

Analisis Perbandingan Metode Klasifikasi Sentimen Berita Saham: Pendekatan Machine Learning, Deep Learning, Transfer Learning, dan Graf

Salsabila Mazya Permataning Tyas¹, Riyanarto Sarno¹, Bagus Setya Rintyarna²

¹Institut Teknologi Sepuluh Nopember

²Universitas Muhammadiyah Jember

e-mail Correspondensi: salsa25mazya@gmail.com

ABSTRAK

Pasar saham merupakan arena yang dinamis, di mana keputusan investasi seringkali dipengaruhi oleh berbagai faktor, termasuk informasi yang diperoleh dari berita saham. Analisis sentimen berita saham menjadi krusial dalam memahami reaksi pasar terhadap berita tertentu. Tujuan penelitian ini untuk mengklasifikasikan sentimen berita saham dalam membantu investor dan pelaku pasar untuk membuat keputusan yang lebih baik. Oleh karena itu, pengembangan metode yang efektif untuk mengklasifikasikan sentimen berita saham menjadi suatu kebutuhan mendesak. Penelitian ini fokus pada pemahaman sentimen di dalam berita saham dan menghadirkan perbandingan antara empat metode klasifikasi yang berbeda, yaitu *Machine Learning*, *Deep Learning*, *Transfer Learning*, dan metode berbasis Graf. Dengan memahami perbedaan kinerja dan kelebihan masing-masing metode, diharapkan penelitian ini dapat memberikan pandangan yang lebih mendalam terkait pilihan teknik yang paling sesuai dalam menghadapi kompleksitas analisis sentimen di pasar saham. Hasil eksperimen dan evaluasi kinerja masing-masing metode diukur dengan menggunakan nilai akurasi. Dari seluruh percobaan yang dilakukan nilai akurasi tertinggi diperoleh menggunakan BERT sebesar 81%.

Kata kunci: machine learning; deep learning; berita saham; transfer learning; graf

ABSTRACT

The stock market is a dynamic arena, where investment decisions are often influenced by various factors, including information obtained from stock news. Stock news sentiment analysis is crucial in understanding market reactions to certain news. The aim of this research is to classify stock news sentiment to help investors and market players make better decisions. Therefore, developing an effective method for classifying stock news sentiment is an urgent need. This research focuses on understanding sentiment in stock news and presents a comparison between four different classification methods, namely Machine Learning, Deep Learning, Transfer Learning, and Graph-based methods. By understanding the differences in performance and advantages of each method, it is hoped that this research can provide a more in-depth view regarding the choice of techniques that are most suitable for dealing with the complexity of sentiment analysis in the stock market. Experimental results and performance evaluation of each method are measured using accuracy values. Of all the experiments carried out, the highest accuracy value was obtained using BERT at 81%.

Keywords: machine learning; deep learning; stock news; transfer learning; graph

PENDAHULUAN

Dalam era informasi digital saat ini, pasar keuangan dipengaruhi oleh berita dan opini yang tersebar luas melalui berbagai platform media. Kondisi pasar saham seringkali dipicu oleh peristiwa-peristiwa tertentu yang tercermin dalam berita, dan para pelaku pasar mencari informasi yang dapat membantu mereka membuat keputusan investasi yang tepat. Analisis sentimen terhadap berita saham menjadi elemen kunci dalam memahami respons pasar terhadap informasi (Haryono dkk., 2022).

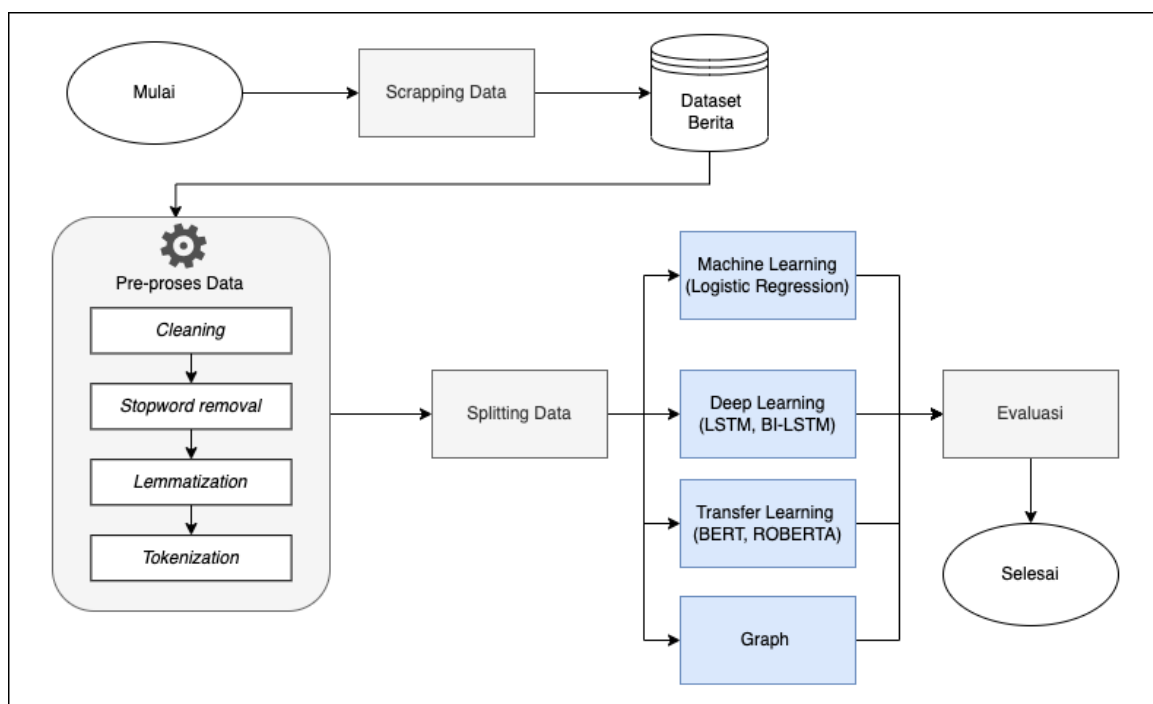
Penelitian ini menyelidiki ranah analisis sentimen dalam konteks berita saham, dengan fokus khusus pada emiten saham Indonesia. Ranah analisis sentimen merupakan bidang yang kompleks, mengingat sifat bahasa yang nuansial dan tergantung pada konteks (Rintyarna dkk., 2019). Menganalisis sentimen dalam berita saham melibatkan dekripsi bukan hanya pada ekspresi eksplisit, tetapi juga pada nada dan implikasi yang mungkin memengaruhi reaksi pasar secara signifikan. Oleh karena itu, klasifikasi sentimen dalam berita saham menjadi alat kritis bagi investor untuk memahami sentimen pasar dan membuat keputusan investasi yang terinformasi. Metode yang dipilih, yaitu

Machine Learning, Deep Learning, Transfer Learning, dan metode berbasis Graf, mewakili pendekatan yang beragam untuk mengatasi tantangan analisis sentimen dalam konteks finansial.

Machine Learning dan *Deep Learning* telah menjadi solusi yang semakin populer dalam mengatasi tugas-tugas kompleks seperti klasifikasi sentimen. Logistic Regression, sebagai metode *Machine Learning* klasik, telah berhasil digunakan dalam beberapa konteks analisis sentimen (Sonkavde dkk., 2023). Sementara itu, LSTM dan BI-LSTM, sebagai salah satu jenis arsitektur *Deep Learning*, dikenal mampu menangkap dependensi jangka panjang dalam data sekuensial, seperti teks berita (Chou dkk., 2021). Pemilihan word embedding juga menjadi aspek penting dalam analisis sentimen. TF-IDF memberikan representasi bobot kata berdasarkan frekuensinya dalam dokumen dan invers frekuensi mereka dalam korpus, sedangkan Word2Vec menangkap struktur semantik dan hubungan antar kata dalam ruang vektor (Permataning Tyas dkk., 2022). Selain itu terdapat metode transfer learning yang cukup populer digunakan sekarang seperti BERT dan RoBERTa. Dan juga terdapat metode berbasis graf seperti GCN yang akan dilakukan dalam penelitian ini.

Tujuan utama dari penelitian ini untuk memberikan kontribusi pada perkembangan pengambilan keputusan keuangan dengan melakukan perbandingan komprehensif terhadap empat metode klasifikasi yang berbeda: *Machine Learning*, *Deep Learning*, *Transfer Learning*, dan metode berbasis Graf. Dengan merinci kontribusi dan metodologi yang diusulkan dalam penelitian ini, diharapkan dapat memberikan kontribusi yang signifikan terhadap pemahaman kita tentang keterkaitan antara sentimen berita saham dan dinamika pasar keuangan. Oleh karena itu, penelitian ini diarahkan untuk menjembatani kesenjangan dalam literatur terkait dan meningkatkan kemampuan prediktif dalam mengklasifikasikan sentimen berita saham.

Gambar 1. Flowchart Penelitian



METODE PENELITIAN

Pada bagian ini akan dijelaskan tentang langkah-langkah, data, metode, dan juga evaluasi yang digunakan di penelitian ini dalam Gambar 1.

Pengumpulan Data

Data berita saham sebanyak 1000 sampel diperoleh melalui web scraping dari portal berita saham, kontan.co.id. Pengambilan sampel dilakukan untuk mencakup berbagai kategori, termasuk investasi, perkembangan industri, dan faktor-faktor yang berpotensi memengaruhi pasar saham. Setiap berita dilabeli secara manual dengan sentimen positif atau negatif oleh ahli,

mempertimbangkan konteks pasar dan dampak berita terhadap pergerakan harga saham. Pengecekan konsistensi label dilakukan dengan memeriksa sampel acak dan mengadakan diskusi berkala untuk memastikan keseragaman pemahaman dalam pemberian label. Dataset yang dihasilkan diharapkan mencerminkan variasi sentimen berita saham dan menyediakan dasar yang kuat untuk analisis sentimen dengan metode klasifikasi yang beragam dalam penelitian ini.

Pre-proses Data

Data teks berita saham melalui tahap pemrosesan untuk memastikan kualitas dan keterbacaan optimal sebelum diolah lebih lanjut. Proses pembersihan data seperti pada Tabel 1 melibatkan langkah-langkah seperti *cleaning*, *stop words*, *lemmatization*, dan tokenisasi, untuk menghilangkan elemen yang tidak relevan dan merapikan struktur teks. Selanjutnya, data teks diubah menjadi representasi numerik melalui proses vektorisasi, memungkinkan model untuk memproses informasi secara efisien. Teknik seperti Term Frequency-Inverse Document Frequency (TF-IDF) atau Word Embeddings dapat digunakan untuk memberikan bobot pada kata-kata berdasarkan relevansinya dalam teks.

Tabel 1. Tahap Pre-proses Data

No	Tahap	Data Awal	Data Bersih
1	Cleaning	bunga deposito bri naik di awal tahun 2020, cek bunga terbaru	bunga deposito bri naik di awal tahun 2020 cek bunga terbaru
2	Stopwords Removal	bunga deposito bri naik di awal tahun 2020 cek bunga terbaru	bunga deposito bri naik awal tahun 2020 cek bunga terbaru
3	Lemmatization	bunga deposito bri naik awal tahun 2020 cek bunga terbaru	bunga deposito bri naik awal tahun 2020 cek bunga baru
4	Tokenisasi	bunga deposito bri naik awal tahun 2020 cek bunga baru	[bunga, deposito, bri, naik, awal, tahun, 2020, cek, bunga, baru]

Pengaturan skala dan normalisasi dapat dilakukan untuk memastikan keseragaman data, sementara pembagian dataset menjadi subset training dan testing menjadi langkah selanjutnya. Dataset yang telah diproses kemudian siap untuk dijadikan input pada berbagai metode klasifikasi, termasuk *Machine Learning*, *Deep Learning*, *Transfer Learning*, dan pendekatan berbasis Graf, yang akan dieksplorasi dalam penelitian ini. Proses pemrosesan data ini menjadi tahap kritis untuk memastikan data yang bersih dan terstruktur sehingga memberikan dasar yang kuat bagi model klasifikasi untuk mengidentifikasi sentimen dalam berita saham dengan akurasi tinggi.

Pembagian Data

Dalam penelitian ini, data dibagi menggunakan metode *splitting* dengan perbandingan 80% untuk data pelatihan (*training*) dan 20% untuk data pengujian (*testing*). Pendekatan ini memiliki tujuan untuk memastikan bahwa model yang dikembangkan memiliki kemampuan untuk menggeneralisasi dengan baik pada data yang tidak pernah dilihat selama proses pelatihan.

Data pelatihan, yang mencakup 80% dari total dataset, digunakan untuk melatih model pada berbagai pola dan karakteristik sentimen berita saham. Dengan memberikan model akses ke sebagian besar data, diharapkan model dapat menangkap informasi yang cukup untuk memahami variabilitas dalam set data.

Data pengujian, sebesar 20% dari total dataset, dipisahkan secara eksklusif untuk mengevaluasi kinerja model yang telah dilatih. Penggunaan data pengujian yang terpisah membantu memvalidasi kemampuan model dalam melakukan prediksi pada data yang belum pernah dilihat sebelumnya. Hal ini memberikan gambaran yang lebih akurat tentang seberapa baik model dapat diterapkan pada situasi dunia nyata.

Penting untuk mencatat bahwa pembagian data 80%:20% adalah suatu kompromi yang umum digunakan di dalam penelitian, memberikan proporsi yang seimbang antara volume data yang digunakan untuk pelatihan dan evaluasi. Proporsi ini memberikan keleluasaan yang memadai untuk melatih model dengan cukup data, sambil tetap menyisakan sebagian data untuk mengukur kinerja model secara obyektif.

Machine Learning

Machine Learning (ML) merupakan paradigma komputasi yang memungkinkan sistem untuk belajar dari data dan meningkatkan kinerjanya seiring waktu tanpa program yang eksplisit. Dalam konteks analisis sentimen berita saham, algoritma ML dapat digunakan untuk mengklasifikasikan teks berita ke dalam kategori sentimen tertentu, seperti positif, negatif, atau netral. *Logistic Regression* adalah salah satu algoritma *Machine Learning* yang sering digunakan dalam tugas klasifikasi biner, di mana data harus dikelompokkan menjadi dua kelas. Meskipun namanya menyiratkan regresi, *Logistic Regression* sebenarnya digunakan untuk klasifikasi. Algoritma ini cocok untuk tugas analisis sentimen karena mampu memberikan probabilitas prediksi yang dapat diinterpretasikan.

Proses *Logistic Regression* dimulai dengan memodelkan hubungan antara variabel independen (fitur dari berita saham) dan variabel dependen (sentimen berita) dengan menggunakan fungsi logistik. Fungsi ini memetakan nilai-nilai fitur ke dalam rentang 0 hingga 1, menciptakan probabilitas bahwa suatu berita termasuk dalam kategori sentimen tertentu. Dalam konteks analisis sentimen berita saham, *Logistic Regression* dapat melibatkan fitur-fitur seperti kata-kata kunci yang muncul dalam teks berita. Setelah dilatih dengan dataset yang sudah dilabeli, model *Logistic Regression* dapat digunakan untuk memprediksi sentimen berita yang belum dilihat sebelumnya dengan tingkat akurasi yang dapat diukur.

Penggunaan *Logistic Regression* dalam analisis sentimen memberikan keuntungan interpretabilitas yang tinggi, di mana kita dapat memahami dampak relatif dari setiap fitur terhadap prediksi sentimen. Meskipun sederhana, *Logistic Regression* dapat menjadi pilihan yang efektif terutama ketika interpretabilitas dan kecepatan pelatihan model menjadi pertimbangan utama.

Deep Learning

Deep Learning (DL) adalah subbidang dari *Machine Learning* yang berfokus pada penggunaan arsitektur jaringan saraf yang lebih kompleks, dikenal sebagai *neural networks*, untuk mengekstrak representasi fitur yang lebih abstrak dari data. Dalam analisis sentimen berita saham, *Deep Learning* dapat memberikan kemampuan untuk menangkap hubungan dan pola yang lebih kompleks dalam teks (Maqsood dkk., 2020), serta menghasilkan hasil yang lebih akurat (Y. Gao dkk., 2021). LSTM dan BiLSTM adalah dua arsitektur jaringan saraf yang umum digunakan dalam tugas-tugas pemrosesan bahasa alami, seperti analisis sentimen. Kedua arsitektur ini dirancang untuk mengatasi tantangan jaringan saraf konvensional dalam menangani dependensi jarak jauh dan memahami konteks yang lebih panjang (Hamayel & Owda, 2021).

Transfer Learning

Transfer Learning adalah paradigma dalam *Deep Learning* di mana model yang sudah dilatih pada tugas tertentu digunakan sebagai dasar untuk tugas serupa yang lebih kompleks atau spesifik. *Transfer Learning* memungkinkan model untuk memanfaatkan pengetahuan yang sudah diperoleh dari sumber tugas untuk meningkatkan kinerja pada tugas target. Dalam analisis sentimen berita saham, *Transfer Learning* dapat membantu model memahami konteks bahasa dan nuansa yang lebih kompleks.

Graf

Dalam konteks analisis sentimen berita saham, representasi data dapat dianggap sebagai graf di mana setiap simpul mewakili berita saham dan sambungan antar simpul menggambarkan keterkaitan antar berita tersebut. Graf ini menciptakan struktur jaringan kompleks yang memungkinkan penggalian informasi lebih dalam terkait korelasi dan dampak antar berita saham.

Graph Convolutional Network atau GCN adalah arsitektur jaringan saraf yang khusus dirancang untuk bekerja dengan data graf. Dalam GCN, setiap simpul dan sambungan dijelajahi secara kontekstual berdasarkan hubungan tetangga mereka dalam graf. GCN memungkinkan pengenalan pola dan informasi kompleks yang muncul dari struktur jaringan.

Langkah-langkah Penggunaan GCN:

- Pembentukan Graf: Setiap berita saham direpresentasikan sebagai simpul, dan hubungan antar berita dibentuk sebagai sambungan dalam graf. Misalnya, jika dua berita memiliki hubungan terkait dampak yang sama pada saham tertentu, sambungan antara simpul-simpul tersebut diperkenalkan.

- Ekstraksi Fitur: Informasi dari setiap simpul dan tetangganya diekstraksi sebagai fitur, mencakup representasi teks dari berita dan hubungan jaringan yang dibentuk.
- Graph Convolution: GCN melakukan proses konvolusi pada representasi fitur setiap simpul berdasarkan informasi dari tetangganya. Proses ini memungkinkan model untuk memahami dampak dan sentimen berita dengan mempertimbangkan konteks jaringan.
- Klasifikasi Sentimen: Hasil akhir dari GCN adalah kemampuan untuk mengklasifikasikan sentimen berita saham. Model ini memungkinkan penilaian yang lebih holistik terhadap dampak berita terhadap sentimen pasar saham.

Pendekatan berbasis graf, khususnya menggunakan GCN, menambah dimensi analisis sentimen dengan mempertimbangkan hubungan antar berita saham. Ini memberikan pandangan yang lebih komprehensif dan kontekstual terhadap pergerakan pasar yang dapat dihasilkan oleh serangkaian berita yang saling terkait.

Evaluasi

Evaluasi adalah langkah kritis dalam menilai kinerja model analisis sentimen, memastikan bahwa model dapat memberikan prediksi yang akurat dan dapat diandalkan. Dalam konteks analisis sentimen berita saham, evaluasi model melibatkan perbandingan hasil prediksi dengan label sebenarnya dari data uji. Salah satu metrik evaluasi yang umum digunakan adalah akurasi.

Akurasi mengukur sejauh mana model dapat memprediksi sentimen dengan benar. Ini dihitung dengan membagi jumlah prediksi yang benar (positif dan negatif) dengan total jumlah prediksi. Akurasi memberikan gambaran umum tentang seberapa baik model dapat mengklasifikasikan sentimen berita saham. Semakin tinggi akurasi, semakin baik model dalam melakukan prediksi yang benar.

HASIL DAN PEMBAHASAN

Bagian ini memuat hasil-hasil dari penelitian serta pembahasan menyeluruh dari masing-masing hasil yang didapatkan. Hasil nilai akurasi pada penelitian ini terdapat pada Tabel 2 berikut.

Tabel 2. Hasil Akurasi Percobaan

No	Metode	Embedding	Akurasi
1.	Logistic Regression	TF-IDF	68.5%
2.	LSTM	TF-IDF	70%
3.	LSTM	Word2Vec	69%
4.	Bi-LSTM	Word2Vec	71%
5.	BERT	-	81%
6.	RoBERTa	-	79%
7.	GCN Text	-	73.1%

Penelitian ini bertujuan untuk mengevaluasi berbagai metode dalam analisis sentimen berita saham. Enam pendekatan yang diuji, yaitu TF-IDF + LR, LSTM, LSTM + Embedding Word2Vec, Bi-LSTM + Embedding Word2Vec, GCN, BERT, dan RoBERTa, memberikan gambaran tentang kinerja model dalam tugas ini.

Metode tradisional seperti TF-IDF + LR mencapai akurasi sebesar 68.5%, menunjukkan bahwa pendekatan ini masih relevan, meskipun kinerja relatifnya lebih rendah dibandingkan dengan metode yang lebih kompleks, diperkuat dengan penelitian sebelumnya (Majumder dkk., 2021) yang menggunakan logistic regression mampu mendapatkan nilai akurasi terbaik dibandingkan dengan metode machine learning lainnya seperti *Random Forest* dan *Naïve Bayes Classifier*. LSTM sebagai pendekatan *Deep Learning*, memberikan akurasi 70%, menunjukkan kemampuan untuk menangkap hubungan temporal dalam teks, sementara penggunaan embedding Word2Vec dengan LSTM menghasilkan akurasi sebesar 69%. Hal ini juga diperkuat dengan pernyataan (Savitri dkk., 2021) melalui percobaan yang telah dilakukan bahwa metode LSTM menghasilkan prediksi yang cukup

baik untuk mengetahui sentimen, serta hasil yang diperoleh lebih baik dibandingkan dengan metode neural network lainnya seperti RNN.

Peningkatan yang lebih signifikan terlihat pada metode yang lebih kompleks, seperti Bi-LSTM + Embedding Word2Vec yang mencapai akurasi 71%. Pendekatan ini menggabungkan keunggulan LSTM dengan kemampuan Bi-LSTM untuk memahami konteks dari kedua arah teks. Jelas terlihat pada penelitian lain yang dilakukan oleh (Zhou & Xu, 2016) bahwa akurasi LSTM lebih rendah 3% dibandingkan dengan BI-LSTM karena BI-LSTM terdapat 2 lapisan LSTM yang bergerak saling berlawanan dari kata pertama hingga kata terakhir (*forward*) dan dari kata terakhir ke kata pertama (*backward*) sehingga pembelajaran model lebih mengenal konteks kalimat. Hasil menarik juga diperoleh dari pendekatan berbasis graf yaitu GCN, karena GCN memanfaatkan struktur jaringan graf untuk memahami hubungan antar berita saham, memberikan wawasan tambahan terkait dampak kolektif dari serangkaian berita terhadap sentimen pasar saham. Hasil yang diperoleh menggunakan GCN mencapai akurasi sebesar 73.1%.

Puncak kinerja tercapai pada metode Transfer Learning, di mana BERT mencapai akurasi 81%, sementara RoBERTa mencapai 79%. Kedua metode ini menunjukkan keunggulan dalam memahami konteks bahasa yang kompleks dan nuansa dalam teks berita saham. Keunggulan tersebut diperjelas oleh (Sousa dkk., 2019) dalam penelitiannya bahwa dengan kemampuannya untuk memahami konteks kata-kata dalam konteks dua arah (maju dan mundur), BERT mampu menangkap hubungan yang lebih kompleks dan kontekstual antar kata-kata dalam suatu kalimat. Sedangkan RoBERTa dalam penelitian (Liao dkk., 2021) lebih fokus pada kemampuan umum pengenalan dan pemahaman bahasa, dengan menghilangkan tugas khusus yang dimiliki BERT, seperti *Next Sentence Prediction*. RoBERTa juga dilatih pada volume data yang lebih besar, menghasilkan representasi fitur yang lebih kuat dan umum. Dengan demikian, keunggulan Transfer Learning terletak pada kemampuannya untuk memanfaatkan pengetahuan yang sudah ada, mengurangi kebutuhan akan dataset pelatihan besar dan meningkatkan kinerja model pada tugas target (Z. Gao dkk., 2019).

Pembahasan menggarisbawahi bahwa keputusan pemilihan metode harus mempertimbangkan kompleksitas model, kebutuhan sumber daya komputasi, dan interpretabilitas hasil. Metode *Transfer Learning*, terutama BERT, muncul sebagai pilihan yang paling efektif dalam analisis sentimen berita saham pada dataset ini karena dari hasil percobaan yang telah dilakukan BERT mampu menghasilkan akurasi terbaik. Namun, GCN juga memberikan hasil yang solid, hal ini dikarenakan kemampuan GCN untuk memodelkan hubungan dan ketergantungan kompleks antara berita saham dalam bentuk grafik yang mungkin sulit diakses oleh metode lain. Keunggulan masing-masing metode memberikan pemahaman yang mendalam tentang kemungkinan penerapan praktis di pasar saham yang dinamis.

KESIMPULAN

Dalam penelitian ini, berbagai metode telah dievaluasi untuk tugas analisis sentimen berita saham pada dataset. Hasil eksperimen menunjukkan bahwa metode Transfer Learning, khususnya BERT memberikan kinerja unggul dalam memahami sentimen pasar saham dari teks berita. BERT mencapai akurasi tertinggi sebesar 81%. Meskipun metode tradisional seperti TF-IDF + LR dan pendekatan *Deep Learning* seperti LSTM memberikan kinerja yang layak, mereka tergolong di bawah metode *Transfer Learning* dan GCN. Penelitian ini memberikan landasan yang kuat untuk penelitian selanjutnya dalam analisis sentimen berita saham seperti optimasi model transfer learning, bisa dilakukan eksplorasi lebih lanjut mengenai teknik ensemble, dan analisis sentimen dengan menggunakan multi faktor.

DAFTAR PUSTAKA

Chou, C., Park, J., & Chou, E. (2021). Predicting Stock Closing Price After COVID-19 Based on Sentiment Analysis and LSTM. *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 2752–2756. <https://doi.org/10.1109/IAEAC50856.2021.9390845>

- Gao, Y., Wang, R., & Zhou, E. (2021). Stock Prediction Based on Optimized LSTM and GRU Models. *Scientific Programming*, 2021, 1–8. <https://doi.org/10.1155/2021/4055281>
- Gao, Z., Feng, A., Song, X., & Wu, X. (2019). Target-Dependent Sentiment Classification With BERT. *IEEE Access*, 7, 154290–154299. <https://doi.org/10.1109/ACCESS.2019.2946594>
- Hamayel, M. J., & Owda, A. Y. (2021). A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms. *AI*, 2(4), 477–496. <https://doi.org/10.3390/ai2040030>
- Haryono, A. T., Sarno, R., & Abdullah, R. (2022). Aspect-Based Sentiment Analysis of Financial Headlines and Microblogs Using Semantic Similarity and Bidirectional Long Short-Term Memory. *International Journal of Intelligent Engineering and Systems*.
- Liao, W., Zeng, B., Yin, X., & Wei, P. (2021). An improved aspect-category sentiment analysis model for text sentiment analysis based on RoBERTa. *Applied Intelligence*, 51(6), 3522–3533. <https://doi.org/10.1007/s10489-020-01964-1>
- Majumder, S., Aich, A., & Das, S. (2021). Sentiment Analysis of People During Lockdown Period of COVID-19 Using SVM and Logistic Regression Analysis. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3801039>
- Maqsood, H., Mehmood, I., Maqsood, M., Yasir, M., Afzal, S., Aadil, F., Selim, M. M., & Muhammad, K. (2020). A local and global event sentiment based efficient stock exchange forecasting using deep learning. *International Journal of Information Management*, 50, 432–451. <https://doi.org/10.1016/j.ijinfomgt.2019.07.011>
- Permataning Tyas, S. M., Rintyarna, B. S., & Suharso, W. (2022). The Impact of Feature Extraction to Naïve Bayes Based Sentiment Analysis on Review Dataset of Indihome Services. *Digital Zone: Jurnal Teknologi Informasi dan Komunikasi*, 13(1), 1–10. <https://doi.org/10.31849/digitalzone.v13i1.9158>
- Rintyarna, B. S., Sarno, R., & Faticah, C. (2019). Evaluating the performance of sentence level features and domain sensitive features of product reviews on supervised sentiment analysis tasks. *Journal of Big Data*, 6(1), 84. <https://doi.org/10.1186/s40537-019-0246-8>
- Savitri, N. L. P. C., Rahman, R. A., Venyutzky, R., & Rakhmawati, N. A. (2021). Analisis Klasifikasi Sentimen Terhadap Sekolah Daring pada Twitter Menggunakan Supervised Machine Learning. *Jurnal Teknik Informatika dan Sistem Informasi*, 7(1). <https://doi.org/10.28932/jutisi.v7i1.3216>
- Sonkavde, G., Dharrao, D. S., Bongale, A. M., Deokate, S. T., Doreswamy, D., & Bhat, S. K. (2023). Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications. *International Journal of Financial Studies*, 11(3), 94. <https://doi.org/10.3390/ijfs11030094>
- Sousa, M. G., Sakiyama, K., Rodrigues, L. de S., Moraes, P. H., Fernandes, E. R., & Matsubara, E. T. (2019). BERT for Stock Market Sentiment Analysis. *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, 1597–1601. <https://doi.org/10.1109/ICTAI.2019.00231>
- Zhou, Z., & Xu, L. (2016). *Amazon Food Review Classification using Deep Learning and Recommender System*.