



**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
RENCANA KENAIKAN PPN 12% DI INDONESIA DENGAN
NAIVE BAYES, KNN, DAN RANDOM FOREST**

LAPORAN TUGAS AKHIR



**ASAFITA DWI PUTRI
41820120038**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA**

2025



**ANALISIS SENTIMEN PENGGUNA TWITTER TERHADAP
RENCANA KENAIKAN PPN 12% DI INDONESIA DENGAN
NAIVE BAYES, KNN, DAN RANDOM FOREST**

LAPORAN TUGAS AKHIR



ASAFITA DWI PUTRI

41820120038

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA**

2025

HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Asafita Dwi Putri

NIM : 41820120038

Program Studi : Sistem Informasi

Judul Laporan Skripsi : Analisis Sentimen Pengguna Twitter Terhadap Rencana Kenaikan PPN 12% di Indonesia dengan Naive Bayes, KNN, dan Random Forest

Menyatakan bahwa Laporan Aplikatif/Tugas Akhir/Jurnal/Media Ilmiah ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 28 Juli 2025



Asafita Dwi Putri

LEMBAR PENGESAHAN

Nama Mahasiswa : Asafita Dwi Putri.
NIM : 41820120038
Judul Tugas Akhir : Analisis Sentimen Pengguna Twitter Terhadap Rencana Kenaikan PPN 12% di Indonesia dengan Naive Bayes, KNN, dan Random Forest

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 28 Juli 2025

Menyetujui:

Pembimbing : Dr. Misbahul Fajri, ST., MTI.
NIDN : 0306077203
Ketua Penguji : Lukman Hakim, ST, M.Kom
NIDN : 0327107701
Penguji 1 : Misni, S.Kom, M.Kom
NIDN : 0413046802
Penguji 2 : Ifan Prihandi, S.Kom, M.Kom
NIDN : 0313098901

()
()
()
()

Mengetahui,


Dr. Bambang Jokonowo, S.Si., M.T.I
Dekan Fakultas Ilmu Komputer


Wawan Gunawan, S.Kom, MT., M.Kom.
Ka.Prodi Sistem Informasi

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan Yang Maha Esa, karena atas berkat dan rahmat-Nya, saya dapat menyelesaikan Laporan Tugas Akhir ini dalam rangka memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer pada Fakultas Ilmu Komputer Universitas Mercu Buana. Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan Laporan Tugas Akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Dr. Misbahul Fajri, ST., MTI. sebagai Dosen Pembimbing Tugas Akhir yang telah membimbing dalam menyelesaikan laporan ini.
2. Dr. Bambang Jokonowo, S.Si., M.T.I sebagai Dekan Fakultas Ilmu Komputer Universitas Mercu Buana.
3. Wawan Gunawan, S.Kom, MT., M.Kom. sebagai Ka.Prodi Sistem Informasi Universitas Mercu Buana.
4. Kedua orang tua yang selama ini selalu menyemangati.

Akhir kata, saya berharap Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Tugas Akhir ini membawa manfaat bagi pengembangan ilmu.

Jakarta, 28 Juli 2025

Asafita Dwi Putri

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Asafita Dwi Putri
NIM : 41820120038
Program Studi : Sistem Informasi
Judul Laporan Skripsi : Analisis Sentimen Pengguna Twitter Terhadap Rencana Kenaikan PPN 12% di Indonesia dengan Naive Bayes, KNN, dan Random Forest

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana Hak Bebas Royalti Non-Eksklusif (Non-exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 28 Juli 2025

Yang menyatakan,



(Asafita Dwi Putri)

ABSTRAK

Nama Mahasiswa : Asafita Dwi Putri
NIM : 41820120038
Pembimbing TA : Dr. Misbahul Fajri, ST., MTI.
Judul : Analisis Sentimen Pengguna Twitter Terhadap Rencana Kenaikan PPN 12% di Indonesia dengan Naive Bayes, KNN, dan Random Forest

Penelitian ini membahas analisis sentimen pengguna Twitter terkait rencana kenaikan Pajak Pertambahan Nilai (PPN) 12% di Indonesia dengan metode machine learning. Sebanyak 2.985 tweet dikumpulkan menggunakan kata kunci terkait, kemudian diproses melalui tahapan preprocessing dan pelabelan otomatis memakai RoBERTa. Distribusi data awal sangat tidak seimbang, didominasi sentimen netral (53%), negatif (35%), dan positif (12%). Tiga algoritma yakni Naive Bayes, K-Nearest Neighbor (KNN), dan Random Forest diuji sebelum dan sesudah penyeimbangan data dengan SMOTE. Setelah dilakukan penyeimbangan data menggunakan SMOTE, model Naive Bayes mencapai akurasi sebesar 88%, precision 88%, recall 88%, dan f1-score 87%. Model KNN memperoleh akurasi 65%, precision 75%, recall 65%, dan f1-score 56%. Sementara itu, model Random Forest menghasilkan hasil terbaik dengan akurasi 89%, precision 90%, recall 89%, dan f1-score 89%.

Kata kunci : Analisis Sentimen, Text Mining, Media Sosial, RoBERTa, SMOTE

ABSTRACT

Name : Asafita Dwi Putri
Student Number : 41820120038
Counsellor : Dr. Misbahul Fajri, ST., MTI.
Title : Sentiment Analysis of Twitter Users Regarding the Planned
12% VAT Increase in Indonesia Using Naive Bayes, KNN, and
Random Forest

This study discusses sentiment analysis of Twitter users regarding the planned 12% Value Added Tax (VAT) increase in Indonesia using machine learning methods. A total of 2,985 tweets were collected using relevant keywords, then processed through a series of preprocessing steps and automatically labeled using RoBERTa. The initial data distribution was highly imbalanced, dominated by neutral sentiment (53%), negative (35%), and positive (12%). Three algorithms—Naive Bayes, K-Nearest Neighbor (KNN), and Random Forest—were tested before and after data balancing with SMOTE. After balancing the data using SMOTE, the Naive Bayes model achieved an accuracy of 88%, precision of 88%, recall of 88%, and an f1-score of 87%. The KNN model obtained an accuracy of 65%, precision of 75%, recall of 65%, and an f1-score of 56%. Meanwhile, the Random Forest model produced the best results, with an accuracy of 89%, precision of 90%, recall of 89%, and an f1-score of 89%.

Keywords : Sentiment Analysis, Text Mining, Social Media, RoBERTa, SMOTE

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENYATAAN KARYA SENDIRI	ii
LEMBAR PENGESAHAN.....	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS.....	v
ABSTRAK.....	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Permasalahan.....	2
1.3 Tujuan Penelitian.....	2
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Teori/Konsep Terkait	5
2.1.1 Machine Learning	5
2.1.2 Analisis sentimen	5
2.1.3 Text Mining.....	5
2.1.4 Cross Industry Standard Process for Data Mining (CRISP-DM).....	6
2.1.5 Pemahaman Bisnis	6
2.1.6 Pemahaman Data	7
2.1.7 Persiapan Data	7
2.1.8 Pemodelan	8
2.1.9 Evaluasi	12
2.1.10 Penerapan.....	13
2.2 Penelitian Terdahulu	14
2.3 Analisis Literature Review.....	24

BAB III METODE PENELITIAN	24
3.1 Deskripsi Sumber Data	25
3.2 Teknik Pengumpulan Data	25
3.3 Diagram Alir Penelitian	25
3.4 Jadwal Penelitian	30
BAB IV HASIL DAN PEMBAHASAN.....	24
4.1 Pembahasan.....	31
4.1.1 Pengumpulan Data	31
4.1.2 Preprocessing Data.....	32
4.1.3 Pelabelan Data	48
4.2 Evaluasi Model.....	51
4.2.1 Evaluasi Model Tanpa Penanganan Ketidakseimbangan Data (Tanpa SMOTE).....	51
4.2.2 Evaluasi Model dengan Penanganan Ketidakseimbangan Data (Dengan SMOTE).....	53
BAB V KESIMPULAN DAN SARAN	57
5.1 Kesimpulan	57
5.2 Saran	58
DAFTAR PUSTAKA	60
LAMPIRAN	63

UNIVERSITAS
MERCU BUANA

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	14
Tabel 3. 1 Jadwal Penelitian	30
Tabel 4. 1 Hasil Preprocessing Data	38
Tabel 4. 2 Perbandingan Akurasi Model Sebelum dan Sesudah SMOTE	54



DAFTAR GAMBAR

Gambar 2. 1 Alur Metodologi CRISP-DM	6
Gambar 2. 2 Alur algoritma Naive Bayes	10
Gambar 2. 3 Alur algoritma K-Nearest Neighbor (KNN).....	11
Gambar 2. 4 Alur Algoritma Random Forest	12
 Gambar 3. 1 Diagram Alir Penelitian	 25
 Gambar 4. 1 Proses Crawling data Twitter 1.....	 31
Gambar 4. 2 Proses Crawling data Twitter 2.....	31
Gambar 4. 3 Data Mentah	32
Gambar 4. 4 Kolom data mentah.....	33
Gambar 4. 5 Tampilan 5 baris pertama data.....	33
Gambar 4. 6 Menghapus data duplikat	34
Gambar 4. 7 Fungsi cleaningText	35
Gambar 4. 8 Fungsi casefoldingText	35
Gambar 4. 9 Fungsi tokenizingText.....	35
Gambar 4. 10 Fungsi filteringText.....	36
Gambar 4. 11 Fungsi toSentence	36
Gambar 4. 12 Fungsi stemmingText	36
Gambar 4. 13 Fungsi fix_slangwords	37
Gambar 4. 14 Hasil Pelabelan menggunakan RoBERTa	48
Gambar 4. 15 WordCloud Sentimen Negatif	49
Gambar 4. 16 WordCloud Sentimen Netral	50
Gambar 4. 17 WordCloud Sentimen Positif	50
Gambar 4. 18 Hasil Evaluasi Model Tanpa Smote Algoritma Naive Bayes	51
Gambar 4. 19 Hasil Evaluasi Model Tanpa Smote Algoritma KNN	52
Gambar 4. 20 Hasil Evaluasi Model Tanpa Smote Algoritma Random Forest.....	52
Gambar 4. 21 Distribusi Label Sesudah Smote.....	53
Gambar 4. 22 Hasil Evaluasi Model dengan SMOTE algoritma Naive Bayes.....	54
Gambar 4. 23 Hasil Evaluasi Model dengan SMOTE algoritma KNN	54
Gambar 4. 24 Hasil Evaluasi Model dengan SMOTE algoritma Random Forest.....	54
Gambar 4. 25 Perbandingan Akurasi Model Sebelum dan Sesudah SMOTE	55

DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi Bimbingan	63
Lampiran 2 Curriculum Vitae	64
Lampiran 3 Sertifikat BNSP	65
Lampiran 4 Hasil Cek Uji Plagiasi	66

