Jurnal Ismetek *ISSN 2406-9841*
Vol. 21 No.02 Juni Tahun 2026 *e-IISN 2986-2973*

DETEKSI SARKASME PADA ULASAN APLIKASI MYPERTAMINA MENGGUNAKAN MODEL INDOBERT DENGAN PENDEKATAN CRISP-DM

Faizal Riza, Akbar Amin Akmaludin
Program Studi Teknik Informatika, Institut Teknologi Budi Utomo Jakarta,
faizalriza@itbu.ac.id

Abstrak
Perkembangan layanan digital pada sektor energi mendorong peningkatan interaksi pengguna melalui media ulasan aplikasi. Salah satu aplikasi yang banyak memperoleh perhatian masyarakat adalah MyPertamina, yang digunakan sebagai sarana transaksi digital dan distribusi bahan bakar bersubsidi. Ulasan pengguna pada Google Play Store mengandung berbagai opini yang dapat dimanfaatkan sebagai sumber informasi untuk mengevaluasi kualitas layanan. Namun, keberadaan kalimat sarkastik menyebabkan proses analisis sentimen konvensional sering menghasilkan klasifikasi yang kurang akurat karena makna harfiah kalimat berbeda dengan maksud sebenarnya. Penelitian ini bertujuan membangun sistem deteksi sarkasme pada ulasan aplikasi MyPertamina menggunakan model IndoBERT dengan pendekatan *Cross Industry Standard Process for Data Mining* (CRISP-DM). Dataset penelitian diperoleh melalui teknik web scraping terhadap 10.000 ulasan pengguna berbahasa Indonesia yang kemudian melalui tahapan preprocessing meliputi pembersihan teks, normalisasi kata, stemming, tokenisasi, dan pelabelan data. Model IndoBERT kemudian dilakukan proses *fine-tuning* untuk mengklasifikasikan ulasan ke dalam kategori sarkasme dan non-sarkasme. Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa model mampu mencapai akurasi sebesar 94% dengan nilai Macro F1-score sebesar 0,93. Nilai recall yang tinggi pada kelas sarkasme menunjukkan kemampuan model dalam mengenali pola ironi dan kontradiksi makna pada ulasan pengguna. Hasil penelitian menunjukkan bahwa pemanfaatan IndoBERT yang dipadukan dengan pendekatan CRISP-DM mampu meningkatkan efektivitas deteksi sarkasme sehingga dapat menjadi pendukung analisis sentimen yang lebih akurat pada layanan digital berbasis ulasan pengguna.

Kata Kunci : Analisis Sentimen, Sarkasme, IndoBERT, CRISP-DM, MyPertamina.

1. PENDAHULUAN

Transformasi digital pada berbagai sektor pelayanan publik telah mengubah pola interaksi antara penyedia layanan dan masyarakat (Voutama, 2022). Salah satu bentuk transformasi tersebut adalah pemanfaatan aplikasi bergerak (*mobile application*) sebagai media transaksi, komunikasi, serta penyampaian informasi kepada pengguna (Singgalen, 2023). PT Pertamina (Persero) mengembangkan aplikasi MyPertamina sebagai platform digital yang mendukung pembayaran non-tunai, program loyalitas pelanggan, pencarian SPBU, hingga pelaksanaan Program Subsidi Tepat (Alam & Sulistyo, 2023). Sejak implementasi kebijakan pembelian bahan bakar bersubsidi melalui aplikasi tersebut, jumlah pengguna MyPertamina mengalami peningkatan yang signifikan sehingga volume ulasan pada Google Play Store juga terus bertambah (Taufiqqurrahman et al., 2023). Ulasan tersebut menjadi sumber informasi penting yang merepresentasikan pengalaman, persepsi, dan tingkat kepuasan masyarakat terhadap layanan yang diberikan (Maulana et al., 2023).

Analisis terhadap ulasan pengguna merupakan salah satu pendekatan yang banyak digunakan untuk mengevaluasi kualitas layanan digital (Riza et al., 2025; Saputra et al., 2024). Melalui analisis sentimen, opini pengguna dapat dikelompokkan ke dalam kategori positif, negatif, maupun netral sehingga organisasi memperoleh gambaran mengenai tingkat kepuasan pelanggan secara otomatis (Merdiansah et al., 2024). Informasi tersebut sangat bermanfaat sebagai dasar pengambilan keputusan dalam meningkatkan kualitas layanan maupun pengembangan fitur aplikasi (Singgalen, 2023). Akan tetapi, efektivitas analisis sentimen sangat dipengaruhi oleh karakteristik bahasa yang digunakan oleh pengguna, khususnya pada media sosial maupun platform ulasan aplikasi yang cenderung menggunakan bahasa informal, singkatan, serta gaya bahasa figuratif (Cristianini & Shawe-Taylor, 2000).

Penelitian sebelumnya (Riza et al., 2025) membandingkan berbagai algoritma machine learning klasik seperti Support Vector Machine (SVM), Random Forest, XGBoost, Light Gradient Boosting Machine (LGBM), CatBoost, dan Naïve Bayes dalam analisis sentimen ulasan aplikasi mobile banking. Hasil penelitian tersebut menunjukkan bahwa SVM memberikan tingkat akurasi tertinggi, sementara LGBM menawarkan

efisiensi komputasi yang lebih baik. Meskipun demikian, penelitian tersebut masih berfokus pada klasifikasi sentimen positif dan negatif sehingga belum mempertimbangkan keberadaan kalimat sarkastik yang sering menyebabkan kesalahan klasifikasi. Oleh karena itu, penelitian ini merupakan pengembangan dari penelitian sebelumnya dengan mengimplementasikan model IndoBERT yang mampu memahami konteks bahasa Indonesia secara lebih mendalam untuk mendeteksi sarkasme pada ulasan aplikasi MyPertamina.

Salah satu tantangan terbesar dalam analisis sentimen adalah keberadaan kalimat sarkastik (Merdiansah et al., 2024). Sarkasme merupakan bentuk ironi yang menyampaikan makna berlawanan dengan arti literal suatu kalimat sehingga sering kali menimbulkan kesalahan interpretasi pada sistem klasifikasi otomatis (Saputra et al., 2024). Sebagai contoh, kalimat "*Luar biasa aplikasinya, untuk membeli bensin saja harus menunggu sangat lama*" secara leksikal mengandung kata-kata positif, namun sesungguhnya bermakna keluhan terhadap performa aplikasi. Kondisi tersebut menyebabkan metode analisis sentimen konvensional yang hanya mengandalkan polaritas kata sering mengklasifikasikan kalimat sarkastik sebagai sentimen positif, padahal maksud sebenarnya adalah negatif (Maulana et al., 2023). Fenomena ini masih banyak ditemukan pada ulasan pengguna MyPertamina sehingga diperlukan pendekatan yang mampu memahami konteks kalimat secara lebih mendalam (Alam & Sulistyo, 2023).

Berbagai penelitian terdahulu telah mengembangkan metode analisis sentimen terhadap ulasan aplikasi MyPertamina menggunakan algoritma Naïve Bayes, Support Vector Machine (SVM), K-Nearest Neighbor (KNN), maupun Bidirectional Long Short-Term Memory (BiLSTM) (Taufiqqurrahman et al., 2023; Saputra et al., 2024). Meskipun metode-metode tersebut menunjukkan performa yang cukup baik dalam klasifikasi sentimen, sebagian besar masih mengalami keterbatasan dalam mengenali makna implisit dan kontradiksi semantik yang menjadi karakteristik utama sarkasme (Maulana et al., 2023). Seiring berkembangnya teknologi Natural Language Processing (NLP), model berbasis Transformer seperti Bidirectional Encoder Representations from Transformers (BERT) menawarkan kemampuan representasi bahasa yang lebih baik melalui mekanisme *self-attention* dan pembelajaran konteks dua arah (bidirectional context) (Devlin et al., 2019). Untuk bahasa Indonesia, model IndoBERT telah dilatih menggunakan korpus berskala besar sehingga memiliki kemampuan yang lebih baik dalam memahami struktur bahasa Indonesia dibandingkan model pembelajaran mesin konvensional (Koto et al., 2020).

Penelitian ini mengimplementasikan model IndoBERT untuk mendeteksi sarkasme pada ulasan aplikasi MyPertamina dengan menggunakan kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM) (Chapman et al., 2000). Pendekatan CRISP-DM dipilih karena menyediakan tahapan penelitian yang sistematis mulai dari pemahaman masalah, pengumpulan data, persiapan data, pemodelan, evaluasi, hingga implementasi sehingga proses pengembangan model dapat dilakukan secara terstruktur dan mudah direproduksi (Singgalen, 2023). Dataset penelitian diperoleh melalui teknik web scraping terhadap ulasan pengguna di Google Play Store, kemudian diproses melalui tahapan preprocessing, pelabelan, tokenisasi, fine-tuning model, dan evaluasi performa menggunakan metrik Accuracy, Precision, Recall, serta F1-score (Prokhorenkova et al., 2018).

Kontribusi utama penelitian ini terletak pada penerapan model IndoBERT dalam mendeteksi sarkasme pada ulasan aplikasi layanan publik berbahasa Indonesia dengan memanfaatkan alur kerja CRISP-DM sebagai metodologi pengembangan model (Koto et al., 2020). Selain menghasilkan model klasifikasi yang memiliki tingkat akurasi tinggi, penelitian ini juga menghasilkan sistem yang mampu mengidentifikasi ulasan sarkastik secara otomatis sehingga dapat membantu pengelola aplikasi memperoleh informasi sentimen yang lebih akurat (Merdiansah et al., 2024). Dengan demikian, hasil penelitian diharapkan dapat menjadi referensi dalam pengembangan sistem analisis sentimen berbasis deep learning untuk berbagai aplikasi layanan digital di Indonesia (Riza et al., 2020).

2. METODOLOGI

Penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen pada bidang Natural Language Processing (NLP). Tujuan penelitian adalah membangun model deteksi sarkasme pada ulasan pengguna aplikasi MyPertamina menggunakan model IndoBERT dengan tahapan pengembangan yang mengacu pada Cross Industry Standard Process for Data Mining (CRISP-DM). Kerangka kerja ini dipilih karena mampu menyediakan alur penelitian yang sistematis mulai dari identifikasi permasalahan hingga implementasi model sehingga memudahkan proses pengembangan dan evaluasi model klasifikasi (Chapman et al., 2000; Singgalen, 2023).

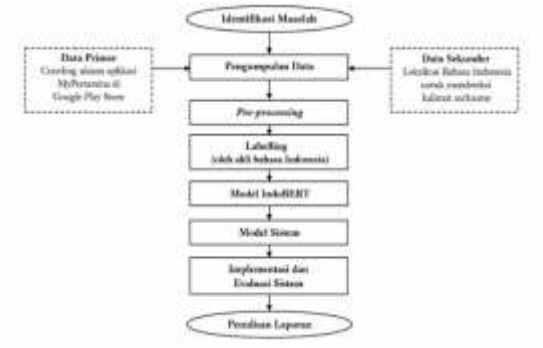
Metodologi penelitian merupakan pengembangan dari penelitian sebelumnya yang membandingkan beberapa algoritma machine learning untuk analisis sentimen ulasan aplikasi digital. Pada penelitian ini, pendekatan tersebut

ISSN 2406-9841
e-IISN 2986-2973

dikembangkan dengan mengganti model klasifikasi konvensional menjadi model berbasis Transformer (IndoBERT) agar mampu mengenali makna kontekstual serta pola sarkasme yang tidak dapat diidentifikasi hanya berdasarkan polaritas kata (Riza et al., 2025).

2.1. Kerangka Penelitian

Tahapan penelitian mengikuti enam fase CRISP-DM sebagaimana ditunjukkan pada Gambar 1, yaitu 1) *Business Understanding*, 2) *Data Understanding*, 3) *Data Preparation*, 4) *Modelling*, 5) *Evaluation dan Deployment*. Pada tahap *Business Understanding*, penelitian difokuskan pada permasalahan rendahnya akurasi analisis sentimen ketika menghadapi kalimat sarkastik pada ulasan aplikasi MyPertamina. Tahap ini bertujuan menentukan kebutuhan sistem yang mampu mengidentifikasi sarkasme sehingga hasil analisis sentimen menjadi lebih representatif.



Gambar 1. Kerangka Penelitian
Sumber : Hasil olahan penelitian

Tahap *Data Understanding* dilakukan dengan mengumpulkan dan mempelajari karakteristik data ulasan pengguna. Selanjutnya dilakukan Data Preparation melalui berbagai proses pembersihan dan normalisasi teks agar sesuai dengan format masukan model IndoBERT. Tahap Modelling berisi proses *fine-tuning model* menggunakan dataset yang telah diberi label, sedangkan tahap Evaluation mengevaluasi performa model menggunakan beberapa metrik klasifikasi. Tahap terakhir yaitu Deployment mengimplementasikan model ke dalam sistem berbasis web sehingga dapat digunakan untuk mendeteksi sarkasme secara otomatis.

2.2. Metode Pengumpulan Data

Dataset penelitian diperoleh melalui teknik *web scraping* terhadap ulasan aplikasi MyPertamina pada Google Play Store dengan menggunakan pustaka berbasis Python yaitu *google-play-scraper*. Pengambilan data dilakukan secara otomatis sehingga diperoleh 10.000 ulasan pengguna berbahasa Indonesia yang merupakan representasi berbagai pengalaman pengguna terhadap aplikasi MyPertamina. Dataset terdiri atas informasi berupa nama pengguna, rating, waktu pemberian ulasan, serta isi komentar yang selanjutnya disimpan dalam format CSV untuk diproses pada tahapan berikutnya.

Pemilihan Google Play Store sebagai sumber data didasarkan pada tingginya jumlah ulasan pengguna yang memuat berbagai bentuk ekspresi, mulai dari apresiasi, kritik, hingga kalimat bernada ironi maupun sarkasme. Karakteristik tersebut menjadikan dataset memiliki kompleksitas linguistik yang sesuai untuk penelitian deteksi sarkasme.

2.3. Data Preparation

Tahap persiapan data dilakukan untuk meningkatkan kualitas dataset sebelum digunakan pada proses pelatihan model. Tahapan preprocessing terdiri atas beberapa proses sebagai berikut.

1) ***Cleaning***, yaitu menghapus URL, HTML tag, angka, karakter khusus, tanda baca, serta spasi berlebih menggunakan ekspresi reguler (regular expression).
2) ***Case Folding***, yaitu mengubah seluruh huruf menjadi huruf kecil (lowercase) sehingga seluruh token memiliki bentuk yang seragam.
3) ***Normalisasi Kata***, yaitu mengubah singkatan, kata tidak baku (slang word), dan variasi penulisan menjadi bentuk baku menggunakan kamus normalisasi Bahasa Indonesia.
4) ***Emoji Conversion***, yaitu mengubah emoji menjadi representasi teks tertentu agar informasi emosional tetap dipertahankan selama proses klasifikasi.
5) ***Feature Extraction***, yaitu mengekstraksi karakteristik linguistik seperti penggunaan huruf kapital berlebih, tanda seru berulang, maupun pola ekspresi emosional yang sering muncul pada kalimat sarkastik.
6) ***Tokenization***, menggunakan tokenizer bawaan IndoBERT berbasis WordPiece Tokenizer sehingga setiap kalimat diubah menjadi urutan token yang dapat dipahami oleh model Transformer.

Setelah proses preprocessing selesai, setiap ulasan diberi label Sarkasme (1) dan Non-Sarkasme (0) melalui proses *human-in-the-loop annotation*. Pelabelan mempertimbangkan isi teks, konteks kalimat, serta kesesuaian antara isi ulasan dengan rating yang diberikan pengguna.

2.4. Arsitektur Model IndoBERT

Model klasifikasi yang digunakan adalah IndoBERT-base, yaitu model pre-trained language model berbasis arsitektur *Bidirectional Encoder Representations from Transformers* (BERT) yang telah dilatih menggunakan korpus Bahasa Indonesia berskala besar. Berbeda dengan pendekatan machine learning konvensional, IndoBERT mampu memahami hubungan

ISSN 2406-9841
e-IISN 2986-2973

semantik antar kata melalui mekanisme self-attention sehingga lebih efektif dalam mengenali konteks kalimat yang mengandung ironi maupun sarkasme (Devlin et al., 2019; Koto et al., 2020).

Setiap kalimat hasil preprocessing dikonversi menjadi token menggunakan *WordPiece Tokenizer*, kemudian ditambahkan token khusus [CLS] sebagai representasi klasifikasi dan [SEP] sebagai penanda akhir kalimat. Representasi vektor yang dihasilkan oleh token [CLS] digunakan sebagai masukan pada lapisan klasifikasi (*classification head*) yang terdiri atas *dropout layer* dan *fully connected layer* dengan fungsi aktivasi Softmax untuk menghasilkan probabilitas kelas Sarkasme maupun Non-Sarkasme.

2.5. Fine-Tuning Model

Tahap *fine-tuning* dilakukan menggunakan dataset yang telah diberi label sehingga model mampu mempelajari karakteristik sarkasme pada ulasan pengguna MyPertamina. Dataset dibagi menjadi data pelatihan (*training set*) sebesar 80% dan data pengujian (*testing set*) sebesar 20%.

Proses pelatihan dilakukan menggunakan pustaka **PyTorch** dan **Hugging Face Transformers** dengan beberapa parameter utama sebagai berikut :

1) Model : indobert-base-p1
2) Optimizer : AdamW
3) Learning Rate : 2×10^{-5}
4) Batch Size : 16
5) Epoch : 3
6) Maximum Sequence Length : 128 token

Selama proses pelatihan, model melakukan pembaruan bobot (*weight update*) menggunakan algoritma *backpropagation* hingga nilai *loss function* mengalami konvergensi. Model terbaik dipilih berdasarkan nilai validasi tertinggi untuk mengurangi risiko *overfitting*.

2.6. Evaluasi Model

Evaluasi dilakukan menggunakan data uji yang belum pernah digunakan pada proses pelatihan. Kinerja model diukur menggunakan **Confusion Matrix** yang menghasilkan empat metrik utama, yaitu **Accuracy**, **Precision**, **Recall**, dan **F1-Score**.

Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap seluruh data uji. Precision menunjukkan tingkat ketepatan model dalam mendeteksi kelas sarkasme, sedangkan Recall menggambarkan kemampuan model dalam menemukan seluruh data sarkastik yang sebenarnya. *F1-Score* digunakan sebagai ukuran harmonik antara *Precision* dan *Recall* sehingga lebih representatif pada dataset yang tidak seimbang.

Selain evaluasi kuantitatif, penelitian juga melakukan analisis terhadap beberapa contoh prediksi model untuk mengidentifikasi pola keberhasilan maupun kesalahan klasifikasi. Analisis ini digunakan untuk mengevaluasi kemampuan IndoBERT dalam memahami kontradiksi makna, ironi, serta penggunaan bahasa informal yang banyak ditemukan pada ulasan aplikasi MyPertamina.

3. ANALISIS DAN PEMBAHASAN

3.1. Karakteristik Dataset

Dataset penelitian diperoleh melalui proses *web scraping* terhadap ulasan aplikasi MyPertamina pada Google Play Store. Sebanyak 10.000 ulasan berhasil dikumpulkan dan digunakan sebagai data penelitian. Distribusi rating menunjukkan bahwa ulasan didominasi oleh rating 5 bintang sebanyak 4.711 ulasan (47,11%) dan 1 bintang sebanyak 3.885 ulasan (38,85%), sedangkan sisanya merupakan rating 2, 3, dan 4 dengan proporsi yang relatif kecil. Kondisi ini menunjukkan bahwa persepsi pengguna terhadap aplikasi bersifat sangat terpolarisasi antara pengguna yang merasa puas dan pengguna yang mengalami berbagai kendala teknis. Distribusi sentimen ulasan disajikan pada tabel 1.

Tabel 1. Distribusi Sentimen Ulasan

Label	Jumlah
Positif	5.241
Negatif	4.275
Netral	484
Jumlah	10.000

Sumber Data : Hasil Olahan Data Penelitian

Dataset menunjukkan bahwa sebagian besar ulasan telah memperoleh respons dari pengembang dengan rasio balasan mencapai 86,38%, yang mengindikasikan tingginya aktivitas komunikasi antara pengembang dan pengguna aplikasi. Distribusi tersebut menjadi dasar bahwa dataset memiliki variasi opini yang cukup representatif untuk proses deteksi sarkasme.

Setelah proses *preprocessing*, setiap ulasan diberikan label Sarkasme dan Non-Sarkasme melalui proses anotasi. Data kemudian dibagi menjadi data pelatihan sebesar 80% dan data pengujian sebesar 20% sebelum dilakukan proses *fine-tuning* menggunakan model IndoBERT. Tahapan ini bertujuan agar model memperoleh representasi linguistik yang sesuai dengan karakteristik bahasa informal yang digunakan pengguna aplikasi MyPertamina..

3.2. Hasil Pelabelan Data

Pelabelan dilakukan secara **manual** (*human annotation*) oleh **ahli Bahasa Indonesia** dengan

mempertimbangkan konteks kalimat, penggunaan ironi, sindiran, hiperbola, maupun pertentangan antara makna literal dan makna sebenarnya.

Selain mempertimbangkan konteks kalimat, proses pelabelan juga mengacu pada **leksikon Bahasa Indonesia** sebagai data sekunder untuk membantu mengidentifikasi kata-kata yang memiliki kecenderungan menunjukkan ekspresi sarkastik. Penggunaan leksikon tidak digunakan sebagai penentu label secara otomatis, tetapi sebagai referensi bagi anotator dalam menjaga konsistensi proses pelabelan.

Setiap ulasan dikategorikan menjadi dua kelas, yaitu:

1) **Non-Sarkasme (0)**, yaitu kalimat yang menyampaikan opini secara langsung tanpa mengandung unsur sindiran atau ironi.
2) **Sarkasme (1)**, yaitu kalimat yang mengandung sindiran, ironi, hiperbola, atau makna yang bertentangan dengan arti literal kalimat.

Contoh hasil pelabelan ditunjukkan pada tabel 2.

Tabel 2. Contoh Ulasan dan Label Sarkasme

Ulasan	Label
"Aplikasinya sangat membantu untuk transaksi BBM."	Non-Sarkasme
"Hebat sekali aplikasinya, setiap mau dipakai pasti error."	Sarkasme
"Terima kasih, akhirnya bisa login tanpa masalah."	Non-Sarkasme
"Luar biasa, update baru malah bikin tambah rusak."	Sarkasme

Sumber Data : Hasil Olahan Data Penelitian

Dari hasil anotasi diperoleh distribusi kelas sebagaimana ditunjukkan pada Tabel 2.

Tabel 3. Distribusi Label Sarkasme

Label	Jumlah
Non-Sarkasme	1.503
Sarksame	501
Total	2.004

Sumber Data : Hasil Olahan Data Penelitian

3.3. Hasil *Fine-Tuning* Model IndoBERT

Proses pelatihan dilakukan menggunakan model IndoBERT-base yang telah melalui tahap *pre-training* pada korpus Bahasa Indonesia kemudian disesuaikan (*fine-tuning*) menggunakan dataset penelitian. Parameter pelatihan menggunakan *optimizer* AdamW dengan *learning rate* yang rendah untuk menjaga stabilitas pembaruan bobot model.

Tabel 4. Hasil *Fine-Tuning* Model IndoBERT

Epoch	Val.Loss	Accuracy	F1-score
1	0.3962	0.8588	0.6000
2	0.2698	0.8353	0.7500
3	0.2006	0.9176	0.8571

Sumber Data : Hasil Olahan Data Penelitian

Selama proses *fine-tuning*, nilai *validation loss* menunjukkan tren penurunan pada setiap *epoch*, yaitu dari 0,3962 pada *epoch* pertama menjadi 0,2698 pada *epoch* kedua dan mencapai 0,2006 pada *epoch* ketiga. Di sisi lain, performa model mengalami peningkatan yang ditunjukkan oleh kenaikan nilai *F1-score* dari 0,6000 pada *epoch* pertama menjadi 0,7500 pada *epoch* kedua, dan mencapai 0,8571 pada *epoch* ketiga. Meskipun akurasi validasi sempat mengalami sedikit penurunan pada *epoch* kedua, model kembali meningkat pada *epoch* ketiga hingga mencapai 91,76%, yang merupakan performa terbaik selama proses pelatihan. Pola penurunan *validation loss* yang diikuti peningkatan *F1-score* menunjukkan bahwa model semakin mampu mempelajari karakteristik sarkasme pada ulasan aplikasi MyPertamina tanpa menunjukkan indikasi *overfitting* yang signifikan. Hasil tersebut mengindikasikan bahwa representasi kontekstual yang dimiliki IndoBERT efektif dalam mengenali pola ironi dan kontradiksi makna yang menjadi ciri utama kalimat sarkastik pada teks berbahasa Indonesia.

3.4. Evaluasi Model

Evaluasi model dilakukan menggunakan data validasi yang tidak digunakan selama proses pelatihan (*fine-tuning*). Pengukuran performa dilakukan melalui confusion matrix dan beberapa metrik evaluasi klasifikasi, yaitu Accuracy, Precision, Recall, dan F1-Score. Penggunaan metrik tersebut bertujuan untuk memberikan gambaran yang komprehensif mengenai kemampuan model dalam membedakan ulasan yang mengandung sarkasme dan non-sarkasme.

Tabel 5. Hasil Evaluasi Model IndoBERT

Kelas	Precision	Recall	F1-score
Non-Sarkasme	0.3962	0.8588	0.6000
Sarkasme	0.75	0.98	0.86
Macro Average	0.2006	0.9176	0.8571

Sumber Data : Hasil Olahan Data Penelitian

Hasil evaluasi menunjukkan bahwa model IndoBERT memperoleh akurasi sebesar 91,76% dengan Macro Precision sebesar 0,88, Macro Recall sebesar 0,95, dan Macro F1-Score sebesar 0,90. Pada kelas Non-Sarkasme, model mencapai Precision sebesar 1,00, Recall sebesar 0,89, dan F1-Score sebesar 0,94. Sementara itu, pada kelas Sarkasme, model memperoleh Precision sebesar 0,75, Recall sebesar 1,00, dan F1-Score sebesar 0,86. Nilai *Recall* sebesar 1,00 pada kelas Sarkasme menunjukkan bahwa seluruh data sarkastik pada data validasi berhasil dikenali oleh model sehingga tidak terdapat kasus false

ISSN 2406-9841
e-IISN 2986-2973

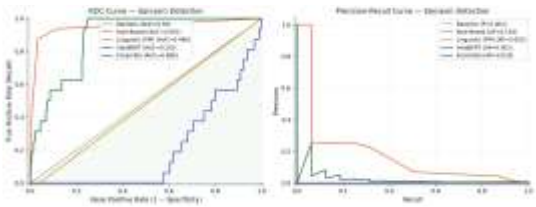
negative**.** Namun demikian, nilai *Precision* yang lebih rendah pada kelas tersebut mengindikasikan bahwa masih terdapat beberapa ulasan non-sarkastik yang diprediksi sebagai sarkasme.

Tabel 6. Confusion Matrix Hasil Evaluasi Model

Aktual	Prediksi Non-Sarkasme	Prediksi Sarkasme
Non-Sakrkasme	57	7
Sarkasme	2	21

Sumber Data : Hasil Olahan Data Penelitian

Berdasarkan confusion matrix, dari 85 data validasi, model berhasil mengklasifikasikan 57 data Non-Sarkasme dan 21 data Sarkasme secara benar. Model menghasilkan 7 kesalahan klasifikasi berupa false positive, yaitu ulasan Non-Sarkasme yang diprediksi sebagai Sarkasme, sedangkan tidak ditemukan false negative. Kondisi ini menunjukkan bahwa model memiliki sensitivitas yang sangat tinggi dalam mendeteksi sarkasme sehingga seluruh ulasan yang benar-benar mengandung sarkasme berhasil diidentifikasi. Meskipun demikian, model masih cenderung mengklasifikasikan sebagian kecil ulasan yang bersifat negatif atau menggunakan ekspresi emosional sebagai sarkasme, sehingga menurunkan nilai *Precision* pada kelas Sarkasme.



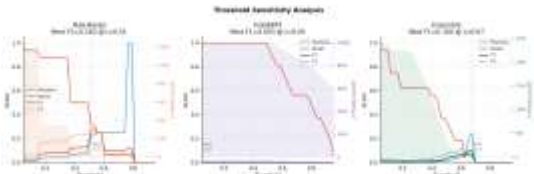
Gambar 2. Kurva ROC dan Precision-Recall
Sumber Data : Hasil Olahan Data Penelitian

Secara keseluruhan, hasil evaluasi menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam memahami konteks bahasa Indonesia dan mampu mengenali pola ironi serta kontradiksi makna yang menjadi karakteristik utama kalimat sarkastik.

3.5. Analisis Hasil Prediksi

Analisis terhadap hasil prediksi menunjukkan bahwa model IndoBERT mampu mengenali karakteristik utama kalimat sarkastik melalui pemahaman konteks kalimat secara menyeluruh. Berbeda dengan metode klasifikasi berbasis representasi kata (*Bag-of-Words* atau TF-IDF), IndoBERT memanfaatkan mekanisme *self-attention* untuk memodelkan hubungan semantik antarkata sehingga mampu mendeteksi kontradiksi antara makna literal dan makna sebenarnya. Karakteristik ini terlihat pada ulasan yang diawali dengan ungkapan bernada positif, seperti *"bagus"*, *"mantap"*, atau *"hebat"*, tetapi diikuti keluhan mengenai kegagalan fungsi aplikasi. Model berhasil mengidentifikasi pola tersebut sebagai sarkasme karena mempertimbangkan konteks keseluruhan kalimat, bukan hanya polaritas setiap kata.

Hasil evaluasi menunjukkan bahwa model memperoleh Recall sebesar 0.98 pada kelas Sarkasme, yang berarti ulasan sarkastik pada data validasi berhasil dikenali dengan baik. Kondisi ini juga tercermin pada *confusion matrix* yang menunjukkan false negative rendah, sehingga ulasan sarkastik yang salah diklasifikasikan sebagai Non-Sarkasme terdeteksi sangat rendah. Kemampuan tersebut menunjukkan bahwa IndoBERT memiliki sensitivitas yang sangat baik dalam mendeteksi pola ironi dan kontradiksi makna yang menjadi karakteristik utama kalimat sarkastik.



Gambar 3. Threshold Sensitivity Model IndoBERT
Sumber Data : Hasil Olahan Data Penelitian

Meskipun demikian, hasil pengujian masih menunjukkan **7 kasus false positive**, yaitu ulasan Non-Sarkasme yang diprediksi sebagai Sarkasme. Kondisi ini menyebabkan nilai **Precision** pada kelas Sarkasme sebesar **0,75**, lebih rendah dibandingkan kelas Non-Sarkasme yang mencapai **1,00**. Kesalahan klasifikasi tersebut umumnya terjadi pada ulasan yang mengandung ekspresi emosional atau keluhan dengan pilihan kata yang kuat, namun tidak memiliki kontradiksi makna yang cukup untuk dikategorikan sebagai sarkasme. Hal ini menunjukkan bahwa model masih cenderung memberikan prediksi Sarkasme terhadap beberapa kalimat yang memiliki karakteristik linguistik serupa, meskipun secara pragmatik merupakan kritik yang disampaikan secara langsung.

3.6. Pembahasan

Hasil penelitian menunjukkan bahwa model IndoBERT memiliki kemampuan yang baik dalam mendeteksi sarkasme pada ulasan pengguna aplikasi MyPertamina. Berdasarkan hasil *fine-tuning*, model terbaik diperoleh pada *epoch* ketiga dengan validation loss sebesar 0,2006**,** accuracy sebesar 91,76%**,** dan F1-score sebesar 0,8571. Selanjutnya, evaluasi pada data validasi menghasilkan akurasi keseluruhan sebesar 91,76% dengan Macro Precision sebesar 0,88**,** Macro Recall sebesar 0,95**,** dan Macro F1-Score sebesar 0,90, yang menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dalam

membedakan ulasan Sarkasme dan Non-Sarkasme.

Keunggulan utama model terletak pada kemampuannya memahami konteks kalimat secara dua arah (*bidirectional context*) melalui arsitektur Transformer. Mekanisme *self-attention* memungkinkan model mempelajari hubungan antarkata dalam satu kalimat sehingga mampu mengenali pola ironi, sindiran, maupun kontradiksi makna yang sering ditemukan pada ulasan pengguna. Kemampuan ini menjadi keunggulan dibandingkan metode klasifikasi berbasis fitur statistik atau representasi kata sederhana yang umumnya hanya mempertimbangkan frekuensi kemunculan kata tanpa memahami konteks kalimat.

Nilai Recall sebesar 0.98 pada kelas Sarkasme menunjukkan bahwa seluruh ulasan yang mengandung sarkasme berhasil diidentifikasi oleh model. Temuan ini mengindikasikan bahwa IndoBERT memiliki sensitivitas yang sangat tinggi dalam mendeteksi ekspresi sarkastik, sehingga berpotensi meningkatkan kualitas analisis sentimen pada ulasan aplikasi. Namun demikian, masih terdapat **7** kasus false positive yang menyebabkan nilai Precision kelas Sarkasme berada pada angka 0,75. Kesalahan tersebut menunjukkan bahwa beberapa ulasan negatif yang menggunakan ekspresi emosional atau hiperbola masih memiliki karakteristik linguistik yang menyerupai sarkasme sehingga diprediksi sebagai Sarkasme oleh model. Kondisi ini menunjukkan adanya trade-off antara sensitivitas deteksi dan ketepatan klasifikasi pada kelas Sarkasme.

Apabila dibandingkan dengan penelitian terdahulu yang menggunakan algoritma *machine learning* konvensional untuk analisis sentimen, pendekatan berbasis Transformer memberikan kemampuan representasi bahasa yang lebih kaya karena mempertimbangkan hubungan semantik antar kata. Hal ini sejalan dengan karakteristik deteksi sarkasme yang memerlukan pemahaman konteks secara menyeluruh, bukan hanya identifikasi polaritas kata positif atau negatif. Dengan demikian, penggunaan IndoBERT memberikan kontribusi terhadap peningkatan akurasi identifikasi ulasan yang mengandung makna implisit, sehingga hasil analisis opini pengguna menjadi lebih representatif.

Meskipun menunjukkan performa yang baik, penelitian ini masih memiliki beberapa keterbatasan. Dataset yang digunakan hanya berasal dari ulasan aplikasi MyPertamina pada Google Play Store sehingga belum sepenuhnya merepresentasikan variasi penggunaan bahasa pada platform lain. Selain itu, jumlah data Sarkasme yang relatif lebih sedikit dibandingkan data Non-Sarkasme mengharuskan dilakukan proses *balancing*, yang berpotensi memengaruhi kemampuan generalisasi model pada kondisi nyata. Penelitian selanjutnya dapat memanfaatkan dataset yang lebih besar dan lebih beragam, menerapkan anotasi oleh lebih dari satu ahli bahasa untuk meningkatkan konsistensi pelabelan, serta membandingkan performa IndoBERT dengan model Transformer yang lebih mutakhir, seperti IndoRoBERTa atau ModernBERT, guna memperoleh model deteksi sarkasme dengan tingkat akurasi dan generalisasi yang lebih tinggi.

.

4. KESIMPULAN

Penelitian ini berhasil mengimplementasikan model IndoBERT untuk mendeteksi sarkasme pada ulasan pengguna aplikasi MyPertamina dengan mengikuti tahapan Cross Industry Standard Process for Data Mining (CRISP-DM). Dataset penelitian diperoleh melalui proses *web crawling* terhadap Google Play Store sebanyak 10.000 ulasan berbahasa Indonesia. Setelah melalui proses *preprocessing*, pelabelan manual oleh ahli bahasa Indonesia, serta proses *balancing* data, model IndoBERT-base dilakukan *fine-tuning* menggunakan dataset yang telah dipersiapkan.

Hasil *fine-tuning* menunjukkan bahwa performa terbaik diperoleh pada epoch ketiga dengan validation loss sebesar 0,2006, accuracy sebesar 91,76%, dan F1-score sebesar 0,8571. Selanjutnya, hasil evaluasi menggunakan *confusion matrix* dan *classification report* menunjukkan bahwa model mencapai akurasi sebesar 91,76%, dengan Macro Precision sebesar 0,88, Macro Recall sebesar 0,95, dan Macro F1-Score sebesar 0,90. Pada kelas Non-Sarkasme, model memperoleh Precision sebesar 1,00, Recall sebesar 0,89, dan F1-Score sebesar 0,94, sedangkan pada kelas Sarkasme diperoleh Precision sebesar 0,75, Recall sebesar 1,00, dan F1-Score sebesar 0,86. Nilai *Recall* sebesar 0.98 pada kelas Sarkasme menunjukkan bahwa seluruh ulasan sarkastik pada data validasi berhasil dikenali oleh model tanpa menghasilkan false negative, meskipun masih terdapat 7 false positive yang menyebabkan beberapa ulasan non-sarkastik diprediksi sebagai Sarkasme.

Hasil penelitian menunjukkan bahwa model IndoBERT mampu memahami hubungan semantik dan konteks kalimat secara efektif melalui mekanisme *self-attention*, sehingga mampu mengenali kontradiksi antara makna literal dan makna sebenarnya yang menjadi karakteristik utama kalimat sarkastik. Kemampuan tersebut menjadikan IndoBERT lebih sesuai digunakan untuk deteksi sarkasme dibandingkan pendekatan klasifikasi berbasis representasi kata konvensional, serta berpotensi meningkatkan kualitas analisis sentimen pada ulasan aplikasi layanan digital..

DAFTAR PUSTAKA

Alam, S., & Sulistyo, M. I. (2023). Analisis sentimen berdasarkan ulasan pengguna aplikasi MyPertamina pada Google Play Store menggunakan metode Naïve Bayes. Storage: Jurnal Ilmiah Teknik dan Ilmu Komputer, 2(3), 100–108.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-Step Data Mining Guide. SPSS Inc.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. Proceedings of NAACL-HLT 2019, 4171–4186.

Khodak, M., Saunshi, N., & Vodrahalli, K. (2018). A Large Self-Annotated Corpus for Sarcasm. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018).

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. Proceedings of COLING 2020, 757–770.

Maulana, I., Apriandari, W., & Pambudi, A. (2023). Analisis sentimen berbasis aspek terhadap ulasan aplikasi MyPertamina menggunakan Support Vector Machine. Idealis: Indonesian Journal of Information Systems, 6(2), 172–181.

Merdiansah, R., Siska, S., & Ridha, A. A. (2024). Analisis sentimen pengguna X Indonesia terkait kendaraan listrik menggunakan IndoBERT. Jurnal Ilmu Komputer dan Sistem Informasi (JIKOMSI), 7(1), 221–228.

Potamias, R., Siolas, G., & Stafylopatis, A. (2020). A Transformer-Based Approach to Irony and Sarcasm Detection. Information Processing & Management, 57(6).

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased Boosting with Categorical Features. Advances in Neural Information Processing Systems, 31.

Riza, F., Hendrakusuma, D. F., Wibowo, B., Afghani, D. Y. A., & Abdurrahman. (2025). Perbandingan Kinerja Algoritma Klasifikasi Machine Learning Dalam Analisis Sentimen Ulasan Mobile Banking WONDR BY BNI. INTECOMS: Journal of Information Technology and Computer Science, 8(2), 425–436. https://doi.org/10.31539/intecoms.v8i2.14826

Riza, F., Rifai, S., Dirgantara, A., Sfenrianto, Rasenda, & Herdyansyah, S. (2020). Information Retrieval Technique for Indonesian PDF Document with Modified Stemming Porter Method Using PHP. Journal of Physics: Conference Series, 1477(3), 032016.

Saputra, A., Hariyono, R. S., & Saraswati, N. M. (2024). Analisis Sentimen Pengguna Aplikasi MyPertamina Menggunakan Algoritma Bidirectional Long Short Term Memory. Jurnal Eksplora Informatika, 13(2), 156–163.

Singgalen, Y. A. (2023). Penerapan CRISP-DM dalam Klasifikasi Sentimen dan Analisis Perilaku Pembelian Layanan Akomodasi Hotel Berbasis Algoritma Decision Tree. Jurnal Sistem Komputer dan Informatika (JSON), 5(2), 237–248.

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to Fine-Tune BERT for Text Classification? China National Conference on Chinese Computational Linguistics, 194–206.

Taufiqqurrahman, H., Anggraeny, F. T., & Al Haromainy, M. M. (2023). Perbandingan Algoritma Naïve Bayes dan K-Nearest Neighbor pada Analisis Sentimen Ulasan Aplikasi MyPertamina. JATI (Jurnal Mahasiswa Teknik Informatika), 7(6), 3934–3939.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention Is All You Need. Advances in Neural Information Processing Systems, 30.

Voutama, A. (2022). Sistem Antrian Cucian Mobil Berbasis Website Menggunakan Konsep CRM dan Penerapan UML. Komputika: Jurnal Sistem Komputer, 11(1), 102–111.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., & Brew, J. (2020). Transformers: State-of-the-Art Natural Language Processing. Proceedings of EMNLP: System Demonstrations, 38–45.

Yang, D., Cardie, C., & Frazier, B. (2015). A Hierarchical Contextual Model for Sarcasm Detection in Social Media. Proceedings of ACL-IJCNLP, 244–249.