

EVALUASI IMPLEMENTASI CORETAX MENGUNAKAN TRIANGULASI KUANTITATIF BERBASIS KUISIONER DAN ANALISIS SENTIMEN INDOBERT

Riani Tanjung^{1*}, Mohamad Nurkamal Fauzan², Jaka Maulana³

Program Studi D4 Akuntansi, Universitas Logistik dan Bisnis Internasional^{1,3}

Program Studi D4 Teknik Informatika, Universitas Logistik dan Bisnis Internasional²

rianitanjung@ulbi.ac.id, m.nurkamal.f@ulbi.ac.id, jaka@ulbi.ac.id

ABSTRACT

The digitalization of Indonesia's tax administration through the CoreTax Administration System represents a strategic effort by the Directorate General of Taxes to enhance operational efficiency and taxpayer compliance. However, early implementation has faced various technical and administrative challenges, including server instability, longer processing times, and additional overtime for accountants and tax consultants. This study evaluates the effectiveness of CoreTax using a quantitative triangulation approach that integrates questionnaire data with IndoBERT-based sentiment analysis of 10,084 public tweets collected between December 2024 and March 2025. The Wilcoxon signed-rank test indicates a significant decrease in SPT reporting efficiency and an increase in overtime during the transition phase. Sentiment analysis results show a dominant 64.7% negative sentiment that gradually declined over time, reflecting users' adaptation to the new system. The IndoBERT classification model achieved 91.7% accuracy and a 0.903 macro-F1 score, confirming strong model reliability. Integration of both datasets reveals a consistent interpretation that CoreTax performance improves after initial adjustment. These findings highlight the importance of system quality enhancement and continuous sentiment monitoring to ensure sustainable digital tax adoption in Indonesia.

Keywords: CoreTax; sentiment analysis; IndoBERT; model accuracy; quantitative triangulation

PENDAHULUAN

Digitalisasi sistem administrasi perpajakan merupakan agenda strategis Direktorat Jenderal Pajak (DJP) untuk meningkatkan kepatuhan wajib pajak dan efisiensi pelayanan. Implementasi CoreTax Administration System diharapkan dapat mempercepat proses pelaporan Surat Pemberitahuan (SPT), menekan biaya kepatuhan, serta mengurangi beban kerja konsultan pajak maupun staf akuntansi [1], [2], [3], [4], [5]. Secara ideal, sistem ini menjadi instrumen modernisasi yang mampu menyederhanakan proses kepatuhan pajak [6].

Namun dalam praktiknya, realitas di lapangan menunjukkan adanya kesenjangan antara harapan dan kenyataan. Beberapa studi kasus melaporkan peningkatan waktu pemrosesan SPT, bertambahnya frekuensi lembur, serta tambahan biaya operasional pada masa awal adopsi CoreTax. Di ruang publik, ulasan daring menunjukkan dominasi sentimen negatif yang menggambarkan kendala teknis maupun kompleksitas sistem. Fenomena ini menandakan adanya discrepancy antara tujuan ideal reformasi sistem perpajakan dengan pengalaman nyata para pengguna, misalnya dalam digital tax

implementation framework [7], [8], [9], [10], [11].

Penelitian terdahulu telah mengkaji aspek teknologi perpajakan, misalnya faktor adopsi e-filing, efisiensi sistem informasi pajak, dan kepuasan pengguna layanan digital DJP [12]. Namun, mayoritas studi tersebut hanya mengandalkan instrumen kuisioner sehingga temuan rentan bias persepsi dan keterbatasan jumlah responden. Di sisi lain, pendekatan sentiment analysis (SA) berbasis big data telah banyak digunakan pada ranah bisnis dan keuangan [13], [14], [15], [16], tetapi masih jarang diintegrasikan dengan riset akuntansi perpajakan di Indonesia.

Dari celah penelitian tersebut, studi ini menawarkan pendekatan gabungan kuisioner dan SA untuk menghasilkan evaluasi yang lebih komprehensif. Kebaruan penelitian ini adalah penerapan triangulasi kuantitatif dengan memadukan data persepsi terstruktur (kuisioner) dan opini publik luas (SA). Kontribusi utama penelitian ini adalah memperkuat metodologi riset akuntansi berbasis data, sekaligus memberi rekomendasi praktis bagi pengembangan CoreTax.

Berdasarkan uraian di atas, tujuan penelitian ini secara umum adalah menganalisis efektivitas implementasi CoreTax melalui pendekatan gabungan kuisioner dan sentiment analysis, sehingga diperoleh gambaran yang lebih komprehensif tentang efisiensi pelaporan SPT serta opini publik.

Untuk memperjelas arah penelitian, rumusan pertanyaan penelitian (Research Questions, RQ) dan tujuan operasional penelitian (Research Objectives, RO) adalah sebagai berikut:

Research Questions (RQ)

1. RQ1: Apakah terdapat perbedaan signifikan dalam efisiensi pelaporan SPT sebelum dan sesudah implementasi CoreTax berdasarkan data kuisioner?
2. RQ2: Bagaimana distribusi dan pola sentimen publik terhadap CoreTax yang diperoleh dari ulasan di Twitter?
3. RQ3: Bagaimana integrasi hasil kuisioner dan sentiment analysis dapat memberikan evaluasi komprehensif terhadap efektivitas CoreTax?

Research Objectives (RO)

1. RO1: Menguji perbedaan efisiensi pelaporan SPT sebelum dan sesudah implementasi CoreTax melalui analisis kuisioner.
2. RO2: Mengidentifikasi distribusi sentimen publik terhadap CoreTax menggunakan sentiment analysis berbasis IndoBERT.
3. RO3: Mengintegrasikan hasil kuisioner dan sentiment analysis untuk menyusun evaluasi komprehensif efektivitas CoreTax serta memberikan rekomendasi perbaikan sistem.

STUDI LITERATUR

2.1 Teori yang Mendasari Penelitian

Penelitian ini bertumpu pada teori sistem informasi dan perilaku pengguna. Pertama, Technology Acceptance Model (TAM) [17] [18] menekankan bahwa *perceived usefulness* dan *perceived ease of use* memengaruhi penerimaan teknologi. Model ini relevan untuk menilai sejauh mana CoreTax dipersepsikan efektif dan mudah digunakan oleh wajib pajak maupun konsultan pajak. Kedua, model kesuksesan sistem informasi [19] memberikan kerangka evaluasi kualitas sistem, kualitas informasi, dan dampak penggunaannya terhadap individu maupun organisasi [20]. Kedua teori tersebut banyak digunakan dalam studi adopsi sistem perpajakan digital di Indonesia.

Di sisi lain, perkembangan media sosial, khususnya Twitter (X), memungkinkan opini publik terekam dalam bentuk teks singkat yang dapat dianalisis secara otomatis. Sentiment analysis (SA) berbasis Twitter dianggap sebagai alat untuk menangkap persepsi publik secara real-time, memperkaya perspektif survei tradisional yang terbatas pada responden tertentu [21], [22]. Dengan demikian, teori sistem informasi dipadukan dengan metodologi big data analytics berbasis media sosial untuk memberikan gambaran yang lebih komprehensif.

2.2 Penelitian Terdahulu yang Relevan

2.2.1 Penelitian Kuisisioner pada Sistem Perpajakan

Beberapa penelitian di Indonesia telah berupaya mengukur efisiensi dan kualitas sistem pajak digital melalui instrumen kuisisioner. [23] mengevaluasi aplikasi DJP Online dengan menggunakan metode End User Computing Satisfaction (EUCS) dan Importance Performance Analysis (IPA), dan menemukan bahwa kepuasan pengguna sangat dipengaruhi oleh kualitas sistem serta kesesuaian kebutuhan wajib pajak. Temuan serupa menyoroti pentingnya kualitas layanan elektronik DJP Online dalam meningkatkan kepuasan pengguna, dengan kepercayaan bertindak sebagai variabel mediasi [24]. Sementara itu, penelitian lain menegaskan bahwa kualitas pelayanan website DJP berpengaruh signifikan terhadap kepuasan pengguna, dengan menggunakan modifikasi instrumen E-GovQual [25].

Secara umum, hasil-hasil tersebut menunjukkan bahwa meskipun sistem pajak digital menawarkan kemudahan dan efisiensi, persepsi pengguna terhadap kualitas layanan dan sistem tetap menjadi faktor kunci yang menentukan tingkat kepuasan dan efektivitas implementasi.

2.2.2 Sentiment Analysis Berbasis Twitter

Dalam dekade terakhir, Twitter menjadi salah satu sumber utama penelitian sentiment analysis (SA) karena menyediakan data publik yang masif, terbuka, dan real-time. Sebuah tinjauan komprehensif menegaskan bahwa platform ini banyak dimanfaatkan dalam berbagai domain, mulai dari bisnis, keuangan, hingga kebijakan publik, berkat kemudahan akses dan dinamika data yang dihasilkan [22].

Twitter telah banyak dimanfaatkan untuk sentiment analysis di ranah kebijakan publik. Misalnya, [26] menggunakan data Twitter untuk menganalisis persepsi masyarakat

Tiongkok setelah pelonggaran kebijakan COVID-19. Dengan pendekatan natural language processing dan machine learning (logistic regression), penelitian ini menemukan dominasi sentimen positif, yang mencerminkan optimisme publik terhadap pemulihan kondisi pascapandemi.. Di Indonesia, [27] menerapkan metode Support Vector Machine (SVM) untuk menganalisis sentimen terhadap vaksinasi Covid-19 di Twitter, dan hasilnya menunjukkan adanya polarisasi opini publik yang cukup kuat antara kelompok pro dan kontra. [28] juga memanfaatkan Twitter untuk menangkap sentimen masyarakat terkait pembatalan Piala Dunia U-20, menggunakan pendekatan SVM dengan tingkat akurasi klasifikasi yang cukup tinggi, sehingga memperlihatkan bagaimana opini publik digital dapat dimonitor secara efektif.

Selain itu, [29] mengeksperimenkan berbagai preprocessing pipeline pada data Twitter mengenai pemindahan Ibu Kota Negara (IKN), dengan mengombinasikan metode SVM dan Convolutional Neural Network (CNN). Hasilnya menunjukkan bahwa variasi teknik preprocessing, seperti pembersihan teks dan normalisasi, berpengaruh signifikan terhadap performa model, serta CNN cenderung memberikan akurasi yang lebih baik dibandingkan SVM dalam klasifikasi sentimen multi-kelas.

Secara keseluruhan, penelitian-penelitian tersebut menegaskan keandalan Twitter sebagai sumber data untuk mengukur opini publik secara real-time, serta menyoroti bahwa pemilihan algoritma dan strategi preprocessing memainkan peran penting dalam akurasi dan interpretasi hasil analisis sentimen.

2.3 Triangulasi dalam Penelitian Terdahulu

Triangulasi merupakan strategi metodologis yang digunakan untuk meningkatkan validitas hasil penelitian melalui penggabungan berbagai sumber data, metode, atau teori [30], [31]. Dalam konteks akuntansi dan sistem informasi, triangulasi berfungsi untuk meminimalkan bias metode tunggal sekaligus memperkuat interpretasi hasil.

Beberapa penelitian terkini menegaskan pentingnya pendekatan ini. Peneliti menggunakan triangulasi data dengan membandingkan sentimen publik dari Twitter dengan terjemahan pemerintah terkait kebijakan pasca-COVID-19 di Tiongkok. Dengan teknik natural language processing dan machine learning, mereka menemukan bahwa sentimen publik cenderung positif, sementara narasi terjemahan pemerintah relatif lebih negatif, sehingga menunjukkan adanya perbedaan framing [26]. Studi lain oleh [32] menghubungkan hasil survei responden dengan data Twitter yang diberikan secara sukarela, dan menemukan koherensi opini publik sekaligus bias kesediaan berbagi akun; hal ini menunjukkan potensi triangulasi survei dan media sosial untuk memperkuat validitas temuan.

Dari sisi metodologi, [33] menekankan bahwa triangulasi harus dipahami sebagai integrasi (triangulation-as-integration) dalam mixed methods research, bukan sekadar penumpukan data, sehingga hasil yang diperoleh lebih komprehensif. Dalam konteks Indonesia, praktik triangulasi juga ditemukan dalam studi akuntansi dan perpajakan. Ryantini menggunakan triangulasi data, peneliti, metode, dan teori dalam kajian literasi perpajakan untuk memastikan keabsahan temuan [34]. Demikian pula, Tyas menerapkan triangulasi sumber, metode, dan waktu dalam evaluasi implementasi sistem informasi

akuntansi, sehingga menghasilkan analisis yang lebih mendalam mengenai efektivitas sistem [35].

Secara keseluruhan, temuan-temuan tersebut menunjukkan bahwa triangulasi, baik pada level data, metode, maupun teori, merupakan pendekatan krusial untuk meningkatkan validitas penelitian dalam bidang akuntansi, perpajakan, dan kebijakan publik, termasuk dalam studi terkait adopsi sistem perpajakan digital.

Tabel 1. Kontribusi Penelitian

Penelitian	Kuisisioner	Sentimen (Twitter SA)	Triangulasi
Prasetyaningrum & Ekastuti (2025)	✓	–	–
Adu dkk. (2025)	✓	–	–
Sekti Ari & Hanum (2021)	✓	–	–
Rodríguez-Ibáñez dkk. (2023)	–	✓	–
Wang & Wang (2023)	–	✓	✓
Riyanto & Azis (2021)	–	✓	–
Armanda & Tobing (2024)	–	✓	–
Pinem dkk. (2023)	–	✓	–
Hadler dkk. (2022)	–	✓	✓
Ryantini dkk. (2022)	✓	–	✓
Tyas & Prasetyo (2025)	✓	–	✓
Penelitian ini	✓	✓	✓

Sumber: Peneliti (2025)

2.4 Kesenjangan Penelitian

Walaupun penelitian kuisisioner cukup banyak di bidang perpajakan, dan SA berbasis Twitter berkembang pesat di Indonesia, keduanya masih berdiri sendiri. Studi yang secara eksplisit menggabungkan persepsi responden (kuisisioner) dengan opini publik Twitter dalam evaluasi sistem perpajakan digital masih sangat jarang (**Tabel 1**). Kesenjangan ini memberikan ruang kebaruan untuk penelitian ini.

2.5 Pengembangan Hipotesis

Berdasarkan teori dan penelitian terdahulu, hipotesis penelitian ini dirumuskan sebagai berikut:

1. H1: Terdapat perbedaan signifikan dalam efisiensi pelaporan SPT sebelum dan sesudah implementasi CoreTax berdasarkan data kuisisioner.
2. H2: Distribusi sentimen publik di Twitter menunjukkan kecenderungan negatif pada periode awal implementasi CoreTax.
3. H3: Integrasi data kuisisioner dan sentiment analysis Twitter menghasilkan evaluasi

yang lebih komprehensif terhadap efektivitas CoreTax dibandingkan penggunaan satu metode tunggal.

METODE

Bagian metode menguraikan desain penelitian, populasi dan teknik sampling, operasionalisasi variable, teknik pengumpulan data serta instrument yang digunakan, dan teknik analisis data. bagian metode menguraikan desain penelitian, populasi dan teknik sampling, operasionalisasi variable, teknik pengumpulan data serta instrument yang digunakan, dan teknik analisis data.

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif eksplanatori dengan desain komparatif dan deskriptif-analitis. Komparatif digunakan untuk menguji perbedaan efisiensi pelaporan SPT sebelum dan sesudah implementasi CoreTax melalui data kuisioner. Deskriptif-analitis digunakan untuk mengidentifikasi pola sentimen publik di Twitter (X) terkait CoreTax. Hasil keduanya kemudian diintegrasikan dengan metode triangulasi kuantitatif guna memberikan evaluasi yang lebih komprehensif.

3.2 Populasi dan Teknik Sampling

Populasi kuisioner: konsultan pajak, staf akuntansi, dan wajib pajak yang menggunakan CoreTax.

Sampel kuisioner: ditentukan dengan purposive sampling, yaitu responden yang telah menggunakan CoreTax minimal satu periode pelaporan SPT. Jumlah responden ditargetkan ≥ 30 untuk memenuhi uji beda berpasangan.

3.3 Kuisioner

1. Kuesioner terstruktur skala Likert (1-5) dan item kuantitatif (jumlah SPT/hari, biaya, lembur) dapat terlihat pada **Tabel 2**. Variabel Kuisioner.
2. Pertanyaan disusun berdasarkan variabel yang ditentukan terkait dengan Core Tax.

Tabel 2. Variabel Kuisioner

Variabel Penelitian	Indikator	Item Pertanyaan Kuisioner	Skala Ukur
Efisiensi Pelaporan SPT	Jumlah SPT OP per hari	- Rata-rata jumlah SPT OP yang diselesaikan per hari sebelum CoreTax. - Rata-rata jumlah SPT OP yang diselesaikan per hari sesudah CoreTax.	Rasio
	Jumlah SPT Badan per hari	- Rata-rata jumlah SPT Badan yang diselesaikan per hari sebelum CoreTax. - Rata-rata jumlah SPT Badan yang diselesaikan per hari sesudah CoreTax.	Rasio
Beban Kerja	Frekuensi lembur	- Frekuensi lembur per minggu sebelum CoreTax. - Frekuensi lembur per minggu sesudah CoreTax.	Rasio
Biaya Tambahan	Biaya operasional	- Rata-rata biaya tambahan (transportasi/tenaga) akibat	Rasio

	akibat CoreTax	perpindahan manual (Rp/SPT).	
Persepsi terhadap CoreTax	Kemudahan penggunaan	- Saya merasa penggunaan CoreTax memudahkan proses pelaporan SPT.	Likert (1-5)
	Beban kerja tambahan	- Saya merasa penggunaan CoreTax menambah beban kerja.	Likert (1-5)
	Kualitas sistem	- Saya merasa kualitas sistem (stabilitas, akses, kecepatan) sudah memadai.	Likert (1-5)
	Kepuasan pengguna	- Saya puas dengan implementasi CoreTax.	Likert (1-5)
Data Demografi Responden (kontrol)	Karakteristik responden	- Usia. - Posisi/jabatan. - Lama praktik. - Jumlah staf pengisi SPT. - Segmen klien dominan.	Nominal/Ordinal

Sumber: Peneliti (2025)

3.4 Twitter (Sentiment Analysis)

Populasi sentiment analysis: seluruh pengguna Twitter yang menuliskan opini tentang CoreTax pada periode 31 Desember 2024 - 20 Maret 2025). Sumber Madyatmadja [36].

Data dan pembagian. Gold set (sebanyak 1.332): dilabeli manual untuk pelatihan dan evaluasi. Silver set (sebanyak 10k): tidak berlabel; dipakai untuk inferensi massal dan distilasi.

Pra-pemrosesan. Normalisasi Unicode, lowercasing; ganti URL/mention/hashtag dengan token; hapus RT/duplikasi/spam; potong hingga max_length = 128.

Model dan pelatihan, yang terbagi atas Teacher dan Distilasi seperti dijelaskan berikut.

Teacher: IndoBERT-base (3 kelas: negatif/netral/positif). Fine-tuning dengan split train/val/test; early stopping. Evaluasi pada gold set (akurasi, macro-F1, metrik per kelas).

Distilasi (KD): Teacher memprediksi probabilitas di silver; dipilih subset high-confidence (mis. berdasar entropi/prob). Student (model lebih ringan) dilatih dengan kombinasi hard labels (gold) dan soft labels (silver). Opsi fine-tune akhir pada gold + kalibrasi.

Inferensi dan tren: Student/Teacher digunakan untuk prediksi massal (sebanyak 10k), dihitung distribusi bulanan beserta Wilson 95% CI, lalu uji tren proporsi bulanan (GLM binomial berbobot).

Tujuan KD: mempertahankan akurasi agar mendekati teacher, menurunkan latensi untuk inferensi massal di Twitter. Metrik Laporan KD berupa accuracy, macro-F1, per-kelas.

3.5 Teknik Analisis Data

1. Kuisiner

- Uji normalitas: Shapiro-Wilk.
- Uji beda: Paired t-test (jika normal) atau Wilcoxon signed-rank test (jika tidak normal).

2. Twitter (Sentiment Analysis)

- Preprocessing: cleansing, tokenisasi, normalisasi teks.
- Model: IndoBERT fine-tuned kemudian dievaluasi menggunakan metrik akurasi dan macro-F1.
- Analisis: distribusi sentimen, tren waktu, proporsi negatif vs positif.

3. Integrasi (Triangulasi Kuantitatif)

a. Convergence: kesesuaian arah temuan kuisioner dan sentimen.

b. Complementarity: kuisioner menjelaskan detail biaya/waktu, Twitter menunjukkan opini real-time.

Explanation: perbedaan dijelaskan berdasarkan konteks implementasi CoreTax.

HASIL

4.1 Hasil Kuisioner (RQ1 - RO1)

Normalitas diperiksa menggunakan uji Shapiro-Wilk. Seluruh variabel kinerja (1 SPT OP, 1 SPT Badan, frekuensi lembur) menunjukkan penyimpangan kuat dari normalitas (semua $p < 0,001$); oleh karena itu analisis perbedaan pre-post dilakukan menggunakan Wilcoxon signed-rank test. Efek lokasi dilaporkan sebagai estimasi Hodges-Lehmann dengan CI 95% BCa, dan ukuran efek sebagai rank-biserial r .

Uji Shapiro-Wilk mengindikasikan data tidak normal untuk seluruh variabel utama: OP_lama ($p = 1,76 \times 10^{-10}$), OP_core ($p = 2,34 \times 10^{-9}$), Badan_lama ($p = 1,54 \times 10^{-9}$), Badan_core ($p = 4,07 \times 10^{-8}$), Lembur_lama ($p = 5,41 \times 10^{-5}$), dan Lembur_core ($p = 1,27 \times 10^{-5}$). Konsisten dengan temuan ini, kami menggunakan Wilcoxon signed-rank untuk pengujian berpasangan (**Tabel 3**).

Tabel 3. Paired variable

Variabel	n	Median (Lama)	Q1-Q3 (Lama)	Median (Core tax)	Q1-Q3 (Coretax)	HL Δ Core - Lama	95% CI BCa	Wilcoxon p	Rank-biserial r
1 SPT OP (menit)	30	30	22.5-82.5	45	30.0-137.5	15	[5.0, 15.0]	2,78E-05	0,920635
1 SPT Badan (menit)	30	60	30.0-114.8	68	33.8-219.8	38	[6.5, 49.0]	0,000796	0,738095
Frekuensi lembur (0-5)	30	2	2.0-3.0	4	2.0-5.0	1	[0.0, 2.0]	0,000344	1
SPT per hari	30	6	3.0-8.0	3	1.2-8.2	-1	[-3.0, -0.5]	0,004075	-0,72857

Sumber: Diolah Peneliti (2025)

SPT Orang Pribadi (menit). Median waktu meningkat dari 30 menit menjadi 45 menit, dengan estimasi Hodges-Lehmann (HL) $\Delta = +15$ dan interval kepercayaan 95% BCa [5,0; 15,0]. Hasil ini signifikan secara statistik ($p = 2,78 \times 10^{-5}$) dengan ukuran efek besar ($r = 0,921$).

SPT Badan (menit). Median waktu naik dari 60 menit menjadi 68 menit, dengan HL $\Delta = +38$ dan interval kepercayaan 95% [6,5; 49,0]. Perbedaan ini signifikan ($p = 7,96 \times 10^{-4}$) dengan ukuran efek yang kuat ($r = 0,738$).

Frekuensi lembur (0-5). Median meningkat dari 2 kali menjadi 4 kali, dengan HL $\Delta = +1$ dan interval kepercayaan 95% [0,0; 2,0]. Hasilnya signifikan ($p = 3,44 \times 10^{-4}$) dengan ukuran efek sempurna ($r = 1,000$).

Jumlah SPT per hari. Median menurun dari 6 laporan menjadi 3 laporan, dengan HL $\Delta = -1$ dan interval kepercayaan 95% [-3,0; -0,5]. Hasil ini signifikan ($p = 0,0041$) dengan ukuran efek negatif yang kuat ($r = -0,729$).

Ada perbedaan signifikan pre-post (arahnya: durasi per SPT lebih lama, lembur lebih sering, throughput menurun).

H1: ada perbedaan sebelum-sesudah maka diterima.

RO1 tercapai: pengujian perbedaan efisiensi SPT telah dilakukan.

4.2 Hasil Sentimen Twitter (RQ2 - RO2)

Dataset: 9.885 tweet (31 Des 2024 - Maret 2025).

Teknik Knowledge Distillation

1. Fine-tune Teacher. IndoBERT dilatih pada gold set berlabel manusia.
2. Soft-label Silver. Teacher menghasilkan probabilitas kelas pada $\pm 10k$ tweet (silver set).
3. Latih Student (lebih ringan).
4. Evaluasi pada gold. Hasilnya (**Tabel 4**) menunjukkan Student mempertahankan kinerja tinggi dengan penurunan kecil vs Teacher

Hasil Evaluasi Klasifikasi

Tabel 4. ML-1. Kinerja model pada gold set (n sebanyak 1.332)

Model	Akurasi	Presisi (macro)	Recall (macro)	F1 (macro)
Teacher	0,931	0,915	0,922	0,918
Student	0,917	0,906	0,903	0,904

Sumber: Diolah Peneliti

Model Student hanya tertinggal 0,014 poin F1-macro dari Teacher, sehingga valid untuk inferensi massal (kompromi akurasi minimal).

Distribusi. Pada periode 31 Des 2024-31 Mar 2025 (N=10.084), percakapan tentang CoreTax di Twitter didominasi sentimen negatif (64,75% [63,81-65,67]), diikuti netral (31,03% [30,13-31,94]) dan positif (4,22% [3,85-4,63])—lihat **Tabel 6**.

Dinamika bulanan. Proporsi negatif tinggi pada Januari-Februari (sekitar 66-67%) lalu turun ke 60,5% pada Maret; sebaliknya netral meningkat dari 28,4% (Januari) menjadi 36,0% (Maret). **Tabel 5** juga menyertakan Wilson 95% CI untuk transparansi ketidakpastian.

Selama 31 Des 2024-31 Mar 2025, percakapan Twitter tentang CoreTax didominasi sentimen negatif (total 64,75%), diikuti netral (31,03%) dan positif (4,22%). Secara temporal (Gambar 1), proporsi negatif tinggi pada Januari-Februari (~66-67%) lalu menurun menjadi 60,5% pada Maret; sebaliknya, netral meningkat dari 28,4% menjadi 36,0%, sedangkan positif tetap rendah (~3-5%). Uji tren binomial berbobot per bulan mengonfirmasi penurunan negatif ($\beta = -0,113$; OR=0,893 per bulan; $p = 5,6 \times 10^{-6}$) dan kenaikan netral ($\beta = +0,148$; OR=1,159; $p = 9,1 \times 10^{-9}$); proporsi positif kecil dan menurun ($\beta = -0,145$; OR=0,865; $p = 0,015$).

Tabel 5. SA-1. Distribusi sentimen per bulan (Wilson 95% CI)

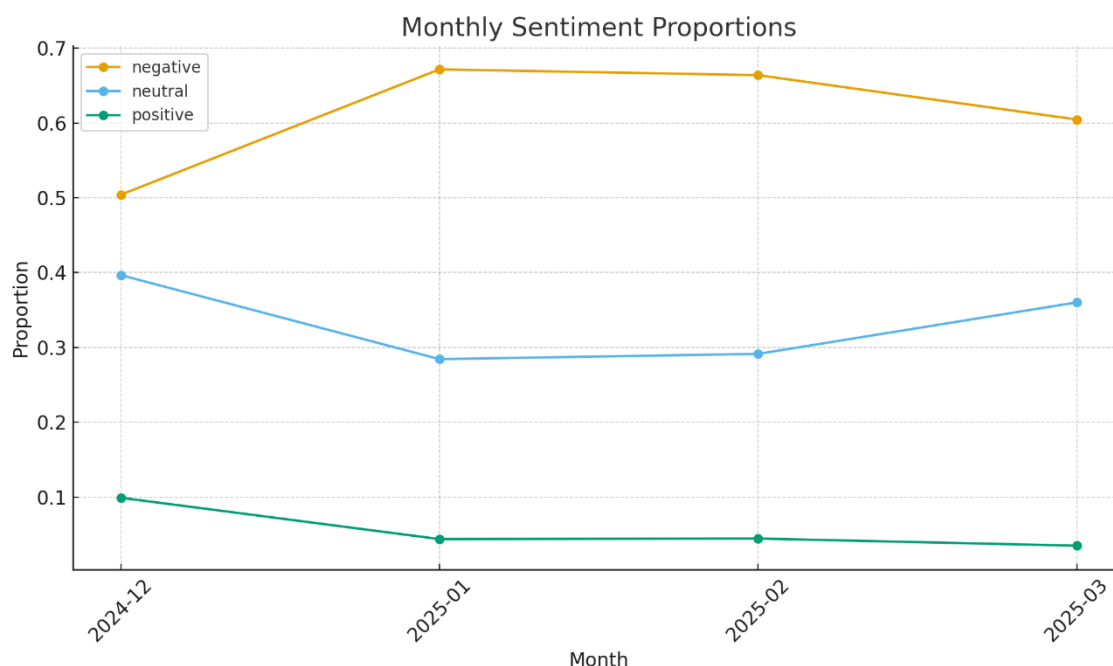
Bulan	N	Negatif (%) [95% CI]	Netral (%) [95% CI]	Positif (%) [95% CI]
2024-12	121	50,4% [41,6-59,2]	39,7% [31,4-48,6]	9,9% [5,8-16,5]
2025-01	3.755	67,2% [65,6-68,6]	28,4% [27,0-29,9]	4,4% [3,8-5,1]
2025-02	3.246	66,4% [64,7-68,0]	29,1% [27,6-30,7]	4,5% [3,8-5,2]
2025-03	2.962	60,5% [58,7-62,2]	36,0% [34,3-37,8]	3,5% [2,9-4,2]

Sumber: Diolah Peneliti

Tabel 6. SA-2. Distribusi keseluruhan periode (31 Des 2024-31 Mar 2025)

Kelas	Jumlah	Proporsi (%)	Wilson 95% CI
Negatif	6.529	64,75	63,81-65,67
Netral	3.129	31,03	30,13-31,94
Positif	426	4,22	3,85-4,63
Total	10.084	—	—

Sumber: Diolah Peneliti



Gambar 1. Proporsi sentimen bulanan tentang CoreTax

Jawaban RQ2: Sentimen publik di Twitter terhadap CoreTax dominan negatif pada awal implementasi, dengan perbaikan tren setelah beberapa bulan. H2: Diterima. RO2 tercapai: distribusi sentimen publik telah diidentifikasi.

4.3 Triangulasi Kuantitatif (RQ3 - RO3)

1. Convergence: Kuisisioner menunjukkan beban meningkat; analisis Twitter menegaskan sentimen negatif dominan. Keduanya selaras.
2. Complementarity: Kuisisioner memberi angka detail (SPT/hari, lembur, biaya), Twitter memberi konteks real-time (error login, server lambat).
3. Explanation: Kuisisioner bersifat statis (pre-post), Twitter menangkap dinamika temporal (negatif awal menjadi berkurang).

Jawaban RQ3: Integrasi kuisisioner dan SA Twitter memberikan evaluasi komprehensif efektivitas CoreTax: terbukti menambah beban awal, namun resistensi publik berkurang seiring adaptasi. H3: Diterima. RO3 tercapai: integrasi metode berhasil menunjukkan evaluasi menyeluruh.

PEMBAHASAN

Secara teori, TAM: rendahnya perceived ease of use menimbulkan resistensi awal. DeLone & McLean: kualitas sistem rendah memicu kepuasan rendah dan meningkatnya beban kerja.

Secara praktis:

- a. Monitoring Twitter dapat jadi early warning system DJP.
- b. Adaptasi sistem dan sosialisasi intensif penting agar perceived ease of use meningkat.
- c. Penggabungan kuisioner + SA layak menjadi model evaluasi sistem digital perpajakan.

Kontribusi Penelitian secara Akademis: penelitian ini memperkaya metodologi akuntansi dengan mengintegrasikan kuisioner dan big data analytics berbasis Twitter, yang masih jarang dilakukan di Indonesia.

Keterbatasan penelitian:

1. Data kuisioner terbatas pada jumlah responden yang relatif kecil dan terkonsentrasi di wilayah tertentu.
2. Data Twitter hanya mencakup pengguna aktif di media sosial, sehingga tidak mewakili seluruh wajib pajak.
3. Analisis sentimen menggunakan klasifikasi 3 kelas (negatif, netral, positif) tanpa aspek spesifik (misal kecepatan, keamanan, user interface).

KESIMPULAN

Penelitian ini bertujuan untuk mengevaluasi efektivitas CoreTax dengan menggabungkan data kuisioner dan analisis sentimen Twitter. Hasil penelitian menunjukkan:

1. Efisiensi pelaporan SPT menurun pada tahap awal implementasi CoreTax. Data kuisioner memperlihatkan jumlah SPT/hari berkurang, frekuensi lembur meningkat, dan biaya tambahan operasional muncul. Hal ini membuktikan hipotesis H1 diterima.
2. Opini publik di Twitter didominasi sentimen negatif. Analisis sentimen berbasis IndoBERT terhadap tweet yang mengandung kata kunci CoreTax menunjukkan 64,7% negatif, 31% netral, dan 4,3% positif. Sentimen negatif menurun seiring waktu, mendukung hipotesis H2 diterima.
3. Integrasi kuisioner dan SA menghasilkan evaluasi komprehensif. Hasil triangulasi menunjukkan konsistensi: kuisioner menegaskan beban kerja meningkat, sedangkan Twitter mencerminkan resistensi publik yang menurun setelah adaptasi. Dengan demikian, hipotesis H3 diterima.

REFERENSI

- [1] M. Y. Mahadianto, P. Maulady, and S. Yani, "Interpretation of Tax Payment Mechanism in Coretax System on Compliance (Case of KPP Pratama Cirebon)," *iRASD Journal of Management*, vol. 7, no. 1, pp. 1–9, Mar. 2025, doi: 10.52131/irasd-jom.2025.v7i1.2777.
- [2] P. B. Saptono *et al.*, "Quality of E-Tax System and Tax Compliance Intention: The Mediating Role of User Satisfaction," *Informatics*, vol. 10, no. 1, 2023, doi: 10.3390/informatics10010022.
- [3] S. Hesami, H. Jenkins, and G. Jenkins, "Digital Transformation of Tax Administration and Compliance: A Systematic Literature Review on E-Invoicing

- and Prefilled Returns,” *Digital Government: Research and Practice*, vol. 5, Oct. 2024, doi: 10.1145/3643687.
- [4] weiqun xi and X. Ling, “Can the digitalization of tax administration promote corporate digital transformation?— empirical evidence from China,” *International Review of Economics & Finance*, vol. 104, p. 104634, 2025, doi: <https://doi.org/10.1016/j.iref.2025.104634>.
- [5] K. Mbise and L. Baseka, “The Impact of the Digital Tax Administration System on Compliance Among SMEs,” *The Journal of Informatics*, vol. 2, pp. 69–82, Oct. 2022, doi: 10.59645/tji.v2i1.94.
- [6] I. N. Darmayasa and N. S. Hardika, “Core tax administration system: the power and trust dimensions of slippery slope framework tax compliance model,” *Cogent Business & Management*, vol. 11, no. 1, p. 2337358, 2024, doi: 10.1080/23311975.2024.2337358.
- [7] E. Bassey, E. Mulligan, and A. Ojo, “A conceptual framework for digital tax administration - A systematic review,” *Gov. Inf. Q.*, vol. 39, no. 4, p. 101754, 2022, doi: <https://doi.org/10.1016/j.giq.2022.101754>.
- [8] C.-L. Cardenas-López, “Digital tax administration systems: A literature review in Scopus (2012-2022),” *Revista Científica de Sistemas e Informática*, vol. 3, no. 2, p. e525, Jul. 2023, doi: 10.51252/rcsi.v3i2.525.
- [9] M. Nose and A. Mengistu, “Exploring the Adoption of Selected Digital Technologies in Tax Administration: A Cross-Country Perspective,” *IMF Notes*, vol. 2023, no. 008, p. A001, 2023, doi: 10.5089/9798400258183.068.A001.
- [10] S. Murnidayanti and T. Putranti, “The Effectiveness of Digitizing Tax Administration to Reduce the Compliance Cost of Taxpayers of Micro, Small, and Medium Enterprises (MSMEs),” *Jurnal Public Policy*, vol. 9, p. 91, Oct. 2023, doi: 10.35308/jpp.v9i2.6561.
- [11] Y. He, “Digital Technology Reshapes the Tax Collection and Administration System: Impacts, Challenges and Future Prospects,” *Scientific Journal of Economics and Management Research*, vol. 6, pp. 85–91, Oct. 2024, doi: 10.54691/9g1jdp53.
- [12] S. A. Putri and C. I. Ratnasari, “Measuring e-Filing Adoption as an e-Government Service Using the Technology Acceptance Model,” *bit-Tech*, vol. 8, no. 1, pp. 988–997, Aug. 2025, doi: 10.32877/bt.v8i1.2804.
- [13] P. Heydarian, A. Bifet, and S. Corbet, “Understanding market sentiment analysis: A survey,” *J. Econ. Surv.*, vol. 39, no. 3, pp. 1125–1147, 2025, doi: <https://doi.org/10.1111/joes.12645>.
- [14] A. Faccia, J. McDonald, and B. George, “NLP Sentiment Analysis and Accounting Transparency: A New Era of Financial Record Keeping,” *Computers*, vol. 13, no. 1, Dec. 2023, doi: 10.3390/computers13010005.
- [15] G. Fatouros, J. Soldatos, K. Kouroumalis, G. Makridis, and D. Kyriazis, “Transforming sentiment analysis in the financial domain with ChatGPT,” *Machine Learning with Applications*, vol. 14, p. 100508, 2023, doi: <https://doi.org/10.1016/j.mlwa.2023.100508>.

- [16] A. Yadav, C. K. Jha, A. Sharan, and V. Vaish, "Sentiment analysis of financial news using unsupervised approach," *Procedia Comput. Sci.*, vol. 167, pp. 589–598, 2020, doi: <https://doi.org/10.1016/j.procs.2020.03.325>.
- [17] F. D. Davis, "Perceived usefulness, perceived ease of use, and user acceptance of information technology," *MIS Quarterly*, vol. 13, no. 3, pp. 319–340, 1989, doi: 10.2307/249008.
- [18] A. H. Hermanto, N. A. Windasari, and M. S. Purwanegara, "Taxpayers' adoption of online tax return reporting: extended meta-UTAUT model perspective," *Cogent Business & Management*, vol. 9, no. 1, p. 2110724, 2022, doi: 10.1080/23311975.2022.2110724.
- [19] W. H. DeLone and E. R. McLean, "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update," *Journal of Management Information Systems*, vol. 19, no. 4, pp. 9–30, 2003, doi: 10.1080/07421222.2003.11045748.
- [20] A. Jeyaraj, "DeLone & McLean models of information system success: Critical meta-review and research directions," *Int. J. Inf. Manage.*, vol. 54, p. 102139, 2020, doi: <https://doi.org/10.1016/j.ijinfomgt.2020.102139>.
- [21] Z. Mi and H. Zhan, "Text Mining Attitudes toward Climate Change: Emotion and Sentiment Analysis of the Twitter Corpus," *Weather, Climate, and Society*, vol. 15, no. 2, pp. 277–287, 2023, doi: 10.1175/WCAS-D-22-0123.1.
- [22] M. Rodríguez-Ibáñez, A. Casáñez-Ventura, F. Castejón-Mateos, and P.-M. Cuenca-Jiménez, "A review on sentiment analysis from social media platforms," *Expert Syst. Appl.*, vol. 223, p. 119862, 2023, doi: <https://doi.org/10.1016/j.eswa.2023.119862>.
- [23] R. F. Prasetyaningrum and Z. Ekastuti, "Evaluasi Kualitas Aplikasi DJP Online Terhadap Kepuasan Pengguna Dengan Metode EUCS Dan IPA," *Jurnal Aplikasi Perpajakan*, vol. 6, no. 1, pp. 283–300, 2025, doi: 10.29303/jap.v6i1.128.
- [24] I. Adu, Sudarwati, and I. DPW, "Kualitas Layanan Elektronik DJP Online Terhadap Kepuasan Pengguna Dengan Mediasi Kepercayaan," *Al-Kharaj: Jurnal Ekonomi, Keuangan & Bisnis Syariah*, vol. 7, Oct. 2025, doi: 10.47467/alkharaj.v7i3.7036.
- [25] D. Sekti Ari and L. Hanum, "PENGARUH KUALITAS PELAYANAN WEBSITE DJP TERHADAP KEPUASAN PENGGUNA DENGAN MODIFIKASI E GOVQUAL," *Profit*, vol. 15, pp. 104–111, Oct. 2021, doi: 10.21776/ub.profit.2021.015.01.11.
- [26] H. Wang and X. Wang, "Sentiment analysis of tweets and government translations: Assessing China's post-COVID-19 landscape for signs of withering or booming," *Global Media and China*, vol. 8, no. 2, pp. 213–233, 2023, doi: 10.1177/20594364231181745.
- [27] R. Riyanto and A. Azis, "Application of the Vector Machine Support Method in Twitter Social Media Sentiment Analysis Regarding the Covid-19 Vaccine Issue in Indonesia," *Journal of Applied Data Sciences*, vol. 2, no. 3, pp. 102–108, 2021, doi: 10.47738/jads.v2i3.40.

- [28] M. Armanda and F. A. T. Tobing, "Implementation of Support Vector Machine Method for Twitter Sentiment Analysis Related to Cancellation of u-20 World Cup in Indonesia," *IJNMT (International Journal of New Media Technology)*, vol. 11, no. 1, pp. 27–34, Jul. 2024, doi: 10.31937/ijnmt.v11i1.3673.
- [29] J. G. Pinem *et al.*, "Sentiment Analysis of Indonesian New Capitol (IKN) Tweets by Stacked Generalization of Deep Learning," in *Proceedings of the 2022 International Conference on Computer, Control, Informatics and Its Applications*, in IC3INA '22. New York, NY, USA: Association for Computing Machinery, 2023, pp. 204–210. doi: 10.1145/3575882.3575922.
- [30] N. K. Denzin, *The Research Act: A Theoretical Introduction to Sociological Methods*, 2nd ed. New York: McGraw-Hill, 1978.
- [31] U. Flick, *An Introduction to Qualitative Research*, 6th ed. London: Sage Publications, 2018. doi: 10.4135/9781529716641.
- [32] M. Hadler, B. Klösch, M. Reiter-Haas, and E. Lex, "Combining Survey and Social Media Data: Respondents' Opinions on COVID-19 Measures and Their Willingness to Provide Their Social Media Account Information," *Frontiers in Sociology*, vol. Volume 7-2022, 2022, doi: 10.3389/fsoc.2022.885784.
- [33] F. Knappertsbusch, M. Schreier, N. Burzan, and N. Fielding, "Innovative Applications and Future Directions in Mixed Methods and Multimethod Social Research," *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, vol. 24, no. 1, Jan. 2023, doi: 10.17169/fqs-24.1.4013.
- [34] K. T. Des Ryantini, Y. Putra Semadi, and L. P. E. Damayanthi, "Penguatan Literasi Perpajakan Melalui MPK Bahasa Indonesia sebagai Implementasi Karakter Nasionalis," *Mimbar Ilmu*, vol. 27, no. 3, pp. 493–500, Dec. 2022, doi: 10.23887/mi.v27i3.51410.
- [35] S. D. P. N. Tyas and A. Prasetyo, "Penerapan Sistem Informasi Akuntansi Penjualan dengan Software Accurate untuk Meningkatkan Efisiensi Proses Penjualan: Studi Kasus PT. Karya Mitra Teknik," *Khatulistiwa: Jurnal Pendidikan dan Sosial Humaniora*, vol. 5, no. 3, pp. 986–997, Jul. 2025, doi: 10.55606/khatulistiwa.v5i3.7300.
- [36] E. D. Madyatmadja, J. Ricardo, B. T. Tanto, and R. Yuristo, "Dataset of 'Challenges and Perceptions of Coretax: A Sentiment Analysis Approach'," Jun. 2025, *Zenodo*. doi: 10.5281/zenodo.15586791.