

**ANALISIS SENTIMEN NETIZEN TERHADAP KEBIJAKAN
KENAIKAN PPN 12% PADA MEDIA SOSIAL TWITTER
MENGUNAKAN METODE *K-NEAREST NEIGHBOR***

SKRIPSI SARJANA

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem
Informasi dari Fakultas Teknologi Komunikasi dan Informatika



Oleh :

Haidar Ali

217006516045

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL**

2025

**ANALISIS SENTIMEN NETIZEN TERHADAP KEBIJAKAN
KENAIKAN PPN 12% PADA MEDIA SOSIAL TWITTER
MENGUNAKAN METODE *K-NEAREST NEIGHBOR***

SKRIPSI SARJANA SISTEM INFORMASI



Oleh :

Haidar Ali

217006516045

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL
2025**

HALAMAN PENGESAHAN

TUGAS AKHIR

ANALISIS SENTIMEN NETIZEN TERHADAP KEBIJAKAN KENAIKAN
PPN 12% PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE K-



Dosen Pembimbing

A handwritten signature in black ink, which appears to read 'Asrul Sani'.

Ir. Asrul Sani, M.Kom., M.T., Ph.D.

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

ANALISIS SENTIMEN NETIZEN TERHADAP KEBIJAKAN KENAikan
PPN 12% PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE K-
NEAREST NEIGHBOR

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 1 September 2025



Haidar Ali

217006516045

LEMBAR PERSETUJUAN REVIEW AKHIR

Tugas Akhir dengan judul :

**ANALISIS SENTIMEN NETIZEN TERHADAP
KEBIJAKAN KENAIKAN PPN 12% PADA
MEDIA SOSIAL TWITTER MENGGUNAKAN
METODE K-NEAREST NEIGHBOR**

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Review Akhir Semester Genap 2024-2025 pada tanggal 27 Agustus Tahun 2025

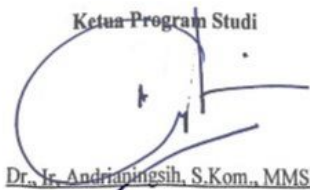
Dosen Pembimbing



Ir. Asrul Sani, M.Kom., M.T., Ph.D.

NIDN. 050023020

Ketua Program Studi



Dr. Ir. Andriyansih, S.Kom., MMSI

NIDN. 0303097902

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Haidar Ali
NPM : 217006516045
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Sistem Informasi
Tanggal Sidang : 27 Agustus 2025

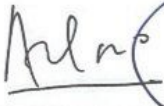
JUDUL DALAM BAHASA INDONESIA :

ANALISIS SENTIMEN NETIZEN TERHADAP KEBIJAKAN KENAIKAN PPN
12% PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE *K-NEAREST
NEIGHBOR*

JUDUL DALAM BAHASA INGGRIS :

ANALYSIS OF NETIZEN SENTIMENT TOWARDS THE 12% VAT INCREASE
POLICY ON TWITTER SOCIAL MEDIA USING THE *K-NEAREST NEIGHBOR*

TANDA TANGAN DAN TANGGAL

Dosen Pembimbing	Ka. Prodi	Mahasiswa
TGL : 27-9-2025	TGL : 27/9/2025	TGL : 2 September 2025
		

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul **“Analisis Sentimen Netizen terhadap Kebijakan Kenaikan PPN 12% pada Media Sosial Twitter Menggunakan Metode *K-Nearest Neighbor*”** dengan baik dan tepat waktu.

Skripsi ini disusun untuk memenuhi salah satu syarat guna memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional. Penulis menyadari bahwa penyusunan skripsi ini tidak akan terselesaikan tanpa adanya bantuan, dukungan, dan doa dari berbagai pihak. Pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT atas segala limpahan rahmat, kesehatan, dan kesempatan yang diberikan.
2. Bapak Hairuloh dan Ibu Mila Rosdianti selaku orang tua tercinta atas doa, kasih sayang, serta dukungan moral maupun material yang tiada henti.
3. Kakek, nenek, adik, dan mamang yang selalu memberikan doa, dukungan, serta semangat selama proses perkuliahan hingga penyusunan skripsi ini.
4. Dr. Agung Triayudi, S.Kom, M.Kom., selaku Dekan Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional.
5. Dr. Andrianingsih, S.Kom., MMSI., selaku Ketua Program Studi Sistem Informasi Universitas Nasional.
6. Ir. Asrul Sani, S.T., M.T, M.Kom., Ph.D., selaku dosen pembimbing yang dengan sabar telah memberikan bimbingan, arahan, dan motivasi selama proses penyusunan skripsi ini.
7. Seluruh dosen pengajar di Program Studi Sistem Informasi FTKI maupun dosen di Program Studi lain yang telah memberikan banyak ilmu.
8. Teman-teman FTKI angkatan 2021 yang selalu memberikan semangat, dukungan, dan kebersamaan selama menempuh perkuliahan hingga proses penyusunan skripsi.

9. Rekan-rekan “inPRESS” dan “kadang-kidding gimmick” yang senantiasa memberikan keceriaan, dukungan, serta motivasi selama menjalani masa kuliah hingga penyusunan skripsi.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi kesempurnaan karya ini. Semoga skripsi ini dapat memberikan manfaat bagi penulis sendiri maupun pihak lain yang membutuhkan.



Jakarta, 27 Agustus 2025

A handwritten signature in black ink, appearing to be "Haidar Ali", written over a light blue grid background.

Haidar Ali

ABSTRAK

Kebijakan kenaikan (PPN) sebesar 12% menimbulkan beragam reaksi masyarakat Twitter, sehingga penelitian ini dilakukan untuk menganalisis sentimen netizen terhadap isu tersebut. Data dikumpulkan melalui *scraping* sebanyak 3.636 *tweet* dan setelah tahap *preprocessing* serta pelabelan tersisa 3.316 data yang terbagi ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Metode yang digunakan adalah *K-Nearest Neighbor* (KNN) sebagai algoritma utama dengan perbandingan terhadap *Support Vector Machine* (SVM). Evaluasi kinerja dilakukan menggunakan *classification report* yang mencakup *accuracy*, *precision recall*, dan *f1-score*, dan hasilnya menunjukkan bahwa SVM memiliki performa yang kompetitif dibandingkan KNN.

Kata kunci : Analisis Sentimen, PPN 12%, Twitter, *K-Nearest Neighbor*, *Support Vector Machine*.



ABSTRACT

The government's policy of increasing the Value Added Tax (VAT) to 12% has sparked various public reactions on Twitter, prompting this research to analyze netizens' sentiments regarding the issue. A total of 3,636 tweets were collected through scraping, and after preprocessing and labeling, 3,316 data remained categorized into positive, negative, and neutral sentiments. The method applied in this study is K-Nearest Neighbor (KNN) as the primary algorithm, compared with Support Vector Machine (SVM). Model performance was evaluated using a classification report covering accuracy, precision, recall, and f1-score, with the results showing that SVM achieved competitive performance compared to KNN..

Kata kunci : *Sentiment Analysis, VAT 12%, Twitter, K-Nearest Neighbor, Support Vector Machine.*



DAFTAR ISI

KATA PENGANTAR.....	ii
ABSTRAK	iv
ABSTRACT	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR.....	ix
DAFTAR TABEL	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan Penelitian.....	4
1.5 Kontribusi	4
BAB II TINJAUAN PUSTAKA.....	6
2.1 Studi Literatur.....	6
2.2 Landasan Teori	12
2.2.1 Pajak Pertambahan Nilai (PPN).....	12
2.2.2 Twitter.....	13
2.2.3 Analisis Sentimen	13
2.2.4 <i>Machine learning</i>	14
2.2.5 <i>Text Preprocessing</i>	14
2.2.6 <i>K-Nearest Neighbor</i>	15
2.2.7 Evaluasi Kinerja Model	16
2.2.8 <i>Confusion matrix</i>	16
2.2.9 Python	16
2.2.10 Google Colab	17

2.2.11 Streamlit.....	17
BAB III METODOLOGI PENELITIAN	19
3.1 Lokasi Penelitian dan Waktu Penelitian.....	19
3.1.1 Lokasi Penelitian.....	19
3.2 Waktu Penelitian	19
3.3 Metode Penelitian.....	19
3.4 Teknik Pengumpulan Data	19
3.5 Fokus Penelitian	20
3.6 Desain Penelitian.....	20
3.6.1 Desain Penelitian Algoritma KNN	22
3.7 Metode Evaluasi Hasil.....	23
3.7.1 Confusion Matrix.....	23
BAB IV HASIL DAN PEMBAHASAN	25
4.1 Pengumpulan Data	25
4.2 <i>Preprocessing</i> Data.....	26
4.2.1 <i>Data Cleaning</i>	26
4.2.2 Normalisasi	27
4.2.3 <i>Tokenizing</i>	29
4.2.4 <i>Stopword Removal</i>	30
4.2.5 <i>Stemming</i>	30
4.3 <i>Exploratory Data Analysis (EDA)</i>	32
4.4 <i>Data Preparation</i>	37
4.4.1 <i>Translating</i>	37
4.4.2 <i>Labeling</i>	38
4.4.3 <i>Splitting Data dan Weighting (TF-IDF)</i>	39
4.5 Implementasi Algoritma <i>K-Nearest Neighbor</i>	41
4.6 Pemodelan Dengan Algoritma <i>K-Nearest Neighbor</i>	45

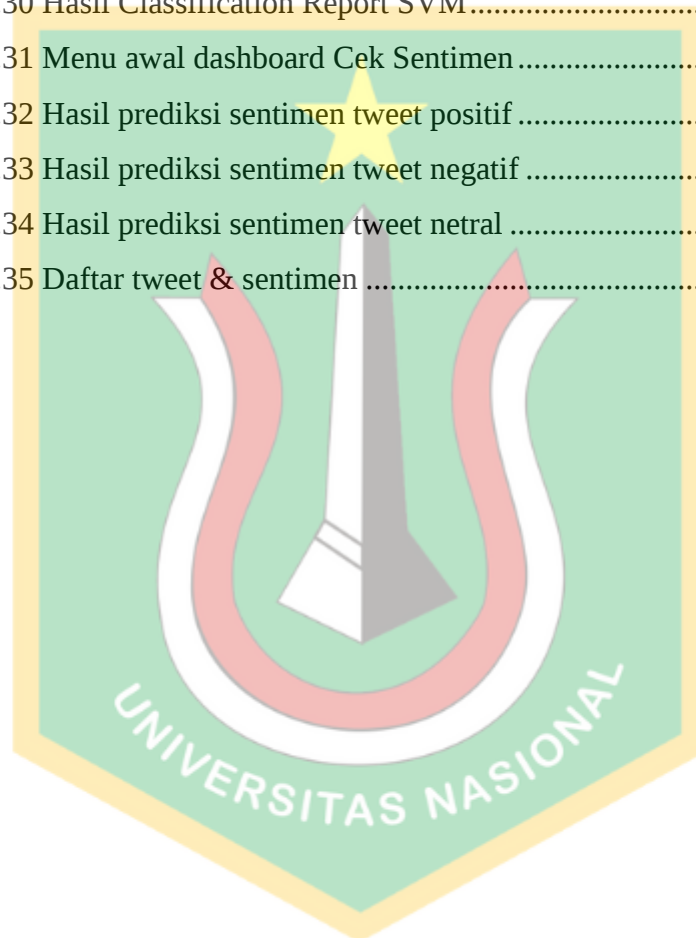
4.7 <i>Confusion Matrix</i>	46
4.7.1 Perhitungan Manual <i>Confusion Matrix</i>	46
4.7.2 Visualisasi <i>Confusion Matrix</i>	47
4.8 Perbandingan Hasil Algoritma KNN dan <i>Support Vector Machine</i>	48
4.9 Deployment	48
BAB V KESIMPULAN DAN SARAN	52
5.1 Kesimpulan.....	52
5.2 Saran.....	52
DAFTAR PUSTAKA	54



DAFTAR GAMBAR

Gambar 1.1 Penerimaan Perpajakan, Penerimaan PPN, dan Pertumbuhan PPN 2010-2021	1
Gambar 2.1 Logo X atau Twitter	13
Gambar 2.2 Logo Python	17
Gambar 2.3 Logo Google Colab	17
Gambar 3.1 Desain Penelitian.....	20
Gambar 3.2 Flowchart KNN	23
Gambar 4.1 Scrapping Data	25
Gambar 4.2 Hasil Scrapping	26
Gambar 4.3 Data Cleaning	27
Gambar 4.4 Normalisasi.....	28
Gambar 4.5 Hasil Normalisasi	28
Gambar 4.6 Tokenizing.....	29
Gambar 4.7 Hasil Tokenizing	29
Gambar 4.8 Stopword removal	30
Gambar 4.9 Hasil Stopword removal.....	30
Gambar 4.10 Stemming	31
Gambar 4.11 Hasil Stemming Data	31
Gambar 4.12 Wordcloud pada data tweet.....	32
Gambar 4.13 Wordcloud pada data berlabel positif.....	32
Gambar 4.14 Wordcloud pada data berlabel negatif.....	33
Gambar 4.15 Wordcloud pada data berlabel netral.....	33
Gambar 4.16 Bar chart 10 kata teratas setelah preprocessing.....	34
Gambar 4.17 Bar chart 10 kata teratas pada data berlabel positif.....	34
Gambar 4.18 Bar chart 10 kata teratas pada data berlabel negatif.....	35
Gambar 4.19 Bar chart 10 kata teratas pada data berlabel netral.....	35
Gambar 4.20 Bar chart 10 jumlah distribusi sentimen.....	36
Gambar 4.21 Translating.....	37
Gambar 4.22 Hasil Translating	38
Gambar 4.23 Labeling.....	38

Gambar 4.24 Hasil Labeling	39
Gambar 4.25 Splitting data dan weighting.....	40
Gambar 4.26 Hasil Pembagian Data Latih dan Data Uji	41
Gambar 4.27 Hasil Classification Report K-Nearest Neighbor	45
Gambar 4.28 Confusion matrix K-Nearest Neighbor	47
Gambar 4.29 Hasil Classification Report KNN	48
Gambar 4.30 Hasil Classification Report SVM	48
Gambar 4.31 Menu awal dashboard Cek Sentimen	49
Gambar 4.32 Hasil prediksi sentimen tweet positif	49
Gambar 4.33 Hasil prediksi sentimen tweet negatif	50
Gambar 4.34 Hasil prediksi sentimen tweet netral	51
Gambar 4.35 Daftar tweet & sentimen	51



DAFTAR TABEL

Tabel 3.1 Waktu Penelitian	19
Tabel 3. 2 Confusion matrix.....	24
Tabel 4.1 Sample data tweet	41
Tabel 4.2 Representasi Numerik pada tweet positif.....	42
Tabel 4.3 Representasi Numerik pada tweet negatif.....	42
Tabel 4.4 Representasi Numerik pada tweet netral.....	42
Tabel 4.5 Representasi Numerik pada tweet baru.....	43

