

SKRIPSI

**KLASIFIKASI SENTIMEN NETIZEN X MENGGUNAKAN
METODE *NAÏVE BAYES* DAN *LINEAR DISCRIMINANT
ANALYSIS (LDA)***

(STUDI KASUS: SENTIMEN NETIZEN TERKAIT AKSI DEMONSTRASI
PADA TANGGAL 1 AGUSTUS - 30 SEPTEMBER 2025)



BUNGA WANDA NURHALISA

NIM. 22106010027

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA

2026

**KLASIFIKASI SENTIMEN NETIZEN X MENGGUNAKAN
METODE *NAÏVE BAYES* DAN *LINEAR DISCRIMINANT
ANALYSIS (LDA)***

(STUDI KASUS: SENTIMEN NETIZEN TERKAIT AKSI DEMONSTRASI
PADA TANGGAL 1 AGUSTUS - 30 SEPTEMBER 2025)

Skripsi

Untuk memenuhi sebagian persyaratan

Mencapai derajat Sarjana S-1

Program Studi Matematika



Diajukan oleh

BUNGA WANDA NURHALISA

22106010027

Kepada

PROGRAM STUDI MATEMATIKA

FAKULTAS SAINS DAN TEKNOLOGI

UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA

YOGYAKARTA



SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal : Persetujuan Skripsi / Tugas Akhir

Lamp :

Kepada

Yth. Dekan Fakultas Sains dan Teknologi

UIN Sunan Kalijaga Yogyakarta

di Yogyakarta

Assalamu 'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

Nama : Bunga Wanda Nurhalisa

NIM : 22106010027

Judul Skripsi : Klasifikasi Sentimen Netizen X Menggunakan Metode Naive Bayes dan Linear Discriminant Analysis (LDA)
(Studi Kasus: Sentimen Netizen Terkait Aksi Demonstrasi pada Tanggal 1 Agustus – 30 September 2025)

sudah dapat diajukan kembali kepada Program Studi Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Program Studi Matematika.

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqasyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu 'alaikum wr. wb.

Yogyakarta, 03 Juni 2026

Pembimbing I

Muchammad Abrori, S.Si., M.Kom.

NIP. 19720423 199903 1 003

Pembimbing II

Muhamad Rashif Hilmi, S.Si., M.Sc.

NIP. 19920309 202012 1 001



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-1394/Un.02/DST/PP.00.9/06/2026

Tugas Akhir dengan judul : KLASIFIKASI SENTIMEN NETIZEN X MENGGUNAKAN METODE NAIVE BAYES DAN LINEAR DISCRIMINANT ANALYSIS (LDA) (STUDI KASUS: SENTIMEN NETIZEN TERKAIT AKSI DEMONSTRASI PADA TANGGAL 1 AGUSTUS - 30 SEPTEMBER 2025)

yang dipersiapkan dan disusun oleh:

Nama : BUNGA WANDA NURHALISA
Nomor Induk Mahasiswa : 22106010027
Telah diujikan pada : Rabu, 10 Juni 2026
Nilai ujian Tugas Akhir : A

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR



Ketua Sidang

Muchammad Abrori, S.Si., M.Kom
SIGNED

Valid ID: 6a43293d684cc



Penguji I

Muhamad Rashif Hilmi, S.Si., M.Sc.
SIGNED

Valid ID: 6a4333a4cd063



Penguji II

Dr. Epha Diana Supandi, S.Si., M.Sc.
SIGNED

Valid ID: 6a3c89bb9f628



Yogyakarta, 10 Juni 2026

UIN Sunan Kalijaga
Dekan Fakultas Sains dan Teknologi

Prof. Dr. Dra. Hj. Khurul Wardati, M.Si.
SIGNED

Valid ID: 6a4383996ada1

SURAT PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

Nama : Bunga Wanda Nurhalisa
NIM : 22106010027
Program Studi : Matematika
Fakultas : Sains dan Teknologi

Dengan ini menyatakan bahwa isi skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu Perguruan Tinggi dan sesungguhnya skripsi ini merupakan hasil pekerjaan penulis sendiri sepanjang pengetahuan penulis, bukan duplikasi atau saduran dari karya orang lain kecuali bagian tertentu yang penulis ambil sebagai bahan acuan. Apabila terbukti pernyataan ini tidak benar, sepenuhnya menjadi tanggung jawab penulis.

Yogyakarta, 04 Juni 2026



Bunga Wanda Nurhalisa

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN PERSEMBAHAN

*Dengan penuh rasa syukur dan kerendahan hati,
skripsi ini saya persembahkan kepada
kedua orang tua tercinta,
kakak tersayang,
keluarga besar,
sahabat,
serta teman-teman seperjuangan.
Terima kasih atas doa dan dukungannya sehingga karya ini dapat terwujud.*



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

HALAMAN MOTTO

“Sebaik-baik manusia adalah yang paling bermanfaat bagi manusia”

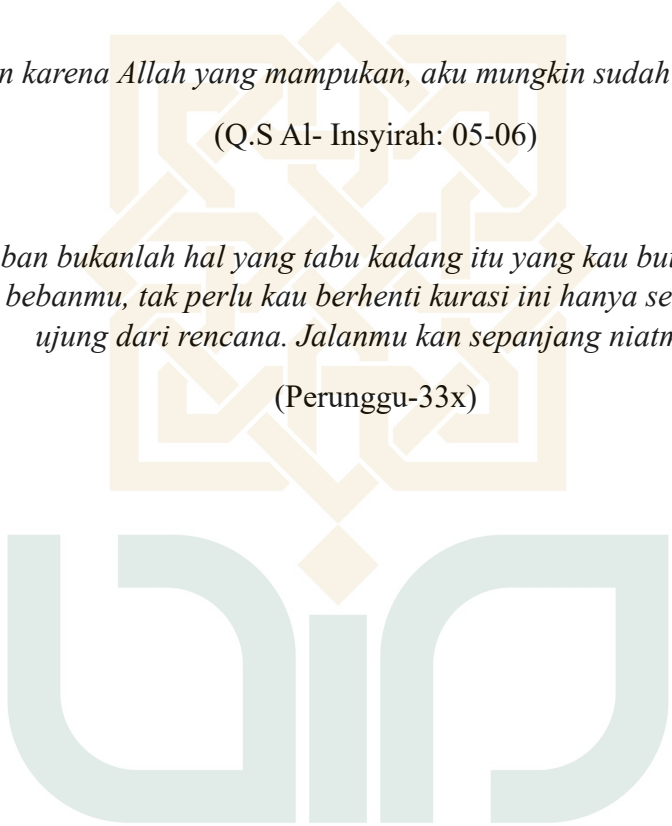
(HR. Ahmad)

“Jika bukan karena Allah yang mampukan, aku mungkin sudah lama menyerah”

(Q.S Al- Insyirah: 05-06)

*“Melamban bukanlah hal yang tabu kadang itu yang kau butuh, bersandar
hibahkan bebanmu, tak perlu kau berhenti kurasi ini hanya sementara, bukan
ujung dari rencana. Jalanmu kan sepanjang niatmu”*

(Perunggu-33x)



STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

KATA PENGANTAR

Alhamdulillah, puji syukur kepada Allah SWT yang telah melimpahkan Rahmat dan karunia-Nya berupa keimanan, kekuatan, kesabaran, dan kelancaran sehingga penulis dapat menyelesaikan skripsi dengan judul “Klasifikasi Sentimen Netizen X Menggunakan Metode Naïve Bayes dan Linear Discriminant Analysis (LDA)”. Sholawat serta salam senantiasa tercurahkan kepada Nabi Muhammad SAW yang kita nantikan syafaatnya di yaumul qiyamah nanti.

Skripsi ini ditulis sebagai syarat dalam menyelesaikan studi S-1 Matematika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta. Penulis menyadari bahwa selama penyusunan skripsi ini telah banyak mendapatkan doa, bantuan, bimbingan serta dukungan dari berbagai pihak. Untuk itu pada kesempatan ini, penulis bermaksud ingin menyampaikan ucapan terima kasih kepada:

1. Almamater UIN Sunan Kalijaga Yogyakarta yang telah memberikan ruang dan kesempatan bagi penulis untuk terus belajar dan berkembang.
2. Bapak Prof. Noorhaidi Hasan, S.Ag., M.A., M.Phil., Ph.D. selaku Rektor UIN Sunan Kalijaga Yogyakarta.
3. Ibu Prof. Dr. Khurul Wardati, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga serta selaku Dosen Penasihat Akademik yang telah memberikan arahan, masukan dan nasihat kepada penulis selama di bangku perkuliahan.
4. Ibu Dr. Epha Diana Supandi, S.Si., M.Sc., selaku Ketua Program Studi Matematika UIN Sunan Kalijaga.
5. Bapak Muchammad Abrori, S.Si., M.Kom., selaku Dosen Pembimbing I yang selalu membimbing, memberi masukan dan arahan kepada penulis dalam menyusun skripsi ini.
6. Bapak Muhamad Rashif Hilmi, S.Si., M.Sc., selaku Dosen Pembimbing II, atas kesabaran dalam membimbing, mengarahkan, serta berkenan berdiskusi dan memberikan masukan kepada penulis selama proses penyusunan skripsi ini.

7. Seluruh dosen matematika dan staff Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta yang selalu memberikan ilmu dan pelayanan yang baik selama perkuliahan.
8. Kedua orang tua tercinta, Bapak Wandoyo Sutrisno dan Ibu Supini yang selalu memberikan kasih sayang, doa, dan dukungannya kepada penulis hingga saat ini.
9. Kakak tersayang penulis, Leiliana, serta kakak ipar Wahyu yang selalu memberikan semangat, dukungan, dan motivasi kepada penulis selama proses penyusunan skripsi ini.
10. Teman-teman Grup Komangs yaitu, Atina, Dzakiyya, Amanda, Tintan, Viga, dan Icha serta teman skripsi bareng yaitu Nafisa dan Laila yang telah memberikan dukungan, kebersamaan, serta semangat selama proses perkuliahan hingga penyusunan skripsi ini. Diskusi, bantuan, dan motivasi yang diberikan turut membantu penulis dalam menghadapi berbagai kendala akademik maupun nonakademik.
11. Teman seperjuangan selama di perantauan, yaitu Hasa, Mutia, Dina, Zulfa, Dea, Liya, Mba Faiza, dan Mba Nunu, yang selalu mendengarkan keluh kesah penulis serta senantiasa memberikan semangat dan dukungan selama menempuh perkuliahan hingga penulis dapat menyelesaikan skripsi ini.
12. Teman-teman Grup Bambang's yaitu Nabila, Beatrice, Diva, dan Ilah yang walaupun sudah sibuk masing-masing tetapi tetap memberikan doa dan *support* kepada penulis untuk mengejar gelar S.Mat ini.
13. Kedua teman penulis, yaitu Alghifario dan Puan, yang hingga saat ini selalu mendengarkan cerita, menemani penulis saat berada di titik lelah, serta memberikan semangat dan dukungan selama proses penyusunan skripsi ini.
14. Teman-teman seperjuangan KKN 117, yaitu Maya, Nadia, Suse, Ratri, Atin, Zumar, Rafi, dan Syauqi. Terima kasih atas pengalaman suka maupun duka bersama mulai dari awal memasuki tahap KKN hingga saat ini yang masih saling memberikan *support* satu sama lain di bangku perkuliahan.
15. Teman-teman Matematika 2022, terutama pada konsentrasi Statistika yang sudah kebersamai selama masa perkuliahan dan saling menyemangati satu sama lain.

16. Diri sendiri. Ya, Bunga Wanda Nurhalisa, anak perempuan yang menjadi harapan orang tuanya. Terimakasih telah bertahan, berjuang, dan tetap melangkah hingga sampai di titik ini. Terima kasih karena tidak menyerah di tengah keraguan, kelelahan, dan berbagai rintangan yang datang silih berganti. Terima kasih telah percaya pada proses, berani mencoba, belajar, dan menyelesaikan apa yang telah dimulai. Semoga setiap langkah yang telah dilewati menjadi awal dari perjalanan yang lebih baik ke depannya.
17. Semua pihak yang tidak dapat disebutkan satu per satu yang telah membantu penulis dalam menyelesaikan skripsi ini, semoga Allah SWT membalas kebajikannya.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan, sehingga penulis mengharapkan kritik dan saran dari semua pihak guna kebaikan dari skripsi ini.

Yogyakarta, 20 Mei 2026

Penulis

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERSETUJUAN.....	Error! Bookmark not defined.
HALAMAN PENGESAHAN.....	Error! Bookmark not defined.
HALAMAN PERNYATAAN	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO	vii
KATA PENGANTAR.....	viii
DAFTAR ISI	xi
DAFTAR TABEL.....	xiv
DAFTAR GAMBAR	xv
DAFTAR LAMBANG.....	xvi
INTISARI.....	1
ABSTRACT	2
BAB I PENDAHULUAN.....	3
1.1 Latar Belakang	3
1.2 Rumusan Masalah	6
1.3 Tujuan Penelitian.....	7
1.4 Manfaat Penelitian	7
1.4.1 Manfaat Teoritis	7
1.4.2 Manfaat Praktis	7
1.5 Batasan Masalah.....	8
1.6 Tinjauan Pustaka	8
1.7 Sistematika Penulisan	14
BAB II LANDASAN TEORI	16
2.1 Probabilitas.....	16
2.1.1 Probabilitas bersyarat.....	18
2.1.2 Distribusi Probabilitas.....	20
2.2 Teorema Bayes	22
2.3 <i>Prior Probability</i>	23
2.4 <i>Likelihood</i>	24
2.5 <i>Evidence</i>	24

2.6 Posterior Probability.....	25
2.7 Matriks	25
2.7.1 Jenis-jenis Matriks	26
2.7.2 Operasi Matriks.....	29
2.8 Vektor	31
2.8.1 Penjumlahan dan Pengurangan Vektor	31
2.8.2 Perkalian Skalar dengan Vektor	32
2.8.3 Panjang Vektor	32
2.9 Mean Vektor	34
2.10 Kovarians Matriks	34
2.11 Eigenvalue dan Eigenvektor	35
2.12 Data Mining	36
2.13 Machine Learning	38
2.13.1 Definisi.....	38
2.13.2 Algoritma Machine Learning.....	38
2.13.3 Langkah-Langkah Machine Learning	40
2.14 Klasifikasi	44
2.15 Data Training dan Data Testing.....	45
2.16 Leksikon Sentimen.....	46
2.17 TF-IDF	47
2.18 Aksi Demonstrasi	49
BAB III METODE PENELITIAN.....	52
3.1 Jenis Penelitian.....	52
3.2 Jenis dan Sumber Data	52
3.3 Variabel Penelitian	52
3.3.1 Variabel Independen (X)	52
3.3.2 Variabel Dependen (Y).....	53
3.4 Metode Pengolahan Data	53
3.5 Alat Pengolahan Data	56
3.6 Alur Penelitian.....	57
BAB IV PEMBAHASAN.....	58
4.1 Naive Bayes	58
4.2 Teorema Naive Bayes	59
4.3 Jenis-jenis Naive Bayes	62
4.3.1 Gaussian Naive Bayes	62

4.3.2 Multinomial Naïve Bayes	63
4.3.3 Bernoulli Naïve Bayes	64
4.4 <i>Linear Discriminant Analysis</i> (LDA).....	65
4.5 Matriks Evaluasi.....	72
BAB V STUDI KASUS	75
5.1 <i>Scraping Data</i>	75
5.2 Penyimpanan Data	77
5.3 <i>Input Data</i>	77
5.4 <i>Preprocessing Data</i>	77
5.4.1 <i>Cleaning text</i>	77
5.4.2 <i>Case Folding</i>	79
5.4.3 <i>Spelling Normalization</i>	80
5.4.4 Tokenisasi	81
5.4.5 <i>Stopword Removal</i>	83
5.4.6 <i>Stemming</i>	84
5.5 Pelabelan Data.....	85
5.6 Split Data.....	87
5.7 TF-IDF	88
5.8 Pembentukan Model.....	91
5. 9 Evaluasi Model.....	93
5.10 Model Terbaik	96
5.11 Implementasi Model Terbaik.....	98
BAB VI PENUTUP	100
6.1 Kesimpulan	100
6.2 Saran	101
DAFTAR PUSTAKA	102
LAMPIRAN.....	105

DAFTAR TABEL

Tabel 1.1 Tinjauan Pustaka	11
Tabel 2.1 Contoh pemberian skor pada sentiment	46
Tabel 4.1 Contoh data latih klasifikasi siswa	68
Tabel 4.2 <i>Confusion matrix</i>	73
Tabel 5.1 Contoh hasil scrapping data X	75
Tabel 5.2 Contoh hasil cleaning text	78
Tabel 5.3 Contoh hasil <i>case folding</i>	79
Tabel 5.4 Contoh hasil <i>Spelling normalization</i>	80
Tabel 5.5 Contoh hasil Tokenisasi	81
Tabel 5.6 Contoh hasil <i>Stopword Removal</i>	83
Tabel 5.7 Contoh hasil <i>stemming</i>	84
Tabel 5.8 Hasil pelabelan data	85
Tabel 5.9 <i>Confusion Matrix Naive Bayes</i>	93
Tabel 5.10 Hasil Evaluasi <i>Naive Bayes</i>	94
Tabel 5.11 <i>Confusion Matrix LDA</i>	95
Tabel 5.12 Hasil Evaluasi LDA	96
Tabel 5.13 Ringkasan hasil evaluasi model	96
Tabel 5.14 Hasil prediksi sentiment pada data uji	98

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Probabilitas	17
Gambar 2.2 Ilustrasi probabilitas bersyarat	18
Gambar 2.3 Vektor 2D	33
Gambar 2.4 Vektor 3D	33
Gambar 2.5 Berita Demonstrasi pada IDN Times	50
Gambar 2.6 Berita Demonstrasi Indonesia pada AP News	50
Gambar 3.1 Flowchart Penelitian.....	57
Gambar 4.1 Ilustrasi <i>Naive Bayes</i>	58
Gambar 5.1 Jumlah data masing-masing kelas sentiment	86
Gambar 5.2 Hasil pembagian data	87
Gambar 5.3 Frekuensi kata pada data <i>training</i>	88
Gambar 5.4 Nilai term TF-IDF berdasarkan sentiment	90



DAFTAR LAMBANG

$P(A)$	= Probabilitas A
$n(A)$	= Jumlah kejadian yang menguntungkan A
$n(S)$	= Jumlah total dari seluruh kejadian dalam ruang sampel
$P(A B)$	= Probabilitas kejadian A dengan syarat B sudah terjadi
$P(A \cap B)$	= Probabilitas A dan B secara bersamaan
$\emptyset$	= Himpunan kosong
$f(x)$	= Nilai fungsi f pada titik x
σ	= Standar deviasi
e	= eksponensial
μ	= Rata-rata (mean)
A^c	= A komplemen
$L(\theta x)$	= Fungsi <i>likelihood</i>
a_{ij}	= Elemen matriks pada posisi (i, j)
I_n	= Matriks identitas berordo n
A^T	= Transpose dari matriks A
$\ w\ $	= Panjang vektor w
N	= Jumlah total sampel
K	= Jumlah kelas
μ^k	= Asumsi mean vektor kelas k
λ	= <i>eigenvalue</i>
ω_k	= Kelas ke-k
S	= Kovarians gabungan

STATE ISLAMIC UNIVERSITY
SUNAN KALIJAGA
YOGYAKARTA

INTISARI

Klasifikasi Sentimen Netizen X Menggunakan Metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA)

(Studi Kasus: Sentimen Netizen X terkait aksi demonstrasi pada tanggal 1 Agustus – 30 September 2025)

Oleh

Bunga Wanda Nurhalisa

22106010027

Demonstrasi merupakan salah satu bentuk partisipasi politik masyarakat yang tidak hanya berlangsung di ruang fisik, tetapi juga memicu pembentukan opini publik di media sosial. Opini netizen terhadap aksi demonstrasi umumnya diekspresikan melalui komentar atau unggahan di media sosial X yang mengandung berbagai sentimen. Penelitian ini bertujuan untuk membandingkan performa metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam klasifikasi sentimen terhadap aksi demonstrasi DPR RI pada media sosial X. Dataset penelitian diperoleh melalui teknik *scraping* data pada media sosial X dengan rentang waktu pengambilan data tanggal 1 Agustus hingga 30 September 2025 dan menghasilkan sebanyak 8.702 data. Data sentimen tersebut diklasifikasikan ke dalam tiga kelas kategori, yaitu positif, negatif, dan netral. Tahapan penelitian meliputi *preprocessing*, *pembobotan TF-IDF*, serta proses klasifikasi menggunakan metode *Naïve Bayes* dan LDA. Evaluasi model dilakukan menggunakan *confusion matrix* dengan metrik *accuracy*, *precision*, dan *F1-score*. Hasil penelitian menunjukkan bahwa metode LDA memperoleh performa yang lebih baik dibandingkan *Naïve Bayes* dengan nilai akurasi sebesar 79,21%, sedangkan *Naïve Bayes* memperoleh akurasi sebesar 74,01%. Temuan ini menunjukkan bahwa metode LDA lebih efektif dalam mengklasifikasikan sentimen pada data teks media sosial yang bersifat tidak terstruktur dan berdimensi tinggi.

Kata kunci: Klasifikasi Sentimen, *Naïve Bayes*, *Linear Discriminant Analysis* (LDA), aksi demonstrasi

ABSTRACT

Sentiment Classification of Netizen on X Using Naïve Bayes and
Linear Discriminant Analysis (LDA)
(Case Study: Netizen's Sentimen Toward Demonstration Activities
from August 1 to September 30, 2025)

By

Bunga Wanda Nurhalisa

22106010027

Demonstrations are a form of public political participation that not only take place in physical spaces but also trigger the formation of public opinion on social media. Netizen's opinions regarding demonstration movements are generally expressed through comments or posts on the social media platform X, containing various sentiments. This study aims to compare the performance of the Naïve Bayes and Linear Discriminant Analysis (LDA) methods in sentiment classification related to the DPR RI demonstration movement on social media X. The dataset was obtained through a data scraping technique on social media X within the period of August 1 to September 30, 2025, resulting in 8.702 data entries. The sentiment data were classified into three categories: positive, negative, and neutral. The research stages included preprocessing, TF-IDF weighting, and sentiment classification using the Naïve Bayes and LDA methods. Model evaluation was conducted using a confusion matrix with accuracy, precision, and F1-score metrics. The results showed that the LDA method achieved better performance than Naïve Bayes, with an accuracy of 79,21%, while Naïve Bayes obtained an accuracy of 74,01%. These findings indicate that the LDA method is more effective in classifying sentiment in unstructured and high-dimensional social media text data.

Keywords: Sentiment Classification, Naïve Bayes, Linear Discriminant Analysis (LDA), Demonstration Movement

BAB I

PENDAHULUAN

1.1 Latar Belakang

Demonstrasi merupakan salah satu bentuk partisipasi politik masyarakat dalam sistem demokrasi untuk menyampaikan aspirasi, kritik, maupun tuntutan terhadap kebijakan pemerintah dan lembaga legislatif. Dalam pelaksanaannya, demonstrasi tidak hanya menjadi sarana penyampaian pendapat di ruang fisik melalui aksi turun ke jalan dan orasi, tetapi juga berkembang sebagai fenomena sosial yang memicu pembentukan opini publik secara luas.

Salah satu peristiwa yang menarik perhatian publik adalah demonstrasi yang terjadi pada tanggal 28-29 Agustus 2025 di Jakarta. Aksi tersebut merupakan kelanjutan dari unjuk rasa yang sebelumnya dilaksanakan pada tanggal 25 Agustus 2025 oleh kelompok serikat buruh di depan gedung Dewan Perwakilan Rakyat Republik Indonesia (DPR RI). Demonstrasi ini dipicu oleh isu tunjangan dan fasilitas anggota DPR yang dinilai tidak sejalan dengan kondisi ekonomi masyarakat. Aksi tersebut kemudian mengalami eskalasi dengan keterlibatan mahasiswa dan pelajar dari berbagai daerah hingga berujung pada keributan di sekitar Gedung DPR RI dan terganggunya aktivitas publik (Rachel Kathryn, 2025).

Peristiwa tersebut tidak hanya berlangsung di ruang fisik, tetapi juga memicu perbincangan luas di ruang digital, khususnya melalui media sosial X. Media sosial berperan sebagai ruang diskusi publik yang memungkinkan masyarakat menyampaikan opini secara cepat, terbuka, dan masif (Ningrum & Hartanto, 2023). Menurut DataIndonesia.id yang merujuk pada laporan *We are Social* dan *Meltware*, Indonesia memiliki sekitar 24,85 juta pengguna X pada tahun 2024, menjadikannya salah satu negara dengan jumlah pengguna terbesar di dunia. Tingginya intensitas percakapan, penggunaan tagar, serta *respons* warganet

terhadap peristiwa demonstrasi menunjukkan bahwa *platform* ini menjadi arena penting dalam pembentukan opini publik digital (Ningrum & Hartanto, 2023).

Dibandingkan dengan media sosial lainnya, X dipilih karena memiliki karakteristik berupa penyebaran informasi yang cepat dan *real-time*, sehingga suatu isu sosial maupun politik dapat dengan mudah menjadi *trending* dan memicu banyak diskusi publik. Selain itu, sebagian besar interaksi pengguna pada *platform* X berbentuk teks yang berisi opini, kritik, maupun tanggapan terhadap suatu peristiwa. Karakteristik tersebut menjadikan media sosial X relevan sebagai sumber data dalam penelitian klasifikasi sentimen karena dapat merepresentasikan opini publik secara spontan dan aktual.

Opini yang berkembang di media sosial umumnya diekspresikan dalam bentuk sentimen, yaitu pandangan atau sikap seseorang terhadap suatu objek atau peristiwa yang dituangkan dalam teks (Indransyah et al., 2022). Dalam konteks demonstrasi, sentimen netizen sering kali bersifat emosional, menggunakan bahasa tidak baku, mengandung kritik tajam, serta dipengaruhi oleh dinamika sosial dan politik yang berkembang secara cepat. Karakteristik tersebut menyebabkan data teks di media sosial bersifat tidak terstruktur dan berdimensi tinggi, sehingga memerlukan pendekatan analisis berbasis komputasi untuk memperoleh gambaran opini publik secara objektif dan terukur.

Salah satu pendekatan yang dapat digunakan adalah analisis sentimen berbasis *Natural Language Processing* (NLP) dan *Machine Learning*. Melalui pendekatan ini, data teks yang tidak terstruktur dapat diproses, diekstraksi fiturnya, dan diklasifikasikan ke dalam kategori sentimen tertentu. Klasifikasi sentimen menjadi penting untuk mengidentifikasi kecenderungan opini masyarakat serta mendukung pengambilan keputusan berbasis data.

Berbagai metode klasifikasi telah dikembangkan dalam pendekatan *supervised learning*, di antaranya *Naïve Bayes*, *Support Vector Machine* (SVM), *Decision Tree*, *K-Nearest Neighbour* (KNN), *Random Forest*, *Linear Discriminant Analysis* (LDA), serta *Logistic Regression*. Setiap metode memiliki karakteristik

yang berbeda dalam mempelajari pola hubungan antara fitur teks dan label sentimen. Dalam penelitian ini, metode yang digunakan adalah *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA).

Naïve Bayes merupakan metode klasifikasi berbasis probabilitas yang dikenal sederhana, efisien secara komputasi, serta mampu bekerja dengan baik pada data berdimensi tinggi (Agustriya et al., 2024). Sementara itu, LDA merupakan metode yang dapat digunakan untuk klasifikasi dengan memaksimalkan keterpisahan antar kelas dan meminimalkan variasi dalam kelas, sehingga berpotensi meningkatkan performa klasifikasi pada data teks (Nurdiansyah et al., 2025). Pemilihan kedua metode tersebut didasarkan pada perbedaan pendekatan klasifikasi yang digunakan, sehingga menarik untuk dibandingkan performanya dalam klasifikasi sentimen pada media sosial X.

Sejumlah penelitian sebelumnya telah mengkaji penggunaan metode klasifikasi dalam analisis sentimen maupun bidang lain. Chrisinta & Simarmata (2023) menerapkan metode *Naïve Bayes* pada 8.000 data Twitter untuk menganalisis opini masyarakat terhadap pejabat publik dan memperoleh tingkat akurasi sebesar 64,55%, dengan mayoritas sentimen bersifat negatif. Selanjutnya, penelitian oleh Hanafi (2023) mengkombinasikan *Support Vector Machine* (SVM) dan *Linear Discriminant Analysis* (LDA) untuk menganalisis ulasan wisata di Kabupaten Gresik dan hasilnya menunjukkan bahwa LDA sebagai teknik reduksi dimensi mampu meningkatkan nilai *F1-score* menjadi 66%, dibandingkan model SVM tanpa LDA yang hanya mencapai 53%. Kemudian, penelitian oleh Nurdiansyah et al (2025) memanfaatkan kombinasi LDA dan KNN dalam klasifikasi citra sayuran dan berhasil mencapai akurasi 92,85%. Sementara itu, Razan et al (2023) membandingkan metode LDA dan *Naïve Bayes* pada dataset *obesity* dan *wine*, menunjukkan bahwa LDA memperoleh akurasi 92,2% dan 97,2%, sedangkan *Naïve Bayes* menunjukkan akurasi sebesar 58,1% dan 97,7%.

Berdasarkan hasil penelitian terdahulu, sebagian besar studi analisis sentimen masih berfokus pada ulasan produk, wisata, maupun opini terhadap tokoh publik, sedangkan penelitian yang membahas komentar netizen terkait aksi

demonstrasi di media sosial X masih terbatas. Selain itu, belum banyak penelitian yang secara khusus membandingkan performa metode *Naïve Bayes* dan LDA pada data teks media sosial dengan karakteristik bahasa yang tidak baku, emosional, serta mengandung kritik. Oleh karena itu, diperlukan penelitian untuk mengetahui metode yang lebih optimal dalam mengklasifikasikan sentimen masyarakat terhadap isu demonstrasi yang berkembang secara cepat di media sosial X.

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk menguji dan membandingkan performa *Naïve Bayes* dan LDA dalam mengklasifikasikan sentimen masyarakat terhadap aksi demonstrasi di sosial media X. Data penelitian diperoleh dari komentar atau unggahan netizen yang berkaitan dengan peristiwa demonstrasi, kemudian melalui tahapan pra-proses teks dan ekstraksi fitur sebelum dilakukan proses pelatihan model dan pengujian. Evaluasi performa kedua metode dilakukan menggunakan *confusion matrix* serta metrik akurasi, *precision*, dan *F1-score* untuk menentukan metode yang paling optimal dalam mengklasifikasikan sentimen.

Melalui penelitian ini, diharapkan dapat memperoleh pemahaman yang lebih komprehensif mengenai efektivitas kedua metode dalam mengolah data teks media sosial yang berdimensi tinggi, serta memberikan kontribusi akademik dalam pengembangan metode klasifikasi berbasis statistika dan *machine learning* pada analisis data teks.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dirumuskan beberapa rumusan masalah dalam penelitian ini yaitu:

1. Bagaimana prosedur penerapan metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam melakukan klasifikasi sentimen pada komentar netizen di X terkait aksi demonstrasi di Indonesia?
2. Bagaimana performa klasifikasi sentimen yang dihasilkan oleh metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) pada komentar netizen di X?

3. Bagaimana perbandingan performa klasifikasi antara metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam mengklasifikasikan sentimen pada komentar netizen di X terkait aksi demonstrasi di Indonesia?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan dalam penelitian ini adalah sebagai berikut:

1. Untuk mengetahui prosedur penerapan metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam klasifikasi sentimen pada komentar netizen di X terkait aksi demonstrasi di Indonesia.
2. Untuk mengetahui performa klasifikasi sentimen yang dihasilkan oleh metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) pada komentar netizen di X.
3. Untuk menentukan metode terbaik antara *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam mengklasifikasikan sentimen netizen X terkait aksi demonstrasi di Indonesia.

1.4 Manfaat Penelitian

Berdasarkan hasil dari penelitian ini, diharapkan dapat memberikan manfaat sebagai berikut:

1.4.1 Manfaat Teoritis

Menambah referensi mengenai penerapan metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam klasifikasi sentimen pada data media sosial, khususnya komentar netizen di X terkait aksi demonstrasi, serta menjadi bahan kajian bagi peneliti lain yang tertarik melakukan penelitian serupa dalam bidang analisis sentimen dan pembelajaran mesin (*machine learning*).

1.4.2 Manfaat Praktis

Memberikan informasi bagi pengambilan kebijakan, peneliti sosial, maupun pihak yang berkepentingan mengenai gambaran sentimen masyarakat terhadap aksi demonstrasi berdasarkan hasil klasifikasi. Selain itu, penelitian ini juga diharapkan

dapat membantu dalam pemilihan metode analisis yang lebih akurat untuk pengolahan data teks, khususnya dalam konteks pemantauan opini publik di media sosial.

1.5 Batasan Masalah

Adapun batasan masalah yang dilakukan pada penelitian ini agar lebih terarah, yaitu sebagai berikut:

1. Penelitian ini menggunakan data komentar netizen di aplikasi X yang berkaitan dengan aksi demonstrasi di Indonesia dan diperoleh melalui proses *scrapping* menggunakan kata kunci tertentu. Data yang digunakan yaitu data komentar netizen di X yang berbahasa Indonesia dan diambil dalam periode 1 Agustus hingga 30 September 2025.
2. Variabel penelitian yang digunakan berupa teks cuitan, yang selanjutnya diolah melalui tahapan pra-pemrosesan seperti *cleaning*, *case folding*, *spelling normalization*, *tokenizing*, *stopword removal*, dan *stemming*. Variabel lain di luar konten teks cuitan tidak digunakan dalam penelitian ini.
3. Metode klasifikasi sentimen yang digunakan dalam penelitian ini adalah metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dengan evaluasi performa model berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score* untuk menilai ketepatan hasil klasifikasi sentimen.
4. Alat pengolahan data menggunakan *software Visual Studio Code* (VS Code) dengan bahasa pemrograman Python versi 3.10.11, serta pustaka pendukung seperti *pandas*, *scikit-learn*, dan *NLTK*.

1.6 Tinjauan Pustaka

Sebelum memulai penelitian, penulis menggunakan beberapa literatur yang relevan dengan topik pembahasan yang telah dilakukan oleh peneliti terdahulu untuk melakukan tinjauan pustaka. Penelitian tersebut antara lain sebagai berikut:

1. Penelitian yang dilakukan Debora Chrisinta dan Justin Eduardo Simarmata (2023) dengan judul “Analisis Sentimen Penilaian Masyarakat Terhadap Pejabat Publik Menggunakan Algoritma *Naïve Bayes Classifier*”. Penelitian ini

menjelaskan tentang penerapan algoritma *Naïve Bayes* dalam melakukan klasifikasi sentimen data Twitter penilaian masyarakat terhadap pejabat publik. Hasil dari penelitian ini didapat bahwa performa algoritma *Naïve Bayes* menunjukkan nilai akurasi sebesar 64,55% dengan *error rate* sebesar 35,45%.

2. Penelitian yang dilakukan Mutmainnah Muchtar dan Rafiqah Arjaliyah Muchtar (2024) dengan judul “Integrasi Fitur Warna, Tekstur dan Renyi Fraktal Untuk Klasifikasi Penyakit Daun Kentang Menggunakan Linear Discriminant Analysis”. Penelitian ini menjelaskan tentang klasifikasi penyakit daun kentang dengan mengintegrasikan berbagai fitur ekstraksi dari citra daun menggunakan *Linear Discriminant Analysis* (LDA). Hasil dari penelitian ini menunjukkan bahwa *Linear Discriminant Analysis* (LDA) memberikan akurasi sebesar 98,46% dalam mengklasifikasikan penyakit daun kentang.
3. Penelitian yang dilakukan oleh Muhammad Irsyad Razan, Muhammad Fatchan, Regan Agam, Gita Suryani, Reza Hadi Asykari, Aldi Darma (2021) dengan judul “Studi Perbandingan Metode *Naïve Bayes* dan Linear Discriminant Analysis Untuk Permasalahan Klasifikasi”. Penelitian ini menjelaskan tentang perbandingan kinerja antara algoritma *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam permasalahan klasifikasi menggunakan data *Wine* dan *Obesity*. Hasil dari perbandingan tersebut untuk data *Wine* dengan algoritma LDA diperoleh tingkat akurasi sebesar 97,2%. Sedangkan, dengan menggunakan algoritma *Naïve Bayes* diperoleh tingkat akurasi sebesar 97,7%. Pada data *obesity*, dengan menggunakan algoritma LDA diperoleh tingkat akurasi sebesar 92,2%. Sedangkan, dengan menggunakan algoritma *Naïve Bayes* diperoleh tingkat akurasi sebesar 58,1%. Sehingga, dari hasil perbandingan metode dengan dua dataset dapat disimpulkan bahwa algoritma LDA memiliki tingkat keakuratan yang lebih baik dibandingkan klasifikasi *Naïve Bayes*.
4. Penelitian yang dilakukan oleh Ni Made Tara Okta Adrian, I Made Agus Dwi Suarjaya, Dwi Putra Githa (2023) dengan judul “Analisis Sentimen Publik Terhadap Aksi Demonstrasi di Indonesia Menggunakan Support Vector Machine dan Random Forest”. Penelitian ini menjelaskan tentang pandangan masyarakat terkait aksi demonstrasi yang terjadi di Indonesia melalui *tweets* pada media

sosial Twitter dengan menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest*. Hasil dari penelitian ini menunjukkan bahwa nilai akurasi paling baik adalah dengan menggunakan algoritma SVM, yaitu sebesar 80,3%. Sedangkan, algoritman *Random Forest* hanya memiliki nilai akurasi sebesar 77,7%.

5. Penelitian yang dilakukan oleh Muhammad Hanafi (2023) dengan judul “Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan *Linear Discriminant Analysis* (LDA) dan *Support Vector Machine* (SVM)”. Penelitian ini menjelaskan tentang bagaimana persepsi pengunjung terhadap destinasi wisata di Kabupaten Gresik dengan cara menganalisis ulasan yang ditulis wisatawan di Google Maps menggunakan metode *Support Vector Machine* (SVM) yang dikombinasikan dengan metode *Linear Discriminant Analysis* (LDA). Dari hasil penelitian tersebut, penggunaan algoritma SVM dengan kombinasi LDA menunjukkan nilai F1-score sebesar 66%, sedangkan penggunaan algoritma SVM tanpa LDA hanya memperoleh F1-score sebesar 53%. Sehingga, dapat disimpulkan bahwa kombinasi SVM dengan LDA dapat meningkatkan kinerja klasifikasi sentimen daripada penggunaan SVM tanpa LDA.
6. Penelitian yang dilakukan oleh Sylvi Anadas dan Ryan Randy Suryono (2025) dengan judul “Analisis Sentimen Terhadap Program Lapor Mas Wapres Periode 2024 di Media Sosial X Menggunakan Metode *Naïve Bayes* dan *Support Vector Machine* (SVM)”. Penelitian ini menjelaskan tentang analisis sentimen masyarakat terhadap program Lapor Mas Wapres melalui *platform* media sosial dengan menerapkan algoritma klasifikasi *Naïve Bayes* dan *Support Vector Machine* (SVM) serta pengaruh *Synthetic Minority Over-sampling Technique* (SMOTE). Hasil dari penelitian ini menunjukkan bahwa akurasi algoritma *Naïve Bayes* sebelum penerapan SMOTE didapat akurasi sebesar 63%, sedangkan setelah penerapan SMOTE didapat akurasi sebesar 78%. Jika menggunakan algoritma SVM didapat akurasi sebelum menerapkan SMOTE yaitu sebesar 80%, sedangkan setelah menerapkan SMOTE didapat akurasi sebesar 91%. Sehingga dapat disimpulkan bahwa algoritma SVM merupakan algoritma

dengan klasifikasi terbaik dan SMOTE terbukti efektif dalam meningkatkan akurasi model, serta hasil analisis sentimen menunjukkan bahwa masyarakat cenderung memberikan tanggapan positif terhadap program Lapor Mas Wapres.

Berikut merupakan rincian persamaan dan perbedaan penelitian ini dengan penelitian yang telah dipaparkan sebelumnya.

Tabel 1.1 Tinjauan Pustaka

Judul	Hasil	Persamaan	Perbedaan
Analisis Sentimen Penilaian Masyarakat Terhadap Pejabat Publik Menggunakan Algoritma <i>Naïve Bayes Classifier</i> (Chrisinta & Simarmata, 2023)	performa algoritma <i>Naïve Bayes</i> menunjukkan nilai akurasi sebesar 64,55% dengan <i>error rate</i> sebesar 35,45%.	Penggunaan metode algoritma <i>Naïve Bayes</i> dalam penelitian khususnya di bidang analisis sentimen.	Tidak menggunakan metode <i>Linear Discriminant Analysis (LDA)</i> sebagai perbandingan hasil akurasi.
Integrasi Fitur Warna, Tekstur dan Renyi Fraktal Untuk Klasifikasi Penyakit Daun Kentang Menggunakan Linear Discriminan Analysis (Muchtar et al., 2024)	<i>Linear Discriminant Analysis (LDA)</i> memberikan akurasi sebesar 98,46% dalam mengklasifikasikan penyakit daun kentang.	Penggunaan metode <i>Linear Discriminant Analysis (LDA)</i> dalam proses klasifikasi data.	Tidak membandingkan dengan metode <i>Naïve Bayes</i> serta tidak digunakan untuk klasifikasi sentimen.

Studi Perbandingan Metode <i>Naïve Bayes</i> dan <i>Linear Discriminant Analysis</i> Untuk Permasalahan Klasifikasi (Razan et al., 2023)	<i>Linear Discriminant Analysis</i> (LDA) memiliki nilai akurasi lebih tinggi (92,2%) dibandingkan dengan <i>Naïve Bayes</i> (58,1%) dalam permasalahan klasifikasi	Penggunaan metode <i>Linear Discriminant Analysis</i> (LDA) dan <i>Naïve Bayes</i> .	Studi kasus yang digunakan dan tidak menerapkan klasifikasi sentimen.
Analisis Sentimen Publik Terhadap Aksi Demonstrasi di Indonesia Menggunakan <i>Support Vector Machine</i> dan <i>Random Forest</i> (Adriana et al., 2023)	<i>Support Vector Machine</i> (SVM) memiliki nilai akurasi paling baik nilai akurasi paling baik yaitu sebesar 80,3%. Sedangkan, algoritman <i>Random Forest</i> hanya memiliki nilai akurasi sebesar 77,7%.	Penggunaan tema studi kasus dalam klasifikasi sentimen	Perbandingan metode yang digunakan dalam klasifikasi sentimen.
Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan <i>Linear Discriminant</i>	Kombinasi Algoritma SVM dengan LDA dapat meningkatkan klasifikasi sentimen dengan	Penggunaan metode <i>Linear Discriminant Analysis</i> (LDA)	Studi kasus yang digunakan dan tidak membandingkan dengan metode <i>Naïve Bayes</i> .

<i>Analysis</i> (LDA) dan <i>Support Vector Machine</i> (SVM) (Hanafi, 2023)	memperoleh F1-score sebesar 66% dibandingkan dengan algoritma SVM tanpa LDA yang hanya memperoleh F1-score sebesar 53%.		
Analisis Sentimen Publik Terhadap Program Laporan Mas Wapres Periode 2024 di Media Sosial X Menggunakan Metode Naïve Bayes dan Support Vector Machine (SVM) (Anadas & Suryono, 2025)	Algoritma SVM menunjukkan hasil akurasi yang paling tinggi yaitu sebesar 80%, sedangkan algoritma <i>Naïve Bayes</i> hanya memiliki akurasi sebesar 63%.	Penggunaan metode <i>Naïve Bayes</i> dalam klasifikasi sentimen.	Tidak menggunakan metode <i>Linear Discriminant Analysis</i> (LDA) sebagai perbandingan hasil akurasi dalam klasifikasi sentimen.

Berdasarkan penelitian terdahulu yang telah dikaji, dapat diketahui bahwa penelitian ini memiliki persamaan dengan penelitian terdahulu dalam penerapan metode klasifikasi pada analisis sentimen. Namun, penelitian tersebut masih memiliki perbedaan pada metode yang digunakan, objek penelitian, studi kasus, serta periode pengambilan data. Berdasarkan hal tersebut, penelitian ini membandingkan kinerja metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam klasifikasi sentimen netizen X terkait aksi demonstrasi periode 1 Agustus hingga 30 September 2025. Berdasarkan penelitian terdahulu yang

digunakan sebagai acuan, belum ditemukan penelitian yang secara khusus membandingkan kedua metode tersebut pada klasifikasi sentimen netizen X terkait aksi demonstrasi dengan periode data tersebut. Dengan demikian, penelitian ini diharapkan dapat memberikan referensi empiris mengenai performa kedua metode dalam mengklasifikasikan sentimen pada konteks tersebut.

1.7 Sistematika Penulisan

Sistematika penulisan disusun untuk memudahkan pembaca dalam memahami isi penelitian ini secara lebih sederhana, runtut, dan jelas. Adapun sistematika penulisan dalam penelitian ini adalah sebagai berikut:

1. BAB I PENDAHULUAN

Bab ini memuat latar belakang yang mendasari dilaksanakannya penelitian, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta tinjauan pustaka yang menjadi acuan teori. Pada bab ini juga disajikan gambaran umum mengenai sistematika penulisan sebagai arah dari keseluruhan isi penelitian.

2. BAB II LANDASAN TEORI

Bab ini berisi rangkuman teori dasar dan teori pendukung yang digunakan untuk memperkuat analisis dalam penelitian, khususnya terkait konsep analisis sentimen, metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA), serta teori-teori relevan lainnya yang menjadi pijakan dalam proses pengolahan data.

3. BAB III METODE PENELITIAN

Bab ini menjelaskan jenis penelitian, jenis dan sumber data yang digunakan, variabel penelitian, alat serta perangkat lunak yang dipakai, tahapan metode analisis, dan *flowchart* penelitian yang menggambarkan alur proses pelaksanaan penelitian dari awal hingga akhir.

4. BAB IV PEMBAHASAN

Bab ini membahas mengenai metode yang digunakan dalam penelitian, sampai dengan pembahasan yang mendalam mengenai metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA).

5. BAB V STUDI KASUS

Bab ini berisi tentang penerapan metode *Naïve Bayes* dan *Linear Discriminant Analysis* (LDA) dalam klasifikasi sentimen netizen di X tentang aksi demonstrasi menggunakan *software* VS Code dengan bahasa pemrograman Python versi 3.10.11.

6. BAB VI PENUTUP

Bab ini berisi tentang kesimpulan yang akan dihasilkan dari pembahasan dan jawaban atas rumusan masalah yang telah dirancang sebelumnya, serta berisikan saran bagi peneliti selanjutnya.



BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Prosedur klasifikasi sentimen pada komentar netizen di media sosial X terkait aksi demonstrasi di Indonesia dilakukan melalui beberapa tahapan, yaitu pengumpulan data yang menghasilkan sebanyak 8.702 data komentar. Data tersebut kemudian dilakukan *preprocessing* data yang meliputi *cleaning*, *case folding*, tokenisasi, *stopword removal*, dan *stemming* sehingga diperoleh sebanyak 6.443 data yang siap digunakan dalam proses klasifikasi. Selanjutnya, dilakukan pembobotan kata menggunakan TF-IDF dan pembagian data menjadi data *training* sebanyak 5.154 data serta data *testing* sebanyak 1.289 data. Kemudian data tersebut akan diklasifikasikan menggunakan metode Naïve Bayes dan *Linear Discriminant Analysis* (LDA) serta dilakukan evaluasi model menggunakan *confusion matrix*, *accuracy*, *precision*, *recall*, dan *f1-score*.
2. Berdasarkan hasil pengujian model, metode Naïve Bayes memperoleh nilai akurasi sebesar 74,01% dengan *weighted average f1-score* sebesar 72,62%, sedangkan metode LDA memperoleh nilai akurasi sebesar 79,21% dengan *weighted average f1-score* sebesar 79%. Hasil tersebut menunjukkan bahwa kedua metode mampu melakukan klasifikasi sentimen pada komentar netizen di media sosial X dengan cukup baik.
3. Berdasarkan perbandingan hasil evaluasi kedua metode tersebut, metode LDA memiliki performa klasifikasi yang lebih baik dibandingkan metode Naïve Bayes. Hal tersebut ditunjukkan oleh nilai akurasi dan *weighted average f1-score* yang lebih tinggi pada metode LDA. Dengan demikian, metode LDA lebih optimal digunakan dalam mengklasifikasikan sentimen komentar netizen di X terkait aksi demonstrasi di Indonesia.

6.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat diberikan untuk penelitian selanjutnya yaitu:

1. Proses pelabelan data disarankan tidak hanya menggunakan pendekatan otomatis, tetapi juga melibatkan pelabelan secara manual (semi-otomatis). Hal ini bertujuan agar label sentimen lebih sesuai dengan konteks komentar dan meningkatkan kualitas data penelitian.
2. Penggunaan kamus leksikon disarankan untuk dikembangkan secara mandiri sesuai dengan studi kasus penelitian, sehingga bobot setiap kata dapat disesuaikan berdasarkan tingkat pengaruhnya terhadap sentimen yang dianalisis.
3. Penelitian berikutnya dapat mempertimbangkan metode lain dalam mendeteksi komentar yang mengandung unsur sarkasme, karena pendekatan berbasis leksikon masih memiliki keterbatasan dalam memahami makna kalimat yang bersifat ironi atau sindiran.
4. Penelitian selanjutnya disarankan untuk dapat membandingkan performa metode yang digunakan dengan metode klasifikasi lainnya guna mengetahui metode yang paling optimal dalam klasifikasi sentimen.

DAFTAR PUSTAKA

- Agustriya, M., Ula, M., & Kurniawati. (2024). Analisis Kinerja Algoritma Klasifikasi Naive Bayes Menggunakan Genetic Algorithm dan Bagging untuk Data Publik Risiko Transaksi Kartu Kredit. *Jurnal Sistem dan Teknologi Informasi (JustIN)*, 12(3), 584. <https://doi.org/10.26418/justin.v12i3.80136>
- Ahzuri, A., Zulfah, Astuti, & Ediputra, K. (2024). Teorema Bayes; Statistika Matematika Kecerdasan dan Pembelajaran Mesin. *Jurnal Ilmiah Multidisiplin*, 1(2), 9–14.
- Amna, Wahyuddin, Sudipa, I. G. I., Putra, T. A. E., Wahidin, A. J., Syukrilla, W. A., Wardhani, A. K., Heryana, N., Indriyani, T., & Santoso, L. W. (2023). *Data Mining* (D. ediana & A. Yanto, Eds.; 1st ed.). PT Global Eksekutif Teknologi. www.globaleksekutifteknologi.co.id
- Anadas, S., & Suryono, R. R. (2025). Analisis Sentimen Publik Terhadap Program Laporan Mas Wapres Periode 2024 di Media Sosial X Menggunakan Metode Naive Bayes dan Support Vector Machine (SVM). *Jurnal Pendidikan Dan Teknologi Indonesia*, 5(8), 2183–2192. <https://doi.org/10.52436/1.jpti.934>
- Anton, Howard., & Rorres, Chris. (2013). *Elementary linear algebra, with applications. Student solutions manual*. J. Wiley.
- Billingsley, P. (1995). *Probability and Measure* (3rd ed.). John Wiley & Sons.
- Chebil, W., Wedyan, M., Alazab, M., Alturki, R., & Elshawees, O. (2023). Improving Semantic Information Retrieval Using Multinomial Naive Bayes Classifier and Bayesian Networks. *Information (Switzerland)*, 14(5). <https://doi.org/10.3390/info14050272>
- Chrisinta, D., & Simarmata, J. E. (2023). Analisis Sentimen Penilaian Masyarakat Terhadap Pejabat Publik Menggunakan Algoritma Naïve Bayes Classifier. *Komputika: Jurnal Sistem Komputer*, 12(1), 93–101. <https://doi.org/10.34010/komputika.v12i1.9638>
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2025). *Bayesian Data Analysis* (3rd ed.). CRC Press.
- Hanafi, M. (2023). *Analisis Sentimen Destinasi Wisata Kabupaten Gresik Menggunakan Linear Discriminant Analysis dan Support Vector Machine (SVM)*. [Skripsi, UIN Sunan Ampel Surabaya].
- Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*.
- Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The Elements of Statistical Learning*.
- Hutaahae, E. S., & Perdini, T. A. (2023). *Metode Penelitian Kuantitatif* (1st ed.). PT. Pena Persada Kerta Utama Redaksi.

- Indransyah, R., Chrisnanto, Y. H., & Sabrina, P. N. (2022). *Klasifikasi Sentimen Pergelaran MotoGP di Indonesia Menggunakan Algoritma Correlated Naive Bayes Classifier*. *Infotech Journal*, 8(2). <https://doi.org/10.31949/infotech.v8i2.3103>
- Karmini, N., & Ibrahim, A. (2025, August 30). Ketegangan meningkat di Indonesia seiring berlanjutnya protes atas kebrutalan polisi dan tunjangan anggota parlemen. *Associated Press*. <https://apnews.com/article/indonesia-students-protests-2b4ad65b836a3b38b6a037b2f45cb309>
- Kathryn, R. (2025, August 30). 11 Fakta Demo 28-29 Agustus 2025 yang Tewaskan Affan Kurniawan. *IDN Times*. <https://www.idntimes.com/news/indonesia/11-fakta-demo-28-29-agustus-2025-yang-tewaskan-affan-kurniawan-00-f48mr-thr5rc>
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Meester, R., & Stevens, L. (2024). Bayesian reasoning and the prior in court: not legally normative but unavoidable. *Law, Probability and Risk*, 23(1). <https://doi.org/10.1093/lpr/mgae001>
- Muchtar, M., & Muchtar, R. A. (2024). Integrasi Fitur Warna, Tekstur dan Renyi Fraktal untuk Klasifikasi Penyakit Daun Kentang Menggunakan Linear Discriminant Analysis. *MNEMONIC*, 7(1).
- Munawwaroh, N. H., Sya'roni, W., & Shudiq, W. J. (2024). *Naive Bayes Metode dan Penerapannya* (1st ed.). Pustaka Nurja.
- Muzakir, A., Adi, K., & Kusumaningrum, R. (2024). *Penerapan Konsep Machine Learning & Deep Learning* (1st ed.). UNDIP Press.
- Nasir, J., Pernando, Y., Wati, M., Yuniarthe, Y., Oktavia, & Nurhayati, S. (2025). *Pengantar Pembelajaran Machine Learning* (1st ed.). CV. Gita Lentera.
- Adriana, N. M. T. O., Suarjaya, I. M. A. D., & Githa, D. P. (2023). Analisis Sentimen Publik Terhadap Aksi Demonstrasi di Indonesia Menggunakan Support Vector Machine Dan Random Forest. *Decode: Jurnal Pendidikan Teknologi Informasi*, 3(2), 257–267. <https://doi.org/10.51454/decode.v3i2.187>
- Nurdiansyah, A., Erlanda, H., Roza, Y. B., & Sovia, R. (2025). Klasifikasi Citra dalam Identifikasi Kol dan Wortel Menggunakan Algoritma LDA dan KNN. *Journal of Science and Social Research*, 8(2), 1895–1902. <http://jurnal.goretanpena.com/index.php/JSSR>
- Razan, M. I., Fatchan, M., Agam, R., Suryani, G., Asykari, R. H., & Darma, A. (2023). Studi Perbandingan Metode Naive Bayes dan Linear Discriminant Analysis Untuk Permasalahan Klasifikasi. *Jurnal Rekayasa Elektro Sriwijaya*, 2(2), 2716–4063.
- Reddi, V. J. (2025). *Machine Learning Systems*.

- Saputra, D. I., & Hakim, D. L. (2022). *Implementasi Algoritma Gaussian Naive Bayes Classifier untuk Prediksi Potensi Tsunami Berbasis Mikrokontroler*.
- Sarker, I. H. (2021). Machine Learning: Algorithms, Real-World Applications and Research Directions. In *SN Computer Science* 2(2). Springer. <https://doi.org/10.1007/s42979-021-00592-x>
- Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding Machine Learning: From Theory to Algorithms*. <http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning>
- Sudirwo, Hadi, A., Judijanto, L., Purwandari, N., Ersela Zain, N. N. L., Rambe, K. H., Mukhlis, I. R., Jihadi, H., Mahliatussikah, H., Baskoro, B. H., & Yusufi, A. (2025). *Artificial Intelligence (Teori, Konsep, dan Implementasi di Berbagai Bidang)*. <https://www.researchgate.net/publication/391369694>
- Supandi, E. (2020). *Statistika dan Penerapannya* (1st ed.). PT Refika Aditama.
- Syaifudin, W. H., & Choiruddin, A. (2021). *Pengantar Teori Probabilitas dan Statistika*.
- Ningrum, M. W., & Hartanto, E. (2023). *Opini Publik pada Aktivisme Tanda Pagar “#” di Media Sosial Twitter (Social Network Analysis dan Sentiment Analysis Pengguna Twitter Terhadap #SayabersamaJokowi)*.
- Yusran, M., Siswanto, S., & Islamiyati, A. (2023). Comparison of Multinomial Naïve Bayes and Bernoulli Naïve Bayes on Sentiment Analysis of Kurikulum Merdeka with Query Expansion Ranking. *Jurnal Sistem Informasi*, 13(1), 96–106.