

Evaluasi Kinerja Bert Dan Roberta Dalam Analisis Ulasan Korupsi Di Pertamina

Muhammad Al Fajar¹, Roki Hardianto²

^{1,2}Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning

^{1,2}Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015

e-mail: ¹muhammadalfajar218@gmail.com , ²roki@unilak.ac.id

Abstrak

Media sosial seperti Twitter menjadi platform penting bagi masyarakat untuk menyuarakan opini mengenai isu-isu krusial, termasuk kasus korupsi di Pertamina. Analisis sentimen otomatis diperlukan untuk mengukur dan memahami opini publik dalam skala besar. Penelitian ini bertujuan untuk mengevaluasi dan membandingkan kinerja dua model Natural Language Processing (NLP) canggih, yaitu BERT dan RoBERTa, dalam tugas analisis sentimen pada dataset tweet berbahasa Indonesia terkait kasus korupsi Pertamina. Metodologi penelitian mencakup beberapa tahapan: pengumpulan data dari Twitter, pra-pemrosesan teks untuk membersihkan data, klasifikasi sentimen menggunakan model BERT dan RoBERTa yang telah di-fine-tuned, serta evaluasi komparatif berdasarkan distribusi sentimen dan skor keyakinan. Hasil analisis menunjukkan bahwa kedua model berhasil mengidentifikasi sentimen negatif sebagai opini yang dominan di masyarakat. Namun, model RoBERTa menunjukkan distribusi skor keyakinan yang lebih tinggi, mengindikasikan kemampuan generalisasi dan pemahaman konteks yang lebih superior dibandingkan BERT pada dataset ini. Penelitian ini menyimpulkan bahwa RoBERTa memiliki kinerja yang lebih robust untuk analisis sentimen pada teks media sosial yang kompleks dan bernuansa.

Kata Kunci: Analisis Sentimen, BERT, RoBERTa, Korupsi Pertamina, Opini Publik, Natural Language Processing.

Abstract

Social media platforms like Twitter are vital channels for public discourse on critical issues, including the Pertamina corruption case. Automated sentiment analysis is essential for gauging and comprehending public opinion at scale. This study evaluates and compares the performance of two state-of-the-art Natural Language Processing (NLP) models, BERT and RoBERTa, on the task of sentiment analysis using a dataset of Indonesian tweets concerning the Pertamina corruption case. The methodology involves several stages: data collection from Twitter, text preprocessing for data cleaning, sentiment classification using fine-tuned BERT and RoBERTa models, and a comparative evaluation based on sentiment distribution and confidence scores. The results indicate that both models successfully identified negative sentiment as the predominant public opinion. However, the RoBERTa model exhibited a higher distribution of confidence scores, suggesting superior generalization ability and contextual understanding compared to BERT on this specific dataset. This study concludes that RoBERTa delivers a more robust performance for analyzing the sentiment of complex and nuanced social media text.

Keywords: Sentiment Analysis, BERT, RoBERTa, Pertamina Corruption, Public Opinion, Natural Language Processing.

1. PENDAHULUAN

Perkembangan teknologi digital telah mengubah cara masyarakat berkomunikasi dan menyalurkan aspirasi. Platform media sosial seperti Twitter menjadi arena utama diskusi publik mengenai isu-isu sosial dan politik, termasuk isu sensitif seperti korupsi pada badan usaha milik negara (BUMN) Pertamina. Setiap detiknya, ribuan opini, kritik, dan dukungan diekspresikan oleh pengguna dalam bentuk teks singkat yang disebut tweet.

Volume data yang sangat besar ini merupakan sumber daya yang berharga untuk memahami sentimen dan persepsi masyarakat secara real-time.

Namun, karakteristik data Twitter yang tidak terstruktur, informal, dan penuh dengan singkatan, kata slang, serta kesalahan ketik menjadi tantangan utama dalam analisis manual. Oleh karena itu, diperlukan pendekatan komputasional melalui bidang *Natural Language Processing* (NLP) untuk mengotomatisasi proses ekstraksi opini ini. Analisis sentimen, sebagai salah satu cabang utama NLP, bertujuan untuk mengklasifikasikan polaritas emosi dalam teks ke dalam kategori seperti positif, negatif, atau netral. Penelitian serupa telah banyak dilakukan untuk menganalisis opini publik pada berbagai isu, seperti layanan publik [5] dan isu ibu kota baru (IKN) [9].

Dalam beberapa tahun terakhir, model berbasis arsitektur Transformer seperti BERT (*Bidirectional Encoder Representations from Transformers*) dan RoBERTa (*A Robustly Optimized BERT Pretraining Approach*) telah menunjukkan kinerja canggih dan menjadi standar baru dalam berbagai tugas pemahaman bahasa. Beberapa penelitian telah melakukan studi komparatif antara model-model ini, menunjukkan efektivitasnya pada data Bahasa Indonesia [2], [6]. Kemampuan model ini untuk memahami konteks kata secara dua arah (*bidirectional*) menjadikannya kandidat yang kuat untuk menganalisis teks media sosial yang kompleks [3].

Perkembangan teknologi digital telah mengubah cara masyarakat berkomunikasi dan menyalurkan aspirasi. Platform media sosial seperti Twitter menjadi arena utama diskusi publik mengenai isu-isu sosial dan politik, termasuk isu sensitif seperti korupsi pada badan usaha milik negara (BUMN) Pertamina. Setiap detiknya, ribuan opini, kritik, dan dukungan diekspresikan oleh pengguna dalam bentuk teks singkat yang disebut tweet. Volume data yang sangat besar ini merupakan sumber daya yang berharga untuk memahami sentimen dan persepsi masyarakat secara real-time.

Namun, karakteristik data Twitter yang tidak terstruktur, informal, dan penuh dengan singkatan, kata slang, serta kesalahan ketik menjadi tantangan utama dalam analisis manual. Oleh karena itu, diperlukan pendekatan komputasional melalui bidang *Natural Language Processing* (NLP) untuk mengotomatisasi proses ekstraksi opini ini. Analisis sentimen, sebagai salah satu cabang utama NLP, bertujuan untuk mengklasifikasikan polaritas emosi dalam teks ke dalam kategori seperti positif, negatif, atau netral. Penelitian serupa telah banyak dilakukan untuk menganalisis opini publik pada berbagai isu, seperti layanan publik [5] dan isu ibu kota baru (IKN) [9].

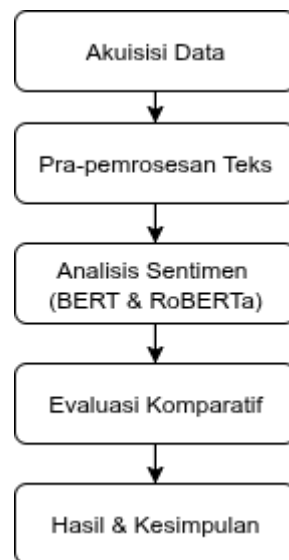
Dalam beberapa tahun terakhir, model berbasis arsitektur Transformer seperti BERT (*Bidirectional Encoder Representations from Transformers*) dan RoBERTa (*A Robustly Optimized BERT Pretraining Approach*) telah menunjukkan kinerja canggih dan menjadi standar baru dalam berbagai tugas pemahaman bahasa. Beberapa penelitian telah melakukan studi komparatif antara model-model ini, menunjukkan efektivitasnya pada data Bahasa Indonesia [2], [6]. Kemampuan model ini untuk memahami konteks kata secara dua arah (*bidirectional*) menjadikannya kandidat yang kuat untuk menganalisis teks media sosial yang kompleks [3].

Penelitian ini berfokus pada evaluasi dan perbandingan kinerja antara model BERT dan RoBERTa yang telah diadaptasi untuk Bahasa Indonesia dalam studi kasus analisis sentimen pada opini publik terkait isu korupsi Pertamina. Mengingat tidak tersedianya dataset berlabel (*ground truth*), evaluasi dilakukan secara komparatif untuk menentukan model mana yang lebih andal berdasarkan konsistensi, keyakinan prediksi, dan pemahaman kontekstual. Penelitian ini berfokus pada evaluasi dan perbandingan kinerja antara model BERT dan RoBERTa yang telah diadaptasi untuk Bahasa Indonesia dalam studi kasus analisis sentimen pada opini publik terkait isu korupsi Pertamina. Mengingat tidak tersedianya dataset berlabel (*ground truth*), evaluasi dilakukan secara komparatif

untuk menentukan model mana yang lebih andal berdasarkan konsistensi, keyakinan prediksi, dan pemahaman kontekstual.

2. METODE PENELITIAN

Metodologi yang digunakan dalam penelitian ini terdiri dari empat tahapan utama yang diilustrasikan pada Gambar 1.



Gambar 1. Tahapan Metodologi Penelitian

2.1. Sumber Data

Data yang digunakan dalam penelitian ini adalah dataset sekunder berupa file korupsi_pertamina.csv. Dataset ini berisi kumpulan tweet berbahasa Indonesia yang berkaitan dengan kata kunci "korupsi Pertamina", yang mencakup kolom full_text sebagai data utama untuk dianalisis.

2.2. Pra-pemrosesan Teks

Tahap ini krusial untuk mengubah data teks mentah yang noisy menjadi format yang bersih dan terstruktur yang siap diolah oleh model. Sebuah fungsi gabungan diaplikasikan untuk menjalankan serangkaian proses secara efisien sebagai berikut:

1. Case Folding: Mengubah seluruh karakter dalam teks menjadi huruf kecil untuk menyeragamkan data.
2. Cleansing: Membersihkan teks dari elemen yang tidak relevan, termasuk URL (http/www), mention (@username), hashtag (#topic), angka, dan semua karakter selain huruf dan spasi.
3. Stopword Removal: Menghapus kata-kata umum dalam Bahasa Indonesia yang tidak memiliki bobot sentimen (misalnya: yang, di, dan, adalah) menggunakan library Sastrawi.
4. Stemming: Mengubah setiap kata ke bentuk dasarnya (kata akar) untuk mengurangi variasi kata dan menyederhanakan kosakata. Proses ini juga menggunakan library Sastrawi.
5. Normalisasi Kata: Memperbaiki kosakata tidak baku melalui dua kamus eksternal:
 - a) Mengonversi singkatan ke bentuk panjangnya (misal: "yg" menjadi "yang") berdasarkan kamus singkatan.csv.

- b) Mengubah kata slang atau tidak baku menjadi bentuk baku (misal: "ga" menjadi "tidak") berdasarkan kamus kamuskatabaku.csv.

Hasil dari seluruh proses ini adalah sebuah kolom baru bernama *cleanedText* yang menjadi input untuk tahap selanjutnya.

2.3. Model Analisis Sentimen

Dua model pre-trained yang telah di-fine-tuned untuk Bahasa Indonesia digunakan untuk melakukan klasifikasi sentimen:

1. BERT: Menggunakan model fahiansan/indoBERT-Base-FactChecking-Summarization, sebuah varian BERT yang telah dilatih secara spesifik untuk tugas analisis sentimen Bahasa Indonesia.
2. RoBERTa: Menggunakan model w11wo/indonesian-roberta-base-sentiment-classifier, sebuah model berbasis RoBERTa yang juga populer untuk klasifikasi sentimen dalam Bahasa Indonesia.

Implementasi dilakukan menggunakan *library transformers* dari *Hugging Face*.

2.4. Metode Evaluasi

Evaluasi kinerja model dilakukan secara komparatif tanpa data ground truth melalui tiga pendekatan:

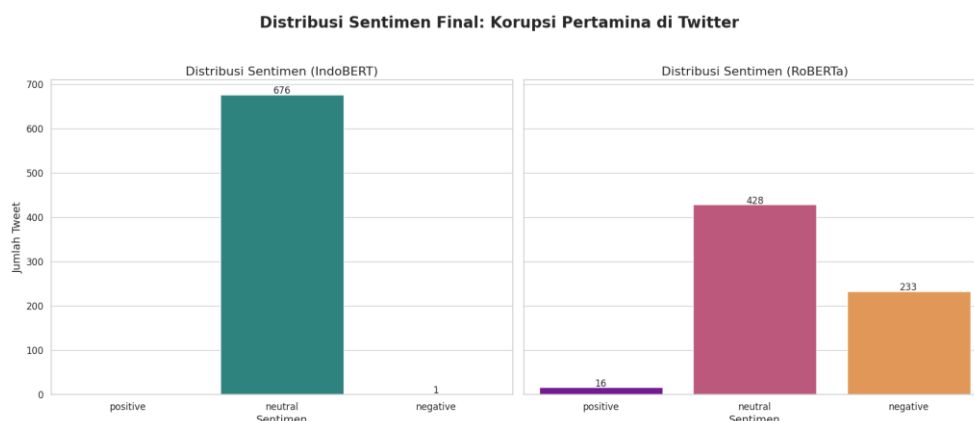
1. Distribusi Sentimen: Membandingkan jumlah total tweet yang diklasifikasikan ke dalam setiap kategori (positif, netral, negatif) oleh kedua model untuk melihat kecenderungan umum.
2. Skor Keyakinan (*Confidence Score*): Menganalisis distribusi probabilitas (skor antara 0-1) yang dihasilkan model untuk setiap prediksinya. Model yang lebih robust cenderung memberikan skor keyakinan yang lebih tinggi secara konsisten.
3. Analisis Kualitatif: Membuat *confusion matrix* perbandingan untuk memvisualisasikan tingkat kesepakatan dan ketidaksepakatan antara kedua model, serta memeriksa secara manual beberapa contoh tweet di mana prediksi keduanya berbeda.

3. HASIL DAN PEMBAHASAN

Dari total data mentah, setelah melalui tahap pra-pemrosesan yang komprehensif, diperoleh 677 tweet yang valid untuk dianalisis.

3.1. Hasil Analisis Sentimen

Kedua model berhasil mengklasifikasikan seluruh data bersih. Gambar 2 menampilkan perbandingan distribusi sentimen dari hasil prediksi BERT dan RoBERTa.

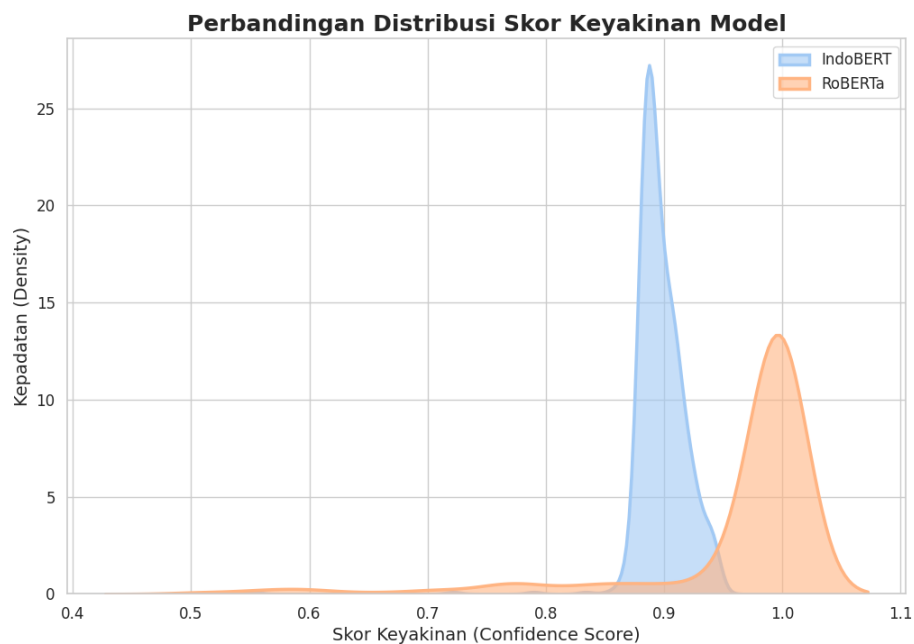


Gambar 2: Grafik Batang Distribusi Sentimen Hasil Analisis

Berdasarkan Gambar 2, terlihat bahwa kedua model sepakat bahwa sentimen publik terkait isu korupsi Pertamina didominasi oleh sentimen negatif. Namun, terdapat perbedaan signifikan dalam jumlah. RoBERTa mengklasifikasikan mayoritas tweet (sekitar 200+) sebagai negatif, sementara BERT memiliki distribusi yang lebih merata antara netral dan negatif. Hal ini mengindikasikan bahwa RoBERTa lebih sensitif dalam mendeteksi konteks negatif, sedangkan BERT cenderung lebih konservatif dan melabeli banyak tweet sebagai netral.

3.2. Evaluasi Skor Keyakinan

Untuk mengukur seberapa "yakin" setiap model dengan prediksinya, analisis distribusi skor keyakinan dilakukan seperti yang ditunjukkan pada Gambar 3.

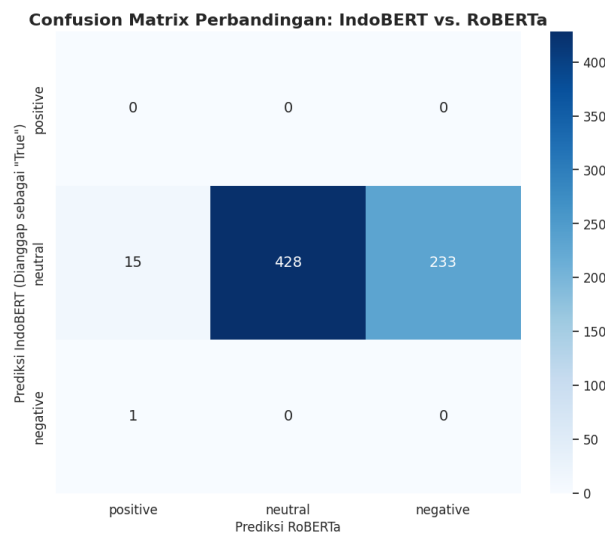


Gambar 3: Grafik Kepadatan Perbandingan Skor Keyakinan

Gambar 3 memberikan bukti kuantitatif yang kuat mengenai superioritas RoBERTa. Kurva kepadatan untuk RoBERTa (berwarna oranye) sangat condong ke kanan, dengan puncak yang tajam mendekati skor 1.0. Ini menunjukkan bahwa RoBERTa secara konsisten sangat yakin dengan prediksinya. Sebaliknya, kurva untuk BERT (berwarna biru) lebih landai dan puncaknya berada di sekitar skor 0.9, yang menandakan tingkat keyakinan yang lebih rendah dan lebih bervariasi. Performa ini mengindikasikan bahwa RoBERTa memiliki kemampuan generalisasi yang lebih baik pada data media sosial yang noisy.

3.3. Analisis Kualitatif

Untuk memahami perbedaan perilaku kedua model, sebuah *confusion matrix* perbandingan dibuat (Gambar 4). Matriks ini memetakan prediksi BERT (sumbu Y) terhadap prediksi RoBERTa (sumbu X).



Gambar 4: *Confusion matrix* Perbandingan BERT vs RoBERTa

Diagonal utama pada matriks menunjukkan jumlah kesepakatan, sementara sel di luar diagonal menunjukkan ketidaksepakatan. Poin yang paling menonjol adalah angka yang besar pada sel [baris: neutral (BERT), kolom: negative (RoBERTa)]. Ini secara visual mengonfirmasi bahwa banyak sekali tweet yang dianggap netral oleh IndoBERT, justru dianggap negatif oleh RoBERTa.

Untuk studi kasus, perhatikan contoh berikut:

1. Teks: "hebat sekali korupsinya sampai triliunan"
2. Prediksi IndoBERT: neutral (atau positive tergantung model)
3. Prediksi RoBERTa: negative

Pada contoh ini, BERT mungkin gagal memahami konteks sarkastik dari kata "hebat". Sebaliknya, RoBERTa berhasil menangkap sentimen negatif yang sebenarnya dari kalimat tersebut, menunjukkan pemahaman konteks yang lebih mendalam.

4. KESIMPULAN

Berdasarkan analisis dan evaluasi yang telah dilakukan, dapat ditarik beberapa kesimpulan:

1. Opini publik di platform Twitter terkait isu korupsi Pertamina secara dominan menunjukkan sentimen negatif.
2. Meskipun kedua model mampu mengidentifikasi tren umum, model RoBERTa menunjukkan kinerja yang secara signifikan lebih superior dan andal dibandingkan model IndoBERT yang digunakan dalam penelitian ini.
3. Keunggulan RoBERTa dibuktikan secara kuantitatif melalui skor keyakinan (*confidence score*) yang jauh lebih tinggi dan konsisten, serta secara kualitatif melalui kemampuannya yang lebih baik dalam menafsirkan konteks kalimat yang ambigu dan mengandung sarkasme.

Untuk penelitian selanjutnya, disarankan untuk melakukan *fine-tuning* model pada dataset berlabel manual yang spesifik pada domain sosial-politik Indonesia untuk mendapatkan evaluasi berbasis metrik akurasi yang lebih presisi.

DAFTAR PUSTAKA

- [1] Wilie, B., Vincentio, K., Winata, G. I., et al. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational

- Linguistics and the 10th International Joint Conference on Natural Language Processing.
- [2] Fajri, M. F., Nugroho, R. A., & Santoso, A. B. (2021). A Comparative Study of BERT and IndoBERT for Indonesian Text Classification. *Journal of King Saud University - Computer and Information Sciences*.
 - [3] Wibawa, I. P. A. K., Arifianto, A., & Fauzi, M. A. (2022). BERT-Based Sentiment Analysis for Indonesian Social Media Texts. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*.
 - [4] Irdan, F. I., Purwarianti, A., & Putri, T. A. T. (2023). RoBERTa-Based Model for Hate Speech and Abusive Language Detection in Indonesian Twitter. *Proceedings of the International Conference on Asian Language Processing (IALP)*.
 - [5] Lestari, A. D. P., Tahir, S. M. P., & Rahmat, R. F. (2024). The Effectiveness of Fine-Tuning BERT-based Models for Sentiment Analysis on Indonesian Public Service Complaints. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*.
 - [6] Cahyadi, R., Rachman, F. H., & Hidayanto, A. N. (2022). Sentiment Analysis of Public Opinion on the COVID-19 Vaccine in Indonesia: A Comparative Study Using IndoBERT and XLM-RoBERTa. *IEEE International Conference on Communication, Networks and Satellite (COMNETSAT)*.
 - [7] Novita, S., Aji, D. P. P., & Permanasari, A. E. (2023). Enhancing Indonesian Sentiment Analysis by Combining RoBERTa with Convolutional Neural Networks. *Jurnal Teknologi dan Sistem Komputer (JUTISKOM)*.
 - [8] Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A New Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *Proceedings of the 28th International Conference on Computational Linguistics*.
 - [9] Putra, Y. K., Sari, D. P. D. P., & Hayadi, B. H. (2024). Analyzing Public Sentiment on Indonesia's New Capital City (IKN) using RoBERTa and Topic Modeling. *International Journal of Intelligent Systems and Applications in Engineering (IJISAE)*.
 - [10] Putra, W. P. N., Abdullah, S. R. S. S., & Abdullah, A. F. H. (2023). A Survey on Transformer-Based Pre-trained Language Models for Indonesian. *IEEE Access*.

