

KLASIFIKASI SENTIMEN HEADLINE BERITA PASAR FINANSIAL CNBC INDONESIA MENGGUNAKAN MULTINOMIAL LOGISTIC REGRESSION DAN TF-IDF

Muhammad Yusra^{*1}, Roberto Kaban²

¹Program Studi Rekayasa Perangkat Lunak, Institut Teknologi dan Bisnis Indonesia, Indonesia

²Program Studi Teknik Informatika, Institut Teknologi dan Bisnis Indonesia, Indonesia

*e-mail: muhammadyusra1102@gmail.com

Abstrak

Penelitian ini bertujuan mengklasifikasikan sentimen headline berita pasar finansial Indonesia dan mengevaluasi performa model Multinomial Logistic Regression berbasis representasi fitur *Term Frequency Inverse Document Frequency* (TF-IDF). Dataset penelitian mencakup 9.819 headline berita CNBC Indonesia yang dipublikasikan selama Januari 2024 hingga Maret 2025 dan dianotasi secara manual ke dalam tiga kategori sentimen, yaitu positif (29,4%), netral (44,4%), dan negatif (26,2%). Representasi teks dibangun menggunakan TF-IDF dengan ukuran kosakata maksimum sebanyak 5.000 kata. Model memperoleh akurasi sebesar 82,43%, dengan nilai macro precision 0,8200, macro recall 0,8233, dan macro F1-score 0,8200. Pengujian menggunakan skema 5-fold cross-validation menghasilkan rata-rata akurasi 82,04%, yang menunjukkan bahwa model memiliki kemampuan generalisasi yang baik. Analisis koefisien menunjukkan bahwa istilah anjlok dan ambruk berkontribusi terhadap prediksi sentimen negatif, sedangkan borong dan melesat menjadi indikator sentimen positif. Kesalahan klasifikasi terutama ditemukan pada headline yang mengandung ambiguitas konteks. Temuan ini menunjukkan bahwa pendekatan Multinomial Logistic Regression berbasis TF-IDF tidak hanya memberikan performa klasifikasi yang baik, tetapi juga menghasilkan model yang interpretatif sehingga dapat mendukung pengembangan analisis sentimen pada berita pasar finansial Indonesia.

Kata kunci: Analisis Sentimen; Berita Pasar Finansial; Klasifikasi Teks; Multinomial Logistic Regression; TF-IDF.

Abstract

This study aims to classify the sentiment of Indonesian financial market news headlines and evaluate the performance of a Multinomial Logistic Regression model based on *Term Frequency–Inverse Document Frequency* (TF-IDF) feature representation. The dataset consists of 9,819 CNBC Indonesia news headlines published between January 2024 and March 2025, which were manually annotated into three sentiment categories: positive (29.4%), neutral (44.4%), and negative (26.2%). Text representation was constructed using TF-IDF with a maximum vocabulary size of 5,000 terms. The proposed model achieved an accuracy of 82.43%, with a macro precision of 0.8200, macro recall of 0.8233, and macro F1-score of 0.8200. Furthermore, 5-fold cross-validation produced an average accuracy of 82.04%, indicating good generalization capability. Coefficient analysis revealed that the terms *anjlok* and *ambruk* contributed to negative sentiment prediction, whereas *borong* and *melesat* served as indicators of positive sentiment. Classification errors were primarily observed in headlines containing contextual ambiguity. These findings demonstrate that the TF-IDF-based Multinomial Logistic Regression approach not only provides good classification performance but also produces an interpretable model, thereby supporting the advancement of sentiment analysis for Indonesian financial market news.

Keywords: Financial Market News; Multinomial Logistic Regression; Sentiment Analysis; Text Classification; TF-IDF.

1. PENDAHULUAN

Transformasi digital telah mengubah cara informasi dihasilkan, disebarluaskan, dan dimanfaatkan dalam berbagai sektor, termasuk pasar keuangan. Perkembangan teknologi informasi memungkinkan investor dan pelaku pasar memperoleh informasi ekonomi secara cepat melalui berbagai media daring sehingga berita menjadi salah satu faktor penting dalam pengambilan keputusan investasi [1]. Kemajuan pada bidang *Natural Language Processing* (NLP) dan pembelajaran mesin mendorong otomatisasi analisis sentimen sehingga opini atau kecenderungan suatu teks dapat diidentifikasi secara lebih efisien [2], [3]. Penelitian terdahulu menunjukkan bahwa sentimen berita memiliki pengaruh terhadap pergerakan pasar saham dan perilaku investor, sehingga analisis sentimen menjadi pendekatan yang banyak digunakan dalam bidang keuangan [4], [5]. Di Indonesia, CNBC Indonesia merupakan salah satu portal berita yang banyak digunakan untuk memperoleh informasi mengenai pasar saham, nilai tukar, dan kondisi ekonomi. Namun, tingginya jumlah berita yang dipublikasikan setiap hari menyebabkan proses identifikasi sentimen secara manual menjadi kurang efisien. Berbagai penelitian telah menerapkan model berbasis *transformer* untuk

analisis sentimen dan menunjukkan hasil yang baik [6], [7]. Di sisi lain, algoritma klasifikasi tradisional seperti Logistic Regression, Support Vector Machine, dan Naïve Bayes juga masih banyak digunakan karena mampu memberikan performa yang kompetitif pada berbagai tugas analisis sentimen [8], [9]. Meskipun demikian, penelitian yang secara khusus membahas headline berita pasar finansial Indonesia masih relatif terbatas. Selain belum tersedianya dataset berskala besar dengan pelabelan sentimen yang konsisten, sebagian besar penelitian juga lebih berfokus pada peningkatan akurasi klasifikasi dibandingkan interpretasi hasil model [9], [6]. Perbandingan karakteristik metode, domain data, dan fokus evaluasi pada penelitian-penelitian tersebut dirangkum pada Tabel 1 berikut, yang sekaligus menegaskan kesenjangan penelitian (*research gap*) yang menjadi dasar penelitian ini.

Tabel 1. Perbandingan Penelitian Terdahulu dan Kesenjangan Penelitian (*Research Gap*)

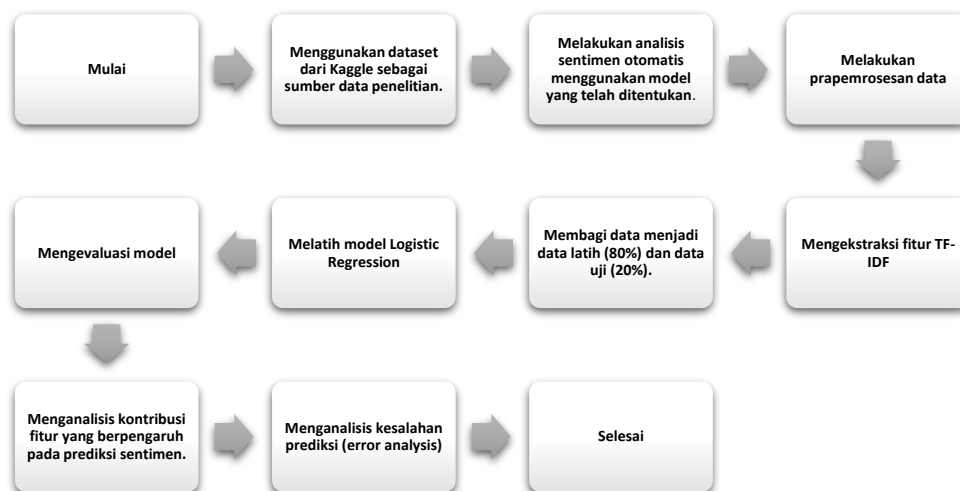
No	Peneliti (Ref.)	Metode	Domain / Bahasa Data	Fokus Evaluasi	Kesenjangan (Gap)
1	Araci [7]	FinBERT (transformer)	Berita finansial, Bahasa Inggris	Akurasi klasifikasi	Tidak membahas interpretabilitas fitur; berbahasa Inggris
2	Kadir & Fairuzabadi [10]	TF-IDF + Logistic Regression	Ulasan Shopee, Bahasa Indonesia	Akurasi	Domain e-commerce, bukan berita finansial
3	Hutagalung & Rasiban [9]	Logistic Regression & SVM	FiQA & Financial PhraseBank, Bahasa Inggris	Akurasi	Berbahasa Inggris; tanpa analisis kesalahan
4	Firizkiansah dkk. [11]	Logistic Regression + TF-IDF + SMOTE	Data teks umum	Akurasi vs. penanganan imbalance	Tidak spesifik pada domain berita finansial
5	Mulyono & Saprudin [12]	Logistic Regression	Komentar YouTube, Bahasa Indonesia	Akurasi	Bukan domain berita finansial; tanpa interpretasi koefisien
6	Sinha dkk. (SEntFiN) [6]	Lexicon & model berbasis BERT	Berita finansial global, Bahasa Inggris	Akurasi & F1-score	Berbahasa Inggris; belum ada studi pada headline finansial Indonesia
7	Penelitian ini	Multinomial Logistic Regression + TF-IDF	Headline CNBC Indonesia, Bahasa Indonesia	Akurasi, precision, recall, F1, analisis koefisien, error analysis	Mengisi kesenjangan domain berita finansial Indonesia dengan penekanan pada interpretabilitas model

Selain itu, sebagian besar penelitian analisis sentimen dalam bidang keuangan masih lebih mengarah pada pengevaluasian kinerja model menggunakan metrik seperti akurasi, presisi, recall, dan F1-score, sedangkan pembahasan mengenai tingkat interpretabilitas model masih terbatas [6], [9]. Meski demikian, dalam bidang pasar finansial, kemampuan untuk menjelaskan bagaimana setiap fitur berkontribusi terhadap hasil klasifikasi merupakan aspek yang penting, karena dapat membantu memahami pola bahasa yang memengaruhi pembentukan sentimen [6], [7]. Oleh karena itu, diperlukan penelitian yang tidak hanya mengevaluasi kemampuan klasifikasi, tetapi juga menganalisis karakteristik yang memengaruhi keputusan model serta pola kesalahan dalam proses klasifikasi yang terjadi. Interpretabilitas menjadi krusial dalam konteks pasar finansial karena keputusan investasi maupun kebijakan yang diambil berdasarkan hasil klasifikasi otomatis perlu dapat dipertanggungjawabkan. Model yang bersifat black-box berisiko menghasilkan keputusan yang sulit ditelusuri penyebabnya, sedangkan model berbasis TF-IDF dan Logistic Regression memungkinkan penelusuran kontribusi setiap term terhadap prediksi secara eksplisit melalui koefisiennya [10], [12].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan menganalisis sentimen headline berita pasar finansial Indonesia menggunakan algoritma Multinomial Logistic Regression dengan ekstraksi fitur TF-IDF. Penelitian ini menggunakan dataset berisi 9.819 headline berita CNBC Indonesia yang telah dianotasi secara manual [13]. Kontribusi penelitian meliputi penyediaan dataset domain-spesifik, evaluasi performa model melalui berbagai metrik pengujian, analisis koefisien TF-IDF untuk mengidentifikasi kata yang paling berpengaruh terhadap setiap kelas sentimen, serta *error analysis* guna mengidentifikasi penyebab kesalahan klasifikasi. Hasil penelitian diharapkan dapat menjadi referensi dalam pengembangan model analisis sentimen berita pasar finansial Indonesia yang lebih akurat dan mudah diinterpretasikan.

2. METODE

Metode penelitian yang digunakan dalam studi ini menerapkan pendekatan kuantitatif berbasis machine learning untuk mengklasifikasikan sentimen headline berita pasar keuangan Indonesia ke dalam tiga kategori, yaitu positif, netral, dan negatif. Proses penelitian disusun secara sistematis melalui beberapa tahapan, meliputi pengumpulan data menggunakan dataset sekunder, prapemrosesan teks, ekstraksi fitur menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), pelatihan model Multinomial Logistic Regression, serta evaluasi performa model menggunakan metrik akurasi, precision, recall, F1-score, dan 5-fold cross-validation. Alur penelitian secara keseluruhan ditunjukkan pada Gambar 1.



Gambar 1. Alur Penelitian (*Flowchart*) Analisis Sentimen Headline CNBC Indonesia

2.1 Populasi dan Sampel

Objek penelitian ini menggunakan dataset sekunder berupa 9.819 entri judul berita pasar saham Indonesia yang dipublikasikan oleh portal CNBC Indonesia dalam rubrik "Pasar". Data penelusuran tersebut mencakup rentang waktu dari 1 Januari 2024 sampai dengan 31 Maret 2025. Seluruh data tekstual tersebut telah dianotasi secara manual oleh pembuat dataset [13] ke dalam tiga kategori target klasifikasi sentimen, yaitu kelas positif, netral, dan negatif.

Tabel 2. Distribusi Hasil Analisis Sentimen Headline CNBC Indonesia.

Kategori	Jumlah Headline	Persentase %	Keterangan
Negatif	2.576	26,2%	Mengindikasikan penurunan, kerugian, risiko
Netral	4.356	44,4%	Bersifat informatif, faktual, tidak ada muatan jelas
Positif	2.887	29,4%	Mengindikasikan kenaikan, keuntungan, prospek positif
TOTAL	9.819	100%	Dataset Lengkap

Orientasi pemilihan CNBC Indonesia sebagai basis sumber data sekunder didasarkan pada tiga argumen metodologis utama, yaitu perannya sebagai media referensi ekonomi terkemuka di sektor domestik yang menjamin validitas serta kredibilitas informasi, representasi data yang inklusif melalui judul berita yang merekam dinamika lintas sektor keuangan mulai dari pasar saham, obligasi, forex, cryptocurrency, hingga ekonomi makro, serta frekuensi sekaligus konsistensi publikasi harian yang teratur guna meminimalkan risiko ketimpangan (*gap*) informasi saat proses akuisisi data dilakukan. Melalui pembatasan jendela waktu penanggalan dari 1 Januari 2024 hingga 31 Maret 2025, diperoleh total sampel akhir yang bersifat masif yaitu sebanyak 9.819 headline berita pasar saham.

2.2 Sumber Data dan Teknik Pengumpulan Data

Penelitian ini memanfaatkan dataset yang tersedia pada platform Kaggle dan disusun oleh pembuat dataset melalui proses *web scraping* serta anotasi sentimen secara manual (<https://www.kaggle.com/datasets/triagung/cnbc-indonesia-stock-news-sentiment-dataset>). Dataset tersebut dikompilasi oleh pembuatnya melalui proses *web scraping* menggunakan Python dan BeautifulSoup untuk memperoleh judul, tanggal publikasi, serta isi berita, kemudian dilanjutkan dengan anotasi sentimen

secara manual menggunakan Google Sheets. Karena anotasi hanya dilakukan oleh satu anotator, *nilai inter-annotator agreement* tidak tersedia sehingga menjadi keterbatasan dalam reliabilitas label sekaligus peluang pengembangan pada penelitian selanjutnya. Pelabelan sentimen didasarkan pada tiga kategori kontekstual, yaitu negatif untuk headline yang mengindikasikan penurunan kinerja finansial, kerugian, risiko, atau kontraksi pasar; netral untuk berita yang bersifat objektif dan faktual tanpa kecenderungan sentimen; serta positif untuk headline yang menunjukkan penguatan nilai, keuntungan, atau prospek pasar yang optimistis.

2.3 Pra-pemrosesan Teks dan Ekstraksi fitur TF-IDF

Seluruh headline berita yang telah dianotasi menjalani tahapan pra-pemrosesan teks sebelum diekstraksi menjadi fitur numerik. Sebelum proses ekstraksi fitur dilakukan, seluruh headline diproses melalui beberapa tahapan pra-pemrosesan yang mencakup konversi huruf menjadi huruf kecil (*lowercasing*), penghapusan tanda baca, angka, dan karakter khusus, penghilangan kata tidak informatif (*stopword removal*) menggunakan daftar *stopword* Sastrawi, serta tokenisasi berbasis spasi (*whitespace tokenization*). Proses *stemming* maupun *lemmatisasi* tidak diterapkan agar bentuk kata asli tetap dipertahankan sehingga interpretasi koefisien model menjadi lebih mudah.

$$TF(t, d) = f(t, d) \quad (1)$$

$$IDF(t) = \log\left(\frac{N}{df(t)}\right) \quad (2)$$

$$TF-IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

Representasi numerik dokumen dibangun menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) dengan pembobotan frekuensi sublinear (sublinear TF) untuk mengurangi dominasi term yang sangat sering muncul. Konfigurasi ekstraksi fitur menggunakan batas maksimum 5.000 term, rentang n-gram (1,2) yang mencakup unigram dan bigram, serta pembobotan logaritmik (sublinear TF). Hasil proses tersebut menghasilkan matriks fitur TF-IDF berdimensi 7.855×5.000 yang digunakan sebagai data latih model. Nilai *max_features* sebesar 5.000 dipilih mengacu pada konfigurasi yang umum digunakan dalam penelitian klasifikasi teks berbahasa Indonesia berbasis TF-IDF [10], [11], dengan tujuan menjaga keseimbangan antara representativitas fitur dan kompleksitas model. Sementara itu, rentang n-gram (1,2) dipilih agar model mampu merepresentasikan baik kata tunggal (*unigram*) maupun pasangan kata (*bigram*) yang umum muncul pada headline berita pasar finansial, seperti "turun ke" atau "anjlok tajam", sehingga informasi kontekstual sederhana yang tidak dapat ditangkap oleh unigram saja tetap dapat dimanfaatkan dalam proses klasifikasi.

2.4 Klasifikasi Multinomial Logistic Regression

Penelitian ini menerapkan Multinomial Logistic Regression (MLR), yaitu pengembangan regresi logistik biner yang mampu menangani klasifikasi multikelas. Pada penelitian ini, MLR digunakan untuk mengklasifikasikan sentimen headline ke dalam tiga kategori, yaitu negatif, netral, dan positif. Model mengestimasi probabilitas setiap dokumen berdasarkan representasi fitur TF-IDF, kemudian kelas dengan probabilitas tertinggi ditetapkan sebagai hasil prediksi. Probabilitas setiap kelas dihitung menggunakan fungsi *softmax* sebagaimana ditunjukkan pada Persamaan (4).

$$P(y = k|x) = \frac{e^{w_k^T x + b_k}}{\sum_{j=1}^K e^{w_j^T x + b_j}} \quad (4)$$

Dengan x merupakan vektor fitur TF-IDF, w_k adalah bobot untuk kelas ke- k , w_k adalah bias, dan k merupakan jumlah kelas sentimen. Model dilatih menggunakan fitur TF-IDF hasil tahap pra-pemrosesan dengan konfigurasi *solver* *L-BFGS*, *max_iter* = 1000, dan *class_weight* = *balanced* untuk mengatasi ketidakseimbangan distribusi kelas. Selanjutnya, dataset dibagi menjadi data latih (80%) dan data uji (20%) menggunakan random sampling dengan *random_state* = 42 untuk menjamin reproduktibilitas penelitian. Pemilihan *solver* *L-BFGS* didasarkan pada efisiensinya untuk permasalahan optimasi berskala menengah dengan fitur berdimensi tinggi namun sparse, sebagaimana karakteristik matriks TF-IDF pada penelitian ini,

serta konsistensinya dalam mendukung fungsi multinomial (*softmax*) secara native tanpa memerlukan skema *one-vs-rest*.

2.5 Confusion Matrix

Evaluasi hasil klasifikasi dilakukan menggunakan *Confusion Matrix* untuk membandingkan label prediksi dengan label aktual sehingga kesalahan klasifikasi dapat diidentifikasi. Matriks ini menampilkan jumlah prediksi yang benar dan salah untuk setiap kelas sehingga dapat digunakan untuk menganalisis pola kesalahan klasifikasi yang dihasilkan model. Dalam penelitian ini, *Confusion Matrix* digunakan untuk mengevaluasi hasil klasifikasi sentimen ke dalam kelas negatif, netral, dan positif. Nilai pada diagonal utama menunjukkan prediksi yang sesuai dengan label aktual, sedangkan nilai di luar diagonal menunjukkan kesalahan klasifikasi.

2.6 Metrik Evaluasi

Performa model diukur menggunakan metrik Accuracy, Precision, Recall, dan F1-Score untuk memberikan gambaran menyeluruh mengenai kemampuan klasifikasi. Accuracy digunakan untuk mengukur proporsi prediksi yang benar terhadap keseluruhan data uji. Precision menunjukkan ketepatan hasil prediksi model, sedangkan Recall mengukur kemampuan model dalam mengidentifikasi data pada kelas yang sesuai. Selanjutnya, F1-Score digunakan sebagai ukuran keseimbangan antara Precision dan Recall yang dihitung menggunakan rata-rata harmonik keduanya sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

$$Recall = \frac{TP}{FP + FN} \quad (7)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

2.7 Spesifikasi Lingkungan Eksperimen

Seluruh tahap pra-pemrosesan, ekstraksi fitur, pelatihan, dan evaluasi model dilakukan menggunakan Python 3.14.0 dengan pustaka *scikit-learn* 1.8.0, *pandas* 3.0.2, *NumPy* 2.4.1, dan Sastrawi. Eksperimen dijalankan pada perangkat MacBook berbasis Apple M3 dengan memori 8 GB yang menjalankan sistem operasi macOS Sequoia 15.6. Seluruh proses komputasi dilakukan menggunakan CPU tanpa akselerator GPU. Pencantuman spesifikasi perangkat keras dan perangkat lunak ini bertujuan meningkatkan reproduktibilitas penelitian, sehingga eksperimen dapat direplikasi pada lingkungan komputasi yang serupa.

2.8 Algoritma Penelitian

Untuk memperjelas alur implementasi metode, tahapan klasifikasi sentimen dirangkum dalam bentuk algoritma penelitian sebagaimana ditunjukkan pada Algoritma 1.

Algoritma 1. Tahapan Klasifikasi Sentimen Headline Berita Pasar Finansial CNBC Indonesia

Input : Dataset headline berita pasar finansial CNBC Indonesia sebanyak 9.819 (data berlabel).

Output : Model Multinomial Logistic Regression terlatih beserta metrik evaluasi performa

1. Baca dataset headline beserta label sentimen.
2. Lakukan pra-pemrosesan yang meliputi *lowercasing*, pembersihan karakter, *stopword removal*, dan tokenisasi.
3. Bentuk representasi fitur menggunakan TF-IDF.
4. Bagi dataset menjadi data latih (80%) dan data uji (20%).
5. Inisialisasi model Multinomial Logistic Regression.
6. Latih model menggunakan data latih.
7. Pemilihan parameter tersebut didasarkan pada praktik umum dalam klasifikasi teks berbasis TF-IDF serta mempertimbangkan keseimbangan antara kualitas representasi fitur, efisiensi komputasi, dan stabilitas proses optimasi.
8. Hitung nilai Accuracy, Precision, Recall, dan F1-Score.
9. Lakukan evaluasi menggunakan 5-fold Cross Validation.
10. Analisis koefisien model untuk mengidentifikasi term yang paling berpengaruh.

Pada implementasinya, representasi fitur TF-IDF menggunakan parameter *max_features* = 5000, *ngram_range* = (1,2), dan *sublinear_tf* = True. Model Multinomial Logistic Regression dikonfigurasi menggunakan solver *lbfgs*, *max_iter* = 1000, dan *class_weight* = *balanced*. Pemilihan parameter tersebut mengacu pada praktik umum dalam klasifikasi teks menggunakan TF-IDF dan Logistic Regression serta mempertimbangkan keseimbangan antara kualitas representasi fitur, efisiensi komputasi, dan stabilitas proses optimasi.

3. HASIL PENELITIAN

3.1 Performa Keseluruhan Model

Model Logistic Regression Multinomial berbasis TF-IDF berhasil mencapai akurasi pengujian sebesar 82,43% pada data uji yang terdiri dari 1.964 sampel. Sementara itu, evaluasi menggunakan metode 5-fold cross-validation menunjukkan rata-rata akurasi sebesar 82,04% dengan variasi antar-lipatan yang relatif kecil, yaitu berkisar antara 81,35% hingga 82,75%. Hasil tersebut menunjukkan bahwa model memiliki tingkat kestabilan yang baik serta kemampuan generalisasi yang memadai. Selain itu, model memperoleh nilai Macro Precision sebesar 0,8200, Macro Recall sebesar 0,8233, dan Macro F1-Score sebesar 0,8200. Capaian performa ini menunjukkan bahwa kombinasi TF-IDF dan Logistic Regression mampu menghasilkan klasifikasi sentimen yang efektif pada data teks berbahasa Indonesia. Temuan ini sejalan dengan beberapa penelitian sebelumnya yang melaporkan bahwa Logistic Regression tetap menjadi metode yang kompetitif dan memiliki tingkat *interpretabilitas* yang tinggi dalam berbagai tugas analisis sentimen berbahasa Indonesia [11], [12].

Tabel 3. Ringkasan Konfigurasi dan Metrik Keseluruhan Model

Metrik	Nilai
Test Accuracy	82,43%
Mean 5-Fold CV Accuracy	82,04% $\pm$ 0,0046
Macro Precision	0,8200
Macro Recall	0,8233
Macro F1-Score	0,8200
Vocabulary TF-IDF	5.000 term

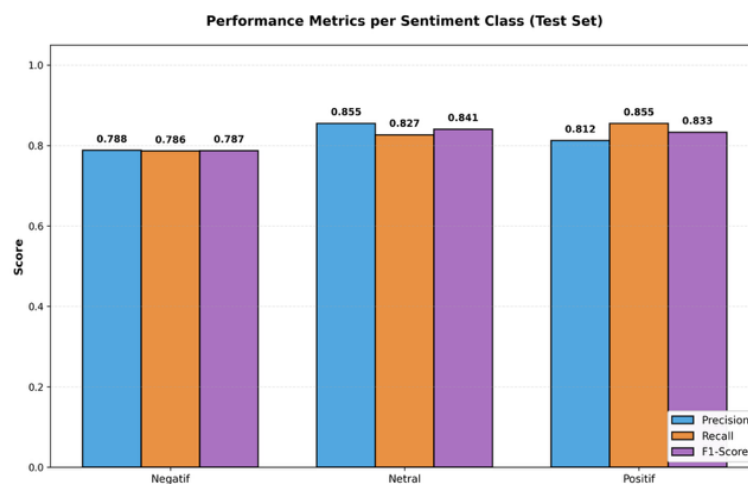
3.2 Performa Model per Kelas

Analisis performa pada setiap kelas menunjukkan adanya perbedaan karakteristik antar kategori sentimen. Detail hasil evaluasi untuk masing-masing kelas pada data pengujian disajikan pada Tabel 3.

Tabel 4. Performa Model Logistic Regression per Kelas Sentimen (Set Uji)

Kelas Sentiment	Presisi	Recall	F1-Score	Sample Uji
Negatif	0,79	0,79	0,79	515
Netral	0,86	0,83	0,84	871
Positif	0,81	0,85	0,83	578
Macro Avg.	0,82	0,82	0,82	1.964

Kelas Netral memperoleh performa terbaik dengan nilai Precision sebesar 0,86, Recall sebesar 0,83, dan F1-Score sebesar 0,84. Tingginya nilai presisi pada kelas Netral diduga dipengaruhi oleh dominasi jumlah data latih pada kategori tersebut, sehingga model mampu mengenali pola fitur netral secara lebih optimal. Sementara itu, kelas Positif mencatat nilai F1-Score sebesar 0,83, Precision sebesar 0,81, dengan Recall sebesar 0,85, yang merupakan nilai Recall tertinggi dibandingkan kelas lainnya. Hasil ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mendeteksi sentimen positif. Di sisi lain, kelas Negatif memperoleh nilai F1-Score terendah, yaitu sebesar 0,79, dengan nilai Precision dan Recall yang sama, yakni 0,79. Penurunan performa pada kelas Negatif mengindikasikan adanya kemiripan leksikal antara headline bersentimen negatif dengan kategori sentimen lainnya.



Gambar 2. Perbandingan Metrik Evaluasi per Kelas Sentimen

3.3 Hasil 5-Fold Cross Validation

Rincian nilai akurasi untuk masing-masing fold pada proses 5-fold cross-validation terhadap data latih sebelum evaluasi akhir disajikan pada Tabel 4.

Tabel 5. Hasil 5-Fold Cross-Validation pada Data Latih

Fold	Akurasi	Keterangan
Fold 1	0,8211	Di atas rata-rata
Fold 2	0,8218	Di atas rata-rata
Fold 3	0,8180	Di bawah rata-rata
Fold 4	0,8275	Tertinggi
Fold 5	0,8135	Terendah
Mean (Std Dev)	0,8204(±0,0047)	Variasi antarlipatan sangat kecil

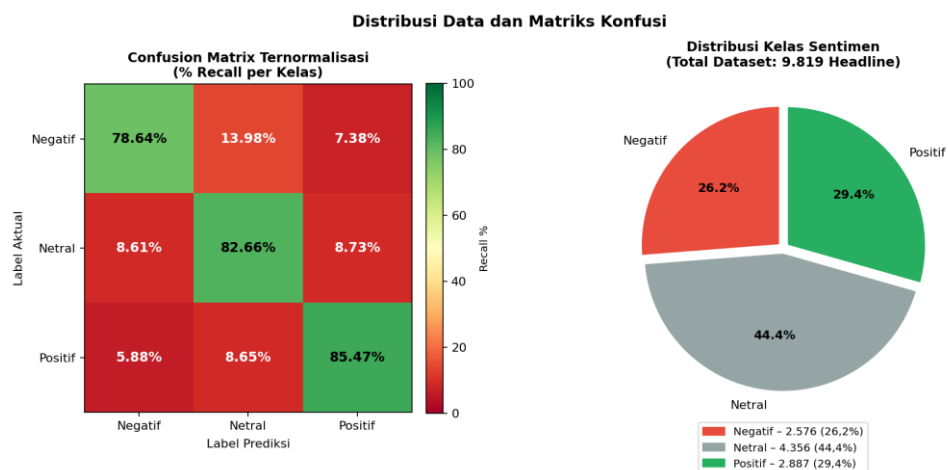
3.4 Analisis Confusion Matrix

Pola prediksi model dianalisis menggunakan *Confusion Matrix* untuk mengidentifikasi distribusi prediksi benar maupun kesalahan klasifikasi pada setiap kelas sentimen yang ditampilkan pada Tabel 5. Hasil pemetaan ini mengungkapkan pola kesalahan klasifikasi (*error pattern*) antar-kelas sentimen secara rinci. Secara keseluruhan, model berhasil memprediksi dengan benar sebanyak 1.619 sampel dari total 1.964 sampel data uji yang digunakan. Hasil ini membuktikan keandalan model dalam mengenali karakteristik

tekstual pada berita. Namun, terdapat pula kesalahan prediksi tertentu yang sebagian besar disebabkan oleh adanya kesamaan kata atau frasa (*lexical overlap*) yang muncul secara berulang dalam korpus berita finansial.

Tabel 6. *Confusion Matrix* Model pada Data Uji

Aktual \ Prediksi	Pred. Negatif	Pred. Netral	Pred. Positif
Aktual Negatif (515)	405 (78,6%)	72 (14,0%)	38 (7,4%)
Aktual Netral (871)	75 (8,6%)	720 (82,7%)	76 (8,7%)
Aktual Positif (578)	34 (5,9%)	50 (8,6%)	494 (85,5%)



Gambar 3. *Confusion Matrix* Ternormalisasi

Dari total 515 headline ber-sentimen Negatif pada data pengujian, sebanyak 405 data berhasil diklasifikasikan dengan benar dengan Recall sebesar 78,6%, sedangkan 72 data salah diprediksi sebagai Netral dan 38 data sebagai Positif. Pada kelas Netral yang terdiri atas 871 sampel, sebanyak 720 sampel berhasil dikenali dengan tepat dengan Recall sebesar 82,7%. Sementara itu, 75 sampel keliru diklasifikasikan sebagai Negatif dan 76 sampel sebagai Positif. Adapun kelas Positif yang berjumlah 578 sampel memperoleh Recall tertinggi, yaitu sebesar 85,5%, dengan 494 prediksi yang sesuai. Kesalahan klasifikasi pada kelas ini terdiri atas 34 sampel yang diprediksi sebagai Negatif dan 50 sampel sebagai Netral.

3.5 Analisis Koefisien Model

Koefisien yang dihasilkan oleh model Logistic Regression dimanfaatkan untuk mengidentifikasi term yang memberikan kontribusi paling besar terhadap masing-masing kelas sentimen.

Tabel 7. Top-5 Term Paling Diskriminatif untuk Kelas Negatif dan Positif
(berdasarkan koefisien model)

No	Term Negatif	Koefisien	Term Positif	Koefisien
1	anjlok	4,248	borong	3,663
2	ambruk	3,538	turun ke	3,420
3	ihsg*	2,975	bri	3,097
4	loyo	2,660	menguat	3,027
5	melemah	2,632	melesat	2,984

Catatan: IHSG pada kelas Negatif mencerminkan konteks berita penurunan indeks saham sehingga term tersebut memiliki koefisien tinggi pada kategori sentimen negatif.

Pada kelas Negatif, term dengan koefisien positif tertinggi adalah anjlok (4,248), ambruk (3,538), IHSG (2,975), loyo (2,660), dan melemah (2,632). Term-term tersebut merepresentasikan penurunan kinerja pasar dan kondisi keuangan yang tidak stabil. Sebaliknya, borong (-2,547) dan frasa turun ke (-2,454) memiliki koefisien negatif sehingga menurunkan peluang prediksi pada kelas Negatif.

Pada kelas Positif, term yang paling berkontribusi adalah borong (3,663), turun ke (3,420), BRI (3,097), menguat (3,027), dan melesat (2,984). Perbedaan arah koefisien ini menunjukkan bahwa polaritas sentimen bergantung pada konteks. Dalam berita pasar finansial, borong umumnya mencerminkan optimisme investor, sedangkan frasa turun ke lebih sering merujuk pada penurunan inflasi, suku bunga, atau nilai Dolar AS yang dipersepsikan positif bagi pasar. Fenomena ini menunjukkan adanya *semantic inversion* pada berita keuangan, yaitu polaritas sentimen ditentukan oleh subjek kalimat, bukan oleh kata kerja yang digunakan [14]. Kondisi tersebut juga menunjukkan keterbatasan TF-IDF dalam merepresentasikan hubungan sintaktis antar kata [10], [14]. Pada kelas Netral, term IHSG (-4,480) dan saham (-2,705) memiliki koefisien negatif sehingga headline yang secara eksplisit membahas keduanya cenderung tidak diklasifikasikan sebagai netral. Sebaliknya, kisah (2,510), sial (2,463), dan prabowo (2,445) memiliki koefisien positif karena lebih sering muncul pada berita yang bersifat informatif atau deskriptif. Perlu diperhatikan bahwa koefisien negatif tidak selalu menunjukkan makna sentimen negatif. Dalam Multinomial Logistic Regression, koefisien hanya menunjukkan arah pengaruh suatu term terhadap peluang dokumen masuk ke kelas tertentu. Koefisien positif meningkatkan peluang klasifikasi pada kelas tersebut, sedangkan koefisien negatif menurunkannya. Dengan demikian, koefisien mencerminkan pengaruh term terhadap kelas yang dianalisis, bukan makna sentimennya secara langsung.

3.6 Analisis Kesalahan (*Error Analysis*)

Pemeriksaan terhadap sampel yang salah diklasifikasikan menunjukkan bahwa kesalahan prediksi terutama disebabkan oleh tiga faktor, yaitu ambiguitas konteks, penggunaan *named entities*, dan headline multiklausal. Ambiguitas konteks terjadi ketika suatu kata memiliki kecenderungan sentimen tertentu, tetapi dalam kalimat hanya berfungsi sebagai deskripsi. Sebagai contoh, headline "Saham Bank X Naik Tipis di Tengah Sentimen Negatif Pasar Global" memiliki label aktual netral, tetapi diprediksi sebagai positif karena term naik memiliki bobot TF-IDF yang kuat terhadap kelas positif, sementara model belum mampu mempertimbangkan konteks keseluruhan kalimat. Sementara itu, *named entities* seperti nama perusahaan atau indeks dapat muncul pada berbagai kelas sentimen sehingga menyulitkan model dalam menentukan polaritas [6]. Selain itu, headline multiklausal yang memuat lebih dari satu kecenderungan sentimen juga meningkatkan potensi kesalahan klasifikasi. Temuan ini menunjukkan bahwa representasi TF-IDF masih memiliki keterbatasan dalam memahami hubungan semantik dan konteks kalimat. Oleh karena itu, metode yang mampu merepresentasikan konteks secara lebih utuh, seperti *contextual embedding* atau arsitektur transformer, dapat menjadi alternatif untuk mengatasi keterbatasan tersebut.

4. DISKUSI

Berdasarkan hasil evaluasi, model Multinomial Logistic Regression dengan representasi fitur TF-IDF mampu mengklasifikasikan headline berita pasar finansial Indonesia dengan akurasi 82,43%, sedangkan evaluasi menggunakan 5-fold cross-validation menghasilkan rata-rata akurasi 82,04%. Hasil tersebut menunjukkan bahwa kombinasi TF-IDF dan Multinomial Logistic Regression efektif untuk klasifikasi teks berbahasa Indonesia pada domain berita pasar finansial. Selain memberikan performa yang kompetitif, model juga memiliki tingkat interpretabilitas yang baik melalui analisis koefisien sehingga faktor-faktor yang memengaruhi hasil klasifikasi dapat dijelaskan secara lebih transparan.

Temuan ini sejalan dengan penelitian Kadir dan Fairuzabadi [10], Firizkiansah et al. [11], Mulyono dan Saprudin [12], serta Juniandika dan Mahendra [15] yang menunjukkan efektivitas kombinasi TF-IDF dan Logistic Regression dalam analisis sentimen berbahasa Indonesia. Berbeda dengan penelitian terdahulu yang umumnya menggunakan dataset non-finansial atau lebih menitikberatkan pada evaluasi performa klasifikasi [10], [11], penelitian ini menggunakan dataset headline berita pasar finansial Indonesia yang telah dianotasi secara manual serta melengkapi evaluasi melalui analisis koefisien dan *error analysis*. Pendekatan tersebut memungkinkan penelitian ini tidak hanya mengevaluasi performa model, tetapi juga mengidentifikasi kontribusi setiap term terhadap masing-masing kelas sentimen serta menjelaskan faktor-faktor utama penyebab kesalahan klasifikasi, yaitu ambiguitas konteks, *named entity*, dan headline multiklausal.

5. KESIMPULAN

Penelitian ini mengevaluasi penerapan Multinomial Logistic Regression berbasis TF-IDF untuk mengklasifikasikan 9.819 headline berita pasar finansial Indonesia dari CNBC Indonesia periode Januari 2024 hingga Maret 2025. Model memperoleh akurasi sebesar 82,43%, dengan Macro Precision 0,8200,

Macro Recall 0,8233, dan Macro F1-Score 0,8200. Evaluasi menggunakan 5-fold cross-validation menghasilkan rata-rata akurasi 82,04%, yang menunjukkan performa model yang stabil dan memiliki kemampuan generalisasi yang baik. Analisis koefisien menunjukkan bahwa model memiliki tingkat interpretabilitas yang baik melalui identifikasi term yang berkontribusi pada setiap kelas sentimen. *Error analysis* menunjukkan bahwa kesalahan klasifikasi terutama dipengaruhi oleh ambiguitas konteks, *named entity*, dan headline multiklausal. Penelitian selanjutnya dapat mengeksplorasi pendekatan berbasis *contextual embedding* atau arsitektur *transformer* untuk meningkatkan pemahaman konteks pada analisis sentimen berita finansial.

DAFTAR PUSTAKA

- [1] R. Shanti, H. Siregar, N. Zulfainarni, and Tony, "Role of Digital Transformation on Digital Business Model Banks," *Sustainability*, vol. 15, no. 23, p. 16293, Nov. 2023, doi: 10.3390/su152316293.
- [2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2019. [Online]. Available: <https://arxiv.org/pdf/1810.04805.pdf>
- [3] S. Minaee, N. Kalchbrenner, E. Cambria, M. Nikola, and K. Chandra, "Deep Learning Based Text Classification: A Comprehensive Review," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, pp. 1–40, 2021, doi: 10.1145/3439726.
- [4] S. Chari, P. H. Desai, N. Borde, and B. George, "Aggregate News Sentiment and Stock Market Returns in India," *Journal of Risk and Financial Management*, vol. 16, no. 8, p. 376, Aug. 2023, doi: 10.3390/jrfm16080376.
- [5] G. Anese, M. Corazza, M. Costola, and L. Pelizzon, "Impact of public news sentiment on stock market index return and volatility," *Computational Management Science*, vol. 20, no. 1, p. 20, Dec. 2023, doi: 10.1007/s10287-023-00454-2.
- [6] A. Sinha, S. Kedas, R. Kumar, and P. Malo, "SEnFiN 1.0: Entity-Aware Sentiment Analysis for Financial News," May 2023, doi: 10.1002/asi.24634.
- [7] D. Araci, "FinBERT: Financial Sentiment Analysis with Pre-trained Language Models," Aug. 2019, [Online]. Available: <http://arxiv.org/abs/1908.10063>
- [8] A. A. Ningtyas, A. Solichin, and R. Pradana, "Analisis Sentimen Komentar Youtube Tentang Prediksi Resesi Ekonomi Tahun 2023 Menggunakan Algoritme Naïve Bayes," *Bit (Fakultas Teknologi Informasi Universitas Budi Luhur)*, vol. 20, no. 1, p. 9, Apr. 2023, doi: 10.36080/bit.v20i1.2317.
- [9] J. S. Hutagalung and Rasiban, "ANALISIS SENTIMEN KEUANGAN (DATA FIQA AND FINANCIAL PHRASEBANK) MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION DAN SUPPORT VECTOR MACHINE," *Jurnal Indonesia : Manajemen Informatika dan Komunikasi*, vol. 4, no. 3, pp. 1654–1669, Sep. 2023, doi: 10.35870/jimik.v4i3.404.
- [10] S. F. Kadir and A. Fairuzabadi, "Analisis Sentimen Ulasan Shopee di Google Play dengan TF-IDF dan Logistic Regression," *RIGGS*, 2025, doi: 10.31004/riggs.v4i2.2850.
- [11] A. Firizkiansah, A. Muhammad, and I. R. Maulana, "Optimasi Klasifikasi Data Teks Menggunakan Algoritma Logistic Regression dengan TF-IDF dan SMOTE," *JIKOMTI : Jurnal Ilmiah Ilmu Komputer dan Teknologi Informasi*, vol. 2, no. 1, pp. 29–36, May 2025, [Online]. Available: <https://ojs.sains.ac.id/index.php/Jikomti/article/view/97>
- [12] H. S. Mulyono and U. Saprudin, "Efektivitas Logistic Regression dalam Analisis Sentimen Berbahasa Indonesia pada Komentar YouTube tentang Isu Ketenagakerjaan," *JIMIK*, 2025, doi: 10.63447/jimik.v6i3.1481.
- [13] T. A. Jiwandono and M. H. Achmad, "Dataset Sentimen Berita Saham CNBC Indonesia Periode 2024-2025Q1," 000959744, Aug. 27, 2025.
- [14] S. Chen, X. Du, J. Zhao, H. Huang, and X. Chen, "A syntactic dependency method for aspect-level sentiment classification by deep learning," *Measurement and Control*, vol. 56, no. 5–6, pp. 1057–1065, May 2023, doi: 10.1177/00202940221090975.
- [15] I Made Juniandika and Ida Bagus Made Mahendra, "Analisis Sentimen Aplikasi Zenius Menggunakan Metode Logistic Regression," *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, vol. 1, no. 4, pp. 1171–1178, Aug. 2023, doi: 10.24843/JNATIA.2023.v01.i04.p20.