

**PENERAPAN ALGORITMA NAIVE BAYES UNTUK
ANALISIS SENTIMEN PENGGUNA APLIKASI CoreTax**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

M. Habib Musthofa Anshori

21.11.3911

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS
SENTIMEN PENGGUNA APLIKASI CoreTax**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

M. Habib Musthofa Anshori

21.11.391

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2025**

HALAMAN PERSETUJUAN

SKRIPSI

PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN

PENGUNA APLIKASI CoreTax

yang disusun dan diajukan oleh

M. Habib Musthofa Anshori

21.11.3911

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Desember 2025

Dosen Pembimbing,



Ferian Fauzi Abdulloh, M. Kom.
NIK. 490302276

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN

PENGUNA APLIKASI CoreTax

yang disusun dan diajukan oleh

M. Habib Musthofa Anshori

21.11.3911

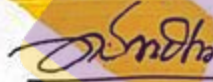
Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Desember 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Windha Mega Pradnya Duhita, M.Kom.
NIK. 190302185



Anna Baita, M. Kom
NIK. 190302290



Ferian Fauzi Abdulloh, M.Kom
NIK. 190302276



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Desember 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama Mahasiswa : M. Habib Musthofa Anshori
NIM : 21.11.3911

Menyatakan bahwa Skripsi dengan judul berikut:

PENERAPAN ALGORITMA NAIVE BAYES UNTUK ANALISIS SENTIMEN PENGGUNA APLIKASI CoreTax

Dosen Pembimbing : Ferian Fauzi Abdulloh, M. Kom.

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 12 Desember 2025

Yang Menyatakan,



METERAI TEMPEL
460ANX191231152

M. Habib Musthofa Anshori

HALAMAN PERSEMBAHAN

Karya ilmiah ini didedikasikan untuk saya dengan rasa syukur dan hormat:

1. Allah SWT yang memberi sumber kekuatan, petunjuk, dan kedamaian penulis. Kasih sayang dan rahmat-Nya membantu penulis menyelesaikan setiap bagian penulisan tesis ini.
2. Orang tua penulis, Abi dan Umi, telah memberikan cinta, doa, dan pengorbanan yang tak ternilai harganya. Mereka membantu dan mendorong kami untuk mencapai semua tujuan ini.
3. Saudari-saudari penulis, Alifah dan Rifka, selalu ada ketika penulis membutuhkannya, memberikan kekuatan dan dukungan. Keberadaan mereka membuat setiap langkah yang penulis ambil menjadi lebih menyenangkan dan menarik.
4. Teman-teman seperti Dika, Renal, Ilham, Naufal, Rakka, Febri, Fakhri, Hasan, Ranga, Joko, dan lainnya. Saya ingin mengucapkan terima kasih atas dukungan kalian, bantuan kalian, dan semangat kalian.

KATA PENGANTAR

Puji syukur saya panjatkan kepada Allah SWT, Tuhan Yang Maha Pengasih dan Penyayang. Atas bantuan, kemudahan, dan ketabahan yang telah Engkau berikan kepada saya untuk menyelesaikan tesis ini. Engkau senantiasa membantu saya dalam segala upaya. Meskipun menghadapi berbagai rintangan, saya telah mengerahkan segala upaya untuk bertekun, menimba ilmu, dan terus berusaha.

Saya ingin menyampaikan berjuta rasa terima kasih kepada ayah saya, Wahyudi, S.Pi. dan ibu saya, Hadining Susilawati, S.Pd., M.Pd. atas doa, dukungan mental, dorongan, dan pengorbanan mereka yang tak henti-hentinya, terutama finansial, selama perjalanan akademis saya. Tanpa kehadiran kalian, saya bukanlah siapa-siapa dan tidak akan ada di tempat ini.

Saya mengucapkan terima kasih kepada kakak perempuan saya, Alifah Rahma Fauzia, S.Pd. dan adik perempuan saya, Rifka Amalia. Bantuan, doa, humor, dan dukungan kalian senantiasa memperlancar perjalanan saya. Kehadiran kalian telah menguatkan saya. Saya menciptakan karya ini untuk mengungkapkan rasa terima kasih dan kasih sayang saya yang mendalam kepada semua orang yang telah mendukung saya dalam hidup dan kegiatan akademis saya.

Yogyakarta, 23 November 2025

Penulis

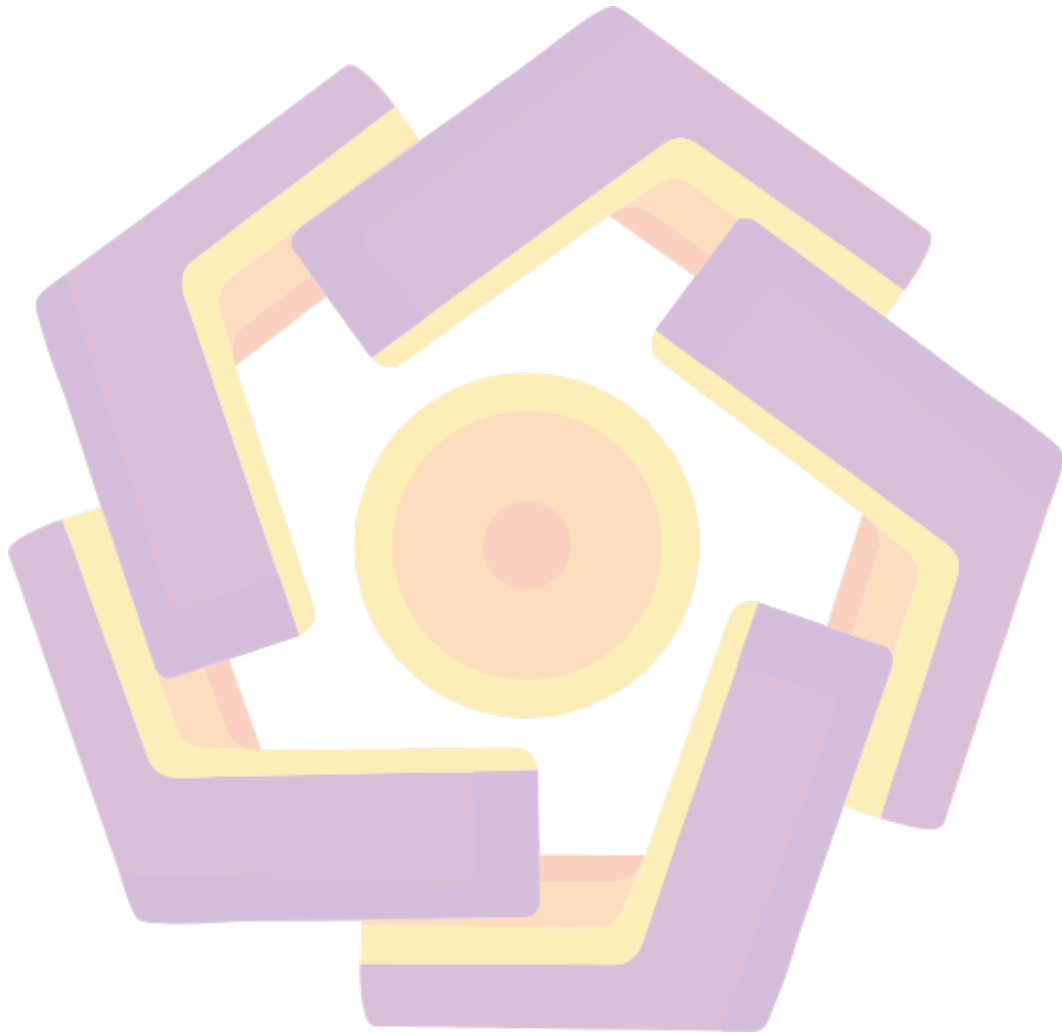


DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI.....	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori.....	13
2.2.1 Analisis Sentimen.....	13
2.2.2 Youtube	13

2.2.2	<i>Exploratory Data Analysis (EDA)</i>	14
2.2.3	<i>Naive Bayes</i>	14
2.2.4	<i>Confusion Matrix</i>	16
BAB III METODE PENELITIAN		18
3.1	Objek Penelitian	18
3.2	Alur Penelitian.....	18
3.2.1	<i>Scraping Data</i>	20
3.2.2	<i>Labeling</i>	22
3.2.3	<i>Pre-processing Data</i>	23
3.2.4	<i>Pre-Modeling</i>	25
3.2.5	<i>Modeling (Naive Bayes)</i>	26
4.2.6	<i>Model Evaluation (Confusion Matrix)</i>	27
3.3	Alat dan Bahan	27
BAB IV HASIL DAN PEMBAHASAN		29
4.1	<i>Scraping Data</i>	29
4.2	<i>Labeling Data</i>	30
4.3	<i>Pre-processing Data</i>	32
4.3.1	<i>Cleaning Data & Case Folding</i>	32
4.3.2	<i>Word Normalization</i>	33
4.3.3	<i>Stopword Removal</i>	34
4.3.4	<i>Tokenizing</i>	35
4.3.5	<i>Stemming</i>	36
4.4	<i>Pre-Modeling</i>	37
4.4.1	<i>Word Cloud</i>	37
4.4.2	<i>Split Data Train & Test</i>	38
4.4.3	<i>TF-IDF (Term Frequency – Inverse Document Frequency)</i>	39
4.4.4	<i>SMOTE (Synthetic Minority Over-sampling Technique)</i>	40
4.5	<i>Modeling Naive Bayes</i>	41
BAB V PENUTUP		46

5.1	Kesimpulan.....	46
5.2	Saran.....	47
REFERENSI.....		48
LAMPIRAN.....		51



DAFTAR TABEL

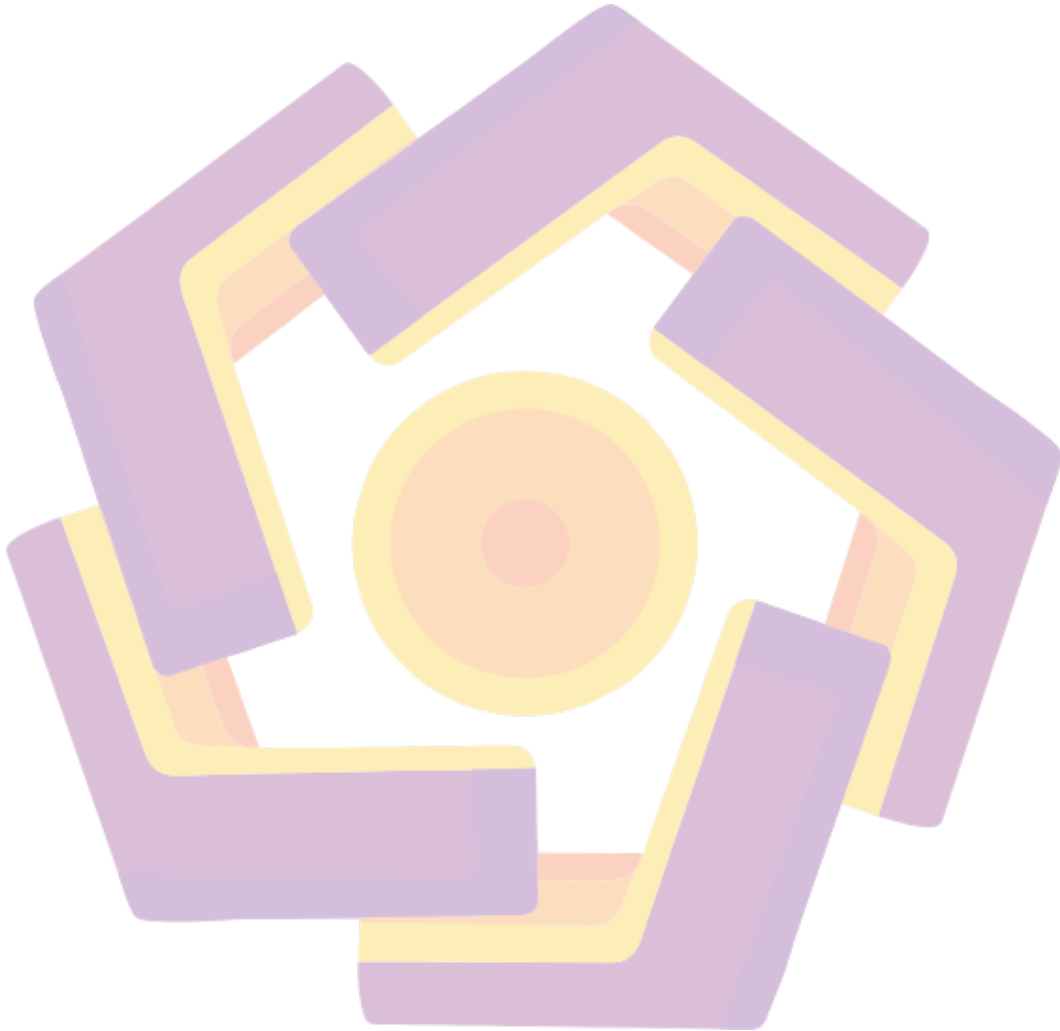
Tabel 2. 1 Keaslian Penelitian	9
Tabel 2. 2 <i>Confusion Matrix</i>	16
Tabel 3. 1 <i>Data Scraping</i>	20
Tabel 4. 1 Top 3 <i>Data Scraping</i>	29
Tabel 4. 2 Hasil <i>Labeling Data</i>	31
Tabel 4. 3 Hasil <i>Cleaning Data & Case Folding</i>	32
Tabel 4. 4 Hasil <i>Word Normalization</i>	33
Tabel 4. 5 Hasil <i>Stopword Removal</i>	34
Tabel 4. 6 Hasil <i>Tokenizing</i>	35
Tabel 4. 7 Hasil <i>Stemming</i>	36
Tabel 4. 8 Hasil Evaluasi Model <i>Naïve Bayes</i>	42
Tabel 4. 9 Distribusi Prediksi Model.....	42

DAFTAR GAMBAR

Gambar 2. 1 Arsitektur <i>Naive bayes</i>	15
Gambar 3. 1 Alur Penelitian	19
Gambar 3. 2 Label Sentimen	23
Gambar 4. 1 <i>WordCloud</i> Positif	37
Gambar 4. 2 <i>WordCloud</i> Negatif	37
Gambar 4. 3 <i>Data Train & Test</i>	39
Gambar 4. 4 <i>Data Train</i> Sebelum <i>SMOTE</i>	40
Gambar 4. 5 <i>Data Train</i> Setelah <i>SMOTE</i>	41
Gambar 4. 6 Evaluasi Model Label Negatif	43
Gambar 4. 7 Evaluasi Model Label Negatif	43
Gambar 4. 8 <i>Confusion Matrix Naive Bayes</i>	44

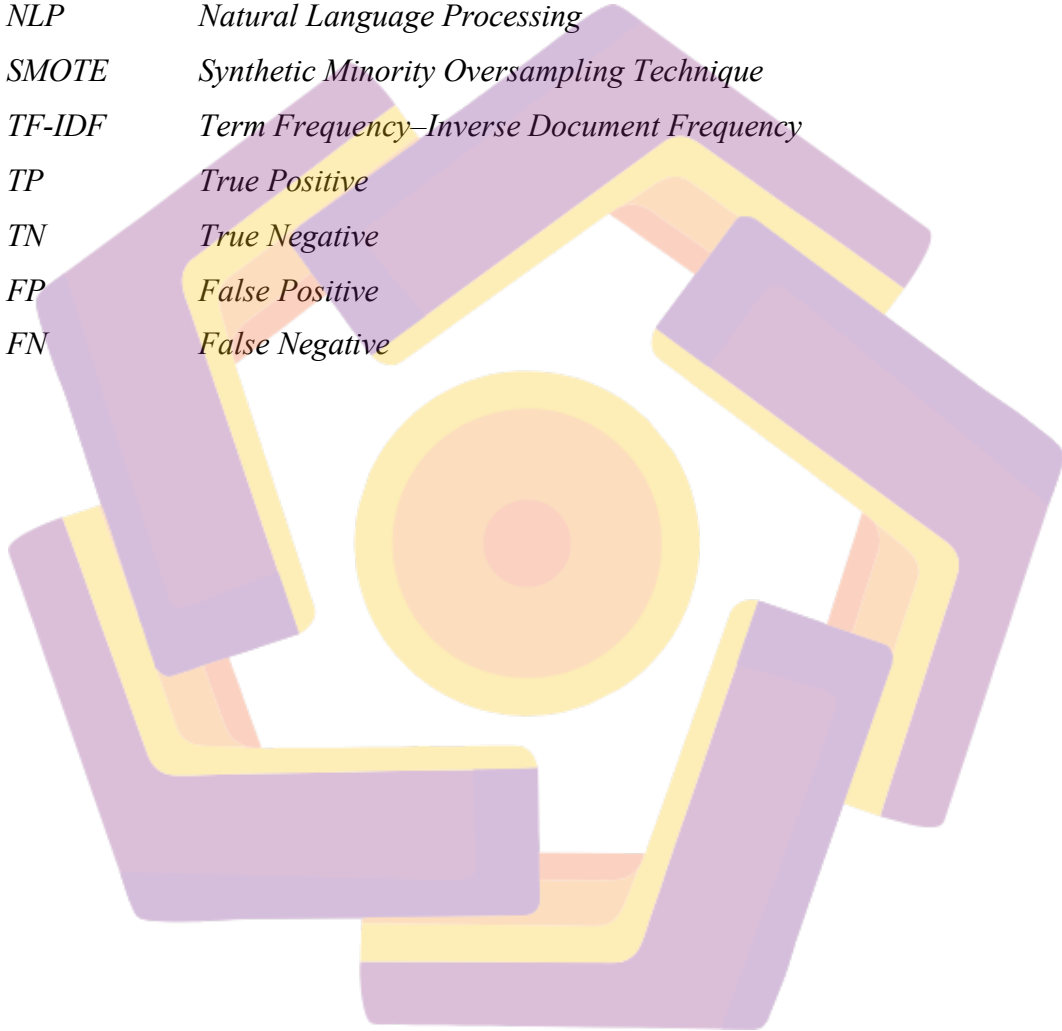
DAFTAR LAMPIRAN

Lampiran 1 Kode <i>Scrapping Data</i>	51
Lampiran 2 Kode <i>Pre-Processing Data</i>	52
Lampiran 3 Kode <i>Modeling Data</i>	55



DAFTAR LAMBANG DAN SINGKATAN

C	Kelas Prediksi
X	Fitur Inputan
P	Probabilitas
NLP	<i>Natural Language Processing</i>
$SMOTE$	<i>Synthetic Minority Oversampling Technique</i>
$TF-IDF$	<i>Term Frequency–Inverse Document Frequency</i>
TP	<i>True Positive</i>
TN	<i>True Negative</i>
FP	<i>False Positive</i>
FN	<i>False Negative</i>



DAFTAR ISTILAH

<i>Naive Bayes</i>	Algoritma klasifikasi berbasis probabilitas
<i>Machine Learning</i>	Metode untuk membuat sistem belajar dari data.
<i>Pre-Processing</i>	Tahap awal untuk membersihkan dan mempersiapkan data.
<i>Labeling</i>	Proses memberikan label atau kategori pada data.
<i>Casefolding</i>	Proses mengubah teks menjadi bentuk standar.
<i>Text Cleaning</i>	Proses membersihkan teks dari karakter, simbol atau noise.
<i>Tokenizing</i>	Proses memecah teks menjadi kata-kata atau token.
<i>Stopword Removal</i>	Menghapus kata-kata umum yang tidak bermakna penting.
<i>Stemming</i>	Proses mengubah kata imbuhan menjadi bentuk dasarnya.
<i>Pre-Modeling</i>	Tahap persiapan data sebelum dimasukkan ke dalam model.
<i>Split Data</i>	Membagi dataset untuk pelatihan dan pengujian.
<i>Data Train</i>	Data yang digunakan untuk melatih model.
<i>Data Test</i>	Data yang digunakan untuk menguji model.
<i>Tokenizing</i>	Proses mengubah teks menjadi numerik.
<i>Modeling</i>	Tahap pembuatan dan pelatihan model untuk analisis data.
<i>Precision</i>	Proporsi prediksi positif yang benar.
<i>Recall</i>	Proporsi data positif yang terdeteksi dengan benar.
<i>F1-Score</i>	Rata-rata harmonis dari precision dan recall.
<i>Accuracy</i>	Proporsi prediksi yang benar.
<i>True Positive</i>	Data positif yang diprediksi benar.
<i>True Negative</i>	Data negatif yang diprediksi benar.
<i>False Positive</i>	Data negatif yang diprediksi sebagai positif.
<i>False Negative</i>	Data positif yang diprediksi sebagai negatif.

INTISARI

Perkembangan teknologi informasi telah mendorong transformasi digital di berbagai sektor, termasuk bidang perpajakan di Indonesia. Direktorat Jenderal Pajak (DJP) mengimplementasikan aplikasi CoreTax sebagai upaya meningkatkan efisiensi, transparansi, dan kenyamanan dalam proses administrasi perpajakan. Meskipun demikian, hadirnya aplikasi ini memunculkan beragam respons dari masyarakat, mulai dari pengalaman positif terkait kemudahan akses hingga keluhan mengenai kendala teknis yang menghambat proses pelaporan. Untuk memahami persepsi publik secara menyeluruh, penelitian ini menerapkan analisis sentimen berbasis teks terhadap komentar pengguna CoreTax yang diperoleh dari platform YouTube menggunakan algoritma Naive Bayes.

Tahapan penelitian mencakup proses preprocessing data seperti cleaning, case folding, tokenizing, stopword removal, word normalization, dan stemming untuk menghasilkan teks yang lebih bersih dan seragam. Representasi fitur dilakukan menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF), sedangkan teknik Synthetic Minority Over-sampling Technique (SMOTE) digunakan untuk menangani ketidakseimbangan pada distribusi kelas sentimen. Hasil pengujian menunjukkan bahwa model Naive Bayes yang dibangun memiliki performa yang baik dengan akurasi sebesar 92%, precision 0.85, recall 0.82, dan f1-score 0.84. Nilai tersebut menunjukkan kemampuan model dalam mengidentifikasi sentimen positif dan negatif secara cukup akurat.

Secara keseluruhan, penelitian ini memberikan gambaran objektif mengenai persepsi masyarakat terhadap aplikasi CoreTax sekaligus menunjukkan efektivitas algoritma Naive Bayes dalam tugas klasifikasi sentimen berbasis teks. Hasil ini dapat menjadi acuan akademis dalam pengembangan penelitian terkait analisis sentimen menggunakan metode pembelajaran mesin

Kata kunci: Analisis Sentimen, Naive Bayes, CoreTax, TF-IDF, SMOTE

ABSTRACT

The rapid development of information technology has driven digital transformation across various sectors, including the taxation system in Indonesia. The Directorate General of Taxes (DJP) introduced the CoreTax application as an effort to improve efficiency, transparency, and convenience in tax administration processes. However, the implementation of this system has received diverse responses from users, ranging from positive feedback regarding ease of access to complaints about technical issues that hinder reporting activities. To obtain an objective understanding of public perception, this study applies text-based sentiment analysis on user comments about CoreTax collected from the YouTube platform, utilizing the Naive Bayes algorithm as the classification method.

The research process involves several data preprocessing stages, including cleaning, case folding, tokenizing, stopword removal, word normalization, and stemming, in order to produce cleaner and more uniform textual data. Text representation is carried out using the Term Frequency–Inverse Document Frequency (TF-IDF) method, while the Synthetic Minority Over-sampling Technique (SMOTE) is employed to address class imbalance in the sentiment dataset. The evaluation results show that the Naive Bayes model demonstrates strong performance with an accuracy of 92%, precision of 0.85, recall of 0.82, and an F1-score of 0.84. These metrics indicate the model's reliable ability to classify both positive and negative sentiments.

Overall, this study provides an objective overview of public sentiment toward the CoreTax application and highlights the effectiveness of the Naive Bayes algorithm for text-based sentiment classification.

Keyword: Sentiment Analysis, Naive Bayes, CoreTax, TF-IDF, SMOTE