

**ANALISIS SENTIMEN PUBLIK TERHADAP KEBIJAKAN PPN 12% DI
INDONESIA MENGGUNAKAN REPRESENTASI TF-IDF DAN
ALGORITMA K-NEAREST NEIGHBOR (KNN)**

SKRIPSI



Disusun Oleh:
Muhammad Syaifudin
210602002

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH GRESIK
2025**

KATA PENGANTAR

Penulis mengucapkan syukur kepada Allah SWT atas rahmat-Nya, sehingga skripsi yang berjudul “ANALISIS SENTIMEN PUBLIK TERHADAP KEBIJAKAN PPN 12% DI INDONESIA MENGGUNAKAN REPRESENTASI TF-IDF DAN ALGORITMA K-NEAREST NEIGHBOR (KNN)” dapat diselesaikan dengan baik. Skripsi ini dibuat sebagai salah satu syarat menyelesaikan studi Teknik Informatika di Universitas Muhammadiyah Gresik.

Dalam proses penyusunan skripsi ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, yang senantiasa memberikan kesehatan, kesabaran dan kekuatan kepada penulis.
2. Kedua orang tua dan keluarga yang selalu memberikan dukungan, baik moral maupun material.
3. Bapak Harunur Rosyid, S.T., M.Kom, dan Ibu Putri Aisyiyah Rakhma Devi, S.Pd., M.Kom, selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta motivasi selama penyusunan skripsi ini.
4. Seluruh rekan Teknik Informatika Angkatan 2021 yang telah mendukung dan menyemangati selama penyusunan tugas akhir ini.

Penulis menyadari skripsi ini masih belum sempurna, sehingga sangat terbuka terhadap kritik dan saran demi perbaikan. Semoga skripsi ini bermanfaat bagi penulis dan pembaca.

Gresik, 6 Januari 2025

Penulis

Muhammad Syaifudin

210602002

ABSTRAK

Kenaikan tarif Pajak Pertambahan Nilai (PPN) dari 11% menjadi 12% pada tahun 2025, sesuai dengan Undang-Undang Nomor 7 Tahun 2021, telah memicu perdebatan terkait dampaknya terhadap perekonomian, khususnya harga barang dan daya beli. Platform media sosial X menjadi salah satu saluran utama dalam menyebarkan opini publik mengenai kebijakan ini, namun tantangan utama terletak pada volume data besar dan variasi bahasa yang digunakan, yang sulit dianalisis secara manual. Penelitian ini bertujuan mengembangkan sistem berbasis Algoritma K-Nearest Neighbor (KNN) untuk menganalisis sentimen masyarakat terhadap kenaikan tarif PPN 12%. Hasil penelitian menunjukkan bahwa pemilihan nilai K memengaruhi kinerja model. Dengan nilai $K = 3$, model menghasilkan performa terbaik dengan akurasi 74,55%, precision 61%, recall 66%, dan F1-score 62%. Pada $K = 5$, akurasi turun menjadi 70,91% dengan precision 60%, recall 66%, dan F1-score 60%. Sementara pada $K = 7$, akurasi menurun lebih jauh menjadi 69,09%, dengan precision 59%, recall 65%, dan F1-score 58%. Temuan ini menegaskan pentingnya pemilihan nilai K yang tepat untuk menghasilkan klasifikasi sentimen yang akurat dan seimbang. Hasil penelitian ini diharapkan dapat membantu pemerintah dalam merespons kebijakan secara lebih tepat dan efektif.

Kata Kunci: PPN, Kenaikan Tarif, Sentimen Masyarakat, X, Algoritma K-Nearest Neighbor (KNN)

ABSTRACT

The increase in the Value Added Tax (VAT) rate from 11% to 12% in 2025, as stipulated by Law No. 7 of 2021, has sparked debates regarding its impact on the economy, particularly on the prices of goods and purchasing power. The X social media platform has become a primary channel for disseminating public opinions on this policy. However, the main challenge lies in the large volume of data and the diversity of language used, which makes manual analysis difficult. This study aims to develop a system based on the K-Nearest Neighbor (KNN) algorithm to analyze public sentiment regarding the VAT increase to 12%. The results show that the choice of the K value significantly affects the model's performance. With $K = 3$, the model achieved the best performance, with an accuracy of 74.55%, precision of 61%, recall of 66%, and F1-score of 62%. With $K = 5$, the accuracy decreased to 70.91%, with precision of 60%, recall of 66%, and F1-score of 60%. At $K = 7$, the accuracy further decreased to 69.09%, with precision of 59%, recall of 65%, and F1-score of 58%. These findings highlight the importance of selecting the optimal K value to achieve accurate and balanced sentiment classification. The results of this study are expected to help the government respond to the policy more effectively and appropriately.

Keywords: VAT, Tax Increase, Public Sentiment, X, K-Nearest Neighbor (KNN)

DAFTAR ISI

HALAMAN JUDUL.....	i
LEMBAR PERSETUJUAN SEMINAR PROPOSAL SKRIPSI	ii
LEMBAR PENGESAHAN SEMINAR PROPOSAL SKRIPSI	iii
LEMBAR PERSETUJUAN SKRIPSI.....	iv
LEMBAR PENGESAHAN SKRIPSI	v
LEMBAR PERNYATAAN KEASLIAN SKRIPSI	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR PERSAMAAN	xiii
DAFTAR LAMPIRAN.....	xiv
ABSTRAK	xv
ABSTRACT.....	xvi
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah.....	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Metodologi Penelitian	4
1.7 Sistematika Penulisan.....	5
BAB 2 LANDASAN TEORI.....	7
2.1 Pajak	7
2.1.1 Pengertian Pajak	7
2.1.2 Ciri – Ciri Pajak	7

2.1.3 Fungsi Pajak.....	8
2.2 Pajak Pertambahan Nilai (PPN)	8
2.3 Analisis Sentimen.....	9
2.4 Text Mining.....	9
2.5 X	10
2.6 <i>Preprocessing</i> Data	11
2.7 <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF).....	11
2.8 <i>K-Nearest Neighbor</i> (KNN)	13
2.9 <i>Confusion Matrix</i>	14
2.10 Penelitian Terdahulu.....	16
BAB 3 ANALISIS DAN PERANCANGAN SISTEM	23
3.1 Analisis Sistem.....	23
3.2 Hasil Analisis	23
3.2.1 <i>Preprocessing</i>	28
3.3 Representasi Model	31
3.3.1 Ekstraksi Fitur.....	31
3.3.2 <i>K-Nearest Neighbor</i> (KNN).....	36
3.4 Perancangan Sistem.....	38
3.4.1 Diagram Konteks	38
3.4.2 Diagram Jenjang Sistem	39
3.4.3 <i>Data Flow Diagram</i> (DFD).....	41
3.5 Perancangan Basis Data	43
3.6 Perancangan Antarmuka Sistem.....	48
3.6.1 Halaman Login	48
3.6.2 Halaman Dashboard.....	49
3.6.3 Halaman Input Data Latih.....	49
3.6.4 Halaman Input Data Uji.....	50
3.6.5 Halaman Hasil Klasifikasi	51
3.7 Skenario Pengujian.....	52
3.8 Spesifikasi Pengembangan Sistem	53
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM	54
4.1 Implementasi Sistem	54

4.1.1 Proses <i>Preprocessing</i>	54
4.1.2 Perhitungan Ekstraksi Fitur TF-IDF	57
4.1.3 Proses Klasifikasi Dengan <i>K-Nearest Neighbor</i> (KNN)	58
4.2 Pengujian Sistem	58
4.2.1 Halaman Login	59
4.2.2 Halaman Dashboard.....	59
4.2.3 Halaman Input Data Latih.....	60
4.2.4 Halaman Input Data Uji	60
4.2.5 Halaman Hasil Klasifikasi	61
4.3 Analisa Hasil Pengujian Sistem	62
4.3.1 Skenario Pengujian	62
BAB 5 Kesimpulan Dan Saran	68
5.1 Kesimpulan.....	68
5.2 Saran.....	68
DAFTAR PUSTAKA	69
LAMPIRAN.....	74

DAFTAR TABEL

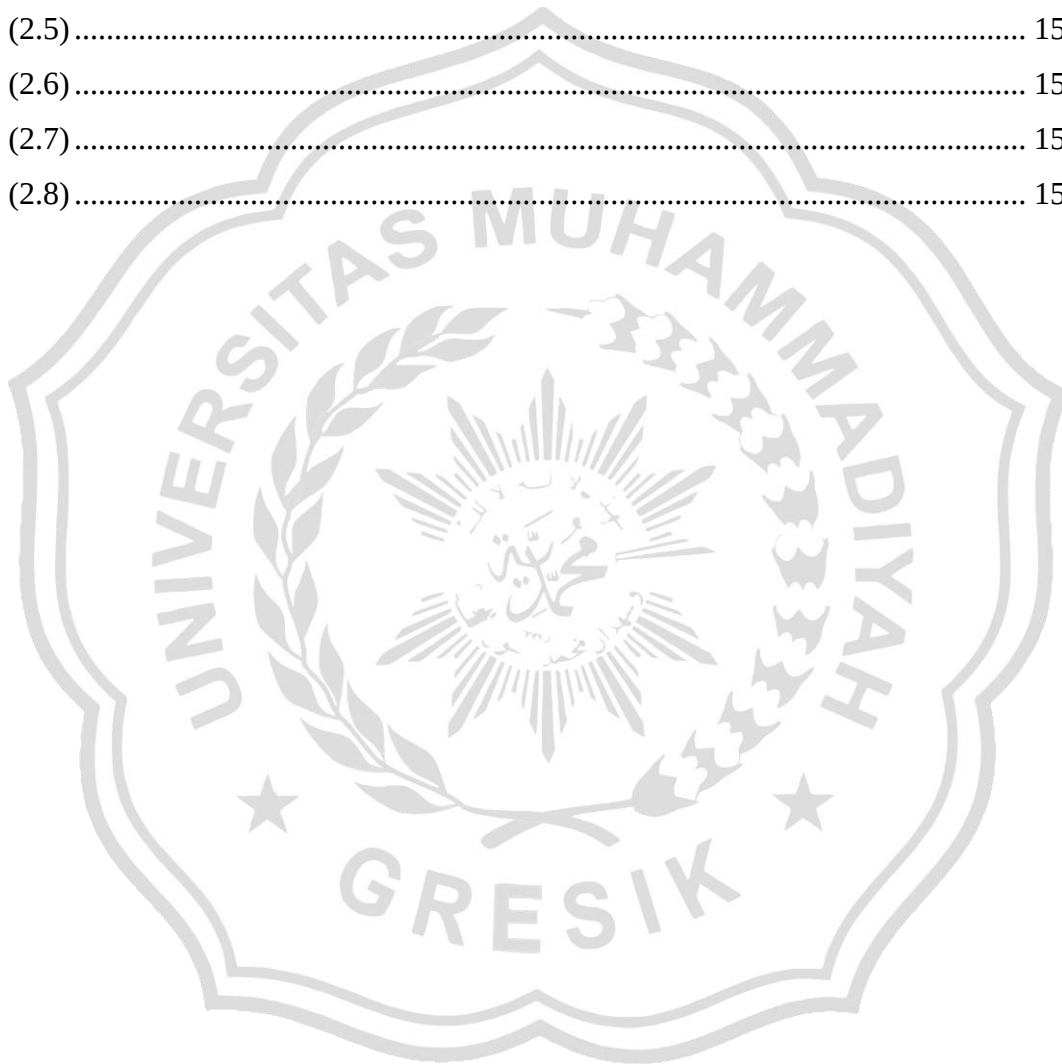
Table 2.1 Confusion Matrix	14
Table 2.2 Review Artikel	17
Table 3.1 Contoh Data Hasil Crawling X	26
Table 3.2 Contoh Data Tweet.....	27
Table 3.3 Contoh Hasil Proses Cleaning.....	29
Table 3.4 Contoh Hasil Proses Case Folding.....	29
Table 3.5 Contoh Hasil Proses Normalisasi.....	30
Table 3.6 Contoh Hasil Proses Tokenizing.....	30
Table 3.7 Contoh Hasil Proses Stopword Removal	31
Table 3.8 Contoh Hasil Proses Stemming.....	31
Table 3.9 Hasil Preprocessing	32
Table 3.10 Contoh Proses Hitung TF	32
Table 3.11 Contoh Proses Hitung IDF	34
Table 3.12 Contoh Proses Hitung TF-IDF	35
Table 3.13 Contoh Hasil Vektor TF-IDF	36
Table 3.14 Contoh Hasil Perhitungan Jarak.....	37
Table 3.15 Contoh Urutan Data Berdasarkan Jarak Terdekat.....	37
Table 3.16 Hasil Perhitungan KNN	38
Table 3.17 Tabel Users.....	44
Table 3.18 Tabel Dataset.....	45
Table 3.19 Tabel Preprocessing	45
Table 3.20 Tabel tfidf_results.....	46
Table 3.21 Tabel knn_results	46
Table 3.22 confusion_matrix.....	47
Table 3.23 confusion_macro	48

DAFTAR GAMBAR

Gambar 3.1 Alur Sistem Analisis Sentimen dengan Algoritma KNN	24
Gambar 3.2 Alur Algoritma KNN.....	25
Gambar 3.3 Distribusi Sentimen Terhadap Kebijakan PPN 12%	28
Gambar 3.4 Diagram Konteks Sistem Analisi Sentimen Menggunakan KNN...	39
Gambar 3.5 Diagram Berjenjang Sistem Analisis sentimen PPN 12% Menggunakan Algoritma KNN.....	40
Gambar 3.6 DFD Level 1 Sistem Analisis Sentimen PPN 12% Menggunakan Algoritma KNN.....	41
Gambar 3.7 ERD Analisis Sentimen PPN 12%	43
Gambar 3.8 Halaman Login	49
Gambar 3.9 Halaman Dashboard.....	49
Gambar 3.10 Halaman Input Data Latih	50
Gambar 3.11 Halaman Input Data Uji.....	51
Gambar 3.12 Halaman Hasil Klasifikasi.....	51
Gambar 4.1 Halaman Login	59
Gambar 4.2 Halaman Dashboard.....	60
Gambar 4.3 Halaman Input Data Latih	60
Gambar 4.4 Halaman Input Data Uji.....	61
Gambar 4.5 Halaman Preprocessing Teks.....	61
Gambar 4.6 Halaman Ekstraksi Fitur TF-IDF.....	62
Gambar 4.7 Halaman Klasifikasi KNN.....	62

DAFTAR PERSAMAAN

(2.1)	12
(2.2)	12
(2.3)	13
(2.4)	14
(2.5)	15
(2.6)	15
(2.7)	15
(2.8)	15



DAFTAR LAMPIRAN

Lampiran 1. Dataset Tweet PPN 12 Persen	74
Lampiran 2. Source Code Dashboard	116
Lampiran 3. Source code Input Data Latih	118
Lampiran 4. Source Code Input Data Uji.....	123
Lampiran 5. Source Code Analisa.....	127

