

PREDIKSI HARGA SAHAM DENGAN *XGBOOST* DAN SENTIMEN *TWITTER*

Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Terapan
Pada Program Pendidikan Teknik Infomatika



Oleh:

ISNAENTI NUR LATIFAH

1214033

**PROGRAM D-IV TEKNIK INFORMATIKA
FAKULTAS SEKOLAH VOKASI
UNIVERSITAS LOGISTIK DAN BISNIS INTERNASIONAL
BANDUNG
2025**

STOCK PRICE PREDICTION WITH XGBOOST AND TWITTER SENTIMENT

*This Report to Partial Fulfillment of the Requirements for the Applied Bachelor's
Degree in the Informatics Engineering Education Program*



Create by,

ISNAENTI NUR LATIFAH

1214033

***APPLIED BACHELOR PROGRAM OF INFORMATICS
ENGINEERING VOCATIONAL SCHOOL
UNIVERSITY LOGISTICS AND INTERNATIONAL BUSINESS
BANDUNG
2025***

HALAMAN PERNYATAAN ORISINILITAS

Yang bertanda tangan dibawah ini:

Nama : Isnaenti Nur Latifah

NPM : 1214033

Program Studi : Diploma IV-Teknik Informatika

Perguruan Tinggi : Universitas Logistik dan Bisnis Internasional

Dengan ini menyatakan bahwa tugas akhir yang telah saya buat dengan judul:

”Prediksi Harga Saham dengan *XGBoost* dan Sentimen *Twitter*” adalah asli

(orisinil) dan belum pernah diterbitkan/dipublikasikan dimanapun dan dalam bentuk

apapun. Demikianlah surat pernyataan ini saya buat dengan sebenar-benarnya tanpa

ada paksaan dari pihak manapun juga. Apabila dikemudian hari ternyata saya

memberikan keterangan palsu dan atau ada pihak lain yang mengklaim bahwa tugas

akhir yang telah saya buat adalah karya milik seseorang atau badan tertentu, saya

bersedia diproses baik secara pidana maupun perdata dan kelulusan saya dari

Universitas Logistik dan Bisnis Internasional dicabutkan/dibatalkan.

Dibuat di : Bandung

Pada tanggal:

Yang menyatakan,

Isnaenti Nur Latifah

LEMBAR PENGESAHAN

PREDIKSI HARGA SAHAM DENGAN *XGBOOST* DAN SENTIMEN *TWITTER*

STOCK PRICE PREDICTION WITH XGBOOST AND TWITTER SENTIMENT

1214033

ISNAENTI NUR LATIFAH

Laporan Program Tugas Akhir ini telah diperiksa, disetujui dan disidangkan di
Bandung, (tanggal sidang)

Oleh:

Pembimbing Utama

Pembimbing Pendamping

Nisa Hanum Harani, S.Kom.,M.T.,CDSP
NIK. 117.89.223

Cahyo Prianto, S.Pd., M.T.,CDSP
NIK. 117.84.222

Menyetujui,
Koordinator Tugas Akhir

M. Yusril Helmi Setyawan, S.Kom.,M.Kom.,SFPC
NIK. 113.74.163

Mengetahui,
Ketua Program Studi D-IV Teknik Informatika

Roni Andarsyah, ST., M.KOM.,SFPC
NIK. 115.88.193

KATA PENGANTAR

Puji dan syukur ke hadirat Tuhan yang Maha Esa, karena atas berkat dan rahmat-Nya, penulis dapat mengerjakan dan menyelesaikan laporan analisis yang berjudul *“Prediksi Harga Saham dengan XGBoost dan Sentimen Twitter”*. Penulis menggunakan bahasa yang sederhana dan mudah untuk dimengerti bagi para pembaca. Adapun Penulisan buku ini dibuat dalam rangka memenuhi salah satu syarat kelulusan Mata kuliah Tugas Akhir di Universitas Logistik dan Bisnis Internasional.

Akhirnya dengan segala kerendahan hati, perkenankanlah penulis mengucapkan terima kasih yang tidak terhingga atas segala bantuan dan bimbingan yang telah diberikan dalam penulisan laporan ini, kepada yang terhormat :

1. Kedua orang tua tercinta, yang telah memberikan doa dan dukungan baik secara moril maupun materil.
2. Roni Andarsyah,. ST., M.KOM.,SFPC. selaku Ketua Proram Studi D4 Teknik Informatika.
3. M. Yusril Helmi Setyawan, S.Kom.,M.Kom.,SFPC, selaku Koordinator Tugas Akhir D4 Teknik Informatika.
4. Nisa Hanum Harani, S.Kom.,M.T.,CDSP,SFPC, selaku dosen pembimbing 1.
5. Cahyo Prianto, S.Pd., M.T.,CDSP, SFPC,selaku dosen pembimbing 2

Penulis berusaha dengan sebaik-baiknya dalam Penulisan laporan ini, namun dalam banyak hal, penulis menyadari bahwa laporan ini masih jauh dari kata sempurna. Maka dari itu, saran dan kritik yang bersifat membangun dari semua pihak sangat penulis harapkan, agar laporan ini menjadi lebih baik dan bermanfaat.

Bandung, 2025

Isnaenti Nur Latifah

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Logistik dan Bisnis Internasional (ULBI), saya yang bertanda tangan dibawah ini:

Nama : Isnaenti Nur Latifah
NPM : 1214033
Program Studi : Diploma IV Teknik Informatika
Jenis Karya : Tugas Akhir

Demi pengembangan ilmu pengetahuan, memberikan kepada ULBI **Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul:

Prediksi Harga Saham dengan *XGBoost* dan Sentimen *Twitter*

beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non eksklusif ini Universitas Logistik dan Bisnis internasional berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pengkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta

Demikianlah pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Bandung
Pada tanggal :

Yang menyatakan,

(Isnaenti Nur Latifah.)

ABSTRAK

Prediksi harga saham yang akurat menjadi kebutuhan penting dalam pengambilan keputusan investasi, terutama di tengah meningkatnya partisipasi investor. Sebagian besar model prediksi masih bergantung pada data historis semata tanpa mempertimbangkan faktor eksternal seperti opini publik yang tersebar melalui media sosial. Penelitian ini mengusulkan pendekatan prediksi harga saham dengan menggabungkan data historis dan sentimen dari Twitter guna meningkatkan akurasi dan relevansi model. Tiga saham dari Bursa Efek Indonesia digunakan sebagai objek kajian, yaitu ISAT.JK, TINS.JK, dan MDKA.JK. Data tweet yang berkaitan dengan masing-masing saham diproses melalui tahapan pembersihan teks, normalisasi, penerjemahan otomatis ke Bahasa Inggris, serta klasifikasi sentimen menggunakan metode *VADER*. Data sentimen kemudian digabungkan dengan data historis saham untuk membentuk fitur prediktif yang digunakan dalam model *XGBoost* dan *LSTM*. Evaluasi dilakukan menggunakan metrik *MSE*, *RMSE*, *MAE*, dan R^2 . Hasil menunjukkan bahwa *XGBoost* memberikan performa terbaik dengan nilai R^2 sebesar 0,93 pada saham MDKA.JK dan validasi stabil sebesar 0,86. Selain itu, ditemukan bahwa lonjakan tweet negatif sering kali bertepatan atau mendahului penurunan harga saham. Dengan mempertimbangkan hasil evaluasi dan kestabilan prediksi, model *XGBoost* dinilai lebih unggul untuk diterapkan dalam sistem prediksi harga saham berbasis data historis dan sentimen media sosial.

Kata Kunci: Prediksi Harga Saham, Sentimen Twitter, *XGBoost*

ABSTRACT

Accurate stock price prediction is essential for supporting investment decision-making, especially amid the growing participation of investors. Most predictive models still rely solely on historical data without considering external factors such as public sentiment expressed through social media. This study proposes a stock price prediction approach by integrating historical data with sentiment information from Twitter to enhance model accuracy and relevance. Three stocks listed on the Indonesia Stock Exchange are used as the study objects: ISAT.JK, TINS.JK, and MDKA.JK. Tweets related to each stock undergo a series of preprocessing steps, including text cleaning, normalization, automatic translation into English, and sentiment classification using the VADER method. The sentiment data is then merged with historical stock prices to form predictive features used by XGBoost and LSTM models. Evaluation is carried out using MSE, RMSE, MAE, and R^2 metrics. The results show that XGBoost delivers the best performance, achieving an R^2 value of 0.93 on MDKA.JK with a stable validation score of 0.86. Furthermore, spikes in negative tweets frequently coincide with or precede stock price declines. Considering the evaluation results and prediction consistency, XGBoost is deemed more suitable for implementation in stock price prediction systems based on historical and social media sentiment data.

Keyword: *Stock Price Prediction, Sentiment Twitter, XGBoost*

DAFTAR ISI

HALAMAN PERNYATAAN ORISINILITAS	i
LEMBAR PENGESAHAN	ii
KATA PENGANTAR	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	iv
ABSTRAK.....	v
ABSTRACT.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	ix
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN.....	xi
BAB I <u>PENDAHULUAN</u>	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah.....	4
1.3 Tujuan dan Manfaat	4
1.4 Ruang Lingkup.....	5
BAB II <u>LANDASAN TEORI</u>	6
3.1 Deskripsi Topik yang Sama	6
3.1.1 Saham.....	6
3.1.2 Keputusan Investasi	6
3.1.3 Analisis Sentimen	7
3.1.4 Twitter Api.....	7
3.1.5 Prediksi saham	7
3.1.6 Machine Learning	8
3.1.7 XGBOOST.....	8
3.1.8 <i>VADER (Valance Aware Dictionary Sentiment Reasioner)</i>	10
3.1.9 <i>LSTM (LongShort-Term Memory)</i>	11
3.2 Deskripsi Metode yang Sama	13
3.2.1 <i>State of the art (SOTA)</i>	13
BAB III <u>METODOLOGI PENELITIAN</u>	19
3.1 Metodologi Penelitian.....	19
3.2 Tahapan Tahapan Diagram Metodologi Penelitian.....	20

BAB IV EKSPERIMEN DAN HASIL.....	23
4.1 Eksperimen	23
4.1.1 Pengumpulan Data	23
4.1.2 <i>Preprocessing</i> Data.....	26
4.1.3 Perancangan Model.....	43
4.1.4 Implementasi Model	43
4.1.5 Evaluasi Model	44
4.2 Hasil	45
4.2.1 Keterkaitan Sentimen Negatif dan Harga Saham.....	46
4.2.2 Evaluasi Hasil Model XGBoost.....	48
4.2.3 Evaluasi Hasil Model LSTM.....	54
4.2.4 Perbandingan Kinerja Model	59
4.2.5 Pemilihan Model untuk Implementasi <i>Dashboard</i>	61
BAB V PENUTUP.....	64
5.1 Kesimpulan	64
5.2 Saran	64
DAFTAR PUSTAKA	lxvi
LAMPIRAN.....	lxxiii

DAFTAR TABEL

Tabel 2. 1 Vader skor[40]	10
Tabel 4. 1 Data harga close & tweet negatif ISAT	46
Tabel 4. 2 Data harga close & tweet negatif MDKA	47
Tabel 4. 3 Data harga close & tweet negatif TINS	48
Tabel 4. 4 Hasil evaluasi model XGBoost pada saham ISAT	48
Tabel 4. 5 Hasil evaluasi model XGBoost pada saham TINS	50
Tabel 4. 6 Hasil evaluasi model XGBoost pada saham MDKA	52
Tabel 4. 7 Evaluasi model LSTM dari ketiga saham	54

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi XGBoost.....	9
Gambar 2. 2 Pohon regresi XGBOOST.....	10
Gambar 3. 1 Diagram alur tahapan penelitian[14].....	20
Gambar 4. 1 Hasil data info, describe dan cek nilai kosong	30
Gambar 4. 2 Distribusi jumlah data per saham (ticker)	31
Gambar 4. 3 Tabel Agresi Bulanan Saham dan Sentimen	32
Gambar 4. 4 Heatmap korelasi antar fitur numerik.....	33
Gambar 4. 5 Boxplot deteksi outlier pada fitur harga dan sentiment.....	34
Gambar 4. 6 Distribusi jumlah tweet per saham	35
Gambar 4. 7 Visualisasi harga penutupan dan jumlah tweet terhadap waktu.....	36
Gambar 4. 8 Boxplot deteksi outlier setelah penanganan	38
Gambar 4. 9 Grafik ISAT.JK harga close vs tweet negatif.....	46
Gambar 4. 10 Grafik harga close vs negatif tweet MDKA.JK	47
Gambar 4. 11 Grafik harga close vs negatif tweet TINS.JK.....	47
Gambar 4. 12 Scaratterplot saham ISAT	49
Gambar 4. 13 Grafik prediksi saham ISAT	50
Gambar 4. 14 Scaratterplot saham TINS	51
Gambar 4. 15 Grafik prediksi saham TINS	52
Gambar 4. 16 Scaratterplot saham MDKA.....	53
Gambar 4. 17 Grafik prediksi saham MDKA	54
Gambar 4. 18 Loss Curva model LSTM di saham ISAT.....	55
Gambar 4. 19 Grafik prediksi model LSTM di saham ISAT.....	56
Gambar 4. 20 Loss Curva model LSTM di saham TINS.....	57
Gambar 4. 21 Grafik prediksi model LSTM di saham TINS.....	57
Gambar 4. 22 Loss Curva model LSTM di saham MDKA	58
Gambar 4. 23 Grafik prediksi model LSTM di saham MDKA	59
Gambar 4. 24 Dashboard Prediksi Harga Saham.....	62
Gambar 4. 25 Tampilan dashboard hasil prediksi.....	63

DAFTAR LAMPIRAN

Lampiran 1 Form Bimbingan Pembimbing 1	lxxiii
Lampiran 2 Form Bimbingan Pembimbing 2	lxxv
Lampiran 3 Hasil Turnitin	lxxvi

BAB I

PENDAHULUAN

1.1 Latar Belakang

Era digital ini, mempunyai berbagai pilihan dalam berinvestasi yang menjadi alternatif untuk menentukan pembagian dengan sumber daya yang dimiliki berguna bagi pengembangan di masa yang akan datang(Firdaus and Ifrochah 2022). Investasi adalah metode untuk menumbuhkan dana yang dimiliki guna memperoleh keuntungan baik dalam waktu dekat maupun di masa mendatang. Keuntungan yang diperoleh dari penempatan dana tersebut dikenal dengan istilah imbal hasil atau return(Saputro and Swanjaya 2023). Investasi memiliki berbagai macam jenis yaitu ada tabungan, deposito, reksadana, obligasi, saham, emas dan juga properti(Apriliansah 2024). Pasar modal di Indonesia, atau yang dikenal sebagai Bursa Efek Indonesia (BEI), tengah mengalami pertumbuhan yang turut memberikan kontribusi terhadap kemajuan perekonomian nasional(Muhidin and Situngkir 2022).

Melalui catatan BEI, jumlah investor saham Indonesia didominasi generasi Z dan milenial dengan investor umur dibawah 30 tahun sebesar 55,38 persen berdasarkan data per Juni 2024(Puspadini 2024). Banyak investor menginvestasikan uang mereka di pasar saham untuk pengembalian tinggi. Saham adalah instrumen keuangan yang menjadi bukti kepemilikan atau partisipasi investor di pasar modal dalam suatu perusahaan. Nilai saham bersifat fluktuatif, artinya dapat mengalami kenaikan maupun penurunan(Putra, Ramadhani, and Burhanuddin 2022). Perusahaan dengan harga saham yang rendah cenderung dipersepsikan memiliki kinerja yang kurang baik. Sebaliknya, harga saham yang tinggi sering kali mencerminkan kinerja perusahaan yang positif. Namun demikian, harga saham yang terlalu tinggi bisa menjadi kendala bagi investor dengan modal terbatas untuk berinvestasi(Maulida and Mahardhika 2021). Menurut berita CNN terdapat beberapa daftar saham yang akan naik di sektor tambang dan telekomunikasi yaitu saham Indosat, Merdeka Copper Gold dan saham timah TINS yang disampaikan oleh Oktavianus dan Herditya yang telah menganalisis teknikal saham(Fitriyani 2025).

Keuntungan dari investasi saham diperoleh melalui selisih antara harga beli dan harga jual saham yang dikenal sebagai capital gain. Setiap investor tentu

mengharapkan return yang positif dari saham yang dimilikinya. Untuk mencapai hasil investasi yang optimal, diperlukan strategi yang selaras dengan tujuan keuangan, tingkat toleransi risiko, serta kondisi finansial secara menyeluruh (Chang Hartono and Dwiyo Widianoro 2024). Apabila investor tidak memahami risiko yang akan dihadapi, maka investor tersebut belum bisa manajemen risiko yang akan datang. Salah satu risiko terburuk dari aktifitas investasi apabila tidak bisa manajemen risikonya yaitu kerugian (Susila 2023). Terkait hal tersebut, terdapat berbagai faktor yang dapat memengaruhi fluktuasi harga saham, baik yang berasal dari dalam perusahaan (faktor internal) maupun dari luar perusahaan (faktor eksternal) (Fitri and Riana 2022). Beberapa faktor yang dapat memengaruhi harga saham antara lain adalah aksi korporasi (seperti akuisisi, merger, penerbitan saham baru, dan divestasi), kinerja keuangan perusahaan (misalnya utang dividen, rasio utang, laba per saham, dan tingkat profitabilitas), kebijakan pemerintah (meliputi kebijakan ekspor-impor, investasi asing, dan utang luar negeri), nilai tukar mata uang, kondisi makroekonomi (seperti inflasi, suku bunga, dan tingkat pengangguran), sentimen serta rumor pasar, stabilitas politik dan ekonomi, hingga praktik manipulasi pasar (Marjuni 2022).

Penggunaan informasi media sosial saat ini memiliki manfaat yang beragam, seperti halnya penggunaan untuk mendapatkan pengetahuan atau informasi dalam saham. Investor ritel saat ini dapat menggunakan informasi di berbagai media sosial bertujuan untuk mencapai keuntungan atau hasil yang positif dalam saham yang mereka investasikan (Christophorus Bintang Saputra and Koesrindartoto 2024). Faktor dari persepsi publik saat ini dapat mempengaruhi harga saham tersebut melalui sentimen yang diungkapkan di media sosial, seperti platform X (twitter) (Taofiqurrohman and Roji 2025). Twitter menjadi salah satu media sosial yang mampu menyebarkan informasi secara cepat dan luas memperoleh popularitas sejak diciptakan pada bulan Maret 2006 (Zakariyya 2020).

Maka dari itu untuk mempermudah mengetahui naik turunnya harga saham diperlukannya prediksi pergerakan harga saham. Dari banyak penelitian yang melakukan experiment tentang memprediksi harga saham saat ini banyak yang menggunakan teknologi machine learning maupun deep learning untuk mengatasi hal tersebut. Seperti pada penelitian oleh Qianyi Xiao and Baha Ihnaini mengusulkan model prediksi tren saham dengan menggabungkan analisis sentimen dari Twitter dan

berita keuangan. Sentimen tweet dianalisis menggunakan algoritma VADER, sedangkan berita dianalisis menggunakan FinBERT. Hasil sentimen kemudian dihitung dalam dua rentang waktu berbeda (jam alami dan jam bursa) dan digunakan sebagai fitur dalam model prediksi. Untuk prediksi tren saham keesokan harinya, enam algoritma machine learning diuji, dan algoritma *Naïve Bayes* terbukti memberikan hasil terbaik dalam sebagian besar kasus (Xiao and Ihnaini 2023).

Dan pada penelitian oleh marah lisanne dan jan farchmin menggabungkan analisis sentimen dari Twitter dengan model prediksi harga saham menggunakan algoritma Long Short-Term Memory (LSTM). Sentimen tweet dianalisis menggunakan TextBlob untuk mengklasifikasikan tweet menjadi positif, negatif, atau netral. Hasil sentimen ini kemudian digunakan sebagai fitur tambahan dalam model *LSTM* untuk meningkatkan akurasi prediksi harga saham. Pendekatan ini menunjukkan bahwa penggabungan data pasar dengan sentimen media sosial dapat meningkatkan kemampuan model dalam memprediksi arah pergerakan harga saham (Thormann et al. 2021).

Penelitian oleh Yifan Zhang menunjukkan bahwa algoritma XGBoost efektif digunakan dalam memprediksi harga saham berbasis data historis harian. Dalam studinya, menggabungkan data time-series dengan indikator teknikal seperti EMA, SMA, RSI, dan MACD untuk membentuk fitur prediksi. Dengan tuning parameter menggunakan Grid SearchCV dan evaluasi menggunakan Mean Squared Error (MSE), model mampu menghasilkan prediksi yang akurat dengan MSE rendah sebesar 0,19. Penelitian ini menjadi dasar penting bahwa XGBoost sangat cocok untuk menangkap pola tren jangka pendek dan fluktuasi harga saham yang cepat (Zhang 2023).

Berdasarkan dari penelitian terdahulu tersebut, penelitian ini akan membahas bagaimana sentimen dari Twitter dapat berpengaruh terhadap prediksi harga saham. Salah satu permasalahan yang sering muncul adalah prediksi harga saham yang hanya mengandalkan data historis sering kali belum mampu menangkap pengaruh faktor eksternal seperti opini publik atau sentimen pasar yang tersebar di media sosial. Untuk mengatasi hal ini, model XGBoost digunakan karena mampu memahami pola data dalam urutan waktu yang kompleks. Sementara itu, untuk mengetahui sentimen dari tweet berbahasa Indonesia, digunakan metode Vader yang dikombinasikan dengan deep-translator agar dapat menenjemahkan tweet terlebih dahulu ke Bahasa Inggris

sebelum diklasifikasikan, sehingga kalimat dapat di evaluasi dengan tepat berdasarkan lexicon-based. Dengan menggabungkan kedua pendekatan ini, diharapkan hasil prediksi harga saham menjadi lebih akurat dan relevan dengan kondisi nyata.

1.2 Identifikasi Masalah

Dalam pembuatan laporan analisis dan prediksi kenaikan penurunan harga saham berikut masalah-masalah yang perlu diselesaikan:

1. Apakah sentimen negatif dari media sosial, khususnya Twitter berpengaruh terhadap pergerakan harga saham?
2. Bagaimana mengimplementasikan data sentimen dari twitter ke dalam model prediksi pergerakan harga saham?
3. Model prediksi apa yang efektif dalam data gabungan sentiment dan historis untuk memprediksi harga saham?

1.3 Tujuan dan Manfaat

a. Tujuan

Tujuan dari laporan ini yaitu:

1. Menganalisis pengaruh sentimen negatif dari media sosial, khususnya Twitter terhadap pergerakan harga saham.
2. Mengembangkan model prediksi harga saham yang mengintegrasikan data sentimen dari Twitter dengan data historis saham.
3. Untuk mengevaluasi dan membandingkan efektivitas model prediksi, khususnya antara XGBoost dan LSTM, dalam memanfaatkan data gabungan (historis dan sentimen) dalam memprediksi harga saham.

b. Manfaat

Manfaat yang diperoleh dari pembuatan laporan ini sebagai berikut:

1. Memberikan pemahaman baru mengenai pengaruh sentimen negatif media sosial terhadap pergerakan harga saham, sehingga dapat menjadi referensi tambahan dalam analisis pasar.
2. Mendukung pengembangan model prediksi harga saham yang lebih akurat dengan menggabungkan data historis dan analisis sentimen dari Twitter, khususnya dalam konteks bahasa Indonesia.
3. Memberikan informasi komparatif terkait performa model prediksi (XGBoost dan LSTM), sehingga dapat membantu pengguna atau pengembang sistem

dalam memilih algoritma yang lebih akurat dan stabil untuk pengambilan keputusan investasi atau analisis saham secara digital.

1.4 Ruang Lingkup

Untuk membatasi fokus dan cakupan pembahasan dalam penelitian ini, maka ruang lingkup penelitian ditentukan agar lebih terarah dan sesuai dengan tujuan yang ingin dicapai. Adapun ruang lingkup penelitian ini meliputi:

1. Data yang digunakan berupa harga saham dan tweet berbahasa Indonesia dalam periode tertentu.
2. Model prediksi yang digunakan adalah XGBOOST, dengan analisis sentimen dilakukan menggunakan VADER.
3. Penelitian difokuskan pada pengaruh sentimen negatif Twitter terhadap pergerakan harga saham, tanpa mempertimbangkan faktor eksternal lainnya.

BAB II

LANDASAN TEORI

3.1 Deskripsi Topik yang Sama

3.1.1 Saham

Saham merupakan instrumen keuangan yang menjadi tanda kepemilikan atau keterlibatan dalam sebuah perusahaan, terutama pada perusahaan terbuka (Hidayana and Ruchjana 2023). Melalui kepemilikan saham tersebut, investor berhak atas bagian dari pendapatan perusahaan, aset yang dimiliki perusahaan, serta memiliki hak untuk ikut serta dalam Rapat Umum Pemegang Saham (RUPS) (Putri and Yustisia 2021).

Harga saham adalah nilai per lembar yang diperdagangkan di pasar modal. Nilai ini sangat krusial dan menjadi perhatian utama bagi investor saat melakukan investasi (Fadila and Nuswandari 2022), karena mencerminkan kinerja suatu emiten. Dalam perdagangan di pasar modal, harga saham diklasifikasikan menjadi tiga jenis: harga tertinggi (high price), harga terendah (low price), dan harga penutup (close price). Harga tertinggi dan terendah menunjukkan nilai tertinggi dan terendah dalam satu hari perdagangan, sedangkan harga penutup adalah harga transaksi terakhir yang terjadi saat penutupan sesi bursa (Indarso and Pangaribuan 2021).

3.1.2 Keputusan Investasi

Keputusan investasi dapat diartikan sebagai proses memilih alternatif penanaman modal yang dianggap paling menguntungkan, dengan tujuan memperoleh keuntungan di masa depan. Istilah ini juga merujuk pada bagaimana seseorang mengalokasikan dana yang dimilikinya ke berbagai bentuk investasi yang diperkirakan akan memberikan imbal hasil di kemudian hari (Pratama, Wijayanto, and Purbawati 2022).

Keputusan investasi merupakan tindakan yang diambil oleh individu atau manajer investasi dalam menanamkan modal pada suatu aset dengan tujuan memperoleh keuntungan di masa mendatang. Keputusan ini mencakup

langkah-langkah seperti membeli, menahan, atau menjual saham(Widyawati and Imronudin 2024).

3.1.3 Analisis Sentimen

Analisis sentimen merupakan studi komputasional yang bertujuan memahami opini, emosi, dan perasaan yang terkandung dalam suatu teks. Teknik ini telah banyak digunakan dalam analisis media sosial untuk keperluan di bidang pemasaran, sosial, maupun politik(Permata Aulia, Arifin, and Mayasari 2021). Tujuan dari analisis sentimen adalah untuk mengevaluasi pandangan dan kecenderungan suatu opini terhadap topik tertentu, apakah bersifat positif atau negatif(Arsi and Waluyo 2021).

3.1.4 Twitter Api

Twitter merupakan salah satu media sosial populer di kalangan pengguna internet karena tampilannya yang sederhana dan kemudahan dalam penggunaannya. Selain itu, platform ini memungkinkan pengguna untuk secara bebas menyampaikan pendapat serta opini mereka.(Azahri, Sulistiyowati, and Jajuli 2023).

Twitter merupakan platform yang dirancang untuk memfasilitasi komunikasi cepat antara teman, keluarga, maupun rekan kerja melalui pesan singkat. Kehadiran media sosial saat ini memudahkan pengguna dalam menyampaikan pandangan atau membagikan pengalaman pribadi. Oleh karena itu, berbagai data yang terdapat di dalam platform media sosial tersebut menjadi sumber informasi yang kaya(Yahyadi and Latifah 2022).

3.1.5 Prediksi saham

Prediksi harga saham merupakan isu krusial dalam dunia perdagangan saham. Harga saham yang berfluktuasi tergantung pada waktu dan informasi historis digunakan sebagai dasar untuk memperkirakan nilai saham suatu perusahaan atau instrumen keuangan lainnya di masa depan yang diperdagangkan di pasar modal(Himawan et al. 2022).

Prediksi saham menjadi hal yang penting agar investor dapat memperkirakan harga saham di masa depan dan menghindari potensi kerugian yang besar. Tujuan utama dari prediksi ini adalah untuk memberikan gambaran

kepada investor dan pemegang saham terkait arah pergerakan aset. Meski demikian, proses prediksi sering kali tidak mudah karena adanya faktor eksternal yang menyebabkan fluktuasi harga saham, sehingga tingkat volatilitasnya menjadi tinggi (Sari et al. 2023).

3.1.6 Machine Learning

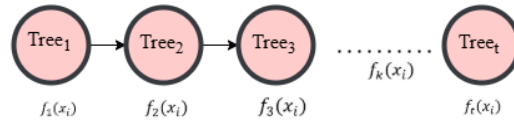
Machine Learning (ML) atau pembelajaran mesin merupakan bagian dari kecerdasan buatan (*AI*) yang berfokus pada kemampuan sistem untuk belajar dari data, yakni membangun sistem yang dapat belajar secara otomatis tanpa perlu pemrograman berulang oleh manusia (Alfarizi et al. 2023).

Dalam konteks peninjauan sistematis, penggunaan machine learning dapat meminimalkan pekerjaan manual yang biasanya diperlukan. Dengan memanfaatkan algoritma yang mampu mengenali pola dan hubungan kompleks dalam data, machine learning menawarkan pendekatan yang lebih fleksibel dibanding metode berbasis aturan, sehingga memungkinkan pengambilan keputusan dengan tingkat akurasi yang lebih tinggi (Nurhalizah, Ardianto, and Purwono 2024).

3.1.7 XGBOOST

XGBoost (eXtreme Gradient Boosting) regression merupakan versi lanjutan dari algoritma gradient boosting yang dirancang dengan struktur pohon regresi yang lebih teratur, sehingga menghasilkan kinerja yang lebih baik serta mengurangi risiko overfitting dengan menyederhanakan kompleksitas model (Chairunisa et al. 2024). Dalam pendekatannya, algoritma *Gradient Boosting* menggabungkan sejumlah model lemah (*weak learner*) biasanya berbentuk *Decision Tree* untuk membentuk model prediktif yang kuat. *XGBoost* memiliki sejumlah keunggulan, seperti adanya fitur regularisasi, pemangkasan pohon (*tree pruning*), kemampuan pemrosesan paralel, serta dukungan terhadap data numerik dan kategorikal (Murdiansyah 2024).

Metode ini membangun beberapa pohon keputusan secara berurutan, di mana setiap pohon selanjutnya dirancang untuk memperbaiki kesalahan dari pohon sebelumnya (Kurnia et al. 2023).



Gambar 2. 1 Ilustrasi *XGBoost*

Dari gambar 2.1 ditunjukkan bahwa $f_t(x_i)$ merepresentasikan model pohon, di mana nilai prediksi pada waktu ke- t disimbolkan sebagai $\hat{y}_i^{(t)}$ yang dinyatakan melalui persamaan (Nugraha and Irawan 2023)

$$\hat{y}_i^{(y)} = \sum_{k=1}^t f_k(x_i) \quad (1)$$

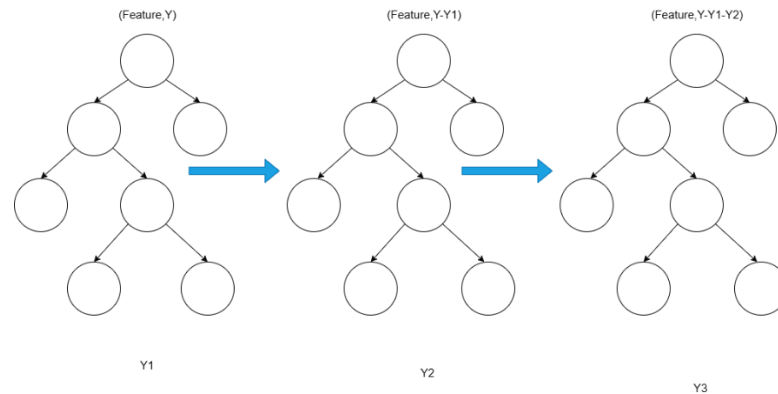
Dimana $\hat{y}_i^{(t)}$ merupakan hasil prediksi dari model pohon pada iterasi ke- t , sedangkan $\hat{y}_i^{(t-1)}$ adalah prediksi dari model pohon pada iterasi sebelumnya. Adapun $f_t(x_i)$ merepresentasikan model baru yang dibangun berdasarkan pohon dasar (base tree models). Dalam algoritma XGBoost, penentuan jumlah pohon (tree) dan kedalamannya (depth) sangatlah krusial. Tantangan utama dalam pencarian model yang optimal adalah meminimalkan fungsi kerugian (loss function), yang dirumuskan dalam bentuk fungsi objektif (Obj) seperti ditunjukkan pada persamaan berikut.

$$Obj(t) = \sum_{i=1}^t l(\hat{y}_i^t, y_i) + \sum_{i=1}^t \Omega(f_i) \quad (2)$$

Dimana $\hat{y}_i^{(t)}$ adalah nilai prediksi pada iterasi ke- t , dan y_i merupakan nilai actual. Simbol l menyatakan fungsi loss yang mengukur selisih antara prediksi dan nilai sebenarnya, sedangkan Ω adalah fungsi regularisasi yang berfungsi untuk mengontrol kompleksitas model. Fungsi Ω ini berperan penting dalam mencegah overfitting dengan memberikan penalti terhadap model yang terlalu rumit.

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2 \quad (3)$$

Dalam hal ini, λ adalah parameter regularisasi dengan nilai default sebesar 1 dan T menunjukkan jumlah daun (leaves) pada pohon keputusan. Fungsi Ω digunakan untuk mengukur kompleksitas model, yang ditentukan melalui nilai γ , ω merepresentasikan bobot pada setiap daun yang berfungsi sebagai nilai output.



Gambar 2. 2 Pohon regresi XGBOOST

Gambar 2.2 ditunjukkan tahapan pembelajaran pohon pertama yang dilakukan menggunakan data latih berupa fitur dan label sebenarnya (feature, Y), menghasilkan estimasi awal $Y1$. Selanjutnya, pohon kedua dilatih dengan data (feature, $|Y - Y1|$), yaitu selisih antara nilai aktual dan prediksi sebelumnya. Pohon ketiga kemudian mempelajari data (feature, $|Y - Y1 - Y2|$) untuk menghasilkan estimasi $Y3$. Dengan pendekatan berulang seperti ini, kesalahan prediksi (error) dapat dikurangi secara efektif (Siringoringo, Perangin-angin, and Purba 2021).

3.1.8 VADER (*Valance Aware Dictionary Sentiment Reasioner*)

VADER merupakan alat analisis sentimen yang dirancang khusus untuk keperluan pemrosesan bahasa alami dalam mengidentifikasi sentimen atau emosi dalam sebuah teks. Alat ini dikembangkan oleh tim dari SocialCog dan tersedia sebagai bagian dari pustaka Python NLTK (Natural Language Toolkit) (Ernawati and Wati 2024). VADER mengandalkan pendekatan berbasis leksikon dan aturan, yang dioptimalkan untuk mendeteksi ekspresi sentimen dalam konteks media sosial. Leksikon tersebut umumnya digunakan untuk mengevaluasi kalimat dan frasa berdasarkan nuansa emosinya tanpa memerlukan referensi eksternal. Dalam klasifikasinya, sistem ini biasanya menggunakan skor seperti negatif, netral, dan positif. Keunikan pendekatan leksikal ini terletak pada kemampuannya melakukan analisis tanpa memerlukan data berlabel untuk pelatihan model (Nanda Sepriadi et al. 2023).

Tabel 2. 1 Vader skor (Effendi; Noviana 2021)

No	Sentimen	Score	Keterangan
1.	Positif	≥ 0.05	Menunjukkan bahwa bobot kata lebih banyak mengandung positif
2.	Netral	$> -0.05 - < 0.05$	Menunjukkan bahwa bobot kata mengandung netral atau tidak terdiri dari positif dan negatif
3.	Negatif	≤ -0.05	Menunjukkan bahwa bobot kata lebih banyak mengandung negatif

3.1.9 LSTM (LongShort-Term Memory)

LSTM merupakan salah satu jenis *Recurrent Neural Network*. Setiap neuron *LSTM* memiliki banyak gerbang yang mengontrol memorinya, sehingga memungkinkan *LSTM* mempelajari data. Data teks, video, dan deret waktu seringkali diproses menggunakan *LSTM*. (Ningrum et al. 2021).

Algoritma yang efektif untuk membuat model deret waktu sekuensial adalah *LSTM*. Elemen dasar penyusun jaringan *LSTM* adalah blok memori yang dirancang untuk menahan gradien yang hilang dengan menghafal parameter jaringan dalam jangka panjang. Blok memori arsitektur *LSTM* sebanding dengan sistem penyimpanan diferensial pada sistem digital. Dengan bantuan fungsi aktivasi (*sigmoid*), gerbang masukan *LSTM* memproses informasi, menghasilkan keluaran dengan rentang 0 hingga 1. (Nugraha, Ariawan, and Arifin 2023).

Komponen pertama dalam arsitektur *LSTM* adalah *forget gate* (f_t). Fungsi utama dari *forget gate* adalah untuk mengatur seberapa besar informasi dari *cell state* sebelumnya (C_{t-1}) yang akan dipertahankan dan diteruskan ke *cell state* saat ini (C_t). Nilai output dari *forget gate* berada dalam rentang 0 hingga 1, karena menggunakan fungsi aktivasi *sigmoid* ($\sigma(\cdot)$). Persamaan matematis untuk *forget gate* diberikan sebagai berikut (Nilsen 2022)

$$f_t = \sigma(W_{hf} \cdot h_{t-1} + W_{xf} \cdot x_t + b_f)$$

Komponen kedua dalam struktur *LSTM* adalah *input gate* (i_t) dan *candidate state*. *Input gate* berfungsi untuk menentukan informasi baru mana yang akan

digunakan dalam memperbarui *cell state*. Sementara itu, *candidate state* bertugas mengumpulkan serta menghasilkan informasi baru dari input saat ini dan digunakan dalam proses pembaruan. Peran *candidate state* serupa dengan *cell state* (ht) pada arsitektur *RNN*. Namun, perbedaannya terletak pada cara informasi disalurkan: pada *RNN*, *cell state* langsung diteruskan ke unit berikutnya, sedangkan pada *LSTM*, *candidate state* digunakan sebagai dasar untuk memperbarui *cell state* C_t . Persamaan untuk *input gate* dan *candidate state* adalah sebagai berikut.

$$\tilde{C}_t = \tan h(W_{hc} \cdot h_{t-1} + W_{xc} \cdot x_t + b_c)$$

$$i_t = \sigma(W_{hi} \cdot h_{t-1} + W_{xi} \cdot x_t + b_i)$$

Komponen keempat dari arsitektur *LSTM* adalah *cell state* (C_t). *Cell state* berfungsi sebagai memori utama jaringan yang diperbarui dengan menggabungkan hasil dari *forget gate*, *input gate*, dan *candidate state*. Proses ini memungkinkan jaringan untuk mempertahankan atau memperbarui informasi penting dari waktu ke waktu. Adapun persamaan matematis yang menggambarkan pembaruan *cell state* adalah sebagai berikut.

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

Komponen kelima dalam struktur *LSTM* adalah *hidden state* (ht) dan *output gate* (ot). *Hidden state* (ht) berfungsi untuk membawa informasi ke *unit LSTM* pada waktu berikutnya serta ke lapisan *output* jaringan. Nilai ht diperoleh dengan menerapkan fungsi aktivasi *hyperbolic tangent* (\tanh) pada *cell state* yang telah diperbarui. Sementara itu, *output gate* bertugas mengontrol seberapa besar informasi dari *cell state* yang perlu disalurkan ke *hidden state* dan diteruskan ke langkah selanjutnya. Persamaan matematis untuk *hidden state* dan *output gate* adalah sebagai berikut.

$$h_t = o_t \odot \tanh(C_t)$$

$$o_t = \sigma(W_{ho} \cdot h_{t-1} + W_{xo} \cdot x_t + b_o)$$

3.2 Deskripsi Metode yang Sama

3.2.1 *State of the art (SOTA)*

Dari berbagai penelitian terdahulu, dapat disimpulkan bahwa integrasi antara analisis sentimen media sosial dan model prediksi harga saham menjadi pendekatan yang cukup efektif. Namun, masih terdapat beberapa *gap* yang dapat dijadikan dasar pengembangan lebih lanjut, antara lain:

Penelitian berjudul “*Optimasi Hyper-Parameter Berbasis Algoritma Genetika Pada Ensemble Learning untuk Prediksi Saham yang Mempertimbangkan Indikator Teknikal & Sentimen Berita*” ini dilakukan untuk membangun model prediksi harga saham yang lebih akurat dengan menggabungkan data teknikal (seperti MACD, RSI, EMA, SMA) dan sentimen berita yang diukur dengan *lexicon Loughran-McDonald*. Peneliti menggunakan pendekatan *stacked ensemble learning* yang menggabungkan empat model berbasis *decision tree* (*Random Forest*, *Extra Trees*, *XGBoost*, dan *LightGBM*) sebagai *base-learner* serta *XGBoost* sebagai *meta-learner*. Untuk mengoptimalkan performa model, algoritma genetika diterapkan dalam proses pencarian kombinasi *hyperparameter* terbaik. Hasilnya menunjukkan bahwa kombinasi fitur teknikal dan sentimen mampu meningkatkan akurasi prediksi hingga 81,63% dan *F1-score* 82,21%, serta menghasilkan return investasi mencapai 121,27% dalam simulasi trading. Kelebihan dari penelitian ini adalah penggunaan strategi optimasi yang efisien dan model ensemble yang kompleks namun efektif. Namun, penelitian ini masih terbatas pada pendekatan klasifikasi (naik/turun) dan *lexicon* bahasa Inggris untuk sentimen, yang mungkin belum sepenuhnya relevan untuk konteks lokal Indonesia. Berbeda dengan penelitian ini, pendekatan yang dilakukan menggunakan model seperti *XGBoost* untuk regresi harga saham secara langsung, serta melibatkan analisis sentimen berbasis bahasa Indonesia dari media sosial seperti Twitter dengan bantuan *VADER*, sehingga prediksi tidak hanya lebih kontekstual namun juga menghasilkan nilai harga aktual, bukan sekadar arah pergerakan (Wiranata and Djunaidy 2021).

Penelitian berjudul “*Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis*” mengusulkan model gabungan *FinBERT*

dan *LSTM* untuk memprediksi harga saham dengan mengintegrasikan data harga saham historis dan analisis sentimen dari berita keuangan. Metode yang digunakan adalah arsitektur *FinBERT-LSTM*, yang menggabungkan embedding sentimen dari teks berita keuangan dengan deret waktu harga penutupan saham untuk memprediksi harga penutupan berikutnya, dan dibandingkan dengan model *LSTM* dan *DNN*. Hasil menunjukkan bahwa model *FinBERT-LSTM* memberikan akurasi terbaik (0.955). Perbedaan dari penelitian ini, peneliti melakukan analisis sentimen menggunakan *vader* untuk melabeli setiap tweet mengenai saham setiap perusahaan dan menggabungkannya dengan data saham lalu dibuatlah model prediksi harga saham (Gu et al. 2024).

Penelitian berjudul "*Using Social Network Sentiment Analysis and Genetic Algorithm to Improve the Stock Prediction Accuracy of the Deep Learning-Based Approach*" mengusulkan metode prediksi harga saham yang menggabungkan analisis sentimen media sosial, algoritma genetika (*GA*), dan *deep learning* berbasis *LSTM*. Penelitian ini memanfaatkan data saham TSMC dan opini publik dari platform media sosial *PTT* untuk membentuk lima variabel sentimen yang kemudian dianalisis menggunakan *Grey Relational Analysis* (*GRA*). Seleksi fitur dilakukan melalui *Hybrid Genetic Algorithm* untuk memilih indikator teknikal (*chip-based indicators*) paling relevan, yang selanjutnya digunakan sebagai input model *LSTM*. Hasilnya menunjukkan bahwa penggabungan data teknikal dan sentimen sosial dapat meningkatkan akurasi prediksi hingga 62,22%. Kelebihan dari penelitian ini terletak pada integrasi multi-metode dan evaluasi terhadap periode sebelum dan sesudah pandemi *COVID-19*. Namun, jurnal ini masih terbatas pada satu sumber media sosial (*PTT*), hanya melakukan prediksi biner naik/turun. Berbeda dengan penelitian ini, peneliti menggunakan data saham Indonesia dan sentimen dari Twitter yang tidak hanya memprediksi arah pergerakan, tetapi juga nilai harga penutupan secara regresi dengan algoritma seperti *XGBoost* serta melibatkan *preprocessing* sentimen berbasis *VADER* yang lebih adaptif terhadap konteks bahasa Indonesia (Huang, Tung, and Lin 2023).

Penelitian berjudul “Analisis Prediksi Saham Tesla Menggunakan Algoritma *Long ShortTerm Memory (LSTM)*” membahas tentang prediksi harga saham tesla dengan menggunakan algoritma *Long ShortTerm Memory (LSTM)*. Dengan data yang digunakan yaitu data historis saham tesla dari 2020 hingga 2024. Penelitian ini menunjukkan perolehan nilai model *LSTM* mencapai nilai *RMSE* 0,0417 dan *MAPE* 0,1866, yang mengindikasikan tingkat akurasi yang baik dalam prediksi harga saham. Perbedaan dari penelitian ini, peneliti menambahkan fitur sentiment publik untuk membantu melihat pengaruh saham terhadap prediksi harga saham (Saputra et al. 2024).

Penelitian berjudul "Prediksi Harga Saham Menggunakan Sentimen Pilkada DKI Jakarta 2017 Dengan Algoritma *Support Vector Machine*" menggunakan metode klasifikasi sentimen dengan algoritma *Support Vector Machine (SVM)* untuk mengolah tweet bertema Pilkada DKI Jakarta 2017, yang kemudian digunakan sebagai pembobot dalam regresi linier untuk memprediksi harga saham. Hasil penelitian menunjukkan bahwa metode *CountVectorizer* menghasilkan akurasi klasifikasi tertinggi sebesar 91%, dan nilai *RMSE* prediksi harga saham bervariasi tergantung emiten, dengan yang terbaik adalah APLN (*RMSE* 17.342). Kelebihan penelitian ini adalah kemampuannya menggabungkan sentimen politik aktual dengan prediksi saham secara efektif, namun masih memiliki *gap* seperti jumlah data terbatas, belum adanya perbandingan algoritma lain, serta belum melibatkan faktor volume perdagangan atau peristiwa politik yang lebih besar. Perbedaan dari penelitian ini, peneliti menggunakan metode regresi dengan model *xgboost* untuk memprediksi harga saham dan sentiment yang dipakai tentang saham setiap perusahaan (Fadliansyah, Basuki, and Azhar 2020).

Penelitian berjudul "Prediksi Harga Saham Menggunakan *BiLSTM* dengan Faktor Sentimen Publik" menggunakan metode *Bidirectional Long Short-Term Memory (BiLSTM)* untuk memprediksi harga saham Bank BCA (BBCA) dengan menggabungkan data harga penutupan dan nilai *compound score* dari sentimen publik di Twitter yang dihitung menggunakan algoritma *VADER*. Hasil penelitian menunjukkan bahwa model pada skenario keempat dengan data latih terpanjang menghasilkan performa terbaik dengan nilai *MSE* 0.094

dan *RMSE* 0.306, sedangkan skenario pertama memiliki akurasi arah pergerakan (*Dstat*) tertinggi sebesar 68%. Kelebihan penelitian ini adalah mengintegrasikan faktor sentimen publik secara langsung ke dalam model *deep learning* dan menguji berbagai skenario panjang data pelatihan. Adapun *gap* penelitian terletak pada belum dipertimbangkannya variabel fundamental lainnya seperti laporan keuangan, rasio keuangan, atau variabel makroekonomi yang bisa meningkatkan akurasi prediksi di masa depan. Perbedaan dari penelitian ini, peneliti menggunakan algoritma *Vader* sebagai metode analisis sentimen dan mentraslatenya agar bisa masuk dalam kamus *lexicon vader* dan menambahkan beberapa kustomisasi untuk mengolah kalimat twitter saham lalu mengolahnya untuk menjadi fitur tambahan yaitu *negatif_tweet* dan *avg_compound* untuk membantu prediksi harga saham menggunakan model *XGBoost*(Afrianto and Nurdi 2022).

Penelitian berjudul “*Prediksi Pergerakan Naik Turun Harga Saham Berdasarkan Headline Berita Menggunakan Support Vector Machine dan Naive Bayes Classifier*” ini memprediksi arah pergerakan saham PT Semen Gresik (SMGR.JK) berdasarkan kombinasi antara sentimen dari judul berita dan data historis harga saham. Sentimen diperoleh dari berita yang dikumpulkan melalui scraping situs bisnis.com dan *Google News*, lalu dianalisis menggunakan pendekatan *lexicon-based* dengan kamus *InSet*. Dua algoritma yang digunakan adalah *SVM*, yang dikenal mampu menangani data kompleks, dan *Naive Bayes* yang unggul dalam kecepatan komputasi. Hasilnya, model *SVM* dengan *kernel linear* menunjukkan performa terbaik dengan akurasi 72,88%, lebih unggul dibanding NBC yang cenderung bias terhadap kelas tertentu. Keunggulan utama dari penelitian ini adalah integrasi data berita dengan fitur *time series* untuk klasifikasi arah harga saham, namun masih terbatas pada jenis data (judul berita saja) dan hanya fokus pada satu emiten. Berbeda dengan penelitian ini yang tidak hanya memprediksi arah naik-turun, tetapi juga nilai harga saham secara langsung (regresi), serta menggabungkan sentimen media sosial seperti Twitter menggunakan metode analisis *VADER*, sehingga memberikan hasil prediksi yang lebih mendalam dan kontekstual(Dwita and Mukarromah 2025).

Penelitian berjudul “Analisis Sentimen Ulasan Aplikasi Pembelajaran Bahasa Menggunakan Metode *VADER*” bertujuan menganalisis persepsi pengguna terhadap aplikasi Duolingo melalui ulasan di *Google Play Store* menggunakan metode *VADER* yang dikombinasikan dengan *Deep-Translator*. Algoritma *VADER* dipilih karena kemampuannya yang cepat dan efisien dalam klasifikasi berbasis leksikon, sementara *Deep-Translator* digunakan untuk menerjemahkan ulasan berbahasa Indonesia ke Inggris agar dapat diproses oleh *VADER*. Hasil penelitian menunjukkan akurasi tinggi (98%), dengan performa sangat baik pada sentimen positif dan netral, namun masih kurang optimal dalam mendeteksi sentimen negatif (*F1-score* 74%). Kelebihan penelitian ini adalah mampu memproses data multibahasa secara efisien, tetapi memiliki gap dalam menangkap nuansa emosional kompleks. Perbedaan dengan penelitian ini, terletak pada penerapannya juga menggunakan teknik *VADER* dan *Deep-Translator*, namun hasil sentimen tersebut tidak hanya dianalisis, melainkan digabungkan dengan data saham sebagai fitur tambahan dalam model prediksi harga saham, sehingga arah penggunaannya lebih bersifat prediktif dan terintegrasi dalam analisis pasar keuangan (Leonardi et al. 2025).

Penelitian berjudul “*Stock Price Prediction Method Based on XGBoost Algorithm*” membahas penggunaan algoritma *XGBoost* untuk memprediksi harga saham berdasarkan data *time series* harian. Penelitian ini memanfaatkan indikator teknikal seperti EMA, SMA, RSI, dan MACD sebagai variabel input, serta mengoptimalkan parameter model menggunakan *GridSearchCV* untuk menghindari *overfitting* dan *underfitting*. Hasilnya menunjukkan bahwa model *XGBoost* mampu menangkap pola fluktuasi harga saham dengan cukup akurat, terutama karena kemampuannya dalam memproses data berfrekuensi tinggi dan mendeteksi tren dari pola historis. Kelebihan utama dari penelitian ini adalah pendekatan sistematis dalam pemilihan parameter dan fokus pada fitur *time series* teknikal. Namun, terdapat beberapa keterbatasan, di antaranya tidak mempertimbangkan faktor eksternal seperti sentimen pasar, berita, atau perilaku investor yang bisa memengaruhi harga saham secara signifikan. Penelitian ini juga terbatas pada data teknikal dan tidak menggabungkan pendekatan berbasis emosi atau opini publik. Berbeda dengan penelitian ini,

peneliti tidak hanya menggunakan *XGBoost*, tetapi juga menggabungkan analisis sentimen dari media sosial seperti Twitter menggunakan algoritma *VADER*. Penelitian ini bertujuan memprediksi harga penutupan saham Indonesia secara regresi dengan mempertimbangkan pengaruh variabel sentimen(Zhang 2023).

Penelitian berjudul “Prediksi Harga Saham pada Bank BCA Tbk dengan Teknik *XGBoost (Extreme Gradient Boosting)*” bertujuan untuk meramalkan harga penutupan saham Bank BCA Tbk menggunakan algoritma *XGBoost* berbasis data historis harian dari tahun 2019–2023. Penelitian ini mencakup enam tahap: pengumpulan data, prapemrosesan, rekayasa fitur, pemilihan model, prediksi, dan evaluasi. *XGBoost* digunakan karena keunggulannya dalam menangani pola data *non-linear* dan kompleks, serta menghasilkan akurasi tinggi, terbukti dari nilai *MAPE* sebesar 2,48%. Kelebihan penelitian ini adalah pemanfaatan indikator teknikal seperti EMA, SMA, RSI, dan MACD dalam proses *feature engineering*. Namun, *gap*-nya adalah belum mempertimbangkan variabel eksternal seperti sentimen pasar atau tren harian sebelumnya. Berbeda dengan penelitian ini, peneliti akan membangun fitur seperti *Close Lag-1* (harga penutupan hari sebelumnya) dan fitur sentimen yang diperoleh dari analisis *VADER* dan *Deep-Translator*, sehingga model tidak hanya mengandalkan indikator teknikal, tetapi juga mengintegrasikan emosi pasar sebagai variabel tambahan untuk meningkatkan prediksi harga saham(Tri Nurmaya, Soesilo, and Abdul Fatah 2025).

BAB III

METODOLOGI PENELITIAN

3.1 Metodologi Penelitian

Pada sub bab metodologi penelitian ini digunakan oleh peneliti untuk menemukan cara menjawab rumusan masalah dan mencapai tujuan penelitian, mulai dari pengumpulan data, analisis, hingga pengujian model. Pada penelitian ini, pendekatan yang digunakan adalah kuantitatif, karena seluruh proses berfokus pada analisis data numerik dan pengukuran kinerja model prediktif secara objektif.

Penelitian ini bertujuan untuk membangun sebuah model prediksi harga saham yang mempertimbangkan dua jenis informasi utama, yaitu data historis harga saham dan data sentimen dari twitter. Untuk mencapai tujuan tersebut, diperlukan pendekatan yang mampu menggabungkan data numerik dan data teks dalam satu sistem prediksi yang terukur dan dapat diandalkan.

Jenis penelitian yang digunakan ialah penelitian eksperimen, dengan melakukan uji suatu metode atau algoritma untuk mengetahui seberapa baik kinerja dalam kondisi tertentu. Dalam situasi ini, model prediksi harga saham berbasis *Long Short-Term Memory (LSTM)* diuji untuk melihat kemampuannya dalam memproses data historis harga saham serta mengolah informasi *nonlinier* dari sentimen yang telah diolah. Peneliti tidak hanya mengamati fenomena, tetapi juga melakukan pengujian terhadap data dan membandingkan hasil prediksi dengan data sebenarnya.

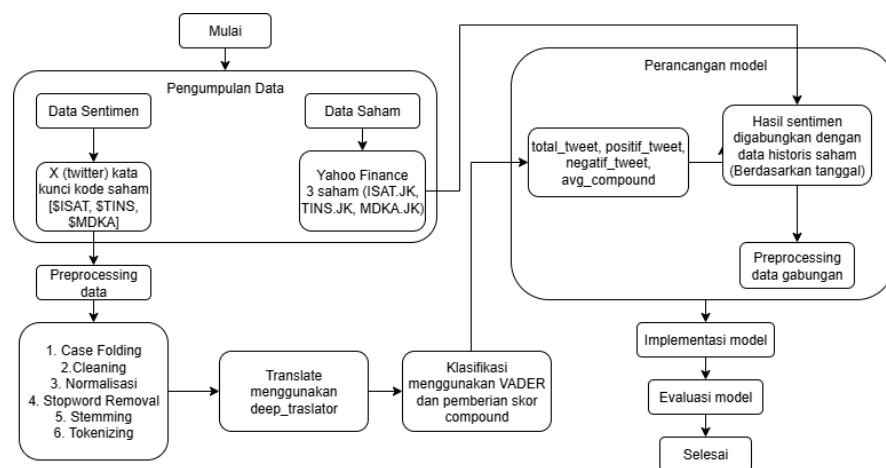
Selain itu, penelitian ini juga berupaya mengembangkan model prediksi yang dapat digunakan dalam pengambilan keputusan keuangan, khususnya terkait saham. Metodologi ini mendukung tujuan utama penelitian, yaitu membangun sistem prediksi harga saham yang mempertimbangkan faktor historis dan sentimen pasar secara bersamaan. Penelitian ini dibantu dengan berbagai alat dan teknologi modern, seperti:

- Python sebagai bahasa pemrograman utama,
- Pustaka (*library*) seperti *TensorFlow*, *Keras*, *Scikit-learn* untuk pengolahan data dan pembangunan model,
- Serta *platform* seperti *Google Colab* sebagai lingkungan kerja komputasi berbasis *cloud* dan juga *Visual Studio Code*.

Dengan pendekatan dan metode ini, diharapkan hasil penelitian dapat memberikan gambaran yang akurat dan bermanfaat, khususnya dalam memahami pengaruh data sentimen terhadap prediksi harga saham.

3.2 Tahapan Tahapan Diagram Metodologi Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang disusun secara sistematis untuk menghasilkan model prediksi harga saham yang optimal. Tahapan-tahapan tersebut digambarkan pada gambar 3.1 berikut:



Gambar 3. 1 Diagram alur tahapan penelitian(Taofiqurrohman and Roji 2025)

1. Pengumpulan Data

Pada tahap pengumpulan data ini terdiri dari 2 bagian:

a. Data Sentimen

Data ini dikumpulkan dari platform X(twitter) dengan menggunakan kata kunci berdasarkan kode saham, seperti ISAT (PT. Indosat Ooredoo Tbk.), TINS (PT.Timah Tbk.), dan MDKA (PT. Merdeka Copper Gold Tbk.) dengan menggunakan hastag atau simbol \$. Tweet yang mengandung kata kunci tersebut dikumpulkan untuk di analisis sentimen. Rentang tahun pengambilan data ini yaitu 2019-2025.

b. Data Saham

Data ini diperoleh dari situs Yahoo Finance untuk 3 saham yang diteliti yaitu ISAT.JK, TINS.JK dan MDKA.JK. Data yang dikumpulkan berupa *date*, *open*, *high*, *low*, *close* dan *volume*. Rentang tahun pengambilan data ini yaitu 2019- 9 Juli 2025.

2. Preprocessing Data

Pada tahap ini dilakukan untuk membersihkan dan menyiapkan data agar dapat digunakan dalam proses permodelan. Untuk data sentimen yang berasal dari Twitter, dilakukan beberapa tahapan pembersihan teks, seperti mengubah huruf menjadi kecil (*case folding*), menghapus karakter tidak penting (*cleaning*), dan mengganti kata tidak baku menjadi baku (*normalisasi*). Selanjutnya, dilakukan penghapusan kata umum yang tidak bermakna penting (*stopword removal*), mengubah kata ke bentuk dasar (*stemming*), dan memecah kalimat menjadi kata-kata (*tokenizing*).

Selanjutnya mentraslate teks yang telah diolah dari *tokenizing* ke Bahasa Inggris agar bisa di olah oleh VADER dalam mengklasifikasikan sentiment dengan memberikan skor *compound*.

Sementara itu, data saham hanya melalui tahap dasar, seperti penyesuaian format tanggal dan pengisian nilai yang hilang (jika ada), agar dapat digabungkan dengan data sentimen. Data yang telah digabung dapat diolah kembali dengan pembersihan data, data eksploratif sampai pada tahap pembuatan fitur tambahan yaitu *close_lag1*. Proses ini penting agar kedua jenis data selaras secara struktur dan bisa dianalisis bersama dalam model prediksi.

3. Perancangan Model

Pada tahap perancangan model yaitu inti dari proses pemodelan, di mana data sentimen dan data harga saham diolah untuk menghasilkan satu set data yang siap digunakan dalam pelatihan model prediksi. Langkah pertama adalah menghitung skor sentimen dari data teks menggunakan metode *Vader*. Hasil dari proses ini berupa nilai *polarity* yang menunjukkan kecenderungan positif atau negatif dari suatu tweet, dan nilai *compound* atau gabungan nilai score keduanya yang menunjukkan pernyataan tersebut.

Setelah itu, dilakukan rekapitulasi harian terhadap seluruh tweet yang telah dianalisis. Data dikumpulkan berdasarkan tanggal dan dihitung dalam bentuk total tweet, jumlah tweet positif dan negatif, rata-rata polaritas, serta rata-rata subjektivitas per hari. Hasil rekapitulasi sentimen ini kemudian digabungkan dengan data historis harga saham berdasarkan tanggal yang sama. Penggabungan ini menghasilkan data terpadu yang memuat aspek harga dan sentimen, sehingga dapat digunakan sebagai input dalam pelatihan model prediksi.

4. Implementasi Model

Pada tahap implementasi model ini data yang telah digabungkan digunakan untuk melatih model prediksi. Dalam hal ini, model XGBoost diterapkan karena kemampuannya dalam memproses data deret waktu (time series).

5. Evaluasi Model

Pada tahap evaluasi dilakukan menggunakan beberapa metrik untuk mengukur akurasi dan performa prediksi menggunakan MSE(Mean Squared Error), RMSE(Root Mean Squared Error), MAE(Mean Absolute Error) dan R^2 (R-Squared).

BAB IV

EKSPERIMEN DAN HASIL

4.1 Eksperimen

4.1.1 Pengumpulan Data

Pengumpulan data dalam penelitian ini terdiri dari dua jenis data utama, yaitu data harga saham historis, data sentimen dari media sosial X (dahulu Twitter) dan data gabungan antara sentimen dan saham.

4.1.1.1 Data Saham

Data harga saham diperoleh melalui platform *Yahoo Finance*, dengan rentang waktu pengambilan mulai dari Januari 2019 hingga Juli 2025. Tiga saham dipilih sebagai objek penelitian, yaitu:

- ISAT.JK (PT. Indosat Ooredoo Tbk.)
- TINS.JK (PT. Timah Tbk.)
- MDKA.JK (PT. Merdeka Copper Gold Tbk.)

```
# Pilih ticker saham
ticker = 'MDKA.JK'
# Tentukan rentang tanggal
start_date = '2019-01-01'
end_date = '2025-07-09'
# Unduh data dari Yahoo Finance
data = yf.download(ticker, start=start_date, end=end_date)
```

Proses pengumpulan data harga saham dilakukan dengan menyesuaikan *ticker code* masing-masing saham pada *platform Yahoo Finance*, dan data yang diperoleh mencakup *Date, Open, High, Low, Close* dan *Volume*.

4.1.1.2 Data Sentimen

Sementara itu, data sentimen diperoleh dari media sosial X melalui proses *crawling* dan *scraping* menggunakan *Twitter API*. Akses dilakukan secara resmi dengan permintaan token otorisasi (*access token*) yang disediakan oleh platform tersebut. Pengambilan data tweet dilakukan berdasarkan pencarian dengan menggunakan simbol dolar (\$) yang diikuti dengan kode saham, yaitu \$ISAT, \$TINS, dan \$MDKA.

```

twitter_auth_token = "ade-----"
# Daftar keyword saham
keywords = ["$ISAT", "$MDKA", "$TINS"]
# Loop untuk crawl masing-masing keyword
for keyword in keywords:
    filename = f"{keyword.lower()}_tweetISAT.csv"
    search_keyword = f"{keyword} since:2025-05-01 until:2025-07-09 lang:id"
    limit = 5000

```

Karena adanya kebijakan pembatasan baru dari platform X pada tahun 2023, pengambilan data dilakukan secara bertahap per saham dalam rentang waktu persatu tahun dari 2019 hingga 2025 untuk menghindari pemblokiran. Setelah seluruh data dari masing-masing saham berhasil dikumpulkan, dilakukan proses penggabungan dan pengurutan berdasarkan waktu untuk membentuk tiga dataset tweet yang terpisah sesuai dengan kode saham terkait. Data yang akan di dapatkan dari tweet ini yaitu *conversation_id_str*, *created_at*, *favorite_count*, *full_text*, *id_str*, *img_url*, *in_reply_to_screen_name*, *lang*, *location*, *quote_count*, *reply_count*, *retweet_count*, *tweet_url*, *user_id_str*, dan *username*.

4.1.1.3 Data Gabungan

Selanjutnya proses penggabungan data saham dan sentimen.

```

# Pastikan format tanggal seragam
saham_gabungan['Date'] =
pd.to_datetime(saham_gabungan['Date']).dt.date
hasil_sent_tins['date'] =
pd.to_datetime(hasil_sent_tins['date']).dt.date
hasil_sent_isat['date'] =
pd.to_datetime(hasil_sent_isat['date']).dt.date
hasil_sent_mdka['date'] =
pd.to_datetime(hasil_sent_mdka['date']).dt.date

```

Kode ini mengubah kolom *Date* atau *date* pada masing-masing *dataframe* (*saham_gabungan*, *hasil_sent_tins*, *hasil_sent_isat*, dan *hasil_sent_mdka*) ke dalam format *datetime*, lalu diambil hanya bagian tanggalnya dengan *dt.date*. Hal ini penting agar proses penggabungan (*merge*) antar data yang menggunakan kolom tanggal bisa dilakukan tanpa error dan dengan hasil yang akurat, karena format yang tidak seragam bisa menyebabkan ketidakcocokan saat pencocokan data berdasarkan tanggal.

```
# Filter saham berdasarkan ticker dengan format yang sesuai
saham_tins = saham_gabungan[saham_gabungan['Ticker'] ==
'TINS.JK']
saham_isat = saham_gabungan[saham_gabungan['Ticker'] ==
'ISAT.JK']
saham_mdka = saham_gabungan[saham_gabungan['Ticker'] ==
'MDKA.JK']
```

Kode tersebut berfungsi untuk memisahkan data saham berdasarkan masing-masing perusahaan dengan menggunakan *ticker*. Data saham dengan *ticker* 'TINS.JK' disimpan dalam *saham_tins*, *ticker* 'ISAT.JK' disimpan dalam *saham_isat*, dan *ticker* 'MDKA.JK' disimpan dalam *saham_mdka*, sehingga masing-masing variabel hanya berisi data dari satu perusahaan yang spesifik.

```
# Gabungkan saham dengan data sentimen berdasarkan tanggal
gabung_tins = pd.merge(saham_tins, hasil_sent_tins,
left_on='Date', right_on='date', how='inner')
gabung_isat = pd.merge(saham_isat, hasil_sent_isat,
left_on='Date', right_on='date', how='inner')
gabung_mdka = pd.merge(saham_mdka, hasil_sent_mdka,
left_on='Date', right_on='date', how='inner')
```

Kode tersebut digunakan untuk menggabungkan data saham dengan data sentimen berdasarkan tanggal yang sama. Proses penggabungan dilakukan untuk masing-masing perusahaan (TINS, ISAT, dan MDKA) menggunakan fungsi *merge* dengan metode *inner*, sehingga hanya data dengan tanggal yang cocok di kedua dataset yang akan digabungkan.

```
# Gabungkan semua menjadi satu dataframe
gabungan_sentimen_saham = pd.concat([gabung_tins, gabung_isat,
gabung_mdka], ignore_index=True)
```

Kode tersebut digunakan untuk menggabungkan seluruh data hasil *merge* dari TINS, ISAT, dan MDKA menjadi satu dataframe tunggal bernama *gabungan_sentimen_saham*, dengan menyusun ulang indeks secara otomatis menggunakan *ignore_index=True*.

Data hasil penggabungan *gabungan_sentimen_saham* memiliki kolom-kolom yang terdiri dari informasi harga saham seperti *Date*, *Ticker*, *Open*, *High*, *Low*, *Close*, dan *Volume*, serta informasi sentimen dari Twitter seperti *total_tweet*, *positif_tweet*, *negatif_tweet*, *netral_tweet*, dan *avg_compound*. Kolom-kolom ini

mencerminkan kondisi pasar dan sentimen publik untuk masing-masing perusahaan pada setiap tanggal tertentu.

4.1.2 *Preprocessing Data*

Tahap ini penting dalam menyiapkan data sebelum digunakan dalam proses analisis maupun pemodelan. Pada penelitian ini, terdapat dua jenis data yang diproses secara terpisah, yaitu data sentimen dari media sosial (Twitter/X) dan data gabungan hasil penggabungan antara data sentimen dan data harga saham. Berikut adalah penjelasan dari masing-masing tahap *preprocessing*.

a. *Preprocessing Data Sentimen*

Data sentimen yang diperoleh dari platform X (Twitter) bersifat mentah dan tidak langsung dapat digunakan dalam proses analisis. Oleh karena itu, dilakukan serangkaian proses pembersihan dan normalisasi untuk menyiapkan teks agar dapat diproses oleh model analisis sentimen menggunakan metode *VADER*. Langkah-langkah *preprocessing* data sentimen ini akan memproses kolom *full_text* dari data twitter meliputi:

1. *Case Folding*

Seluruh teks diubah menjadi huruf kecil (*lowercase*) untuk menyamakan format penulisan dan menghindari duplikasi kata berdasarkan kapitalisasi.

```
step1 = text.lower()
```

2. *Cleaning*

Menghapus elemen yang tidak relevan seperti tautan (*URL*), simbol, emotikon, tanda baca, serta karakter khusus yang tidak dibutuhkan dalam analisis sentimen.

```
step2 = re.sub(r"http\S+|www\S+|pic\.twitter\.com/\S+", "",
step1)
step2 = re.sub(r"^\w\s]", "", step2)
step2 = re.sub(r"\s+", " ", step2).strip()
```

3. *Normalisasi*

Mengganti kata tidak baku atau kata slang dengan padanan kata yang sesuai menggunakan kamus slang yang disusun secara manual. Selain itu, juga dilakukan perluasan makna terhadap istilah-istilah khas pasar saham seperti “cuan”, “cutloss”, “breakout”, dan sebagainya.

```
def normalize_kata(text):
    for slang, normal in kamus_slang.items():
        text = re.sub(r'\b' + re.escape(slang) + r'\b',
                      normal, text)
    return text
```

4. Stopword Removal

Menghapus kata-kata umum yang tidak memiliki nilai makna signifikan (seperti "dan", "yang", "ke", dan "dari") menggunakan Sastrawi *StopWordRemoverFactory*, yaitu library bahasa Indonesia yang umum digunakan dalam proses *NLP* (*Natural Language Processing*).

```
stopword_remover =
StopWordRemoverFactory().create_stop_word_remover()
```

5. Stemming

Mengubah kata ke bentuk dasar menggunakan Sastrawi *StemmerFactory*. Misalnya, kata “dibeli” akan distem menjadi “beli”, atau “membeli” menjadi “beli”.

```
stemmer = StemmerFactory().create_stemmer()
```

6. Tokenizing

Memecah kalimat menjadi kata-kata tunggal agar dapat dikenali oleh tokenisasi.

```
step6 = step5.split()
"tokens": step6
```

Setelah tahap *preprocessing* selesai, data teks diproses dengan model *VADER* untuk memperoleh nilai *score compound* dari setiap tweet. Kemudian dilakukan proses klasifikasi sentimen menjadi tiga kategori: positif (2), negatif (1) dan netral (0) menggunakan pendekatan *rule-based* yang mempertimbangkan hasil skor sentimen dan konteks frasa.

Sebelum melakukan proses *VADER text* perlu di *translate* ke Bahasa Inggris guna menyesuaikan aturan dari *lexicon-based*, *Vader* hanya tersedia dalam Bahasa Inggris (Gatti and van Stegeren 2020) maka dari itu peneliti melakukan penerjemah *Deep-Translator* dengan mengintegrasikan:


```
def translate_text(text):
    try:
        return GoogleTranslator(source='auto',
                                target='en').translate(text)
    except:
        return ""
df['text_en'] =
df['tokens'].astype(str).progress_apply(translate_text)
```

Fungsi *translate_text* digunakan untuk menerjemahkan sebuah teks ke dalam bahasa Inggris. Fungsi ini secara otomatis mendeteksi bahasa asal teks, lalu menerjemahkannya menggunakan *GoogleTranslator*. Jika terjadi kesalahan selama proses penerjemahan, fungsi akan menangkap error tersebut dan mengembalikan string kosong sebagai gantinya.

```
analyzer = SentimentIntensityAnalyzer()
custom_lexicon = {
    'nice': 1.5, 'good': 1.5, 'hold': 0.5, 'sell': -0.7, 'rise':
    1.6, 'maximum': 1.0, 'drop': -1.6, 'risk': -1.0, 'out': -0.5}
analyzer.lexicon.update(custom_lexicon)
```

Selanjutnya menjalankan kode untuk menambahkan *custom_lexicon* agar istilah saham seperti ‘*maximum*’ = +1.0 dan ‘*drop*’ = -1.6 dengan bobot yang ditentukan secara manual agar sesuai konteks pasar saham. Pendekatan ini sesuai dengan (Du et al. 2023), menunjukkan bahwa analisis sentimen keuangan efektif menggunakan *lexicon* yang dibuat secara manual oleh ahli domain, bukan sepenuhnya otomatis.

```
def get_vader_label(text):
    scores = analyzer.polarity_scores(text)
    compound = scores['compound']
    if compound >= 0.05:
        return 2 # Positif
    elif compound <= -0.05:
        return 1 # Negatif
    else:
        return 0 # Netral
```

Hasil labeling di dapatkan dari perhitungan numerik *polarity_score* ini, akan menghasilkan *score_compound* yaitu skor gabungan positif dan negatif. Nilai ini mengukur kecenderungan emosional dalam tweet:

- Label 2 jika positif ($compound \geq 0.05$)
- Label 1 jika negatif ($compound \leq -0.05$)
- Label 0 jika netral ($compound$ di antara -0.05 dan 0.05)

```
df['vader_compound'] = df['text_en'].apply(lambda x:
analyzer.polarity_scores(x)['compound'])
df['vader_label'] = df['text_en'].apply(get_vader_label)
```

Kode tersebut digunakan untuk menghitung nilai skor sentimen *compound* dari setiap teks dalam kolom *text_en* menggunakan *VADER*, lalu menyimpannya di kolom baru bernama *vader_compound*. Kode selanjutnya digunakan untuk mengklasifikasikan sentimen (positif, negatif, atau netral) dari setiap teks dalam kolom *text_en* menggunakan fungsi *get_vader_label*, lalu hasil labelnya disimpan di kolom baru *vader_label*.

Data yang telah diolah ini akan menghasilkan beberapa kolom yaitu *Date*, *full_text*, *lower*, *cleaned*, *normalized*, *stopword*, *steemed*, *tokens*, *text_en*, *vader_compound*, dan *vader_label*.

```
rekap_harian = df.groupby('date').agg(
    total_tweet=('vader_label', 'count'),
    positif_tweet=('vader_label', lambda x: (x == 2).sum()),
    negatif_tweet=('vader_label', lambda x: (x == 1).sum()),
    netral_tweet=('vader_label', lambda x: (x == 0).sum()),
    avg_compound=('vader_compound', 'mean')
).reset_index()
```

Proses selanjutnya dilakukan proses peringkasan per hari atau tanggal dengan membari perintah *groupby('date')* digunakan untuk mengelompokkan data berdasarkan tanggal. Kemudian, fungsi *agg()* digunakan untuk menghitung beberapa ringkasan statistik per tanggal, yaitu:

total_tweet = jumlah total tweet dalam satu hari

positif_tweet = jumlah tweet positif dalam satu hari (*vader_label*=2)

netral_tweet = jumlah tweet netral dalam satu hari (*vader_label*=0)

negatif_tweet = jumlah tweet negatif dalam satu hari (*vader_label*=1)

avg_compound = jumlah nilai rata-rata *compound* skor vader dari semua tweet di hari tersebut.

b. Preprocessing Data Gabungan

count	
Ticker	
MDKA.JK	328
ISAT.JK	256
TINS.JK	213
dtype: int64	

Gambar 4. 2 Distribusi jumlah data per saham (*ticker*)

Gambar 4.2 menunjukkan jumlah entri data saham untuk masing-masing *ticker*. MDKA.JK merupakan *ticker* dengan jumlah data terbanyak di 3, sedangkan TINS.JK paling sedikit.

2. Pembuatan Fitur Derivatif

```
df['YearMonth'] = df['Date'].dt.to_period('M')
```

Ditambahkan kolom *Year*, *Month*, dan *YearMonth* untuk kebutuhan analisis tren bulanan.

3. Analisis Bulanan

```
first_close =
df.sort_values('Date').groupby('YearMonth')['Close'].first()
last_close =
df.sort_values('Date').groupby('YearMonth')['Close'].last()

monthly_agg = df.groupby('YearMonth').agg({
    'positif_tweet': 'sum',
    'negatif_tweet': 'sum',
    'avg_compound': 'mean'
}).reset_index()
```

Dilakukan agregasi per bulan dengan setiap baris pada tabel mewakili ringkasan data untuk satu bulan (*YearMonth*) yang diurutkan berdasarkan tanggal. Dan menghitung total tweet positif, negatif dengan *sum*, rata-rata compound dengan *mean*.

```
# Gabungkan semua hasil ke dataframe akhir
monthly_agg['Harga_Awal'] = first_close.values
monthly_agg['Harga_Akhir'] = last_close.values
monthly_agg['Kenaikan_Harga_%'] =
((monthly_agg['Harga_Akhir'] - monthly_agg['Harga_Awal'])
 / monthly_agg['Harga_Awal'])
```

Kode ini digunakan Harga_Awal dan Harga_Akhir ditambahkan dari nilai harga penutupan pertama dan terakhir setiap bulan. Kemudian, dihitung

persentase kenaikan atau penurunan harga saham tiap bulan melalui rumus:

$$(\text{Harga_Akhir} - \text{Harga_Awal}) / \text{Harga_Awal}$$

Dari analisis bulanan ini akan menghasilkan informasi berikut:

- **Harga_Awal & Harga_Akhir:** Harga saham di awal dan akhir bulan tersebut.
- **Kenaikan_Harga_%:** Persentase perubahan harga selama bulan tersebut.
- **Sentimen_Positif & Sentimen_Negatif:** Jumlah tweet berisi sentimen positif dan negatif terkait saham dalam bulan itu.
- **Rata2_Compound:** Nilai rata-rata *compound* dari semua tweet, menunjukkan seberapa positif atau negatif secara umum sentimen yang beredar.

	YearMonth	Harga_Awal	Harga_Akhir	Kenaikan_Harga_%	Sentimen_Positif	Sentimen_Negatif	Rata2_Compound
0	2019-01	347.079747	858.398613	1.473203	0	0	0.000000
1	2019-02	723.519659	786.434462	0.086957	3	0	0.169862
2	2019-04	488.637953	488.637953	0.000000	0	0	0.000000
3	2019-05	416.286010	370.148501	-0.110831	1	0	0.386950
4	2019-06	503.318087	503.318087	0.000000	0	0	0.000000
...
72	2025-02	2220.000000	1045.000000	-0.529279	17	6	0.114783
73	2025-03	1480.000000	905.000000	-0.388514	9	5	0.176880
74	2025-05	1814.579956	2020.000000	0.113205	19	4	0.280564
75	2025-06	1060.116455	2000.000000	0.886585	33	10	0.229172
76	2025-07	2030.000000	2080.000000	0.024631	26	0	0.364903

Gambar 4. 3 Tabel Agresi Bulanan Saham dan Sentimen

Gambar 4.3 menunjukkan ringkasan bulanan harga saham beserta metrik sentimen seperti jumlah tweet positif/negatif dan nilai rata-rata *compound*, yang digunakan dalam analisis hubungan antara sentimen pasar dan presentase pergerakan harga.

4. Eksplorasi Data (EDA)

Visualisasi dilakukan dengan:

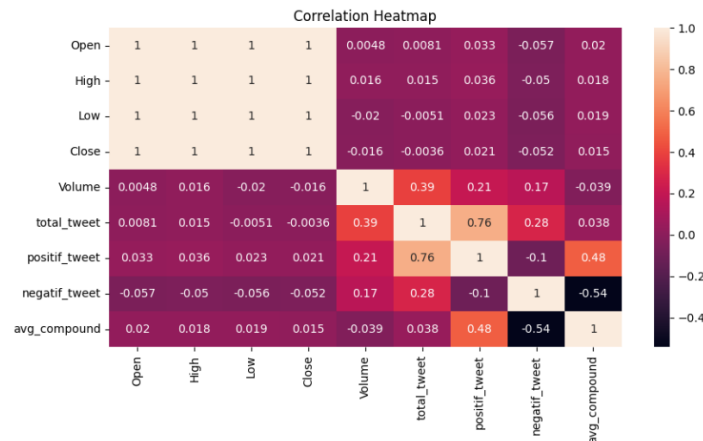
Heatmap korelasi antar fitur numerik pada dataset yang digunakan.

```
sns.heatmap(df[['Open', 'High', 'Low', 'Close', 'Volume', 'total_tweet', 'positif_tweet', 'negatif_tweet', 'avg_compound']].corr(), annot=True)
```

Kode *corr()* menghitung korelasi antar kolom seperti *Open*, *High*, *Low*, *Close*, *Volume*, *total_tweet*, *positif_tweet*, *negatif_tweet*, *avg_compound*. *sns.heatmap(..., annot=True)* membuat visualisasi korelasi dalam bentuk

matriks warna, dengan nilai korelasi ditampilkan di dalam kotaknya. Nilai korelasi berkisar antara -1 hingga 1, di mana:

- +1 menunjukkan korelasi positif sempurna
- 0 menunjukkan tidak ada korelasi
- -1 menunjukkan korelasi negatif sempurna



Gambar 4. 4 *Heatmap* korelasi antar fitur numerik

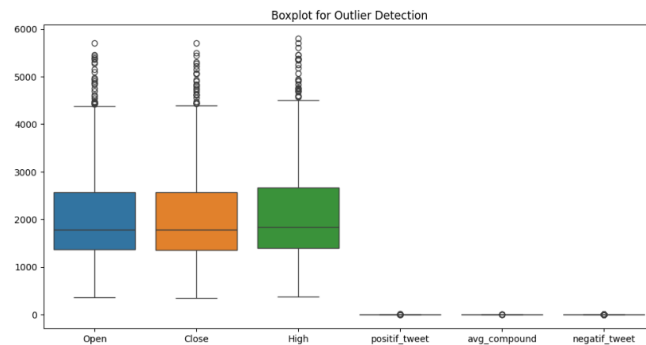
Gambar 4.4 menunjukkan hasil analisis korelasi, fitur *negatif_tweet* memiliki hubungan yang sangat lemah terhadap harga penutupan saham (*Close*) secara *linier*. Nilai korelasi *negatif_tweet* terhadap *Close* sebesar -0.052, sementara *avg_compound* berkorelasi sebesar -0.015. Hal ini menunjukkan bahwa pergerakan harga saham tidak secara langsung dipengaruhi oleh jumlah tweet negatif maupun nilai rata-rata polaritas sentimen. Meskipun demikian, perubahan ekstrem dalam fitur ini tetap berpotensi sebagai sinyal awal pergerakan harga, terutama bila dianalisis dengan pendekatan *non-linier* seperti *LSTM* maupun *XGBOOST* yang mempertimbangkan pola sekuensial dan hubungan tidak langsung antar variabel.

Boxplot untuk mendeteksi *outlier* pada fitur saham dan sentimen.

```
sns.boxplot(data=df[['Open', 'Close', 'High', 'positif_tweet', 'avg_compound', 'negatif_tweet']])
```

Kode *boxplot()* digunakan untuk membuat visualisasi *boxplot* dari beberapa kolom numerik, dengan tujuan melihat ringkasan statistik seperti median,

kuartil, serta mengidentifikasi data yang menyimpang (*outlier*) pada masing-masing variabel.



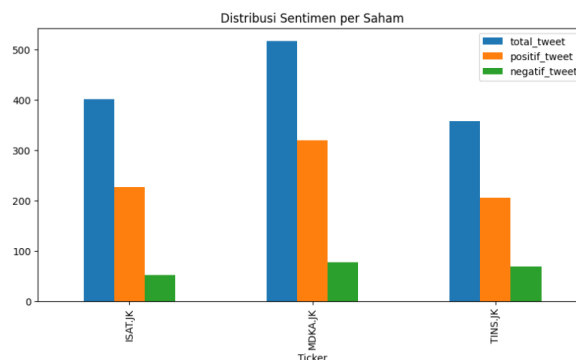
Gambar 4. 5 *Boxplot* deteksi *outlier* pada fitur harga dan sentiment

Gambar 4.5 menunjukkan keberadaan nilai-nilai ekstrem (*outlier*) pada fitur harga saham (*Open*, *Close*, *High*) dan fitur sentimen (*avg_compound*, *negatif_tweet*, *total_tweet*). Terlihat bahwa outlier banyak muncul pada fitur harga, sedangkan fitur sentimen memiliki skala nilai yang rendah dengan sebaran sempit.

Bar chart untuk melihat distribusi sentiment per saham.

```
df.groupby(['Ticker'])[['total_tweet', 'positif_tweet', 'negatif_tweet']].sum().plot(kind='bar', figsize=(10,5))
```

Kode tersebut digunakan untuk membuat diagram batang (*bar chart*) yang menunjukkan total jumlah tweet, tweet positif, dan tweet negatif untuk setiap perusahaan (*Ticker*). Data dikelompokkan *groupby()* berdasarkan kolom *Ticker*, kemudian dijumlahkan (*sum()*) untuk kolom *total_tweet*, *positif_tweet*, dan *negatif_tweet*, lalu hasilnya divisualisasikan dengan *plot(kind='bar')* dalam ukuran grafik 10x5. Visualisasi ini membantu membandingkan tingkat aktivitas dan sentimen publik antar saham.



Gambar 4. 6 Distribusi jumlah tweet per saham

Gambar 4.6 menunjukkan distribusi total tweet, tweet positif, dan tweet negatif untuk masing-masing saham. MDKA.JK memiliki jumlah tweet tertinggi, diikuti oleh ISAT.JK dan TINS.JK. Ketiga saham secara umum lebih banyak menerima sentimen positif dibandingkan negatif.

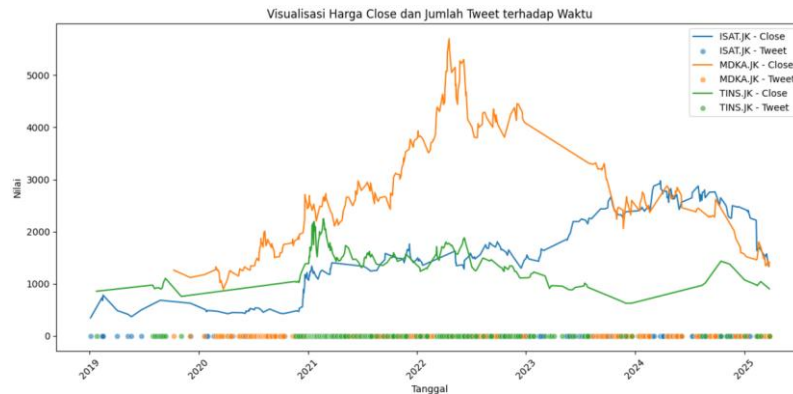
```
for ticker in TICKERS:
    subset = df[df['Ticker'] == ticker]
    sns.lineplot(x='Date', y='Close', data=subset,
label=f"{ticker} - Close")
    sns.scatterplot(x='Date', y='total_tweet', data=subset,
label=f"{ticker} - Tweet", marker='o', alpha=0.6)
plt.title("Visualisasi Harga Close dan Jumlah Tweet terhadap Waktu")
plt.xlabel("Tanggal")
plt.ylabel("Nilai")
```

Kode tersebut digunakan untuk memvisualisasikan harga penutupan (*Close*) dan jumlah tweet (*total_tweet*) dari setiap saham dalam daftar *TICKERS* terhadap waktu (*Date*).

Untuk setiap *ticker*:

- Data difilter sesuai perusahaan,
- Grafik garis (*lineplot*) digunakan untuk menampilkan tren harga penutupan,
- Titik-titik (*scatterplot*) digunakan untuk menunjukkan jumlah tweet per tanggal.

Judul dan label sumbu ditambahkan untuk memperjelas isi grafik. Visualisasi ini membantu membandingkan dinamika harga dan aktivitas tweet dari tiap saham sepanjang waktu.



Gambar 4. 7 Visualisasi harga penutupan dan jumlah tweet terhadap waktu

Grafik gambar 4.7 menunjukkan tren harga penutupan saham ISAT.JK, MDKA.JK, dan TINS.JK selama periode 2019–2025, disertai distribusi waktu kemunculan tweet yang berkaitan dengan masing-masing saham. Titik-titik pada grafik mewakili tanggal saat terdapat tweet terkait ticker tertentu, dan distribusinya menunjukkan bahwa aktivitas media sosial cenderung meningkat saat terjadi fluktuasi harga saham.

```
subset = df_isat # atau df_mdka, df_tins

fig, ax1 = plt.subplots(figsize=(14, 5))
ax1.plot(subset['Date'], subset['Close'], 'g-', label='Harga Close')
ax1.set_ylabel('Harga Close', color='green')
ax1.tick_params(axis='y', labelcolor='green')

ax2 = ax1.twinx()
ax2.bar(subset['Date'], subset['negatif_tweet'], width=1,
alpha=0.3, color='red', label='Tweet Negatif')
ax2.set_ylabel('Jumlah Tweet Negatif', color='red')
ax2.tick_params(axis='y', labelcolor='red')
ax2.set_ylim(0, 2.5)
ax1.legend(loc='upper left')
ax2.legend(loc='upper right')

plt.title(f"{subset['Ticker'].iloc[0]} - Harga Close vs Tweet Negatif")
plt.xlabel("Tanggal")
```

- `subset = df_isat`
→ Memilih data saham ISAT, MDKA, atau TINS untuk divisualisasikan.

- `fig, ax1 = plt.subplots(...)` dan `ax1.plot(...)`
→ Membuat grafik garis untuk menampilkan harga penutupan (*Close*).
- `ax1.set_ylabel(...)` dan `tick_params(...)`
→ Mengatur label dan warna sumbu Y kiri untuk harga.
- `ax2 = ax1.twinx()` dan `ax2.bar(...)`
→ Menambahkan sumbu Y kanan dan membuat grafik batang jumlah tweet negatif.
- `ax2.set_ylim(0, 2.5)`
→ Mengatur batas atas-bawah sumbu Y kanan agar lebih terbaca.
- `legend(...)`, `title`, dan `xlabel`
→ Menambahkan keterangan grafik, judul, dan label sumbu X.

Fungsinya adalah untuk membandingkan pergerakan harga saham dengan jumlah tweet negatif dalam satu grafik waktu yang sama.

5. Transformasi dan Penanganan *Outlier*

```
df['Open_log'] = np.log1p(df['Open'])
df['Close_log'] = np.log1p(df['Close'])
df['High_log'] = np.log1p(df['High'])
```

Transformasi logaritmik (*log1p*) diterapkan pada kolom harga (*Open*, *Close*, *High*) untuk mengurangi skewness dan distribusi data menjadi lebih normal dan mengurangi pengaruh nilai ekstrem (*outlier*). Diubah menjadi *Open_log*, *Close_log*, dan *High_log*.

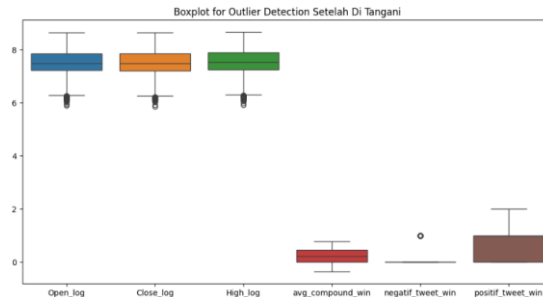
```
df['avg_compound_win'] = winsorize(df['avg_compound'],
limits=[0.05, 0.05])
df['positif_tweet_win'] = winsorize(df['positif_tweet'],
limits=[0.0, 0.10])
df['negatif_tweet_win'] = winsorize(df['negatif_tweet'],
limits=[0.0, 0.05])
```

Winsorizing dilakukan pada fitur *negatif_tweet* dan *total_tweet* untuk memotong nilai ekstrem atau membatasi nilai ekstrim pada batas tertentu:

- **avg_compound_win**: membatasi 5% nilai terendah dan tertinggi pada kolom *avg_compound*,
- **positif_tweet_win**: membatasi 10% nilai tertinggi pada kolom *positif_tweet*,

- **negatif_tweet_win**: membatasi 5% nilai tertinggi pada kolom negatif_tweet.

Tujuannya adalah agar data ekstrem tidak terlalu memengaruhi analisis atau model.



Gambar 4. 8 Boxplot deteksi outlier setelah penanganan

Gambar 4.11 menampilkan distribusi fitur-fitur numerik setelah dilakukan transformasi logaritmik (Open_log, Close_log, High_log) dan winsorizing (negatif_tweet, avg_compound). Sebaran data menjadi lebih normal, outlier berkurang, dan siap digunakan dalam pelatihan model prediksi.

6. Validasi Konsistensi Data

Untuk memastikan integritas data, dilakukan validasi terhadap kolom harga *Close*.

```
invalid_close = df[df['Close'] < 0]

unreasonable_jump = df.copy()
unreasonable_jump['prev_close'] =
unreasonable_jump.groupby('Ticker')['Close'].shift(1)
unreasonable_jump['pct_change'] =
(unreasonable_jump['Close'] -
unreasonable_jump['prev_close']) /
unreasonable_jump['prev_close'] * 100
outliers =
unreasonable_jump[np.abs(unreasonable_jump['pct_change']) >
1000]
```

Kode tersebut digunakan untuk mendeteksi lonjakan harga saham yang tidak wajar (lebih dari 1000%) amupun harga close kurang dari 0. `shift(1)` menghitung harga penutupan sebelumnya per *ticker*, `'pct_change'` menghitung persentase perubahan harga, baris terakhir memfilter data yang

mengalami lonjakan lebih dari $\pm 1000\%$, yang dianggap sebagai outlier ekstrim.

Hasil menunjukkan tidak ditemukan nilai harga yang negatif maupun tidak masuk akal. Selain itu, dilakukan juga pengecekan perubahan harga harian, dan hasilnya menunjukkan tidak terdapat lonjakan ekstrem (lebih dari 1000%) antar hari pada seluruh *ticker* saham. Dengan demikian, data dinyatakan layak untuk digunakan dalam pemodelan.

7. Pembuatan Fitur (*Feature Engineering*)

```
df = df.sort_values(['Ticker', 'Date'])
df['Close_lag1'] =
df.groupby('Ticker')['Close_log'].shift(1)
```

Sebagai bagian dari proses *feature engineering*, dibuat fitur baru bernama *Close_lag1*, yaitu harga penutupan (*Close*) pada hari sebelumnya. Tujuan dari pembuatan fitur ini adalah untuk membantu model mempelajari pola historis pergerakan harga saham. Karena model seperti *XGBoost* maupun *LSTM* membutuhkan urutan waktu sebagai input, fitur *lag* menjadi penting dalam membentuk struktur data sekuensial yang merepresentasikan dinamika harga dari waktu ke waktu.

Setelah pembuatan fitur *Close_lag1*, baris pertama dari setiap ticker akan memiliki nilai kosong (*NaN*) karena tidak memiliki data hari sebelumnya. Nilai-nilai *NaN* ini perlu dihapus agar tidak menimbulkan *error* saat proses pelatihan model. Oleh karena itu, dilakukan penghapusan seluruh baris yang mengandung *NaN*, baik akibat dari fitur lag maupun proses lainnya.

```
feature_cols = ['Open_log',
'High_log', 'negatif_tweet', 'Close_lag1']
target_col = 'Close_log'

data = df_lag[df_lag['Ticker'] == 'ISAT.JK'].copy()
data['Date'] = pd.to_datetime(data['Date'])
data = data.sort_values('Date')
data.set_index('Date', inplace=True)

X = data[feature_cols]
y = data[target_col]
```

Dari kode tersebut *feature_cols* berfungsi untuk menentukan daftar kolom yang akan digunakan sebagai fitur input dalam model prediksi. *Target_cols* menentukan kolom target yang akan diprediksi oleh model, yaitu logaritma dari harga penutupan. *data* berfungsi untuk mengambil data saham dengan kode *ticker* ISAT.JK dari dataset utama dan membuat salinan dengan *copy()* agar data aslinya tidak terpengaruh. *to_datetime()* kolom tanggal menjadi format *datetime* agar dapat digunakan untuk pengurutan waktu dan analisis *time series*. *Sort_values* mengurutkan data berdasarkan tanggal secara kronologis agar urutan waktu tetap terjaga dalam analisis. Menetapkan kolom *Date* sebagai *index* dataframe untuk memudahkan pemodelan *time series* dan visualisasi berdasarkan waktu. *Set.index()* menetapkan kolom *Date* sebagai *index* dataframe untuk memudahkan pemodelan *time series* dan visualisasi berdasarkan waktu, *X* sebagai kolom fitur yang telah ditentukan sebagai input model. *Y* sebagai kolom target yang akan diprediksi oleh model dari data yang telah difilter.

Data disaring berdasarkan *ticker* saham yang akan dianalisis, dalam hal ini 'TINS.JK', 'MDKA.JK' dan 'ISAT.JK'. Data yang digunakan kemudian dikonversi ke format tanggal (*datetime*) dan dijadikan indeks agar lebih mudah dikelola dalam analisis berbasis waktu. Selanjutnya, hanya kolom-kolom relevan yang dipilih sebagai fitur (*Open_log*, *High_log*, *negatif_tweet*, *Close_lag1*) dan target (*Close*).

8. Split Data (Train-Val-Test Split)

Untuk *split* data ini terdapat kode untuk proses *split* data pada *LSTM* dan *XGBoost* dalam file yang berbeda sebagai bahan *experiment* mana yang paling cocok untuk data yang peneliti miliki.

```

n_total = len(X)
train_size = int(n_total * 0.7)
val_size = int(n_total * 0.15)
test_size = n_total - train_size - val_size # sisa 15%

X_temp = X[:-test_size]
y_temp = y[:-test_size]

X_test = X[-test_size:]
y_test = y[-test_size:]

X_train = X_temp[:-val_size]
y_train = y_temp[:-val_size]

X_val = X_temp[-val_size:]
y_val = y_temp[-val_size:]

print(f"Train: {len(X_train)}, Validation: {len(X_val)},
Test: {len(X_test)}")

```

a.) *XGBoost*

Kode ini digunakan untuk membagi dataset time series menjadi tiga bagian, yaitu *training*, *validation*, dan *testing*, tanpa mengacak urutan data (karena ini penting untuk data deret waktu/saham).

Pertama-tama, program menghitung total jumlah data (*n_total*) berdasarkan panjang data fitur *X*. Kemudian, ditentukan bahwa 70% data akan digunakan untuk pelatihan (*train_size*), 15% untuk validasi (*val_size*), dan 15% sisanya untuk pengujian (*test_size*). Perhitungan ini dilakukan secara eksplisit agar proporsi pembagian data tetap terjaga.

Setelah itu, dilakukan pemisahan data:

- 15% data paling akhir digunakan sebagai data pengujian (*test set*), dengan *X[-test_size:]* dan *y[-test_size:]*.
- 85% data awal disimpan dalam variabel *X_temp* dan *y_temp* sebagai kandidat untuk data pelatihan dan validasi.

Kemudian, dari 85% awal tersebut:

- 70% bagian pertama dari 85% diambil sebagai data pelatihan (*train set*).
- 15% bagian terakhir dari 85% digunakan sebagai data validasi (*validation set*).

Akhirnya, kode mencetak jumlah data pada masing-masing *set* (*train*, *val*, *test*) agar pengguna bisa memverifikasi proporsi pembagian yang dihasilkan.

b.) *LSTM*

Perbedaan dari pembagian data pada model *LSTM* setelah data difilter dan dipilih fitur serta target-nya, dilakukan proses *windowing* yaitu mengubah data menjadi bentuk sekuensial yang dibutuhkan oleh model *LSTM*. *Windowing* dilakukan dengan cara mengambil sejumlah hari ke belakang (*window* = 3) sebagai input untuk memprediksi hari berikutnya. Proses ini menghasilkan array tiga dimensi dengan struktur (jumlah sampel, panjang *window*, jumlah fitur).

9. *Scaling* (Normalisasi)

Langkah selanjutnya adalah melakukan normalisasi atau penskalaan data. Normalisasi dilakukan agar seluruh fitur numerik berada dalam rentang yang sama dan tidak mendominasi proses pelatihan model karena perbedaan skala. Proses ini menggunakan metode *Min-Max Scaling* yang mengubah nilai-nilai fitur ke dalam rentang 0–1. Penskalaan ini penting khususnya pada model *deep learning* seperti *LSTM* maupun *XGBoost* yang sensitif terhadap skala *input*.

```
from sklearn.preprocessing import MinMaxScaler

scaler_X = MinMaxScaler()
scaler_y = MinMaxScaler()

X_train_scaled = scaler_X.fit_transform(X_train)
X_val_scaled = scaler_X.transform(X_val)
X_test_scaled = scaler_X.transform(X_test)

y_train_scaled =
scaler_y.fit_transform(y_train.values.reshape(-1, 1))
y_val_scaled = scaler_y.transform(y_val.values.reshape(-1,
1))
y_test_scaled = scaler_y.transform(y_test.values.reshape(-
1, 1))
```

Setelah seluruh tahap *preprocessing* selesai, data gabungan siap digunakan dalam tahap implementasi model prediksi menggunakan arsitektur *LSTM* (*Long Short-Term Memory*).

4.1.3 Perancangan Model

Pada tahap ini, dilakukan perancangan model prediksi untuk memproyeksikan harga saham atau arah pergerakannya berdasarkan data historis dan fitur-fitur yang telah direkayasa pada tahap *preprocessing* sebelumnya. Penelitian ini menggunakan dua pendekatan utama, yaitu model *deep learning* berbasis urutan (*LSTM*) dan model machine learning berbasis pohon (*XGBoost*). Pemilihan kedua model ini didasarkan pada karakteristik data dan tujuan yang ingin dicapai, yaitu memanfaatkan informasi temporal dari data historis saham serta menggali hubungan *non-linear* antara fitur-fitur dengan target. Adapun proses perancangan model meliputi:

A) *XGBoost*

Arsitektur model : Jumlah pohon (*estimators*) dan kedalaman pohon (*max_depth*) merupakan parameter utama yang menentukan kompleksitas model.

Hyperparameter : *n_estimators*, *max_depth*, *learning_rate*, *subsample*, *colsample_bytree*

B) *LSTM*

Arsitektur model :

- *Input layer*: menerima urutan data (misalnya: 30 hari sebelumnya)
- *LSTM layer*: 1 atau lebih *layer* dengan sejumlah unit memori 32
- *Dense layer*: koneksi penuh layer untuk menggabungkan *output LSTM*
- *Output layer*: *neuron* tunggal jika prediksi nilai kontinu harga

Hyperparameter : *Epoch*, *batchsize*, *optimizer*, *learning rate*, *loss function*, *time steps*

Setelah melakukan perancangan model proses selanjutnya bisa langsung mengimplementasikan model.

4.1.4 Implementasi Model

Implementasi dilakukan dengan membangun model terpisah untuk masing-masing saham, yaitu TINS, ISAT, dan MDKA. Hal ini karena karakteristik pergerakan harga dan sentimen ketiga saham tersebut memiliki pola yang unik sehingga diperlukan pendekatan model yang independen untuk masing-masing *ticker*.

a) Model *XGBoost*

Beberapa kombinasi parameter yang dieksplorasi dalam eksperimen *XGBoost* antara lain:

- *n_estimators*: 60, 70, 80
- *max_depth*: 2, 3
- *learning_rate*: 0.03, 0.05
- *subsample*: 0.6, 0.7, 0.8
- *colsample_bytree*: 0.6, 0.7, 0.8
- *reg_alpha*: 0.1, 0
- *reg_lambda*: 0.1, 0

b) Model LSTM

Eksperimen dilakukan dengan mencoba berbagai arsitektur dan kombinasi *hyperparameter*, yaitu:

- *units LSTM*: 32,
- *batch_size*: 16, 32
- *epochs*: 50, 100, 150
- *time_steps* (jendela *input*): 3
- *optimizer*: Adam
- *loss function*: Mean Squared Error (MSE)

4.1.5 Evaluasi Model

Evaluasi model dilakukan dengan tujuan untuk menilai sejauh mana model yang telah dibangun mampu menghasilkan prediksi yang akurat terhadap data harga saham berdasarkan input yang diberikan. Evaluasi dilakukan terhadap kedua jenis model yang digunakan, yaitu *LSTM* dan *XGBoost*, dengan penerapan pada tiga *ticker* saham: TINS, ISAT, dan MDKA.

Evaluasi dilakukan secara terpisah untuk data training, validasi, dan testing guna menilai konsistensi performa dan potensi *overfitting* atau *underfitting* dari masing-masing model.

a) Model *XGBoost*

Model *XGBoost* dievaluasi menggunakan metrik *MSE*, *RMSE*, *MAE*, dan R^2 pada data *training*, validasi, dan *testing*. Evaluasi ini dilakukan untuk masing-

masing saham, yaitu TINS, ISAT, dan MDKA, guna menilai kemampuan generalisasi model.

Selain itu, dilakukan *3-fold cross-validation* untuk memperoleh rata-rata nilai R^2 (*Mean R^2 CV*) sebagai indikator kestabilan model pada data yang berbeda.

Untuk mendukung analisis visual, ditambahkan *scatter plot* antara nilai aktual dan prediksi pada data uji, guna melihat kesesuaian antara hasil prediksi model dan nilai sebenarnya.

b) Model *LSTM*

Model *LSTM* dievaluasi menggunakan metrik *MSE*, *RMSE*, *MAE*, dan R^2 pada data training, validasi, dan testing untuk masing-masing saham (TINS, ISAT, dan MDKA). Evaluasi ini membantu mengukur sejauh mana model mampu memprediksi nilai secara akurat.

Selain metrik kuantitatif, dilakukan juga visualisasi *loss curve* (*training* dan *validation loss*) selama proses pelatihan untuk memantau konvergensi model dan potensi *overfitting*.

Rangkaian eksperimen telah dilakukan secara terstruktur, mulai dari pengumpulan dan *preprocessing* data, pemilihan fitur serta target, perancangan model, implementasi, hingga evaluasi. Kedua model yang digunakan, yaitu *LSTM* dan *XGBoost*, telah disesuaikan dengan karakteristik data *time series* saham serta diuji pada tiga emiten berbeda (TINS, ISAT, dan MDKA) untuk menguji konsistensi dan fleksibilitas model terhadap berbagai pola data.

Proses eksperimen juga mencakup pemilihan kombinasi parameter yang relevan serta evaluasi dengan berbagai metrik dan visualisasi pendukung. Secara umum, model yang digunakan menunjukkan tingkat kesesuaian terhadap dataset yang dimiliki peneliti, baik dari sisi struktur data, sekuensialitas, maupun hasil *preprocessing* yang telah diterapkan sebelumnya. Temuan dan pembahasan lebih lanjut mengenai performa masing-masing model akan dijelaskan pada subbab selanjutnya.

4.2 Hasil

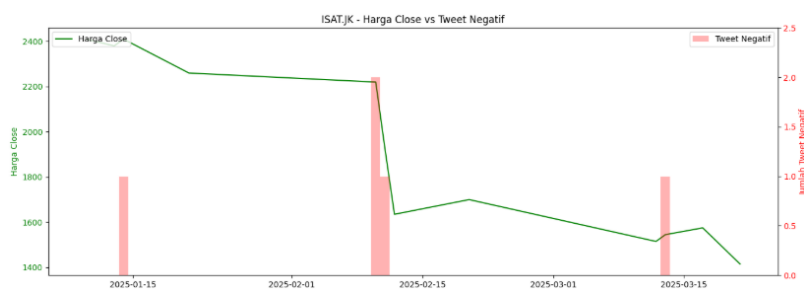
Bagian ini menyajikan hasil eksperimen yang mencakup analisis hubungan antara sentimen negatif di media sosial Twitter dengan pergerakan harga saham dari beberapa emiten, yakni ISAT.JK, MDKA.JK, dan TINS.JK. Selain visualisasi data yang menunjukkan pola keterkaitan tersebut, subbab ini juga membahas performa model

prediksi yang dikembangkan, sehingga dapat memberikan gambaran menyeluruh mengenai pengaruh sentimen terhadap harga saham serta efektivitas model dalam menangkap pola tersebut.

4.2.1 Keterkaitan Sentimen Negatif dan Harga Saham

Bagian ini menyajikan hasil eksperimen yang mencakup analisis hubungan antara sentimen negatif di media sosial Twitter dengan pergerakan harga saham dari beberapa emiten, yakni ISAT.JK, MDKA.JK, dan TINS.JK. Selain visualisasi data yang menunjukkan pola keterkaitan tersebut, subbab ini juga membahas performa model prediksi yang dikembangkan, sehingga dapat memberikan gambaran menyeluruh mengenai pengaruh sentimen terhadap harga saham serta efektivitas model dalam menangkap pola tersebut.

a) Saham ISAT.JK



Gambar 4. 9 Grafik ISAT.JK harga *close* vs tweet negatif

Gambar 4.8 menampilkan hubungan antara harga penutupan saham ISAT.JK dan jumlah tweet negatif dari Januari hingga Maret 2025. Garis hijau menunjukkan harga *Close*, sedangkan batang merah menunjukkan jumlah negatif_tweet per tanggal.

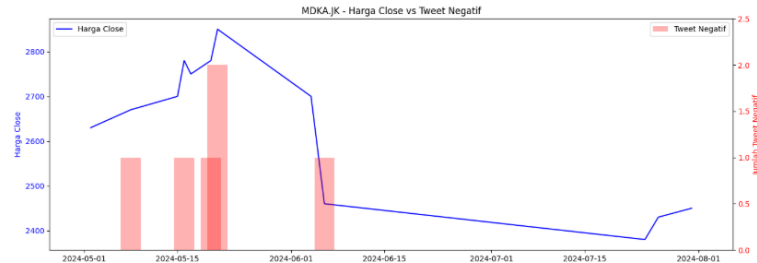
Tabel 4. 1 Data harga *close* & tweet negatif ISAT

Tanggal	Harga Penutupan (<i>Close</i>)	Tweet Negatif
2025-02-10	2220.0	2
2025-02-11	1920.0	1
2025-02-12	1635.0	0

Tabel 4.1 menyajikan nilai numerik *Close* dan negatif_tweet pada periode yang sama. Terlihat bahwa jumlah tweet negatif yang cukup tinggi pada 10 Februari diikuti penurunan harga saham dalam dua hari berikutnya. Hal ini mengindikasikan adanya potensi keterkaitan antara sentimen negatif di media

sosial dengan penurunan harga saham, meskipun tidak bersifat linier atau langsung.

b) Saham MDKA.JK



Gambar 4. 10 Grafik harga *close* vs negatif tweet MDKA.JK

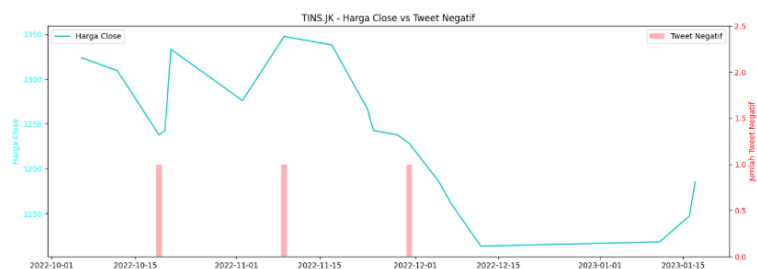
Gambar 4.9 memperlihatkan tren penurunan harga saham MDKA.JK dari bulan Mei 2024 hingga Juli 2024 yang diiringi dengan munculnya tweet negatif. Lonjakan sentimen negatif terutama terlihat menjelang titik terendah harga saham.

Tabel 4. 2 Data harga *close* & tweet negatif MDKA

Tanggal	Harga Penutupan (<i>Close</i>)	Tweet Negatif
2024-05-21	2850.0	2
2024-06-04	2700.0	0
2025-06-06	2460.0	1

Tabel 4.2 menyajikan nilai harga penutupan dan jumlah tweet negatif untuk saham MDKA.JK pada beberapa tanggal penting. Data ini memperkuat indikasi hubungan antara penurunan harga dan peningkatan tweet negatif.

c) Saham TINS.JK



Gambar 4. 11 Grafik harga *close* vs negatif tweet TINS.JK

Gambar 4.10 menunjukkan tren harga penutupan saham TINS.JK dari Oktober 2022 hingga Januari 2023, bersama dengan jumlah tweet negatif harian. Terlihat bahwa peningkatan jumlah tweet negatif mulai muncul saat harga saham mencapai puncaknya dan kemudian turun secara bertahap,

menunjukkan potensi hubungan antara sentimen negatif dan penurunan harga saham.

Tabel 4. 3 Data harga *close* & tweet negatif TINS

Tanggal	Harga Penutupan (<i>Close</i>)	Tweet Negatif
2022-11-28	1237.81	0
2022-11-30	1228.25	1
2022-12-05	1185.24	0

Tabel 4.3 menampilkan data numerik harga penutupan saham dan jumlah tweet negatif pada periode November-Desember 2022. Analisis ini kembali memperlihatkan adanya potensi keterkaitan antara sentimen negatif dan penurunan harga saham, meskipun tidak selalu linier atau instan. Tercatat lonjakan tweet negatif pada akhir November 2022, bertepatan dengan dimulainya tren penurunan harga dari Rp1.228 ke Rp1.185.

Berdasarkan hasil analisis data harga saham dan jumlah tweet negatif dari tiga emiten (ISAT.JK, MDKA.JK, dan TINS.JK), terlihat adanya indikasi hubungan antara meningkatnya sentimen negatif di media sosial dengan penurunan harga saham. Meskipun hubungan tersebut tidak bersifat linier atau langsung, pola yang muncul menunjukkan bahwa lonjakan tweet negatif cenderung berdekatan dengan fase penurunan harga. Hal ini menunjukkan bahwa sentimen negatif di Twitter dapat menjadi salah satu indikator yang perlu diperhatikan dalam analisis pergerakan harga saham, khususnya dalam konteks peringatan dini terhadap potensi penurunan nilai.

4.2.2 Evaluasi Hasil Model XGBoost

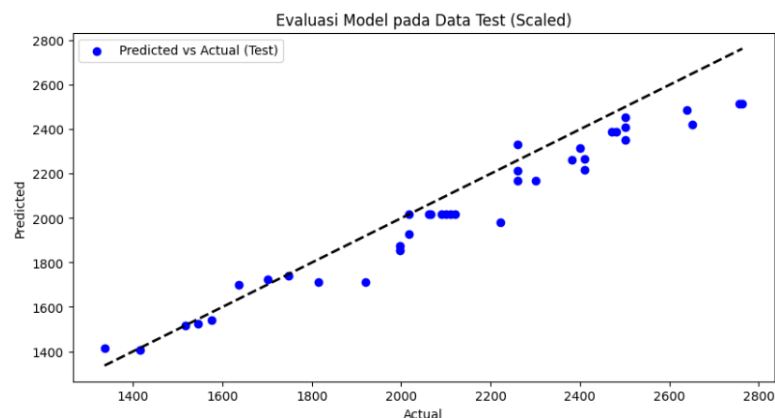
a) Saham ISAT.JK

Tabel 4. 4 Hasil evaluasi model *XGBoost* pada saham ISAT

Parameter	Saham ISAT					
<i>n_estimators</i>	60	60	70	70	60	60
<i>max_depth</i>	2	2	2	2	2	2
<i>learning_rate</i>	0.05	0.04	0.04	0.04	0.05	0.05
<i>subsample</i>	0.7	0.7	0.6	0.7	0.7	0.7
<i>colsample_bytree</i>	0.8	0.6	0.8	0.6	0.6	0.8
<i>reg_alpha</i>	0	0	0	0	0.1	0.1

<i>reg_lambda</i>	0	0	0	0	0	0.1
<i>MSE</i>	14897.17	29365.73	20268.06	17822.07	16349.10	14729.17
<i>RMSE</i>	122.05	171.36	141.37	133.50	127.86	121.36
<i>MAE</i>	100.61	143.73	121.79	113.18	107.04	100.31
<i>R² Score</i>	89.35%	79.00%	85.5%	97.26%	88.31%	89.47%
<i>Val R² Score</i>	-4.06	-12.76	-5.64	-5.60	-6.77	-4.92

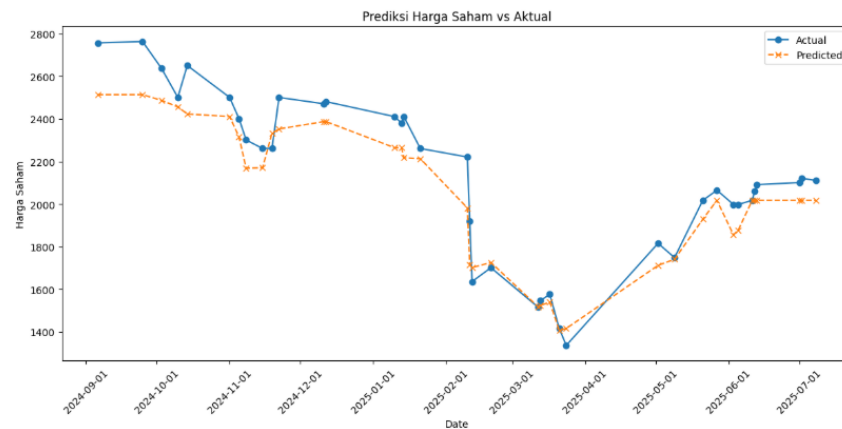
Berdasarkan table 4.4 hasil tuning model *XGBoost* untuk saham ISAT, seluruh model menunjukkan performa yang sangat baik pada data *training* (R^2 hingga 97.26%), namun gagal melakukan generalisasi terhadap data validasi. Hal ini ditunjukkan oleh nilai *Val R²* yang negatif pada seluruh konfigurasi model, menandakan adanya overfitting yang serius. Model dengan parameter $n_estimators = 60$, $learning_rate = 0.05$, dan $reg_lambda = 0.1$ memberikan error terendah ($RMSE = 121.36$), namun *Val R²* tetap negatif (-4.92). Maka dari itu, model ini dipilih sebagai model sementara terbaik, meskipun perlu dilakukan optimasi lanjutan untuk meningkatkan kemampuan generalisasi model terhadap data baru.



Gambar 4. 12 Scarterplot saham ISAT

Gambar 4.12 merupakan scatter plot antara nilai aktual dan nilai prediksi harga saham ISAT pada data uji. Setiap titik merepresentasikan hasil prediksi model terhadap harga aktual pada satu waktu tertentu. Garis putus-putus hitam menunjukkan garis ideal $y = x$, yaitu di mana hasil prediksi sempurna akan berada. Sebagian besar titik berada cukup dekat dengan garis identitas, menunjukkan bahwa model berhasil menghasilkan prediksi yang cukup akurat. Namun, terdapat beberapa deviasi, terutama pada kisaran harga menengah dan tinggi, yang menandakan adanya sedikit perbedaan antara hasil prediksi dan nilai aktual. Secara keseluruhan, plot ini menunjukkan bahwa model memiliki kecenderungan yang

baik dalam mengikuti pola harga aktual, dengan penyimpangan yang masih dalam batas wajar.



Gambar 4. 13 Grafik prediksi saham ISAT

Gambar 4.13 menunjukkan perbandingan antara harga saham aktual dan hasil prediksi untuk saham ISAT selama periode waktu tertentu. Secara umum, model mampu mengikuti pola pergerakan harga saham dengan cukup baik, terutama pada tren penurunan di awal 2025 dan tren kenaikan menjelang pertengahan tahun. Meskipun terdapat beberapa deviasi antara prediksi dan nilai aktual, terutama pada fase penurunan tajam dan lonjakan harga, garis prediksi tetap berada dalam arah yang serupa dengan data aktual. Hal ini menunjukkan bahwa model memiliki performa prediksi yang cukup stabil dan mampu merepresentasikan pergerakan harga saham secara umum.

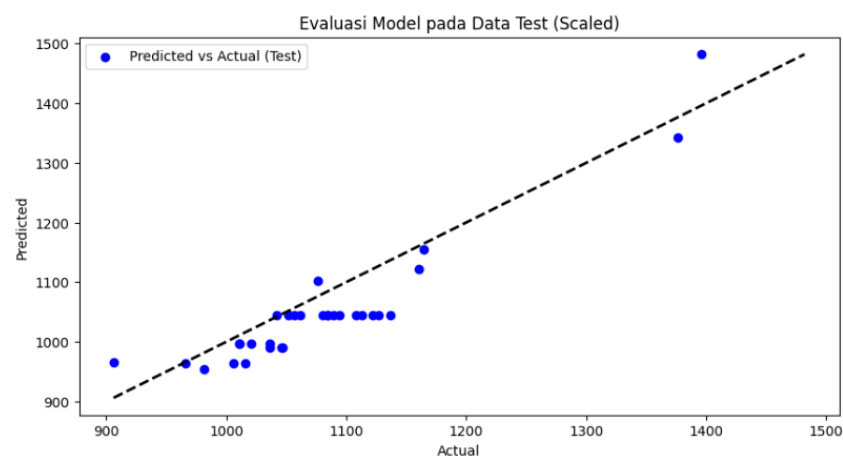
b) Saham TINS.JK

Tabel 4. 5 Hasil evaluasi model *XGBoost* pada saham TINS

Parameter	Saham TINS					
<i>n_estimators</i>	60	60	60	70	70	70
<i>max_depth</i>	2	2	3	3	2	2
<i>learning_rate</i>	0.05	0.05	0.04	0.04	0.04	0.05
<i>subsample</i>	0.7	0.7	0.7	0.7	0.7	0.07
<i>colsample_bytree</i>	0.8	0.8	0.8	0.8	0.8	0.8
<i>reg_alpha</i>	0.1	0.1	0.1	0.1	0.1	0.1
<i>reg_lambda</i>	0.1	0.1	0.1	0.1	0.1	0.1
<i>MSE</i>	1660.49	1925.86	1400.90	1906.58	1667.55	2235.87
<i>RMSE</i>	40.75	43.88	37.43	43.66	40.84	47.28
<i>MAE</i>	32.75	36.45	28.28	36.18	32.40	40.42

R^2 Score	82.57%	79.79%	85.30%	79.99%	82.50%	76.53%
Val R^2 Score	0.69	0.66	0.61	0.63	0.67	0.71

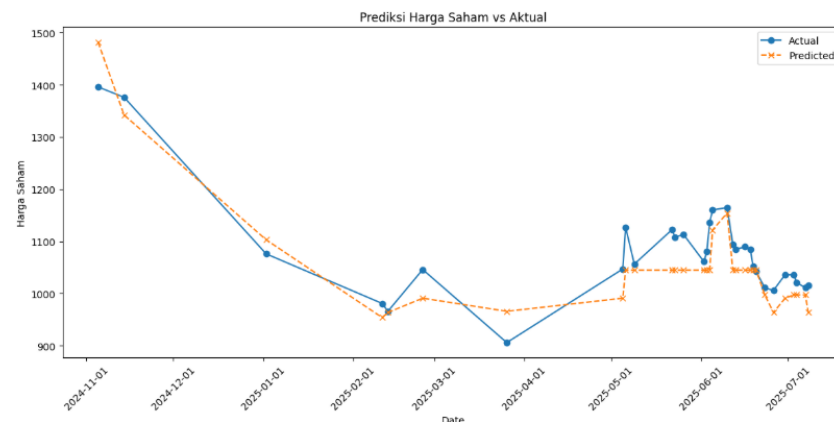
Berdasarkan tabel 4.5 hasil tuning model *XGBoost* untuk saham TINS, model terbaik dipilih dengan mempertimbangkan kemampuan generalisasi terhadap data baru, yang tercermin dari nilai *validation R² Score* (Val R^2). Meskipun model ke-3 memiliki R^2 Score tertinggi pada data pelatihan sebesar 85.30%, nilai Val R^2 -nya hanya 0.61, yang menunjukkan adanya potensi *overfitting*. Sebaliknya, model ke-6 menunjukkan performa yang lebih stabil, dengan Val R^2 tertinggi sebesar 0.71, meskipun R^2 pada data pelatihannya lebih rendah, yaitu 76.53%. Hal ini mengindikasikan bahwa model ke-6 lebih mampu melakukan generalisasi dan memprediksi data yang belum pernah dilihat sebelumnya secara lebih akurat. Oleh karena itu, model ke-6 lebih direkomendasikan untuk digunakan dalam implementasi di *dashboard*, karena menawarkan keseimbangan yang lebih baik antara akurasi dan stabilitas prediksi terhadap data baru.



Gambar 4. 14 *Scatterplot* saham TINS

Gambar 4.14 menunjukkan *scatter plot* antara harga aktual dan prediksi saham TINS pada data uji. Titik-titik biru merepresentasikan hasil prediksi dibandingkan dengan nilai sebenarnya, sedangkan garis putus-putus hitam menunjukkan garis ideal di mana prediksi = aktual. Sebagian besar titik memang berada cukup dekat dengan garis, namun terlihat adanya kumpulan titik yang stagnan di sekitar angka 1050 pada sumbu prediksi. Hal ini mengindikasikan bahwa model cenderung memberikan prediksi yang relatif seragam pada beberapa data, sehingga kurang sensitif terhadap variasi harga aktual. Meski demikian, arah penyebaran titik masih mengikuti pola umum, yang menunjukkan bahwa model tetap mampu mempelajari

hubungan antara input dan output secara dasar, namun perlu peningkatan agar lebih presisi.



Gambar 4. 15 Grafik prediksi saham TINS

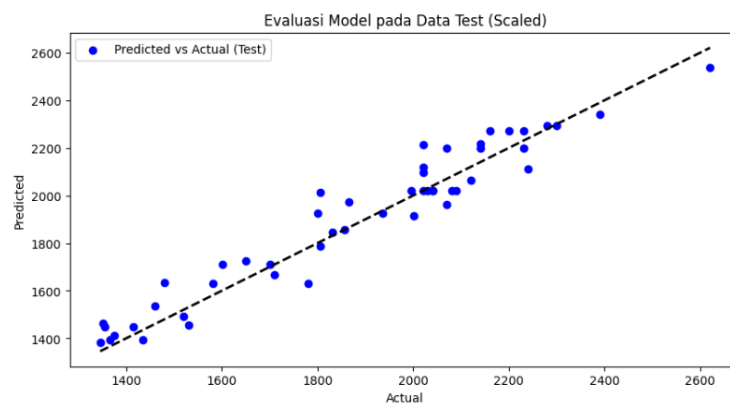
Gambar 4.15 memperlihatkan perbandingan antara harga saham aktual dan hasil prediksi model untuk saham TINS dalam rentang waktu tertentu. Secara umum, model berhasil mengikuti arah tren pergerakan harga, terutama pada fase penurunan di awal 2025. Namun, pada beberapa bagian seperti periode Mei hingga Juli 2025, prediksi terlihat kurang responsif terhadap fluktuasi tajam pada harga aktual. Hal ini menunjukkan bahwa model masih cenderung menghasilkan prediksi yang stabil namun kurang menangkap perubahan mendadak, meskipun tetap mampu memberikan gambaran tren secara keseluruhan.

c) Saham MDKA.JK

Tabel 4. 6 Hasil evaluasi model *XGBoost* pada saham MDKA

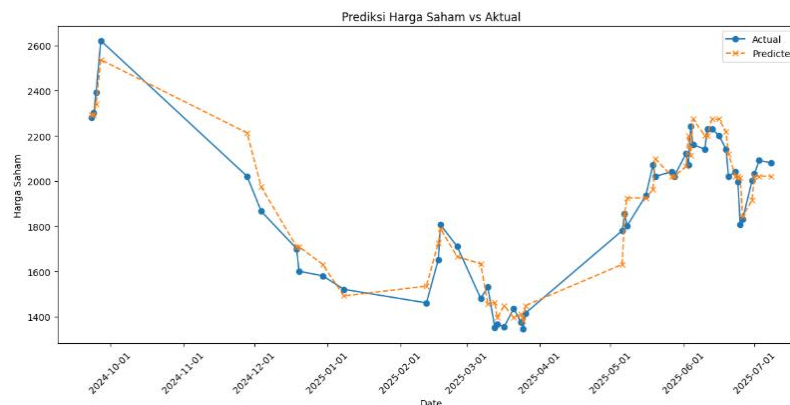
Parameter	Saham MDKA					
<i>n_estimators</i>	60	60	70	70	70	80
<i>max_depth</i>	2	2	2	3	3	3
<i>learning_rate</i>	0.4	0.05	0.05	0.04	0.04	0.04
<i>subsample</i>	0.8	0.8	0.8	0.8	0.7	0.8
<i>colsample_bytree</i>	0.8	0.8	0.8	0.8	0.7	0.6
<i>reg_alpha</i>	0.1	0.1	0.1	0.1	0.1	0.1
<i>reg_lambda</i>	0.1	0.1	0.1	0.1	0.1	0.1
<i>MSE</i>	8991.41	7288.06	7016.16	6868.41	7185.37	6962.44
<i>RMSE</i>	94.82	86.32	83.76	82.88	84.77	83.44
<i>MAE</i>	73.76	66.38	65.55	66.58	66.94	67.12
<i>R² Score</i>	91.03%	92.73%	93.00%	93.15%	92.83%	93.05%
<i>Val R² Score</i>	0.87	0.85	0.86	0.85	0.86	0.85

Dalam pemilihan model terbaik untuk saham MDKA, evaluasi tidak hanya berfokus pada nilai akurasi di data pelatihan (R^2 Score), tetapi juga pada kemampuan model dalam melakukan generalisasi terhadap data baru, yang tercermin dari nilai $Val R^2$ Score dan kesalahan prediksi ($RMSE$). Dari tujuh kandidat model yang diuji, model ke-4 dengan parameter $n_estimators=70$, $max_depth=3$, dan $learning_rate=0.04$ menunjukkan performa paling seimbang. Model ini memiliki R^2 Score sebesar 93.15%, $RMSE$ terendah sebesar 82.88, serta $Val R^2$ Score tinggi sebesar 0.86. Meskipun model ke-1 memiliki $Val R^2$ tertinggi (0.87), kesalahan prediksinya ($RMSE = 94.82$) jauh lebih besar dibanding model ke-4. Oleh karena itu, model ke-4 lebih direkomendasikan karena mampu memberikan prediksi yang lebih akurat dan stabil terhadap data baru yang belum pernah dilihat sebelumnya, sehingga lebih andal untuk diterapkan pada dashboard atau sistem prediksi nyata.



Gambar 4. 16 *Scartterplot* saham MDKA

Gambar 4.16 menunjukkan *scatter plot* antara harga aktual dan prediksi saham MDKA pada data uji. Sebagian besar titik berada dekat dengan garis identitas (putus-putus hitam), yang berarti hasil prediksi mendekati nilai sebenarnya. Pola sebaran yang mengikuti garis tersebut mengindikasikan bahwa model memiliki akurasi yang baik dan generalisasi yang stabil. Meskipun ada beberapa deviasi, terutama pada nilai tinggi, secara keseluruhan model menunjukkan performa yang cukup andal dalam memprediksi harga saham MDKA.



Gambar 4. 17 Grafik prediksi saham MDKA

Gambar 4.17 menunjukkan perbandingan antara harga saham aktual dan hasil prediksi model untuk saham MDKA. Secara umum, model mampu mengikuti pola pergerakan harga dengan baik, terutama saat terjadi penurunan tajam di awal 2025 dan kenaikan signifikan menjelang pertengahan tahun. Meskipun terdapat sedikit perbedaan pada beberapa titik, garis prediksi tetap berada di jalur tren yang serupa dengan data aktual. Hal ini menunjukkan bahwa model memiliki performa yang cukup akurat dan stabil dalam menangkap dinamika harga saham MDKA.

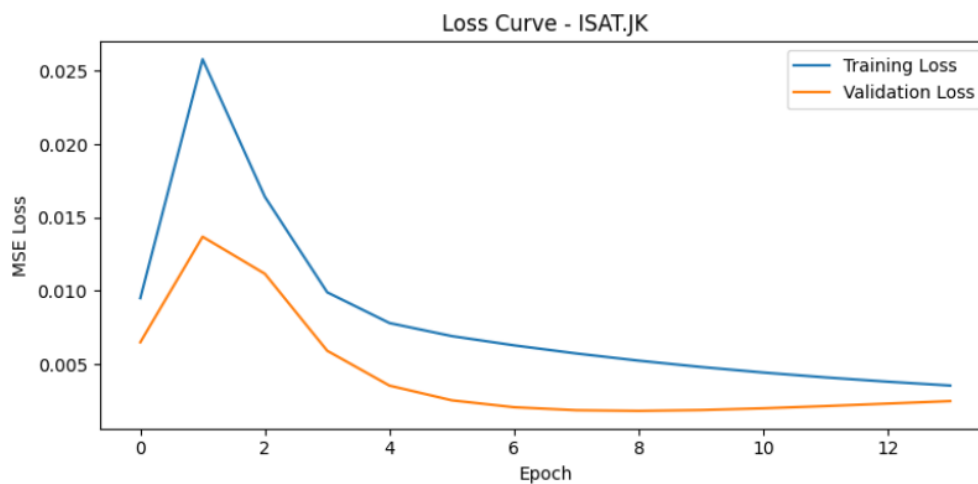
4.2.3 Evaluasi Hasil Model LSTM

Tabel 4. 7 Evaluasi model *LSTM* dari ketiga saham

Parameter	Saham		
	ISAT	MDKA	TINS
<i>unit</i>	32	32	32
<i>window</i>	3	3	3
<i>batch_size</i>	4	4	4
<i>epoch</i>	14	19	100
<i>optimizer</i>	Adam	Adam	Adam
<i>time_step</i>	1	1	1
<i>MSE</i>	25297.57	0.01	0.01
<i>RMSE</i>	159.05	0.09	0.12
<i>MAE</i>	133.80	0.07	0.08
<i>R² Score</i>	0.84	0.79	0.43
<i>Val R² Score</i>	-2.20	0.36	72

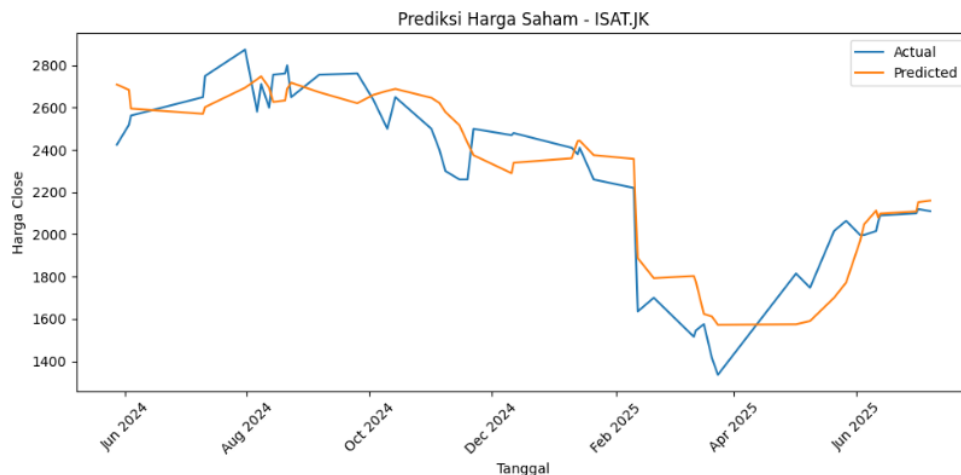
a) Saham ISAT.JK

Model *LSTM* untuk saham ISAT menggunakan 3 *timestep* (*window*) dan 32 *unit neuron*, serta dioptimasi dengan algoritma Adam dan *batch size* 4. Meskipun model menunjukkan performa sangat baik pada data latih dengan nilai R^2 sebesar 0.84, hasil pada data validasi menunjukkan indikasi *overfitting* yang cukup parah, ditandai dengan nilai *Val R²* sebesar -2.20 dan nilai *error* yang tinggi (*MSE* 25297.57, *RMSE* 159.05, *MAE* 133.80). Hal ini menunjukkan bahwa model terlalu menyesuaikan diri terhadap pola data latih namun gagal memprediksi data baru. Perlu dilakukan perbaikan dengan menambahkan teknik regularisasi seperti dropout atau memperluas jumlah data latih untuk meningkatkan generalisasi model.



Gambar 4. 18 *Loss Curva* model *LSTM* di saham ISAT

Gambar 4.18 *Loss Curve* untuk ISAT.JK menunjukkan perbandingan antara nilai *training loss* dan *validation loss* sepanjang proses pelatihan model *LSTM*. Terlihat bahwa pada awal *epoch* (khususnya *epoch* ke-1), nilai *loss* sempat meningkat tajam pada data latih, namun kemudian menurun secara konsisten. Sementara itu, *validation loss* mengalami penurunan yang cepat dan stabil, lalu mulai mendatar hingga akhir pelatihan. Meskipun tren ini tampak baik, perbedaan nilai antara *training* dan *validation loss* cukup besar di akhir *epoch*, yang mengindikasikan potensi *overfitting* di mana model terlalu menyesuaikan data latih namun kurang efektif memprediksi data baru. Hal ini sejalan dengan nilai *Val R²* yang negatif (-2.20) yang telah ditemukan sebelumnya. Grafik ini menegaskan perlunya perbaikan model, seperti regularisasi atau penambahan data latih.

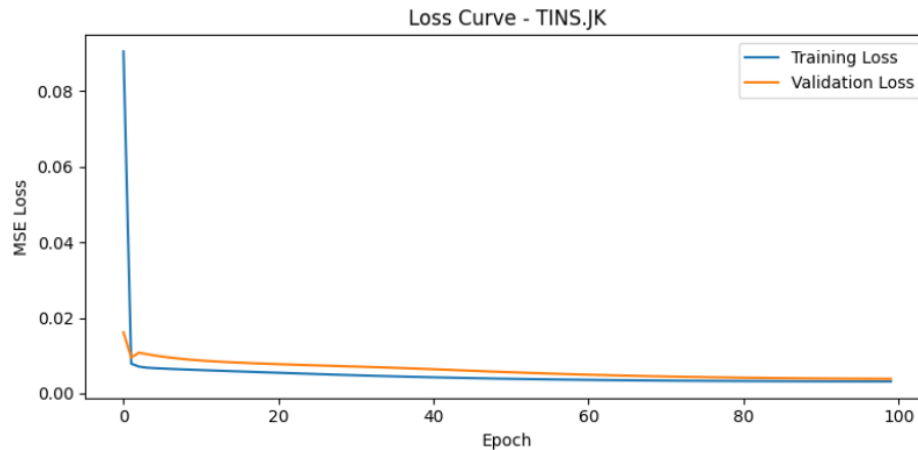


Gambar 4. 19 Grafik prediksi model *LSTM* di saham ISAT

Gambar 4.19 grafik prediksi harga saham ISAT.JK menunjukkan perbandingan antara harga aktual (garis biru) dan hasil prediksi model *LSTM* (garis oranye) selama periode waktu tertentu. Terlihat bahwa model mampu mengikuti arah tren umum pergerakan harga, terutama pada periode naik di pertengahan 2024 dan pertengahan 2025. Namun, pada periode penurunan tajam di awal 2025, model gagal menangkap perubahan drastis dan cenderung memprediksi nilai yang lebih stabil atau datar. Hal ini menunjukkan bahwa model kurang sensitif terhadap fluktuasi tajam dalam data harga saham, kemungkinan akibat *overfitting* atau keterbatasan jumlah *window (timestep)* yang hanya 3. Secara keseluruhan, meskipun model bisa mengikuti pola umum, akurasi prediksi untuk pergerakan ekstrem masih perlu ditingkatkan.

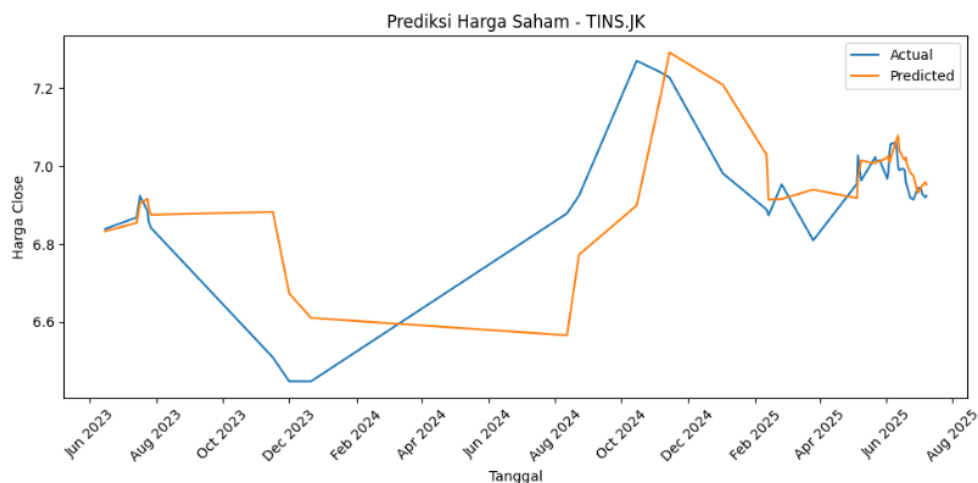
b) Saham TINS.JK

Model *LSTM* untuk saham TINS berhasil menunjukkan kemampuan generalisasi yang baik, meskipun performa pada data latih tidak setinggi model lainnya. Model dilatih hingga 100 *epoch* dan menghasilkan nilai R^2 sebesar 0.43 pada data latih, namun memberikan hasil validasi yang sangat memuaskan dengan *Val R^2* sebesar 0.72. Nilai *error* yang dihasilkan juga rendah (*MSE* 0.01, *RMSE* 0.12, *MAE* 0.08), yang mengindikasikan bahwa model mampu mempelajari pola-pola penting dalam data tanpa berlebihan menyesuaikan terhadap data latih. Model ini cenderung lebih optimal dalam menghadapi data baru dan memiliki potensi yang baik untuk digunakan dalam prediksi harga saham jangka pendek.



Gambar 4. 20 *Loss Curve* model *LSTM* di saham TINS

Gambar 4.20 grafik *Loss Curve* untuk saham TINS.JK menunjukkan proses pelatihan model *LSTM* yang berjalan dengan sangat stabil. Baik *training loss* maupun *validation loss* mengalami penurunan tajam di awal *epoch* dan terus menurun secara konsisten hingga akhir pelatihan di *epoch* ke-100. Kedua garis loss juga berada cukup dekat, menandakan bahwa model tidak mengalami *overfitting* dan memiliki kemampuan generalisasi yang baik terhadap data baru. Stabilitas dan konvergensi yang baik ini sejalan dengan hasil evaluasi sebelumnya, di mana model TINS memiliki nilai *Val R²* sebesar 0.72, menunjukkan bahwa model cukup andal dalam memprediksi harga saham pada data validasi. Dengan demikian, model ini layak dipertimbangkan untuk digunakan dalam prediksi jangka pendek.



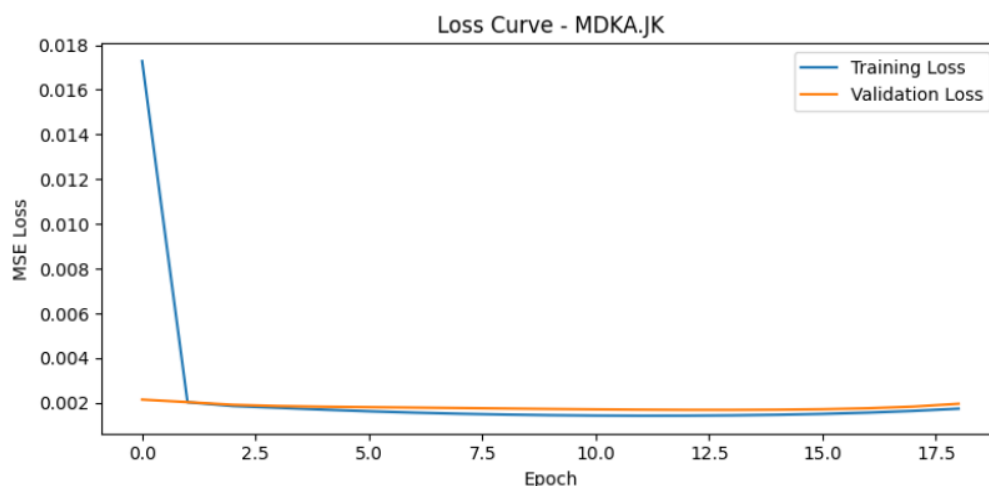
Gambar 4. 21 Grafik prediksi model *LSTM* di saham TINS

Gambar 4.21 grafik prediksi harga saham TINS.JK menunjukkan bahwa model *LSTM* mampu mengikuti pola tren harga secara umum dengan cukup baik. Garis

oranye (hasil prediksi) menunjukkan kemiripan arah dengan garis biru (harga aktual), terutama pada periode tren naik dan turun yang lebih halus. Namun, terdapat beberapa momen di mana model kurang responsif terhadap perubahan harga yang tajam, seperti pada lonjakan signifikan di akhir 2024 yang diprediksi terlalu tinggi dibandingkan harga sebenarnya. Meskipun begitu, model tetap menunjukkan performa yang cukup stabil dan tidak terlalu menyimpang drastis dari nilai aktual dalam sebagian besar periode. Hal ini konsisten dengan hasil evaluasi sebelumnya yang menunjukkan $Val R^2$ sebesar 0.72, menandakan bahwa model *LSTM* untuk saham TINS memiliki kemampuan generalisasi yang baik dan layak untuk digunakan dalam prediksi harga jangka pendek.

c) Saham MDKA.JK

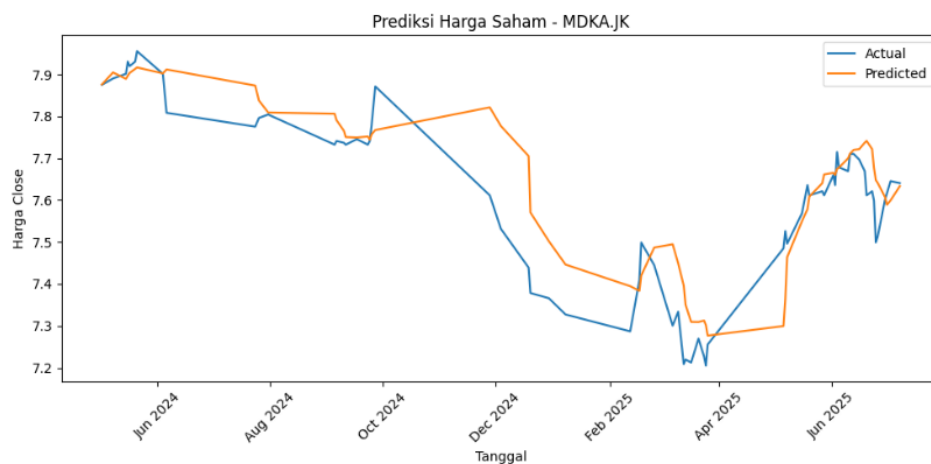
Model *LSTM* pada saham MDKA menunjukkan performa yang cukup baik dan stabil. Dengan parameter yang sama seperti ISAT ($timestep = 3$, $unit = 32$), model berhenti pada *epoch* ke-19 karena *early stopping*, dan menghasilkan nilai R^2 pada data latih sebesar 0.79 serta $Val R^2$ sebesar 0.36. Nilai error-nya juga sangat rendah (MSE 0.01, $RMSE$ 0.09, MAE 0.07), menunjukkan bahwa model mampu memahami pola pergerakan harga saham MDKA dengan baik dan tetap memiliki kemampuan generalisasi terhadap data baru. Model ini tergolong cukup andal dan layak dipertahankan untuk kebutuhan prediksi.



Gambar 4. 22 *Loss Curva model LSTM* di saham MDKA

Gambar 4.22 grafik *Loss Curve* untuk saham MDKA.JK menunjukkan bahwa model *LSTM* mengalami penurunan *MSE loss* yang sangat cepat pada awal pelatihan, kemudian stabil pada level rendah baik untuk data latih maupun data validasi. Garis

training loss dan *validation loss* terlihat sangat dekat sepanjang *epoch*, tanpa adanya perbedaan mencolok ataupun pola divergen. Kondisi ini menunjukkan bahwa model memiliki fit yang baik dan seimbang, tidak mengalami *overfitting* maupun *underfitting*. Stabilitas *loss* pada nilai yang sangat rendah juga sejalan dengan hasil evaluasi sebelumnya, di mana error sangat kecil dan *Val R²* cukup baik (0.36). Dengan demikian, model *LSTM* untuk saham MDKA dapat dianggap cukup akurat dan andal dalam memprediksi harga saham di data baru.



Gambar 4. 23 Grafik prediksi model *LSTM* di saham MDKA

Gambar 4.23 grafik prediksi harga saham MDKA.JK menunjukkan bahwa model *LSTM* mampu mengikuti arah tren umum pergerakan harga dengan cukup baik, terutama pada periode pertengahan 2024 hingga pertengahan 2025. Garis prediksi (oranye) terlihat cukup dekat dengan garis harga aktual (biru), meskipun terdapat beberapa momen di mana model terlambat atau kurang tepat dalam menangkap perubahan tajam, khususnya saat harga mengalami penurunan drastis pada awal 2025. Secara keseluruhan, model menunjukkan kemampuan prediksi yang stabil, dengan akurasi yang cukup memadai untuk memetakan pola jangka pendek maupun menengah. Hal ini konsisten dengan nilai *error* yang rendah dan *Val R²* sebesar 0.36, yang menunjukkan bahwa model cukup andal meskipun belum sepenuhnya menangkap fluktuasi ekstrem secara presisi.

4.2.4 Perbandingan Kinerja Model

Dalam penelitian ini, dilakukan evaluasi terhadap dua model prediksi harga saham, yaitu *LSTM* dan *XGBoost*, dengan penerapan pada tiga saham berbeda: ISAT.JK, TINS.JK, dan MDKA.JK. Hasil evaluasi menunjukkan bahwa masing-masing model

memiliki karakteristik dan performa yang berbeda, tergantung pada kompleksitas dan pola data saham yang dianalisis.

Saham	Model	R^2 Train	Val R^2	RMSE	Karakteristik Prediksi	Kesimpulan
ISAT.JK	<i>XGBoost</i>	89%	-4.92	121.36	<i>Overfitting</i> berat, tren umum diikuti, tapi prediksi tidak tajam	Perlu perbaikan, tidak stabil
	<i>LSTM</i>	84%	-2.20	159.05	<i>Overfitting</i> , kurang tangkap fluktuasi tajam	Belum layak pakai
TINS.JK	<i>XGBoost</i>	76%	0.71	47.28	Stabil, prediksi cukup akurat meskipun kurang balus	Layak digunakan
	<i>LSTM</i>	43%	0.72	0.12	Prediksi halus, ikuti tren dengan baik	Sama baiknya, unggul di tren
MDKA.JK	<i>XGBoost</i>	93%	0.85	82.88	Akurat, stabil, <i>responsive</i> terhadap perubahan	Paling direkomendasikan
	<i>LSTM</i>	79%	0.36	0.09	Cenderung rata, kurang akurat pada fluktuasi	Kurang cocok untuk MDKA

Berdasarkan hasil evaluasi, *XGBoost* dan *LSTM* memiliki performa yang berbeda tergantung pada saham yang dianalisis. Untuk ISAT.JK, keduanya mengalami

overfitting, ditandai dengan *Val R²* yang negatif meskipun akurasi pelatihan tinggi, sehingga belum direkomendasikan tanpa perbaikan lebih lanjut. Pada TINS.JK, kedua model menunjukkan kemampuan prediksi yang baik dan setara; *XGBoost* lebih stabil, sementara *LSTM* lebih halus dalam mengikuti tren. Sementara itu, untuk MDKA.JK, *XGBoost* jauh lebih unggul dengan akurasi validasi tinggi dan prediksi yang akurat terhadap perubahan tajam, menjadikannya model paling direkomendasikan.

4.2.5 Pemilihan Model untuk Implementasi *Dashboard*

Berdasarkan hasil evaluasi terhadap tiga saham (ISAT.JK, TINS.JK, dan MDKA.JK), model *XGBoost* dipilih karena menunjukkan performa yang lebih stabil dan mampu melakukan generalisasi dengan baik dibandingkan *LSTM*. Meskipun *LSTM* unggul dalam memproses data berurutan, hasil validasinya cenderung tidak konsisten dan rentan *overfitting*. *XGBoost* memberikan hasil terbaik terutama pada saham MDKA dan TINS, serta tetap lebih terkendali pada ISAT. Oleh karena itu, model ini digunakan sebagai dasar dalam pembuatan *dashboard* prediksi harga saham, karena selain akurat, juga mudah diintegrasikan dan diimplementasikan secara efisien.

Model *XGBoost* yang telah dipilih diimplementasikan dalam bentuk *dashboard* prediksi harga saham. *Dashboard* ini dirancang agar mudah digunakan oleh pengguna awam maupun analis, dengan tampilan antarmuka yang sederhana namun informatif.

Prediksi Harga Saham

Masukkan Data Saham

Tanggal Hari Ini
2025/07/22

TINS.JK
Open TINS_JK
1200
High TINS_JK
1225
Negatif Tweet TINS_JK
2

MDKA.JK
Open MDKA_JK
1300
High MDKA_JK
1345
Negatif Tweet MDKA_JK
2

ISAT.JK
Open ISAT_JK
1360
High ISAT_JK
1400
Negatif Tweet ISAT_JK
2

Prediksi

Gambar 4. 24 *Dashboard* Prediksi Harga Saham

Tampilan *dashboard* gambar 4.24 merupakan antarmuka pengguna untuk sistem prediksi harga saham berbasis model *XGBoost*. Pengguna cukup mengisi data input harian yang dibutuhkan untuk masing-masing saham, yaitu TINS.JK, MDKA.JK, dan ISAT.JK. Data yang dimasukkan berupa:

- Harga *Open*
- Harga *High*
- Jumlah tweet negatif (sentimen negatif dari Twitter)

Selain itu, pengguna juga diminta memilih tanggal prediksi (secara otomatis menyesuaikan dengan hari ini). Setelah semua data diisi, pengguna dapat menekan tombol "Prediksi", dan sistem akan menampilkan hasil prediksi harga penutupan saham (*Close*) untuk masing-masing emiten berdasarkan model *XGBoost* yang telah dilatih sebelumnya. *Dashboard* ini dirancang agar sederhana, responsif, dan mudah digunakan, bahkan oleh pengguna non-teknis. Implementasi seperti ini sangat cocok untuk mendukung keputusan investor atau analis pasar dalam memperkirakan pergerakan harga saham secara harian.

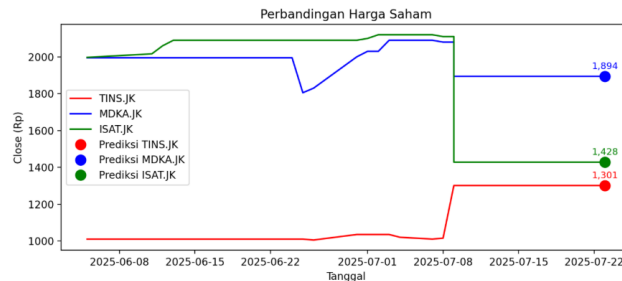
Hasil Prediksi

TINS.JK: Rp 1.301 (Naik 28.18%)

MDKA.JK: Rp 1.894 (Turun 8.94%)

ISAT.JK: Rp 1.428 (Turun 32.32%)

Perbandingan Harga Saham (7 Data Terakhir)



Gambar 4. 25 Tampilan *dashboard* hasil prediksi

Gambar 4.25 merupakan tampilan hasil prediksi yang dihasilkan oleh dashboard. Dalam contoh ini, prediksi dilakukan terhadap tiga saham, yaitu TINS.JK, MDKA.JK, dan ISAT.JK. Masing-masing saham menampilkan hasil prediksi harga penutupan berdasarkan data yang diinput oleh pengguna, seperti harga *Open*, *High*, dan jumlah tweet negatif.

Hasil prediksi ditampilkan secara langsung bersama informasi persentase kenaikan atau penurunan dibandingkan data sebelumnya. Selain itu, ditampilkan pula grafik perbandingan harga saham selama tujuh data terakhir, termasuk titik prediksi yang ditandai secara jelas. Tampilan ini akan berubah secara otomatis sesuai dengan data terbaru yang dimasukkan oleh pengguna, sehingga dapat digunakan untuk memantau dan memperkirakan pergerakan harga saham secara praktis dan cepat.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan beberapa hal sebagai berikut:

1. Sentimen negatif dari media sosial, khususnya Twitter, memiliki pengaruh terhadap pergerakan harga saham pada waktu-waktu tertentu. Meskipun demikian, hubungan antara sentimen negatif dan harga saham secara keseluruhan tergolong lemah dan belum konsisten di semua saham. Namun, kehadiran data sentimen tetap memberikan kontribusi penting sebagai informasi tambahan yang dapat memperkaya konteks dari data historis yang digunakan dalam prediksi harga saham.
2. Data sentimen diperoleh dari jumlah tweet negatif harian yang kemudian diolah sebagai fitur tambahan, bersama dengan data harga saham historis, dalam proses pelatihan model prediksi menggunakan algoritma *XGBoost*. Model ini selanjutnya diterapkan ke dalam sebuah *dashboard* interaktif yang memudahkan pengguna dalam memantau dan memprediksi pergerakan harga saham berdasarkan gabungan data historis dan sentimen.
3. Hasil pengujian menunjukkan bahwa integrasi antara data sentimen dan data historis mampu meningkatkan kualitas prediksi harga saham. Model *XGBoost* menunjukkan kinerja yang lebih stabil dan akurat dibandingkan dengan model *LSTM*, serta lebih tepat untuk diimplementasikan dalam sistem prediksi yang berorientasi pada pengguna. Dengan pendekatan ini, proses analisis pasar dan pengambilan keputusan investasi dapat dilakukan dengan lebih baik dan relevan terhadap kondisi aktual di lapangan.

5.2 Saran

- 1) Penggunaan data sentimen sebaiknya dilengkapi dengan lebih banyak variasi sumber, seperti berita keuangan, forum diskusi saham, atau media sosial lainnya, agar hasil analisis sentimen menjadi lebih kuat dan representatif.

- 2) Model prediksi dapat ditingkatkan dengan menambahkan fitur tambahan seperti indikator teknikal (*moving average*, *RSI*, *volume* tren) atau faktor makroekonomi untuk memperkuat akurasi prediksi harga saham.
- 3) Perlu dilakukan analisis sentimen yang lebih mendalam, misalnya dengan mengelompokkan tweet berdasarkan sektor atau topik tertentu, agar pengaruh sentimen terhadap saham menjadi lebih terfokus dan akurat.

DAFTAR PUSTAKA

- Afrianto, and Nurdi. 2022. “Prediksi Harga Saham Menggunakan BiLSTM Dengan Faktor Sentimen Publik.” *Resti* 1(1):41–46.
- Alfarizi, M. Riziq Sirfatullah, Muhamad Zidan Al-farish, Muhamad Taufiqurrahman, Ginan Ardiansah, and Muhamad Elgar. 2023. “Penggunaan Python Sebagai Bahasa Pemrograman Untuk Machine Learning Dan Deep Learning.” *Karya Ilmiah Mahasiswa Bertauhid (KARIMAH TAUHID)* 2(1):1–6.
- Apriliansah, Lalu. 2024. “Analisis Pengaruh Investasi Terhadap Pertumbuhan Ekonomi.” *JICN: Jurnal Intelek Dan Cendekiawan Nusantara* 1(2):2401–13.
- Arsi, Primandani, and Retno Waluyo. 2021. “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (SVM).” *Jurnal Teknologi Informasi Dan Ilmu Komputer* 8(1):147. doi: 10.25126/jtiik.0813944.
- Azahri, Mutiara, Nina Sulistiyowati, and Mohamad Jajuli. 2023. “Analisis Sentimen Pengguna Kereta Api Indonesia Melalui Sosial Media Twitter Dengan Algoritma Naïve Bayes Classifier.” *JATI (Jurnal Mahasiswa Teknik Informatika)* 7(3):1671–75. doi: 10.36040/jati.v7i3.6886.
- Chairunisa, Ghevira, Mohamad Khoirun Najib, Sri Nurdiati, Salsabila F. Imni, Wardah Sanjaya, Rizka D. Andriani, Henriyansah, Renda S. P. Putri, and Dhea Ekaputri. 2024. “Life Expectancy Prediction Using Decision Tree, Random Forest, Gradient Boosting, and XGBoost Regressions.” *Jurnal Sintak* 2(2):71–82. doi: 10.62375/jsintak.v2i2.249.
- Chang Hartono, Pieter, and Albertus Dwiyoğa Widianoro. 2024. “Analisis Prediksi Harga Saham Unilever Menggunakan Regresi Linier Dengan RapidMiner.” *Journal of Computer and Information Systems Ampera* 5(3):2775–2496.
- Christophorus Bintang Saputra, and Deddy Priatmodjo Koesrindartoto. 2024. “Pemanfaatan Analisis Sentimen Youtube Untuk Prediksi Harga Saham: Studi Pada Investor Retail Indonesia.” *Jurnal Manajemen* 21(1):1–17. doi: 10.25170/jm.v21i1.5184.

- Du, Kelvin, Frank Xing, Rui Mao, and Erik Cambria. 2023. "FinSenticNet: A Concept-Level Lexicon for Financial Sentiment Analysis." *2023 IEEE Symposium Series on Computational Intelligence, SSCI 2023* 109–14. doi: 10.1109/SSCI52147.2023.10371970.
- Dwita, Olga Mauri Sandi, and Adatul Mukarromah. 2025. "Prediksi Pergerakan Naik Turun Harga Saham Berdasarkan Headline Berita Menggunakan Support Vector Machine Dan Naive Bayes Classifier." *Jurnal Sains Dan Seni ITS* 13(6). doi: 10.12962/j23373520.v13i6.156225.
- Effendi; Noviana, Rina. 2021. "Perancangan Web Sistem Analisis Sentimen Media Sosial Twitter Dengan Metode Valence Aware Dictionary And Sentimen Reasoner (Vader) Menggunakan PHP & MySQL Pada Pemerintah Kota Bekasi." *Jurnal Ilmiah Komputasi* 20(1):1–14. doi: 10.32409/jikstik.20.1.369.
- Ernawati, Siti, and Risa Wati. 2024. "Evaluasi Performa Kernel SVM Dalam Analisis Sentimen Review Aplikasi ChatGPT Menggunakan Hyperparameter Dan VADER Lexicon." *Jurnal Buana Informatika* 15(01):40–49. doi: 10.24002/jbi.v15i1.7925.
- Fadila, Anisa Nur, and Cahyani Nuswandari. 2022. "Apa Saja Faktor - Faktor Yang Mempengaruhi Harga Saham." *E-Bisnis : Jurnal Ilmiah Ekonomi Dan Bisnis* 15(2):283–93. doi: 10.51903/e-bisnis.v15i2.837.
- Fadliansyah, Muhammad, Setio Basuki, and Yufis Azhar. 2020. "Prediksi Harga Saham Menggunakan Sentimen Pilkada DKI Jakarta 2017 Dengan Algoritma Support Vector Machine." *Jurnal Repositor* 2(12):1623–30. doi: 10.22219/repositor.v2i12.444.
- Firdaus, Rizky Achmad, and Nur Ifrochah. 2022. "Pengaruh Pengetahuan Investasi Dan Motivasi Investasi Terhadap Minat Investasi Mahasiswa Politeknik Keuangan Negara Stan Di Pasar Modal." *Jurnal Acitya Ardana* 2(1):16–28. doi: 10.31092/jaa.v2i1.1434.
- Fitri, Evita, and Dwiza Riana. 2022. "Analisa Perbandingan Model Prediction Dalam Prediksi Harga Saham Menggunakan Metode Linear Regression, Random Forest Regression Dan Multilayer Perceptron." *METHOMIKA Jurnal Manajemen*

Informatika Dan Komputerisasi Akuntansi 6(1):69–78. doi: 10.46880/jmika.vol6no1.pp69-78.

Fitriyani, Dela Naufalia. 2025. “Daftar Saham Cuan Pekan Ini, Tambang Dan Telekomunikasi Bisa Dipilih.” *CNN Indonesia*. Retrieved (<https://www.cnnindonesia.com/ekonomi/20250317063145-92-1209526/daftar-saham-cuan-pekan-ini-tambang-dan-telekomunikasi-bisa-dipilih>).

Gatti, Lorenzo, and Judith van Stegeren. 2020. “Improving Dutch Sentiment Analysis in Pattern.” *Computational Linguistics in the Netherlands Journal* 10:73–89.

Gu, Wen jun, Yi hao Zhong, Shi zun Li, Chang song Wei, Li ting Dong, Zhuo yue Wang, and Chao Yan. 2024. “Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis.” *ACM International Conference Proceeding Series* 67–72. doi: 10.1145/3694860.3694870.

Hidayana, Rizki Apriva, and Budi Nurani Ruchjana. 2023. “Peramalan Return Saham Menggunakan Model Integrated Moving Average.” *Jambura Journal of Mathematics* 5(1):199–209. doi: 10.34312/jjom.v5i1.17381.

Himawan, Irvan, Odi Nurdiawan, Gifthera Dwilestari, and ... 2022. “Prediksi Harga Saham Dengan Algoritma Regresi Linier Dengan RapidMiner.” *JURSIMA (Jurnal ...* 10(3):239–47.

Huang, Jia Yen, Chun Liang Tung, and Wei Zhen Lin. 2023. “Using Social Network Sentiment Analysis and Genetic Algorithm to Improve the Stock Prediction Accuracy of the Deep Learning-Based Approach.” *International Journal of Computational Intelligence Systems* 16(1). doi: 10.1007/s44196-023-00276-9.

Indarso, Andhika Octa, and Artambo Benjamin Pangaribuan. 2021. “Penggunaan Metode Multilayer Perceptron Pada Prediksi Indeks Saham LQ45.” *Informatik : Jurnal Ilmu Komputer* 17(1):38. doi: 10.52958/iftk.v17i1.2225.

Kurnia, Deni, Muhammad Itqan Mazdadi, Dwi Kartini, Radityo Adi Nugroho, and Friska Abadi. 2023. “Seleksi Fitur Dengan Particle Swarm Optimization Pada Klasifikasi Penyakit Parkinson Menggunakan XGBoost.” *Jurnal Teknologi Informasi Dan Ilmu Komputer* 10(5):1083–94. doi: 10.25126/jtiik.2023107252.

- Leonardi, Veronica Hertensia, Ali Ibrahim, Rizka Dhini Kurnia, and Mira Afrina. 2025. “Analisis Sentimen Ulasan Aplikasi Pembelajaran Bahasa Menggunakan Metode VADER.” doi: 10.33364/algorithm/v.22-1.2285.
- Marjuni, Aris. 2022. “Peramalan Harga Saham Serentak Menggunakan Model Multivariate Singular Spectrum Analysis.” *Jurnal Sistem Informasi Bisnis* 12(1):17–25. doi: 10.21456/vol12iss1pp17-25.
- Maulida, Dina, and Arya Samudra Mahardhika. 2021. “Analisis Perbedaan Harga Saham, Volume Perdagangan Saham, Dan Return Saham Sebelum Dan Sesudah Stock Split.” *Jurnal Akuntansi* 1(1):1–7. doi: 10.37058/jak.v1i1.2754.
- Muhidin, Muhidin, and Tiar Lina Situngkir. 2022. “Pengaruh Rasio Profitabilitas Terhadap Harga Saham Perusahaan Perbankan Yang Terdaftar Di Bursa Efek Indonesia Pada Tahun 2015 - 2021.” *Transformasi Manageria: Journal of Islamic Education Management* 3(1):15–27. doi: 10.47467/manageria.v3i1.2093.
- Murdiansyah, Danang Triantoro. 2024. “Prediksi Stroke Menggunakan Extreme Gradient Boosting.” *JIKO (Jurnal Informatika Dan Komputer)* 8(2):419. doi: 10.26798/jiko.v8i2.1295.
- Nanda Sepriadi, Elvia Budianita, Muhammad Fikry, and Pizaini. 2023. “Analisis Sentimen Review Aplikasi Mypertamina Menggunakan Word Embedding Fasttext Dan Algoritma K-Nearest Neighbor.” *INFORMASI (Jurnal Informatika Dan Sistem Informasi)* 15(1):91–109. doi: 10.37424/informasi.v15i1.222.
- Nilsen, Andrew. 2022. “Perbandingan Model RNN, Model LSTM, Dan Model GRU Dalam Memprediksi Harga Saham-Saham LQ45.” *Jurnal Statistika Dan Aplikasinya* 6(1):137–47. doi: 10.21009/jsa.06113.
- Ningrum, Ayu Ahadi, Iwan Syarif, Agus Indra Gunawan, Edi Satriyanto, and Rosmaliati Muchtar. 2021. “Algoritma Deep Learning-LSTM Untuk Memprediksi Umur Transformator.” *Jurnal Teknologi Informasi Dan Ilmu Komputer* 8(3):539–48. doi: 10.25126/jtiik.2021834587.
- Nugraha, Alan Catur, and Mohammad Isa Irawan. 2023. “Komparasi Deteksi Kecurangan Pada Data Klaim Asuransi Pelayanan Kesehatan Menggunakan

- Metode Support Vector Machine (SVM) Dan Extreme Gradient Boosting (XGBoost).” *Jurnal Sains Dan Seni ITS* 12(1). doi: 10.12962/j23373520.v12i1.107032.
- Nugraha, Yoga Estu Nugraha, Ishak Ariawan, and Willdan Aprizal Arifin. 2023. “Weather Forecast From Time Series Data Using Lstm Algorithm.” *Jurnal Teknologi Informasi Dan Komunikasi* 14(1):144–52. doi: 10.51903/jtikp.v14i1.531.
- Nurhalizah, Ria Suci, Rian Ardianto, and Purwono Purwono. 2024. “Analisis Supervised Dan Unsupervised Learning Pada Machine Learning: Systematic Literature Review.” *Jurnal Ilmu Komputer Dan Informatika* 4(1):61–72. doi: 10.54082/jiki.168.
- Permata Aulia, Thalita Meisya, Nur Arifin, and Rini Mayasari. 2021. “Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinasi Covid-19.” *SINTECH (Science and Information Technology) Journal* 4(2):139–45. doi: 10.31598/sintechjournal.v4i2.762.
- Pratama, Ananda Willies, Andi Wijayanto, and Dinalestari Purbawati. 2022. “Pengaruh Motivasi Dan Pengetahuan Investasi Saham Terhadap Keputusan Berinvestasi Saham Di Bursa Efek Indonesia Selama Pandemi Covid-19 (Studi Pada Investor Saham Kota Semarang).” *Jurnal Ilmu Administrasi Bisnis* 11(4):710–21. doi: 10.14710/jiab.2022.36015.
- Puspadini, Mentari. 2024. “Investor Gen Z-Millennial Dominasi Pasar Modal, Tapi Aset Paling Kecil.” *CNBC Indonesia*. Retrieved (<https://www.cnbcindonesia.com/market/20240812163415-17-562426/investor-gen-z-millennial-dominasi-pasar-modal-tapi-aset-paling-kecil>).
- Putra, Janur Syah, Rima Dias Ramadhani, and Auliya Burhanuddin. 2022. “Prediksi Harga Saham Bank Bri Menggunakan Algoritma Linear Regresion Sebagai Strategi Jual Beli Saham.” *Journal of Dinda : Data Science, Information Technology, and Data Analytics* 2(1):1–10. doi: 10.20895/dinda.v2i1.273.
- Putri, Via Andani, and Natali Yustisia. 2021. “Dampak Rasio Keuangan Terhadap Harga Saham Perusahaan Barang Konsumsi Yang Terdaftar Di Bursa Efek

- Indonesia.” *Lisyabab: Jurnal Studi Islam Dan Sosial* 2(1):1–16. doi: 10.58326/jurnallisyabab.v2i1.71.
- Saputra, Rendy, Alfin Gusti Alamsyah, Dandi, Michael Tjoanda, Kent Nick, and Alten Cornelius. 2024. “Analisis Prediksi Saham Tesla Menggunakan Algoritma Long Short Term Memory (Lstm).” *Journal of Computer Science and Information Technology* 2(1):81–90. doi: 10.70248/jcsit.v2i1.1482.
- Saputro, Daniel, and Daniel Swanjaya. 2023. “Analisa Prediksi Harga Saham Menggunakan Neural Network Dan Net Foreign Flow.” *Generation Journal* 7(2):96–104. doi: 10.29407/gj.v7i2.20001.
- Sari, RR. Kurnia Novita, Wilan Sutisna, Martha Joanadiva Majesty Wororomi, and Venansius Ryan Tjahjono. 2023. “Komparasi Model Gerak Brown Geometrik Termodifikasi Dan Model Kecerdasan Buatan Untuk Prediksi Harga Saham Sektor Kesehatan Di Indonesia.” *JST (Jurnal Sains Dan Teknologi)* 12(2):446–56. doi: 10.23887/jstundiksha.v12i2.48960.
- Siringoringo, Rimbun, Resianta Perangin-angin, and Mufria J. Purba. 2021. “Segmentasi Dan Peramalan Pasar Retail Menggunakan Xgboost Dan Principal Component Analysis.” *METHOMIKA Jurnal Manajemen Informatika Dan Komputerisasi Akuntansi* 5(1):42–47. doi: 10.46880/jmika.vol5no1.pp42-47.
- Susila, Muktar Redy. 2023. “Dampak COVID-19 Terhadap Hasil Analisis Kelayakan Investasi Pada Saham Perusahaan BUMN.” *E-Jurnal Akuntansi* 33(7):1796–1808. doi: 10.24843/eja.2023.v33.i07.p08.
- Taofiqurrohman, Hendra, and Fikri Fahu Roji. 2025. “Telkom Stock Price Prediction Using Prophet : Analysis of the Effect of Public Sentiment on the Presence of Starlink Prediksi Harga Saham Telkom Menggunakan Prophet : Analisis Pengaruh Sentimen Publik Terhadap Kehadiran Starlink.” 5(April):484–95.
- Thormann, Marah Lisanne, Jan Farchmin, Christoph Weisser, René Marcel Kruse, Benjamin Safken, and Alexander Silbersdorff. 2021. “Stock Price Predictions with LSTM Neural Networks and Twitter Sentiment.” *Statistics, Optimization and Information Computing* 9(2):268–87. doi: 10.19139/soic-2310-5070-1202.

- Tri Nurmaya, Septia, Budi Soesilo, and Doni Abdul Fatah. 2025. "Prediksi Harga Saham Pada Bank Bca Tbk Dengan Teknik Xgboost (Extreme Gradient Boosting)." *JURSIMA (Jurnal Sistem Informasi & Manajemen)* 370–79.
- Widyawati, Lia Ratna, and Imronudin. 2024. "Pengaruh Sentimen Investor Dan Antecedennya Terhadap Pengambilan Keputusan Investasi." *Journal of Management (YUME)* 7(3):321–31.
- Wiranata, Rico Bayu, and Arif Djunaidy. 2021. "Optimasi Hyper-Parameter Berbasis Algoritma Genetika Pada Ensemble Learning Untuk Prediksi Saham Yang Mempertimbangkan Indikator Teknikal & Sentimen Berita." *JATISI: Jurnal Teknik Informatika Dan Sistem Informasi* 8(3):1442–56.
- Xiao, Qianyi, and Baha Ihnaini. 2023. "Stock Trend Prediction Using Sentiment Analysis." *PeerJ Computer Science* 9. doi: 10.7717/PEERJ-CS.1293.
- Yahyadi, Adi, and Fitri Latifah. 2022. "Analisis Sentimen Twitter Terhadap Ppkm Di Tengah Pandemi Covid-19 Mode Lstm." *Journal of Information System, Applied, Management, Accounting and Research*. 6(2):464–70. doi: 10.52362/jisamar.v6i2.791.
- Zakariyya, Q. Ahmad Riza. 2020. "Analisis Pengaruh Traffic Sosial Media Twitter Terhadap Pergerakan Harga Saham Pada Pt. Garuda Indonesia." Pp. 37–44 in *Konferensi Tren Bisnis dan Manajemen Praktis di Pasar Indonesia*. Vol. 2006.
- Zhang, Yifan. 2023. *Stock Price Prediction Method Based on XGboost Algorithm*. Atlantis Press International BV.

LAMPIRAN

Lampiran 1 Form Bimbingan Pembimbing 1



FORMULIR BIMBINGAN KEGIATAN TUGAS AKHIR

Nama Mahasiswa : Imanti Nur Latifah
 NPM : 1214033
 Pembimbing : Nisa Hanum Harani S.Kom., M.T., CISP, SFPC
 Judul : PREDIKSI HARGA SAHAM DENGAN XGBOOST DAN SENTIMEN TWITTER

Kegiatan : Proyek 1/Proyek 2/Proyek 3/Internship 1/Internship 2/**Tugas Akhir***)

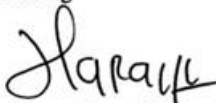
Pelaksanaan Bimbingan

No	Tanggal/Jam Bimbingan	Catatan Kemajuan/Materi	Paraf Pembimbing
1.	Jumat, 25 April 2025	- Menjelaskan topik & judul pengejuan proposal TA	
2.	Jumat, 9 Mei 2025	- Mencoba simulasi algoritma berdasarkan data stock jurnal dr watase - Mencari data tweet dengan keyword kode saham masing-masing perusahaan	
3.	Jumat, 23 Mei 2025	- Mengimplementasikan data train & test mengunci data tersebut agar tidak aeak - Mendapatkan data tweet mencoba menganalisis sentiment menggunakan library VADER - Mendapatkan jurnal sebagai acuan jurnal	
4.	Senin, 26 Mei 2025	- Memperlihatkan data hasil analisis sentiment twitter menggunakan metode vader - Mendapatkan 3 bulan data tweet - Mencoba mengolah data hasil sentiment dan saham	
5.	Senin, 2 Juni 2025	- Mencari jurnal acuan internasional untuk mengetahui alur sentiment bagaimana saat preprocessing dengan menyesuaikan topik saat mencari jurnal - Data digabung dari berbagai tipe saham dengan menambahkan kolom kode saham dan sentiment	
6.	Jumat, 13 Juni 2025	- Mulai menganalisis data saham & data sentiment - Masuk ke data training & test pada setiap saham pembagian data	

7.	Jumat, 20 Juni 2025	- Mulai penyusunan Laporan Bab 1	D.
8.	Senin, 23 Juni 2025	- Penyusunan laporan bab 2 - Mulai membangun model LSTM dari data gabungan sentien dan saham	D.
9.	Jumat, 11 Juli 2025	- Menampilkan hasil evaluasi model LSTM melalui table experiment yang telah dilakukan - Terdapat akurasi yang kurang memenuhi, dicoba untuk melakukan model algoritma lain machine learning - Menampilkan progress laporan bab 3	D.
10.	Senin, 14 Juli 2025	- Konsultasi judul laporan dan identifikasi masalah - Progres laporan bab 4 bagian eksperimen - Menampilkan hasil evaluasi menggunakan algoritma XGBoost pada data gabungan	D.
11.	Jumat, 18 Juli 2025	- Memperbaiki tahapan preprocessing analisis sentiment dengan menggunakan vader - Menyatukan hasil analisis dengan saham melakukan prediksi dan membandingkan algoritma LSTM dan XGBoost	D.
12.	Selasa, 22 Juli 2025	- Memberikan hasil final laporan - Melakukan pelaporan hasil evaluasi untuk dimasukan ke dalam laporan bab 4 dan hasil - Membuat dashboard prediksi harga saha menggunakan algoritma XGBoost	D.

Bandung,²³ Juli 2025

Pembimbing



(Nisa Hanum Harani S.Kom., M.T., CDSP, SFPC)

Keterangan :

1. Formulir kegiatan Tugas Akhir harus selalu dibawa pada saat bimbingan berlangsung.
2. Formulir kegiatan Tugas Akhir tidak boleh hilang, kotor dan tidak boleh rusak.
3. Formulir kegiatan Tugas Akhir dikumpulkan kembali pada saat pengajuan sidang.

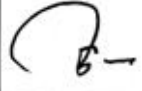

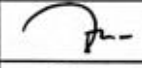

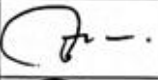


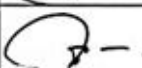
* : Pilih satu kegiatan yang sedang diambil :
isi jenis kegiatan yang diambil Jika kurang silahkan
tambah halaman baru

Lampiran 2 Form Bimbingan Pembimbing 2

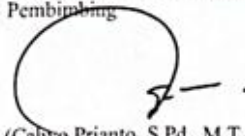
FORMULIR BIMBINGAN KEGIATAN TUGAS AKHIR *

Nama Mahasiswa : Isnaenti Nur Latifah
 NPM : 1214033
 Pembimbing : Cahyo Prianto, S.Pd., M.T.,CDSP, SFPC
 Judul : PREDIKSI HARGA SAHAM DENGAN XGBOOST DAN SENTIMEN TWITTER
 Kegiatan : Proyek 1/Proyek 2/Proyek 3/Internship 1/Internship 2/Tugas Akhir*)

Pelaksanaan Bimbingan

No	Tanggal/Jam Bimbingan	Catatan Kemajuan/Materi	Paraf Pembimbing
1.	06/05/2025	Menyampaikan topik yang diusulkan Mencari sumber dataset harga saham dan data Twitter terkait saham.	
2.	20/05/2025	Menyampaikan dataset yang digunakan dan melakukan pembersihan preprocessing sentiment	
3.	27/05/2025	Menggabungkan dataset sentiment dan saham	
4.	17/06/2025	Melakukan analisis terhadap sentiment dan harga saham pengaruh ke duanya	
5.	24/06/2025	Menyampaikan penggunaan Vader untuk metode analisis sentiment	
6.	09/07/2025	Menyampaikan menggunakan algoritma yang digunakan untuk prediksi harga saham	
7.	12/07/2025	Melakukan diskusi penggunaan algoritma xgboost	
8.	22/07/2025	Menampilkan dashboard visualisasi hasil prediksi menggunakan algoritma xgboost	

Bandung, 23 Juli 2025
 Pembimbing


 (Cahyo Prianto, S.Pd., M.T.,CDSP, SFPC)

Keterangan :
 1. Formulir kegiatan Tugas Akhir harus selalu dibawa pada saat bimbingan berlangsung.
 2. Formulir kegiatan Tugas Akhir tidak boleh hilang, kotor dan tidak boleh rusak. 3. Formulir kegiatan Tugas Akhir dikumpulkan kembali pada saat pengajuan sidang.

* : Pilih satu kegiatan yang sedang diambil :
 isi jenis kegiatan yang diambil Jika kurang silahkan tambah halaman baru


Lampiran 3 Hasil Turnitin

Feedback Studio - Google Chrome
ev.turnitin.com/app/carta/en_us/?o=2722207424&student_user=1&lang=en_us&u=1183931989&s=1&ro=103

feedback studio Eliseo Coe Laporan TA - Cek2_removed.pdf

PREDIKSI HARGA SAHAM DENGAN XGBOOST DAN SENTIMEN *TWITTER*

49
Diajukan Sebagai Salah Satu Syarat untuk Memperoleh Gelar Sarjana Terapan
Pada Program Pendidikan Teknik Informatika


ULBI
Universitas Logistik & Bisnis Internasional

Match Overview

13%

1	ejournal.mediantarik...	Internet Source	2%
2	Muhamad Rizki, Aditya ...	Publication	1%
3	link.springer.com	Internet Source	<1%
4	Submitted to Internatio...	Student Paper	<1%
5	www.coursehero.com	Internet Source	<1%
6	ejournal.atmajaya.ac.id	Internet Source	<1%
7	Submitted to Universita...	Student Paper	<1%
8	ejurnal.ung.ac.id	Internet Source	<1%
9	123dok.com	Internet Source	<1%

Page: 1 of 69 Word Count: 15471 Text-Only Report High Resolution On 11:32 29/07/2025