

**ANALISIS SENTIMEN PUBLIK TERHADAP KUALITAS
LAYANAN PT PLN DI X MENGGUNAKAN
ALGORITMA *NAÏVE BAYES* DAN
*K-NEAREST NEIGHBOR***

SKRIPSI

Oleh:

**Nurul Zahra
202110715022**



**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BHAYANGKARA JAKARTA RAYA
2025**

LEMBAR PERSETUJUAN PEMBIMBING

Judul Skripsi : Analisis Sentimen Publik Terhadap Kualitas Layanan
PT PLN di X Menggunakan Algoritma *Naïve Bayes*
dan *K-Nearest Neighbor*

Nama Mahasiswa : Nurul Zahra
Nomor Pokok Mahasiswa : 202110715022
Program Studi/Fakultas : Informatika / Ilmu Komputer
Tanggal Lulus Ujian Skripsi : 29 Juli 2025

Jakarta, 05 Agustus 2025
MENYETUJUI,

Pembimbing I

Wowon Priatna, S.T., M.TI
NIDN: 0429118007

Pembimbing II

Dr. Dra. Tyastuti Sri Lestari, M.M.
NIDN: 0327036701

Ketua Program Studi

Ahmad Fathurrozi, S.E., M.M.S.I
NIP: 2012486

Program Studi Informatika
Fakultas Ilmu Komputer
Universitas Bhayangkara Jakarta Raya
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir : Analisis Sentimen Publik Terhadap Kualitas Layanan PT
PLN Di X Menggunakan Algoritma Naïve Bayes
Nama Mahasiswa : Nurul Zahra
Nomor Pokok Mahasiswa : 202110715022
Program Studi/Fakultas : Informatika / Ilmu Komputer
Tanggal Lulus Ujian Tugas Akhir : 29 Juli 2025

Jakarta, 05 Agustus 2025

MENGESAHKAN,

Ketua Tim Penguji : Dr.Rakhmi Khalida, S.T., M.M.S.I.

NIDN: 0304099201

Penguji I : Allan D. Alexander, S.T., M. Kom

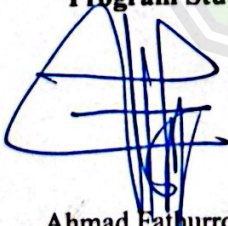
NIDN: 0305127404

Penguji II : Wowon Priatna, S.T., M.T.I

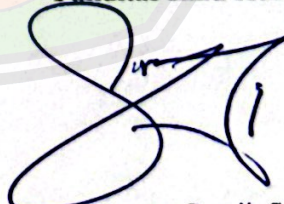
NIDN: 0429118007

MENGETAHUI,

Ketua
Program Studi Informatika


Ahmad Fathurrozi, S.E., M.M.S.I
NIP: 2012486

Dekan
Fakultas Ilmu Komputer


Dr. Robertus Suraji, S.S., M.A
NIP: 1904406



UNIVERSITAS BHAYANGKARA JAKARTA RAYA
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA

LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Nurul Zahra
NPM : 202110715022
Program Studi : Informatika
Fakultas : Ilmu Komputer
Judul Tugas Akhir : Analisis Sentimen Publik Terhadap Kualitas Layanan
PT.PLN Di X Menggunakan Algoritma Naïve Bayes
Dan K-Nearest Neighbo

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan hasil karya saya sendiri dan benar keasliannya. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak mana pun.

Bekasi, 5 Agt 2025

Penulis,

Nurul Zahra
202110715022



ABSTRAK

Nurul Zahra, 202110715022. Analisis Sentimen Publik Terhadap Kualitas Layanan PT PLN di X Menggunakan Algoritma *Naïve Bayes* dan *K-Nearest Neighbor*. Fakultas Ilmu Komputer, Universitas Bhayangkara Jakarta Raya, 2025.

PT PLN (Persero) sebagai penyedia layanan ketenagalistrikan milik negara memiliki tanggung jawab penting dalam memberikan kualitas pelayanan terbaik, mengingat listrik merupakan kebutuhan pokok masyarakat. Namun demikian, masih banyak keluhan publik yang menyoroti kurangnya responsivitas PLN, terutama yang disampaikan melalui platform X (sebelumnya Twitter). Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kualitas layanan PLN yang diungkapkan di media sosial X, serta membandingkan performa algoritma *Naïve Bayes* dan *K-Nearest Neighbor* (KNN) dalam mengklasifikasikan sentimen menjadi kategori positif, negatif, dan netral. Penelitian ini menggunakan pendekatan *Knowledge Discovery in Databases* (KDD) yang meliputi tahapan pengumpulan data, preprocessing (case folding, tokenizing, filtering, stemming), transformasi dengan pembobotan TF-IDF, serta klasifikasi menggunakan algoritma *Naïve Bayes* dan KNN. Evaluasi dilakukan menggunakan confusion matrix untuk mengukur nilai akurasi, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa algoritma *Naïve Bayes* menghasilkan akurasi sebesar 87%, precision 88%, recall 86%, dan F1-score 87%. Sedangkan algoritma KNN menghasilkan akurasi sebesar 86%, precision 87%, recall 85%, dan F1-score 86%. Nilai-nilai ini menunjukkan bahwa *Naïve Bayes* sedikit lebih unggul dalam mengklasifikasikan sentimen publik terhadap layanan PLN. Hasil analisis ini diharapkan tidak hanya memberikan wawasan tentang persepsi publik terhadap kinerja layanan PLN, tetapi juga menjadi dasar bagi implementasi nyata, seperti peningkatan kecepatan respon terhadap keluhan pelanggan, perbaikan komunikasi digital, serta pengambilan kebijakan berbasis data opini masyarakat secara real-time. Penelitian ini juga dapat menjadi referensi bagi pengembangan lanjutan dalam bidang analisis sentimen berbasis machine learning.

Kata Kunci: Analisis Sentimen, *Naïve Bayes*, *K-Nearest Neighbor*, Kualitas Layanan PLN

ABSTRACT

Nurul Zahra, 202110715022. *Public Sentiment Analysis on the Quality of PT PLN Services on X Using Naïve Bayes and K-Nearest Neighbor Algorithms. Faculty of Computer Science, Univercity of Bhayangkara Jakarta Raya, 2025.*

PT PLN (Persero), as a state-owned electricity service provider, holds a crucial responsibility in delivering high-quality services, considering that electricity is a basic necessity for society. However, numerous public complaints have highlighted PLN's lack of responsiveness, particularly voiced through the X platform (formerly Twitter). This study aims to analyze public sentiment regarding the quality of PLN's services expressed on social media X, and to compare the performance of the Naïve Bayes and K-Nearest Neighbor (KNN) algorithms in classifying sentiments into positive, negative, and neutral categories. This research adopts the Knowledge Discovery in Databases (KDD) approach, consisting of data collection, preprocessing (case folding, tokenizing, filtering, stemming), transformation using TF-IDF weighting, and classification using Naïve Bayes and KNN algorithms. Evaluation is conducted using confusion matrix to measure accuracy, precision, recall, and F1-score values. The results show that the Naïve Bayes algorithm achieved an accuracy of 87%, precision of 88%, recall of 86%, and F1-score of 87%. Meanwhile, the KNN algorithm achieved an accuracy of 86%, precision of 87%, recall of 85%, and F1-score of 86%. These results indicate that Naïve Bayes slightly outperforms KNN in classifying public sentiment towards PLN services. The findings are expected to not only provide insight into public perceptions of PLN's service performance but also serve as a basis for practical implementation, such as improving response speed to customer complaints, enhancing digital communication, and supporting data-driven decision-making based on real-time public opinion. This research may also serve as a reference for further development in sentiment analysis using machine learning techniques.

Keywords: *Sentiment Analysis, Naïve Bayes, K-Nearest Neighbor, PT PLN Service Quality*

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

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan dibawah ini:

Nama : Nurul Zahra
NPM : 202110715022
Program Studi : Informatika
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Eksklusif (*Non-Exclusive Royalty Free Rights*)**, atas karya ilmiah saya yang berjudul:

Analisis Sentimen Publik Terhadap Kualitas Layanan PT PLN di X Menggunakan Algoritma *Naïve Bayes* dan *K-Nearest Neighbor*

Beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta
Pada tanggal : 05 Agustus 2025

Yang Menyatakan



Nurul Zahra

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul **“Analisis Sentimen Publik terhadap Kualitas Layanan PT PLN di X (Twitter) menggunakan Algoritma Naïve Bayes dan K-Nearest Neighbor”** dengan baik dan tepat waktu. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika.

Penulis menyadari bahwa terselesaikannya skripsi ini tidak lepas dari bantuan, bimbingan, dukungan, serta doa dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Irjen Pol (Pum) Prof. Dr. Drs. H. Bambang Karsono, S.H., M.M., Ph.D.,D.Crim(Hc) selaku Rektor Universitas Bhayangkara Jakarta.
2. Bapak Dr. Robertus Suraji, S.S., M.A., selaku Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I., selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.

4. Bapak Wowon Priatna, S.T., M.T.I., selaku Dosen Pembimbing I, atas bimbingan, arahan dan pembelajaran yang telah diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini.
5. Ibu Dr. Dra.Tyastuti Sri Lestari, M.M. selaku Dosen Pembimbing II, atas bimbingan, arahan dan pembelajaran yang telah diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini.
6. Ibu Prima Dina Atika. S.Kom., M.Kom., selaku Dosen Pembimbing Akademik, atas bimbingan, arahan dan pembelajaran yang telah diberikan kepada penulis sehingga dapat menyelesaikan skripsi ini.
7. Kedua orang tua, Alm. Ayahanda Syamsuri dan Ibunda Mahrojah yang sangat penulis cintai, yang selalu memberikan dukungan juga dan semangat kepada penulis dan tidak henti mendoakan penulis agar diberikan kelancaran dan kemudahan dalam menyelesaikan skripsi ini. Terimakasih atas setiap lelah dan upaya mamah dalam mencari rezeki untuk anak-anak mamah agar bisa mendapatkan Pendidikan yang layak, semoga mamah Panjang umur dan diberikan kesehatan, agar bisa menemani setiap proses anak-anak mamah.
8. Adnan Syam Itsnain, selaku adik penulis yang turut memberikan dukungan serta semangat kepada penulis.
9. Tia Aulia, Tata Arya, Nabila Ramadhita, Anisya yang selalu memberikan semangat dan dukungan serta kebersamaan penulis dalam proses penulisan skripsi ini.

10. Teman -Teman, Khoirunnisa, Erinta, Bella, Syntia, Tasya, Navisha, yang selalu memberikan dukungan selama proses penulisan skripsi.
11. Terakhir, saya sangat berterimakasih kepada diri saya sendiri Nurul Zahra, karena telah bertahan dan bertanggung jawab untuk menyelesaikan apa yang telah dimulai. Terimakasih karena terus berusaha dan berjuang serta menikmati setiap prosesnya dan juga membuktikan kepada semua orang bahwa saya bisa melalui proses yang tidak mudah ini, Terimakasih sudah bertahan.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa mendatang. Semoga skripsi ini dapat memberikan manfaat dan kontribusi bagi pengembangan ilmu pengetahuan, khususnya di bidang analisis data dan data mining.

Akhir kata, semoga Allah SWT senantiasa melimpahkan rahmat dan ridho-Nya kepada kita semua.

Bekasi,



Nurul Zahra

NPM: 202110715022

x

DAFTAR ISI

LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN BUKAN PLAGIASI	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xv
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	4
1.4 Batasan Masalah.....	4
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
1.7 Sistematika Penulisan	6
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terdahulu	7
2.2 <i>Data Science</i>	9
2.3 <i>Data Mining</i>	9
2.4 <i>Text Mining</i>	11
2.5 <i>Text preprocessing</i>	12
2.6 <i>Machine learning</i>	13

2.7	<i>Knowledge Discovery in Database</i>	14
2.8	<i>Term Frequency-Inverse Document Frequency TF-IDF</i>	16
2.9	<i>Analisis Sentimen</i>	17
2.10	<i>Lexicon Based</i>	17
2.11	<i>Naïve Bayes</i>	19
2.12	<i>K-Nearest Neighbor</i>	20
2.13	<i>Confusion Matrix</i>	20
2.14	<i>Word Could</i>	22
2.15	<i>Media Sosial</i>	23
2.16	<i>X</i>	24
2.17	<i>Pyhton</i>	24
2.18	<i>Google Colab</i>	25
BAB III METODOLOGI PENELITIAN		26
3.1	<i>Waktu dan Tempat Penelitian</i>	26
3.2	<i>Kerangka Penelitian</i>	26
3.3	<i>Metode Pengumpulan Data</i>	27
3.4	<i>Data Mining</i>	31
3.4	<i>Kebutuhan Perangkat Lunak</i>	33
BAB VI HASIL DAN PEMBAHASAN		35
4.1	<i>Data Selection</i>	35
4.1.1	<i>Crawling Data</i>	35
4.2	<i>Pelabelan</i>	38
4.3	<i>Preprocessing</i>	40
4.3.1	<i>Case Folding</i>	40
4.3.2	<i>Tokenizing</i>	41
4.3.3	<i>Filtering/Stopword Removing</i>	42
4.3.4	<i>Stemming</i>	43

4.4	<i>Transformation</i>	44
4.5	<i>Data Mining</i>	59
4.5.1	Implementasi <i>Naïve Bayes</i>	59
4.5.2	Implementasi <i>K-Nearest Neighbor</i>	67
4.5	<i>Evaluation</i>	70
4.6.1	<i>Evaluation Naïve Bayes</i>	70
4.6.2	<i>Evaluation K-Nearest Neighbor</i>	73
4.6	<i>Knowledge Presentation</i>	75
4.7	Pembahasan.....	78
BAB V PENUTUP		81
5.1	Kesimpulan	81
5.2	Saran.....	81
DAFTAR PUSTAKA		83
LAMPIRAN		87



DAFTAR TABEL

Tabel 2.1 Studi Literatur	7
Tabel 2.2 Confusion Matrix	21
Tabel 3.1 Kebutuhan Perangkat Lunak	33
Tabel 4.1 Hasil Proses Crawling Data.....	37
Tabel 4.2 Hasil Pelabelan Sumber:	39
Tabel 4.3 Contoh Hasil <i>Case Folding</i>	40
Tabel 4.4 Contoh Hasil <i>Tokenizing</i>	41
Tabel 4.5 Contoh Hasil <i>Filtering</i>	42
Tabel 4.6 Contoh Hasil <i>Stemming</i>	43
Tabel 4.7 Tahapan Word Vector Manual	47
Tabel 4.8 Perhitungan TF dan DF	50
Tabel 4.9 Perhitungan TF-IDF	51
Tabel 4.10 Perhitungan Bobot.....	54
Tabel 4.11 Sampel Data Train	59
Tabel 4.12 Sampel Data Test.....	60
Tabel 4.13 Penentuan Prediksi dan Aktual.....	60
Tabel 4.14 Prior Probability	62
Tabel 4.15 Sampel data Test.....	64
Tabel 4.16 Sample Data Test.....	66
Tabel 4.18 Sampel Data Train.....	67
Tabel 4.19 Sampel Data Test.....	68
Tabel 4.20 Data Test “Puasa mati lampu jam kerja”	68
Tabel 4.21 Hasil Jarak	69
Tabel 4.23 Perbandingan Confusion Matrix Algoritma Naïve Bayes dan K-Nearest Neighbor.....	79

DAFTAR GAMBAR

Gambar 2.1 Tahapan KDD.....	14
Gambar 3.1 Tahapan Kerangka Penelitian.....	26
Gambar 3.2 Scrip Crawling Data.....	28
Gambar 3.3 Case Folding	29
Gambar 3.4 Tahapan Tokenize.....	30
Gambar 3.5 Tahapan Filtering/Stopwords Removing.....	30
Gambar 3.6 Tahapan Stemming.....	30
Gambar 3.7 Tahapan TF-IDF	31
Gambar 3.8 Tahapan Naive Bayes.....	32
Gambar 3.9 Tahapan K-Nearest Neighbor.....	33
Gambar 4.1 Script Auth Token.....	36
Gambar 4.2 Script Install Library Pandas dan Node.js.....	37
Gambar 4.3 Script Proses Crawling Data	37
Gambar 4.4 Script Pelabelan Data	39
Gambar 4.5 Script Tahapan Case Flolding	41
Gambar 4.6 Script Tahapan <i>Tokenizing</i>	42
Gambar 4.7 Script Tahapan Filtering.....	43
Gambar 4.8 <i>Script</i> Tahapan <i>Stemming</i>	44
Gambar 4.9 Script Tahapan Mengubah Data Menjadi Numpy Array.....	44
Gambar 4.10 Hasil Tahapan Merubah Data Menjadi Numpy Array.....	44
Gambar 4.11 Tahapan Splitting Data dan Hasil Splitting Data.....	45
Gambar 4.12 Script Tahapan Pembobotan TF-IDF	45
Gambar 4.13 Hasil Pembobotan Kata.....	46
Gambar 4.14 Syntax Tahapan Pembobotan TF IDF	58
Gambar 4.15 Hasil Syntax Pembobotan Kata.....	59
Gambar 4.16 Script Tahapan Modeling Naive Bayes	67
Gambar 4.17 Script Tahapan Modeling K-Nearest Neighbor.....	70
Gambar 4.18 Hasil Confusion Matrix Algoritma Naïve Bayes	71
Gambar 4.19 Script Menampilkan Confusion Matrix Naïve Bayes	72
Gambar 4.20 Hasil Confusion Matrix Naïve Bayes.....	73
Gambar 4.21 Hasil Confusion Matrix Algoritma K-Nearest Neighbor	73

Gambar 4.22 Script Menampilkan Confusion Matrix K-Nearest Neighbor	74
Gambar 4.23 Hasil Confusion Matrix K-Nearest Neighbor	75
Gambar 4.24 Wordcloud Positive	76
Gambar 4.25 Wordcloud Negative.....	77
Gambar 4.26 Wordcloud Neutral	78



DAFTAR LAMPIRAN

Plagiarisme

Biodata Mahasiswa

Kartu Konsultasi Bimbingan

Kartu Bimbingan Skripsi Dosen Pembimbing I

Kartu Bimbingan Skripsi Dosen Pembimbing II

Rekomendasi Dari Pembimbing I Untuk Tugas Akhir

Rekomendasi Dari Pembimbing II Untuk Tugas Akhir

Dataset

