

BAB II

LANDASAN TEORI

2.1 Saham Perbankan

Saham perbankan adalah jenis saham yang diterbitkan oleh perusahaan-perusahaan di sektor perbankan. Saham perbankan menjadi salah satu instrumen investasi penting, karena kinerjanya tidak hanya mencerminkan performa individu perusahaan tetapi juga memberikan gambaran kondisi ekonomi yang lebih luas (Syahid et al., 2024). Saham-saham ini sering dipandang sebagai tolak ukur kesehatan sektor keuangan secara keseluruhan karena peran krusial bank dalam menyediakan likuiditas, memfasilitasi transaksi keuangan, dan mendorong pertumbuhan ekonomi (Novianti et al., 2024). Saham perbankan merupakan bagian dari sektor keuangan yang memainkan peran penting dalam stabilitas perekonomian suatu negara (Athari et al., 2023).

Dalam konteks penelitian, saham perbankan memiliki karakteristik khusus yang menjadikannya menarik untuk diteliti, seperti tingkat sensitivitas terhadap kondisi makroekonomi dan fluktuasi pasar (Farich & Kornitasari, 2024). Penelitian terkait saham perbankan juga tidak hanya berfokus pada analisis teknikal dan fundamental tetapi juga pada teknik modern seperti analisis sentimen dan model peramalan berbasis data historis (Pradiptyo et al., 2024)..

2.2 Text Mining

Text mining adalah proses mengekstraksi informasi dan pola bermakna dari data teks tidak terstruktur. Sebagai salah satu cabang dari *data science*, *text mining* menggabungkan teknik dari *Natural Language Processing* (NLP), *machine*

learning, dan statistika untuk menganalisis, memahami, serta menginterpretasikan teks (Shaffiei et al., 2023). Dengan semakin besarnya *volume* data dari berbagai sumber seperti media sosial, berita, ulasan pelanggan, serta dokumen bisnis *text mining* menjadi alat penting dalam memperoleh wawasan yang bernilai dari teks yang sebelumnya sulit diolah secara otomatis (Ranjan et al., 2023).

Secara umum, text mining mencakup beberapa tahapan utama:

- a. Pengumpulan Data – Mengambil teks dari berbagai sumber, seperti artikel berita, media sosial, forum diskusi, atau laporan bisnis.
- b. Prapemrosesan Teks – Membersihkan dan menyiapkan teks agar lebih mudah dianalisis. Proses ini mencakup *case folding*, tokenisasi, penghapusan stopwords, stemming/lemmatization, serta transformasi teks ke dalam format numerik seperti TF-IDF atau *word embeddings*.
- c. Pemodelan Teks – Menggunakan teknik seperti *clustering*, klasifikasi, atau topic modeling untuk menemukan pola dan hubungan dalam teks.
- d. Analisis & Interpretasi – Menganalisis hasil model untuk mendapatkan *insight* yang dapat digunakan dalam pengambilan keputusan.

Dalam berbagai bidang, *text mining* banyak digunakan untuk berbagai keperluan, seperti analisis tren pasar, pemantauan media sosial, ekstraksi informasi dari dokumen hukum, serta pemrosesan otomatis dokumen bisnis.

2.3 Analisis Sentimen

Analisis sentimen adalah metode untuk mengevaluasi opini, emosi, atau perasaan yang terkandung dalam suatu teks. Teknologi ini bertujuan untuk menentukan apakah sentimen dalam teks tersebut bersifat positif, negatif, atau

netral (Lourdusamy et al., 2024). Dengan menggunakan pemrosesan bahasa alami (*Natural Language Processing*), analisis sentimen telah berkembang dari pendekatan manual yang mengandalkan pengelompokan opini berdasarkan pengamatan, ke teknik berbasis pembelajaran mesin yang memungkinkan pengolahan dataset dalam skala besar secara cepat dan efisien.

Analisis sentimen pasar modal diterapkan untuk mengevaluasi persepsi pasar terhadap berita, laporan keuangan, atau interaksi media sosial yang berkaitan dengan saham (Begum & Awalakki, 2024). Pendekatan ini mampu memberikan wawasan penting tentang bagaimana berbagai faktor memengaruhi pergerakan harga saham. Misalnya, peningkatan kinerja keuangan seperti ROA atau EPS biasanya mendorong kenaikan harga saham karena mencerminkan profitabilitas perusahaan. Sebaliknya, tekanan eksternal seperti inflasi tinggi, kenaikan suku bunga BI7DRR, atau pelemahan nilai tukar rupiah dapat menurunkan ekspektasi pasar dan berdampak negatif pada harga saham (Sukartaatmadja et al., 2023).

Khususnya pada saham sektor perbankan, analisis sentimen menjadi semakin relevan karena sentimen pasar sangat dipengaruhi oleh berita suku bunga, kebijakan moneter, atau kebijakan pemerintah (Anjum et al., 2024). Dengan bantuan model modern seperti *BERT* dan *RoBERTa*, analisis sentimen saham perbankan dapat memberikan prediksi yang lebih akurat tentang bagaimana berita keuangan dan opini pasar memengaruhi harga saham.

2.4 Peramalan

Peramalan adalah proses memprediksi kejadian di masa depan berdasarkan data historis dan pola yang teridentifikasi. Berbagai metode telah digunakan untuk

peramalan, mulai dari pendekatan statistik tradisional seperti Regresi Linear dan *ARIMA*, hingga metode yang lebih kompleks seperti *machine learning* dan *deep learning* (Fisher et al., 2024). Teknik tradisional umumnya efektif untuk data yang memiliki pola linier sederhana, tetapi sering kali kurang optimal dalam menangani data dengan hubungan non-linear atau pola yang kompleks (Poongadan & Lineesh, 2024).

Penelitian terakhir pada kemajuan pembelajaran mesin dan pembelajaran mendalam (*Deep Learning*) telah membawa perubahan besar dalam metode peramalan. Metode pembelajaran mesin seperti *Random Forest* dan *Gradient Boosting* mampu menemukan pola-pola sulit yang tidak dapat diidentifikasi oleh teknik konvensional (Chang et al., 2024). Selain itu, model *deep learning* seperti *Long Short-Term Memory (LSTM)* dan *Gated Recurrent Unit (GRU)* sangat efektif untuk data deret waktu, karena mampu menangkap hubungan temporal jangka panjang (Sivadasan et al., 2025).

Model *deep learning* untuk peramalan harga saham menjadi sangat relevan karena sifat data saham yang bersifat dinamis, kompleks, dan dipengaruhi oleh berbagai faktor eksternal (Dong & Hao, 2024).

2.5 Pra-Pemrosesan Data

Prapemrosesan data merupakan tahap penting dalam analisis sentimen dan *forecasting* saham, yang bertujuan untuk membersihkan, menyusun, serta mentransformasikan data agar lebih sesuai untuk digunakan dalam model pembelajaran mesin (Maharana et al., 2022).

2.5.1 Analisis Sentimen

Analisis sentimen berbasis teks, prapemrosesan melibatkan beberapa tahap untuk memastikan data teks siap digunakan dalam model (Jabbarova & Nasrullayev, 2023).

a. *Text Cleaning*

Proses membersihkan teks dengan menghapus karakter khusus, tanda baca, angka, URL, dan elemen yang tidak relevan. Langkah ini memastikan bahwa teks dalam kondisi optimal sebelum diproses lebih lanjut oleh model.

b. *Tokenization*

Proses memecah teks menjadi unit yang lebih kecil (tokens) agar dapat dipahami oleh model. Berbeda dengan metode tokenisasi tradisional yang memisahkan teks berdasarkan spasi atau kata, RoBERTa menggunakan *subword* tokenization berbasis Byte-Pair Encoding (BPE).

c. *Truncation*

Setelah teks diubah menjadi token, panjang setiap teks akan bervariasi, sementara model *deep learning* memerlukan input dengan ukuran yang seragam. Oleh karena itu, jika teks memiliki jumlah token yang melebihi batas maksimum, sistem akan memangkas (truncate) token dari akhir teks agar tetap berada dalam batas yang dapat diproses oleh model (Shu, 2024).

2.5.2 Peramalan

Prapemrosesan untuk peramalan saham menggunakan data historis bertujuan untuk memastikan kualitas dan relevansi data sebelum digunakan dalam model prediksi.

a. Datetime Index

Mengubah indeks dataset menggunakan Tanggal dengan format *datetime* memastikan bahwa model dapat memahami urutan kronologis data. Dengan menjadikan tanggal sebagai kolom eksplisit, data menjadi lebih mudah untuk difilter, digabung, atau dipetakan berdasarkan waktu, sambil tetap mempertahankan informasi temporal yang penting untuk keperluan forecasting.

b. Normalisasi

Normalisasi dilakukan untuk memastikan bahwa skala data seragam, sehingga model dapat lebih mudah memahami pola pergerakan harga. Untuk data close price, metode normalisasi yang umum digunakan adalah *Min-Max Scaling*, yang menskalakan harga saham ke rentang $[0,1]$ (Samal & Dash, 2023). Formula *Min-Max Scaling* dapat dilihat pada persamaan (2.1).

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.1)$$

Dengan menggunakan metode ini, nilai harga saham tetap dalam skala yang lebih kecil tanpa kehilangan pola pergerakan yang penting bagi model prediksi.

c. Sliding window

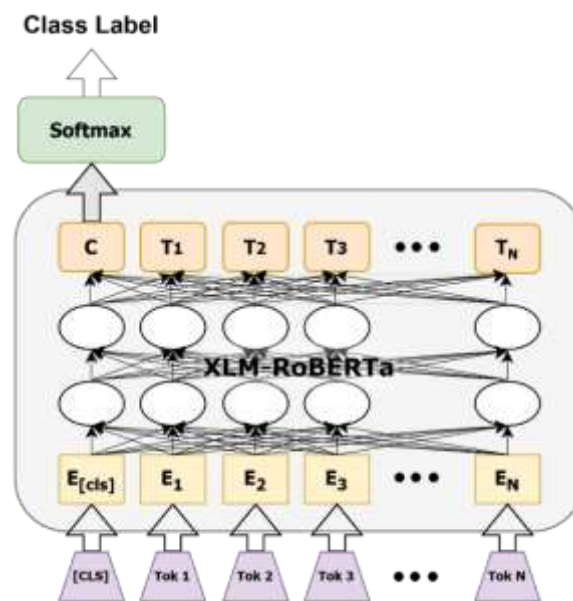
Metode *window* digunakan untuk membentuk struktur input-output yang diperlukan dalam pelatihan model prediksi, seperti yang diterapkan oleh (Chengcheng Chen et al., 2022). Teknik ini membagi data historis secara bertahap ke dalam jendela (*window*) berukuran tetap, di mana setiap *window* terdiri dari sejumlah titik data masa lalu yang digunakan untuk memprediksi titik data berikutnya. Setelah satu *window* terbentuk, jendela tersebut digeser satu langkah ke depan untuk menghasilkan pasangan input-output baru.

Dengan pendekatan ini, model dapat mempelajari pola perubahan nilai seiring waktu dan memahami hubungan temporal antar data. *Sliding window* menjadi dasar dalam membangun hubungan sekuensial yang kuat dalam model berbasis deep learning seperti GRU, khususnya untuk tugas peramalan jangka pendek maupun jangka panjang.

2.6 *Financial Robustly Optimized BERT Pretraining Approach*

Adaptasi dari model *RoBERTa* yang dirancang khusus untuk memahami teks dalam domain keuangan. *RoBERTa* sendiri merupakan pengembangan dari *BERT* dengan berbagai peningkatan, seperti pelatihan lebih lama pada dataset lebih besar, penghapusan tugas *Next Sentence Prediction* (NSP), serta penggunaan masking dinamis (Semary et al., 2023). Dalam analisis sentimen keuangan, model ini dapat dioptimalkan agar lebih sesuai dengan konteks pasar saham Indonesia.

XLNet-RoBERTa merupakan versi multibahasa dari *RoBERTa* yang telah dilatih pada 100 bahasa, termasuk bahasa Indonesia (Hidayatullah et al., 2023). Agar lebih spesifik dalam memahami sentimen berita keuangan dan pasar modal Indonesia, dilakukan *fine-tuning* pada model ini dengan menggunakan dataset *Financial PhraseBank* berbahasa Indonesia. Dengan pendekatan ini, diharapkan model dapat meningkatkan akurasi dalam mengenali terminologi serta pola bahasa yang digunakan dalam berita keuangan lokal.



Gambar 2.1 Arsitektur *XLM-RoBERTa*

Gambar 2.1. menunjukkan arsitektur *XLM-RoBERTa*, yang menggunakan *transformer encoder* untuk memproses teks. Berikut penjelasan komponen utama dalam model:

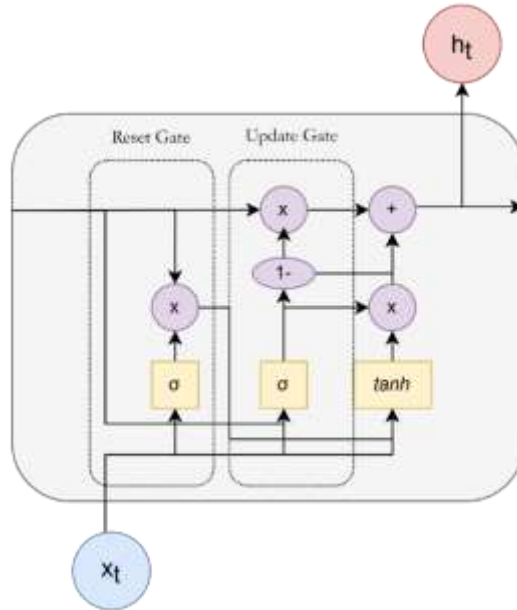
- a. [CLS]: Token spesial di awal input, digunakan untuk representasi keseluruhan kalimat.
- b. Token_{1-N}: Tokenisasi teks input menjadi *subword* unit.
- c. $E_{[cls]}$, E_{1-N} : Embedding dari masing-masing token setelah melalui lapisan embedding.
- d. C, T_{1-N} : Representasi akhir dari token setelah diproses oleh transformer. C merupakan keluaran dari token [CLS], sering digunakan untuk klasifikasi teks.
- e. Softmax: Fungsi aktivasi di lapisan output, mengubah nilai logits menjadi probabilitas untuk klasifikasi.

XLM-RoBERTa dirancang dengan arsitektur berbasis transformer dan terdiri atas sekitar 125 juta parameter, 12 lapisan, 768 dimensi *hidden states*, serta 8 *attention heads*. Model ini dapat menerima input hingga 512 token, di mana token pertama selalu berupa [CLS] yang berfungsi sebagai representasi keseluruhan teks. Untuk tugas klasifikasi sentimen, hidden state terakhir dari [CLS] kemudian diteruskan ke lapisan *softmax* guna memprediksi probabilitas kelas (Ranasinghe & Zampieri, 2020).

Dengan pretraining tambahan menggunakan dataset keuangan berbahasa Indonesia, model akan lebih akurat dalam memahami berita saham dan laporan ekonomi lokal.

2.7 *Gated Recurrent Unit*

Salah satu varian dari *Recurrent Neural Networks* (RNN) yang dirancang untuk mengatasi masalah *vanishing gradient* pada RNN, yang sering terjadi saat memproses urutan data yang panjang (Noh, 2021). Dibandingkan dengan *Long Short-Term Memory* (LSTM), *GRU* memiliki arsitektur yang lebih sederhana karena tidak menggunakan *output gate* (Abumohsen et al., 2023). Dengan hanya menggunakan *reset gate* dan *update gate*, *GRU* tetap mampu menangkap dependensi jangka panjang dalam data sekuensial sambil mempertahankan efisiensi komputasi yang lebih baik. Ilustrasi mengenai arsitektur *GRU* dapat dilihat pada gambar 2.1.



Gambar 2.2 GRU Arsitektur

Arsitektur *GRU* memiliki dua jenis "gating mechanisms" digunakan untuk mengontrol aliran informasi: *reset gate* (r_t) dan *update gate* (z_t). *Reset gate* menentukan seberapa banyak informasi dari status sebelumnya yang perlu dilupakan, sedangkan *update gate* mengatur proporsi informasi yang akan diperbarui dari status sebelumnya (Kopsiaftis et al., 2024). Rumus matematis *GRU* dapat dilihat pada persamaan (2.2), (2.3), (2.4), dan (2.5).

a. *Reset Gate*

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (2.2)$$

r_t adalah vektor *reset gate*, σ adalah fungsi sigmoid, W_r dan b_r adalah bobot dan bias reset gate, h_{t-1} adalah status tersembunyi sebelumnya, dan x_t adalah input saat ini.

b. *Update Gate*

$$z_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_z) \quad (2.3)$$

z_t adalah vektor *update gate*, dengan parameter W_r dan b_z .

c. *Candidate Hidden State*

$$\tilde{h}_t = \tanh(W_h \cdot [r_t \odot h_{t-1}, x_t] + b_h) \quad (2.4)$$

$\tilde{h}_t$ adalah kandidat status tersembunyi, $r_t \odot h_{t-1}$ merepresentasikan elemen *wise multiplication*, dan W_h , b_h adalah parameter terkait.

d. *Final Hidden State*

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (2.5)$$

h_t adalah status tersembunyi yang diperbarui, menggabungkan informasi dari status sebelumnya dan kandidat status berdasarkan kontrol dari *update gate*.

Keunggulan utama dari *GRU* adalah kesederhanaannya dibandingkan *LSTM*, karena tidak memiliki "*cell state*" tambahan. Hal ini mengurangi jumlah parameter, sehingga memungkinkan pelatihan yang lebih cepat dengan kinerja yang sebanding, terutama pada tugas yang tidak memerlukan pemrosesan dependensi sangat panjang.

2.8 Evaluasi Model

Evaluasi model merupakan langkah yang dilakukan untuk mengukur kinerja model. Dalam penelitian ini evaluasi dilakukan menggunakan akurasi, presisi, recall, F1-score untuk analisis sentimen dan RMSE, MAE, MAPE untuk *forecasting* saham.

2.8.1 Analisis Sentimen

Evaluasi model analisis sentimen dilakukan untuk mengukur kemampuan model dalam mengklasifikasikan teks ke dalam kategori tertentu, seperti positif, negatif, atau netral. Evaluasi ini menggunakan beberapa matrik utama, yaitu akurasi, presisi, recall, dan F1-score (Ullah et al., 2023). Akurasi mengukur persentase prediksi yang benar terhadap total data uji, dapat dilihat pada (2.6).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.6)$$

Presisi mengukur seberapa tepat model dalam memberikan prediksi positif, persamaannya ada pada (2.7).

$$Presisi = \frac{TP}{TP + FP} \quad (2.7)$$

Sementara itu, recall menghitung sejauh mana model mampu mendeteksi data positif dari seluruh data positif yang ada, dapat dilihat pada (2.8)

$$Recall = \frac{TP}{TP + FN} \quad (2.8)$$

F1-score adalah rata-rata harmonis antara presisi dan recall, yang memberikan evaluasi performa secara lebih seimbang, terutama pada dataset yang tidak seimbang, Formulanya dapat dilihat pada (2.9)

$$F1 - Score = 2 \times \frac{Presisi \times Recall}{Presisi + Recall} \quad (2.9)$$

2.8.2 Peramalan

Evaluasi peramalan saham berfokus pada pengukuran tingkat kesalahan antara nilai prediksi dan nilai aktual (Thakurdesai et al., 2024). Untuk peramalan biasa menggunakan tiga matrik evaluasi seperti, *Root Mean Squared Error (RMSE)*,

Mean Absolute Percentage Error (MAPE), dan *Mean Squared Error (MAE)* (Kervanci et al., 2024). *MAE* adalah rata-rata dari kesalahan absolut antara nilai aktual (y) dan prediksi ($\hat{y}$), dengan formula yang dapat dilihat pada (2.10)

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.10)$$

MAPE menghitung rata - rata persentase kesalahan relatif terhadap nilai aktual, dapat dilihat pada (2.11).

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left(\frac{|y_i - \hat{y}_i|}{y_i} \times 100 \right) \quad (2.11)$$

Sedangkan *RMSE* adalah akar dari rata - rata kuadrat selisih antara nilai aktual dan nilai prediksi. Metode ini memberikan penalti yang lebih besar terhadap kesalahan prediksi yang besar, ditunjukkan pada persamaan (2.12).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.12)$$

Kombinasi matrik evaluasi pada analisis sentimen dan *forecasting* memberikan dasar yang kuat untuk menilai performa masing-masing pendekatan. Evaluasi ini memastikan model tidak hanya akurat dalam klasifikasi sentimen, tetapi juga mampu menghasilkan prediksi harga saham dengan tingkat kesalahan yang dapat diterima.

2.9 Penelitian Terkait

Beberapa penelitian sebelumnya telah menggunakan berbagai pendekatan untuk memprediksi pasar saham, baik melalui analisis sentimen, analisis teknikal, maupun kombinasi keduanya. Pendekatan berbasis analisis sentimen

memanfaatkan data dari berita atau media sosial untuk menangkap opini dan emosi pasar yang dapat mempengaruhi harga saham. Sebaliknya, pendekatan berbasis analisis teknikal mengandalkan data historis saham, seperti harga, *volume* perdagangan, dan indikator teknikal, untuk mengidentifikasi pola dan tren masa lalu yang relevan dengan prediksi.

Penelitian yang mengintegrasikan analisis sentimen dengan analisis teknikal menunjukkan potensi yang lebih besar dalam meningkatkan akurasi prediksi pasar saham. Kombinasi ini memungkinkan pendekatan yang lebih holistik dengan memanfaatkan faktor internal dari data historis saham dan faktor eksternal seperti sentimen publik. Integrasi ini telah menjadi landasan penting dalam pengembangan metode prediksi yang lebih akurat dan relevan untuk memantau dinamika pasar saham. Tabel 2.1 berikut merangkum penelitian-penelitian terkait yang menjadi landasan untuk penelitian ini.

Tabel 2.1 Penelitian Terkait

Penulis	Algoritma	Pendekatan	Fokus Pembahasan
(John & Latha, 2023)	<i>HyRNN</i> + <i>VADER</i>	TA + SA	Algoritma <i>hybrid Recurrent Neural Network (HyRNN)</i> , yang menggabungkan <i>Bi-LSTM</i> , <i>GRU</i> , dan <i>sLSTM</i> , dirancang untuk memprediksi tren harga saham dengan mengintegrasikan sentimen berita dan fitur saham guna meningkatkan akurasi prediksi.

Penulis	Algoritma	Pendekatan	Fokus Pembahasan
(Sharaf et al., 2023)	<i>Stacked – LSTM + Naive Bayes, SVM</i>	TA + SA	Penggunaan algoritma <i>Stacked-LSTM</i> dalam analisis sentimen berita sosial dan data historis saham, hasil membuktikan kombinasi SA dan TA mengungguli akurasi dari pendekatan SA dan TA saja.
(B L & B R, 2023)	<i>Neural Network + Deep Belief Network + Self-Improved Whale Optimization Algorithm</i>	TA + SA	Memperkenalkan kerangka prediksi saham melalui analisis sentimen yang menggabungkan data saham dan sentimen berita, dengan menggunakan algoritma deep neural network dan optimasi whale untuk memprediksi harga saham dengan akurasi yang lebih tinggi.
(Haryono, Sarno, & Sungkono, 2024)	<i>TFGRU</i>	TA	Membandingkan kinerja beberapa arsitektur jaringan saraf dalam memprediksi harga saham, yaitu CNN, GRU, LSTM, dan GCN, dan menemukan bahwa arsitektur TFGRU merupakan yang terbaik pada 315 perusahaan di Bursa Efek Indonesia (IDX).
(Karanikola et al., 2023)	<i>RoBERTa</i>	SA	Membandingkan kinerja beberapa algoritma klasik dan

Penulis	Algoritma	Pendekatan	Fokus Pembahasan
			deep learning, termasuk <i>BERT</i> , <i>RoBERTa</i> , dan <i>FinBERT</i> , dalam analisis sentimen pada teks keuangan, dan menemukan bahwa <i>RoBERTa</i> memiliki performa paling baik.
(Lengkeek et al., 2023)	<i>RoBERTa</i>	SA	Model two-step yang menggunakan struktur hierarkis data dan <i>RoBERTa</i> sebagai model dasar dapat meningkatkan skor F1 sebesar 7,6% menjadi 0,70 dan mengurangi skor <i>MAE</i> sebesar 28% dan 33% pada set validasi silang dan set tes, masing-masing, sehingga menjadi lebih akurat dalam memprediksi kelas aspek keuangan.
(Venkatarao et al., 2023)	<i>LSTM – GRU</i>	TA	Mengimplementasikan model prediksi harga saham yang menggabungkan <i>LSTM</i> (<i>Long Short-Term Memory</i>) dan <i>GRU</i> (<i>Gated Recurrent Unit</i>) untuk memprediksi harga saham perusahaan di masa depan dengan lebih akurat.
(Chi Chen et al., 2023)	<i>GRU</i>	TA	Mengatasi masalah overfitting dalam prediksi harga saham dengan mengusulkan model

Penulis	Algoritma	Pendekatan	Fokus Pembahasan
			prediksi berbasis <i>Gated Recurrent Unit (GRU)</i> dengan dataset yang direkonstruksi, dan hasil eksperimen menunjukkan peningkatan yang signifikan dalam akurasi prediksi pada berbagai industri.
(Rahmadeyan & Mustakim, 2024)	<i>GRU, LSTM</i>	TA	Mengembangkan model prediksi harga saham Bank Rakyat Indonesia menggunakan <i>Long Short-Term Memory (LSTM)</i> dan <i>Gated Recurrent Unit (GRU)</i> , dan menemukan bahwa model <i>GRU</i> merupakan yang terbaik dengan nilai <i>MAE</i> 4958,9168, <i>RMSE</i> 70,4195, dan <i>MAPE</i> 1,1699%, yang memprediksi penurunan harga saham.
(Rahmadeyan & Mustakim, 2024)	<i>Hybrid EMD-LSTM</i>	TA	Metode hybrid Akima-EMD- <i>LSTM</i> yang menggabungkan teknik Empirical Mode Decomposition (EMD) dengan jaringan Long-Short Memory (<i>LSTM</i>) untuk memprediksi harga saham dengan kinerja yang lebih baik.
(Maqbool et al., 2022)	<i>MLP Regressor</i> + (<i>VADER</i> ,	TA + SA	Model machine learning yang menggabungkan data berita

Penulis	Algoritma	Pendekatan	Fokus Pembahasan
	<i>FLAIR, TextBlob</i>)		keuangan dengan data harga saham historis untuk memprediksi harga saham mendatang, dan menemukan bahwa model tersebut dapat mencapai akurasi tertinggi sebesar 0,90 untuk memprediksi tren dan tren masa depan selama 10 hari.
(Rizinski et al., 2024)	<i>Transformer & SHapley Additive exPlanations (SHAP)</i>	SA	Memperkenalkan metode baru bernama <i>eXplainable Lexicons (XLex)</i> yang menggabungkan kelebihan metode <i>lexicon-based</i> dan model transformer untuk melakukan analisis sentimen pada teks keuangan dengan lebih akurat dan efisien.
(Pan et al., 2023)	<i>MarIA, BETO</i>	SA	Melakukan analisis sentimen pada sektor ekonomi menggunakan data dari tweet dan berita keuangan, serta mengevaluasi kinerja berbagai <i>Large Language Model</i> Spanyol dalam tugas klasifikasi sentimen dan deteksi target.
(Md Salim Chowdhury et al., 2024)	<i>MLP, RNN, LSTM, CNN, ARIMA</i>	TA	Melakukan analisis komparatif terhadap berbagai model deep learning untuk <i>forecasting</i> pasar saham, serta

Penulis	Algoritma	Pendekatan	Fokus Pembahasan
			mengevaluasi performa model dalam mengenali pola dinamis dan tren nonlinier dalam data seri waktu keuangan.
(Sonkavde et al., 2023)	<i>Ensemble Model: Random Forest + XG-Boost + LSTM</i>	TA	Membandingkan penerapan algoritma <i>machine learning</i> dan <i>deep learning</i> di sektor keuangan, mencakup deskripsi model, kerangka generik untuk <i>forecasting</i> harga saham dan klasifikasi, serta analisis komparatif kinerja model ensemble dan model populer lainnya dalam <i>forecasting</i> harga saham.
(Srijiranon et al., 2022)	<i>PCA-EMD-LSTM</i>	TA+SA	Mengembangkan model prediksi hybrid <i>PCA-EMD-LSTM</i> untuk memprediksi harga saham di Thailand dengan menggabungkan analisis sentimen berita menggunakan <i>FinBERT</i> untuk meningkatkan kinerja model.
(Wu et al., 2022)	<i>S_I_LSTM</i>	TA + SA	Membangun model untuk prediksi harga saham yang menggabungkan multiple data sumber dan sentimen investor, menggunakan CNN untuk analisis sentimen dan jaringan

Penulis	Algoritma	Pendekatan	Fokus Pembahasan
			<i>LSTM</i> untuk prediksi harga saham di pasar A-share Shanghai, Cina.
(Koukaras et al., 2022)	<i>KNN, SVM LR, NB, DT, RF, MLP, VADER</i>	TA + SA	Mengembangkan model prediksi pergerakan saham menggunakan analisis sentimen pada data Twitter dan StockTwits, serta mengintegrasikan beberapa metode analisis sentimen dan mesin pembelajaran untuk meningkatkan akurasi prediksi saham.
(Cristescu et al., 2022)	<i>VADER, Linear Autoregression, Quadratic Autoregression, Cubic Autoregression, Nonlinear Autoregression</i>	TA + SA	Menggunakan analisis sentimen untuk meningkatkan akurasi prediksi harga saham dengan menggunakan model regresi, dan menunjukkan bahwa penggunaan skor sentimen sebagai faktor eksogen dapat meningkatkan kinerja model autoregressions dan meningkatkan koefisien R-square.
Penelitian ini	<i>GRU + RoBERTa</i>	TA + SA	<i>Financial RoBERTa</i> digunakan untuk menganalisis sentimen karena unggul dalam memahami konteks keuangan. GRU dipilih untuk peramalan

Penulis	Algoritma	Pendekatan	Fokus Pembahasan
			saham karena efektif dalam menangani data deret waktu. Kombinasi keduanya memungkinkan model menangkap pengaruh sentimen pasar terhadap pergerakan harga saham.

Berdasarkan temuan ini, penelitian yang akan dilakukan yaitu mengintegrasikan *RoBERTa* untuk analisis sentimen dan *GRU* untuk *forecasting* harga saham, karena integrasi kedua pendekatan berpotensi meningkatkan akurasi prediksi berdasarkan penelitian (B L & B R, 2023; John & Latha, 2023; Sharaf et al., 2023). *RoBERTa* dipilih karena kemampuannya dalam mengekstraksi sentimen keuangan secara akurat (Karanikola et al., 2023), sedangkan *GRU* dipilih karena efektivitasnya dalam memodelkan dependensi jangka panjang pada data historis saham (Haryono, Sarno, & Sungkono, 2024). Dengan pendekatan ini, penelitian ini bertujuan untuk menguji efektivitas kombinasi *RoBERTa* dan *GRU* dalam meningkatkan akurasi prediksi saham BBRI, BBCA, dan BMRI.

Tabel 2.2 Matriks Penelitian

Penulis	Algoritma	Pendekatan	
		Technical Analysis	Sentiment Analysis
(John & Latha, 2023)	<i>HyRNN + VADER</i>	√	√
(Sharaf et al., 2023)	<i>Stacked – LSTM + Naive Bayes, SVM</i>	√	√

Penulis	Algoritma	Pendekatan	
		Technical Analysis	Sentiment Analysis
(B L & B R, 2023)	<i>Neural Network + Deep Belief Network + Self-Improved Whale Optimization Algorithm</i>	√	√
(Haryono, Sarno, & Sungkono, 2024)	<i>TFGRU</i>	√	-
(Karanikola et al., 2023)	<i>RoBERTa</i>	-	√
(Lengkeek et al., 2023)	<i>RoBERTa</i>	-	√
(Venkatarao et al., 2023)	<i>LSTM - GRU</i>	√	-
(Chi Chen et al., 2023)	<i>GRU</i>	√	-
(Rahmadeyan & Mustakim, 2024)	<i>GRU, LSTM</i>	√	-
(Rahmadeyan & Mustakim, 2024)	<i>Hybrid EMD-LSTM</i>	√	-
(Maqbool et al., 2022)	<i>MLP Regressor + (VADER, FLAIR, TextBlob)</i>	√	√
(Rizinski et al., 2024)	<i>Transformer & SHapley Additive exPlanations (SHAP)</i>	-	√
(Pan et al., 2023)	<i>MarIA, BETO</i>	-	√
(Md Salim Chowdhury et al., 2024)	<i>MLP, RNN, LSTM, CNN, ARIMA</i>	-	√
(Sonkavde et al., 2023)	<i>Ensemble Model: Random Forest + XG-Boost + LSTM</i>	√	-
(Srijiranon et al., 2022)	<i>PCA-EMD-LSTM</i>	√	-

Penulis	Algoritma	Pendekatan	
		Technical Analysis	Sentiment Analysis
(Wu et al., 2022)	<i>S_I_LSTM</i>	√	√
(Koukaras et al., 2022)	<i>KNN, SVM LR, NB, DT, RF, MLP, VADER</i>	√	√
(Cristescu et al., 2022)	<i>VADER, Linear Autoregression, Quadratic Autoregression, Cubic Autoregression, Nonlinear Autoregression</i>	√	√
Penelitian ini	<i>GRU + RoBERTa</i>	√	√