



Analisis Sentimen Kutipan Media Sosial Berbahasa Indonesia Menggunakan Convolutional Neural Network

M. Fajar Ramadhan*, Febriyanti Panjaitan, Winarnie, Hery Oktafiandi, Yohanes

Teknik Informatika, Fakultas Teknologi Kreatif, Satu University, Palembang

Abstrak: Media sosial menjadi salah satu ruang utama ekspresi opini dan emosi masyarakat Indonesia. Berbagai pandangan, keluhan, dukungan, maupun refleksi pribadi sering dituangkan dalam bentuk teks pendek seperti caption dan kutipan (quotes). Teks-teks ini tidak hanya memuat informasi, tetapi juga sentimen yang dapat menggambarkan sikap pengguna terhadap suatu isu. Penelitian ini bertujuan untuk mengembangkan model analisis sentimen kutipan media sosial berbahasa Indonesia menggunakan Convolutional Neural Network (CNN) dan membandingkannya dengan pendekatan klasifikasi klasik berbasis Term Frequency–Inverse Document Frequency (TF-IDF) dan Logistic Regression. Data yang digunakan berasal dari dataset terbuka opini publik di media sosial terkait pasar saham Indonesia. Dari dataset tersebut dipilih secara seimbang 1.200 kutipan berbahasa Indonesia, masing-masing 400 berlabel negatif, 400 netral, dan 400 positif. Tahapan penelitian meliputi pembersihan teks, normalisasi, tokenisasi, pembagian data latih dan uji (80:20), pembentukan vektor TF-IDF sebagai baseline, pembentukan urutan indeks kata untuk CNN, perancangan arsitektur text-CNN dengan tiga kernel konvolusi paralel, serta evaluasi model menggunakan akurasi, precision, recall, F1-score, dan confusion matrix. Model baseline TF-IDF + Logistic Regression menghasilkan akurasi 66% dengan macro F1-score 0,66 pada data uji. CNN varian awal (CNN V1) memberikan performa sebanding dengan akurasi 65% dan macro F1-score 0,65, namun menunjukkan gejala overfitting. Setelah kapasitas model dikurangi dan regularisasi diperkuat (CNN V2), kinerja meningkat menjadi akurasi 67,5% dan macro F1-score 0,67, sedikit melampaui baseline klasik. Hasil ini mengindikasikan bahwa CNN mampu menangkap pola n-gram lokal pada kutipan secara efektif, meskipun keunggulannya terhadap pendekatan klasik masih terbatas pada skala data yang relatif kecil.

Kata Kunci: Analisis Sentimen, Kutipan Media Sosial, Convolutional Neural Network, TF-IDF, Google Colab

DOI:

<https://doi.org/10.53697/jkomitek.v6i1.3627>

*Correspondence: M. Fajar Ramadhan

Email: m.ramadhan@univ.satu.ac.id

Received: 08-04-2026

Accepted: 08-05-2026

Published: 08-06-2026



Copyright: © 2025 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: Social media has become a primary platform for Indonesians to express their opinions and emotions. Various views, complaints, support, and personal reflections are often expressed in short texts such as captions and quotes. These texts not only contain information but also sentiments that can describe users' attitudes towards an issue. This study aims to develop a sentiment analysis model for Indonesian-language social media quotes using a Convolutional Neural Network (CNN) and compare it with classical classification approaches based on Term Frequency–Inverse Document Frequency (TF-IDF) and Logistic Regression. The data used comes from an open dataset of public opinion on social media related to the Indonesian stock market. From this dataset, 1,200 Indonesian-language quotes were selected in a balanced manner, with 400 labeled negative, 400 neutral, and 400 positive. The research phase included text cleaning, normalization, tokenization, splitting the training and test data (80:20), generating TF-IDF vectors as a baseline, generating word index sequences for the CNN, designing a text-CNN architecture with three parallel convolution kernels, and evaluating the model using accuracy, precision, recall, F1-score, and confusion matrix. The baseline TF-IDF + Logistic Regression model achieved 66% accuracy with a macro F1-score of 0.66 on the test data. The initial CNN variant (CNN V1) provided comparable performance with 65% accuracy and a macro F1-score of 0.65, but exhibited signs of overfitting. After reducing the model capacity and strengthening regularization (CNN V2), performance improved to 67.5% accuracy and a macro F1-score of 0.67, slightly surpassing the classical baseline. These results indicate that CNNs are capable of effectively capturing local n-gram patterns in quotations, although their superiority over classical approaches is still limited to relatively small data scales.

Keywords: Sentiment Analysis, Social Media Quotes, Convolutional Neural Network, TF-IDF, Google Colab

Pendahuluan

Perkembangan teknologi informasi dan internet mendorong media sosial menjadi salah satu sarana utama bagi masyarakat untuk menyampaikan pendapat, berdiskusi, dan mengekspresikan emosi. Platform seperti X (Twitter), Instagram, dan TikTok dipenuhi oleh teks pendek dalam bentuk caption, komentar, dan kutipan (quotes). Teks-teks ini tidak hanya memuat informasi, tetapi juga sentimen yang dapat menggambarkan sikap pengguna terhadap suatu isu, misalnya terkait layanan digital, kebijakan publik, maupun kondisi pasar saham (Nip & Berthelie, 2024).

Kutipan seperti “capek jadi kuat terus”, “bahagia itu sederhana, asal bukan sama kamu”, atau “portofolio hijau, rezeki pagi ini alhamdulillah” merupakan contoh teks pendek yang mengandung muatan emosional kuat, meskipun tidak selalu menyebutkan kata “senang” atau “sedih” secara eksplisit. Kondisi ini membuat analisis sentimen terhadap kutipan menjadi menarik sekaligus menantang, karena model tidak cukup hanya mengenali kata kunci tertentu, tetapi perlu memahami kombinasi kata dan konteks kalimat secara menyeluruh (Li et al., 2020).

Analisis sentimen (sentiment analysis) atau opinion mining merupakan cabang dari pemrosesan bahasa alami (Natural Language Processing/NLP) yang bertujuan mengidentifikasi polaritas sentimen dalam teks, misalnya negatif, netral, atau positif (Li et al., 2020). Sejumlah tinjauan pustaka sistematis mencatat bahwa analisis sentimen di media sosial umumnya memanfaatkan algoritma klasifikasi klasik seperti Naive Bayes, Support Vector Machine (SVM), dan Logistic Regression dengan representasi teks berbasis n-gram atau TF-IDF (Singgalen, 2021; Li et al., 2020). Pada konteks Bahasa Indonesia, pendekatan tersebut telah banyak digunakan untuk menganalisis opini publik di berbagai domain, mulai dari ulasan aplikasi, layanan transportasi daring, hingga komentar di media sosial (Wahyudi & Kusumawardhana, 2021).

Penelitian terdahulu oleh penulis menunjukkan bahwa kombinasi TF-IDF dan algoritma klasifikasi klasik mampu menghasilkan performa yang baik pada teks pendek berbahasa Indonesia, khususnya judul berita. Dengan melakukan penyesuaian parameter TF-IDF dan augmentasi data, sistem klasifikasi topik dan sentimen judul berita mampu mencapai akurasi yang tinggi (Ramadhan, 2023). Hasil ini menegaskan bahwa representasi fitur dan pemilihan algoritma yang tepat sangat berpengaruh terhadap kualitas model analisis sentimen (Li et al., 2020). Namun, judul berita cenderung lebih terstruktur dan informatif dibanding kutipan media sosial yang lebih bebas, informal, dan sering mengandung singkatan (Nip & Berthelie, 2024).

Seiring berkembangnya deep learning, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), dan model hybrid mulai banyak diterapkan untuk analisis sentimen teks (Li et al., 2020). CNN bekerja dengan mengekstraksi pola lokal n-gram melalui operasi konvolusi di atas embedding kata, sehingga mampu mengenali kombinasi kata penting tanpa perlu merancang fitur secara manual (Gao et al., 2022). Berbagai penelitian menunjukkan bahwa model berbasis CNN dan variannya mampu memberikan peningkatan performa dibanding metode klasik untuk teks pendek di berbagai bahasa, termasuk Bahasa Indonesia (Juwiantho et al., 2020; Lin & Nuha, 2023).

Dalam konteks pasar saham Indonesia, Hartanto et al. (2025a) mengembangkan ID-SMSA (Indonesian Stock Market Dataset for Sentiment Analysis), yaitu dataset berbahasa Indonesia yang berisi 3.288 tweet terkait sepuluh emiten dengan kapitalisasi terbesar, masing-masing diberi label positif, netral, atau negatif oleh dua anotator dengan panduan psikologi klinis. Dataset ini dirilis secara terbuka dalam bentuk artikel data (Hartanto et al., 2025a), dataset terkurasi di Mendeley Data (Hartanto et al., 2025b), dan repositori kode/dokumentasi pendukung di GitHub (Hartanto et al., 2025c). Ketersediaan ID-SMSA membuka peluang bagi peneliti untuk mengkaji opini publik terhadap pasar saham secara lebih sistematis dan terukur (Hartanto et al., 2025a).

Berdasarkan latar belakang tersebut, penelitian ini diarahkan untuk: (1) menyusun dataset kutipan media sosial berbahasa Indonesia yang seimbang untuk tiga kelas sentimen dengan memanfaatkan ID-SMSA sebagai sumber utama (Hartanto et al., 2025a); (2) mengembangkan model CNN untuk klasifikasi sentimen kutipan tersebut (Gao et al., 2022); (3) membandingkan performa CNN dengan baseline TF-IDF + Logistic Regression (Ramadhan, 2023); serta (4) mengkaji kelebihan dan keterbatasan CNN pada dataset kutipan berukuran menengah (Li et al., 2020).

Tinjauan Pustaka

Analisis Sentimen dan Deep Learning

Berbagai studi telah mengkaji perkembangan analisis sentimen dari pendekatan klasik hingga deep learning. Li et al. (2020) merangkum bahwa awalnya penelitian banyak memanfaatkan fitur berbasis bag-of-words atau TF-IDF yang diproses oleh algoritma seperti Naive Bayes, SVM, dan Logistic Regression. Seiring meningkatnya ketersediaan data dan daya komputasi, pendekatan berbasis word embedding dan jaringan saraf dalam (deep neural networks) mulai mendominasi, terutama CNN, LSTM, dan model hybrid (Li et al., 2020).

Singgalen (2021) menyusun systematic literature review yang memetakan metode dan algoritma dalam analisis sentimen di media sosial. Hasil kajian tersebut menunjukkan bahwa metode klasik masih banyak digunakan, tetapi tren riset mulai bergeser menuju pemanfaatan neural network dan deep learning (Singgalen, 2021). Nip dan Berthelie (2024) meninjau lebih luas konsep social media sentiment analysis, mencakup jenis data (teks, gambar, video), tantangan noise dan bias, serta beragam aplikasi di bidang sosial, politik, dan ekonomi.

Pada Bahasa Indonesia, Juwiantho et al. (2020) melakukan analisis sentimen pada Twitter menggunakan Word2Vec dan deep CNN. Mereka menunjukkan bahwa kombinasi embedding pra-latih dan CNN mampu memberikan performa yang baik pada teks pendek, dengan tahapan preprocessing dan evaluasi yang mirip dengan penelitian ini (Juwiantho et al., 2020). Lin dan Nuha (2023) mengusulkan strategi hybrid deep learning berbasis BERT, LSTM, dan CNN untuk beberapa dataset sentimen Bahasa Indonesia, dan melaporkan peningkatan kinerja dibanding model tunggal.

CNN dan Model Hybrid untuk Analisis Sentimen

Penggunaan CNN untuk analisis sentimen teks telah dikembangkan dalam berbagai bentuk. Gao et al. (2022) mengusulkan model aspect-based sentiment analysis (ABSA) berbasis CNN + BiGRU untuk short text dan menunjukkan bahwa kombinasi konvolusi dan memori jangka panjang mampu menangkap aspek dan polaritas secara lebih akurat. Rajesh dan Hiwarkar (2023) mengembangkan model multi-channel deep learning yang menggabungkan CNN dan BiLSTM untuk meningkatkan akurasi analisis sentimen teks. Yan et al. (2021) menggunakan model CNN-BiGRU-AT dengan mekanisme attention untuk menganalisis sentimen teks siswa, sementara Ghorbani et al. (2020) memperkenalkan arsitektur ConvLSTMConv untuk analisis sentimen di domain cloud computing.

Berbagai penelitian tersebut memperlihatkan bahwa CNN dapat dijadikan blok bangunan utama, baik sebagai model mandiri maupun sebagai bagian dari arsitektur hybrid (Li et al., 2020). Penelitian ini memilih CNN murni sebagai model utama, dengan mempertimbangkan ukuran dataset yang relatif tidak terlalu besar, serta kebutuhan menjaga kompleksitas model agar tetap proporsional (Gao et al., 2022).

Analisis Sentimen pada Media Sosial dan Ulasan Aplikasi

Pada ranah media sosial dan ulasan aplikasi, banyak penelitian yang memanfaatkan Bahasa Indonesia sebagai objek kajian. Alfin et al. (2023) menggunakan CNN untuk menganalisis klasifikasi sentimen berbasis topik pada ulasan layanan dompet digital Dana dan Sakuku. Zulfiqri et al. (2024) mengkaji sentimen ulasan pengguna aplikasi Instagram di Google Play Store menggunakan Naive Bayes classifier. Praneswara dan Cahyono (2023) menganalisis sentimen ulasan aplikasi TikTok Shop Seller Center dengan algoritma Naive Bayes, sementara Wijaya et al. (2024) membandingkan kinerja Naive Bayes dan K-NN pada domain yang sama. Wahyudi dan Kusumawardhana (2021) menerapkan SVM untuk analisis sentimen ulasan aplikasi Grab di Google Play Store.

Secara umum, hasil penelitian tersebut menunjukkan bahwa ulasan pengguna aplikasi maupun komentar di media sosial merupakan sumber data yang kaya untuk analisis sentimen, dengan metode klasik yang masih mampu memberikan performa kompetitif (Singgalen, 2021). Penelitian-penelitian tersebut memperkuat argumen bahwa teks pendek di media sosial dan ulasan aplikasi—yang gaya bahasanya mirip dengan kutipan (quotes)—dapat diolah menjadi dataset analisis sentimen yang bermanfaat (Nip & Berthelie, 2024). Penelitian ini memanfaatkan ID-SMSA (Hartanto et al., 2025a; Hartanto et al., 2025b; Hartanto et al., 2025c) untuk mengekstraksi kutipan opini publik terkait pasar saham, sehingga berada dalam jalur yang sama namun dengan fokus pada domain finansial.

IndoBERT, BERT, dan Aspect-Based Sentiment Analysis

Perkembangan model pra-latih berbasis Transformer juga membawa dampak besar pada analisis sentimen Bahasa Indonesia. Wilie et al. (2020) memperkenalkan IndoNLU sebagai benchmark dan sumber daya untuk evaluasi pemahaman bahasa alami Indonesia, sekaligus memperkenalkan berbagai varian IndoBERT. Nugroho et al. (2021) menguji fine-

tuning BERT dan IndoBERT untuk analisis sentimen ulasan aplikasi mobile berbahasa Indonesia, dan menunjukkan bahwa model berbasis Transformer mampu mengungguli metode klasik.

Mandhasiya et al. (2023) mengusulkan hybrid BERT + deep learning untuk analisis sentimen Bahasa Indonesia dan mendapatkan peningkatan performa di berbagai dataset. Kusumaningrum et al. (2025) mengoptimalkan aspect-based sentiment analysis (ABSA) dengan memanfaatkan BERT pada feedback mahasiswa di Indonesia, sedangkan Hoang et al. (2019) menunjukkan bahwa BERT dapat digunakan secara efektif untuk ABSA pada bahasa lain.

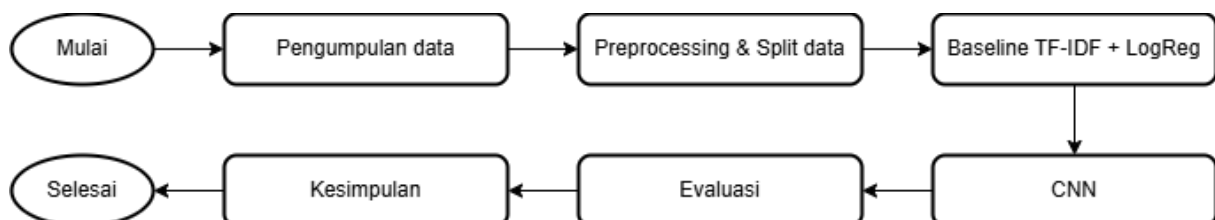
Berbagai penelitian tersebut menunjukkan arah perkembangan riset ke depan, yaitu pemanfaatan embedding kontekstual dan model Transformer untuk analisis sentimen yang lebih kaya konteks, baik pada level kalimat maupun level aspek (Wilie et al., 2020). Dalam penelitian ini, CNN diposisikan sebagai langkah antara: lebih kuat daripada model klasik murni, namun masih lebih sederhana dan ringan dibanding model Transformer (Li et al., 2020). Hasil yang diperoleh dapat menjadi landasan awal sebelum beralih ke pendekatan berbasis IndoBERT atau model Transformer lainnya pada penelitian lanjutan.

Metodologi

Secara umum, penelitian ini menggunakan pendekatan kuantitatif eksperimental dengan metode supervised text classification. Eksperimen dilakukan di lingkungan Google Colab menggunakan Python dan beberapa pustaka pendukung seperti scikit-learn dan TensorFlow/Keras.

Secara garis besar, tahapan penelitian meliputi:

1. Pengumpulan dan pembentukan dataset kutipan.
2. Pra-pemrosesan (preprocessing) teks.
3. Pembagian data menjadi data latih dan data uji.
4. Pembentukan representasi TF-IDF untuk baseline dan urutan indeks kata untuk CNN.
5. Pelatihan model baseline TF-IDF + Logistic Regression.
6. Perancangan dan pelatihan model CNN.
7. Evaluasi dan analisis hasil.



Gambar 1. Flowchart Alur Penelitian

Pengumpulan dan Pembentukan Dataset

Data yang digunakan dalam penelitian ini diambil dari dataset terbuka opini publik di media sosial X/Twitter terkait pasar saham Indonesia. Dataset asli berisi lebih dari tiga ribu tweet berbahasa Indonesia yang telah diberi label sentimen Positive, Neutral, dan Negative oleh peneliti sebelumnya.

Pada penelitian ini, penulis hanya menggunakan kolom teks dan label sentimen. Label bahasa Inggris tersebut kemudian disesuaikan menjadi label Bahasa Indonesia, yaitu: positif, netral, dan negatif. Untuk menjaga keseimbangan kelas, diambil masing-masing 400 kutipan dari setiap kelas secara acak tanpa pengulangan (random sampling). Dengan demikian, total terdapat 1.200 kutipan yang digunakan sebagai dataset akhir, dengan distribusi:

- 400 kutipan negatif,
- 400 kutipan netral,
- 400 kutipan positif.

Dataset disimpan dalam berkas quotes_dataset.csv dengan dua kolom utama: text dan label.

Tabel 1. Contoh Data Kutipan

No	Teks Kutipan (Tweet)	Label
1	IHSG makin ga jelas arah, saham gorengan terbang lagi	negatif
2	Portofolio hijau, rezeki pagi ini alhamdulillah	positif
3	Pasar sepi, menunggu keputusan The Fed malam ini	netral

Preprocessing Data

Pra-pemrosesan teks dilakukan untuk mengurangi noise dan menyeragamkan format teks. Langkah-langkah preprocessing yang diterapkan adalah sebagai berikut:

1. Case folding, yaitu mengubah seluruh huruf menjadi huruf kecil.
2. Pembersihan karakter khusus, seperti penghapusan URL, mention (@username), hashtag (#tag), angka, dan beberapa tanda baca yang tidak diperlukan menggunakan regular expression.
3. Normalisasi spasi, yaitu menghapus spasi ganda dan spasi di awal atau akhir teks.
4. Stopword removal, yaitu menghapus kata-kata umum (seperti "dan", "yang", "di") agar model lebih fokus pada kata yang lebih informatif.

Hasil preprocessing disimpan pada kolom baru clean_text.

Tabel 2. Contoh Hasil Preprocessing

No	Teks Asli	clean_text	Label
1	IHSG makin ga jelas arah, saham gorengan terbang lagi	ihsg makin ga jelas arah saham gorengan terbang lagi	negatif
2	Portofolio hijau, rezeki pagi ini alhamdulillah	portofolio hijau rezeki pagi ini alhamdulillah	positif
3	Pasar sepi, menunggu keputusan The Fed malam ini	pasar sepi menunggu keputusan the fed malam ini	netral

Pembagian Data dan Encoding Label

Setelah preprocessing, data dibagi menjadi data latih dan data uji menggunakan fungsi train_test_split dengan rasio **80:20** dan parameter stratify berdasarkan label agar distribusi kelas tetap seimbang. Hasil pembagian data adalah:

- 960 kutipan sebagai data latih,
- 240 kutipan sebagai data uji.

Distribusi per kelas setelah pembagian data ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Data Latih dan Uji

Kelas	Data Latih	Data Uji	Total
Negatif	320	80	400
Netral	320	80	400
Positif	320	80	400
Total	960	240	1.200

Label teks (negatif, netral, positif) dikonversi menjadi label numerik menggunakan LabelEncoder, misalnya:

- negatif → 0
- netral → 1
- positif → 2

Untuk keperluan CNN, label numerik ini selanjutnya diubah menjadi representasi one-hot.

Representasi Teks

Penelitian ini menggunakan dua jenis representasi teks:

1. TF-IDF

Representasi TF-IDF digunakan untuk model baseline Logistic Regression. Teks `clean_text` diubah menjadi vektor TF-IDF menggunakan `TfidfVectorizer` dengan pengaturan:

- `ngram_range = (1,2)` (unigram dan bigram),
- `max_features = 5000`.

- #### 2. Urutan indeks kata (sequences) untuk CNN
- Untuk CNN, digunakan `Tokenizer` dari Keras dengan jumlah kata maksimum (`num_words`) sebesar 20.000 dan token OOV (`oov_token`). Teks `clean_text` diubah menjadi urutan indeks kata, kemudian dilakukan padding sehingga semua urutan memiliki panjang tetap `max_len = 40` token.

Model Baseline TF-IDF + Logistic Regression

Model baseline dibangun menggunakan kombinasi TF-IDF dan Logistic Regression. Langkah-langkahnya:

1. Mengubah data latih dan data uji menjadi vektor TF-IDF.
2. Melatih model `LogisticRegression` dengan parameter `max_iter = 1000` menggunakan data latih TF-IDF.
3. Menggunakan model tersebut untuk memprediksi label data uji.
4. Menghitung akurasi, precision, recall, F1-score, dan confusion matrix pada data uji.

Model ini dijadikan acuan (benchmark) untuk menilai apakah penggunaan CNN memberikan peningkatan performa yang signifikan.

Perancangan Model CNN

Model utama yang digunakan adalah Convolutional Neural Network (CNN) dengan pendekatan text-CNN. Dua varian arsitektur CNN yang diuji adalah:

1. CNN V1

- Embedding dimension: 128
- Tiga `Conv1D` paralel dengan kernel size 3, 4, dan 5, masing-masing 128 filter.

- o Masing-masing Conv1D diikuti oleh Global Max Pooling.
- o Hasil pooling digabungkan (concatenate), dilanjutkan dengan Dropout 0,5 dan Dense 128 neuron (aktivasi ReLU).
- o Lapisan output berupa Dense 3 neuron dengan aktivasi softmax.

Tabel 4. Arsitektur model CNN Quotes Sentiment

Layer (type)	Output Shape	Param #	Connected to
input_text (InputLayer)	(None, 40)	0	-
embedding (Embedding)	(None, 40, 128)	436,352	input_text[0][0]
conv_3 (Conv1D)	(None, 38, 128)	49,280	embedding[0][0]
conv_4 (Conv1D)	(None, 37, 128)	65,664	embedding[0][0]
conv_5 (Conv1D)	(None, 36, 128)	82,048	embedding[0][0]
pool_3 (GlobalMaxPooling1D)	(None, 128)	0	conv_3[0][0]
pool_4 (GlobalMaxPooling1D)	(None, 128)	0	conv_4[0][0]
pool_5 (GlobalMaxPooling1D)	(None, 128)	0	conv_5[0][0]
concat (Concatenate)	(None, 384)	0	pool_3[0][0], pool_4[0][0], pool_5[0][0]
dropout (Dropout)	(None, 384)	0	concat[0][0]
dense_1 (Dense)	(None, 128)	49,280	dropout[0][0]
output (Dense)	(None, 3)	387	dense_1[0][0]

2. CNN V2 (Model Final)

- o Embedding dimension: 64
- o Tiga Conv1D paralel dengan kernel size 3, 4, dan 5, masing-masing 64 filter, dilengkapi regularisasi L2.
- o Masing-masing Conv1D diikuti Global Max Pooling.
- o Hasil pooling digabungkan, kemudian diberikan Dropout 0,6.
- o Lapisan Dense 64 neuron (ReLU) dengan regularisasi L2.
- o Lapisan output Dense 3 neuron (softmax).
- o Learning rate optimizer Adam diturunkan menjadi 5×10^{-4} dan validation_split diperbesar menjadi 0,2.

Tabel 5. Arsitektur model CNN V2 untuk analisis sentimen kutipan media sosial

Layer (type)	Output Shape	Param #	Connected to
input_text (InputLayer)	(None, 40)	0	-
embedding (Embedding)	(None, 40, 64)	218,176	input_text[0][0]
conv_3 (Conv1D)	(None, 38, 64)	12,352	embedding[0][0]
conv_4 (Conv1D)	(None, 37, 64)	16,448	embedding[0][0]
conv_5 (Conv1D)	(None, 36, 64)	20,544	embedding[0][0]
pool_3 (GlobalMaxPooling1D)	(None, 64)	0	conv_3[0][0]
pool_4 (GlobalMaxPooling1D)	(None, 64)	0	conv_4[0][0]
pool_5 (GlobalMaxPooling1D)	(None, 64)	0	conv_5[0][0]
concat (Concatenate)	(None, 192)	0	pool_3[0][0], pool_4[0][0], pool_5[0][0]
dropout (Dropout)	(None, 192)	0	concat[0][0]
dense_1 (Dense)	(None, 64)	12,352	dropout[0][0]
output (Dense)	(None, 3)	195	dense_1[0][0]

Model dikompilasi dengan fungsi loss `categorical_crossentropy` dan optimizer Adam. Pelatihan dilakukan dengan `batch_size = 32`, maksimum 20 epoch, dan callback `EarlyStopping` berdasarkan `val_loss` untuk mencegah overfitting.

Evaluasi Model

Evaluasi model dilakukan pada data uji yang berjumlah 240 kutipan. Metrik yang digunakan adalah:

- Akurasi,
- Precision, recall, dan F1-score per kelas,
- Macro-average F1-score,
- Confusion matrix.

Confusion matrix untuk masing-masing model divisualisasikan dalam bentuk heatmap dan akan ditempatkan sebagai:

- Gambar 3: Confusion Matrix model TF-IDF + Logistic Regression,
- Gambar 4: Confusion Matrix model CNN V2.

Hasil dan Pembahasan

Hasil Deskripsi Data

Dataset yang digunakan merupakan subset dari **ID-SMSA (Indonesian Stock Market Dataset for Sentiment Analysis)** yang berisi 3.288 tweet berbahasa Indonesia mengenai pasar saham, masing-masing berlabel positif, netral, atau negatif. Dataset ini dikembangkan oleh Hartanto dkk. dan tersedia secara terbuka melalui Mendeley Data serta repositori GitHub.

Pada penelitian ini, penulis hanya menggunakan 1.200 tweet versi Bahasa Indonesia yang kemudian dipandang sebagai kutipan (quotes) opini publik di media sosial. Data tersebut dipilih secara acak namun seimbang, masing-masing 400 negatif, 400 netral, dan 400 positif. Setelah pembagian data dengan rasio 80:20 dan stratified sampling, diperoleh 960 kutipan sebagai data latih dan 240 kutipan sebagai data uji, dengan distribusi setiap kelas yang tetap seimbang sebagaimana tercantum pada Tabel 3.

Untuk menggambarkan komposisi dataset secara visual, kakak bisa menambahkan grafik batang distribusi sentimen (opsional) sebagai Gambar 3a atau langsung melanjutkan ke confusion matrix baseline pada Gambar 3.

Hasil Evaluasi CNN

Model CNN V1 menggunakan embedding berdimensi 128 dan tiga lapisan Conv1D paralel dengan kernel 3, 4, dan 5 (masing-masing 128 filter). Selama proses pelatihan, akurasi data latih meningkat cepat mendekati 100%, sementara akurasi validasi cenderung stagnan di sekitar 70%, menunjukkan adanya indikasi overfitting.

Pada data uji, CNN V1 menghasilkan:

- Akurasi: 66%
- Macro F1-score: 0,66

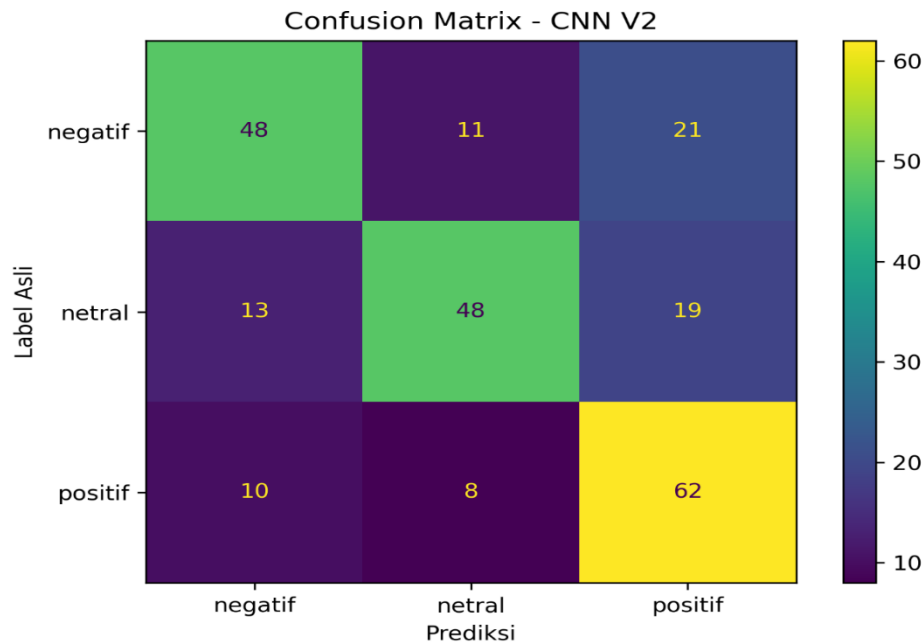
Ringkasan hasil per kelas ditunjukkan pada Tabel 6.

Tabel 6. Hasil Evaluasi CNN

Kelas	Precision	Recall	F1-score	Support
Negatif	0,68	0,60	0,64	80
Netral	0,72	0,60	0,65	80
Positif	0,61	0,78	0,68	80
Macro	0,67	0,66	0,66	240

rata-rata

Secara keseluruhan, baseline menghasilkan akurasi 66% dan macro F1-score 0,66. Kelas negatif memiliki recall tertinggi (0,78), artinya sebagian besar kutipan yang benar-benar negatif berhasil dikenali. Di sisi lain, kelas netral memiliki recall yang lebih rendah (0,5), sehingga masih cukup sering tertukar dengan kelas lain.



Gambar 2. Confusion Matrix Logreg + TF-IDF

Gambar 2 diisi menggunakan file cm_logreg_tfidf.png yang sudah kakak unduh dari Google Colab. Matriks kebingungan yang mendasarinya adalah:

$$CM_{CNN} = \begin{bmatrix} 48 & 11 & 21 \\ 13 & 48 & 19 \\ 10 & 8 & 62 \end{bmatrix}$$

Dari matriks tersebut tampak bahwa:

- 48 dari 80 kutipan negatif terprediksi benar,
- 48 dari 80 kutipan netral terprediksi benar,
- 62 dari 80 kutipan positif terprediksi benar,

Hal ini menunjukkan kecenderungan model baseline untuk kutipan positif terprediksi benar.

Hasil Model Baseline TF-IDF + Logistic Regression

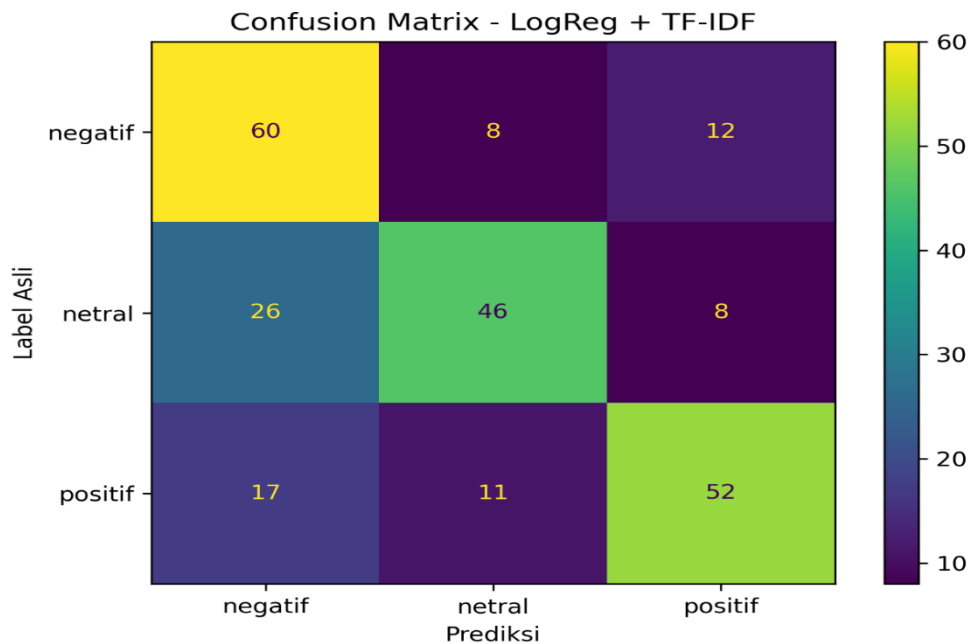
Model baseline menggunakan representasi TF-IDF dengan ngram_range = (1,2) dan max_features = 5000, kemudian diklasifikasikan menggunakan Logistic Regression (max_iter = 1000). Evaluasi dilakukan pada 240 data uji.

Hasil evaluasi per kelas dirangkum pada Tabel 7.

Tabel 7. Hasil Evaluasi TF-IDF + Logistic Regression

Kelas	Precision	Recall	F1-score	Support
Negatif	0,58	0,75	0,66	80
Netral	0,71	0,57	0,63	80
Positif	0,72	0,65	0,68	80
Macro	0,67	0,66	0,66	240
rata-rata				

Secara keseluruhan, baseline menghasilkan akurasi 66% dan macro F1-score 0,66. Kelas negatif memiliki recall tertinggi (0,75), artinya sebagian besar kutipan yang benar-benar negatif berhasil dikenali. Di sisi lain, kelas netral memiliki recall yang lebih rendah (0,57), sehingga masih cukup sering tertukar dengan kelas lain.



Gambar 3. Confusion Matrix Logreg + TF-IDF

Gambar 3 diisi menggunakan file `cm_logreg_tfidf.png` yang sudah kakak unduh dari Google Colab. Matriks kebingungan yang mendasarinya adalah:

$$CM_{LogReg} = \begin{bmatrix} 60 & 8 & 12 \\ 26 & 46 & 8 \\ 17 & 11 & 52 \end{bmatrix}$$

Baris menyatakan label aktual (negatif, netral, positif) dan kolom menyatakan label prediksi. Terlihat bahwa:

- 60 kutipan negatif terklasifikasi benar,
- 26 kutipan netral salah diprediksi sebagai negatif,
- 17 kutipan positif salah diprediksi sebagai negatif.

Hal ini menunjukkan kecenderungan model baseline untuk “bermain aman” dengan memprediksi negatif ketika informasi kurang jelas.

Diskusi

Berdasarkan keseluruhan hasil, beberapa poin penting dapat dicatat:

1. Baseline klasik tetap kuat. Kombinasi TF-IDF + Logistic Regression menghasilkan akurasi 66% dan macro F1-score 0,66. Hal ini sejalan dengan temuan pada penelitian sebelumnya bahwa model klasik dengan representasi TF-IDF masih sangat kompetitif untuk teks pendek berbahasa Indonesia.

2. CNN V1 belum optimal. CNN V1 dengan embedding dan jumlah filter yang besar menghasilkan akurasi 65% dan macro F1-score 0,65, sedikit di bawah baseline. Perbedaan besar antara akurasi latih dan validasi menunjukkan bahwa kapasitas model terlalu besar untuk ukuran dataset 1.200 kutipan.
3. Penyesuaian arsitektur penting untuk dataset kecil/menengah. Setelah kapasitas model dikurangi (embedding 64, 64 filter per kernel) dan regularisasi diperkuat (Dropout 0,6, L2), CNN V2 mampu meningkatkan akurasi menjadi 67,5% dan macro F1-score menjadi 0,67. Peningkatan ini memang tidak drastis, tetapi konsisten dan menunjukkan bahwa CNN dapat mengungguli baseline klasik apabila arsitektur disesuaikan dengan skala data.
4. Kelas netral tetap menjadi tantangan. Pada ketiga model, kelas netral memiliki F1-score yang lebih rendah dibanding negatif dan positif. Hal ini wajar mengingat banyak kutipan netral yang berisi pernyataan faktual atau campuran informasi dan opini tipis, sehingga mudah tertukar ke negatif atau positif.
5. Pengaruh jenis teks dan domain. Performa akurasi di kisaran 65–68% pada tiga kelas sentimen untuk kutipan pasar saham berbeda dengan hasil pada judul berita yang sebelumnya mampu mencapai akurasi di atas 90% setelah augmentasi. Perbedaan ini menegaskan bahwa jenis teks (judul berita vs kutipan media sosial) dan domain (berita umum vs diskusi pasar saham) sangat mempengaruhi tingkat kesulitan tugas analisis sentimen.

Secara umum, hasil ini menunjukkan bahwa CNN tidak otomatis lebih unggul, tetapi memerlukan desain yang proporsional dengan data. Baseline klasik tetap menjadi pondasi yang kuat, sementara CNN memberikan nilai tambah ketika kapasitas dan regularisasi diatur dengan tepat.

Simpulan

Penelitian ini mengkaji analisis sentimen kutipan media sosial berbahasa Indonesia yang berasal dari dataset ID-SMSA. Dari dataset awal yang berisi 3.288 tweet terkait pasar saham, dipilih 1.200 kutipan secara seimbang untuk tiga kelas sentimen (negatif, netral, positif), kemudian dibagi menjadi 960 data latih dan 240 data uji.

Berdasarkan hasil eksperimen, dapat disimpulkan beberapa hal:

1. Model baseline TF-IDF + Logistic Regression menghasilkan akurasi 66% dan macro F1-score 0,66, sehingga terbukti menjadi baseline yang kuat untuk analisis sentimen kutipan berbahasa Indonesia.
2. Model CNN V1 dengan kapasitas besar cenderung mengalami overfitting dan memberikan performa yang tidak lebih baik dari baseline, dengan akurasi 65% dan macro F1-score 0,65.
3. Model CNN V2 yang arsitekturnya dirampingkan dan dilengkapi regularisasi tambahan mampu meningkatkan akurasi menjadi 67,5% dan macro F1-score menjadi 0,67, sehingga sedikit melampaui baseline klasik.

4. Kelas netral secara konsisten menjadi kelas yang paling sulit diklasifikasikan dengan tepat oleh semua model, menunjukkan sifat ambigu dari banyak kutipan netral di media sosial.
5. Secara keseluruhan, CNN dengan desain arsitektur yang tepat dapat menjadi alternatif yang kompetitif untuk analisis sentimen kutipan media sosial berbahasa Indonesia, namun metode klasik tetap relevan dan efisien terutama ketika jumlah data tidak terlalu besar.

Saran

Berdasarkan hasil dan keterbatasan penelitian ini, beberapa saran pengembangan yang dapat dilakukan pada penelitian selanjutnya adalah:

1. Perluasan dan Diversifikasi Dataset. Menggunakan seluruh tweet pada ID-SMSA (3.288 data) dan menggabungkannya dengan sumber lain seperti komentar Instagram, forum diskusi, atau platform media sosial lain. Dengan demikian, model akan belajar dari ragam bahasa yang lebih luas dan diharapkan memiliki kemampuan generalisasi yang lebih baik.
2. Pengayaan Label Emosi. Selain tiga label polaritas (negatif, netral, positif), penelitian lanjutan dapat menambahkan kategori emosi seperti senang, sedih, marah, takut, atau kecewa. Hal ini memungkinkan analisis emosi publik yang lebih rinci, misalnya untuk memetakan reaksi investor terhadap peristiwa tertentu di pasar saham.
3. Eksplorasi Embedding dan Model Transformer. Mengganti embedding acak dengan embedding pra-latih seperti FastText atau word2vec Bahasa Indonesia, dan menguji model berbasis Transformer seperti IndoBERT atau IndoBERTtweet. Model-model tersebut dapat dipadukan dengan lapisan CNN atau Dense sederhana untuk melihat sejauh mana embedding kontekstual mampu meningkatkan performa di atas model CNN V2 saat ini.
4. Aspect-Based Sentiment Analysis (ABSA). Mengembangkan model aspect-based sentiment analysis yang mampu mengidentifikasi aspek spesifik (misalnya nama emiten, jenis kebijakan, atau peristiwa tertentu) yang menjadi sasaran sentimen di dalam kutipan. Pendekatan ini akan memberikan informasi lebih kaya dibanding sekadar label sentimen global.
5. Interpretabilitas Model. Menambahkan teknik interpretabilitas seperti analisis kata penting (feature importance) atau visualisasi perhatian (attention) agar diketahui kata atau frasa mana yang paling berkontribusi terhadap prediksi model. Hal ini penting untuk meningkatkan kepercayaan pengguna dan dapat digunakan sebagai bahan edukasi literasi investasi di media sosial.
6. Pengembangan Aplikasi Pemantauan Opini. Mengintegrasikan model terbaik (misalnya CNN V2 atau model berbasis Transformer di masa depan) ke dalam prototipe aplikasi pemantauan opini publik terkait pasar saham. Aplikasi ini dapat menampilkan tren sentimen harian, contoh kutipan representatif, dan potensi hubungan antara sentimen dengan pergerakan pasar, sehingga bermanfaat bagi peneliti, analis, maupun praktisi pasar modal.

Referensi

- Alfin, A. A., Kurniasari, I., & Yanuartanti, I. (2023). Analisis klasifikasi sentimen berbasis topik pada ulasan layanan Dana dan Sakuku dengan Convolutional Neural Network. *INFORMASI – Jurnal Informatika dan Sistem Informasi*, 15(2), 225–236.
- Gao, Z., Li, Z., Luo, J., & Li, X. (2022). Short text aspect-based sentiment analysis based on CNN + BiGRU. *Applied Sciences*, 12(5), 2707.
- Ghorbani, M., Bahaghighat, M., Xin, Q., & Özen, F. (2020). ConvLSTMConv network: A deep learning approach for sentiment analysis in cloud computing. *Journal of Cloud Computing*, 9, 16.
- Hartanto, J., Liundi, T., Sutoyo, R., & Andangsari, E. W. (2025). ID-SMSA: Indonesian stock market dataset for sentiment analysis. *Data in Brief*, 60, 111571.
- Hartanto, J., Liundi, T., Sutoyo, R., & Andangsari, E. W. (2025). ID-SMSA: Indonesian Stock Market Dataset for Sentiment Analysis [Dataset, versi 3]. Mendeley Data. <https://doi.org/10.17632/tn4vzs8tdw.3>
- Hartanto, J., Liundi, T., Sutoyo, R., & Andangsari, E. W. (2025). ID-SMSA: Indonesian stock market dataset for sentiment analysis [Repositori kode & dokumentasi]. GitHub. Diakses dari <https://github.com/jasonh14/ID-SMSA>
- Hoang, M., Bihorac, O. A., & Rouces, J. (2019). Aspect-based sentiment analysis using BERT. Dalam *Proceedings of the 22nd Nordic Conference on Computational Linguistics (NoDaLiDa 2019)* (pp. 187–196).
- Juwiantho, H., Setiawan, E. I., Santoso, J., & Purnomo, M. H. (2020). Sentiment analysis Twitter bahasa Indonesia berbasis Word2Vec menggunakan Deep Convolutional Neural Network. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 7(1), 181–188.
- Kusumaningrum, R., dkk. (2025). Optimizing aspect-based sentiment analysis using BERT for comprehensive analysis of Indonesian student feedback. *Applied Sciences*, 15(1), 172.
- Li, W., Jin, B., & Quan, Y. (2020). Review of research on text sentiment analysis based on deep learning. *Open Access Library Journal*, 7(3), 1–8.
- Li, W., Jin, B., & Quan, Y. (2020). Review of research on text sentiment analysis based on deep learning. *Open Access Library Journal*, 7, 1–8.
- Lin, C.-H., & Nuha, U. (2023). Sentiment analysis of Indonesian datasets based on a hybrid deep-learning strategy. *Journal of Big Data*, 10, 88.

- Mandhasiya, D. G., Murfi, H., & Bustamam, A. (2023). The hybrid of BERT and deep learning models for Indonesian sentiment analysis. *Indonesian Journal of Electrical Engineering and Computer Science*, 33(1), 475–483.
- Nip, J. Y. M., & Berthelie, B. (2024). Social media sentiment analysis. *Encyclopedia*, 4(4), 1590–1598.
- Nugroho, K. S., Sukmadewa, A. Y., Wuswilahaken, H., Bachtiar, F. A., & Yudistira, N. (2021). BERT fine-tuning for sentiment analysis on Indonesian mobile apps reviews. arXiv preprint arXiv:2107.06802.
- Praneswara, A. O., & Cahyono, N. (2023). Analisis sentimen ulasan aplikasi TikTok Shop Seller Center di Google Playstore menggunakan algoritma Naive Bayes. *Indonesian Journal of Computer Science*, 12(6), 3925–3931.
- Rajesh, A., & Hiwarkar, T. (2023). Sentiment analysis from textual data using multiple channels deep learning models. *Journal of Electrical Systems and Information Technology*, 10(1), 56.
- Ramadhan, M. F. (2023). Klasifikasi topik dan sentimen judul berita berbahasa Indonesia menggunakan metode pembobotan TF-IDF dan algoritma klasifikasi. *RIGGS – Research of Informatics, Governance, and Social Science Journal*, 1(1).
- Singgalen, Y. A. (2021). Pemilihan metode dan algoritma dalam analisis sentimen di media sosial: Systematic literature review. *Journal of Information Systems and Informatics*, 3(2), 288–300.
- Wahyudi, R., & Kusumawardhana, G. (2021). Analisis sentimen pada aplikasi Grab di Google Play Store menggunakan Support Vector Machine. *Jurnal Informatika*.
- Wijaya, M., Al Mudzakir, T., & Lestari, S. A. P. (2024). Penerapan algoritma Naive Bayes dan K-NN dalam menganalisis sentimen aplikasi TikTok Shop Seller Center berdasarkan review Google Playstore. *Scientific Student Journal for Information, Technology and Science*, 5(2).
- Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (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*, 843–857.
- Yan, W., Zhou, L., Qian, Z., Xiao, L., & Zhu, H. (2021). Sentiment analysis of student texts using the CNN-BiGRU-AT model. *Scientific Programming*, 2021, 8405623.

Zulfiqri, R., Sari, B. N., & Padilah, T. N. (2024). Analisis sentimen ulasan pengguna aplikasi media sosial Instagram pada situs Google Play Store menggunakan Naïve Bayes classifier. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 12(3), 2965–2973.