

**ANALISIS SENTIMEN SISTEM PAJAK CORETAX PADA
MEDIA SOSIAL X DENGAN PERBANDINGAN METODE
NAIVE BAYES DAN *SUPPORT VECTOR MACHINE***

Tugas Akhir

Diajukan untuk memenuhi persyaratan mencapai derajat

Sarjana Sistem Informasi



Zelika Awine Windarti

NPM: 211711470

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ATMA JAYA YOGYAKARTA**

2025

HALAMAN PENGESAHAN

Tugas Akhir Berjudul

ANALISIS SENTIMEN SISTEM PAJAK CORETAX PADA MEDIA SOSIAL X DENGAN PERBANDINGAN
METODE NAIVE BAYES DAN SUPPORT VECTOR MACHINE

yang disusun oleh

Zelika Awine Windarti

211711470

dinyatakan telah memenuhi syarat pada tanggal 22 Agustus 2025

		Keterangan
Dosen Pembimbing 1	: Citra Yayu' Palangan, S.T., M.Sc.	Telah Menyetujui
Dosen Pembimbing 2	: Elisabeth Marsella, S.S., M.Li.	Telah Menyetujui
Tim Penguji		
Penguji 1	: Citra Yayu' Palangan, S.T., M.Sc.	Telah Menyetujui
Penguji 2	: Generosa Lukhayu Pritalia, S.T., M.Eng.	Telah Menyetujui
Penguji 3	: Dr. Fl. Spty Rahayu, S.T., M.Kom.	Telah Menyetujui

Yogyakarta, 22 Agustus 2025

Universitas Atma Jaya Yogyakarta

Teknologi Industri

Dekan

ttd.

Dr. Ir. Parama Kartika Dewa SP., S.T., M.T.

Dokumen ini merupakan dokumen resmi UAJY yang tidak memerlukan tanda tangan karena dihasilkan secara elektronik oleh Sistem Bimbingan UAJY. UAJY bertanggung jawab penuh atas informasi yang tertera di dalam dokumen ini

LEMBAR PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan di bawah ini:

Nama Lengkap : Zelika Awine Windarti
NPM : 211711470
Program Studi : Sistem Informasi
Fakultas : Teknologi Industri
Judul Penelitian : Analisis Sentimen Sistem Pajak Coretax pada Media Sosial X dengan Perbandingan Metode *Naive Bayes* dan *Support Vector Machine*.

Menyatakan dengan ini:

1. Skripsi ini adalah benar merupakan hasil karya sendiri dan tidak merupakan salinan sebagian atau keseluruhan dari karya orang lain.
2. Memberikan kepada Universitas Atma Jaya Yogyakarta, berupa Hak Bebas Royalti non eksklusif (*Non-Exclusive-Royalty-Free Right*) atas Penelitian ini, dan berhak menyimpan, mengelola dalam pangkalan data, mendistribusikan, serta menampilkan untuk kepentingan akademis, tanpa perlu meminta izin selama tetap mencantumkan nama penulis.
3. Bersedia menanggung secara pribadi segala bentuk tuntutan hukum yang mengikuti atas pelanggaran Hak Cipta dalam pembuatan skripsi ini.

Demikianlah pernyataan ini dibuat dan dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 06 Agustus 2025

Yang menyatakan,



Zelika Awine Windarti
211711470

PRAKATA

Puji dan syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi ini yang berjudul "*Analisis Sentimen pada Media Sosial X (Twitter) Terhadap Sistem Pajak Coretax Menggunakan Naive Bayes Classifier dan Support Vector Machine* " sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Teknologi Industri, Universitas Atma Jaya Yogyakarta.

Dalam proses penyusunan skripsi ini, penulis telah menghadapi berbagai tantangan dan menyadari bahwa tanpa bantuan, bimbingan, dan dukungan dari berbagai pihak, penyusunan skripsi ini tidak akan dapat diselesaikan dengan baik. Dari tahap awal hingga akhir, keberhasilan penyusunan skripsi ini tidak terlepas dari kontribusi dan peran banyak pihak yang telah memberikan arahan, bimbingan, motivasi, serta bantuan moril maupun materil. Oleh karena itu, dengan segala kerendahan hati, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Allah SWT, penulis memanjatkan puji dan syukur yang tak terhingga atas segala limpahan rahmat, nikmat, kekuatan, serta kemudahan yang diberikan selama proses penyusunan skripsi ini. Tanpa petunjuk dan pertolongan-Nya, segala usaha dan pencapaian ini tentu tidak akan mungkin terwujud.
2. Orang tua dan adik tercinta, yang senantiasa menjadi sumber kekuatan dan inspirasi. Terima kasih atas doa yang tidak pernah putus, kasih sayang yang tulus, serta dukungan moral dan material yang telah diberikan sepanjang perjalanan pendidikan ini.
3. Citra Yuyu' Palangan, S.T., M.Sc., selaku dosen pembimbing skripsi pertama, penulis menyampaikan rasa terima kasih yang sebesar-besarnya atas bimbingan, arahan, serta masukan yang sangat berarti selama proses penyusunan skripsi ini. Kesabaran dan dedikasi Ibu dalam membimbing telah memberikan kontribusi besar terhadap penyelesaian skripsi ini.
4. Elisabeth Marsella, S.S., M.Li., penulis mengucapkan terima kasih atas waktu, perhatian, serta saran dan koreksi yang telah diberikan, yang sangat membantu dalam memperbaiki dan menyempurnakan isi skripsi ini.

5. Seluruh dosen Program Studi Sistem Informasi yang telah menjadi pendidik sekaligus pembimbing selama masa perkuliahan. Terima kasih atas ilmu, wawasan, serta arahan dan dukungan yang diberikan, yang telah memberikan bekal berharga bagi penulis hingga berhasil menyelesaikan tahap penelitian ini.
6. Kepada sahabat-sahabat terbaik penulis, Wisni, Syahra, dan Elsa. Terima kasih atas setiap canda tawa, dukungan, dan semangat yang telah diberikan dan menjadi alasan mengapa setiap tantangan dalam penulisan skripsi ini terasa lebih ringan.
7. Kepada anggota grup WIB, terima kasih telah menjadi *support system* terbaik yang selalu menemani dan menyemangati penulis di setiap proses penulisan skripsi ini. Dan untuk seluruh teman-teman KKN Kelompok 44, terima kasih atas kenangan, pengalaman, dan kebersamaan yang tak terlupakan.
8. Teman-teman seperjuangan dan rekan satu angkatan, terima kasih atas semangat, kerja sama, dan kebersamaan yang telah terjalin selama masa kuliah hingga tahap penyusunan skripsi ini. Dukungan moral, saling berbagi pengalaman dalam menghadapi berbagai tantangan menjadi bagian penting dalam perjalanan akademik penulis.
9. Seluruh pihak yang tidak dapat disebutkan satu per satu, penulis mengucapkan terima kasih atas segala bantuan, dukungan, dan kontribusi yang telah diberikan, baik secara langsung maupun tidak langsung.

INTISARI

Media sosial X menjadi wadah bagi pengguna untuk menyuarakan pandangan mengenai isu-isu yang sedang hangat, seperti sistem pajak Coretax. Coretax merupakan inovasi teknologi informasi yang dirancang oleh DJP untuk mengintegrasikan semua aspek administrasi perpajakan di seluruh Indonesia. Data penelitian dikumpulkan dari platform X melalui teknik *scraping*, menghasilkan 7.632 *tweet* berbahasa Indonesia yang membahas Coretax selama periode Januari hingga April 2025. Tahapan penelitian meliputi *preprocessing data*, pelabelan sentimen menggunakan *Indonesia Sentiment Lexicon*, dan klasifikasi menggunakan *Naïve Bayes* dan *Support Vector Machine* (SVM). Hasil analisis menunjukkan bahwa mayoritas *tweet* mengandung sentimen negatif. Model SVM menghasilkan akurasi sebesar 90%, lebih tinggi dibandingkan *Naïve Bayes* yang memperoleh akurasi 87%. Selain itu, SVM lebih unggul dalam mengklasifikasikan sentimen negatif, positif dan netral.

Kata Kunci: X, Coretax, *Indonesia Sentiment Lexicon*, *Naïve Bayes*, *Support Vector Machine* (SVM).

ABSTRACT

Social media platform X serves as a space for users to express opinions on trending issues, including the Coretax system. Coretax is an information technology innovation developed by the Directorate General of Taxes (DJP) to integrate all aspects of tax administration across Indonesia. The research data were collected from platform X using scraping techniques, resulting in 7,632 Indonesian-language tweets discussing Coretax from January to April 2025. The research stages included data preprocessing, sentiment labeling using the Indonesia Sentiment Lexicon, and classification using Naïve Bayes and Support Vector Machine (SVM). The analysis results show that the majority of tweets contain negative sentiment. The SVM model achieved an accuracy of 90%, higher than Naïve Bayes which reached 87%. Furthermore, SVM performed better in classifying negative, positive and neutral sentiments..

Keywords: *X, Coretax, Indonesia Sentiment Lexicon, Naïve Bayes, Support Vector Machine (SVM).*

DAFTAR ISI

HALAMAN PENGESAHAN	ii
LEMBAR PERNYATAAN KEASLIAN PENELITIAN	iii
PRAKATA	iv
INTISARI.....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah.....	3
1.3. Pertanyaan Penelitian.....	3
1.4. Tujuan Penelitian	3
1.5. Batasan Masalah.....	3
1.6. Manfaat Penelitian	4
1.7. Bagan Keterkaitan.....	4
BAB II TINJAUAN PUSTAKA	5
2.1. Studi Sebelumnya	5
2.2. Dasar Teori.....	9
2.2.1 Analisis Sentimen	9
2.2.2 Media Sosial	10
2.2.3 X	10
2.2.4 Pajak.....	11
2.2.5 Coretax.....	12
2.2.6 <i>Naïve Bayes Classifier</i> (NBC)	13
2.2.7 <i>Support Vector Machine</i> (SVM)	14
2.2.8 <i>Confusion Matrix</i>	15
BAB III METODOLOGI PENELITIAN	16
3.1. Tahapan Penelitian	16

3.1.1	Identifikasi Masalah	17
3.1.2	Studi Literatur	17
3.1.3	Pengumpulan Data	17
3.1.4	<i>Preprocessing Data</i>	17
3.1.5	Pelabelan Data	20
3.1.6	Klasifikasi.....	20
3.1.7	Evaluasi Model.....	21
3.1.8	Interpretasi Hasil.....	21
3.1.9	Visualisasi Data	21
BAB IV HASIL DAN PEMBAHASAN		22
4.1	Pengumpulan Data	22
4.2	<i>Preprocessing Data</i>	23
4.3	Pelabelan Data.....	33
4.4	Klasifikasi.....	34
4.5	Evaluasi Model.....	35
4.6	Interpretasi Hasil.....	40
4.7	Visualisasi Data	43
BAB V KESIMPULAN DAN SARAN		49
5.1	Kesimpulan	49
5.2	Saran	49
5.3	Implikasi Manajerial.....	50
DAFTAR PUSTAKA		51
DAFTAR REVISI		56

DAFTAR GAMBAR

Gambar 1. 1 Bagan Keterkaitan	4
Gambar 3. 1 Tahapan Penelitian	16
Gambar 4. 1 Script Scraping data	22
Gambar 4. 2 Script Cleaning Data 1	24
Gambar 4. 3 Script Cleaning Data 2	24
Gambar 4. 4 Script Case Folding	26
Gambar 4. 5 Script Normalization	27
Gambar 4. 6 Script Tokenizing	29
Gambar 4. 7 Script Stopword Removal.....	30
Gambar 4. 8 Script Stemming	32
Gambar 4. 9 Script Pelabelan Data	33
Gambar 4. 10 Jumlah Analisis Sentimen.....	34
Gambar 4. 11 Script Naïve Bayes.....	34
Gambar 4. 12 Script Suppor Vector Machine	35
Gambar 4. 13 Confusion Matrix Naïve Bayes	36
Gambar 4. 14 Learning Curve Naïve Bayes Classifier	37
Gambar 4. 15 Confusion Matrix SVM	39
Gambar 4. 16 Learning Curve Support Vector Machine.....	40
Gambar 4. 17 Distribusi Sentimen Prediksi SVM.....	42
Gambar 4. 18 Word Cloud	43
Gambar 4. 19 Chart Frekuensi Kata	44
Gambar 4. 20 Word Cloud Sentimen Positif.....	44
Gambar 4. 21 Word Cloud Sentimen Negatif	46
Gambar 4. 22 Word Cloud Sentimen Netral.....	47

DAFTAR TABEL

Tabel 2. 1. Studi Sebelumnya.....	7
Tabel 2. 2. Confusion Matrix.....	15
Tabel 3. 1. Cleaning.....	18
Tabel 3. 2. Case Folding	18
Tabel 3. 3. Normalization.....	19
Tabel 3. 4. Tokenizing	19
Tabel 3. 5. Stopword Removal.....	19
Tabel 3. 6. Stemming	20
Tabel 4. 1. Hasil Pengumpulan Data	23
Tabel 4. 2. Hasil Proses Cleaning	25
Tabel 4. 3. Hasil Proses Case Folding	26
Tabel 4. 4. Hasil Proses Normalization	28
Tabel 4. 5. Hasil Proses Tokenizing	29
Tabel 4. 6. Hasil Proses Stopword Removal.....	31
Tabel 4. 7. Hasil Proses Stemming.....	32
Tabel 4. 8. Hasil Perhitungan Naïve Bayes Classifier	35
Tabel 4. 9. Hasil Perhitungan Support Vector Machine	38
Tabel 4. 10. Perbandingan Algoritma Naïve Bayes Classifier dan Support Vector Machine (SVM)	40
Tabel 4. 11 Perbandingan Learning Curve Algoritma Naïve Bayes Classifier dan Support Vector Machine (SVM).....	42
Tabel 4. 12. Sentimen Positif	45
Tabel 4. 13. Sentimen Negatif.....	46
Tabel 4. 14. Sentimen Netral	48