

SKRIPSI

**ANALISIS SENTIMEN CHATGPT PADA TWITTER DENGAN
MENGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER
(NBC) DENGAN LAPLACE ESTIMATOR**



Intelligentia - Dignitas

Disusun Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana
Program Studi Pendidikan Teknik Informatika dan Komputer

Oleh:

ADITYA NUGRAHA

NIM : 1512619035

UNIVERSITAS NEGERI JAKARTA

FAKULTAS TEKNIK

PROGRAM STUDI

PENDIDIKAN TEKNIK INFORMATIKA DAN KOMPUTER

TAHUN 2026

LEMBAR PENGESAHAN TUGAS AKHIR BENTUK : SKRIPSI

Nama Mahasiswa : Aditya Nugraha
 NIM : 1512619035
 Program Studi/Fakultas : S1 Pendidikan Teknik Informatika dan Komputer/Teknik
 Judul Penelitian : ANALISIS SENTIMEN CHATGPT PADA TWITTER
 DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER (NBC) DENGAN
 LAPLACE ESTIMATOR

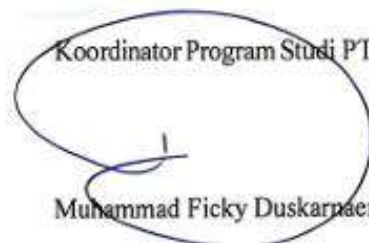
Ini telah dipertahankan dihadapan dewan penguji dan dinyatakan *Lulus*/
~~*Tidak Lulus*~~ * Ujian Tugas Akhir oleh Tim Penguji pada Tanggal 7 Juli 2026

Tim Penguji,

Ketua Penguji	Z.E Ferdi Fauzan Putra, S.Pd., M.Pd.T. NIDN. 0003029004	
Anggota Penguji	Wiranti Kusuma Hapsari, S.Kom., M.Cs. NIDN. 0416079403	
Dosen Ahli	Vannisa Irma Dewi, S.Pd., M.T.I. NUPTK. 2440771672230322	
Dosen Pembimbing 1	Dr. Widodo, M.Kom. NIDN. 0025037206	
Dosen Pembimbing 2	Ressy Dwitias Sari, S.T, M.T.I NIDN. 0115098902	



Mengetahui,
 Dekan, Fakultas Pendidikan,
 Dr. Neneng Siti Silfi
 Ambarwati, S.Si., Apt., M.Si.
 NIDN. 0029027204



Koordinator Program Studi PTIK,
 Muhammad Ficky Duskarnaen, M.Sc
 NIDN. 0029027304

LEMBAR PERSETUJUAN PEMBIMBING TUGAS AKHIR
BENTUK : SKRIPSI

Nama Mahasiswa : Aditya Nugraha

NIM : 1512619035

Judul Penelitian : ANALISIS SENTIMEN CHATGPT PADA TWITTER
DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER
(NBC) DENGAN LAPLACE ESTIMATOR

ini telah disetujui dan dinyatakan memenuhi syarat untuk diajukan dalam ujian
tugas akhir

Jakarta, 31 Juli 2026

Pembimbing I

Pembimbing II



Dr. Widodo, M.Kom.



Ressa Dwitias Sari, S.T, M.T.I

NIDN. 0025037206

NIDN. 0115098902

LEMBAR PERNYATAAN KEASLIAN
SURAT PERNYATAAN KEASLIAN KARYA

Nama Mahasiswa : Aditya Nugraha
NIM : 1512619035
Program Studi : Pendidikan Teknik Informatika dan Komputer
Fakultas : Teknik
Judul TA : ANALISIS SENTIMEN CHATGPT PADA TWITTER
DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES CLASSIFIER
(NBC) DENGAN LAPLACE ESTIMATOR

Menyatakan bahwa Tugas Akhir ini beserta seluruh isinya adalah benar-benar karya saya sendiri. Saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika ilmu yang berlaku dalam masyarakat keilmuan. Atas pernyataan ini, saya siap menanggung risiko/sanksi apabila di kemudian hari ditemukan adanya pelanggaran etika keilmuan atau ada klaim dari pihak lain terhadap keaslian karya tulis ilmiah ini.

Jakarta, 31 Juli 2026

Yang menyatakan,



Aditya Nugraha

NIM. 1512619035

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadiran Tuhan Yang Maha Esa, karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul “Analisis Sentimen *ChatGPT* Pada *Twitter* Dengan Menggunakan Algoritma *Naive Bayes Classifier* (NBC) Dengan *Laplace Estimator*” dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan pada Program Studi Pendidikan Teknik Informatika dan Komputer, Fakultas Teknik, Universitas Negeri Jakarta.

Dalam proses penyusunan skripsi ini, penulis banyak mendapatkan bantuan, bimbingan, serta dukungan dari berbagai pihak, baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih kepada:

1. Prof. Dr. Komarudin, M.Si., selaku Rektor Universitas Negeri Jakarta, atas kesempatan dan fasilitas yang diberikan sehingga penulis dapat menempuh dan menyelesaikan pendidikan dengan baik.
2. Neneng Siti Silfi Ambarwati, S.Si, Apt., M.Si., selaku Dekan Fakultas Teknik Universitas Negeri Jakarta, atas dukungan serta kebijakan akademik yang membantu kelancaran studi penulis.
3. Bapak Muhammad Ficky Duskarnaen, S.T., M.Sc., selaku Koordinator Program Studi Pendidikan Teknik Informatika dan Komputer atas arahan dan perhatian yang diberikan dalam mendukung mahasiswa menyelesaikann studi tepat waktu.
4. Bapak Dr. Widodo, M.Kom., selaku Dosen Pembimbing I, yang dengan penuh dedikasi dan kesabaran telah memberikan bimbingan, arahan, serta masukan yang sangat berharga selama proses penyusunan skripsi ini.
5. Ibu Ressy Dwitias Sari, S.T, M.Ti., selaku Dosen Pembimbing II, yang dengan ketulusan dan perhatian telah memberikan bimbingan, motivasi, serta dukungan selama penulis menyelesaikan skripsi ini.

6. Seluruh dosen dan staff Program Studi Pendidikan Teknik Informatika dan Komputer, yang telah membekali penulis dengan ilmu pengetahuan serta bantuan selama masa perkuliahan.
7. Kedua orang tua tercinta, Alm. Rahardjo dan Ibu Siti Indasah, atas doa, kasih sayang, serta dukungan moral yang tiada henti diberikan kepada penulis.
8. Indra Utami, dan Teguh Rahadian, selaku kakak penulis yang selalu memberikan semangat dan dukungan.
9. Teman-teman seperjuangan Program Studi Pendidikan Teknik Informatika dan Komputer angkatan 2019, yang telah memberikan semangat, bantuan dan kebersamaan selama masa studi.
10. Serta seluruh pihak yang terlibat secara langsung maupun tidak langsung yang namanya tidak bisa penulis sebutkan satu-persatu

Penulis menyadari bahwa skripsi ini masih memiliki berbagai kekurangan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun guna penyempurnaan di masa mendatang. Semoga karya ini dapat memberikan manfaat serta kontribusi bagi perkembangan ilmu pengetahuan.

Jakarta, Juni 2026



Aditya Nugraha

No. Reg. 1512619035

ABSTRAK

Aditya Nugraha, Analisis Sentimen *ChatGPT* Pada *Twitter* Dengan Menggunakan Algoritma *Naive Bayes Classifier* (NBC) Dengan *Laplace Estimator* . Dosen Pembimbing : Dr. Widodo, M.Kom. dan Ressy Dwitias Sari, S.T, M.Ti Program Studi Pendidikan Teknik Informatika dan Komputer . Fakultas Teknik. Universitas Negeri Jakarta. 2026

Perkembangan teknologi kecerdasan buatan, khususnya *ChatGPT*, telah memicu beragam respons masyarakat Indonesia yang banyak diungkapkan melalui media sosial *Twitter/X*. Analisis sentimen terhadap opini tersebut menjadi penting untuk memahami persepsi publik, namun klasifikasi teks menggunakan *Naive Bayes Classifier* (NBC) masih menghadapi kendala berupa *zero probability* problem dan ketidakseimbangan distribusi kelas yang dapat menurunkan kualitas prediksi. Penelitian ini bertujuan mengklasifikasikan sentimen tweet berbahasa Indonesia tentang *ChatGPT* ke dalam tiga kelas, yaitu Positif, Negatif, dan Netral, serta mengevaluasi pengaruh penerapan *Laplace Estimator* dan teknik *resampling* terhadap performa klasifikasi. Penelitian menggunakan 4.590 tweet yang diproses melalui tahapan *preprocessing*, pembobotan fitur menggunakan TF-IDF, serta penanganan ketidakseimbangan data menggunakan kombinasi *Synthetic Minority Oversampling Technique* (SMOTE) dan *Random Undersampling*. Evaluasi dilakukan pada empat konfigurasi model, yaitu NBC tanpa *Laplace Estimator* dan tanpa *resampling*, NBC dengan *resampling*, NBC dengan *Laplace Estimator* tanpa *resampling*, serta NBC dengan *Laplace Estimator* dan *resampling*. Hasil penelitian menunjukkan bahwa penerapan *Laplace Estimator* meningkatkan akurasi model pada data tanpa *resampling* dari 19,06% menjadi 72,66%. Namun, model tersebut masih menunjukkan kecenderungan memprediksi kelas mayoritas sehingga performanya kurang seimbang antarkelas. Sementara itu, model NBC dengan *Laplace Estimator* dan *resampling* dipilih sebagai model terbaik karena menghasilkan performa klasifikasi yang lebih seimbang dengan akurasi sebesar 64,72%, presisi 65,55%, recall 64,72%, dan F1-score 64,67%, serta F1-score antarkelas yang relatif merata (Negatif 0,67; Netral 0,61; Positif 0,66). Hasil penelitian menunjukkan bahwa kombinasi *Laplace Estimator* dan teknik *resampling* efektif dalam mengatasi *zero probability* problem serta mengurangi dampak ketidakseimbangan kelas, sehingga menghasilkan performa klasifikasi sentimen yang lebih seimbang pada data tweet berbahasa Indonesia.

Kata Kunci: Analisis Sentimen; *ChatGPT*; *Naive Bayes Classifier*; *Laplace Estimator*; *Resampling*.

ABSTRACT

Aditya Nugraha, “*ChatGPT Sentiment Analysis on Twitter Using the Naive Bayes Classifier (NBC) Algorithm with the Laplace Estimator.*” Supervisors: Dr. Widodo, M.Kom., and Ressy Dwitias Sari, S.T., M.Ti., *Study Program of Informatics and Computer Engineering Education, Faculty of Engineering, Universitas Negeri Jakarta.* 2026

The rapid advancement of artificial intelligence, particularly ChatGPT, has generated diverse public responses in Indonesia, many of which are expressed through the Twitter/X social media platform. Sentiment analysis is essential for understanding public perception; however, text classification using the Naive Bayes Classifier (NBC) is often challenged by the zero probability problem and class imbalance, both of which may reduce prediction performance. This study aims to classify Indonesian-language tweets about ChatGPT into three sentiment categories, namely Positive, Negative, and Neutral, while evaluating the impact of the Laplace Estimator and resampling techniques on classification performance. A dataset consisting of 4,590 tweets was processed through text preprocessing, TF-IDF feature weighting, and class imbalance handling using a combination of the Synthetic Minority Oversampling Technique (SMOTE) and Random Undersampling. Four model configurations were evaluated: NBC without Laplace Estimator and without resampling, NBC with resampling, NBC with Laplace Estimator without resampling, and NBC with both Laplace Estimator and resampling. The results show that applying the Laplace Estimator increased the classification accuracy on the non-resampled dataset from 19.06% to 72.66%. However, the model remained biased toward the majority class, resulting in less balanced performance across sentiment categories. In contrast, the NBC model with Laplace Estimator and resampling was selected as the best-performing model because it achieved the most balanced classification performance, with an accuracy of 64.72%, precision of 65.55%, recall of 64.72%, and an F1-score of 64.67%, while producing relatively balanced class-wise F1-scores (Negative 0.67, Neutral 0.61, and Positive 0.66). These findings demonstrate that combining the Laplace Estimator with resampling effectively addresses the zero probability problem and mitigates the impact of class imbalance, thereby improving the overall balance of sentiment classification performance on Indonesian tweet datasets.

Keywords: *Sentiment Analysis; ChatGPT; Naive Bayes Classifier; Laplace Estimator; Resampling.*

DAFTAR ISI

LEMBAR PENGESAHAN TUGAS AKHIR.....	ii
LEMBAR PERSETUJUAN PEMBIMBING TUGAS AKHIR.....	iii
LEMBAR PERNYATAAN KEASLIAN	iv
KATA PENGANTAR.....	v
ABSTRAK.....	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Identifikasi Masalah	7
1.3 Pembatasan Masalah.....	7
1.4 Perumusan Masalah.....	8
1.5 Tujuan Penelitian	8
1.6 Manfaat Penelitian	8
BAB II KAJIAN TEORI.....	9
2.1 Kerangka Teoritik.....	9
2.1.1 Analisis sentimen.....	9
2.1.2 Text Mining.....	10
2.1.3 Text Preprocessing.....	10
2.1.4 Machine Learning.....	14
2.1.5 Data Mining	15
2.1.6 Twitter	16
2.1.7 Naive Bayes.....	19
2.1.8 Laplace Smoothing.....	21
2.1.9 Confusion Matrix	22
2.1.10 PySastrawi.....	23
2.1.11 Kendala yang dihadapi	24
2.2 Penelitian yang Relevan	25

2.3	Kerangka Berpikir	31
BAB III METODE PENELITIAN		32
3.1	Tempat dan Waktu Penelitian	32
3.2	Alat dan Bahan Penelitian	33
3.2.1	Perangkat Keras (<i>Hardware</i>)	33
3.2.2	Perangkat Lunak (<i>Software</i>)	33
3.3	Diagram Alir Penelitian	34
3.3.1	Melakukan <i>Crawling Data</i>	35
3.3.2	Melakukan Pelabelan Data	36
3.3.3	Melakukan <i>Preprocessing Data</i>	38
3.4	Menghitung TF-IDF	43
3.5	Metode <i>Naive Bayes</i>	43
3.6	Metode Evaluasi	46
BAB IV HASIL PENELITIAN DAN ANALISA DATA		48
4.1	Deskripsi Hasil Penelitian	48
4.1.1	Pengumpulan Data	48
4.1.2	<i>Preprocessing Data</i>	51
4.1.3	Pembobotan TF-IDF	53
4.1.4	<i>Resampling</i>	54
4.1.5	<i>Naive Bayes Classifier</i>	54
4.1.6	<i>Naive Bayes Classifier Dengan Laplace Estimator</i>	56
4.1.7	<i>Naive Bayes Classifier Dengan Resampling</i>	58
4.1.8	<i>Naive Bayes Classifier Dengan Laplace Estimator Dan Resampling</i>	60
4.2	Pembahasan	62
4.2.1	Distribusi Data Sebelum dan Sesudah <i>Resampling</i>	62
4.2.2	Hasil <i>Naive Bayes Classifier</i>	63
4.2.3	Hasil <i>Naive Bayes Classifier Dengan Laplace Estimator</i>	65
4.2.4	Hasil <i>Naive Bayes Classifier Dengan Resampling</i>	68
4.2.5	Hasil <i>Naive Bayes Classifier Dengan Resampling Dan Laplace Estimator</i> 70	
4.2.6	Perbedaan Keseluruhan 4 Model	73
4.2.7	Analisis Top Kata TF-IDF	77
4.3	Aplikasi Hasil Penelitian	78

BAB V PENUTUP	79
5.1. Kesimpulan	79
5.2. Saran	80
DAFTAR PUSTAKA	81
LAMPIRAN	88



DAFTAR TABEL

Tabel 2. 1 Confusion Matrix.....	22
Tabel 2. 2 Penelitian Relevan	27
Tabel 3. 1 Tempat dan Waktu Penelitian	32
Tabel 3. 2 Pelabelan Data.....	36
Tabel 3. 3 Cleansing.....	38
Tabel 3. 4 Case Folding.....	39
Tabel 3. 5 Tokenizing.....	39
Tabel 3. 6 Normalisasi	40
Tabel 3. 7 Stemming	41
Tabel 3. 8 Stopword Removal	42
Tabel 3. 9 Prediksi Kelas.....	46
Tabel 4. 1 Contoh Data Tweet yang Diberi Label	49
Tabel 4. 2 Hasil Perbandingan Data Tweet	50
Tabel 4. 3 Proses Cleaning dan Case Folding	51
Tabel 4. 4 Proses Tokenizing.....	51
Tabel 4. 5 Proses Normalisasi	52
Tabel 4. 6 Proses Stemming	52
Tabel 4. 7 Proses Stopword Removal	53
Tabel 4. 8 Contoh Proses TF-IDF.....	53
Tabel 4. 9 Hasil Distribusi Data.....	54
Tabel 4. 10 Confusion Matrix NBC.....	54
Tabel 4. 11 Ringkasan Hasil NBC	56
Tabel 4. 12 Confusion Matrix NBC Dengan Laplace Estimator.....	56
Tabel 4. 13 Ringkasan Hasil NBC Dengan Laplace Estimator Tanpa Resampling	58
Tabel 4. 14 Confusion Matrix Naive Bayes Classifier Dengan Resampling	58
Tabel 4. 15 Ringkasan hasil Naive Bayes Classifier Dengan Resampling	60
Tabel 4. 16 Confusion Matrix Naive Bayes Classifier dengan Laplace Estimator dan Resampling.....	60
Tabel 4. 17 Ringkasan hasil Naive Bayes Classifier dengan Laplace Estimator dan Resampling	62
Tabel 4. 18 Perbandingan Hasil Evaluasi Keempat Model.....	74

DAFTAR GAMBAR

Gambar 2. 1 Text Preprocessing	11
Gambar 2. 2 Data Twitter Tiap Tahun (Recurpost.com)	17
Gambar 2. 3 Data Pengguna Twitter di Tiap Negara (Recurpost.com)	18
Gambar 2. 4 Kerangka Berpikir	31
Gambar 3. 1 Diagram Alir Penelitian	34
Gambar 3. 2 Auth Token	35
Gambar 3. 3 Package Python	35
Gambar 3. 4 Crawling Data.....	36
Gambar 3. 5 Perbandingan Diagram Alir Naive Bayes Classifier dan Naive Bayes Classifier dengan Laplace Smoothing.....	45
Gambar 4. 1 Hasil Crawling Data Tweet	48
Gambar 4. 2 Distribusi Label Sentimen.....	62
Gambar 4. 3 Confusion Matrix NBC.....	64
Gambar 4. 4 Hasil Klasifikasi NBC.....	65
Gambar 4. 5 Confusion Matrix NBC Dengan Laplace Estimator	66
Gambar 4. 6 Hasil Klasifikasi NBC Dengan Laplace Estimator.....	67
Gambar 4. 7 Confusion Matrix NBC Dengan Resampling.....	68
Gambar 4. 8 Hasil Klasifikasi NBC Dengan Resampling	69
Gambar 4. 9 Confusion Matrix NBC Dengan Resampling dan Laplace Estimator	71
Gambar 4. 10 Hasil Klasifikasi NBC dengan Resampling dan Laplace Estimator	72
Gambar 4. 11 Perbandingan Hasil Akurasi Keempat Model	73
Gambar 4. 12 Perbandingan 4 Metrik Semua Model	73
Gambar 4. 13 Pengaruh Laplace Estimator Terhadap Performa Model	74
Gambar 4. 14 Pengaruh Resampling Terhadap Performa Model.....	75
Gambar 4. 15 Kata Paling Berpengaruh per Kelas Sentimen	77

DAFTAR LAMPIRAN

Lampiran 1 Contoh Sentimen dari Berbagai Sosial Media	88
Lampiran 2 Hasil <i>Crawling</i> Data Twitter	91
Lampiran 3 <i>Source Code Crawling</i> Data.....	92
Lampiran 4 <i>Source Code</i> Penggabungan Data	94
Lampiran 5 <i>Source Code Preprocessing</i>	103
Lampiran 6 <i>Source Code</i> TF-IDF	116
Lampiran 7 <i>Source Code Resampling</i>	123





KEMENTERIAN PENDIDIKAN TINGGI, SAINS DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
PERPUSTAKAAN DAN KEARSIPAN

Jalan Rawamangun Muka Jakarta 13220

Telepon/Faksimili: 021-4894221

Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Aditya Nugraha

NIM : 1512619035

Fakultas/Prodi : Teknik/Pendidikan Teknik Informatika dan Komputer

Alamat email : adigrah8@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan dan Kearsipan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (.....)

yang berjudul :

ANALISIS SENTIMEN CHATGPT PADA TWITTER DENGAN MENGGUNAKAN
ALGORITMA NAIVE BAYES CLASSIFIER (NBC) DENGAN LAPLACE ESTIMATOR

Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan dan Kearsipan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara **fulltext** untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 10 Agustus 2026

Penulis

(**Aditya Nugraha**)
nama dan tanda tangan