



**SKRIPSI**

**IMPLEMENTASI METODE BERNOULLI NAÏVE BAYES  
PADA ANALISIS SENTIMEN TERHADAP LESBIAN,  
GAY, BISEKSUAL, DAN TRANSGENDER (LGBT) PADA  
MEDIA SOSIAL TWITTER**

**ANGGUN SAFIRA KUSMINDASARI**

NPM 18081010003

**DOSEN PEMBIMBING**

Wahyu Syaifullah J. S, S.Kom., M.Kom

Hendra Mulana, S.Kom, M.Kom

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**SURABAYA**

**2025**

Halaman ini sengaja dikosongkan



## **SKRIPSI**

# **IMPLEMENTASI METODE BERNOULLI NAÏVE BAYES PADA ANALISIS SENTIMEN TERHADAP LESBIAN, GAY, BISEKSUAL, DAN TRANSGENDER (LGBT) PADA MEDIA SOSIAL TWITTER**

**ANGGUN SAFIRA KUSMINDASARI**

NPM 18081010003

## **DOSEN PEMBIMBING**

Wahyu Syaifullah J. S., S.Kom., M.Kom

Hendra Mulana, S.Kom, M.Kom

**KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI INFORMATIKA**

**SURABAYA**

**2025**

Halaman ini sengaja dikosongkan

## LEMBAR PENGESAHAN

### IMPLEMENTASI METODE BERNOULLI NAIVE BAYES PADA ANALISIS SENTIMEN TERHADAP LESBIAN, GAY, BISEKSUAL, DAN TRANSGENDER (LGBT) PADA MEDIA SOSIAL TWITTER

Oleh :

ANGGUN SAFIRA KUSMINDASARI

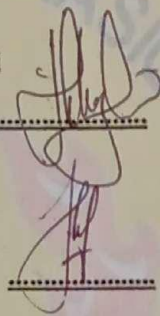
NPM. 18081010003

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 21 Januari 2025

Menyetujui

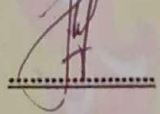
Wahyu Syaifullah J. S., S.Kom., M.Kom

NIP. 19860825 2021211 003

  
.....  
(Pembimbing I)

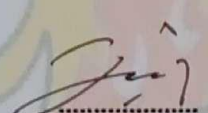
Hendra Maulana, S.Kom, M.Kom

NPT. 201198 31 223248

  
.....  
(Pembimbing II)

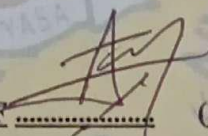
Fawwaz Ali Akbar, S.Kom, M.Kom

NIP. 19920317 2018031 002


  
.....  
(Ketua Penguji)

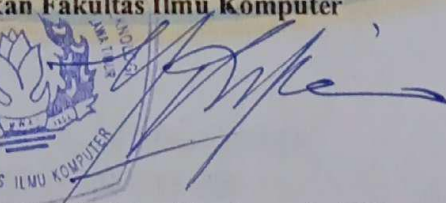
Firza Prima Aditiawan, S.Kom, M.T.I, M.C.F

NIP. 19860523 2021211 003

  
.....  
(Anggota Penguji)

Mengetahui

  
Dekan Fakultas Ilmu Komputer

  
.....  
Prof. Dr. Ir. Novirina Hendrasarie, MT.

NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

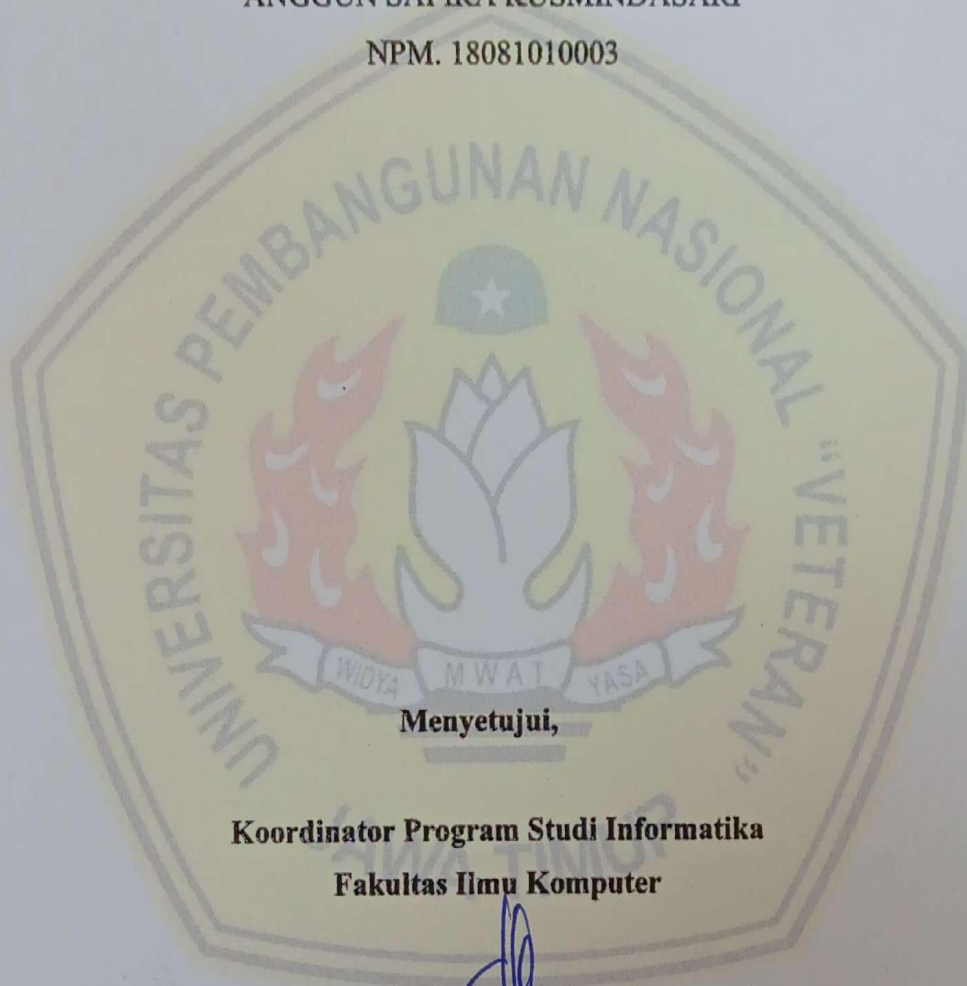
**LEMBAR PERSETUJUAN**

**IMPLEMENTASI METODE BERNOULLI NAIVE BAYES PADA ANALISIS  
SENTIMEN TERHADAP LESBIAN, GAY, BISEKSUAL, DAN  
TRANSGENDER (LGBT) PADA MEDIA SOSIAL TWITTER**

Oleh :

ANGGUN SAFIRA KUSMINDASARI

NPM. 18081010003



**Menyetujui,**

**Koordinator Program Studi Informatika  
Fakultas Ilmu Komputer**

**Fetty Tri Anggraeny, S.Kom., M.Kom.**

**NIP. 19820211 2021212 005**

Halaman ini sengaja dikosongkan



## SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini :

Nama mahasiswa / NPM : Anggun Safira Kusmindasari / 18081010003  
Program Studi : Informatika  
Dosen Pembimbing : 1. Wahyu Syaifullah J.S, S.Kom., M.Kom  
2. Hendra Mulana, S.Kom, M.Kom

Dengan ini menyatakan bahwa Skripsi dengan judul “Implementasi Metode Bernoulli Naive Bayes Pada Analisis Sentimen Terhadap Lesbian, Gay, Biseksual, dan Transgender (LGBT) Pada Media Sosial Twitter” adalah hasil karya sendiri, bersifat orisinal, dan ditulis dengan mengikuti kaidah penulisan ilmiah.

Bilamana dikemudian hari ditemukan ketidak sesuaian dengan pernyataan ini, maka saya bersedia menerima sanksi sesuai dengan ketentuan yang berlaku di Universitas Pembangunan Nasional Veteran Jawa Timur. Bukti hasil pengecekan plagiasi dokumen dapat di telusuri melalui QR Code dibawah.



Surabaya, 24 Januari 2025

Mahasiswa



(Anggun Safira Kusmindasari)

NPM. 18081010003

Halaman ini sengaja dikosongkan

## ABSTRAK

Nama Mahasiswa / NPM : Anggun Safira Kusmindasari / 18081010003  
Judul Skripsi : Implementasi Metode Bernoulli Naive Bayes Pada Analisis Sentimen Terhadap Lesbian, Gay, Biseksual, dan Transgender (LGBT) Pada Media Sosial Twitter  
Dosen Pembimbing : 1. Wahyu Syaifullah J. S, S.Kom., M.Kom  
2. Hendra Mulana, S.Kom, M.Kom

Saat ini teknologi berkembang dengan sangat pesat dan mempermudah pengguna untuk mengakses berbagai macam informasi. Salah satu media yang sering digunakan untuk mencari informasi selain google adalah twitter atau yang sekarang orang menyebut dengan X. X merupakan salah satu media sosial atau platform yang banyak digunakan untuk menyampaikan pendapat dan mencari informasi berita terbaru. Banyak hal dapat dilakukan di platform X salah satunya melakukan pembahasan topik dengan sebuah *hashtag* (#) agar topik itu menjadi trending. Salah satu pembahasan topik yang masih dibahas hingga saat ini adalah mengenai Lesbian, Gay, Biseksual dan Transgender (LGBT). Ulasan atau tweet pada pembahasan topik LGBT memuat banyak sentimen dari pengguna X. Untuk mendeteksi ulasan atau tweet yang dikirimkan pengguna pada X penting untuk dilakukan suatu teknik pengolahan data yang disebut dengan teknik analisis sentimen. Analisis sentimen menggunakan algoritma *Bernoulli Naive Bayes* yang akan di gunakan pada tweet atau ulasan pada X dalam penelitian ini. Tujuan penelitian ini untuk mengukur performa dari algoritma *Bernoulli Naive Bayes* dalam pengklasifikasian sentimen menjadi dua kelas berupa sentimen positif dan sentimen negatif. Sebanyak 2.000 data tweet pada X digunakan pada penelitian ini dibagi dengan perbandingan data latih dan data uji sebesar 70:30, 80:20, dan 90:10 . Hasil pengujian menunjukkan *Bernoulli Naive Bayes* menghasilkan nilai akurasi diatas 90%.

**Kata kunci :** Analisis sentimen, Bernoulli Naive Bayes, Twitter, Lesbian Gay Biseksual Transgender (LGBT)

Halaman ini sengaja dikosongkan

## ABSTRACT

Student Name / NPM : Anggun Safira Kusmindasari / 18081010003

Thesis Title : Implementation of Bernoulli Naive Bayes Method on Sentiment Analysis Towards Lesbian, Gay, Bisexual, and Transgender (LGBT) on Social Media Twitter

Advisors : 1. Wahyu Syaifullah J. S, S.Kom., M.Kom  
2. Hendra Mulana, S.Kom, M.Kom

Currently, technology is developing very rapidly and making it easier for users to access various kinds of information. One of the media that is often used to search for information other than Google is Twitter or what people now call X. X is one of the social media or platforms that is widely used to express opinions and find the latest news information. Many things can be done on the X platform, one of which is discussing topics with a hashtag (#) so that the topic becomes trending. One of the topics discussed that is still being discussed today is about Lesbian, Gay, Bisexual and Transgender (LGBT). Reviews or tweets on the discussion of the LGBT topic contain a lot of sentiment from X users. To detect reviews or tweets sent by users on X, it is important to carry out a data processing technique called sentiment analysis. Sentiment analysis uses the Bernoulli Naive Bayes algorithm which will be used on tweets or reviews on X in this study. The purpose of this study is to measure the performance of the Bernoulli Naive Bayes algorithm in classifying sentiment into two classes, namely positive sentiment and negative sentiment. A total of 2,000 tweet data on X were used in this study divided by a ratio of training data and test data of 70:30, 80:20, and 90:10. The test results showed that Bernoulli Naive Bayes produced an accuracy value above 90%.

**Keywords:** Sentiment analysis, Bernoulli Naive Bayes, Twitter, Lesbian Gay Bisexual Transgender (LGBT).

Halaman ini sengaja dikosongkan

## KATA PENGANTAR

Puji syukur kepada Allah SWT atas segala berkat rahmat dan petunjuk-Nya sehingga penulis dapat menyelesaikan penelitian dan menyusun Laporan Akhir Mata Kuliah Skripsi yang berjudul **”Implementasi Metode Bernoulli Naive Bayes Pada Analisis Sentimen Terhadap Lesbian, Gay, Biseksual dan Transgender (LGBT) Pada Media Sosial Twitter”**. Skripsi ini merupakan bagian dari persyaratan untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur.

Penulis mengucapkan terima kasih kepada Bapak Wahyu Syaifullah Jauharis Saputra, S.Kom., M.Kom selaku Dosen Pembimbing Pertama dan Bapak Hendra Maulana, S.Kom, M.Kom selaku Dosen Pembimbing Kedua yang telah meluangkan waktu untuk memberikan bimbingan, motivasi dan nasehat kepada penulis. Penulis juga banyak menerima bantuan secara moril, spiritual dan materi kepada berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada :

1. Prof. Dr. Ir. Akhmad Fauzi, MMT selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Tri Anggraeny, S.Kom., M. Kom. selaku Koordinator Program Studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Staff Dosen dan Tenaga Kependidikan program studi Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur yang telah memberikan ilmu dan pengalaman selama masa perkuliahan.
5. Orang tua dan adik tercinta penulis yang dengan penuh kasih sayang mendukung dan memotivasi penulis dalam penyusunan skripsi/tugas akhir. Terimakasih atas kesabaran dan cinta yang luar biasa yang telah diberikan kepada penulis.
6. Teman-teman tersayang penulis Binar Indira Mahendra, Almh. Aktivasi Inas Almira, Fifi Nurainina, Ayunda Noviala Dwijayati, Winarti, dan Fadila

Choirinnisa, Nuril Madania telah memberikan semangat motivasi dan perhatian kepada penulis,

7. Kepada salah satu boygrup yang lagu-lagunya selalu menemani penulis untuk menyusun skripsi/tugas akhir yaitu SEVENTEEN.

Penulis menyadari bahwa skripsi ini belum mencapai kata sempurna dan masih ada kekurangan dalam aspek penulisan maupun pembahasan. Maka dari itu penulis sangat mengharapkan adanya masukan, kritik dan saran yang bersifat membangun dan dapat menyempurnakan skripsi ini.

Penulis mengucapkan banyak terima kasih yang sebesar-besarnya kepada semua pihak yang turut membantu dalam penyusunan laporan skripsi ini hingga dapat diselesaikan dengan baik. Penulis berharap agar skripsi ini memberikan manfaat dan menambahkan informasi bagi semua pihak yang membaca. Penulis juga berharap semoga skripsi ini dapat menjadi pelajaran bagi penulis ataupun pembaca dikemudian hari.

Surabaya, Januari 2025

Penulis



## DAFTAR ISI

<b>LEMBAR JUDUL SKRIPSI</b> .....	<b>i</b>
<b>LEMBAR PENGESAHAN</b> .....	<b>ii</b>
<b>LEMBAR PERSETUJUAN</b> .....	<b>iii</b>
<b>SURAT PERNYATAAN ORISINALITAS</b> .....	<b>iv</b>
<b>ABSTRAK</b> .....	<b>v</b>
<b>ABSTRACT</b> .....	<b>vi</b>
<b>KATA PENGANTAR</b> .....	<b>vii</b>
<b>DAFTAR ISI</b> .....	<b>ix</b>
<b>DAFTAR GAMBAR</b> .....	<b>xii</b>
<b>DAFTAR TABEL</b> .....	<b>xiv</b>
<b>LAMPIRAN</b> .....	<b>xv</b>
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian .....	4
1.5. Manfaat Penelitian .....	4
<b>BAB II TINJAUAN PUSTAKA</b> .....	<b>5</b>
2.1. <i>Text Processing</i> .....	5
2.2.1. <i>Case Folding</i> .....	6
2.2.2. <i>Cleaning</i> .....	6
2.2.3. <i>Normalization</i> .....	6
2.2.4. <i>Tokenizing</i> .....	7
2.2.5. <i>Stopword removal</i> .....	7
2.2.6. <i>Stemming</i> .....	7
2.2. Analisis Sentimen .....	8
2.3. TF-IDF .....	8
2.4. Naïve Bayes .....	9
2.4.1 Kelebihan Naive Bayes .....	10
2.4.2 Kekurangan Naive Bayes .....	10
2.5. <i>Bernoulli Naïve Bayes</i> .....	11
2.5.1 Kelebihan Bernoulli Naïve Bayes .....	12
2.5.2 Kekurangan Bernoulli Naïve Bayes .....	12

2.6.	<i>Confusion Matix</i> .....	12
2.7.	X (Twitter) .....	14
2.8.	<i>RapidMiner</i> .....	15
2.9.	Kaggle.....	16
2.10.	<i>Synthetic data</i> .....	16
2.11.	Python .....	17
2.12.	LGBT ( <i>Lesbian Gay Bisexual Transgender</i> ).....	17
2.13.	Penelitian Terdahulu .....	19
<b>BAB III METODOLOGI PENELITIAN .....</b>		<b>23</b>
3.1.	Identifikasi Masalah.....	23
3.2.	Studi Literatur .....	24
3.3.	Pengumpulan Data dan Pelabelan Data .....	24
3.4.	Analisis Proses .....	25
3.4.1	<i>Preprocessing Data</i> .....	25
3.4.2	Pembobotan TF-IDF .....	29
3.4.3	Bernoulli Naïve Bayes Classifier.....	33
3.5.	Perancangan Sistem .....	35
3.6.	Evaluasi Model .....	35
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>37</b>
4.1.	Kebutuhan Sistem .....	37
4.1.1	Spesifikasi Perangkat Keras .....	37
4.1.2	Spesifikasi Perangkat Lunak.....	37
4.2.	Pengumpulan Data .....	37
4.3.	Pelabelan Data .....	38
4.4.	Text Processing.....	39
4.4.1	Casefolding .....	42
4.4.2	Cleaning .....	43
4.4.3	Normalisasi .....	44
4.4.4	Tokenizing.....	46
4.4.5	Stopword Removal.....	46
4.4.6	Stemming .....	48
4.5.	Pembobotan TF IDF .....	49
4.6.	Proses Implementasi <i>Bernoulli Naïve Bayes</i> .....	49
4.7.	Evaluasi Model .....	52
4.8.	WordCloud.....	58
<b>BAB V PENUTUP.....</b>		<b>61</b>

5.1	Kesimpulan .....	61
5.2	Saran .....	61
<b>DAFTAR PUSTAKA .....</b>		<b>63</b>
<b>LAMPIRAN.....</b>		<b>67</b>

Halaman ini sengaja dikosongkan

## DAFTAR GAMBAR

Gambar 2.1 Proses Text Processing.....	6
Gambar 3.1 Alur Penelitian.....	23
Gambar 3.2 Analisis Proses .....	25
Gambar 4.1 Proses Text Crawling .....	37
Gambar 4.2 Stigma Negaitf Untuk Pelabelan .....	38
Gambar 4.3 Data Kaggle Untuk Acuan Pelabelan.....	39
Gambar 4.4 Hasil Pelabelan Data .....	39
Gambar 4.5 Kode Persiapan Pemanggilan Data .....	39
Gambar 4.6 Implementasi Dataframe .....	40
Gambar 4.7 Kode Sintetis Data.....	41
Gambar 4.8 Hasil Sintetis Data.....	41
Gambar 4.9 Pemanggilan Atribut .....	42
Gambar 4.10 Hasil Pemanggilan Atribut.....	42
Gambar 4.11 Kode Melihat Jumlah Sentimen .....	42
Gambar 4.12 Jumlah Sentimen .....	42
Gambar 4.13 Kode Proses Case folding .....	43
Gambar 4.14 Hasil Case folding .....	43
Gambar 4.15 Kode Cleaning.....	43
Gambar 4.16 Hasil Cleaning .....	44
Gambar 4.17 Potongan Kamus Normalisasi .....	44
Gambar 4.18 Kode Normalisasi.....	45
Gambar 4.19 Hasil Normalisasi .....	45
Gambar 4.20 Kode Tokenizing .....	46
Gambar 4.21 Hasil Tokenizing .....	46
Gambar 4.22 Kamus Stopword Removal .....	47
Gambar 4.23 Kode Stopword removal.....	47
Gambar 4.24 Hasil Stopword removal.....	47
Gambar 4.25 Kode Stemming.....	48
Gambar 4.26 Hasil Stemming .....	48
Gambar 4.27 Kode Pembobotan TF IDF .....	49
Gambar 4.28 Hasil Pembobotan TF IDF .....	49
Gambar 4.29 Potongan Kode Klasifikasi.....	50
Gambar 4.30 Hasil Prediksi Pertama .....	51

Gambar 4.31 Hasil Prediksi Kedua.....	51
Gambar 4.32 Hasil Prediksi Ketiga.....	52
Gambar 4.33 Kode Confusion Matrix.....	52
Gambar 4.34 Hasil Confusion Matrix Skenario Pertama.....	53
Gambar 4.35 Classification Report Bernoulli Naïve Bayes Skenario 1.....	53
Gambar 4.36 Hasil Confusion Matrix Skenario Kedua .....	54
Gambar 4.37 Classification Report Bernoulli Naïve Bayes Skenario 2 .....	55
Gambar 4.38 Hasil Confusion Matrix Skenario Ketiga.....	56
Gambar 4.39 Classification Report Bernoulli Naïve Bayes Skenario 3 .....	56
Gambar 4.40 Perbandingan Tingkat Akurasi.....	58
Gambar 4.40 Word cloud negatif.....	58
Gambar 4.41 Word cloud positif.....	59

## DAFTAR TABEL

Tabel 2.1 Confusion Matrix .....	13
Tabel 3.1 Atribut dan Pelabelan.....	24
Tabel 3.2 Case folding .....	26
Tabel 3.3 Cleaning .....	26
Tabel 3.4 Normalisasi .....	27
Tabel 3.5 Tokenizing .....	27
Tabel 3.6 Stopword removal .....	28
Tabel 3.7 Stemming .....	28
Tabel 3.8 Contoh Perhitungan TF IDF .....	29
Tabel 3.9 Frekuensi Kata (DF).....	29
Tabel 3.10 TF IDF .....	31
Tabel 3.11 Hasil TF IDF .....	32
Tabel 3.12 Probabilitas Fitur.....	34
Tabel 4.1 Hasil Akurasi.....	57

Halaman ini sengaja dikosongkan



## LAMPIRAN

Lampiran Data Hasil Preprocessing.....	67
--	----

Halaman ini sengaja dikosongkan