

**ANALISIS SENTIMEN PADA KASUS LGBT MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

SKRIPSI

Diajukan untuk salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



Disusun oleh :

ERNI DEWI UTAMI

17.11.1488

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

**ANALISIS SENTIMEN PADA KASUS LGBT MENGGUNAKAN
ALGORITMA NAÏVE BAYES**

SKRIPSI

Diajukan untuk salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh
ERNI DEWI UTAMI
17.11.1488

Kepada

FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2024

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS SENTIMEN KASUS LGBT MENGGUNAKAN ALGORITMA NAÏVE
BAYES**

yang disusun dan diajukan oleh

ERNI DEWI UTAMI

17.11.1488

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal <22 Mei 2024>

Dosen Pembimbing,



Kusnawi, S.Kom., M.Eng
NIK. 190302112

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS SENTIMEN PADA KASUS LGBT MENGGUNAKAN ALGORITMA
NAÏVE BAYES**

yang disusun dan diajukan oleh

ERNI DEWI UTAMI

17.11.1488

Telah dipertahankan di depan Dewan Penguji
pada tanggal < 22 Mei 2024 >

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Heri Sismoro, M.Kom
NIK. 190302057



Agung Nugroho, M.Kom
NIK. 190302242



Kusnawi, S.Kom, M.Eng
NIK. 190302112



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal < 22 Mei 2024 >

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Erni Dewi Utami
NIM : 17.11.1488

Menyatakan bahwa Skripsi dengan judul berikut:

Tuliskan Judul Skripsi

Dosen Pembimbing : Kusnawi, S.Kom, M. Eng.

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, <22 Mei 2024>

Yang Menyatakan,



Erni Dewi Utami

HALAMAN PERSEMBAHAN

Alhamdulillah pujisyukur penulis panjatkan kepada Allah SWT, atas segala rahmat, taufiq, serta hidayah-Nya sehingga diberi kesempatan untuk dapat menyelesaikan skripsi ini sebagai salah satu syarat untuk mendapatkan gelar sarjana dengan segala kekurangan penulis. Segala syukur penulis ucapkan kepada-Mu karena telah menghadirkan mereka yang memberikan semangat dan doa disaat menjalankan proses pembuatan skripsi ini. Dengan segala kerendahan hati saya persembahkan skripsi ini kepada.

1. Kedua Orang Tua saya Bapak Endarto dan Ibu Bawon. Yang selalu mendoakan untuk kebaikan anak-anaknya, selalu memberi semangat serta motivasi supaya dapat menyelesaikan skripsi ini dengan lancar serta bermanfaat bagi semua.
2. Bapak Kusnawi, S.Kom, M.Eng. selaku dosen pembimbing saya dalam penyusunan skripsi ini. yang tidak lelah untuk selalu membimbing dan mengingatkan saya dari awal hingga akhir proses pembuatan skripsi.
3. Dosen-dosen Universitas Amikom Yogyakarta yang telah memberikan banyak ilmu baik ilmu akademik maupun ilmu non-akademik selama masa kuliah.
4. Kepada adek saya tercinta Eti Widayani dan Zulfa Qiroatun Nafisyah untuk segala doa, dukungan dan motivasinya untuk saya menyelesaikan skripsi ini.
5. Keluarga besar kelas 17-S1IF-09 yang telah menemani saya selama menempuh pendidikan di sini.
6. Keluarga besar kos Delima Raya yang telah menemani dan mendukung saya selama menyelesaikan skripsi.

KATA PENGANTAR

Bismillahirrohmanirrohim.

Puji dan syukur saya panjatkan kepada Allah SWT. Atas segala rahmat, taufiq, serta hidayah-Nya kepada saya dapat menyelesaikan skripsi yang berjudul “ANALISIS SENTIMEN PADA KASUS LGBT MENGGUNAKAN ALGORITMA NAÏVE BAYES”.

Selama proses pengerjaan skripsi ini penulis menyadari bahwa dalam proses pengerjaan ini memiliki banyak kendala, namun dengan bantuan, bimbingan dari berbagai pihak sehingga kendala tersebut bias diatasi. Oleh karena ini pada kesempatan ini penulis ingin mengucapkan banyak terimakasih kepada yang terhormat:

1. Bapak Prof. Dr. M. Suyanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Bapak Kusnawi, S.Kom, M.Eng. selaku dosen pembimbing yang telah memberikan banyak saran dan membntu membimbing dalam menyelesaikan skripsi ini.
3. Bapak Hanif Al Fatta, S,Kom, M.Kom. selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Ibu Windha Mega D, M.Kom selaku Kepala Prode Informatika Universitas Amikom Yogyakarta.
5. Dosen Penguji Bapak Heri Sismoro, M.Kom dan Bapak Agung Nugroho, M.Kom yang telah memberikan masukan kepada penelitian saya.
6. Kedua orang tua dan keluarga saya yang memberikan doa, dukungan, motivasi dan semangat.

Yogyakarta, <22 Mei 2024>

Erni Dewi Utami

DAFTAR ISI

ANALISIS SENTIMEN PADA KASUS LGBT MENGGUNAKAN ALGORITMA NAÏVE BAYES	1
HALAMAN PERSETUJUAN.....	ii
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR	x
INTISARI	xi
ABSTRACT.....	xii
BAB 1	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	2
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	3
BAB II.....	4
LANDASAN TEORI.....	4
2.1 Studi Literatur	4
2.2 Dasar Teori.....	10
2.2.1 Data Mining	10
2.2.2 Text Mining	11
2.2.3 Sentimen Analisis	11
2.2.4 Level Analisis Sentimen	12
2.2.5 Data Crawling	12
2.2.6 Preprocessing	13
2.2.7 Algoritma Naïve Bayes Classifier	14
2.2.8 Term Frequency – Inverse Document Frequency (TF-IDF).....	15
2.2.9 Confusion Matrix	16
BAB III	19

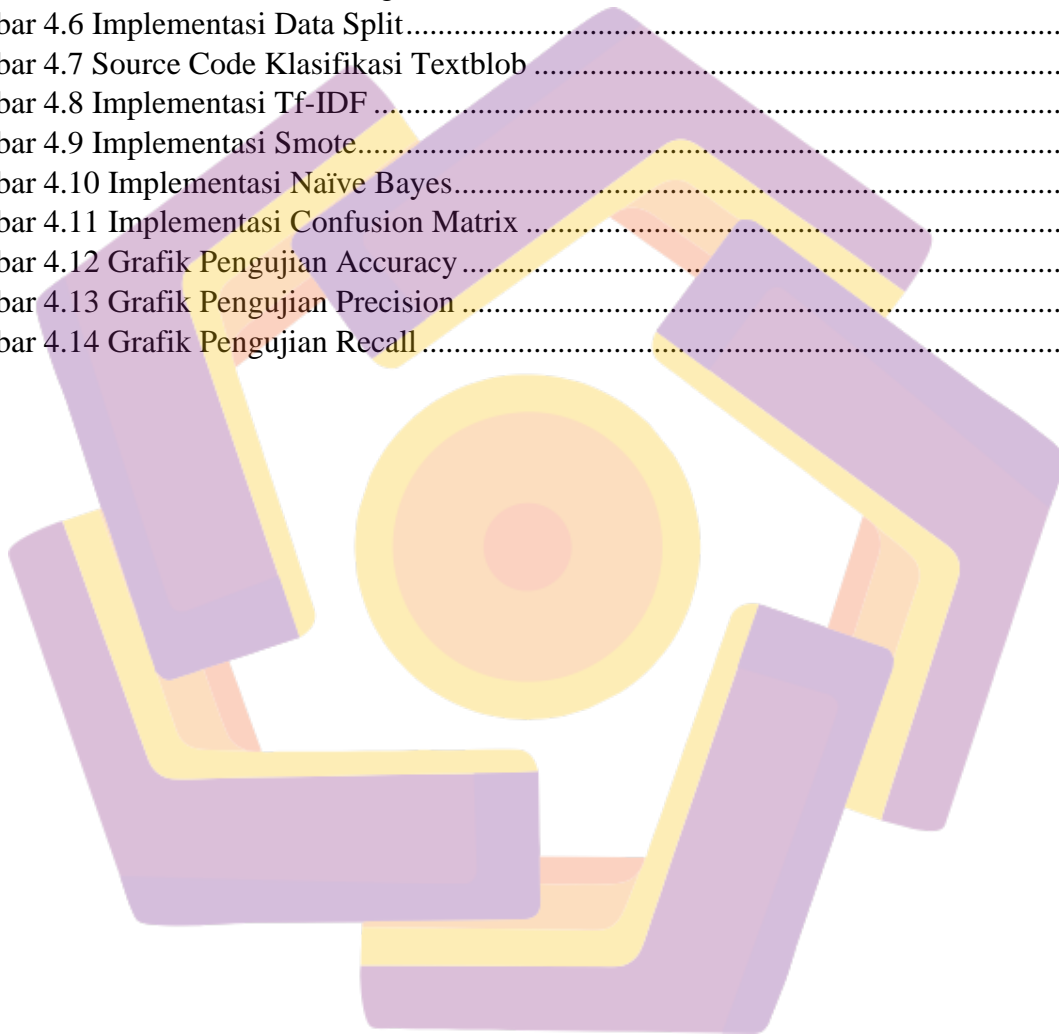
ANALISIS DAN PERANCANGAN	19
3.3 Objek Penelitian	19
3.2 Alur Penelitian.....	19
3.2.1 Crawling Data Twitter	20
3.2.2 <i>Preprocessing</i> Dataset	23
3.2.3 Data Split	28
3.2.4 Feature Extraction.....	28
3.2.5 Naïve Bayes Classifier	31
3.2.6 Confusion Matrix	41
3.3 Alat dan Bahan Penelitian	41
BAB IV	43
IMPLEMENTASI DAN PEMBAHASAN.....	43
4.1 Deskripsi Implementasi.....	43
4.2 Implementasi Pengumpulan Data.....	43
4.3 Implementasi <i>Preprocessing</i> Data	43
4.3.1 Proses <i>Preprocessing</i>	43
4.4 Implementasi Klasifikasi.....	45
4.4.1 Implementasi Data Split.....	45
4.4.2 Klasifikasi Data Menggunakan Textblob	46
4.4.3 Implementasi TF-IDF	47
4.4.4 Implementasi Smote.....	47
4.4.5 Implementasi Algoritma Naïve Bayes Classifier.....	48
4.4.6 Implementasi Confusion Matrix	48
4.6 Hasil Klasifikasi Naïve Bayes.....	49
4.7 Grafik Hasil Pengujin.....	51
BAB V	53
KESIMPULAN.....	53
5.1 Kesimpulan.....	53
5.2 Saran.....	53
REFERENSI	54

DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	6
Tabel 2.2 Rumus Confesion Matrix.....	16
Tabel 3.1 Dataset Utuh	20
Tabel 3.2 Labeling Data Tweet.....	22
Tabel 3.3 Contoh Username Removal	23
Tabel 3.4 Contoh Case Folding	24
Tabel 3.5 Punctuation Removal.....	24
Tabel 3.6 Contoh Cleaning	25
Tabel 3.7 Contoh <i>Tokenzing</i>	26
Tabel 3.8 Contoh Stemming	27
Tabel 3.9 Contoh Stopword Removal.....	27
Tabel 3.10 Term Kata	29
Tabel 3.11 Perhitungan IDF.....	29
Tabel 3.12 Perhitungan TF-IDF.....	30
Tabel 3.13 Data Tweet.....	32
Tabel 3.14 Proses Tf-Idf pada data Tweet	32
Tabel 3.15 Hasil probabilitas kata disetiap kelas	39
Tabel 3.16 Preprocessing data uji	40
Tabel 4.1 Hasil Klasifikasi Textblob	47
Tabel 4.2 Hasil Confusion Matrix Pengujian 1	49
Tabel 4.3 Hasil Confusion Matrix Pengujian 2	49
Tabel 4.4 Hasil Confusion Matrix Pengujian 3	50

DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	20
Gambar 3.2 Source Code Data Spilt	28
Gambar 4.1 Source Code Case Folding	43
Gambar 4.2 Source Code Punctuation Removal.....	44
Gambar 4.3 Source Code Remove.....	44
Gambar 4.4 Source Code Stopword Removal	45
Gambar 4.5 Source Code Stemming.....	45
Gambar 4.6 Implementasi Data Split.....	45
Gambar 4.7 Source Code Klasifikasi Textblob	46
Gambar 4.8 Implementasi Tf-IDF	47
Gambar 4.9 Implementasi Smote.....	48
Gambar 4.10 Implementasi Naïve Bayes.....	48
Gambar 4.11 Implementasi Confusion Matrix	48
Gambar 4.12 Grafik Pengujian Accuracy	51
Gambar 4.13 Grafik Pengujian Precision	51
Gambar 4.14 Grafik Pengujian Recall	52



INTISARI

LGBT adalah singkatan dari lesbian, gay, biseksual, dan transgender. Komunitas LGBT mulai berani mengumbar ke masyarakat, di Amerika Serikat pernikahan sesama jenis telah disahkan pada 27 Juni 2015. Sejauh ini komunitas LGBT masih menimbulkan pro dan kontra ditengah-tengah masyarakat. Oleh sebab itu banyak masyarakat di seluruh dunia memberikan berbagai jenis respon terhadap komunitas LGBT baik itu respon positif, netral atau negatif yang mereka sampaikan melalui social media salah satunya *Twitter*. Dengan munculnya berbagai opini tersebut, penelitian melakukan analisis sentiment untuk mengetahui kecenderungan sentimen dari masyarakat terhadap komunitas LGBT, banyaknya opini yang dituliskan dijadikan sebagai dataset yang berjumlah 3200 data. dataset tersebut akan diklasifikasikan apakah dataset tersebut beropini positif, netral atau negatif.

Adapun beberapa tahapan yang dilakukan dalam pengumpulan data atau crawling dataset secara manual melalui website kaggle, preprocessing untuk menghilangkan noise yang tidak diperlukan. Feature extraction dengan metode TF-IDF, klasifikasi dengan algoritma Naive Bayes dan dievaluasi dengan menggunakan metode Confussion Matrix.

Berdasarkan hasil penelitian yang telah dilakukan dengan algoritma Naive Bayes didapatkan akurasi sebesar 68,58%, precision score 60,1% dan recall score sebesar 53,96%.

Keywords: LGBT, Naive Bayes, Twitter.

ABSTRACT

LGBT is an abbreviation for lesbian, gay, bisexual and transgender. The LGBT community is starting to dare to speak out in society, in the United States same-sex marriage was legalized on June 27 2015. So far the LGBT community still raises pros and cons in society. For this reason, many people around the world provide various types of responses to the LGBT community, whether positive, neutral or negative, which they convey via social media, one of which is Twitter. With the emergence of these various opinions, research conducted sentiment analysis to determine the tendency of public sentiment towards LGBT community, the number of opinions written was used as a dataset totaling 3200 data. The dataset will be classified whether the dataset has a positive, neutral or negative opinion.

There are several stages carried out in collecting data or crawling datasets manually via the Kaggle website, preprocessing to remove unnecessary noise. Feature extraction using the TF-IDF method, classification using the Naive Bayes algorithm and evaluated using the Confusion Matrix method.

Based on the results of research conducted using the Naïve Bayes algorithm, an accuracy of 68.58%, a precision score of 60.1% and a recall score of 53.96% were obtained.

Keywords: *LGBT, Naïve Bayes, Twitter.*