VOL. 3. NO. 04 JULI 2026

JOURNAL OF COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE (JCSAI)

PERBANDINGAN KINERJA MODEL TF-IDF SVM DAN INDOBERT DALAM KLASIFIKASI SENTIMEN CAMPURAN JUDUL BERITA IHSG MENGGUNAKAN PEMBELAJARAN

Muhammad Ismail Marzuki[1], Martanto[2], Fatihanursari Dikananda[3], Ade Rizki Rinaldi[4].

Program Studi Teknik Informatika[13]
Program Studi Manajemen Informatika[2]
Program Studi Rekayasa Perangkat Lunak[4]

STMIK IKMI Cirebon
https://ikmi.ac.id/page/18/?lang=de
muhammadismailmarzuki7@gmail.com

(*) Corresponding Author : muhammadismailmarzuki7@gmail.com
Published : 30 Juni 2026

Abstract—*This study compares the performance of the traditional TF-IDF + SVM model with the transformer-based IndoBERT model for sentiment classification of IHSG news headlines collected between 1 January and 30 September 2025 (1,380 samples). The dataset underwent a comprehensive preprocessing pipeline consisting of case folding, removal of numbers and punctuation, tokenization, stopword removal, and stemming, followed by semi-automatic keyword-based labeling into four sentiment categories: Positive, Negative, Neutral, and Mixed. The comparative experiments demonstrate that the fine-tuned IndoBERT model consistently outperforms TF-IDF + SVM—IndoBERT achieved an accuracy of 0.9614 and an F1-Macro score of 0.9608, whereas TF-IDF + SVM reached 0.8647 accuracy and 0.8323 F1-Macro. These results highlight the superiority of contextual transformer-based representations in capturing semantic relations, subtle nuances, and linguistic ambiguity frequently present in financial news headlines. IndoBERT's most notable advantage is its strong ability to detect Mixed sentiment, a minority class (8.41% of the dataset) that often carries dual or conflicting polarity cues. Despite the strong performance, this study identifies several limitations, including class imbalance, potential bias from pseudo-labeling, and the risk of overfitting, which necessitate cross-validation and external corpus testing. Based on these empirical findings, practical recommendations include adopting IndoBERT for financial sentiment analysis systems in Indonesia, enhancing label verification, applying data augmentation for minority classes, and exploring multi-label or hybrid approaches to improve the detection of mixed sentiment. The contribution of this research lies in providing empirical evidence comparing both approaches in the context of Indonesian IHSG news and offering practical guidelines for applying transformer models in the financial domain.*

Keywords*: Sentiment Analysis, Indobert, TF-IDF, SVM, IHSG, Mixed Sentiment.*

Abstrak—Penelitian ini membandingkan kinerja model tradisional TF-IDF + SVM dengan model berbasis transformer IndoBERT pada tugas klasifikasi sentimen judul berita IHSG (1 Januari–30 September 2025; 1.380 sampel), dengan penekanan khusus pada deteksi sentimen campuran (mixed sentiment). Data menjalani tahapan preprocessing komprehensif (case folding, penghapusan angka/tanda baca, tokenisasi, stopword removal, dan stemming) lalu dilabeli secara semi-otomatis berbasis kata kunci menjadi empat kelas: Positif, Negatif, Netral, dan Campuran. Eksperimen komparatif menunjukkan IndoBERT yang di-fine-tune secara konsisten mengungguli TF-IDF + SVM—IndoBERT mencapai akurasi 0,9614 dan F1-Macro 0,9608, sementara TF-IDF + SVM tercatat akurasi 0,8647 dan F1-Macro 0,8323—menunjukkan superioritas representasi kontekstual dalam menangkap relasi semantik dan ambiguitas linguistik pada headline finansial. Keunggulan IndoBERT paling nyata pada kemampuan mendeteksi kelas Campuran, yang meskipun minoritas (8,41% dari dataset) sering memiliki nuansa polaritas ganda. Meski demikian, penelitian ini mengidentifikasi keterbatasan penting: ketidakseimbangan kelas, potensi bias dari pelabelan pseudo-otomatis, serta risiko overfitting yang memerlukan validasi silang dan pengujian pada korpus eksternal. Berdasarkan temuan empiris, rekomendasi praktis meliputi penerapan IndoBERT untuk sistem analisis sentimen finansial di Indonesia dengan prosedur verifikasi label, augmentasi data untuk kelas minoritas, serta eksperimen multi-label atau hibrid untuk memperkuat deteksi mixed sentiment.

Kontribusi penelitian ini terletak pada bukti empiris perbandingan kedua pendekatan dalam konteks berita IHSG berbahasa Indonesia dan pedoman praktis untuk penerapan model berbasis transformer di domain finansial.

Kata Kunci: Analisis Sentimen, Indobert, TF-IDF, SVM, IHSG, Mixed Sentiment.

INTRODUCTION

Perkembangan teknologi informasi telah mendorong analisis sentimen menjadi komponen krusial dalam memahami dinamika pasar keuangan modern. Sentimen dalam berita ekonomi terbukti berkorelasi kuat dengan pergerakan Indeks Harga Saham Gabungan (IHSG). Analisis terhadap judul berita yang singkat dan informatif sangat penting karena sering kali memuat persepsi instan pelaku pasar terhadap kondisi ekonomi terkini.

Penelitian terkini menunjukkan dominasi model berbasis *transformer* seperti FinBERT dan IndoBERT dalam memahami konteks semantik teks keuangan dibandingkan model klasik seperti *Support Vector Machine* (SVM). Studi oleh [1]memperlihatkan bahwa model bahasa besar yang di-*fine-tune* pada domain keuangan mampu mengungguli *baseline* tradisional dalam klasifikasi multibahasa. Hal ini diperkuat oleh temuan [2] serta [3] bahwa sentimen berita berpengaruh langsung terhadap volatilitas pasar. Namun, sebagian besar studi tersebut masih berfokus pada polaritas sentimen tunggal (positif, negatif, atau netral).

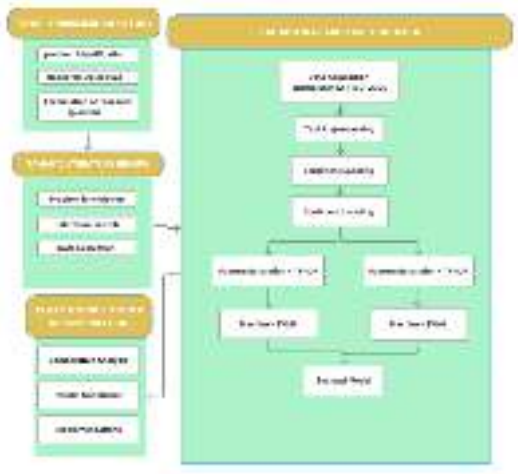
Dalam konteks Bahasa Indonesia, penelitian oleh [4]dan [5] mengidentifikasi tantangan besar berupa keterbatasan korpus spesifik domain dan adanya *semantic drift* pada model *transformer* [6]). Meskipun teknik *machine learning* klasik seperti SVM tetap menjadi *baseline* yang relevan [7]. model ini memiliki keterbatasan signifikan dalam menangani bahasa keuangan yang kompleks dan ambivalen.

Permasalahan utama yang sering terabaikan dalam penelitian sebelumnya adalah fenomena *mixed sentiment* (sentimen campuran) pada judul berita IHSG. Sering kali, sebuah judul berita memuat dua sentimen yang kontradiktif dalam satu kalimat pendek, misalnya berita yang mengabarkan kenaikan indeks namun disertai aksi jual asing. Model berbasis fitur leksikal seperti TF-IDF dengan SVM sering kali gagal menangkap nuansa kontradiktif ini karena hanya mengandalkan frekuensi kata tanpa memahami konteks kalimat secara utuh. Sementara itu, efektivitas IndoBERT dalam menangani kelas "campuran" pada domain finansial Indonesia masih memerlukan pengujian empiris yang lebih mendalam.

Oleh karena itu, penelitian ini bertujuan untuk melakukan kajian komparatif antara pendekatan klasik (TF-IDF Dengan SVM) dan modern (IndoBERT). Kebaruan (*novelty*) dari penelitian ini terletak pada pengujian model terhadap klasifikasi empat kelas, dengan penekanan khusus pada kemampuan model dalam mengenali sentimen campuran. Analisis ini tidak hanya mengukur perbedaan performa secara statistik melalui *McNemar Test*, tetapi juga mengevaluasi sejauh mana *contextual embedding* dapat menyelesaikan ambiguitas pada teks berita pasar keuangan Indonesia.

MATERIALS AND METHODS

Desain penelitian yang digunakan dalam studi ini untuk menganalisis sentimen pada judul berita IHSG. Penelitian menerapkan pendekatan kuantitatif dengan metode eksperimen komparatif, melibatkan serangkaian tahapan mulai dari pengumpulan data, *preprocessing*, pelabelan, hingga pelatihan serta evaluasi model. Alur penelitian divisualisasikan melalui Gambar 1 berikut :



Gambar 1. Alur Penelitian

Desain penelitian yang ditunjukkan pada Gambar 1 menggambarkan kerangka metodologis eksperimental komparatif yang digunakan untuk membandingkan kinerja dua pendekatan utama dalam analisis sentimen, yaitu model *machine learning* klasik (Jalur A) dan model *deep learning* berbasis *transformer* (Jalur B). Penelitian ini bersifat kuantitatif karena memungkinkan

pengukuran performa model secara terstandar melalui metrik seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Secara rinci, alur dalam desain penelitian tersebut dijelaskan sebagai berikut:

Tahap 1 (*Preliminary Study*): Merupakan fase awal yang difokuskan pada identifikasi masalah utama terkait ambiguitas sentimen berita, penetapan tujuan penelitian, serta perumusan pertanyaan riset sebagai landasan dasar studi.

Tahap 2 (*Literature Review*): Melibatkan aktivitas formulasi masalah secara akademik, pencarian literatur yang relevan, serta evaluasi data dari penelitian terdahulu guna memperkuat basis teoritis dan mengidentifikasi celah penelitian.

Tahap 3 (*Research Methodology*): Merupakan inti dari proses eksperimen yang dimulai dengan akuisisi data judul berita IHSG periode 2025 melalui teknik *web scraping*. Data mentah tersebut kemudian melewati tahap *text preprocessing* untuk pembersihan dan pelabelan sentimen secara semi-otomatis. Bagian krusial pada tahap ini adalah pengujian dua jalur paralel: Jalur A menggunakan ekstraksi fitur TF-IDF yang dilatih dengan algoritma SVM, sementara Jalur B menerapkan teknologi *transformer* melalui *fine-tuning* indobert.

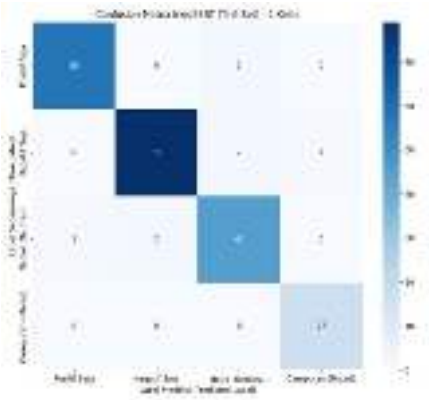
Tahap 4 (*Conclusion & Recommendation*): Tahap akhir ini mencakup analisis komparatif hasil performa kedua model untuk membuktikan hipotesis penelitian. Hasil tersebut kemudian digunakan untuk menarik kesimpulan metodologis dan memberikan rekomendasi praktis bagi pengembangan analisis sentimen di domain finansial di masa mendatang.

RESULTS AND DISCUSSION

Pada tahap kedua, penelitian dilakukan dengan mengimplementasikan model *State-of-the-Art* (SOTA) untuk Bahasa Indonesia, yaitu IndoBERT. Berbeda dengan SVM yang berbasis frekuensi kata, IndoBERT melalui proses *fine-tuning* mampu memahami hubungan semantik antar kata dan konteks kalimat secara bidireksional. Hasil evaluasi model IndoBERT pada 207 data uji dipaparkan sebagai berikut:

Matriks Kebingungan (*Confusion Matrix*) IndoBERT

Performa prediksi model IndoBERT terhadap label sebenarnya (*true labels*) dapat dilihat pada Gambar 2

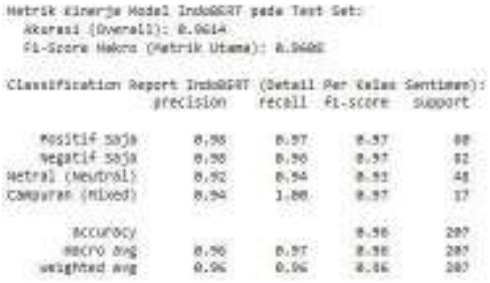


Gambar 2 Visualisasi Confusion Matrix IndoBERT

Berdasarkan Gambar 2 terlihat bahwa IndoBERT memiliki tingkat presisi yang sangat tinggi di seluruh kelas dengan sebaran kesalahan yang sangat minim:

Sentimen Positif: Dari 60 data, 58 berhasil diprediksi benar, dengan hanya 2 data yang salah dikategorikan sebagai Netral. Sentimen Negatif: Dari 82 data, 79 berhasil diprediksi benar. Terdapat 2 data meleset ke Netral dan 1 data ke Campuran. Sentimen Netral: Dari 48 data, 45 berhasil diprediksi benar. Sentimen Campuran: Model berhasil memprediksi seluruh 17 data secara sempurna (100% akurat) tanpa ada satu pun data yang salah klasifikasi.

Laporan Klasifikasi (*Classification Report*) IndoBERT Secara kuantitatif, keunggulan IndoBERT tercermin dalam metrik evaluasi yang disajikan pada Gambar 3



Metrik Kinerja Model IndoBERT pada Test Set:
Akurasi (Overall): 0.9614
F1-Score Makro (Metrik Utama): 0.9608

Classification Report IndoBERT (Detail Per Kelas Sentimen):

	precision	recall	f1-score	support
Positif Saja	0.98	0.97	0.97	60
Negatif Saja	0.98	0.96	0.97	82
Netral (Neutral)	0.92	0.94	0.93	48
Campuran (Mixed)	0.94	1.00	0.97	17
accuracy			0.96	207
macro avg	0.96	0.97	0.96	207
weighted avg	0.96	0.96	0.96	207

Gambar 3 Hasil Evaluasi Model Fine-Tuning IndoBERT

Berdasarkan tabel pada Gambar 3 model IndoBERT mencapai performa yang sangat impresif , Model IndoBERT mencapai Akurasi sebesar 96,14% dan Macro F1-Score sebesar 96,08%. Temuan yang paling krusial pada tahap ini adalah kemampuan IndoBERT dalam mengatasi

kelemahan model SVM pada kelas Campuran (Mixed). Jika sebelumnya pada SVM nilai *recall* kelas campuran hanya 0,59, pada IndoBERT nilai *recall* meningkat tajam menjadi 1,00.

Hal ini membuktikan bahwa arsitektur *self-attention* pada IndoBERT sangat efektif dalam menangkap ambiguitas dan kontradiksi kata dalam satu kalimat (seperti teks yang berisi pujian sekaligus keluhan), yang merupakan karakteristik utama dari sentimen campuran. Secara keseluruhan, IndoBERT menunjukkan konsistensi performa di angka 90% ke atas untuk semua metrik, yang mengindikasikan bahwa model ini sangat handal untuk klasifikasi sentimen multi-kelas pada data pasar modal yang bersifat dinamis.

Setelah melakukan pengujian pada model *baseline* (TF-IDF + SVM) dan model *fine-tuning* IndoBERT, dilakukan analisis komparatif untuk melihat sejauh mana peningkatan performa yang dihasilkan oleh arsitektur berbasis *Transformer*. Ringkasan perbandingan metrik kinerja kedua model disajikan pada Tabel 4.1:

Tabel 1 Ringkasan Komparatif Kinerja Model

Metrik Evaluasi	TF-IDF + SVM (Baseline)	IndoBERT (Fine-Tuning)	Selisih Peningkatan
Akurasi (Accuracy)	0,8647 (86,47%)	0,9614 (96,14 %)	+ 9,67 %
Macro F1-Score	0,8323 (83,23%)	0,9608 (96,08 %)	+ 12,85 %

Berdasarkan Tabel 1 dapat ditarik beberapa poin analisis komparatif sebagai berikut:

Peningkatan Akurasi Keseluruhan: Model IndoBERT memberikan peningkatan akurasi sebesar 9,67% dibandingkan dengan SVM. Hal ini menunjukkan bahwa IndoBERT secara umum lebih handal dalam mengklasifikasikan data sentimen pasar modal yang memiliki karakteristik bahasa yang tidak baku, singkatan, dan istilah teknis bursa.

Stabilitas pada Macro F1-Score: Peningkatan paling signifikan terlihat pada metrik Macro F1-Score, yaitu sebesar 12,85%. Karena Macro F1-Score menghitung rata-rata performa tiap kelas tanpa mempedulikan jumlah sampel, peningkatan besar ini membuktikan bahwa IndoBERT jauh lebih stabil dalam menangani seluruh kelas sentimen secara adil, termasuk kelas yang memiliki jumlah data sedikit seperti kelas Campuran (Mixed).

Kemampuan Menangkap Konteks (Semantik vs Frekuensi): SVM yang mengandalkan TF-IDF hanya menghitung frekuensi kemunculan kata tanpa memahami hubungan antar kata. Sebaliknya, IndoBERT mampu memahami konteks kalimat secara utuh. Sebagai contoh, pada kelas Campuran yang sering kali memiliki kata-kata kontradiktif (seperti "IHSG ambruk tapi tetap borong saham"), SVM cenderung bingung dan salah mengklasifikasikan teks tersebut, sementara IndoBERT mampu mengenalinya sebagai sentimen campuran dengan akurasi 100% (*recall* 1.00).

Efektivitas Model: Meskipun model SVM sudah memberikan hasil yang cukup baik (di atas 80%), penggunaan *fine-tuning* pada model bahasa IndoBERT terbukti membawa performa ke tingkat yang lebih tinggi (*state-of-the-art*). Hal ini mengonfirmasi bahwa untuk dataset yang memiliki kompleksitas bahasa tinggi seperti opini publik di media sosial/berita ekonomi, model berbasis *Transformer* merupakan pilihan yang lebih unggul dibandingkan metode *machine learning* konvensional.

CONCLUSION

Berdasarkan hasil penelitian, pengujian, dan analisis yang telah dilakukan mengenai perbandingan kinerja model TF-IDF + SVM dan *fine-tuning* IndoBERT dalam klasifikasi sentimen campuran pada judul berita IHSG, maka dapat ditarik beberapa kesimpulan sebagai berikut:

Performa Keseluruhan Model: Model IndoBERT menunjukkan performa yang jauh lebih unggul dibandingkan model *baseline* TF-IDF + SVM. Hal ini dibuktikan dengan perolehan akurasi sebesar 96,14% untuk IndoBERT, sementara model SVM hanya mencapai 86,47%. Penggunaan arsitektur *Transformer* pada IndoBERT terbukti mampu menangkap konteks bahasa Indonesia yang kompleks pada berita pasar modal jauh lebih baik daripada metode berbasis frekuensi kata (TF-IDF).

Efektivitas Klasifikasi Sentimen Campuran: IndoBERT berhasil mengatasi keterbatasan utama model SVM dalam mengidentifikasi sentimen "Campuran" (*Mixed*). Pada kelas ini, IndoBERT mencapai nilai *recall* sebesar 1,00 (100%), yang berarti seluruh data sentimen campuran berhasil

diidentifikasi dengan benar. Sebaliknya, model SVM hanya mampu mencapai *recall* sebesar 0,59 (59%), di mana model tersebut sering salah mengklasifikasikan sentimen campuran ke dalam kategori positif atau negatif karena hanya mengandalkan keberadaan kata kunci tanpa memahami konteks kontradiksi dalam kalimat.

Signifikansi Statistik: Berdasarkan uji validasi menggunakan McNemar Test, diperoleh nilai Chi-Square sebesar 15,041 yang lebih besar dari nilai tabel (3,841) dengan *p-value* < 0,05. Hal ini memberikan bukti ilmiah yang kuat bahwa perbedaan performa antara kedua model bersifat signifikan secara statistik dan bukan terjadi karena faktor kebetulan pada dataset. Peningkatan performa sebesar 9,67% pada akurasi dan 12,85% pada *Macro F1-Score* menjustifikasi transisi dari pembelajaran mesin konvensional ke metode *deep learning* berbasis konteks.

REFERENCE

[1] A. M. Ardekani, J. Bertz, C. Bryce, M. Dowling, and S. Long, "FinSentGPT: A universal financial sentiment engine?," *International Review of Financial Analysis*, vol. 94, p. 103291, 2024, doi: https://doi.org/10.1016/j.irfa.2024.103291.

[2] M. P. Cristescu, R. A. Nerisanu, D. A. Mara, and S.-V. Oprea, "Using Market News Sentiment Analysis for Stock Market Prediction," *Mathematics*, vol. 10, no. 22, 2022, doi: 10.3390/math10224255.

[3] L. T. Vu, D. N. Pham, H. T. Kieu, and T. T. T. Pham, "Sentiments Extracted from News and Stock Market Reactions in Vietnam," *International Journal of Financial Studies*, vol. 11, no. 3, Sep. 2023, doi: 10.3390/ijfs11030101.

[4] S. Cahyawijaya *et al.*, "NusaCrowd: A Call for Open and Reproducible NLP Research in Indonesian Languages," Aug. 2022.

[5] N. P. I. Maharani, Y. Yustiawan, F. C. Rochim, and A. Purwarianti, "Domain-Specific Language Model Post-Training for Indonesian Financial NLP," Oct. 2023.

[6] N. H. Setiono and Y. Sari, "Exploring the Impact of Back-Translation on BERT's Performance in Sentiment Analysis of Code-Mixed Language Data," *Indonesian Journal of Computing and Cybernetics Systems*, vol. 15, no. 2, pp. 101–118, 2024, doi: 10.22146/ijccs.104757.

[7] W. Long, J. Gao, K. Bai, and Z. Lu, "A hybrid model for stock price prediction based on multi-view heterogeneous data," *Financial Innovation*, vol. 10, no. 1, p. 48, 2024, doi: 10.1186/s40854-023-00519-w.