Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 5 Iss. 3 July 2025, pp: 1037-1048
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Sentiment Analysis of Coretax: A Comparison of Manual, Transformers-Based, and Lexicon-Based Data Labeling on IndoBERT *Performance*

Analisis Sentimen Coretax: Perbandingan Pelabelan Data Manual, *Transformers-Based,* dan *Lexicon-Based* pada Performa IndoBERT

Agnia Suci Rizkia[1*], Wufron[2], Fikri Fahru Roji[3]

[1,3]Program Studi Bisnis Digital, Fakultas Ekonomi, Universitas Garut, Jawa Barat, Indonesia
[2]Program Studi Manajemen, Fakultas Ekonomi, Universitas Garut, Jawa Barat, Indonesia

E-Mail: [1]agniasucirizkia822@gmail.com, [2]wufron@uniga.ac.id, [3]fikri@uniga.ac.id

Received Jun 2nd 2025; Revised Jul 12th 2025; Accepted Jul 21th 2025; Available Online Jul 31th 2025, Published Aug 15th 2025
Corresponding Author: Agnia Suci Rizkia

Abstract

Sentiment analysis of public opinion on social media presents significant challenges due to informal language and large data volume. This study aims to evaluate the impact of five labeling approaches manual, IndoBERT , IndoBERT weet, RoBERTa , and InSet Lexicon on the performance of the Indonesian Bidirectional Encoder Representations from Transformers (IndoBERT) model in classifying sentiments related to the Coretax issue. A total of 8.035 tweets were collected, processed, and labeled using each method. The labeled datasets were then used to retrain the IndoBERT model and evaluated using accuracy, F1-Score, confusion matrix, and Receiver Operating Characteristic-Area Under the Curve (ROC-AUC) curve. Results show that Indonesian Bidirectional Encoder Representations from Transformers for Tweet (IndoBERTweet) achieved the highest metric performance F1-Score (0.9802) but suffered from a strong dominance of the neutral class, indicating a potential model bias towards that class or significant data imbalance. Manual labeling produced a more balanced class distribution despite lower performance F1-Score (0.8684), while Robustly Optimized BERT Pretraining Approach (RoBERTa) demonstrated the best balance between metric performance and label distribution. In contrast, InSet Lexicon and IndoBERT tended to overpredict specific sentiment classes. The study concludes that labeling effectiveness should not be assessed solely by metric scores, but also by the fairness and balance of class distribution to produce reliable and generalizable models.

Keyword: Coretax, Data Labeling, Indobert, *Sentiment Analysis, Transformers*

Abstrak

Analisis sentimen terhadap opini publik di media sosial menjadi tantangan signifikan karena kompleksitas bahasa informal dan volume data yang besar. Penelitian ini bertujuan untuk mengevaluasi pengaruh lima pendekatan pelabelan data manual, IndoBERT , IndoBERT *weet*, RoBERTa , dan InSet Lexicon terhadap performa model *Indonesian Bidirectional Encoder Representations from Transformers* (IndoBERT) dalam klasifikasi sentimen terkait isu Coretax. Sebanyak 8.035 tweet dikumpulkan, diproses, dan dilabeli menggunakan masing-masing pendekatan. Dataset hasil pelabelan kemudian digunakan untuk melatih ulang model IndoBERT, yang dievaluasi menggunakan metrik akurasi, F1-score, confusion matrix, dan kurva *Receiver Operating Characteristic-Area Under the Curve* (ROC-AUC). Hasil menunjukkan bahwa pelabelan otomatis menggunakan *Indonesian Bidirectional Encoder Representations from Transformers for Tweet* (IndoBERTweet) menghasilkan metrik tertinggi *F1-Score* (0,9802), tetapi mengalami dominasi kelas netral yang menunjukkan overfitting. Pelabelan manual menghasilkan distribusi kelas yang lebih merata meskipun dengan metrik lebih rendah *F1-Score* (0,8684), sedangkan *Robustly Optimized BERT Pretraining Approach* (RoBERTa) menunjukkan keseimbangan terbaik antara performa metrik dan distribusi label. InSet Lexicon dan IndoBERT menunjukkan kecenderungan bias terhadap kelas tertentu. Simpulan dari penelitian ini menegaskan bahwa efektivitas pelabelan tidak hanya ditentukan oleh skor metrik, tetapi juga oleh distribusi kelas yang seimbang untuk menghasilkan model yang adil dan dapat digeneralisasi.

Kata Kunci: Analisis Sentiment, Coretax, Indobert, Pelabelan Data, *Transformer*

1. PENDAHULUAN

Perkembangan teknologi digital telah mendorong perubahan besar dalam cara masyarakat berpartisipasi serta berinteraksi dalam diskusi kebijakan publik [1], [2]. Media sosial, sebagai salah satu manifestasi utama dari transformasi ini, telah menjadi arena terbuka bagi masyarakat untuk menyampaikan opini secara langsung dan spontan. Di Indonesia, salah satu kebijakan yang menuai perhatian besar adalah Coretax, sebuah sistem administrasi perpajakan terintegrasi yang dirancang untuk meningkatkan efisiensi, transparansi, dan kepatuhan wajib pajak [3], [4], [5], [6], [7], [8]. Sejak wacana peluncurannya, Coretax telah menjadi topik perbincangan luas di berbagai platform media sosial, khususnya *Twitter/X*, yang dikenal dengan dinamika opini publiknya yang tinggi dan kemampuannya merekam beragam tanggapan, mulai dari apresiasi hingga kritik tajam [9].

Karakteristik unik dari komunikasi di media sosial, seperti penggunaan bahasa informal, slang, sarkasme, dan singkatan, menjadi tantangan tersendiri dalam upaya memahami sentimen publik secara akurat [10]. Untuk mengatasi tantangan ini, analisis sentimen—sebuah metode komputasional untuk mengidentifikasi dan mengkategorikan opini yang diekspresikan dalam teks—telah menjadi pendekatan yang populer [11]. Salah satu tahapan paling krusial dan fundamental dalam analisis sentimen adalah pelabelan data, yaitu proses pemberian label sentimen (misalnya, positif, negatif, atau netral) pada data teks yang akan digunakan untuk melatih model klasifikasi [12].

Secara tradisional, pelabelan data dilakukan secara manual oleh anotator manusia. Metode ini dianggap sebagai standar emas (*gold standard*) karena kemampuannya dalam memahami konteks, nuansa, dan sarkasme yang seringkali sulit ditangkap oleh mesin. Namun, pelabelan manual membutuhkan sumber daya yang besar, baik dari segi waktu maupun biaya, sehingga menjadi tidak praktis untuk dataset berskala besar yang umum dijumpai dalam analisis media sosial [13], [14]. Sebagai alternatif, berbagai metode pelabelan otomatis telah dikembangkan. Salah satu pendekatan yang populer adalah metode berbasis leksikon (*lexicon-based*), yang menggunakan kamus kata-kata dengan skor polaritas sentimen yang telah ditentukan sebelumnya. Di Indonesia, InSet Lexicon adalah salah satu sumber daya leksikon yang sering digunakan dan terbukti cukup efektif untuk tugas-tugas analisis sentimen dasar [15], [16]. Meskipun mudah diimplementasikan, pendekatan ini memiliki keterbatasan dalam menangani kata-kata yang maknanya bergantung pada konteks kalimat.

Dalam beberapa tahun terakhir, model berbasis *transformer*, seperti *Bidirectional Encoder Representations from Transformers* (BERT), telah merevolusi bidang *Natural Language Processing* (NLP) dengan menunjukkan performa canggih dalam memahami konteks bahasa. Untuk Bahasa Indonesia, model seperti IndoBERT dan IndoBERT *weet* telah dikembangkan secara khusus dan menunjukkan keunggulan dalam berbagai tugas, termasuk analisis sentimen pada teks media sosial yang kompleks [17], [18], [19]. Model lain seperti *Robustly Optimized BERT Pretraining Approach* (RoBERTa) juga telah menunjukkan performa yang kuat dalam berbagai bahasa dan domain [20]. Kemampuan model-model ini untuk melakukan pelabelan *zero-shot* (melabeli data tanpa pelatihan spesifik pada domain target) menawarkan solusi yang efisien dan canggih.

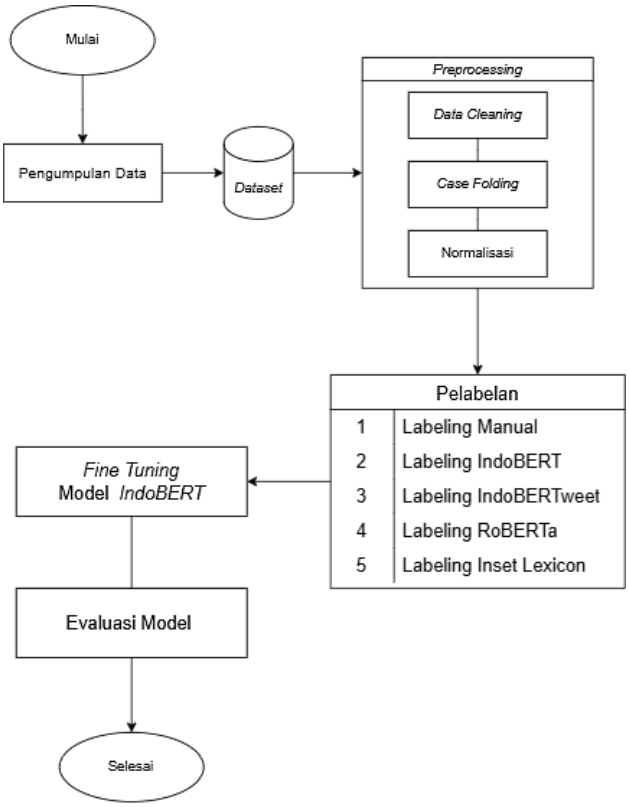
Meskipun berbagai penelitian telah membandingkan metode pelabelan manual dengan otomatis [21], atau membandingkan berbagai algoritma klasifikasi [22], perbandingan yang komprehensif antara efektivitas pelabelan manual, berbasis leksikon, dan berbagai model *transformer* (IndoBERT , IndoBERTweet, RoBERTa) dalam satu kerangka penelitian yang sama untuk konteks kebijakan publik di Indonesia masih sangat terbatas. Kesenjangan ini menjadi krusial karena pilihan metode pelabelan secara langsung mempengaruhi kualitas dataset pelatihan, yang pada gilirannya akan menentukan performa, keadilan (*fairness*), dan kemampuan generalisasi dari model sentimen yang dihasilkan. Tanpa pemahaman yang jelas mengenai kelebihan dan kekurangan masing-masing pendekatan pada domain spesifik seperti opini kebijakan publik, praktisi dan peneliti berisiko memilih metode yang kurang optimal, yang dapat mengarah pada kesimpulan yang bias atau tidak akurat mengenai sentimen publik.

Oleh karena itu, penelitian ini bertujuan untuk mengisi kesenjangan tersebut dengan mengevaluasi dan membandingkan secara sistematis lima pendekatan pelabelan data sentimen yang berbeda terkait isu Coretax. Kelima pendekatan tersebut adalah pelabelan manual, pelabelan berbasis leksikon menggunakan InSet Lexicon, serta pelabelan *zero-shot* menggunakan tiga model *transformer*: IndoBERT , IndoBERTweet, dan RoBERTa . Dataset yang dihasilkan dari masing-masing pendekatan pelabelan ini kemudian digunakan untuk melatih ulang (*fine-tuning*) model IndoBERT yang sama. Performa model dievaluasi secara komprehensif menggunakan metrik akurasi, *F1-score, confusion matrix*, dan kurva ROC-AUC. Penelitian ini diharapkan dapat memberikan wawasan mendalam mengenai strategi pelabelan data yang paling efektif dan efisien untuk menganalisis opini publik berbahasa Indonesia di media sosial, khususnya dalam konteks isu kebijakan nasional yang dinamis.

2. METODE PENELITIAN

Penelitian ini mengadopsi pendekatan eksperimental untuk membandingkan dampak berbagai strategi pelabelan data terhadap performa model analisis sentimen IndoBERT dalam konteks opini publik mengenai

Coretax. Fokus utama adalah pada klasifikasi sentimen ke dalam tiga kategori: positif, negatif, dan netral. Metodologi penelitian ini dirancang untuk secara sistematis mengevaluasi bagaimana perbedaan dalam proses pelabelan data memengaruhi kemampuan model IndoBERT dalam mengidentifikasi sentimen pada teks berbahasa Indonesia yang informal dan bervariasi. Tahapan penelitian ini disusun mengikuti alur visual dalam Gambar 1, dengan langkah-langkah utama meliputi pengumpulan data, pra-pemrosesan data, implementasi berbagai metode pelabelan data, *fine-tuning* model IndoBERT, dan evaluasi komparatif.



Gambar 1. Diagram Alur Tahapan Penelitian

2.1. Pengumpulan dan Pra-pemrosesan Data

Data opini publik dikumpulkan dari platform media sosial *Twitter/ X*, yang dikenal sebagai repositori kaya akan ekspresi sentimen spontan dan real-time terhadap isu-isu kontemporer [10]. Pengambilan data (*data crawling*) dilakukan menggunakan tool *Tweet Harvest* yang terhubung dengan *API Twitter*, dengan fokus pada kata kunci "coretax" atau "core_tax" dalam rentang waktu Oktober 2024 hingga Maret 2025. Pembatasan pada tweet berbahasa Indonesia diterapkan untuk memastikan relevansi linguistik. Proses ini menghasilkan 8.571 tweet yang kemudian disimpan dalam format CSV untuk diproses lebih lanjut. Teknik *crawling* ini telah terbukti efektif dalam studi serupa pada analisis sentimen isu nasional di media sosial [10], [12], [22] serta pada pengambilan data *Twitter* atribut lengkap melalui *API* dan *tools* terkait [23], [24].

2.2. *Preprocessing* Data

Setelah tahap pengumpulan, dilakukan pemilahan dan pembersihan data, menghasilkan tweet yang memenuhi kriteria untuk analisis lebih lanjut. Pemilahan ini mempertimbangkan relevansi konten, kejelasan bahasa, dan keberhasilan dalam tahap pra-pemrosesan.

Tahap *preprocessing* data bertujuan untuk membersihkan dan menormalisasi teks, menjadikannya siap untuk analisis sentimen. Tahapan yang diterapkan meliputi:

1. *Case Folding*: Mengubah seluruh teks menjadi huruf kecil untuk menyeragamkan data.
2. *Data Cleaning*: Menghapus elemen-elemen tidak relevan seperti URL, mention (@username), hashtag (#), tanda baca, emoji, dan spasi berlebih. Duplikasi tweet juga dieliminasi.
3. Normalisasi: Mengganti kata-kata tidak baku atau slang ke bentuk standar Bahasa Indonesia. Proses ini krusial mengingat karakteristik bahasa di media sosial.

Penelitian ini tidak menerapkan stemming dan tokenisasi pada tahap pra-pemrosesan, sejalan dengan temuan [25] yang menunjukkan bahwa model IndoBERT dapat memberikan hasil yang lebih optimal tanpa tahapan tersebut, karena model *transformer* mampu menangani variasi leksikal secara internal [25]. Dengan

melakukan tahapan *preprocessing* ini, data yang dihasilkan menjadi lebih bersih, terstruktur, dan siap digunakan dalam analisis sentimen yang lebih baik [24].

2.3. Pelabelan Data

Untuk mengeksplorasi dampak metode pelabelan terhadap performa model, penelitian ini mengimplementasikan lima pendekatan pelabelan data yang berbeda pada dataset Coretax. Setiap pendekatan menghasilkan dataset berlabel yang kemudian akan digunakan secara independen untuk melatih model IndoBERT . Kelima pendekatan tersebut adalah:

1. Pelabelan Manual: Sebanyak 8.035 dataset dilabeli secara manual oleh anotator manusia yang terlatih. Anotator mengklasifikasikan setiap tweet ke dalam kategori positif, negatif, atau netral berdasarkan pemahaman kontekstual [26]. Proses ini melibatkan pedoman pelabelan yang ketat untuk memastikan konsistensi dan kualitas *gold standard*. Dataset berlabel manual ini berfungsi sebagai tolak ukur utama untuk memvalidasi akurasi metode otomatis.

2. Pelabelan Berbasis Lexicon (InSet Lexicon): Pendekatan ini memanfaatkan InSet Lexicon, sebuah kamus sentimen berbahasa Indonesia yang telah dikembangkan sebelumnya [19]. Setiap kata dalam teks dicocokkan dengan entri dalam leksikon, dan skor sentimen agregat dihitung untuk setiap tweet [27]. Ambang batas tertentu digunakan untuk mengklasifikasikan tweet sebagai positif, negatif, atau netral. Metode ini cepat dan tidak memerlukan pelatihan, namun memiliki keterbatasan dalam menangani nuansa kontekstual dan ekspresi non-literal [28].

3. Pelabelan *Zero-shot* Berbasis *Transformer (*IndoBERT , IndoBERTweet, RoBERTa): Tiga model *transformer* yang telah dilatih sebelumnya (pre-trained) untuk bahasa Indonesia digunakan dalam mode *zero-shot* classification [19], [29], [30]. Ini berarti model model tersebut langsung digunakan untuk memprediksi sentimen tanpa fine-tuning tambahan pada dataset Coretax. Model yang digunakan adalah:
    a. IndoBERT : Model BERT yang dilatih pada korpus teks umum berbahasa Indonesia .
    b. IndoBERTweet: Varian IndoBERT yang secara khusus dilatih pada data Twitter berbahasa Indonesia, sehingga lebih adaptif terhadap karakteristik bahasa media sosial.
    c. RoBERTa : Model *Transformer* yang dioptimalkan dari BERT, dikenal karena performanya yang kuat dalam berbagai tugas NLP. Meskipun RoBERTa asli dilatih pada bahasa Inggris, terdapat versi yang telah diadaptasi atau dilatih pada bahasa Indonesia yang dapat digunakan untuk tujuan ini [31].

Masing-masing model ini menerima teks tweet sebagai input dan menghasilkan probabilitas untuk setiap kelas sentimen (positif, negatif, netral). Kelas dengan probabilitas tertinggi kemudian ditetapkan sebagai label sentimen untuk tweet tersebut. Pendekatan *zero-shot* ini mengevaluasi kemampuan intrinsik model *transformer* dalam memahami sentimen tanpa paparan data spesifik domain

2.4. Fine Tuning Model IndoBERT

Setelah dataset Coretax dilabeli menggunakan kelima pendekatan di atas, setiap dataset berlabel digunakan secara terpisah untuk melatih ulang (*fine-tune*) model IndoBERT yang sama. Ini merupakan aspek krusial dari penelitian ini untuk mengisolasi dan membandingkan dampak langsung dari kualitas dan karakteristik pelabelan data terhadap performa model akhir. Model IndoBERT dipilih sebagai model target karena kemampuannya yang telah terbukti dalam memahami konteks linguistik Bahasa Indonesia secara mendalam dan unggul dalam berbagai tugas NLP, termasuk klasifikasi sentimen sosial-politik [10], [19], [32]. Setiap dataset dibagi menggunakan rasio 80% untuk pelatihan dan 20% untuk pengujian [21], [33]. Proses fine-tuning dilakukan pada lingkungan Google Colab, memanfaatkan pustaka *Transformer*s dari Hugging Face [21]. Parameter hyperparameter seperti learning rate, ukuran batch, dan jumlah epoch dioptimalkan untuk setiap dataset untuk memastikan performa terbaik dari model yang dilatih [34].

2.5. Evaluasi Model

Tahap akhir adalah evaluasi model, evaluasi menyeluruh dilakukan menggunakan dataset uji untuk mengukur kinerjanya [35]. Evaluasi dilakukan dengan menghitung metrik performa standar dalam klasifikasi teks, yaitu:

1. Akurasi
   Akurasi mengacu pada rasio instansi yang diprediksi dengan benar terhadap jumlah total observasi dalam dataset.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Dalam evaluasi model klasifikasi, *True Positive* (TP) adalah instansi positif yang teridentifikasi benar oleh model. *True Negative* (TN) adalah instansi negatif yang diprediksi benar. Kesalahan terjadi pada *False Positive* (FP), yaitu instansi negatif yang salah diklasifikasikan sebagai positif, dan *False Negative* (FN), yaitu instansi positif yang gagal diidentifikasi oleh model. Pemahaman keempat metrik ini penting untuk menilai kinerja model secara komprehensif.

2. Presisi
Presisi mengukur proporsi observasi positif yang diprediksi dengan benar terhadap total prediksi positif.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Presisi yang tinggi menunjukkan tingkat *False Positive* (FP) yang lebih rendah, menjadikannya metrik penting untuk kasus di mana *False Positive*s (FP) harus diminimalkan [36].

3. *Recall* (Sensitivitas)

$$Recall = \frac{TP}{TP+FN} \quad (1)$$

Recall mengacu pada efektivitas model dalam mengambil semua sampel positif aktual dalam data uji. Nilai *Recall* yang tinggi menunjukkan bahwa model menangkap sebagian besar instansi positif, mengurangi *False Negative*s (FN) [37].

4. *F1-Score*
F1-Score menangkap keseimbangan antara presisi dan *Recall*, menjadikannya indikator komprehensif kinerja klasifikasi.

$$F1\text{-}Score = 2x\frac{\text{Presisi}\times Recall}{\text{Presisi}+Recall} \quad (1)$$

Metrik ini sangat berguna ketika dataset tidak seimbang [38].

5. Kurva ROC-AUC
Dengan memplot *True Positive (TP)* Rate terhadap *False Positive* (FP) Rate di berbagai ambang batas, kurva ROC menawarkan wawasan tentang kinerja diskriminatif model. Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC adalah grafik dua dimensi dengan false positif sebagai garis horizontal dan true positif sebagai garis vertikal. Pedoman umum untuk mengklasifikasikan keakuratan pengujian menggunakan AUC [22].

$$\text{TPR} = \frac{TP}{TP+FN}, \quad \text{FPR} = \frac{FP}{FP+TN} \quad (1)$$

Kurva ROC yang lebih tinggi menunjukkan kinerja model yang lebih baik dalam membedakan antara kelas [39].

3. TINJAUAN PUSTAKA

Analisis sentimen dalam konteks bahasa Indonesia, khususnya pada data media sosial yang informal, menghadapi tantangan signifikan akibat penggunaan bahasa gaul, singkatan, dan sarkasme [10]. Ketersediaan dan kualitas data berlabel menjadi fondasi krusial untuk membangun model analisis sentimen yang efektif [12]. Secara tradisional, pelabelan manual oleh anotator manusia dianggap sebagai standar emas karena kemampuannya menangkap nuansa dan konteks yang kompleks. Namun, metode ini sangat tidak efisien dan memakan biaya serta waktu yang besar untuk dataset berskala besar [13], [14]. Alternatifnya, pendekatan berbasis leksikon, seperti InSet Lexicon, menawarkan efisiensi tetapi seringkali gagal memahami ambiguitas dan sarkasme, menghasilkan label yang tidak akurat dan bias [15], [16]. Kesenjangan antara akurasi manual dan efisiensi otomatis ini mendorong eksplorasi strategi pelabelan yang lebih canggih, terutama dengan munculnya model bahasa berbasis *transformer*.

Revolusi *transformer* telah mengubah lanskap Pemrosesan Bahasa Alami (NLP) secara fundamental. Model *Bidirectional Encoder Representations from Transformer*s (BERT) menjadi terobosan karena kemampuannya memahami konteks kata secara dua arah, menghasilkan representasi bahasa yang kaya dan kontekstual [17]. Kemampuan ini telah mendorong kinerja *state-of-the-art* dalam berbagai tugas NLP, termasuk analisis sentimen. Untuk mengatasi kekhasan linguistik Bahasa Indonesia, model *transformer*

spesifik seperti IndoBERT telah dikembangkan, dilatih pada korpus data lokal yang masif untuk menangkap struktur gramatikal dan semantik yang relevan [18]. Penelitian terbaru secara konsisten menunjukkan bahwa IndoBERT dan variannya, seperti IndoBERTweet yang dioptimalkan untuk media sosial, mencapai kinerja canggih dalam analisis sentimen berbahasa Indonesia, melampaui metode tradisional [40], [41]. Model lain seperti RoBERTa juga menunjukkan kinerja kuat dalam berbagai bahasa dan domain [42]. Model-model *transformer* ini tidak hanya berfungsi sebagai alat, melainkan menjadi subjek utama dalam pengembangan dan perbandingan pendekatan pelabelan data otomatis yang inovatif.

Penelitian ini secara khusus mengevaluasi lima pendekatan pelabelan data yang berbeda: pelabelan manual, pelabelan berbasis leksikon, dan tiga metode pelabelan *zero-shot* yang memanfaatkan kekuatan model *transformer* (IndoBERT, IndoBERTweet, dan RoBERTa). Pelabelan manual menjadi tolak ukur kualitas dan keakuratan kontekstual [13], [14]. Pelabelan berbasis leksikon menawarkan kecepatan namun terbatas dalam menangani konteks dan ambiguitas [15], [16], [43]. Sementara itu, pelabelan *zero-shot* dengan model *transformer* menawarkan solusi efisien dan canggih, memanfaatkan pengetahuan pra-pelatihan untuk mengklasifikasikan teks tanpa pelatihan eksplisit pada domain target [44].

Meskipun banyak penelitian telah membandingkan model *transformer* dengan metode lain [21], [22], masih terdapat kesenjangan literatur yang signifikan terkait perbandingan sistematis dari berbagai strategi pelabelan data—khususnya manual, leksikon, dan berbagai model *transformer* dalam mode *zero-shot*—dalam satu kerangka penelitian yang koheren untuk bahasa Indonesia. Studi yang ada seringkali hanya membandingkan dua metode pelabelan atau berfokus pada performa model akhir tanpa menganalisis secara mendalam bagaimana kualitas dataset yang dilabeli secara berbeda memengaruhi hasil tersebut. Kesenjangan ini sangat penting karena pilihan metode pelabelan secara langsung berdampak pada kualitas, bias, dan keandalan dataset pelatihan, yang pada akhirnya menentukan performa dan kemampuan generalisasi model sentimen [43]. Oleh karena itu, penelitian ini secara eksplisit mengisi kesenjangan ini dengan menyoroti peran krusial dari metode pelabelan data dan dampaknya terhadap hasil analisis sentimen, khususnya dalam konteks bahasa Indonesia dan isu kebijakan publik.

Dengan demikian, penelitian ini secara eksplisit bertujuan untuk mengisi kesenjangan tersebut dengan secara sistematis mengevaluasi lima pendekatan pelabelan data yang berbeda pada isu kebijakan publik Coretax. Dengan melatih model IndoBERT yang sama pada dataset yang dihasilkan oleh setiap metode pelabelan, kami dapat mengisolasi dan menganalisis dampak dari strategi pelabelan itu sendiri. Kontribusi penelitian ini terletak pada penyediaan analisis komparatif yang mendalam, tidak hanya pada metrik performa akhir, tetapi juga pada distribusi kelas dan potensi bias yang diperkenalkan oleh setiap metode pelabelan. Wawasan ini diharapkan dapat memberikan panduan praktis bagi para peneliti dan praktisi dalam memilih strategi pelabelan yang paling efektif untuk analisis sentimen pada domain spesifik berbahasa Indonesia.

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Penelitian ini mengumpulkan tweet dari *Twitter/X* dengan kata kunci “coretax” dan “core_tax” pada rentang 1 Oktober 2024 hingga 31 Maret 2025. Rentang ini mencakup masa pra-peluncuran (Oktober-November 2024), peluncuran dan reaksi awal (Desember 2024 - Januari 2025), evaluasi awal dampak (Februari 2025), serta periode kritis pelaporan SPT Tahunan (Maret 2025). Dari proses ini berhasil dikumpulkan sebanyak 8.571 tweet berbahasa Indonesia. Namun, setelah dilakukan proses pemilahan dan pembersihan data, hanya 8.035 tweet yang memenuhi kriteria dan digunakan dalam penelitian ini. Pemilahan ini mempertimbangkan aspek relevansi konten, kejelasan bahasa, dan keberhasilan dalam tahap *preprocessing*. Hasil pengumpulan data dapat dilihat pada Tabel 1.

Tabel 1. Hasil Pengumpulan Data

No	Full_Text
1	#KawanPajak Coretax akan segera hadir untuk mempermudah pelayanan pajak! Berbekal sistem yang modern dan terintegrasi Coretax akan memberikan pengalaman baru yang lebih cepat mudah dan efisien dalam menjalankan kewajiban perpajakan. https://t.co/MfoGPvq1Bg
2	Halo #KawanPajak Edukasi Coretax Tahap II telah dibuka. Kelas pajak akan dilaksanakan 2 kali dalam seminggu pada hari Rabu dan Kamis. Buruan daftar jangan sampai ketinggalan. Kelas Pajak ini tidak dipungut biaya. #coretax #taxreform #PajakKitaUntukKita #PajakKuatAPBNSehat https://t.co/cfS5kGITcz
…	…
8035	coretax sialan sialan

4.2. *Preprocessing* Data

Tahapan *preprocessing* meliputi *case folding*, pembersihan simbol dan URL (*cleaning*), serta normalisasi bahasa informal (*slang*) ke dalam bentuk baku. Tabel 2 berikut menyajikan hasil *preprocessing*.

Tabel 2. Hasil *Preprocessing*

Tahapan	Hasil
Original Text	#KawanPajak Coretax akan segera hadir untuk mempermudah pelayanan pajak! Berbekal sistem yang modern dan terintegrasi Coretax akan memberikan pengalaman baru yang lebih cepat mudah dan efisien dalam menjalankan kewajiban perpajakan. https://t.co/MfoGPvq1Bg
Cleaned Text	Coretax akan segera hadir untuk mempermudah pelayanan pajak Berbekal sistem yang modern dan terintegrasi Coretax akan memberikan pengalaman baru yang lebih cepat mudah dan efisien dalam menjalankan kewajiban perpajakan
Lowercased Text	Coretax akan segera hadir untuk mempermudah pelayanan pajak berbekal sistem yang modern dan terintegrasi Coretax akan memberikan pengalaman baru yang lebih cepat mudah dan efisien dalam menjalankan kewajiban perpajakan
Normalized Text	Coretax akan segera hadir untuk mempermudah pelayanan pajak berbekal sistem yang modern dan terintegrasi Coretax akan memberikan pengalaman baru yang lebih cepat mudah dan efisien dalam menjalankan kewajiban perpajakan

4.3. Pelabelan Data

Setelah tahap *preprocessing*, data diberi label sentimen menggunakan lima pendekatan berbeda, yaitu pelabelan manual, tiga model berbasis *transformer* (IndoBERT , IndoBERTweet, dan RoBERTa), serta pendekatan berbasis Lexicon yang menggunakan InSet Lexicon. Tabel 3 berikut menunjukkan distribusi kelas sentimen yang dihasilkan dari setiap metode pelabelan, yang selanjutnya dievaluasi untuk menilai keakuratan dan efektivitas masing-masing pendekatan dalam mengklasifikasikan sentimen data.

Tabel 3. Distribusi Kelas Sentimen

Metode Pelabelan	Positif	Netral	Negatif
Manual	72	5889	2074
IndoBERT	192	7810	33
IndoBERTweet	-	7936	99
RoBERTa	349	4387	2849
InSet Lexicon	599	268	7168

Dari Tabel 3, terlihat bahwa pelabelan manual menghasilkan distribusi yang relatif seimbang antara kelas netral dan negatif, dengan sentimen positif sebagai kelas minoritas. Sebaliknya, model IndoBERT dan IndoBERT *weet* menunjukkan dominasi kelas netral yang sangat kuat, dengan IndoBERT *weet* bahkan tidak mengklasifikasikan satu pun tweet sebagai positif. RoBERTa memberikan distribusi yang lebih seimbang dibandingkan dua model *transformer* lainnya, meskipun masih cenderung pada kelas netral dan negatif. Pendekatan berbasis leksikon (InSet Lexicon) menunjukkan bias yang kuat terhadap kelas negatif, yang mengindikasikan sensitivitas tinggi terhadap kata-kata bernuansa negatif dalam kamus yang digunakan.

4.4. Fine-Tuning dan Evaluasi Model

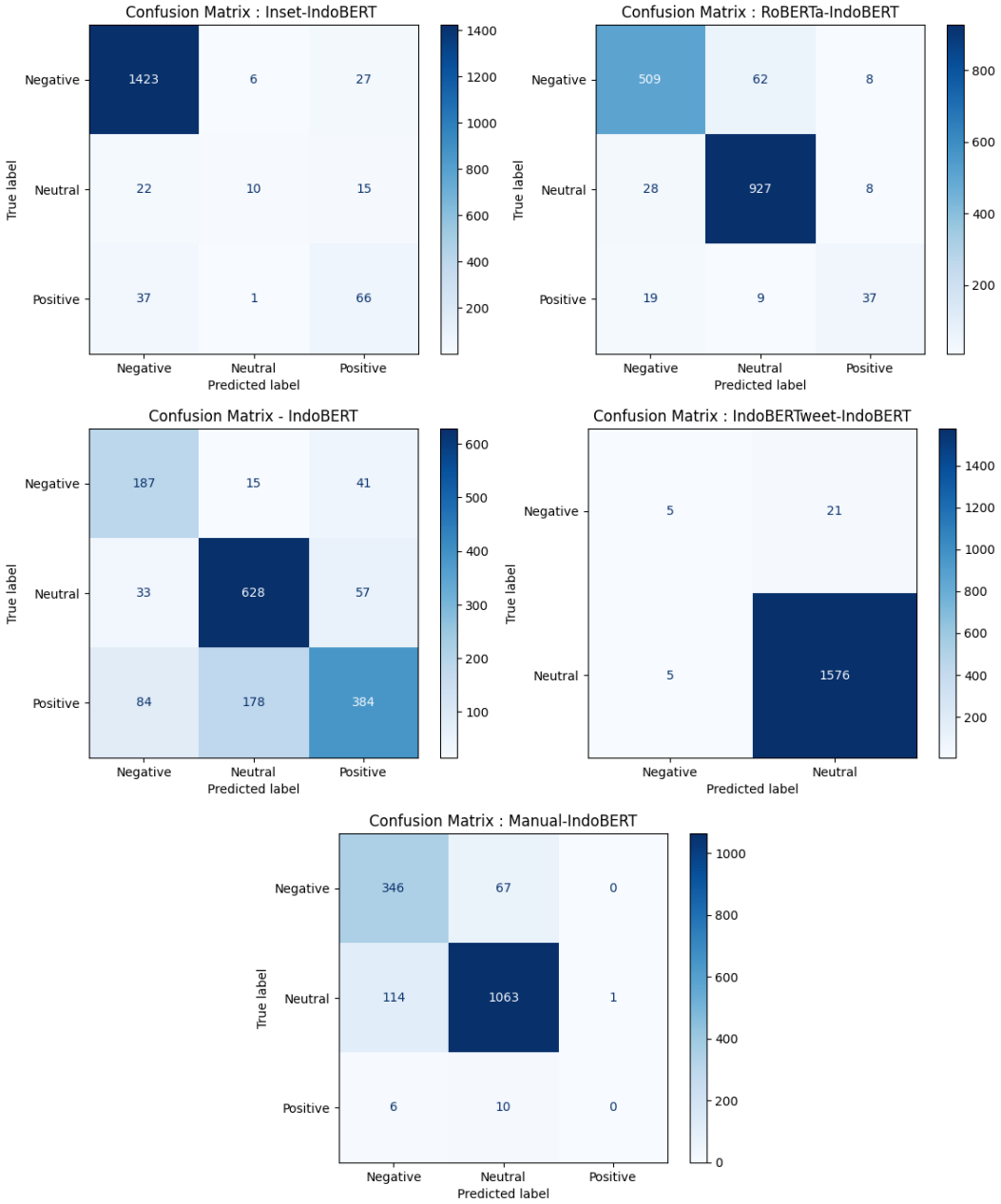
Untuk mengevaluasi dampak metode pelabelan, model IndoBERT dilatih ulang (*fine-tune*d) pada masing-masing dari lima dataset berlabel. Performa model dievaluasi menggunakan metrik akurasi, presisi, *Recall*, dan *F1-score*. Tabel 4 merangkum hasil evaluasi performa model IndoBERT pada kelima dataset tersebut.

Tabel 4. Evaluasi Perform Model IndoBERT

Dataset	Accuracy	Precision	*Recall*	*F1-Score*	Training Time
Manual	0.8724	0.8681	0.8724	0.8684	260.4844
IndoBERT	0.7461	0.7550	0.7461	0.7417	264.5360
IndoBERT *weet*	0.9838	0.9789	0.9838	0.9802	266.8433
RoBERTa	0.9166	0.9147	0.9166	0.9150	254.7875
InSet Lexicon	0.9327	0.9267	0.9327	0.9270	266.3536

Hasil pada Tabel 4 menunjukkan bahwa model yang dilatih pada dataset berlabel IndoBERT *weet* mencapai metrik performa tertinggi (*F1-Score*, 0.9802), diikuti oleh InSet Lexicon (0.9270), dan RoBERTa (0.9150). Model yang dilatih pada dataset berlabel manual menunjukkan performa yang lebih rendah (0.8684), sementara model yang dilatih pada dataset berlabel IndoBERT memiliki performa terendah (0.7417). Namun, metrik performa yang tinggi tidak selalu mencerminkan kualitas model yang baik, terutama jika distribusi kelas sangat tidak seimbang, seperti yang terlihat pada kasus IndoBERTweet.

Untuk analisis yang lebih mendalam, confusion matrix dan kurva ROC-AUC digunakan untuk mengevaluasi kemampuan model dalam membedakan antar kelas sentimen. Gambar 2 menyajikan confusion matrix untuk masing-masing model yang dilatih.



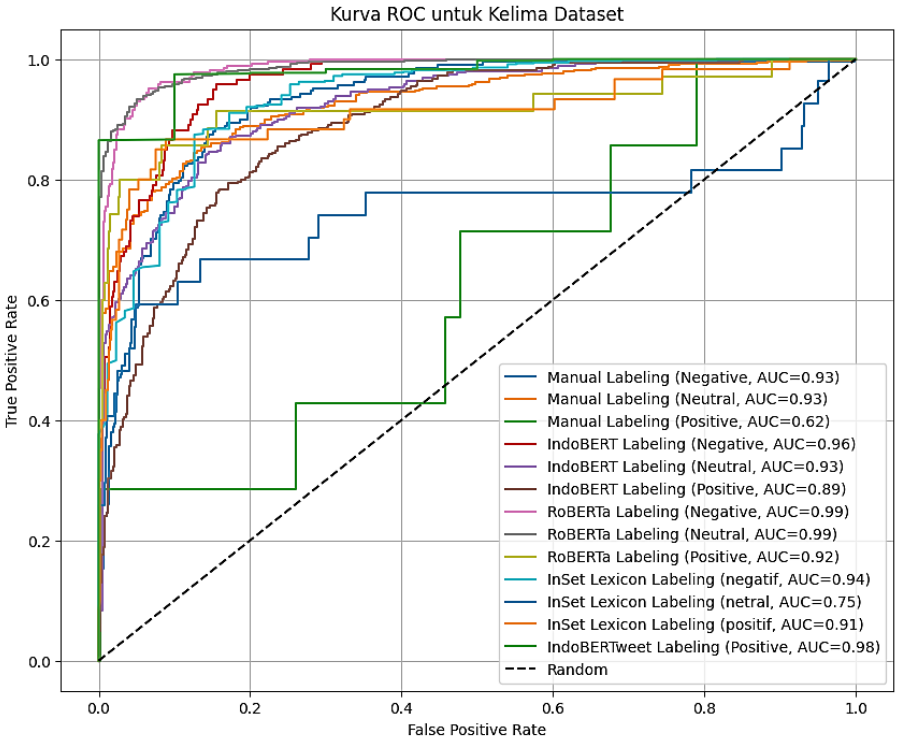
Gambar 2. Confusion Matrix antar Dataset

Confusion matrix menunjukkan bahwa model yang dilatih pada dataset berlabel IndoBERT *weet* sangat cenderung memprediksi kelas netral, yang mengindikasikan overfitting pada kelas mayoritas. Sebaliknya, model yang dilatih pada dataset berlabel manual dan RoBERTa menunjukkan kemampuan yang lebih baik dalam membedakan antara kelas negatif dan netral, meskipun masih kesulitan dengan kelas positif yang minoritas. Model yang dilatih pada dataset berlabel InSet Lexicon menunjukkan bias yang kuat terhadap kelas negatif, sejalan dengan distribusi label pada Tabel 3.

Kurva ROC-AUC pada Gambar 3 memberikan gambaran lebih lanjut tentang kemampuan diskriminatif model. Model yang dilatih pada dataset berlabel RoBERTa menunjukkan nilai AUC yang tinggi dan seimbang di ketiga kelas, mengindikasikan kemampuan generalisasi yang baik. Model yang dilatih pada dataset berlabel manual juga menunjukkan performa yang baik, meskipun sedikit di bawah RoBERTa . Model yang dilatih pada dataset berlabel IndoBERTweet, meskipun memiliki akurasi tinggi, menunjukkan kurva ROC yang kurang ideal karena dominasi kelas netral.

Secara keseluruhan, hasil ini menunjukkan bahwa meskipun pelabelan otomatis dengan IndoBERTweet menghasilkan metrik performa tertinggi, model yang dihasilkan sangat tidak seimbang dan cenderung *overfit*. Pelabelan dengan RoBERTa , di sisi lain, menghasilkan model dengan keseimbangan terbaik antara performa metrik dan kemampuan generalisasi, menjadikannya pendekatan pelabelan otomatis

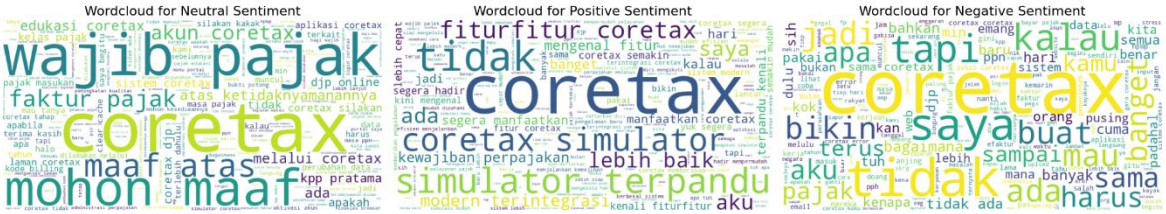
yang paling menjanjikan dalam penelitian ini. Pelabelan manual, meskipun menghasilkan metrik yang lebih rendah, tetap menjadi *gold standard* dalam hal distribusi kelas yang realistis dan interpretasi kontekstual.



Gambar 3. Kurva ROC

4.5. Word Cloud

Untuk memberikan gambaran visual mengenai karakteristik sentimen, dibuat *word cloud* menggunakan data pelabelan RoBERTa karena performa dan distribusi kelas yang seimbang. *Word cloud* ini mewakili opini publik tentang Coretax berdasarkan sentimen netral, positif, dan negatif. Lihat Gambar 4.



Gambar 4. Word cloud sentiment Neutral, Positive, dan Negative

Visualisasi word cloud dari dataset pelabelan RoBERTa memberikan gambaran komprehensif tentang persepsi publik terhadap Coretax. Pada sentimen Netral, kata-kata dominan seperti "wajib pajak", "coretax", "pajak", "faktur", dan "mohon" menyoroti fokus pada aspek administratif, kewajiban, dan permintaan klarifikasi terkait sistem perpajakan. Sementara itu, sentimen Positif didominasi oleh kata-kata seperti "coretax", "fitur-fitur", "simulator", "terpandu", dan "lebih baik", yang mencerminkan apresiasi terhadap inovasi, kemudahan penggunaan, dan potensi perbaikan sistem. Sebaliknya, sentimen Negatif menampilkan kata-kata seperti "coretax", "tidak", "tapi", "kalau", "bikin", dan "pajak", yang mengindikasikan keluhan, kritik, ketidakpuasan, atau masalah teknis yang dirasakan pengguna.

Secara keseluruhan, temuan ini menegaskan bahwa evaluasi performa model tidak cukup hanya berdasarkan metrik agregat, melainkan harus mempertimbangkan distribusi label, keseimbangan klasifikasi, dan konteks penggunaan yang tercermin dalam kata-kata kunci. Meskipun pelabelan manual tetap menjadi tolok ukur kualitas data, pendekatan otomatis seperti RoBERTa menunjukkan potensi besar sebagai solusi efisien dengan hasil yang kompetitif dan stabil, sangat relevan untuk analisis sentimen yang akurat dalam konteks kebijakan dan ekonomi.

5. KESIMPULAN

Penelitian ini menunjukkan bahwa metode pelabelan data memiliki pengaruh signifikan terhadap performa model klasifikasi sentimen berbasis IndoBERT dalam menganalisis opini publik terkait implementasi sistem perpajakan Coretax. Pelabelan otomatis menggunakan IndoBERTweet memberikan metrik evaluasi tertinggi, namun distribusi kelas yang sangat tidak seimbang menyebabkan risiko overfitting dan mengurangi kemampuan model mengenali sentimen positif dan negatif secara proporsional. Sebaliknya, pelabelan manual menghasilkan distribusi sentimen yang lebih seimbang dan representatif meskipun metriknya lebih rendah, sedangkan metode RoBERTa menjadi alternatif pelabelan otomatis yang optimal dengan keseimbangan terbaik antara akurasi dan distribusi kelas. Selain itu, analisis sentimen melalui visualisasi *word cloud* memperlihatkan pola persepsi masyarakat yang beragam, mulai dari permintaan informasi hingga kritik terhadap aspek teknis dan aksesibilitas Coretax. Temuan ini memberikan wawasan tambahan untuk mendukung pengembangan kebijakan dan layanan perpajakan digital. Secara keseluruhan, studi ini menegaskan pentingnya pemilihan metode pelabelan yang tepat guna meningkatkan akurasi analisis sentimen sekaligus memberikan kontribusi pemahaman yang lebih baik mengenai opini publik terhadap Coretax.

Implikasi praktis dari penelitian ini adalah bahwa dalam pengembangan sistem analisis sentimen untuk isu-isu kebijakan publik, tidak cukup hanya berfokus pada metrik performa agregat. Keseimbangan distribusi kelas dan kemampuan model untuk menggeneralisasi sentimen di seluruh kategori harus menjadi pertimbangan utama. Pelabelan manual, meskipun memakan waktu, tetap menjadi tolok ukur untuk kualitas data yang seimbang, sementara RoBERTa menawarkan solusi otomatis yang menjanjikan untuk mencapai keseimbangan tersebut.

REFERENSI

[1] J. Mannayong, M. R. S, H. Herling, And M. Faisal, "Transformasi Digital Dan Partisipasi Masyarakat: Mewujudkan Keterlibatan Publik Yang Lebih Aktif," *J. Adm. Publik*, Vol. 20, No. 1, Pp. 53–75, Jun. 2024, Doi: 10.52316/Jap.V20i1.260.

[2] Y. O. Nainggolan, E. S. Sihombing, And S. Gulo, "Media Sebagai Agen Perubahan : Studi Peran Media Dalam Pemberdayaan Masyarakat Dalam Era Digital," Vol. 01, No. 03, Pp. 156–163, 2025.

[3] F. Arianty, "Implementation Challenges And Opportunities Coretax Administration System On The Efficiency Of Tax Administration," *J. Vokasi Indones.*, Vol. 12, No. 2, P. 98, Dec. 2024, Doi: 10.7454/Jvi.V12i2.1227.

[4] H. T. Ilyas, S. D. Devano, And S. H. Herdianti, "The Effect Of Tax Planning And The Implementation Of The Core Tax Administration System On Taxpayer Compliance," *Eduvest - J. Univers. Stud.*, Vol. 5, No. 3, Pp. 3326–3338, Mar. 2025, Doi: 10.59188/Eduvest.V5i3.44798.

[5] C. Korat And A. Munandar, "Penerapan Core Tax Administration System (Ctas) Langkah Meningkatkan Kepatuhan Perpajakan Di Indonesia," *J. Ris. Akunt. Politala*, Vol. 8, No. 1, Pp. 16–29, Mar. 2025, Doi: 10.34128/Jra.V8i1.453.

[6] D. T. Della Nabila, L. T. Jumaidi, B. Anggun, H. Lestari, And M. Firmansyah, "Jurnal Abdimas : Pengabdian Dan Pengembangan Masyarakat Penyederhanaan Proses Perpajakan Melalui Penggunaan Core Tax Administration System Sebagai Sistem Pajak Terbaru," Vol. 6, No. 2, Pp. 89–93, 2024, Doi: Https://Doi.Org/10.30630/Jppm.V6i2.1635.

[7] T. Purnomo, A. Sadiqin, And R. Arvita, "Analisis Implementasi Aplikasi Pajak Coretax Dalam Meningkatkan Kepatuhan Dan Efisiensi Pelaporan Pajak Di Indonesia," Vol. 3, No. 2, Pp. 114–118, 2025, Doi: Https://Doi.Org/10.63200/Jebmass.V3i2.181.

[8] R. Rahmawati And N. Nurcahyani, "Coretax System Dalam Upaya Reformasi Administrasi Perpajakan, Apa Urgensinya?," *J. Financ.*, Vol. 6, No. 1, Pp. 1–8, Jan. 2025, Doi: 10.51977/Financia.V6i1.1980.

[9] D. T. Attaulah And D. Soyusiawaty, "Analisis Sentimen Program Makan Siang Gratis Di Twitter/X Menggunakan Metode Bi-Lstm," *Edumatic J. Pendidik. Inform.*, Vol. 9, No. 1, Pp. 294–303, Apr. 2025, Doi: 10.29408/Edumatic.V9i1.29725.

[10] A. D. H. Setiawan And W. Maharani, "Understanding Public Sentiments On The 2024 Presidential Election Through Bert-Powered Analysis," *Edumatic J. Pendidik. Inform.*, Vol. 9, No. 1, Pp. 89–98, Apr. 2025, Doi: 10.29408/Edumatic.V9i1.29267.

[11] T. M. Permata Aulia, N. Arifin, And R. Mayasari, "Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinisasi Covid-19," *Sintech (Science Inf. Technol. J.*, Vol. 4, No. 2, Pp. 139–145, Oct. 2021, Doi: 10.31598/Sintechjournal.V4i2.762.

[12] Y. Asri, W. N. Suliyanti, D. Kuswardani, And M. Fajri, "Pelabelan Otomatis Lexicon Vader Dan Klasifikasi Naive Bayes Dalam Menganalisis Sentimen Data Ulasan Pln Mobile," *Petir*, Vol. 15, No. 2, Pp. 264–275, Nov. 2022, Doi: 10.33322/Petir.V15i2.1733.

[13] P. Ayuningtyas, S. Khomsah, And S. Sudianto, "Pelabelan Sentimen Berbasis Semi-Supervised Learning Menggunakan Algoritma Lstm Dan Gru," *Jiska (Jurnal Inform. Sunan Kalijaga)*, Vol. 9,

No. 3, Pp. 217–229, Sep. 2024, Doi: 10.14421/Jiska.2024.9.3.217-229.
[14] M. Jafarlou And M. M. Kubek, “Reducing Labeling Costs In Sentiment Analysis Via Semi-Supervised Learning,” In *The 2024 8th International Conference On Natural Language Processing And Information Retrieval (Nlpir 2024), Okayama, Japan, 2024.*, Oct. 2024, Pp. 1–12.
[15] H. Firda, P. Putra, N. R. Oktadini, P. E. Sevtiyuni, And A. Meiriza, “Comparison Of Rating-Based And InSet Lexicon-Based Labeling In Sentiment Analysis Using Svm (Case Study: Gobiz Application Reviews On Google Play Store),” *Sistemasi*, Vol. 14, No. 2, P. 516, Mar. 2025, Doi: 10.32520/Stmsi.V14i2.4795.
[16] D. Musfiroh, U. Khaira, P. E. P. Utomo, And T. Suratno, “Analisis Sentimen Terhadap Perkuliahan Daring Di Indonesia Dari Twitter Dataset Menggunakan InSet Lexicon,” *Malcom Indones. J. Mach. Learn. Comput. Sci.*, Vol. 1, No. 1, Pp. 24–33, Mar. 2021, Doi: 10.57152/Malcom.V1i1.20.
[17] J. F. Kusuma And A. Chowanda, “Indonesian Hate Speech Detection Using Indobertweet And Bilstm On Twitter,” *Joiv Int. J. Informatics Vis.*, Vol. 7, No. 3, Pp. 773–780, Sep. 2023, Doi: 10.30630/Joiv.7.3.1035.
[18] H. Imaduddin, F. Y. A’la, And Y. S. Nugroho, “Sentiment Analysis In Indonesian Healthcare Applications Using Indobert Approach,” *Int. J. Adv. Comput. Sci. Appl.*, Vol. 14, No. 8, Pp. 113–117, 2023, Doi: 10.14569/Ijacsa.2023.0140813.
[19] F. Koto, J. H. Lau, And T. Baldwin, “Indobertweet: A Pretrained Language Model For Indonesian Twitter With Effective Domain-Specific Vocabulary Initialization,” In *Proceedings Of The 2021 Conference On Empirical Methods In Natural Language Processing*, Stroudsburg, Pa, Usa: Association For Computational Linguistics, 2021, Pp. 10660–10668. Doi: 10.18653/V1/2021.Emnlp-Main.833.
[20] N. A. Semary, W. Ahmed, K. Amin, P. Pławiak, And M. Hammad, “Improving Sentiment Classification Using A RoBERTa -Based Hybrid Model,” *Front. Hum. Neurosci.*, Vol. 17, No. December, Pp. 1–10, Dec. 2023, Doi: 10.3389/Fnhum.2023.1292010.
[21] D. C. Febrianto, M. A. Fitriani, M. Afrad, And M. A. Khadija, “Aspect Based Sentiment Analysis Menggunakan Indobert Model Terhadap Review Pengunjung Objek Wisata Baturraden,” *Melek It Inf. Technol. J.*, Vol. 10, No. 2, Pp. 157–166, Dec. 2024, Doi: 10.30742/Melekitjournal.V10i2.358.
[22] H. S. Rifai, S. Febrianti, And I. Santoso, “Analisis Sentimen Tanggapan Masyarakat Terhadap Cyberbullying Di Media Sosial Menggunakan Algoritma Naïve Bayes (Nb),” *J. Ikraith-Informatika*, Vol. 7, No. 2, Pp. 183–196, 2023.
[23] D. Normawati And S. A. Prayogi, “Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter,” *J. Sains Komput. Inform. (J-Sakti*, Vol. 5, No. 2, Pp. 697–711, 2021.
[24] H. Taofiqurrohman, W. Wufron, And F. F. Roji, “Prediksi Harga Saham Telkom Menggunakan Prophet: Analisis Pengaruh Sentimen Publik Terhadap Kehadiran Starlink,” *Malcom Indones. J. Mach. Learn. Comput. Sci.*, Vol. 5, No. 2, Pp. 484–495, Mar. 2025, Doi: 10.57152/Malcom.V5i2.1796.
[25] U. Khairani, V. Mutiawani, And H. Ahmadian, “Pengaruh Tahapan Preprocessing Terhadap Model Indobert Dan Indobertweet Untuk Mendeteksi Emosi Pada Komentar Akun Berita Instagram,” *J. Teknol. Inf. Dan Ilmu Komput.*, Vol. 11, No. 4, Pp. 887–894, Aug. 2024, Doi: 10.25126/Jtiik.1148315.
[26] M. R. Saputra And S. Agustian, “Bulletin Of Computer Science Research Klasifikasi Sentimen Pada Dataset Yang Terbatas Menggunakan Algoritma Convolutional Neural Network,” Vol. 5, No. 4, Pp. 522–531, 2025, Doi: Https://Doi.Org/10.47065/Bulletincsr.V5i4.613.
[27] S. Adi Nugraha, “Penerapan Lexicon Based Untuk Analisis Sentimen Masyarakat Indonesia Terhadap Danantara,” *Jati (Jurnal Mhs. Tek. Inform.*, Vol. 9, No. 3, Pp. 4949–4957, May 2025, Doi: 10.36040/Jati.V9i3.13836.
[28] R. Firdaus, I. Asror, And A. Herdiani, “Lexicon-Based Sentiment Analysis Of Indonesian Language Student Feedback Evaluation,” *Indones. J. Comput.*, Vol. 6, No. 1, Pp. 1–12, 2021, Doi: Https://Doi.Org/10.34818/Indojc.2021.6.1.408.
[29] D. I. Putri, A. N. Alfian, M. Y. Putra, And P. D. Mulyo, “Indobert Model Analysis: Twitter Sentiments On Indonesia’s 2024 Presidential Election,” *J. Appl. Informatics Comput.*, Vol. 8, No. 1, Pp. 7–12, Jul. 2024, Doi: 10.30871/Jaic.V8i1.7440.
[30] Y. Wiciaputra, J. Young, And A. Rusli, “Bilingual Text Classification In English And Indonesian Via Transfer Learning Using Xlm-RoBERTa ,” *Int. J. Adv. Soft Comput. Its Appl.*, Vol. 13, No. 3, Pp. 73–87, Dec. 2021, Doi: 10.15849/Ijasca.211128.06.
[31] E. M. Pusung And I. N. Dewi, “Optimasi RoBERTa Dengan Hyperparameter Tuning Untuk Deteksi Emosi Berbasis Teks,” *J. Nas. Teknol. Dan Sist. Inf.*, Vol. 10, No. 3, Pp. 240–248, Feb. 2025, Doi: 10.25077/Teknosi.V10i3.2024.240-248.
[32] A. H. Wildan And V. R. S. Nastiti, “Perbandingan Kinerja Pre-Trained Indobert-Base Dan Indobert-Lite Pada Klasifikasi Sentimen Ulasan Tiktok Tokopedia Seller Center Dengan Model Indobert,” *Jsii*

(Jurnal Sist. Informasi), Vol. 11, No. 2, Pp. 13–20, Sep. 2024, Doi: 10.30656/Jsii.V11i2.9168.

[33] R. Illahi, S. Agustian, S. K. Riau, S. Baru, And K. Pekanbaru, "Klasifikasi Sentimen Menggunakan Bidirectional Lstm Dan Indobert Dengan Dataset Terbatas," *J. Sist. Inf.*, Vol. 7, No. 1, Pp. 74–84, 2025, Doi: Https://Doi.Org/10.31849/Zn.V7i1.25091.

[34] Andriani Marshanda Putri, Widya Khafa Nofa, And Dewi Anggraini Puspa Hapsari, "Penerapan Metode Bert Untuk Analisis Sentimen Ulasan Pengguna Aplikasi Segari Di Google Play Store," *J. Ilm. Tek.*, Vol. 4, No. 1, Pp. 89–104, Jan. 2025, Doi: 10.56127/Juit.V4i1.1902.

[35] S. S. Sabrina, D. F. Shiddieq, And F. F. Roji, "Comparative Analysis Of Svm And Bert For Sentiment And Sarcasm Detection In The Boycott Of Israeli Products On Platform X," *Sinkron*, Vol. 9, No. 2, Pp. 872–883, May 2025, Doi: 10.33395/Sinkron.V9i2.14723.

[36] B. Couvy-Duchesne *Et Al.*, "Linear Mixed Models Minimise False Positive Rate And Enhance Precision Of Mass Univariate Vertex-Wise Analyses Of Grey-Matter," In *2020 Ieee 17th International Symposium On Biomedical Imaging (Isbi)*, Ieee, Apr. 2020, Pp. 404–407. Doi: 10.1109/Isbi45749.2020.9098719.

[37] R. Bold, H. Al-Khateeb, And N. Ersotelos, "Reducing False Negatives In Ransomware Detection: A Critical Evaluation Of Machine Learning Algorithms," *Appl. Sci.*, Vol. 12, No. 24, P. 12941, Dec. 2022, Doi: 10.3390/App122412941.

[38] S. Riyanto, I. S. Sitanggang, T. Djatna, And T. D. Atikah, "Comparative Analysis Using Various Performance Metrics In Imbalanced Data For Multi-Class Text Classification," *Int. J. Adv. Comput. Sci. Appl.*, Vol. 14, No. 6, Pp. 1082–1090, 2023, Doi: 10.14569/Ijacsa.2023.01406116.

[39] E. Richardson, R. Trevizani, J. A. Greenbaum, H. Carter, M. Nielsen, And B. Peters, "The Receiver Operating Characteristic Curve Accurately Assesses Imbalanced Datasets," *Patterns*, Vol. 5, No. 6, P. 100994, Jun. 2024, Doi: 10.1016/J.Patter.2024.100994.

[40] F. S. Mulyo, "Building A Sentiment Classification Model Using Indobert," Medium.Com. Accessed: Apr. 29, 2025. [Online]. Available: Building A Sentiment Classification Model Using Indobert

[41] M. R. Alfarid, "Sentiment Analysis Of Tapera Policy Using Indobertweet," Medium.Com. Accessed: Apr. 29, 2025. [Online]. Available: Https://Medium.Com/@Ridhoalfarid95/Sentiment-Analysis-Of-Tapera-Policy-Using-Indobertweet-43c332701efe

[42] A. Agustyawan, "Comprehensive Guide To Training Your Own Sentiment Analysis Model With RoBERTa ," Medium.Com. Accessed: Apr. 29, 2025. [Online]. Available: Https://Arifagustyawan.Medium.Com/Comprehensive-Guide-To-Training-Your-Own-Sentiment-Analysis-Model-With-RoBERTa -0eea180d78b2

[43] N. Rana, "Lexicon Vs. Transformer-Based Models For Sentiment Analysis," Medium.Com. Accessed: Jul. 11, 2025. [Online]. Available: Https://Rananeha.Medium.Com/Lexicon-Vs-Transformer-Based-Models-For-Sentiment-Analysis-56a80c8f4b10

[44] A. Kaba, "How To Use Zero-Shot Classification For Sentiment Analysis," Towards Data Science. Accessed: Jul. 11, 2025. [Online]. Available: Https://Towardsdatascience.Com/How-To-Use-Zero-Shot-Classification-For-Sentiment-Analysis-Abf7bd47ad25/