. Contoh <i>Confusion Matrix</i> .....	43
3. Kriteria Kualitas Akurasi.....	44
4. Kriteria Kualitas Presisi.....	45
5. Kriteria Kualitas Recall.....	45
6. Kriteria Kualitas F1-Score.....	46
7. Variabel Penelitian.....	47
8. Struktur Data Penelitian Sebelum <i>Preprocessing</i> .....	48
9. Struktur Data Penelitian Sesudah <i>Preprocessing</i> .....	48
10. Contoh Ulasan Twitter.....	54
11. Contoh <i>Cleaning Data</i> .....	56
12. Contoh <i>Case Folding</i> .....	57
13. Contoh <i>Stopword Removal</i> .....	57
14. Contoh <i>Spelling Normalization</i> .....	58
15. Contoh <i>Stemming</i> .....	59
16. Contoh <i>Tokenizing</i> Menggunakan <i>Bigram</i> .....	59
17. Contoh Nilai <i>Compound Score</i> Dan Label.....	61
18. Contoh Ekstraksi Fitur <i>Bigram</i> .....	63
19. Pembagian Data set.....	65
20. Contoh Frekuensi Kemunculan Pada Setiap Kelas Sentimen.....	65
21. Probabilitas <i>Bigram</i> Setiap Kelas.....	67

22. Hasil Sentimen Data <i>Training</i> Dengan NBC.....	68
23. <i>Confusion Matrix</i> .....	69

## DAFTAR GAMBAR

<b>Gambar</b>	<b>Halaman</b>
1. Grafik Negara Pengguna Twitter.....	2
2. <i>Flowchart</i> Penelitian.....	52
3. <i>Flowchat</i> Metode NBC.....	53
4. Sebaran Sentimen Positif Dan Negatif.....	62
5. Diagam Hasil Klasifikasi NBC Dengan N-Gram.....	68

## DAFTAR LAMPIRAN

<b>Lampiran</b>	<b>Halaman</b>
1. <i>Crawling Data</i> .....	79
2. <i>Preprocessing Text</i> .....	79
3. <i>Translate ke Bahasa Inggris</i> .....	83
4. <i>Pelabelan Tweet</i> .....	83
5. <i>Menghapus Sentimen Netral</i> .....	84
6. <i>Sebaran Kelas Sentimen Menggunakan VADER</i> .....	84
7. <i>Pembagian Data set</i> .....	84
8. <i>Klasifikasi NBC Dengan N-Gram</i> .....	85
9. <i>Confusion Matrix</i> .....	85
10. <i>Sebaran Frekuensi Sentimen Dengan NBC</i> .....	86

# BAB I

## PENDAHULUAN

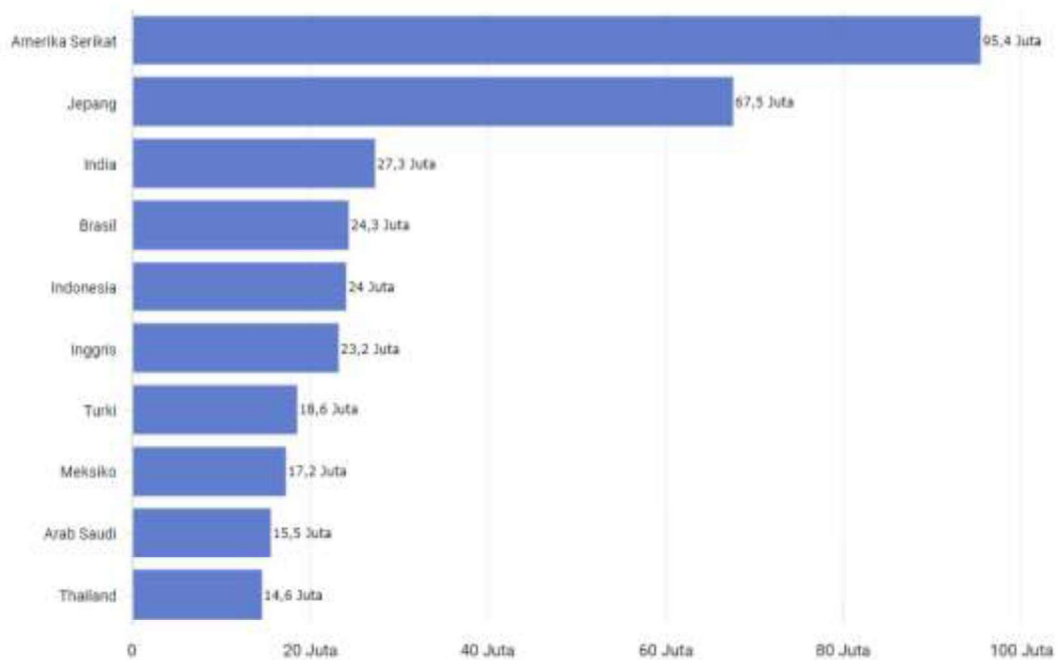
### A. Latar Belakang Masalah

Perkembangan media sosial di era globalisasi tidak terlepas dari adanya internet. Melalui internet, media sosial dapat digunakan sebagai sarana komunikasi tanpa terbatas pada jarak dan waktu. Media sosial telah digunakan banyak orang untuk berkomunikasi, berbagi konten, dan berinteraksi dalam bentuk teks, gambar, video, dan suara. Begitu banyak manfaat media sosial bagi pengguna, sehingga ada beberapa hal yang ingin dikemukakan, seperti media sosial telah memungkinkan penyebaran informasi yang cepat dan luas, pertukaran ide antara individu dan kelompok, serta menjadi alat promosi yang efektif bagi bisnis dan organisasi. Media sosial juga berperan dalam gerakan sosial dengan memperluas pemahaman tentang isu-isu penting di seluruh dunia. Banyak aplikasi media sosial yang digunakan oleh manusia, seperti Facebook, Instagram, WhatsApp, Twitter, dan Line.

Salah satu media sosial yang banyak digunakan adalah Twitter. Twitter adalah media sosial berbasis *microblogging* yang dapat digunakan untuk mengirim dan membaca pesan yang disebut “*tweet*” dengan batasan maksimal 280 karakter. Twitter memungkinkan pengguna untuk berbagi pemikiran, informasi, dan pengalaman secara real-time. Pengguna Twitter dapat memposting *tweet* dengan berbagai bentuk singkatan, *emoticon*, dan karakter lain yang mengekspresikan arti khusus dari *tweet* tersebut. Terdapat beberapa atribut menarik di Twitter, termasuk kemampuan untuk menggunakan berbagai elemen dalam *tweet* seperti bahasa asing, bahasa daerah,

*emoji*, dan elemen lainnya. Menurut Fitri (2019) Twitter menyediakan data yang dapat diakses secara bebas menggunakan *Twitter API (Application Programming Interface)*, sehingga lebih mudah untuk mengumpulkan *tweet* dalam jumlah banyak.

Pada bulan Januari 2023, jumlah pengguna Twitter di seluruh dunia mencapai 556 juta orang. Jumlah tersebut meningkat 27,4% dibandingkan pada tahun 2022 (databoks, 2023). Sepuluh negara pengguna Twitter terbanyak, disajikan pada Gambar 1.



**Gambar 1.** Grafik Negara pengguna Twitter

Sumber : Laporan Statista, 2022

Berdasarkan Gambar 1, Indonesia berada di peringkat ke-5 dengan 24 juta pengguna Twitter per Januari 2023. Jumlah pengguna Twitter yang begitu banyak membuat Twitter disebut gudang data. Menurut Mailoa (2021) pengguna Twitter biasanya meningkat saat terjadi peristiwa penting dan menarik untuk dikomentari seperti piala dunia, pemilihan presiden, kematian selebritas, dan lainnya.



Salah satu isu yang menjadi *trending topic* di Twitter sampai akhir tahun 2022 yaitu gerakan *Lesbian, Gay, Biseksual, Transgender* (LGBT). Menurut Mukhid (2019) istilah LGBT telah digunakan sejak tahun 1990-an, menggantikan istilah komunitas *gay* karena istilah tersebut dianggap lebih mewakili kelompok yang disebutkan. Di Indonesia isu LGBT merupakan isu yang belum terselesaikan, karena masih tergolong hal yang tabu dan belum sepenuhnya dipahami oleh masyarakat Indonesia (Hamzah & Maharani, 2021). Keberadaan kaum LGBT di Indonesia merupakan salah satu contoh pergeseran budaya yang telah terjadi dan menjadi perhatian masyarakat. Kebebasan berbicara kaum LGBT di media sosial Twitter memiliki berbagai pengaruh terhadap masyarakat. Melalui media sosial, LGBT dapat dengan mudah berinteraksi dengan sesama LGBT. Keberadaan LGBT tidak disebabkan oleh masalah genetik, ketidakseimbangan hormon, masalah mental, atau kejahatan, namun disebabkan oleh pengaruh lingkungan dan pengalaman pribadi.

Terdapat beberapa kejadian yang melibatkan komunitas LGBT di Indonesia. Salah satunya berawal dari pemberitaan yang diterbitkan oleh *magdelan.co* yang menyebutkan bahwa kelompok LGBT di Indonesia menghadiri Paris Parade (Festival LGBT) pada tanggal 29 Juni 2019 (Amalia, 2019). Selain kasus tersebut, terdapat juga laporan dari Polsek Megamendung Bogor yang menggagalkan rencana kegiatan yang diduga merupakan pesta yang akan diadakan oleh komunitas LGBT di sebuah vila di kawasan Puncak, Megamendung, Kabupaten Bogor. Kegiatan tersebut dikemas dengan tema *gathering* dan penyuluhan HIV/AIDS (Solihin, 2022).

Dalam konteks kasus-kasus LGBT yang sedang terjadi di Indonesia, masyarakat memiliki beragam pendapat. Pendapat tersebut nantinya dituangkan dalam sebuah

*tweet* yang dituliskan pada Twitter. Setiap *tweet* banyak mengandung berbagai Bahasa seperti Bahasa daerah, Bahasa Inggris, dan Bahasa gaul, bahkan terdapat *emoji* untuk memperkuat ekspresi mereka dalam mengartikan sebuah kejadian.

Dengan menentukan sentimen dari masyarakat tentang isu LGBT memudahkan untuk melihat bagaimana pendapat, emosi, atau persepsi orang terhadap isu tersebut (Pak & Paroubek, 2010). Informasi ini dapat digunakan untuk menganalisis pola pikir dan sikap masyarakat terhadap komunitas LGBT. Oleh karena itu, diperlukan analisis sentimen untuk menentukan sebuah *tweet* mengandung sentimen positif, sentimen negatif, dan sentimen netral. analisis sentimen adalah proses pengolahan data teks seperti ulasan, *tweet*, dan *postingan* media sosial untuk mendapatkan informasi sentimen dari dokumen (Zuhri & Saputri, 2020). Analisis sentimen berguna untuk mengklasifikasikan data dengan cepat, sehingga data sentimen dapat mengandung kalimat opini. Data opini tersebut dapat diolah menggunakan konsep *text mining*. Menurut Zuhri & Saputri (2020) *text mining* dapat memberikan penjelasan yang bermanfaat dari sekumpulan data. Oleh karena itu, berkas teks dalam format yang sistematis atau tidak sistematis adalah data sumber yang digunakan dalam *text mining*.

Menurut Permatasari dkk. (2021) banyak metode yang dapat digunakan dalam analisis sentimen di antaranya, *Naïve Bayes Classifier* (NBC), *Support Vector Machine* (SVM), *K-Nearest Neighbors* (K-NN), *Recurrent Neural Networks* (RNN), *C4.5*, dan *Lexicon Based*. Namun metode yang sering digunakan dalam analisis sentimen adalah *Naïve Bayes Classifier* (NBC). Menurut Chandra dkk. (2023) metode NBC dianggap sederhana namun memiliki tingkat akurasi yang tinggi dalam

hasil klasifikasinya, yang dipengaruhi oleh banyaknya data latih yang digunakan. Metode NBC merupakan teknik pembelajaran mesin berdasarkan probabilitas yang efektif dan akurat dalam klasifikasi teks.

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. Pada penelitian tersebut diperoleh nilai akurasi untuk metode NBC sebesar 83,43%, *Decision Tree* sebesar 82,91%, dan *Random Forest* sebesar 82,91%. Dapat disimpulkan bahwa NBC memiliki akurasi yang lebih tinggi dibandingkan algoritma yang lainnya pada kasus LGBT di Indonesia. Penelitian lain juga dilakukan oleh Ardras & Voutama (2023) mengenai pengklasifikasian sentimen masyarakat mengenai anti LGBT. Pada penelitian ini diperoleh nilai akurasi metode NBC sebesar 76,84%.

Dari beberapa penelitian yang sudah dilakukan, digunakan ekstraksi fitur *N-Gram* yang menggantikan ekstraksi fitur yang sering digunakan pada analisis sentimen yaitu *Term Frequency-Inverse Document Frequency* (TF-IDF). Menurut Indarbensyah & Rochmawati (2021) *N-gram* memiliki kelebihan yang tidak dimiliki TF-IDF yaitu *N-Gram* memperhitungkan urutan kata yang bersebelahan dalam teks yang nantinya model akan menangkap konteks lokal dan ketergantungan antara kata-kata yang berdekatan. Dalam analisis sentimen konteks lokal sangat penting dalam memahami makna teks. TF-IDF hanya memperhitungkan frekuensi kata secara terisolasi tanpa memperhatikan urutan atau konteksnya. Oleh karena itu, penelitian menggunakan *N-*

*Gram* diharapkan mampu memberikan hasil yang baik dalam analisis sentimen tentang kampanye anti-LGBT.

*N-Gram* digunakan pada tahap *text preprocessing*, selanjutnya dilakukan proses pengklasifikasian analisis sentimen pada data yang diperoleh menggunakan metode NBC. Penggunaan metode ini bekerja pada analisis sentimen yang diterapkan pada fitur *N-Gram* agar mendapatkan akurasi tertinggi menggunakan algoritma NBC (Nugroho, 2018).

Berdasarkan uraian di atas, dilakukan penelitian dengan judul “**Analisis Sentimen Kampanye Anti-LGBT Pada Twitter Menggunakan Metode *Naïve Bayes Classifier* Dengan Ekstraksi Fitur *N-Gram*”**”.

## **B. Batasan Masalah**

Batasan masalah penelitian ini adalah sebagai berikut.

1. Bahasa asing, bahasa daerah, dan penggunaan *emoji* pada *tweet* tidak digunakan.
2. Klasifikasi sentimen *tweet* dibagi menjadi dua, yaitu sentimen positif dan sentimen negatif.
3. Data yang digunakan merupakan *tweet* berbahasa Indonesia dari bulan Januari-Maret 2023.

## **C. Rumusan Masalah**

Rumusan masalah dalam penelitian ini adalah sebagai berikut.

1. Bagaimana hasil klasifikasi sentimen tentang kampanye Anti-LGBT di Indonesia menggunakan metode *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur *N-Gram*?

2. Bagaimana performa model dalam mengklasifikasikan sentimen tentang kampanye Anti-LGBT di Indonesia menggunakan *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur *N-Gram*?

#### **D. Tujuan Penelitian**

Adapun tujuan dari penelitian ini adalah sebagai berikut.

1. Mengetahui hasil klasifikasi sentimen tentang kampanye Anti-LGBT di Indonesia menggunakan metode *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur *N-Gram*.
2. Mengetahui performa model dalam mengklasifikasikan sentimen tentang kampanye Anti-LGBT di Indonesia menggunakan *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur *N-Gram*.

#### **E. Manfaat Penelitian**

Manfaat penelitian ini adalah sebagai berikut.

1. Bagi penulis, menambah pengetahuan lebih kepada penulis dalam menganalisis sentimen masyarakat tentang kampanye Anti-LGBT menggunakan Algoritma *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur *N-Gram*.
2. Bagi pembaca, sebagai bahan bacaan dan sumber informasi mengenai analisis sentimen masyarakat tentang kampanye anti-LGBT menggunakan algoritma *Naïve Bayes Classifier* (NBC) dengan ekstraksi fitur *N-Gram*.
3. Bagi peneliti selanjutnya, sebagai bahan referensi untuk peneliti selanjutnya.