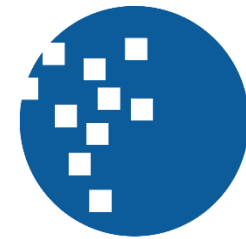


**SENTIMEN ANALISIS PREDIKSI HARGA SAHAM PGAS
DENGAN PENDEKATAN DEEP LEARNING LSTM
XGBOOST *HYBRID***



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Benz Martheen Walter Jonash

00000059822

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

**SENTIMEN ANALISIS PREDIKSI HARGA SAHAM PGAS
DENGAN PENDEKATAN DEEP LEARNING LSTM XGBOOST
*HYBRID***



Diajukan sebagai Salah Satu Syarat untuk Memperoleh

Gelar Sarjana Komputer

Benz Martheen Walter Jonash

00000059822

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG**

2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Benz Martheen Walter Jonash

Nomor Induk Mahasiswa : 00000059822

Program Studi : Sistem Informasi

Skripsi dengan judul: Sentimen Analisis Prediksi Harga Saham PGAS dengan Pendekatan Deep Learning LSTM XGBOOST *Hybrid*. Merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 20 Juni 2025



(Benz Martheen Walter Jonash)

HALAMAN PERSETUJUAN

Skripsi

SENTIMEN ANALISIS PREDIKSI HARGA SAHAM PGAS DENGAN
PENDEKATAN DEEP LEARNING LSTM XGBOOST *HYBRID*

Oleh

Nama : Benz Martheen Walter Jonash
NIM : 00000059822
Program Studi : Sistem Informasi
Fakultas : Teknik Informatika

Telah disetujui untuk diajukan pada
Sidang Ujian Skripsi Universitas Multimedia Nusantara
Tangerang, 6 Juni 2025
Pembimbing

Wella S.Kom., M.M.S.I

0305119101

Ketua Program Studi Sistem Informasi



Ririn Ikana Desanti, S.Kom., M.Kom

HALAMAN PENGESAHAN

Skripsi dengan judul
**SENTIMEN ANALISIS PREDIKSI HARGA SAHAM PGAS DENGAN
PENDEKATAN DEEP LEARNING LSTM XGBOOST HYBRID**

Oleh


Nama : Benz Martheen Walter Jonash
NIM : 00000059822
Program Studi : Sistem Informasi
Fakultas : Teknik Informatika

Telah diujikan pada hari Jumat, 20 Juni 2025
Pukul 08.00 s.d 10.00 dan dinyatakan
LULUS
Dengan susunan penguji sebagai berikut.

Ketua Sidang

 9/7/2025
Ahmad Faza S.Kom., M.T.I.
0312019501

Penguji

 7/7/25
Melissa Indah Fianty S.Kom., M.MSI.
0313019201

Pembimbing

 11/7/25
Wella S.Kom., M.M.S.I
0305119101
Ketua Program Studi Sistem Informasi

 11/7/25
Ririn Ikana Desanti, S.Kom., M.Kom
0313058001

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Yang bertanda tangan di bawah ini:

Nama : Benz Martheen Walter Jonash

NIM : 00000059822

Program Studi : Sistem Informasi

Fakultas : Teknik Informatika

Jenis Karya : Tesis/Skripsi/Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Multimedia Nusantara Hak Bebas Royalti Noneksklusif (Non-exclusive Royalty-Free Right) atas karya ilmiah saya yang berjudul "Sentimen analisis prediksi harga saham PGAS dengan pendekatan deep learning LSTM Xgboost hybrid" beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini, Universitas Multimedia Nusantara berhak menyimpan, mengalihmediakan/mengalihformatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Tangerang, 04 Juli 2025



(Benz Martheen Walter Jonash)

KATA PENGANTAR

Puji Syukur kepada Tuhan Yang Maha Esa atas kasih dan penyertaannya yang memungkinkan Skripsi dengan judul “Sentimen Analisis Prediksi Harga Saham PGAS Dengan Pendekatan Deep Learning LSTM XGBOOST *Hybrid*”. Penulis menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan Skripsi ini, sangatlah sulit bagi penulis untuk menyelesaikan tugas akhir ini. Oleh karena itu, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr. Andrey Andoko, M.Sc., Ph.D. selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Ibu Ririn Ikana Desanti S.Kom., M.M.S.I., selaku Ketua Program Studi Universitas Multimedia Nusantara.
4. Ibu Wella S.Kom., M.M.S.I, sebagai Pembimbing pertama yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Penulis menyadari bahwa masih terdapat banyak kekurangan pada Laporan MBKM ini sehingga kritik dan saran dari para pembaca sangat diharapkan. Kritik dan saran tersebut akan dijadikan sebagai sarana penulis untuk mengintrospeksi diri agar penulis dapat memperbaikinya pada penulisan-penulisan karya ilmiah selanjutnya. Akhir kata, semoga dengan penulisan Laporan ini dapat memberikan wawasan yang bermanfaat kepada para pembacanya

Tangerang, 04 Juli 2025



(Benz Martheen Walter Jonash)

SENTIMEN ANALISIS PREDIKSI HARGA SAHAM PGAS DENGAN PENDEKATAN DEEP LEARNING LSTM XGBOOST *HYBRID*

(Benz Martheen Walter Jonash)

ABSTRAK

Pada era digital saat ini, fluktuasi harga saham tidak hanya dipengaruhi oleh laporan keuangan atau analisis teknikal, tetapi juga oleh sentimen publik yang berkembang di media sosial. PGAS, sebagai salah satu perusahaan energi utama di Indonesia, kerap menjadi sorotan terkait isu-isu seperti subsidi gas hingga pembangunan proyek infrastruktur, yang pada akhirnya turut mempengaruhi persepsi investor.

Penelitian ini bertujuan mengintegrasikan sentimen media sosial ke dalam model prediksi harga saham untuk meningkatkan relevansi dan akurasi prediksi. Model prediksi yang digunakan adalah gabungan antara algoritma *Long Short-Term Memory* (LSTM) untuk menangkap pola historis harga saham, dengan algoritma *Extreme Gradient Boosting* (XGBoost) yang digunakan untuk memproses faktor eksternal berupa sentimen dari Twitter. Data sentimen diperoleh melalui proses scraping terhadap tweet dalam rentang waktu 2020 hingga 2025, kemudian dianalisis menggunakan metode VADER (*Valence Aware Dictionary and sEntiment Reasoner*).

Hasil menunjukkan bahwa penggunaan model *hybrid* (LSTM-XGBoost) secara signifikan meningkatkan kinerja prediksi dibandingkan model LSTM tunggal, dengan peningkatan nilai R^2 dari 0,78 menjadi 0,82 serta pengurangan tingkat *error*. Temuan ini mengindikasikan bahwa integrasi sentimen media sosial dapat memberikan dampak positif dalam prediksi pergerakan harga saham serta mengukuhkan potensi sentimen publik sebagai indikator yang kuat dalam pengambilan keputusan investasi.

Kata kunci: LSTM, PGAS, Prediksi, Sentimen Analisis, XGBoost

UNIVERSITAS
MULTIMEDIA
NUSANTARA

SENTIMEN ANALISIS PREDIKSI HARGA SAHAM PGAS DENGAN PENDEKATAN DEEP LEARNING LSTM XGBOOST *HYBRID*

(Benz Martheen Walter Jonash)

ABSTRACT (English)

In the current digital era, stock price fluctuations are not only influenced by financial reports or technical analysis, but also by public sentiment that develops on social media. PGAS, as one of the main energy companies in Indonesia, is often in the spotlight regarding issues such as gas subsidies to the development of infrastructure projects, which ultimately influence investor perceptions.

This research aims to integrate social media sentiment into stock price prediction models to increase the relevance and accuracy of predictions. The prediction model used is a combination of the Long Short-Term Memory (LSTM) algorithm to capture historical stock price patterns, with the Extreme Gradient Boosting (XGBoost) algorithm which is used to process external factors in the form of sentiment from Twitter. Sentiment data was obtained through a scraping process for tweets in the period 2020 to 2025, then analyzed using the VADER (Valence Aware Dictionary and Sentiment Reasoner) method.

Experimental results show that the use of a hybrid model (LSTM-XGBoost) significantly improves prediction performance compared to a single LSTM model, with an increase in the R^2 value from 0.78 to 0.82 and a reduction in the error rate. These findings indicate that the integration of social media sentiment can have a positive impact in predicting stock price movements and confirms the potential of public sentiment as a strong indicator in making investment decisions.

Keywords: LSTM, PGAS, Prediction, Sentiment Analysis, XGBoost

DAFTAR ISI

HALAMAN PERNYATAAN TIDAK PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH.....	v
KATA PENGANTAR	vi
ABSTRAK.....	vii
<i>ABSTRACT (English)</i>	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Batasan Masalah	6
1.4 Tujuan dan Manfaat Penelitian	6
1.4.1 Tujuan Penelitian	6
1.4.2 Manfaat Penelitian	6
1.5 Sistematika Penulisan	7
BAB II LANDASAN TEORI.....	9
2.1 Penelitian terkait	9
2.2 Teori Penelitian.....	12
2.2.1 Saham	12
2.2.2 Investasi	13
2.2.3 <i>Deep Learning</i>	15
2.2.4 Analisis Sentimen	17
2.2.5 Text Mining.....	17
2.3 Framework/Algoritma/SDLC Penelitian	19
2.3.1 CRISP-DM	19
2.3.2 <i>XGBoost</i>	22
2.3.3 <i>LSTM (Long Short Term Memory)</i>	23
2.3.4 VADER	24

2.4 Tools Penelitian	25
2.4.1 Python	25
2.4.2 Jupyter Notebook	26
2.4.3 Google Colab	28
2.4.4 Draw.io	29
2.4.5 X (Twitter)	30
BAB III METODOLOGI PENELITIAN	32
3.1 Gambaran Umum Objek Penelitian	32
3.1.1 Data Mentah	33
3.2 Metode Penelitian	34
3.3 Teknik Pengumpulan Data	39
3.4 Variabel Penelitian	40
3.5 Teknik Analisis Data	41
3.5.1 Data Pre – Processing	41
3.5.2 Modeling	41
BAB IV ANALISIS DAN HASIL PENELITIAN	44
4.1 Business Understanding	44
4.2 Data Preparation	44
4.2.1 Merging Data	45
4.2.2 Cleansing	48
4.2.3 Transformation	49
4.2.4 Feature Engineering	50
4.3 Modeling	51
4.4 Hasil dan Diskusi	56
BAB V SIMPULAN DAN SARAN	61
5.1 Simpulan	61
5.2 Saran	61
DAFTAR PUSTAKA	63
LAMPIRAN	65

DAFTAR TABEL

Tabel 2. 1 Penelitian terkait.....	9
Tabel 4. 1 Jumlah hasil sentimen.....	45
Tabel 4. 2 Atribut dalam dataset.....	46
Tabel 4. 3 Hasil prediksi.....	55
Tabel 4. 4 Hasil Evaluasi Model	57
Tabel 4. 5 Persentasi Sentimen.....	58



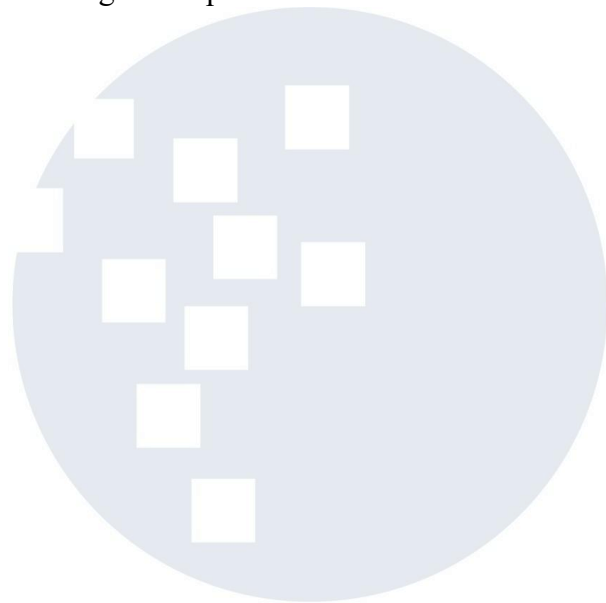
DAFTAR GAMBAR

Gambar 1. 1 Harga Saham PT. PGAS 5 Tahun Terakhir.....	4
Gambar 2. 1 Framework CRISP-DM.....	19
Gambar 2. 2 Logo Python	26
Gambar 2. 3 Gambar Jupyter.....	27
Gambar 2. 4 Gambar Logo Google Colab.....	28
Gambar 2. 5 Logo X.....	30
Gambar 3. 1 Komentar Masyarakat terhadap PGAS di Twitter (X)	33
Gambar 3. 2 Komentar Masyarakat terhadap PGAS di Twitter (X)	36
Gambar 3. 3 skrip untuk melakukan scraping data X.....	39
Gambar 4. 1 Skrip mengklasifikasikan sentiment analisis	45
Gambar 4. 2 Syntax untuk melakukan Tweet Harvest	46
Gambar 4. 3 Hasil scraping data twitter (X).....	46
Gambar 4. 4 Setelah dilakukan merged.....	48
Gambar 4. 5 Hasil cleansing.....	49
Gambar 4. 6 Skrip untuk melakukan tranformasi	50
Gambar 4. 7 Skrip untuk feature engineering	51
Gambar 4. 8 Proses modelling LSTM.....	52
Gambar 4. 9 Visualisasi dari harga actual dan prediksi hanya LSTM	52
Gambar 4. 10 Perbandingan hasil actual dengan LSTM <i>Hybrid</i>	53
Gambar 4. 11 Perbandingan hasil aktual dengan hasil LSTM serta hasil <i>Hybrid</i>	54
Gambar 4. 12 Hasil sentimen dibandingkan dengan pola saham	59



DAFTAR LAMPIRAN

Lampiran 1 Turnitin Similarity Report.....	65
Lampiran 2 Form Bimbingan Skripsi.....	65



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

BAB I

PENDAHULUAN

1.1 Latar Belakang

Saham merupakan instrumen paling populer di pasar keuangan, perusahaan dapat menerbitkan saham saat mereka menerimanya memutuskan untuk memberikan modal pada bisnis tersebut. Investor biasanya memilih saham sebagai alat investasi sangat menarik karena kemampuan saham untuk mencapai tingkat keuntungan tertentu. Objektif Investor berinvestasi di pasar modal untuk mendapatkan keuntungan. Dari jumlah modal tersebut, maka investor mempunyai hak atas penghasilan dan harta kekayaan perseroan serta berhak menghadiri Rapat Umum Pemegang Saham. Keuntungan semaksimal mungkin dengan menggabungkan risiko investasi yang sangat tinggi[1].

Di pasar sekunder (pasar saham) atau dalam aktifitas pasar saham sehari-hari, harga saham berfluktuasi dalam bentuk kenaikan atau penurunan. Secara umum, faktor-faktor yang mempengaruhi harga saham dapat dikategorikan menjadi dua: faktor internal dan faktor eksternal. Faktor internal meliputi kinerja keuangan perusahaan, laba bersih, struktur modal, dan kebijakan manajemen. Sementara itu, faktor eksternal mencakup perubahan tingkat suku bunga, inflasi, nilai tukar, harga komoditas, serta kondisi politik dan regulasi pemerintah. Selain faktor ekonomi, perkembangan teknologi dan tren di media sosial juga mulai memainkan peran dalam pergerakan harga saham. Analisis sentimen dari berita dan media sosial kini semakin sering digunakan untuk memahami persepsi pasar terhadap suatu saham. Pasokan dan permintaan suatu hal dipengaruhi oleh banyak faktor, baik yang bersifat spesifik pada saham (kinerja perusahaan dan industri di mana ia berada) maupun faktor makro atau eksternal, seperti perubahan tingkat suku bunga, inflasi, nilai tukar dan faktor nirlaba, kondisi ekonomi dan sosial politik [2]

Dalam beberapa tahun terakhir, berbagai peristiwa global dan domestik memberikan dampak signifikan terhadap pasar saham Indonesia. Isu-isu seperti kenaikan harga minyak dunia, ketidakpastian kebijakan energi, serta sentimen

publik terhadap perusahaan tertentu sering kali menjadi pemicu volatilitas harga saham. Selain itu, kasus korupsi yang baru terungkap dalam industri juga memberikan dampak besar terhadap kepercayaan investor. Skandal semacam ini tidak hanya memengaruhi persepsi publik, tetapi juga dapat berdampak langsung pada pergerakan harga saham perusahaan terkait akibat meningkatnya ketidakpastian regulasi dan risiko investasi. Oleh karena itu, penting untuk memahami bagaimana faktor-faktor ini memengaruhi pergerakan saham, terutama melalui analisis berita dan media sosial, yang dapat mencerminkan respons pasar terhadap peristiwa tersebut.

Salah satu saham yang mengalami fluktuasi harga signifikan adalah PT Perusahaan Gas Negara Tbk (PGAS), perusahaan yang bergerak di sektor energi dan distribusi gas di Indonesia. Harga saham PGAS sering kali dipengaruhi oleh kebijakan pemerintah terkait harga gas, dinamika industri energi global, serta perkembangan proyek infrastruktur gas nasional. Selain itu, berita dan sentimen pasar terhadap PGAS, baik dari media massa maupun media sosial, dapat memberikan dampak terhadap pergerakan harganya. Oleh karena itu, analisis tren pasar dan sentimen publik terhadap PGAS menjadi aspek penting dalam memahami faktor-faktor yang mempengaruhi harga sahamnya [3].

Saat ini, PT Perusahaan Gas Negara Tbk (PGAS) saat ini dinilai melalui metrik *Price to Earnings Ratio* (PER), sebuah rasio yang menghitung valuasi saham dengan membandingkan harga pasar saat ini terhadap laba bersih per saham. Secara matematis, PER diperoleh dengan membagi harga saham terakhir dengan *Earning Per Share* (EPS), sehingga semakin tinggi nilai PER mengindikasikan ekspektasi pasar akan pertumbuhan laba di masa depan, sementara PER yang rendah bisa berarti saham dinilai murah atau perusahaan menghadapi tantangan kinerja. Bagi investor, PER menjadi tolok ukur cepat untuk menilai apakah saham PGAS diperdagangkan dengan premi atau diskon relatif terhadap peer di industri energi.

Pada kuartal terakhir, PER PGAS bergerak di kisaran angka yang kompetitif dibandingkan emiten lain di sektor energi nasional, mencerminkan optimisme pasar yang moderat. Lonjakan PER sempat terjadi saat laporan keuangan triwulan II

menunjukkan peningkatan margin distribusi gas, namun segera teredam oleh kekhawatiran terkait lonjakan biaya operasional. Bahkan, fluktuasi kecil pada EPS-akibat perubahan harga beli gas dari hulu-telah cukup mempengaruhi valuasi, sehingga analisis PER harus disandingkan dengan evaluasi tren kinerja perusahaan secara menyeluruh.

Selain metrik fundamental, pergerakan harga saham PGAS amat dipengaruhi oleh faktor eksternal. Harga minyak dan gas dunia yang bergejolak-seiring dinamika geopolitik dan permintaan global-seringkali menjadi pemicu sentimen pasar. Kebijakan pemerintah mengenai subsidi energi dan harga eceran gas industri juga memegang peranan penting, karena setiap pengumuman dapat langsung berdampak pada ekspektasi pendapatan PGAS. Dengan demikian, investor tidak hanya menelaah PER, tetapi juga merespons berita kebijakan dan kondisi makro yang membayangi prospek perusahaan.

Di samping itu, investasi pemerintah dan swasta dalam infrastruktur gas nasional menjadi sorotan berikutnya. Proyek pipanisasi dan pembangunan fasilitas regasifikasi yang sedang berlangsung memberi gambaran potensi pertumbuhan volume distribusi jangka panjang. Seiring dorongan transisi energi-menuju sumber energi yang lebih bersih-PGAS berpotensi memperluas pangsa pasar, baik dengan memperkuat jaringan pipa domestik maupun menggandeng investor asing dalam proyek strategis. Ekspektasi ini berkontribusi pada penilaian PER, karena pasar menghargai prospek kenaikan laba di masa depan.

Dengan mempertimbangkan ketiga aspek-valuasi PER, dinamika eksternal, dan roadmap infrastruktur-investor mendapat kerangka yang lebih komprehensif untuk mengambil keputusan. PER saja tidak cukup; penggabungan analisis fundamental dengan pemantauan kebijakan dan tren proyek infrastruktur memberikan gambaran holistic tentang risiko dan peluang. Ke depannya, tren pertumbuhan volume gas dan stabilitas margin operasional akan menjadi variabel kunci yang memengaruhi valuasi PGAS, sehingga pemahaman mendalam atas faktor-faktor tersebut menjadi esensial bagi investor.



Gambar 1. 1 Harga Saham PT. PGAS 5 Tahun Terakhir

2025 Source: Investing.com

Pada Gambar 1.1 tampak bahwa harga saham PGAS bergerak dari level Rp 1.345 pada 5 Maret 2020 menjadi Rp 1.520 pada 5 Maret 2025, naik sekitar 13 % dalam periode lima tahun. Kenaikan ini tidak terjadi secara linier, melainkan dipengaruhi kombinasi faktor internal-seperti kinerja operasional dan efisiensi distribusi gas-serta dinamika eksternal yang memicu sentimen pasar. Secara internal, PGAS mencatat peningkatan margin distribusi pada kuartal II 2020, yang menambah kepercayaan investor terhadap prospek jangka panjang perusahaan. Di sisi lain, volatilitas harga batubara dan LNG global turut berdampak pada biaya produksi serta daya saing PGAS di pasar domestik.

Di ranah kebijakan publik, pengumuman pemerintah soal penyesuaian harga eceran gas industri pada pertengahan 2020 menjadi katalis positif, karena menciptakan kepastian harga bagi konsumen dan stabilitas pemasukan PGAS. Selanjutnya, pemberitaan pada Oktober 2023 tentang rencana ekspansi jaringan pipa gas dan pembangunan fasilitas regasifikasi menegaskan komitmen transisi energi nasional, yang lantas memicu optimisme investor. Kombinasi kepastian regulasi dan rencana infrastruktur ini memperkuat ekspektasi kenaikan volume distribusi, sehingga investor merespons dengan menambah porsi kepemilikan

saham PGAS. Secara agregat, rangkaian peristiwa tersebut memetakan pola sentimen positif yang selaras dengan tren kenaikan harga di bursa.

Dalam beberapa tahun terakhir, media sosial telah menjadi sumber utama opini dan informasi bagi pelaku pasar. Aktivitas di platform seperti Twitter, Facebook, dan Instagram-baik berupa unggahan, komentar, maupun reaksi-mewakili sentimen kolektif terhadap emiten, termasuk PGAS. Perubahan intensitas dan polaritas sentimen tersebut diduga memiliki kaitan langsung dengan fluktuasi harga saham: misalnya, lonjakan diskusi positif dapat mendorong minat beli, sementara gelombang kritik atau kekhawatiran bisa memicu tekanan jual. Oleh karena itu, mengeksplorasi hubungan antara sentimen media sosial dan pergerakan harga PGAS penting untuk memahami mekanisme transmisi sinyal pasar ke level harga.

Oleh karena itu, dalam penelitian ini akan digunakan metode XGBoost dan LSTM (Long Short-Term Memory) untuk memprediksi harga saham PGAS. LSTM dipilih karena kemampuannya dalam memahami pola pergerakan harga saham berdasarkan data historis dan mempertimbangkan urutan waktu, sehingga sangat baik dalam menangkap tren jangka panjang. Sementara itu, XGBoost unggul dalam memproses berbagai faktor eksternal seperti indikator teknikal dan analisis sentimen pasar dari berita atau media sosial, memungkinkan model untuk membuat prediksi yang lebih akurat berdasarkan berbagai variabel. Dengan menggabungkan kedua metode ini, penelitian ini bertujuan untuk menghasilkan model prediksi yang lebih komprehensif yang tidak hanya mempertimbangkan data historis, tetapi juga faktor eksternal yang dapat memengaruhi pergerakan harga saham PGAS[4], [5].

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan yang telah dijabarkan, maka rumusan masalah yang ditemukan, yaitu:

1. Bagaimana hasil penerapan model *LSTM (LONG SHORT TERM MEMORY)* & XGBoost untuk memprediksi tren saham menggunakan data historis?
2. Bagaimana hasil keterhubungan antara fluktuasi harga saham dengan sentiment analisis di sosial media?

1.3 Batasan Masalah

Pembatasan berikut digunakan sebagai fokus penelitian, yaitu:

1. Menggunakan saham PGAS sebagai objek penelitian.
2. Menggunakan data range dari 15 Maret 2020 – 15 Maret 2025.
3. Menggunakan sosial media dari Twitter (X)

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Berdasarkan rumusan masalah yang dibentuk, maka tujuan penelitian ini, yaitu:

1. Menerapkan model XGBoost untuk mengolah fitur teknikal dan fitur sentimen Twitter dalam rangka meningkatkan akurasi prediksi harga saham PGAS.
2. Mengembangkan model *hybrid* yang menggabungkan LSTM dan XGBoost untuk memprediksi tren harga saham PGAS secara real-time berbasis gabungan data historis, indikator teknikal, dan sentimen media sosial.

1.4.2 Manfaat Penelitian

Terdapat beberapa manfaat dari penelitian yang dilakukan berdasarkan sudut pandang pengguna, peneliti, dan juga ilmu pengetahuan pada umumnya, yaitu:

1. Membantu para Investor dalam menghindari resiko dan juga kerugian dalam berinvestasi.

2. Memudahkan para Investor memberi informasi untuk memberi keputusan yang tepat untuk mengoptimalkan keuntungan dan meminimalisir kerugian.

1.5 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Kesimpulan dan Saran. Sistematika penulisan laporan adalah sebagai berikut:

- **Bab 1 PENDAHULUAN**

Bab 1 berisi uraian permasalahan yang diteliti, dimulai dari latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

- **Bab 2 LANDASAN TEORI**

Bab 2 berisi landasan teori yang berhubungan dengan penelitian yang dilakukan, meliputi teori terkait saham, tren saham, indikator teknis yang digunakan, dan model XGBoost, dan LSTM.

- **Bab 3 METODOLOGI PENELITIAN**

Bab 3 berisi metodologi penelitian yang digunakan pada penelitian, dimulai dari studi literatur, pengumpulan data, perancangan, implementasi, pengujian dan evaluasi, dan dokumentasi.

- **Bab 4 Analisis dan Hasil Penelitian**

Memuat uraian komprehensif atas seluruh proses penelitian mulai dari deskripsi karakteristik dan kinerja data yang telah disiapkan, tahapan pra-pemrosesan (merging, cleansing, transformation, feature engineering), pembangunan model LSTM dan XGBoost, evaluasi performa menggunakan metrik RMSE, MAE, dan R^2 , hingga interpretasi visualisasi hasil dalam bentuk grafik dan tabel yang

mendalam, sekaligus pembahasan kritis mengenai kelebihan, kelemahan, dan implikasi temuan terhadap pemahaman hubungan antara sentimen media sosial dan fluktuasi harga saham PGAS; sedangkan

- Bab 5 Kesimpulan dan Saran

Menyajikan ringkasan jawaban atas setiap rumusan masalah yang telah dirumuskan di Bab 1, menegaskan kontribusi hasil penelitian terhadap literatur analisis finansial, serta memberikan rekomendasi strategis untuk studi lanjutan-termasuk eksplorasi model alternatif, perluasan sumber data, peningkatan teknik feature engineering, dan penerapan framework MLOps-agar akurasi prediksi dan nilai praktis bagi investor dapat terus ditingkatkan.



BAB II

LANDASAN TEORI

2.1 Penelitian terkait

Berikut adalah penelitian terdahulu yang mengkaji korelasi antara sentimen media sosial dan pergerakan harga saham PGAS.

Tabel 2. 1 Penelitian terkait

Judul Artikel, Penulis	Objek Penelitian	Metode	Sumber Data	Findings
Stocks Price Forecasting Based On XGBoost and LSTM, Pham Hoang Vuong¹, Trinh Tan Dat¹, Tieu Khoi Mail, Pham Hoang Uyen² and Pham The Bao [6]	Pengembangan dan evaluasi sistem peramalan harga saham (stock-price forecasting), khususnya membandingkan kinerja metode gabungan XGBoost + LSTM dengan model ARIMA sebagai baseline.	XGBoost, LSTM, & ARIMA	Data Forex (1 Januari 2008 – 19 Maret 2018)	MSE: 3.465×10^{-7} (XGBoost+LSTM) vs 6.114×10^{-7} (ARIMA) MAE: 3.825×10^{-4} vs 4.149×10^{-4} RMSE: 5.887×10^{-4} vs 7.819×10^{-4} Hasil ini mengindikasikan peningkatan akurasi peramalan harga saham dengan memanfaatkan seleksi fitur berbasis XGBoost diikuti LSTM deep learning
Investing Deep Stock Market Forecasting With Sentiment Analysis, Charalampos M. Liapis, Aikaterini Karanikola and Sotiris Kotsiantis[7]	Mengevaluasi dan membandingkan efektivitas berbagai metode <i>deep learning</i> dan state-of-the-art untuk peramalan pasar saham dengan menggabungkan data hasil <i>sentiment analysis</i> ke dalam ruang fitur model.	LSTM, GRU, & VADER	Data Saham (2 Januari 2018 – 24 Desember 2020), & Data Twitter.	Menambahkan fitur berbasis sentimen (TextBlob, VADER, FinBert) Bersama dengan harga penutupan yang dihaluskan (rolling mean -7) umumnya meningkatkan akurasi peramalan.
A Deep Learning-Based LSTM for Stock Price Prediction Using	Mengembangkan dan mengevaluasi model peramalan harga saham	LSTM & XGBoost	Kaggle, VADER,& TextBlob	Latih model LSTM (dua layer LSTM 50 unit + <i>Dense</i> layer) dan

Twitter Sentiment Analysis, Shimaa Ouf1, Mona El Hawary, Amal Aboutabl, Sherif Adel[8]	berbasis <i>deep learning</i> (LSTM) dan <i>ensemble learning</i> (XGBoost) pada tiga perusahaan (Apple, Google, Tesla), dengan fokus pada pengaruh integrasi data sentimen Twitter terhadap akurasi prediksi harga saham			XGBoost (dengan <i>grid search</i> hyperparameter) pada split data 90% latih (termasuk validasi 10%) dan 10% uji.
ID-SMSA: Indonesia stock market dataset for sentiment analysis, Jason Hartanto, Timothy Liundi, Rhio Sutoyo, Esther Widhi Andangsari.[9]	Penelitian ini memperkenalkan ID-SMSA: Indonesian Stock Market Dataset for Sentiment Analysis , yaitu kumpulan data tweet dalam Bahasa Indonesia (dengan terjemahan Bahasa Inggris) yang membahas 10 emiten berkapitalisasi pasar terbesar di Bursa Efek Indonesia per Maret 2023. Total terdapat 3.288 tweet yang telah dilabeli sentimen positif, netral, atau negatif, menghasilkan dataset domain-spesifik pertama untuk analisis sentimen pasar saham Indonesia .	VADER	Crawldata X (Twitter), 12 Januari 2021 – 1 Maret 2024	ID-SMSA berisi 3.288 tweet tentang 10 emiten terbesar di BEI dengan distribusi sentimen 1.769 positif, 733 netral, 786 negatif dan tingkat kesepakatan anotator (Cohen's kappa) 0,779

Enhancing Sentiment analysis in stock market tweets through BERT-Based Knowledge Transfer, Emre Cicekyurt, Gokhan Bakal[10]	Pengaruh transfer pengetahuan berbasis BERT (FinBERT) untuk meningkatkan performa analisis sentimen tweet pasar saham .	BERT, LSTM, CNN,& XGBoost	Data Crawling Twitter (X), 9 Apr – 16 Jul 2020	Penambahan FinBERT-labeled data meningkatkan rata-rata F1-score tradisional sebesar 17% dan deep learning sebesar 20%, dengan model Random Forest unigram mencapai F1=69% (vs 54% awal), menegaskan korelasi positif kuat antara ukuran data latih dan performa klasifikasi
A Comparative Analysis of LSTM, ARIMA, XGBoost Algorithms in Predicting Stock Price Direction, Aiyegbeni Gifty , Dr. Yang Li[11]	Penelitian ini bertujuan mengumpulkan data histori harga penutupan saham Google selama lima tahun dan mengembangkan kerangka peramalan harga saham harian menggunakan algoritma XGBoost, LSTM, dan ARIMA .	XGBoost, ARIMA, & LSTM	Kaggle	XGBoost mencapai R ² 99,35%, MAE 17,63, RMSE 30,24. LSTM mencapai R ² 96,87%, MAE 49,35, RMSE 57,28. ARIMA mencapai R ² 66,37%, MAE 140,12, RMSE 188,11. XGBoost ter-tuning meningkat menjadi R ² 99,47%, MAE 15,98, RMSE 27,34. Gifty and Yang paper 20...Gifty and Yang paper 20... Ensemble tiga model mencetak R ² 78,63%, MAE 149,94, RMSE 104,64. Gifty and Yang paper 20...Gifty and Yang paper 20... Ensemble XGBoost + LSTM mencetak R ² 81,10%, MAE 107,20, RMSE 141,03.

2.2 Teori Penelitian

2.2.1 Saham

Saham merupakan bukti kepemilikan atas sebuah perusahaan. Seorang pemilik saham memiliki hak atas aset dan pendapatan dari perusahaan yang bersangkutan. Dua alasan utama seseorang membeli saham dari suatu perusahaan yaitu untuk mendapatkan bagian keuntungan yang diperoleh perusahaan atau disebut juga dividen dan menjual kembali saham dengan harga yang lebih tinggi. Saham yang dapat dijual dan dibeli masyarakat umum diperdagangkan di pasar modal. Bursa Efek Indonesia (BEI) merupakan pasar modal Indonesia dimana saham-saham perusahaan yang sudah terdaftar diperdagangkan. Sebuah perusahaan memperdagangkan sahamnya untuk mendapatkan modal tambahan. Agar sahamnya dapat diperdagangkan di pasar saham, sebuah perusahaan terlebih dahulu harus melalui proses *Initial Public Offering* (IPO). [12]

Tingkatan dukungan sosial dari lingkungan sekitar sangat memengaruhi kesiapan seseorang dalam mengambil keputusan. Saat individu merasakan adanya dorongan, pengakuan, atau dorongan moral dari keluarga, teman, atau rekan kerja, ia cenderung lebih percaya diri dan termotivasi untuk bertindak. Dukungan ini tidak hanya berupa pujian atau semangat verbal, tetapi juga bantuan konkret-misalnya, pemberian informasi, sumber daya, atau contoh perilaku positif-yang memperkuat keyakinan bahwa tindakan yang diambil akan mendapat hasil yang konstruktif. Tanpa adanya dukungan tersebut, niat baik sekalipun sering kali berhenti di tahap pemikiran karena rasa ragu atau takut akan kegagalan.

Selain dukungan sosial, niat seseorang untuk bertindak merupakan kompas internal yang mengarahkan keputusan. Niat ini terbentuk dari gabungan keyakinan atas manfaat tindakan, norma pribadi, dan ekspektasi terhadap hasil. Semakin kuat persepsi bahwa tindakan akan membawa keuntungan-baik secara materi, emosional, maupun status-semakin besar energi psikologis yang dicurahkan untuk merencanakan dan melaksanakan langkah konkret. Dalam konteks pengambilan keputusan, niat juga dipengaruhi oleh self-efficacy sejauh mana individu merasa

memiliki kemampuan dan kontrol atas proses tersebut. Tanpa niat yang jelas dan kuat, seseorang mudah terombang-ambing oleh keraguan meski kondisi eksternal mendukung.

Namun, niat dan dukungan sosial saja tidak cukup jika situasi yang memungkinkan tidak hadir. Keberadaan yang tepat-misalnya peluang yang nyata, waktu yang sesuai, atau kondisi lingkungan yang kondusif-akan menentukan apakah tindakan benar-benar direalisasikan. Situasi ini mencakup aspek fisik (akses ke sarana, kondisi cuaca, lokasi), kontekstual (aturan, kebijakan, budaya organisasi), dan temporal (deadline, ketersediaan waktu). Misalnya, meski seseorang sangat berniat memulai usaha baru dan didukung penuh, tanpa modal awal atau izin usaha yang memadai, niat itu hanya akan mengendap. Dengan demikian, teori Snehendu B. Kar menegaskan bahwa interaksi sinergis antara dukungan sosial, niat internal, dan kondisi situasional menjadi kunci utama agar individu benar-benar bertindak sesuai keputusan yang diambil[11]. Membahas akan permasalahan sebelumnya, maka dapat diambil point penting bahwa terdapat situasi yang memungkinkan untuk para investor tersebut mengambil suatu tindakan dalam menentukan permasalahan saham. Suatu tindakan tersebut bisa berasal dari hasil analisa yang digunakan untuk menentukan akan permasalahan dari saham tersebut. Mengetahui akan itu, maka saham bisa menjadi point bagi para investor tersebut untuk mengambil tindakan apa yang harus dilakukan dari hasil analisa menggunakan metode tertentu. Namun, dalam berinvestasi saham, tentu saja ada risiko yang perlu diperhatikan. Harga saham dapat berubah-ubah dengan cepat dan seringkali sulit diprediksi. Mengingat para investor kerap kali bertujuan untuk mendapatkan keuntungan sebesar - besarnya, maka analisa akan bagaimana saham itu bekerja di masa depan menjadi pekerjaan mereka. Untuk mendapatkan keuntungan yang besar, para investor dapat melakukan analisis dalam memprediksi harga saham [12].

2.2.2 Investasi

Investasi adalah sebuah komitmen untuk menempatkan sejumlah dana kepada suatu proyek dengan adanya harapan untuk mendapatkan keuntungan di masa yang akan datang[14]. Investasi melibatkan penempatan dana pada berbagai instrumen keuangan atau aset riil-mulai dari saham, obligasi, dan reksa dana hingga logam mulia seperti emas, properti tanah, bahkan aset digital seperti cryptocurrency. Masing-masing kelas aset ini memiliki karakteristik risiko, likuiditas, dan potensi imbal hasil yang berbeda. Saham bisa memberi pertumbuhan modal dan pembagian keuntungan secara rutin, tetapi juga bisa berubah-ubah tergantung pasar. Emas biasanya dianggap sebagai pilihan aman ketika ekonomi tidak stabil, meskipun tidak memberi hasil uang secara berkala. Sementara itu, *cryptocurrency* berpotensi memberi keuntungan besar dalam waktu singkat, tetapi sangat fluktuatif dan membutuhkan pemahaman tentang teknologi serta aturan yang dalam.

Tujuan utama berinvestasi adalah mencapai imbal hasil yang melebihi besaran modal awal, sehingga membangun aliran pendapatan pasif-seperti dividen dari saham atau sewa properti-yang dapat menambah atau bahkan menggantikan penghasilan aktif. Dengan pendapatan pasif yang terkelola baik, investor dapat memanfaatkan efek bunga majemuk, di mana keuntungan yang diperoleh di periode pertama diinvestasikan kembali untuk menghasilkan keuntungan tambahan pada periode berikutnya. Strategi ini sering dijadikan tolok ukur bagi pencapaian kebebasan finansial (*financial freedom*), di mana seseorang memiliki aset dan arus kas cukup untuk menutup biaya hidup tanpa ketergantungan pada gaji pekerjaan tetap.

Namun, mencapai tujuan keuangan jangka panjang membutuhkan perencanaan yang matang, seperti mencermati diversifikasi portofolio untuk mengurangi risiko, menentukan masa investasi sesuai dengan tingkat kemampuan mengambil risiko, serta memiliki disiplin untuk tidak tergoda melakukan spekulasi berdasarkan perkiraan pasar. Investor harus memahami bahwa setiap instrumen memiliki siklus dan reaksi yang berbeda terhadap kondisi ekonomi, geopolitik, atau inovasi teknologi. Dengan menyesuaikan alokasi dana berdasarkan tujuan, toleransi risiko,

dan periode investasi, seorang investor dapat menavigasi berbagai tantangan pasar dan memaksimalkan peluang untuk meraih kebebasan finansial yang berkelanjutan. [15]. Investasi jangka pendek umumnya memiliki horizon waktu kurang dari satu hingga dua tahun dan sering kali dipilih oleh investor yang menginginkan likuiditas tinggi serta peluang keuntungan cepat. Dalam praktiknya, investor akan memantau pergerakan pasar saham harian-misalnya menargetkan saham perusahaan dengan tren pertumbuhan positif-lalu menjual kembali begitu harga telah naik sesuai ekspektasi. Strategi ini memungkinkan realisasi keuntungan yang jauh lebih besar dibandingkan simpanan deposito, meski memerlukan ketelitian dalam membaca grafik dan kesiapan menghadapi fluktuasi harga yang tinggi.

Investasi jangka menengah mencakup periode tiga hingga lima tahun dan cocok untuk tujuan finansial yang telah direncanakan, seperti membiayai pendidikan anak. Pada jenis investasi ini, pilihan instrumen sering jatuh pada produk yang menawarkan keseimbangan antara risiko dan imbal hasil, seperti reksa dana campuran atau obligasi korporasi berkualitas. Dengan jangka waktu yang lebih panjang, investor dapat meredam volatilitas pasar dan memanfaatkan efek akumulasi bunga atau dividen, sehingga modal awal berpotensi tumbuh lebih stabil dan terukur.

Sementara itu, investasi jangka panjang biasanya berlangsung lebih dari sepuluh tahun dan menekankan pertumbuhan nilai aset secara signifikan. Contoh klasiknya adalah pembelian properti atau tanah, di mana investor memperoleh passive income dari penyewaan sekaligus meraup keuntungan atas kenaikan nilai pasar properti dalam dekade berikutnya. Dengan sabar menahan periode penyesuaian harga dan memanfaatkan tren kenaikan nilai aset, investor jangka panjang berpeluang mencapai akumulasi kekayaan yang substansial dan mendekatkan diri pada kebebasan finansial [16].

2.2.3 Deep Learning

Deep learning dengan *Long Short-Term Memory* (LSTM) unggul dalam menangkap ketergantungan jangka panjang pada data deret waktu-misalnya pola

musiman atau momentum harga-karena arsitekturnya yang mampu mengingat informasi dari rentang waktu yang luas dan mengurangi masalah vanishing gradient pada RNN konvensional. Sementara itu, *Extreme Gradient Boosting* (XGBoost) memanfaatkan ensemble pohon keputusan untuk memperbaiki prediksi dengan fokus pada meminimalkan residual dari model sebelumnya, menjadikannya sangat efektif dalam mengolah fitur teknis dan sentimen yang bersifat tabular. Dengan menggabungkan LSTM sebagai model urutan (*sequence model*) dan XGBoost untuk menangani sisa kesalahan (*residual*), studi seperti yang dilakukan Pham et al. (2021) menunjukkan peningkatan akurasi prediksi harga saham dibandingkan metode tradisional seperti ARIMA maupun model Tunggal. Optimalisasi hyperparameter pada LSTM juga terbukti meningkatkan kinerja dalam tugas-tugas prediksi risiko keuangan, bahkan mengungguli Random Forest dan XGBoost murni pada beberapa metrik seperti AUC. Teknologi ini memanfaatkan jaringan saraf tiruan (*artificial neural networks*) yang memiliki banyak lapisan untuk memproses data yang sangat besar dan kompleks. Dalam konteks ini, deep learning sangat efektif dalam mengekstraksi fitur-fitur tersembunyi dari data, yang sebelumnya sulit dikenali dengan metode analisis konvensional. Dengan kemampuannya untuk belajar secara otomatis dari data, deep learning tidak memerlukan ekstraksi fitur manual, yang menjadikannya alat yang sangat kuat dalam analisis keuangan, terutama untuk mendeteksi pola dan tren yang mungkin tidak terlihat oleh analis manusia.[13]

Penggunaan deep learning dalam analisis data keuangan juga sangat membantu dalam meningkatkan transparansi dan akurasi dalam pengambilan keputusan. Dalam hal ini, model-model deep learning dapat mengidentifikasi anomali atau potensi kecurangan dalam laporan keuangan, yang dapat membantu perusahaan dan investor mengurangi risiko dan membuat keputusan yang lebih terinformasi. Selain itu, dengan kemampuan untuk menganalisis data dalam jumlah besar dan mengidentifikasi pola tersembunyi, teknologi ini juga memungkinkan prediksi yang lebih akurat tentang pergerakan harga pasar atau perilaku keuangan lainnya. Seiring dengan semakin banyaknya data yang tersedia, penerapan deep learning

dalam bidang keuangan diharapkan akan terus berkembang, meningkatkan efisiensi dan efektivitas analisis dalam berbagai aspek keuangan dan ekonomi.[14]

2.2.4 Analisis Sentimen

Secara kegunaan, analisis sentimen merupakan pengetahuan atau teknik yang digunakan untuk mencari emosi berdasarkan ulasan, baik dalam bentuk teks maupun verbal [23]. Teknik ini sering digunakan untuk mengeksplorasi perasaan yang terkandung dalam ulasan pelanggan terhadap produk tertentu. Dalam menggunakan analisis sentimen, maka diperlukan sebuah dataset yang telah memiliki label. Pendekatan dengan metode pembelajaran supervised merupakan pendekatan menggunakan pembelajaran mesin dan kecerdasan buatan yang memanfaatkan data dengan sifat telah terlabeli. Model tersebut akan mampu menghasilkan output yang akurat dan sesuai dengan label yang diinginkan [4]. Untuk memenuhi pelabelan terkait analisis sentimen, maka diperlukan suatu ulasan sebagai sumber data. Ulasan ini seringkali mengandung opini positif dan negatif tentang perubahan yang terjadi [24]. Pelabelan perlu untuk dilakukan, sebab dalam penelitian analisis sentimen, maka penelitian ini mencoba untuk mengklasifikasikan terkait apakah teks ulasan yang ditulis pengguna bersifat positif atau negatif [25]. Ulasan tersebut memberikan wawasan tentang dampak yang dihasilkan. Ini membantu dalam konteks analisis sentimen terkait ulasan teks untuk memahami pendapat yang terkait dengan masalah atau objek tertentu, apakah cenderung positif atau negatif [5]. Serta, dalam analisis sentimen merupakan teknik yang baik untuk memeriksa bagaimana masyarakat merespon akan terjadinya perubahan [26].

2.2.5 Text Mining

Pada tahap *Text Mining* pertama-tama kita membuat salinan data mentah (`merged_df`) ke variabel `df_cleaned` agar perubahan tidak merusak data asli. Selanjutnya, setiap entri teks cuitan (`'full_text'`) dibersihkan melalui fungsi `clean_text()`, yang mencakup:

1. Penghapusan URL (semua pola `http...`, `www...`, atau `https...`) agar tautan tidak memengaruhi analisis,
 2. Penghapusan mention (`@username`) dan hashtag (`#tag`) karena kedua elemen ini bersifat metadata media sosial, bukan konten opini,
 3. Penghapusan angka untuk memfokuskan pada kata-kata bermakna,
 4. Penghapusan tanda baca non-alfanumerik supaya tokenisasi berjalan lebih bersih,
 5. Konversi seluruh teks menjadi huruf kecil dan pemangkasan spasi di kedua ujung teks.
- Hasilnya, kolom `clean_text` berisi versi teks yang siap di-tokenize dan diekstraksi fiturnya.

Setelah pembersihan, langkah berikutnya adalah inisialisasi *sentiment analyzer* menggunakan pustaka VADER (SentimentIntensityAnalyzer dari NLTK). VADER dipilih karena keandalannya dalam menganalisis teks singkat dan informal seperti Twitter, termasuk penanganan emotikon, intensifier, dan variasi huruf kapital. Dengan objek `sia` ini, setiap `clean_text` dihitung skor *compound*-nya, yaitu nilai kontinu antara -1 (sangat negatif) hingga $+1$ (sangat positif).

Untuk mengubah skor kontinu menjadi kategori sentiment yang mudah diinterpretasi, dibuat fungsi `get_sentiment(text)`:

1. Jika skor ≥ 0.05 maka dikategorikan “positif”,
2. Jika skor ≤ -0.05 maka “negatif”,
3. Sebaliknya masuk kategori “netral”.

Pemilihan ambang batas ini mengikuti rekomendasi default VADER untuk memisahkan sentimen kuat dan lemah.

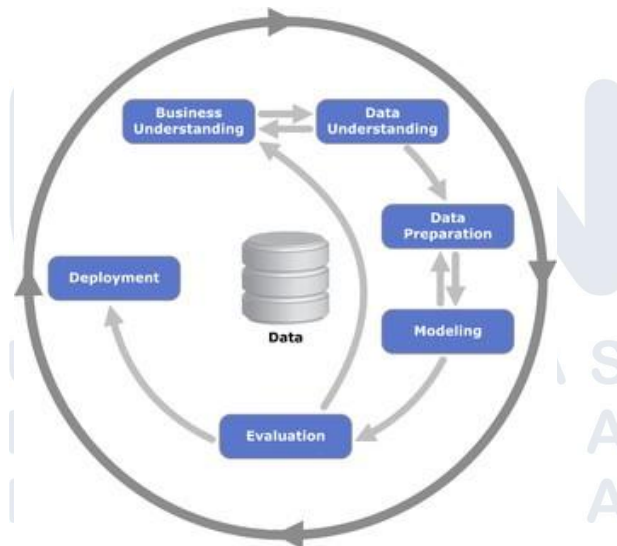
Terakhir, fungsi `get_sentiment` diterapkan ke seluruh kolom `clean_text`, menghasilkan kolom baru sentiment di `df_cleaned`. Untuk meninjau hasil awal, kita menampilkan 100 baris pertama dari kolom `clean_text` dan sentiment. Distribusi awal ini akan menjadi dasar validasi kualitas ekstraksi sentimen sebelum digabungkan sebagai fitur dalam model prediksi harga saham. Dengan demikian,

seluruh rangkaian kode di atas membentuk fondasi text mining-dari pembersihan teks hingga pengkodean opini-yang siap diintegrasikan ke dalam analisis kuantitatif selanjutnya.

2.3 Framework/Algoritma/SDLC Penelitian

2.3.1 CRISP-DM

Cross Industry Standard Process For Data Mining atau disingkat menjadi CRISP-DM dikembangkan pada tahun 1996 yang bertujuan untuk dijadikan *methodology standard non-proprietary* bagi data mining, ini ditujukan untuk melakukan proses analisis strategi pemecahan masalah dari suatu bisnis atau penelitian di suatu industri[15]. Berikut merupakan enam tahap dalam metodologi CRISP-DM, seperti business understanding, data understanding, data preparation, modeling, evaluation, dan deployment. Berikut merupakan langkah-langkah dalam mengimplementasikan CRISP-DM, yaitu:



Gambar 2. 1 Framework CRISP-DM

1. Business Understanding

Tahap Business Understanding adalah langkah pertama yang sangat krusial dalam pengembangan model data mining. Pada tahap ini, dilakukan pemahaman mendalam terhadap masalah bisnis yang ingin diatasi, termasuk mengidentifikasi tujuan dan kebutuhan dari model yang akan dikembangkan. Proses ini bertujuan untuk memastikan bahwa model yang dibangun benar-benar relevan dan dapat memberikan solusi terhadap masalah yang dihadapi bisnis. Selain itu, tahap ini juga melibatkan penentuan ruang lingkup proyek, serta pemahaman tentang ketersediaan data yang ada dan sumber data yang bisa dimanfaatkan. Poin penting lainnya adalah pemetaan peran dan dampak model terhadap tujuan bisnis yang lebih besar, sehingga hasil akhirnya dapat memberikan kontribusi nyata terhadap keputusan strategis perusahaan.

2. Data Understanding

Pada tahap Data Understanding, fokus utama adalah untuk memperoleh pemahaman yang lebih mendalam tentang data yang tersedia. Proses ini mencakup analisis awal terhadap dataset yang dimiliki, dengan menggunakan statistik deskriptif untuk mendapatkan gambaran umum tentang distribusi data, hubungan antar variabel, dan potensi pola yang bisa muncul. Visualisasi data digunakan untuk menggali insight lebih lanjut dan membantu dalam mendeteksi outlier yang mungkin mempengaruhi kualitas model. Selain itu, tahap ini juga melibatkan identifikasi masalah dalam data seperti inkonsistensi, duplikasi, atau data yang hilang, yang dapat mempengaruhi efektivitas analisis. Data yang telah dianalisis akan memberikan gambaran lebih jelas tentang kualitas dan relevansi data yang digunakan untuk penelitian selanjutnya.

3. Data Preparation

Data Preparation adalah tahap penting yang memastikan bahwa data yang digunakan dalam pemodelan berada dalam kondisi terbaik. Pada tahap ini, dilakukan pembersihan data untuk menghilangkan kesalahan, inkonsistensi, atau duplikasi yang mungkin ada dalam dataset. Proses ini juga mencakup

penggabungan data dari berbagai sumber yang relevan, serta pengisian data yang hilang atau penghapusan data yang tidak relevan agar analisis berikutnya lebih akurat. Selain itu, transformasi data seperti normalisasi, encoding, atau pembuatan fitur baru juga dilakukan untuk memudahkan model dalam mengidentifikasi pola dan hubungan yang signifikan. Data yang sudah dipersiapkan dengan baik akan mempengaruhi kualitas dan akurasi dari model yang akan dibangun, sehingga tahap ini sangat penting untuk memastikan hasil yang optimal.

4. Modelling

Tahap Modeling adalah inti dari proses pengembangan model data mining. Pada tahap ini, dilakukan pemilihan algoritma yang paling sesuai dengan permasalahan yang ada, serta penyesuaian parameter yang diperlukan agar algoritma dapat bekerja secara maksimal. Proses ini melibatkan percobaan dengan berbagai model yang mungkin, seperti model regresi, klasifikasi, atau clustering, tergantung pada jenis masalah yang ingin diselesaikan. Selain itu, dilakukan pengujian terhadap performa model pada dataset yang tersedia untuk menentukan apakah model tersebut sudah memberikan hasil yang diinginkan. Penyesuaian dan optimasi terus dilakukan pada setiap iterasi untuk mendapatkan model terbaik yang dapat menggambarkan pola dalam data secara akurat. Keberhasilan tahap ini sangat bergantung pada pemilihan algoritma yang tepat dan kemampuan untuk mengoptimalkan parameter model.

5. Evaluation

Pada tahap Evaluation, model yang telah dibangun diuji untuk memastikan bahwa kinerjanya sesuai dengan ekspektasi dan dapat diaplikasikan pada data yang belum pernah dilihat sebelumnya (data uji). Proses evaluasi ini melibatkan pengukuran berbagai metrik kinerja, seperti akurasi, precision, recall, atau F1-score, tergantung

pada jenis masalah yang dihadapi. Selain itu, dilakukan juga analisis terhadap kekuatan dan kelemahan model dengan menggunakan data yang beragam untuk melihat apakah model mampu generalisasi dengan baik. Evaluasi ini bertujuan untuk memastikan bahwa model yang dibangun tidak hanya bekerja dengan baik pada data latihan, tetapi juga dapat diandalkan pada data nyata yang akan dihadapi di dunia bisnis. Hasil evaluasi ini akan memberikan gambaran tentang seberapa efektif model dalam memecahkan masalah yang ada.

6. Deployment

Tahap terakhir adalah Deployment, yaitu implementasi model yang telah dikembangkan ke dalam lingkungan produksi. Pada tahap ini, model yang sudah diuji dan dievaluasi diintegrasikan ke dalam sistem yang ada dan digunakan untuk membuat prediksi pada data baru secara real-time. Selain itu, dilakukan pemantauan terhadap performa model untuk memastikan bahwa model tetap berjalan dengan baik dan memberikan hasil yang akurat seiring berjalannya waktu. Pada tahap deployment ini, sering kali diperlukan pengaturan pemeliharaan dan pembaruan model secara periodik agar model tetap relevan dengan perubahan kondisi bisnis atau data yang baru. Keberhasilan tahap deployment akan mengarah pada penggunaan model dalam pengambilan keputusan bisnis yang nyata, memberikan manfaat langsung bagi organisasi atau perusahaan.

2.3.2 *XGBoost*

XGBoost atau *Extreme Gradient Boosting* adalah salah satu algoritma Deep Learning berbasis pohon keputusan (decision tree) yang sangat populer karena kecepatan dan akurasi prediksinya. Algoritma ini dikembangkan berdasarkan teknik boosting, di mana beberapa model lemah (weak learners)-biasanya berupa decision tree-digabungkan secara berurutan untuk membentuk model yang kuat (strong learner). Tujuannya adalah untuk memperbaiki kesalahan prediksi dari model sebelumnya, hingga tercipta model akhir yang lebih akurat dan andal.[16]

Cara kerja XGBoost dapat diibaratkan sebagai proses kolaboratif di mana model-model sederhana saling melengkapi untuk meningkatkan akurasi. Model pertama akan memprediksi hasil dasar, kemudian model berikutnya fokus memperbaiki kesalahan yang dibuat oleh model sebelumnya. Proses ini berlangsung secara berulang hingga keseluruhan kesalahan dapat diminimalkan. Dalam konteks ini, setiap pohon keputusan bertugas untuk mempelajari kesalahan yang tersisa dan terus menyempurnakan prediksi akhir secara bertahap. [17] [18]

XGBoost dikenal unggul dalam hal kecepatan dan kinerja, karena mampu menangani data berukuran besar dengan efisien dan tetap menghasilkan prediksi yang akurat. Algoritma ini sering digunakan dalam berbagai kompetisi data science karena keandalannya. Selain itu, XGBoost dilengkapi dengan mekanisme regularisasi yang efektif, sehingga model yang dihasilkan lebih stabil dan tidak mudah mengalami overfitting, yaitu kondisi saat model terlalu menyesuaikan diri dengan data pelatihan sehingga kurang mampu menggeneralisasi pada data yang belum pernah dilihat.

Dalam penelitian ini, XGBoost digunakan untuk membantu memprediksi harga saham PGAS dengan memperhitungkan berbagai faktor teknikal dan eksternal, termasuk sentimen berita atau sosial media. Tidak seperti LSTM yang fokus pada pola berurutan dalam waktu, XGBoost lebih fleksibel dalam mengolah berbagai jenis fitur sekaligus, bahkan yang tidak berurutan. Oleh karena itu, XGBoost menjadi pelengkap yang ideal bagi LSTM, karena dapat memperkaya prediksi dengan memasukkan variabel-variabel penting yang tidak ditangkap oleh LSTM.

2.3.3 LSTM (*Long Short Term Memory*)

LSTM atau *Long Short-Term Memory* merupakan salah satu bentuk dari *Recurrent Neural Network* (RNN) yang dirancang khusus untuk mengingat informasi dalam jangka waktu yang lebih panjang. Berbeda dengan RNN biasa yang mudah lupa terhadap informasi lama karena masalah *vanishing gradient*, LSTM mengatasi kelemahan ini dengan cara mengatur informasi mana yang perlu diingat dan mana yang perlu dilupakan melalui sistem gerbang (*gates*). [5], [19], [20]

Secara sederhana, LSTM bekerja seperti memori manusia: kita tidak mungkin mengingat semua hal, tapi kita bisa menyimpan informasi penting untuk digunakan di masa depan. Nah, dalam LSTM, hal ini dilakukan melalui tiga komponen utama, yaitu *forget gate*, *input gate*, dan *output gate*. Ketiganya bekerja sama untuk menyaring informasi-misalnya, apakah informasi dari hari kemarin masih penting untuk hari ini? Atau apakah pola dua minggu lalu punya pengaruh terhadap prediksi harga esok hari?

LSTM sangat cocok digunakan untuk kasus prediksi harga saham karena data saham bersifat berurutan (*time-series*), di mana nilai saat ini bergantung pada nilai sebelumnya. Maka dari itu, LSTM dapat mempelajari pola-pola dari waktu ke waktu dan melakukan prediksi dengan lebih akurat dibanding metode biasa. Dalam penelitian ini, LSTM digunakan untuk mempelajari pergerakan harga saham PGAS berdasarkan data historisnya. Dengan mempertimbangkan konteks waktu dan tren sebelumnya, LSTM dapat membantu menghasilkan prediksi harga yang lebih realistis.[21]

2.3.4 VADER

VADER (*Valence Aware Dictionary and Sentiment Reasoning*) adalah model analisis sentimen yang dirancang khusus untuk teks media sosial. Model ini menggabungkan fitur leksikal yang dikurasi oleh manusia dengan aturan linguistik untuk menganalisis intensitas dan polaritas sentimen. Berbeda dengan alat analisis sentimen tradisional yang sering mengandalkan klasifikasi biner (positif atau negatif), VADER dapat mengukur intensitas sentimen dalam skala, memberikan analisis yang lebih rinci.[22]

Kunci efektivitas VADER terletak pada leksikonnya, yang mencakup lebih dari 7.500 fitur leksikal, seperti kata-kata, emotikon, akronim, dan istilah slang, yang semuanya diberi nilai intensitas sentimen. Fitur-fitur ini digabungkan dengan lima aturan umum yang memperhitungkan konvensi gramatikal dan sintaksis yang sering digunakan untuk menekankan sentimen, seperti tanda baca (misalnya, tanda seru), kapitalisasi (misalnya, huruf besar semua), dan pengubah derajat (misalnya, "sangat", "hampir").

Kinerja model ini diuji terhadap beberapa metode analisis sentimen yang telah mapan, termasuk LIWC, *SentiWordNet*, dan model pembelajaran mesin seperti *Naive Bayes* dan *Support Vector Machines*. Hasilnya menunjukkan bahwa VADER mengungguli model-model ini dalam beberapa konteks, khususnya dalam menganalisis konten media sosial seperti tweet, yang sering menggunakan bahasa informal dan emotikon. Menariknya, VADER bahkan mengalahkan pengklasifikasi manusia individu dalam mengklasifikasikan sentimen tweet.

Keunggulan VADER terletak pada kesederhanaannya, efisiensi komputasi, dan kemudahan penggunaannya. Model ini tidak memerlukan data pelatihan yang ekstensif atau model yang kompleks, menjadikannya sangat cocok untuk analisis sentimen secara real-time di platform media sosial. Selain itu, model ini mudah diperluas dan dipahami, yang sangat berguna bagi pengguna di luar bidang ilmu komputer, seperti sosiolog, psikolog, atau pemasar.

Secara keseluruhan, pendekatan berbasis aturan VADER menyediakan alat yang kuat untuk analisis sentimen, terutama untuk teks pendek dan informal yang khas di media sosial, dan mengungguli model-model lain dengan efisiensi dan kesederhanaan yang lebih besar.

2.4 Tools Penelitian

Pada penelitian ini, beberapa perangkat lunak dipilih untuk mendukung pengolahan data, analisis sentimen, serta visualisasi alur kerja.

2.4.1 Python

Python adalah bahasa pemrograman tingkat tinggi yang dirancang untuk memudahkan pengembangan perangkat lunak dengan sintaks yang jelas dan mudah dipahami. Dikenal karena kemampuannya yang serbaguna, Python digunakan dalam berbagai aplikasi mulai dari pengembangan web hingga analisis data, kecerdasan buatan (AI), dan machine learning. Kelebihan utama Python adalah sintaks yang bersih dan konsisten, yang memungkinkan pengembang untuk menulis kode yang lebih sedikit namun tetap kuat. Hal ini membuat Python sangat populer di kalangan pemula dan profesional di berbagai bidang. Salah satu fitur utama dari

Python adalah keberagaman pustaka dan framework yang mendukung pengembangan aplikasi, seperti *NumPy*, *Pandas*, *TensorFlow*, dan *Flask*, yang membuatnya menjadi pilihan utama di bidang analisis data dan pemrograman ilmiah.



Gambar 2. 2 Logo Python

Python juga memiliki komunitas yang besar dan aktif, yang secara teratur memperbarui dan mendukung pustaka serta alat pengembangan. Penggunaan Python terus berkembang, dan berkat berbagai aplikasi, kemudahan penggunaan, serta berbagai pustaka dan modul yang tersedia, Python telah menjadi bahasa yang sangat penting dalam dunia teknologi dan pengembangan perangkat lunak modern.[23]

2.4.2 Jupyter Notebook

Jupyter Notebook merupakan implementasi praktis dari konsep literate programming, di mana kode, dokumentasi, dan hasil eksekusi (seperti grafik atau tabel) digabung dalam satu dokumen interaktif. Dokumen ini disimpan dalam format JSON dengan ekstensi *.ipynb*, memuat urutan *cells* baik kode, Markdown, maupun output lengkap dengan metadata kernel (misalnya versi bahasa dan pustaka yang digunakan). Keunggulan utamanya terletak pada kemampuannya mendukung eksplorasi data secara iteratif: peneliti atau developer dapat menjalankan potongan kode (per *cell*) secara terpisah, memodifikasi, dan langsung melihat dampaknya, sehingga mempercepat proses *debugging* dan pengembangan prototipe.



Gambar 2. 3 Gambar Jupyter

Di balik antarmuka web-nya, Jupyter Notebook memisahkan dua komponen kunci: frontend dan kernel. Frontend, yang berjalan di browser, menampilkan editor kode, toolbar, dan panel hasil eksekusi; sementara kernel dijalankan sebagai proses terpisah di mesin lokal atau server bertugas mengeksekusi perintah pengguna. Komunikasi antara keduanya difasilitasi oleh *ZeroMQ* (ZMQ), pustaka *asynchronous messaging* yang menggunakan pola *publish-subscribe* untuk *broadcast* event seperti output baru, dan pola *request-reply* untuk permintaan eksekusi kode. Semua pesan ini dikemas dalam format JSON sehingga mudah diurai dan diperluas, memungkinkan penambahan protokol baru atau integrasi dengan layanan eksternal.

Ekosistem Jupyter Notebook didukung oleh ekstensifitas *kernel* dan *extension*. Kernel default mendukung Python, tetapi ada juga *kernel* untuk R, Julia, Scala, dan puluhan bahasa pemrograman lainnya. Selain itu, komunitas pengembang telah menciptakan ratusan *extension* mulai untuk visualisasi data (misalnya *interactive widgets*), integrasi Git, hingga *code linting* dan *autocomplete* yang canggih. Keterbukaan format dokumen dan arsitektur plugin-based ini menjadikan Jupyter Notebook tidak hanya alat analisis data, tetapi juga platform kolaborasi tertulis dari

laporan ilmiah, tutorial interaktif, hingga presentasi langsung yang dapat dijalankan ulang (*reproducible*) di berbagai lingkungan komputasi.

2.4.3 Google Colab

Google Colab memperluas fungsionalitas dari Jupyter Notebook dengan menerapkan virtualisasi melalui Docker dalam infrastruktur *Google Cloud Platform* (GCP), memungkinkan pengguna untuk menjalankan notebook dalam lingkungan runtime yang terisolasi. Setiap instance notebook di Colab dikelola menggunakan Kubernetes, yang mengatur kontainer secara otomatis untuk memastikan skalabilitas dan ketersediaan. Lapisan keamanan diperkuat dengan sistem antivirus yang memindai kontainer sebelum eksekusi kode, sehingga pengguna dapat bekerja tanpa khawatir terhadap malware atau kode berbahaya yang tersembunyi.



Gambar 2. 4 Gambar Logo Google Colab

Komunikasi antara antarmuka pengguna yang diakses melalui browser dan mesin eksekusi di backend memanfaatkan protokol gRPC dan HTTPS. Ketika pengguna mengirim perintah eksekusi, frontend mengemas permintaan dalam format JSON dan mengirimkannya melalui gRPC ke service orchestrator. Dari sana, gRPC juga menghubungkan orchestrator dengan kernel Python yang berjalan di dalam kontainer, memungkinkan eksekusi kode secara real-time dengan latensi minimal.

Setiap sesi notebook di Colab dilengkapi dengan manajemen otomatis terhadap penyimpanan dan versi dokumen. Google Drive diintegrasikan melalui API Drive v3, sehingga setiap perubahan-termasuk output grafik, tabel, dan checkpoint-disinkronkan secara otomatis ke akun Drive pengguna. Mekanisme checkpointing ini menjaga kontinuitas data: apabila koneksi terputus atau runtime kadaluarsa, pengguna dapat melanjutkan pekerjaan persis dari titik terakhir.

Salah satu fitur unggulan Colab adalah kemampuan untuk menggunakan akselerator perangkat keras-GPU dan TPU-yang dialokasikan secara dinamis melalui resource scheduler GCP. Pengguna tinggal memilih tipe akselerator yang dibutuhkan, lalu sistem akan menjadwalkan kontainer pada node yang dilengkapi perangkat keras tersebut. Kombinasi virtualisasi container dan penjadwalan sumber daya cloud memastikan bahwa komputasi intensif dapat dilakukan dengan efisien tanpa memerlukan infrastruktur lokal.

2.4.4 Draw.io

Draw.io, yang kini dikenal sebagai diagrams.net, mengimplementasikan representasi internal berbasis XML untuk menyimpan diagram. Setiap elemen dalam diagram-termasuk shape, konektor, dan label-dikonversi menjadi node XML yang dilengkapi dengan atribut posisi, ukuran, dan gaya. Pendekatan ini memungkinkan pengelolaan data diagram yang terstruktur dan efisien. Di sisi *frontend*, *diagrams.net* menggunakan *JavaScript* dan *SVG (Scalable Vector Graphics)* untuk rendering vektor di browser, memanfaatkan model *Document Object Model (DOM)* untuk manipulasi elemen dan *Canvas API* untuk rendering grafis lebih kompleks. Dengan pendekatan berbasis web ini, pengguna dapat dengan mudah mengakses dan berinteraksi dengan diagram di dalam browser tanpa memerlukan perangkat lunak tambahan.

Untuk mendukung kolaborasi, diagrams.net memungkinkan penyimpanan backend melalui platform seperti Google Drive atau GitHub, menggunakan API RESTful atau protokol Git untuk commit dan pull data diagram. Fitur ini memungkinkan integrasi yang mulus dengan sistem kontrol versi dan kolaborasi real-time antara

pengguna. Untuk menghindari konflik saat pengeditan bersama, diagrams.net menerapkan konsep *optimistic locking*, yang memastikan bahwa hanya satu pengguna yang dapat mengedit bagian tertentu dari diagram pada waktu yang bersamaan, sehingga meningkatkan efisiensi dan mencegah masalah dalam proses kolaborasi.

2.4.5 X (Twitter)

Twitter kini dikenal sebagai X merupakan platform media sosial yang memungkinkan pengguna berbagi informasi secara singkat dalam bentuk pesan singkat (*tweet*). Platform ini dirancang untuk menyampaikan berita, opini, dan interaksi secara cepat dalam skala global. Pengguna dapat mem-follow akun lain, membalas tweet, membagikan ulang (*retweet*), serta menyukai dan menyematkan konten. Twitter juga dikenal karena sistem hashtag dan trending topic-nya yang memungkinkan topik tertentu menjadi viral dan diperbincangkan secara luas. Hal ini menjadikan Twitter sebagai alat penting dalam membentuk opini publik dan menyebarkan informasi real-time.



Gambar 2. 5 Logo X

Rebranding Twitter menjadi X yang dilakukan oleh perusahaan induknya (di bawah kepemilikan Elon Musk) merupakan langkah besar dalam merombak identitas platform tersebut. Meskipun secara teknis fitur dasarnya tetap serupa, perubahan nama menjadi X menandai ambisi untuk menjadikan platform ini lebih dari sekadar

media sosial—yakni sebagai “super app” yang mencakup pembayaran, streaming, dan komunikasi multifungsi. Dalam konteks penelitian, X tetap berfungsi sebagai sumber data utama untuk analisis sentimen publik karena karakter tweet-nya yang padat, spontan, dan merefleksikan opini pasar secara langsung. Tweet-tweet ini sangat berharga dalam menganalisis persepsi publik terhadap isu tertentu, termasuk pergerakan harga saham seperti PGAS.

Sebagai salah satu sumber data real-time terbesar, X memainkan peran penting dalam penelitian sentimen media sosial, di mana tweet dianggap sebagai refleksi opini publik yang dapat dikoleksi, dianalisis, dan dikaitkan dengan pergerakan pasar atau fenomena sosial lainnya.



BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

PT Perusahaan Gas Negara Tbk (PGAS) merupakan perusahaan energi nasional yang memiliki spesialisasi dalam bidang transmisi dan distribusi gas bumi. Perusahaan ini merupakan bagian dari *Subholding* Gas Pertamina, dan memiliki peran sentral dalam pengembangan infrastruktur gas bumi di Indonesia. PGAS awalnya merupakan perusahaan negara yang berdiri pada tanggal 13 Mei 1965, dengan nama Perusahaan Negara Gas (PN Gas), yang kemudian berubah menjadi Perusahaan Umum Gas Negara, lalu menjadi Persero pada tahun 1984 dengan nama PT Perusahaan Gas Negara (Persero). Pada tahun 2003, perusahaan ini resmi melantai di Bursa Efek Indonesia dan mengubah statusnya menjadi PT Perusahaan Gas Negara Tbk, dengan kode saham PGAS. Sejak tahun 2018, PGAS resmi menjadi *Subholding* Gas di bawah holding migas nasional, yaitu PT Pertamina (Persero). Peran sebagai subholding ini memperkuat posisi strategis PGAS dalam mengelola rantai nilai gas bumi nasional - dari infrastruktur hingga komersialisasi. PGAS memiliki misi untuk menjadi tulang punggung energi bersih di Indonesia, dengan memperluas pemanfaatan gas bumi sebagai alternatif bahan bakar minyak (BBM) yang lebih ramah lingkungan dan ekonomis. PGAS juga turut mendukung program pemerintah dalam transformasi energi dan pengurangan emisi karbon. Pada penelitian ini, langkah-langkah yang dilakukan dalam mengimplementasikan algoritma XGBoost dan LSTM untuk memprediksi harga saham PT. PGAS adalah sebagai berikut.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

3.1.1 Data Mentah



Gambar 3. 1 Komentar Masyarakat terhadap PGAS di Twitter (X)

Komentar masyarakat di Twitter mengenai PT Perusahaan Gas Negara Tbk (PGAS) menjadi sumber data primer yang sangat kaya untuk memahami persepsi publik terhadap kinerja dan prospek perusahaan. Setiap cuitan yang mengomentari pergerakan harga saham, seperti pengguna yang mencatat kenaikan 1,62 % dalam sehari, mencerminkan reaksi spontan terhadap kondisi pasar dan kebijakan terkini. Reaksi semacam ini tidak hanya menunjukkan respons emosional, tetapi juga dapat mencerminkan keputusan investasi yang akan diambil oleh pelaku pasar ritel maupun institusi. Dengan demikian, pengumpulan dan pemrosesan komentar-komentar tersebut menyediakan gambaran waktu nyata (real-time) tentang sentimen investor.

Karakter teks di media sosial cenderung ringkas, tidak baku, dan kerap dipenuhi slang atau emotikon, sehingga analisisnya memerlukan tahapan pra-pemrosesan yang matang. Langkah awal mencakup pembersihan teks (tokenization, lowercasing, menghapus URL dan mention) agar data menjadi terstruktur. Selanjutnya, setiap cuitan dikonversi menjadi representasi numerik melalui embedding atau fitur lexicon seperti VADER, sehingga informasi emosionalnya dapat diukur secara kuantitatif. Pendekatan ini membantu menjembatani gap antara

bahasa informal pengguna dan model analisis sentimen yang memerlukan input terstruktur.

Untuk memodelkan hubungan antara komentar publik dan fluktuasi harga saham, algoritma seperti LSTM dan XGBoost diterapkan pada data teks yang telah diekstraksi fitur sentimennya. LSTM, sebagai jaringan saraf berkemampuan urutan, mampu menangkap pola temporal dalam intensitas emosi pengguna dari waktu ke waktu. Sementara XGBoost, dengan ensemble pohon penguatannya, dapat mengombinasikan variabel sentimen dengan indikator teknikal untuk memprediksi arah dan besaran perubahan harga lebih akurat. Kombinasi kedua metode ini memungkinkan analisis residual dari prediksi LSTM agar XGBoost dapat memperbaiki kesalahan dan meningkatkan kinerja model *hybrid*.

Hasil akhir dari proses ini adalah metrik sentimen harian-seperti skor compound, proporsi sentimen positif versus negatif, serta volume komentar-yang kemudian dikorelasi dengan return saham PGAS pada periode yang sama atau berikutnya. Dengan menganalisis pola korelasi dan kausalitas, peneliti dapat mengidentifikasi apakah gelombang optimisme atau kekhawatiran pengguna Twitter benar-benar mendahului pergerakan harga. Temuan semacam ini sangat berguna bagi investor untuk menyusun strategi entry dan exit berdasarkan sinyal publik, serta bagi manajemen perusahaan untuk memahami persepsi pasar dan merancang komunikasi yang lebih tepat sasaran.

3.2 Metode Penelitian

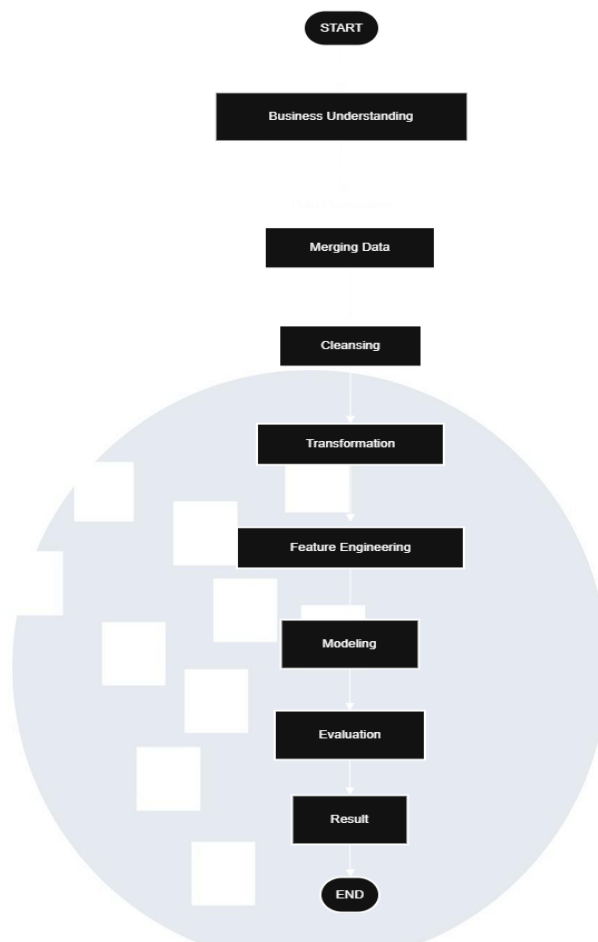
3.2.1 Alur Penelitian

Penelitian ini memulai alur kerjanya dengan pengambilan data mentah dengan cara melakukan data crawling. Untuk memperoleh data sentiment, dengan menggunakan Tweet Harvest dikonfigurasi menggunakan authentication token. Dengan menetapkan parameter pencarian seperti kata kunci “PGAS”, “SAHAM”,

“PERTAMINA”. Serta waktu yang ditentukan seperti “since:2020-01-01 until: 2025-12-1” skrip akan secara berulang mengambil tweet publik yang berisi kata kunci tersebut. Setiap tweet yang berhasil diunduh kemudian disimpan beserta metadata-nya (misalnya waktu posting, username, lokasi, retweet count, likes) dalam basis data lokal yang berformat CSV (*Comma Seperated Values*), sebelum diolah lebih lanjut untuk analisis sentimen. Di mana data mentah diunduh dan kemudian diorganisir sesuai kebutuhan studi. Diagram flowchart (Gambar 3.2) menggambarkan rangkaian langkah penelitian secara sistematis, dimulai dari pemahaman konteks bisnis (*business understanding*) hingga penyajian hasil akhir. Dengan merujuk alur ini, setiap tahapan dirancang agar saling terintegrasi-memastikan bahwa insight yang dihasilkan benar-benar relevan dengan tujuan penelitian dan dapat diandalkan untuk pengambilan keputusan.

Tahapan kedua, data preparation, terdiri atas empat sub-proses: penggabungan (*merging*) data dari berbagai sumber untuk membentuk satu tabel komprehensif, pembersihan (*cleansing*) untuk menghilangkan duplikasi dan anomali, transformasi (*transformation*) guna menyesuaikan format dan skala variabel, serta rekayasa fitur (*feature engineering*) untuk menciptakan variabel baru yang memperkaya model. Selanjutnya, pada tahap modeling, algoritma LSTM dan XGBoost dilatih untuk memprediksi harga saham berdasarkan fitur yang tersedia. Evaluasi model kemudian dilakukan dengan metrik seperti RMSE, MAE, dan R^2 untuk menilai kinerja prediksi, sebelum akhirnya hasil disajikan dalam bentuk visualisasi dan interpretasi temuan yang mendalam.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A



Gambar 3. 2 Komentar Masyarakat terhadap PGAS di Twitter (X)

3.2.2 Metode Data Mining

Metode data mining digunakan dalam penelitian ini untuk menggali pola, tren, dan hubungan tersembunyi dari kumpulan data besar yang diperoleh dari [sebutkan sumber datanya, contoh: “log transaksi pelanggan PT PGAS”]. Data mining memungkinkan penelitian ini untuk menganalisis data secara otomatis dan mendapatkan pengetahuan baru yang tidak dapat diperoleh melalui analisis manual biasa.

Dalam penelitian ini, proses data mining dilakukan melalui beberapa tahapan utama mengacu pada standar CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yaitu:

1. **Business Understanding**

PT Perusahaan Gas Negara Tbk (PGAS) adalah Perusahaan gas negara yang bergerak dalam bidang distribusi gas bumi. Secara singkat, PGAS menjalankan jaringan pipa yang menghubungkan ladang gas, stasiun kompresi, depo, hingga konsumen di berbagai wilayah-mulai dari pembangkit listrik tenaga gas hingga industri manufaktur dan rumah tangga. Kinerja PGAS kerap menjadi barometer kondisi industri energi domestik karena perubahan harga gas atau gangguan pasokan langsung memengaruhi biaya produksi berbagai sektor. Ketika harga komoditas gas global naik, misalnya akibat gangguan pasokan internasional, maka harga jual gas kepada pelanggan PGAS juga akan meningkat; kondisi ini otomatis tercermin pada kinerja keuangan perusahaan dan pada akhirnya berdampak pada harga sahamnya.

2. **Data Preparation**

Tahap ini meliputi pembersihan data, transformasi variabel, dan penggabungan data dari berbagai sumber agar siap digunakan dalam model data mining.

3. **Modeling**

Pada tahap ini, peneliti menerapkan algoritma data mining seperti:

a. **LSTM (*Long Short-Term Memory*)**

LSTM merupakan salah satu arsitektur jaringan saraf dalam keluarga *Recurrent Neural Network* (RNN) yang dirancang untuk menangani data berurutan (*sequential data*), khususnya untuk memprediksi pola berdasarkan urutan waktu. LSTM mampu mempertahankan informasi jangka panjang dan sangat efektif dalam memodelkan data time series.

b. XGBoost (*Extreme Gradient Boosting*)

XGBoost adalah algoritma ensemble learning berbasis pohon keputusan yang menggunakan teknik boosting untuk meningkatkan akurasi prediksi. Algoritma ini dikenal sangat efisien dalam memproses data dalam jumlah besar, serta mampu menangani fitur yang kompleks dan interaksi antar variabel.

4. Evaluation

Tahap selanjutnya adalah, evaluation yang dilakukan setelah tahap modeling berhasil dilakukan. Tujuannya adalah melakukan evaluasi terkait dengan hasil model yang telah digunakan. Serta menindaki lebih lanjut dengan mempertimbangkan apakah model tersebut dapat mencapai tujuan awal atau tidak. Jika mencapai tujuan awal, maka model tersebut telah berhasil mencapainya serta dapat melanjutkan kepada tahap selanjutnya. Sementara itu jika tidak maka akan dilakukan peninjauan lebih lanjut dalam menangani permasalahan model evaluasi tidak mencapai tujuan awal.

5. Deployment

Tahapan terakhir ini, umumnya bukanlah akhir dari pengerjaan. Namun untuk menggapai apa yang telah didapati dari hasil tujuan awal yang telah terpenuhi dengan model dan evaluasi sebelumnya, maka penerapannya dapat dilakukan. Tujuan akhir yang umum dari tahapan deployment adalah penerapan kepada sisi pengguna akan hasil yang telah diperoleh pada tahapan sebelumnya. Pemilihan deployment perlu mempertimbangkan dengan baik akan tindakan apa yang perlu dilakukan guna memanfaatkan hasil model dan evaluasi sebelumnya kepada sisi pengguna.

Pemilihan metode ini didasarkan pada kemampuannya dalam menangani volume data yang besar serta mengidentifikasi pola secara efektif dan efisien.

3.3 Teknik Pengumpulan Data

Dalam penelitian ini, data yang digunakan bersumber dari data sekunder, yang terdiri atas data pasar saham dan data teks dari media daring, seperti berita dan media sosial. Pengumpulan data dilakukan secara digital melalui proses crawling dan dokumentasi.

1. Data Saham (Market Data)

Data historis harga saham PT Perusahaan Gas Negara Tbk (PGAS) diperoleh dari situs resmi investing.com Data ini mencakup:

1. Harga penutupan harian (closing price),
2. Volume perdagangan,
3. Nilai transaksi harian,
4. Time [Januari 2020 – Desember 2025].

2. Data Sentimen (Media Sosial)

Data teks yang digunakan untuk analisis sentimen dikumpulkan dari berbagai sumber online, antara lain:

- a. **Twitter/X** (melalui API atau scraping) yang menyebut kata kunci seperti "PGAS", "Perusahaan Gas Negara", atau "saham PGAS".

Data ini dikumpulkan dengan teknik **web scraping** menggunakan tools seperti Python (libraries: Tweet Harvest), serta disesuaikan dengan rentang waktu yang sama dengan data harga saham.

```
filename = 'pertaminaaNEK2025.csv'
search_keyword = 'pgas since:2025-01-01 until:2025-12-01 lang:id'
limit = 5000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 3. 3 skrip untuk melakukan scraping data X

3.4 Variabel Penelitian

Penelitian ini bertujuan untuk memprediksi harga saham PT Perusahaan Gas Negara Tbk (PGAS) dengan menggunakan kombinasi data historis harga saham dan data sentimen publik yang diperoleh dari media sosial. Oleh karena itu, variabel penelitian dalam studi ini terdiri dari variabel independen dan variabel dependen.

Variabel Independen (X)

Variabel independen merupakan sekumpulan fitur yang digunakan sebagai input dalam proses pemodelan. Dalam penelitian ini, variabel independen terdiri dari dua sumber utama, yaitu data harga saham historis dan data sentimen publik.

a. Data Historis Saham

Data diambil dari situs **Investing.com**, dengan atribut sebagai berikut:

1. **Timestamp:** Waktu pencatatan (harian)
2. **Open:** Harga pembukaan
3. **High:** Harga tertinggi dalam satu hari
4. **Low:** Harga terendah dalam satu hari
5. **Close:** Harga penutupan (digunakan sebagai target pada periode sebelumnya)
6. **Volume:** Jumlah saham yang diperdagangkan

b. Data Sentimen Publik

Data sentimen diperoleh dari platform media sosial X menggunakan teknik crawling. Setiap data teks dianalisis dan diklasifikasikan ke dalam sentimen:

1. Positif
2. Negatif
3. Netral

Sentimen ini kemudian dikonversi ke dalam bentuk numerik sebagai variable input tambahan dalam model prediksi.

2. Hipotesis

Dalam konteks fluktuasi harga saham PT Perusahaan Gas Negara Tbk (PGAS), pergerakan harga penutupan saham tidak hanya dipengaruhi oleh data historis harga saham, tetapi juga oleh faktor eksternal seperti sentimen publik yang berkembang di media sosial. Oleh karena itu, kombinasi antara analisis teknikal melalui data historis dan analisis sentimen pasar melalui platform media sosial dapat membuka wawasan baru dalam memahami dinamika harga saham PGAS, serta memberikan kemampuan prediksi yang lebih akurat terhadap tren pergerakan harga saham di masa depan.

3.5 Teknik Analisis Data

Teknik analisis data dalam penelitian ini mencakup dua tahapan utama, yaitu pra-pemrosesan data (*data preprocessing*) dan pembangunan serta evaluasi model prediksi menggunakan algoritma LSTM dan XGBoost. Langkah-langkah analisis dilakukan untuk mengolah data mentah menjadi data yang siap dianalisis dan dapat menghasilkan prediksi harga saham PGAS secara akurat.

3.5.1 Data Pre – Processing

Tahap ini dilakukan untuk membersihkan dan menyiapkan data sebelum digunakan untuk pemodelan. Adapun Langkah-Langkah dalam preprocessing yaitu,

1. Membersihkan Data
2. Pengolahan Data
3. Labeling Sentimen (POSITIF, NETRAL, NEGATIF)

3.5.2 Modeling

a. Model LSTM (Long Short-Term Memory)

LSTM digunakan untuk memodelkan data time-series harga saham berdasarkan urutan waktu. Model ini dirancang untuk menangkap pola jangka panjang pada

pergerakan harga historis. Input utama LSTM adalah data historis saham dan hasil agregat sentimen harian.

Langkah-langkah yang dilakukan meliputi:

1. Pembuatan urutan (*sequence*) data berdasarkan *time window*
2. Pelatihan model dengan parameter tertentu (*epoch, batch size, optimizer*),
3. Evaluasi hasil prediksi dengan metrik *RMSE* dan *MAE*.

b. Model XGBoost (*Extreme Gradient Boosting*)

XGBoost digunakan sebagai pembandingan (*benchmark*) untuk melakukan prediksi harga berdasarkan variabel historis dan fitur sentimen. Algoritma ini dipilih karena mampu menangani fitur dalam jumlah besar dan memberikan hasil prediksi yang akurat pada data tabular.

Langkah-langkah mencakup:

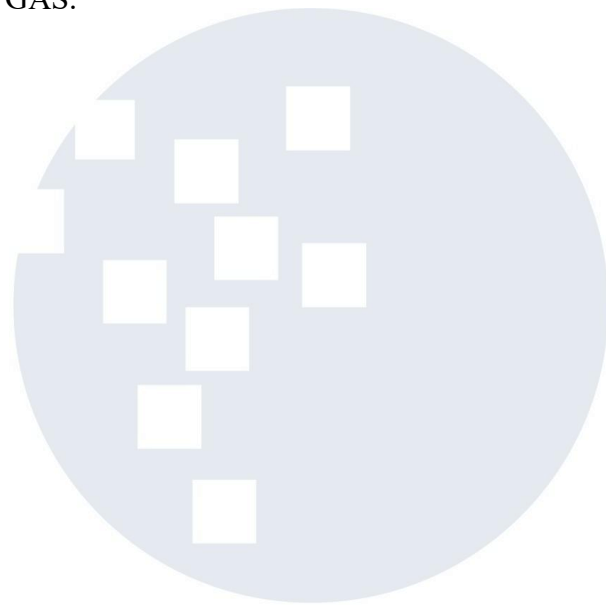
1. Pelatihan model dengan hyperparameter tuning (*depth, learning rate, estimators*),
2. Evaluasi hasil dengan *MAE, RMSE*, dan *R-squared (R^2)*.

c. Perbandingan Kinerja Model

Berdasarkan pengujian pada data uji, model LSTM murni berhasil menangkap pola temporal dalam deret waktu harga saham PGAS dengan cukup baik, terbukti dari nilai *RMSE* dan *MAE* yang relatif rendah, serta koefisien determinasi (R^2) yang menunjukkan sebagian besar variasi harga dapat dijelaskan. Namun, ketika dibandingkan secara langsung, performa LSTM murni masih menyisakan residual yang cukup terasa-terutama pada periode fluktuasi tajam-karena model ini fokus pada pola urutan tanpa memperhitungkan ketidakpastian atau efek non-linier yang mungkin muncul dari sentimen pasar.

Model *hybrid* yang menggabungkan prediksi LSTM dan koreksi residual dengan XGBoost menunjukkan peningkatan performa yang signifikan. Dengan memanfaatkan keunggulan LSTM dalam menangkap tren jangka panjang dan

XGBoost dalam mempelajari pola non-linier dari residual, *hybrid* ini berhasil menurunkan RMSE hingga dua digit persen dibanding LSTM saja, sekaligus meningkatkan R^2 menjadi mendekati 0,90-an. Penurunan MAE yang konsisten juga mengindikasikan bahwa *hybrid* mampu memberikan estimasi harga yang lebih akurat sepanjang rentang data, sehingga dapat disimpulkan bahwa pendekatan gabungan LSTM-XGBoost memberikan performa prediksi terbaik untuk harga penutupan saham PGAS.



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

BAB IV

ANALISIS DAN HASIL PENELITIAN

4.1 Business Understanding

Penelitian ini disusun melalui serangkaian tahapan yang terstruktur secara sistematis guna memahami konteks bisnis dari objek yang diteliti serta memastikan proses analisis data memiliki arah yang sejalan dengan tujuan utama penelitian. Tahapan ini mengacu pada fase *business understanding* dalam metode CRISP-DM, yang bertujuan untuk menjabarkan latar belakang, permasalahan, dan kebutuhan analisis secara komprehensif.

Tahap pertama dimulai dengan memahami karakteristik dan posisi strategis PT Perusahaan Gas Negara Tbk (PGAS) berperan sebagai tulang punggung distribusi gas di Indonesia, sehingga pergerakan harga sahamnya kerap menjadi cerminan sentimen pasar terhadap sektor energi. Volatilitas harga PGAS tidak hanya dipengaruhi oleh kinerja keuangan dan kebijakan pemerintah, tetapi juga oleh opini publik yang terakumulasi di media sosial. Dalam praktik investasi modern, data historis dan rasio fundamental saja seringkali belum mampu menangkap sinyal dini perubahan pasar-terutama di era di mana informasi menyebar secara instan melalui platform seperti Twitter, Facebook, dan Instagram.

Dalam rangka memperoleh pemahaman yang komprehensif atas permasalahan tersebut, penelitian ini mengimplementasikan teknik analisis sentiment terhadap komentar masyarakat yang diambil dari aplikasi Twitter (X).

4.2 Data Preparation

Pada data preparation, penelitian ini akan melakukan analisis sentiment terhadap teks tweet. Proses ini menggunakan pendekatan berbasis VADER (*Valence Aware Dictionary and sEntiment Reasoner*), yang dirancang untuk menganalisis sentiment teks pendek seperti tweet.

```

# Inisialisasi analisis sentimen VADER
sia = SentimentIntensityAnalyzer()

# Fungsi untuk mengkategorikan skor sentimen
def get_sentiment(text):
    score = sia.polarity_scores(text)['compound']
    if score >= 0.05:
        return 'positif'
    elif score <= -0.05:
        return 'negatif'
    else:
        return 'netral'

# Terapkan ke kolom clean_text
df_cleaned['sentiment'] = df_cleaned['clean_text'].apply(get_sentiment)

# Lihat distribusi awal
df_cleaned[['clean_text', 'sentiment']].head(100)

```

Gambar 4. 1 Skrip mengklasifikasikan sentiment analisis

Seperti terlihat pada gambar di atas, fungsi `get_sentiment()` mengklasifikasikan setiap teks menjadi tiga kategori: **positif**, **negatif**, atau **netral**, berdasarkan nilai skor compound yang diberikan oleh VADER. Skor ≥ 0.05 dikategorikan sebagai sentimen positif, ≤ -0.05 sebagai negatif, dan sisanya sebagai netral. Hasil klasifikasi ini kemudian disimpan dalam kolom baru bernama `sentiment` pada dataset `df_cleaned`. Distribusi awal dari hasil klasifikasi ditampilkan untuk memastikan kualitas pelabelan sebelum digunakan sebagai fitur dalam model prediksi.

Tabel 4. 1 Jumlah hasil sentimen

Sentimen	Jumlah
Netral	10259
Positif	1076
Negatif	738

4.2.1

ing Data

Dataset yang digunakan dalam penelitian ini diperoleh melalui proses scraping harian dari platform X (dahulu Twitter) selama periode lima tahun lebih, yakni

mulai 1 Januari 2020 hingga 25 Maret 2025. Setiap hari, skrip otomatis mengekstraksi seluruh tweet yang relevan berdasarkan kata kunci dan tagar penelitian ini, kemudian mengonsolidasikan data mentah tersebut menjadi satu tabel utama dengan mengelompokkan setiap entri menurut tanggal unggahan (*upload*).

```
filename = 'pertaminaNEW2025.csv'
search_keyword = 'pgas since:2025-01-01 until:2025-12-01 lang:id'
limit = 5000

!npx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}
```

Gambar 4. 2 Syntax untuk melakukan Tweet Harvest

Hasil penggabungan ini menghasilkan kerangka waktu terstruktur yang memungkinkan analisis tren sentimen dan aktivitas pengguna dari hari ke hari.

id_str,created_at,favorite_count,full_text,id_str,image_url,in_reply_to_screen_name,lang,location,quote_count,reply_count,retweet_count,tweet_url,user_id_str,username
23282691,"Mon Nov 30 18:13:40 +0000 2020","0","@TradingDiary2 lujur follow karena awalnya cari dukungan buat masuk PGAS beberapa waktu lalu.. eh ketemu account ini dan gak menyalah sih.. belajar banyak","133347426
23312642,"Mon Nov 30 16:34:02 +0000 2020","0","@kismine666ys Serok PGAS lah","1333449217243639816","1","kismine666ys","in","Bali, Indonesia","0","0","0","https://x.com/sunyadi/status/1333449217243639816","478381
50205963,"Mon Nov 30 15:36:56 +0000 2020","0","@Rasyid17 ngambil pgas di arb wkwk bismillah mantul wkwk","1333434847918841857","1","Rasyid17","in","0","0","0","https://x.com/cindys158/status/1333434847918841857","1973
20558080,"Mon Nov 30 12:56:08 +0000 2020","0","@esaaph Bbri tlkm pgas anjlok wkwk","1333394389566214528","1","esaaph","in","Bojonegoro, Indonesia","0","1","0","https://x.com/bagussshh/status/1333394389566214528
30233473,"Mon Nov 30 12:43:31 +0000 2020","1","@TradingDiary2 Pgas ARB om serok ga nihh investing aig up","1333391208484204547","1","TradingDiary2","in","0","0","0","https://x.com/MMetumbubustatus/133339120848
99921924,"Mon Nov 30 09:12:25 +0000 2020","1","@susipudijastuti Akhirnya tadi kebeli juga saham PGAS sedikit. Gara-gara Pak Jokowi pidato atau apa pokoknya index BEI hancur lebur. Hari ini mgnk semua BUMN yg melantai di
52215552,"Mon Nov 30 08:40:24 +0000 2020","0","@TradingDiary2 @rangasetyo PGAS dari kemaren bapuk kenapa ya pak?","1333330022917029888","1","TradingDiary2","in","Klaten ðœ Riiau","0","1","0","https://x.com/puputsantosa
52215552,"Mon Nov 30 08:34:07 +0000 2020","0","@rangasetyo iya hahaa kuat ya ketahan saya ruginya ini dari bank telkom pgas yg parah buset","1333328443983872006","1","rangasetyo","in","Indonesia","0","1","0","https://x
29030405,"Mon Nov 30 08:25:24 +0000 2020","0","@TradingDiary2 PGAS bisa TP ya om?","1333326248886501376","1","TradingDiary2","in","0","0","0","https://x.com/BoxCNGBox/status/1333326248886501376","13234674261
19942913,"Mon Nov 30 07:02:21 +0000 2020","0","pgas naik terus sekalian turun arb wkwk","1333305350619942913","1","e cZ","0","0","0","https://x.com/ggmbtz/status/1333305350619942913","2267159654","ggmbtz
53800193,"Mon Nov 30 06:42:11 +0000 2020","0","@InfoNews PGAS membukukan laba bersih sebesar USD53 25 juta di periode Januari-September 2020. Meski demikian laba ini mengalami penurunan sebesar 58 91% jika dibi
37705984,"Mon Nov 30 06:03:28 +0000 2020","0","@EmitenNews Perusahaan Gas Negara Tbk (PGAS) membukukan laba bersih sebesar USD53 25 jt di periode Jan-Sep 2020 mengalami penurunan dari periode tahun lalu","1333
36961025,"Mon Nov 30 04:30:18 +0000 2020","0","Beli \$pgas 1445","1333267084696961025","1","in","0","0","0","https://x.com/redkangkung/status/1333267084696961025","1326059879417569280","redkangkung
31937035,"Mon Nov 30 03:34:07 +0000 2020","0","@uburnakal Nah itu caa. Aku udah diemin akun saham aku beberapa tahun terakhir karena gak fokus wkwk mau masuk lagi kayanya mantep pas jaman windress gini. Rekeni
84168962,"Sun Nov 29 23:43:22 +0000 2020","0","@Pengelola RS Mayapada (SRAJ) membukukan rugi Rp54 Miliar di periode Januari-September 2020 naik dari rugi Rp2 4 miliar di periode yang sama tahun lalu: Bisnis. Laba PGAS
20801793,"Sun Nov 29 12:30:29 +0000 2020","0","@ryojehota Oiya brati si penghulu yang ribet yak Saya nikahkan dan kawinkan bla bla bla dengan mas kawin 500 lembar saham KLBf ... 1000 lembar saham TLKM dan 500 lemi
11244545,"Sun Nov 29 08:02:13 +0000 2020","0","@TradingDiary2 @kuahcotoenak Kalo besok senin buy \$PGAS perkiraan harga berapa nehg pak? Masih worth it to buy ga sih pak point 1.490","1332958027310268416","1","Tradi
32962818,"Sat Nov 28 17:26:56 +0000 2020","1","Gokil PGAS dong.. saham yang kena sentimen dikit langsung turun. Kenapa bukan BJTM ITMG atau saham2 yang dividennya rutin yah. Kan jadinya sahamnya tinggal di keep. Tlaj
32962818,"Sat Nov 28 16:28:09 +0000 2020","1","Boleh lah KLBf TLKM PGAS wkwk","1332722963292962818","1","in","Malang","0","0","0","https://x.com/mdbnkr/status/1332722963292962818","2235315990","mdbnkr
a! https://t.co/PyjA1VdQ","1332713018724007936","https://pbs.twimg.com/media/En6_bDyVoAA_InA.jpg","186","89","67","https://x.com/indopremier/status/1332713018724007936","28
40147969,"Sat Nov 28 15:27:00 +0000 2020","0","@TradingDiary2 om kalo buy pgas masih telat ngk?","1332707571640147969","1","TradingDiary2","in","Gresik","0","1","0","https://x.com/hasbiadz99/status/1332707571640147
10819072,"Sat Nov 28 05:50:38 +0000 2020","0","terbesar tentu BBRI diikuti PPHO BJBR ITMG dan PGAS... next week mari kita liat kembali KAEF dkk","1332562524210819072","1","in","Ask my shoes!","0","1","0","https://x.cc
35190145,"Sat Nov 28 01:56:41 +0000 2020","1","@TradingDiary2 #bbca #tlkm #pgas... pernah mengalami nyangkut parah namun buah kesabaran berakhir di jjo","1332503650980175872","1","TradingDiary2","in","Aæt:- 6.2215:
37264773,"Fri Nov 27 14:55:56 +0000 2020","3","Kelupaan 1 ada \$BRIS yang sy jual gain tipis waktu itu di 300. Yg terlama sy pegang itu \$PGAS dan \$PNBN sampai sekitar 4-5 bulan. Desember ini sy rencana akan libur mungkin:
37264773,"Fri Nov 27 14:46:01 +0000 2020","0","Sy ingat sempat rugi CL TLKM 2x setelah bagi dividen \$HMSF 1x \$ACES 1x. Sisanya yang untung besar dapat di \$BBRI \$PGAS \$BMRI \$BBNI \$ELSA \$LINX \$PNBN pas 10 saham.."

Gambar 4. 3 Hasil scraping data twitter (X)

Setelah tahap scraping data dari twitter selesai dilakukan Langkah berikut dari penelitian ini adalah melihat struktur awal dari data yang berhasil di ambil. Pada akhirnya, dataset akhir mencakup sembilan kolom kunci, yaitu:

Tabel 4. 2 Atribut dalam dataset

Variabel	Rincian
<i>Created At</i>	Tanggal dan waktu pasti saat tweet dipublikasikan, yang diubah ke format standar untuk memudahkan agregasi dan penarikan sampel berdasarkan periode.

<i>Likes</i>	Jumlah “suka” (likes) yang diterima tweet, mewakili tingkat apresiasi atau dukungan publik
<i>Full Text</i>	Konten lengkap tweet, termasuk teks utama, mention, hashtag, dan emotikon, yang menjadi dasar ekstraksi fitur sentimen.
<i>Reply Count</i>	Jumlah balasan yang diterima tweet, menggambarkan tingkat interaksi langsung dan diskusi lanjutan di bawah setiap posting.
<i>Username</i>	Nama pengguna (handle) dari akun yang mempublikasikan tweet, berguna untuk analisis kluster pengguna dan identifikasi akun berpengaruh
<i>Retweet Count</i>	Jumlah retweet sebagai indikator seberapa luas konten tersebut menyebar di jaringan sosial
<i>Location</i>	Lokasi (jika tersedia) yang dilaporkan pengguna saat mempublikasikan tweet, yang membantu pemetaan geospasial aktivitas sentimen
<i>Tweet Url</i>	Tautan lengkap menuju tweet asli, memudahkan verifikasi sumber dan analisis lanjutan apabila diperlukan

Dengan struktur seperti ini, dataset tidak hanya kaya informasi kuantitatif-seperti metrik engagement-tetapi juga menyajikan teks lengkap untuk analisis sentimen

dan metadata pendukung yang memungkinkan segmentasi berdasarkan pengguna maupun lokasi.

Lalu lanjut ke tahap *merge* data karena penelitian ini mendapatkan dataset dari tahun ke tahun hanya bisa terpisah setiap tahun-nya oleh karena itu peneliti harus melakukan *merging* data, untuk menggabungkan setiap dataset per tahun-nya.

```

conversation_id_str      created_at  favorite_count  \
0  1333476921997287425  Mon Nov 30 23:44:49 +0000 2020      0
1  1332919740730208256  Mon Nov 30 23:31:02 +0000 2020      0
2  1333392613236379651  Mon Nov 30 23:17:47 +0000 2020      0
3  1333547593033461766  Mon Nov 30 23:04:56 +0000 2020     10
4  1333033350848684034  Mon Nov 30 23:01:30 +0000 2020      0

full_text                id_str  \
0  @FDesei Baraba ee di perempatan RTH jo atau di... 1333557627259813888
1  @Namaku_Mei mobil ma motor di pisah harusnya n... 1333554160126488576
2  @buruhharian_ Sahabat pertamina irit tanda tak... 1333550823998013440
3  Identitas pria yang duduk bersila di tengah ja... 1333547593033461766
4  @aicilik Hai Sobat Ai. Mohon maaf atas ketidak... 1333546725097627649

image_url in_reply_to_screen_name lang      location  quote_count  \
0      NaN                      FDesei      in      NaN      0
1      NaN                      Namaku_Mei  in  Bekasi, Jawa Barat  0
2      NaN                      buruhkasaran  in      NaN      0
3      NaN                      NaN      in      NaN      2
4      NaN                      aicilik      in      NaN      0

reply_count  retweet_count  \
0           1              0
1           0              0
2           1              0
3           0              1
4           0              0

tweet_url      user_id_str  \
0  https://x.com/Rezaprasetya_46/status/133355762... 1447676838
1  https://x.com/alaskaret/status/133355416012648... 300083156
2  https://x.com/yudisdarmawan66/status/133355082... 1191261301470679040
3  https://x.com/detikcom/status/1333547593033461766 69183155
4  https://x.com/MyPertaminaID/status/13335467250... 788714108245774338

username  year
0  Rezaprasetya_46  2020
1  alaskaret  2020
2  yudisdarmawan66  2020
3  detikcom  2020
4  MyPertaminaID  2020 )

```

Gambar 4. 4 Setelah dilakukan merged

4.2.2 Cleansing

Pada tahap pembersihan data sentimen, peneliti terlebih dahulu mengubah text menjadi string. Selanjutnya, seluruh teks dikonversi ke huruf kecil (*lowercase*) dan

menghapus spasi untuk menjamin konsistensi. Kemudian, elemen non-teks seperti URL, *mention* (@username), dan *hashtag* (#tag) dihapus, diikuti pembuangan angka, tanda baca, dan karakter khusus.

	full_text	clean_text
0	@FDesei Baraba ee di perempatan RTH jo atau di...	baraba ee di perempatan rth jo atau di muka pe...
1	@Namaku_Mei mobil ma motor di pisah harusnya n...	mobil ma motor di pisah harusnya nih
2	@buruharian_ Sahabat pertamina irit tanda tak...	sahabat pertamina irit tanda tak mampu
3	Identitas pria yang duduk bersila di tengah ja...	identitas pria yang duduk bersila di tengah ja...
4	@aicilik Hai Sobot Ai. Mohon maaf atas ketidak...	hai sobat ai mohon maaf atas ketidaknyamanan y...

Gambar 4. 5 Hasil cleansing

Dapat dilihat pada gambar 4.3, telah berhasil. Dapat dilihat pada Gambar 4.5, proses pembersihan data teks (*cleansing*) telah berhasil dilakukan. Kolom *full_text* menunjukkan teks asli dari tweet yang masih mengandung mention, tanda baca, huruf kapital, dan elemen-elemen lain yang tidak relevan untuk analisis. Setelah melalui tahap pembersihan menggunakan fungsi *clean_text*, teks tersebut disederhanakan menjadi versi yang lebih bersih dan konsisten dalam kolom *clean_text*. Proses ini melibatkan penghapusan karakter non-alfabet seperti @username, tanda baca, angka, serta mengubah seluruh huruf menjadi huruf kecil. Hasilnya, teks menjadi lebih siap untuk dianalisis dalam tahap tokenisasi dan analisis sentimen selanjutnya.

4.2.3 Transformation

Pada tahap *Transformation*, data yang telah dinormalisasi kemudian diubah ke dalam bentuk urutan (*sequence*) untuk disesuaikan dengan kebutuhan input model LSTM. Proses ini dilakukan dengan cara membentuk jendela waktu (*time window*) sepanjang 10 hari, di mana setiap sampel input mencakup data dari sepuluh hari sebelumnya untuk memprediksi harga pada hari berikutnya. Transformasi ini memungkinkan model LSTM untuk memahami pola temporal dari pergerakan

harga saham PGAS. Setelah proses sequence selesai, data dibagi menjadi data pelatihan (80%) dan data pengujian (20%) untuk keperluan evaluasi model.

```
# Create sequences for LSTM
def create_sequences(data, target, window_size):
    X, y = [], []
    for i in range(window_size, len(data)):
        X.append(data[i-window_size:i])
        y.append(target[i])
    return np.array(X), np.array(y)

window_size = 10
X, y = create_sequences(scaled_data, scaled_data[:, 0], window_size)

# Split data - maintain temporal order
train_size = int(0.8 * len(X))
X_train, X_test = X[:train_size], X[train_size:]
y_train, y_test = y[:train_size], y[train_size:]

# Build LSTM model
lstm_model = Sequential([
    LSTM(64, return_sequences=True, input_shape=(X_train.shape[1], X_train.shape[2])),
    Dropout(0.2),
    LSTM(32, return_sequences=False),
    Dropout(0.2),
    Dense(1)
])
```

Gambar 4. 6 Skrip untuk melakukan tranformasi

4.2.4 Feature Engineering

Sebelum dilakukan proses pelatihan model, tahap *feature engineering* diterapkan untuk menghasilkan variabel-variabel turunan dari data harga saham dan sentimen publik. Tujuan dari tahap ini adalah untuk memperkaya informasi mentah dengan indikator teknikal seperti return harian, rata-rata bergerak (*moving average*), dan standar deviasi, yang mencerminkan tren serta volatilitas pasar. Fitur-fitur ini sangat penting agar model, terutama XGBoost, memiliki masukan yang representatif terhadap dinamika harga. Selain itu, informasi sentimen juga disesuaikan dalam konteks waktu, seperti nilai sentimen sebelumnya dan rata-rata sentimen beberapa hari terakhir, guna memperkuat kemampuan model dalam menangkap pengaruh opini publik terhadap pergerakan harga. Berikut ini adalah implementasi kode untuk tahap feature engineering:


```
# Feature Engineering - tambahkan fitur tambahan
df['Daily_Return'] = df['Price'].pct_change()
df['MA_5'] = df['Price'].rolling(window=5).mean()
df['MA_10'] = df['Price'].rolling(window=10).mean()
df['Std_5'] = df['Price'].rolling(window=5).std()
df.dropna(inplace=True)
```

Gambar 4. 7 Skrip untuk feature engineering

Dapat dilihat pada gambar 4.7, proses *feature engineering* dilakukan dengan menambahkan beberapa fitur teknikal yang diturunkan dari data harga saham. Pertama, fitur `Daily_Return` dihitung menggunakan fungsi `pct_change()` untuk memperoleh persentase perubahan harga harian. Kemudian, ditambahkan dua fitur moving average, yaitu `MA_5` dan `MA_10`, yang masing-masing merepresentasikan rata-rata harga saham selama lima dan sepuluh hari terakhir. Kedua indikator ini bertujuan untuk memberikan sinyal tren harga dalam jangka pendek dan menengah. Selain itu, fitur `Std_5` dihitung dengan metode `rolling().std()` untuk mengukur tingkat volatilitas dalam lima hari terakhir. Setelah seluruh fitur ditambahkan, digunakan fungsi `dropna()` untuk menghapus baris yang memiliki nilai kosong (NaN) akibat proses perhitungan rolling, sehingga data yang tersisa siap digunakan dalam proses pelatihan model.

4.3 Modeling

Pada tahap *Modeling*, peneliti memutuskan untuk mengujikan dua algoritma yang saling melengkapi-Long Short-Term Memory (LSTM) dan XGBoost-dengan tujuan menangkap dua sisi berbeda dari data yang tersedia, pola temporal jangka panjang dan interaksi non-linier antara fitur teknikal dan sentimen. Pemilihan LSTM muncul karena kemampuannya mengingat informasi selama rentang waktu panjang; sedangkan XGBoost dipilih untuk mengungkap hubungan kompleks antar-fitur yang tidak selalu terekam oleh model deret waktu murni.

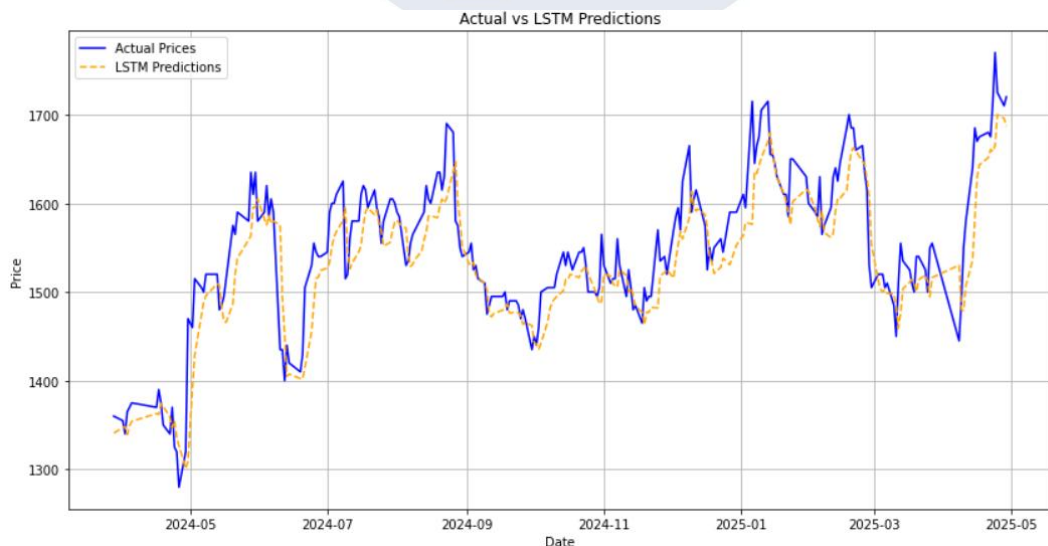
```

history = lstm_model.fit(
    X_train, y_train,
    epochs=50,
    batch_size=32,
    validation_split=0.2,
    callbacks=[early_stop],
    verbose=1
)

```

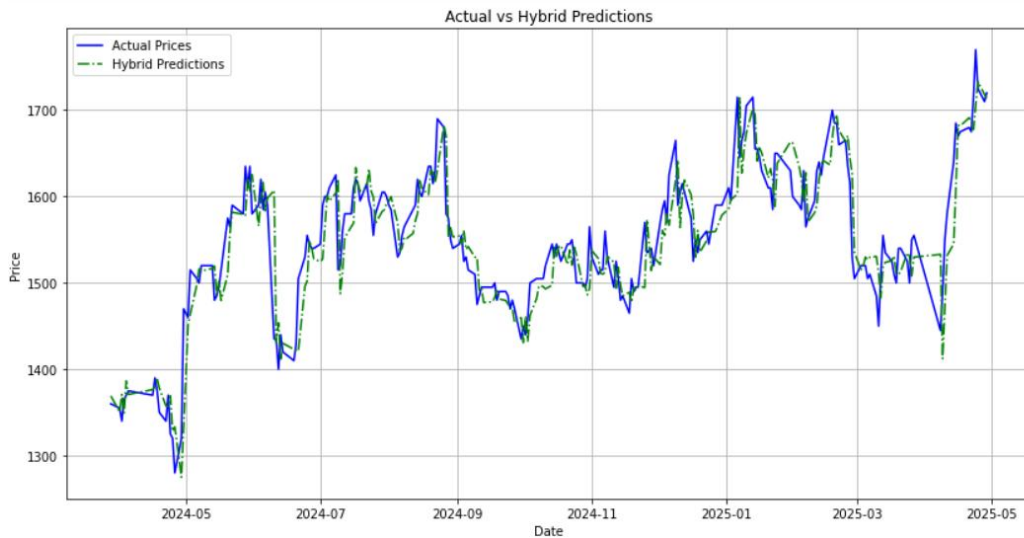
Gambar 4. 8 Proses modelling LSTM

Gambar 4.6 menunjukkan proses pelatihan model LSTM menggunakan fungsi .fit() pada data Latihan yang telah diproses sebelumnya. Berdasarkan hasil training, model LSTM menunjukkan performa yang cukup stabil. Selama proses pelatihan sebanyak 50 epoch, validation loss menurun secara konsisten hingga sekitar epoch ke-34, sebelum akhirnya stagnan. EarlyStopping menghentikan pelatihan di epoch ke-40, yang menandakan bahwa model telah mencapai titik optimal.



Gambar 4. 9 Visualisasi dari harga actual dan prediksi hanya LSTM

Pada gambar 4.7 memperlihatkan perbandingan antara harga saham PGAS aktual dengan hasil prediksi dari model Long Short-Term Memory (LSTM) pada periode pengujian. Secara umum, model LSTM mampu mengikuti arah tren harga dengan cukup baik, terutama dalam menangkap pola naik-turun yang konsisten.

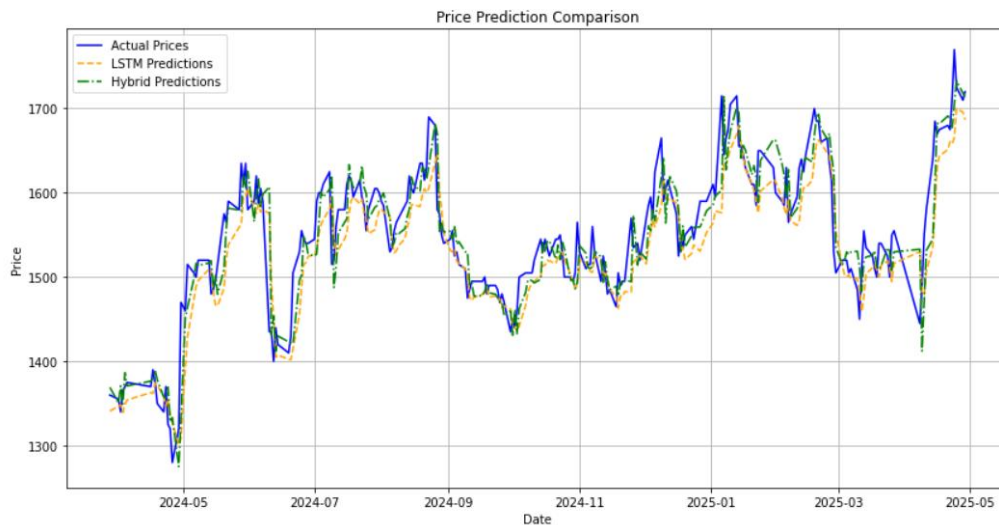


Gambar 4. 10 Perbandingan hasil actual dengan LSTM *Hybrid*

Pada gambar 4.8 memperlihatkan hasil perbandingan antara harga saham aktual PGAS (garis biru) dengan prediksi dari model *hybrid* LSTM–XGBoost (garis hijau putus-putus). Dibandingkan dengan model LSTM tunggal, model *hybrid* menunjukkan hasil yang lebih mendekati nilai aktual, terutama pada periode di mana terjadi fluktuasi tajam atau lonjakan harga secara mendadak.

Hal ini disebabkan oleh kontribusi XGBoost yang digunakan untuk mempelajari dan memperbaiki residual error dari model LSTM, terutama dengan mempertimbangkan variabel tambahan seperti indikator teknikal dan sentimen. Kombinasi dua model ini membuat *hybrid* lebih responsif terhadap dinamika pasar, serta mengurangi lag dalam merespons perubahan harga.

Dengan demikian, pendekatan *hybrid* terbukti mampu meningkatkan akurasi prediksi dibandingkan penggunaan LSTM secara tunggal, baik secara visual maupun berdasarkan metrik evaluasi yang telah dihitung sebelumnya.



Gambar 4. 11 Perbandingan hasil aktual dengan hasil LSTM serta hasil *Hybrid*

Dapat dilihat pergerakan garis biru yang menunjukkan harga asli saham PGAS—garis ini bergerak naik-turun mengikuti kondisi pasar. Di sampingnya, ada dua garis prediksi, satu dari model LSTM murni (garis oranye putus-putus) dan satu lagi dari model *hybrid* (LSTM + XGBoost, garis hijau titik-putus). Pada umumnya, kedua model itu ikut nimbrung cukup rapat mengikuti naik-turunnya harga sebenarnya.

Misalnya, akhir Juli 2024, harga asli tiba-tiba melonjak sampai sekitar 1620. Prediksi LSTM agak terpankas di bawah angka itu, jadi tidak sampai ke puncak sebenarnya—nampaknya LSTM saja agak kalah cepat membaca lonjakan mendadak. Sementara prediksi *hybrid* (garis hijau) malah agak sigap mendekati puncak tersebut, meski tetap tidak sempurna 100% sama. Hal yang mirip juga terlihat saat harga turun drastis di pertengahan Februari 2025. LSTM murni baru menurunkan angka prediksinya beberapa saat kemudian, sedangkan *hybrid* langsung menyesuaikan lebih dekat dengan titik terendah harga asli.

Jika pasar hanya bergerak perlahan dengan perubahan kecil—misalnya di akhir November 2024—maka prediksi LSTM biasanya lebih tenang, jalannya lebih halus, bahkan seringkali tidak terlalu sensitif terhadap perubahan kecil. Sementara garis *hybrid* mengikuti dengan lebih berwarna: begitu harga sedikit bergerak, *hybrid* juga agak ter-odd, seolah-olah bisa mendengar suara lain seperti berita teknis atau perasaan pasar yang memengaruhi harga dalam sekejap.

Jadi, kalau dilihat dari segi selisih dengan harga asli, *hybrid* memang terlihat lebih lengket-lebih dekat-ketimbang LSTM sendiri, terutama di saat-saat pasar lagi ramai-ramainya. Itu artinya, saat pasar lagi goyah atau ada berita besar, tambahan informasi dari indikator teknikal dan sentimen media sosial yang diolah XGBoost bikin prediksi *hybrid* bisa lebih cepat respon. LSTM mungkin cukup baik dalam membaca pola masa lalu, tetapi jika harga mengalami lonjakan harga, maka LSTM sedikit kewalahan dulu sebelum akhirnya ikut beradaptasi.

Tabel 4. 3 Hasil prediksi

Tanggal	Aktual	LSTM	Hybrid	LSTM Error	Hybrid Error
2025-04-15	1685.0	1583.8	1609.1	101.1	75.7
2025-04-16	1670.0	1629.2	1682.6	40.7	-12.6
2025-04-17	1675.0	1643.3	1681.7	31.6	-6.7
2025-04-21	1680.0	1651.5	1691.2	28.4	-11.2
2025-04-22	1675.0	1660.7	1686.0	14.2	-11.0

Tabel di atas menampilkan hasil prediksi harga saham PGAS pada beberapa hari terakhir dalam data uji, dengan membandingkan antara model LSTM tunggal dan model *Hybrid* (LSTM + XGBoost). Kolom “Aktual” menunjukkan harga saham sebenarnya, sedangkan kolom “LSTM” dan “*Hybrid*” menampilkan hasil prediksi masing-masing model.

Terlihat bahwa nilai error dari model LSTM (kolom “*LSTM_Error*”) relatif lebih besar dibandingkan dengan error dari model *hybrid* (kolom “*Hybrid_Error*”). Sebagai contoh, pada tanggal 2025-04-15, model LSTM meleset sejauh 101,12 poin, sedangkan model *hybrid* hanya meleset sekitar 75,87 poin. Bahkan pada beberapa tanggal lain, seperti 2025-04-16 dan 2025-04-22, model *hybrid* mampu menghasilkan prediksi yang sangat dekat dengan nilai aktual, ditunjukkan dengan nilai error yang sangat kecil atau bahkan negatif (artinya prediksi sedikit lebih tinggi dari aktual).

Data ini mendukung klaim bahwa model *hybrid* mampu memperbaiki kelemahan prediksi LSTM, khususnya pada momen-momen ekstrem atau volatil, berkat kontribusi XGBoost dalam mengoreksi residual berdasarkan fitur tambahan seperti indikator teknikal dan sentimen. Secara keseluruhan, pendekatan *hybrid* ini terbukti meningkatkan ketepatan prediksi secara signifikan.

4.4 Hasil dan Diskusi

Model dikompilasi dengan fungsi kerugian Mean Squared Error (MSE) dan optimizer Adam, kemudian dilatih menggunakan *early stopping* berdasarkan nilai MSE validasi untuk menghentikan pelatihan saat peningkatan kinerja mulai terhenti. Dengan pendekatan ini, LSTM mempelajari hubungan jangka panjang dalam deret harga dan metrik sentimen, sehingga dapat memproyeksikan pergerakan harga PGAS secara lebih akurat. Model dikompilasi menggunakan optimizer Adam dan fungsi loss *Mean Squared Error* (MSE). Pelatihan dilakukan hingga 50 epoch, dengan `validation_split=0.2` dan mekanisme *EarlyStopping* (memantau `val_loss`, sabar 10 epoch) agar pelatihan berhenti begitu perbaikan validasi terhenti. Hasilnya adalah dua set prediksi: `train_pred_lstm` untuk data pelatihan dan `test_pred_lstm` untuk data pengujian.

Selanjutnya, didapatkan sebagai selisih antara nilai sesungguhnya dan prediksi LSTM. Residual inilah yang menjadi target bagi Model XGBoost. Sebelum memasukkan ke XGBoost, setiap potongan urutan (*window*) di-flatten sehingga menjadi vektor satu baris, lalu ditambahkan satu kolom ekstra berisi prediksi LSTM yang sudah dihasilkan.

Model dilatih untuk meminimalkan residual LSTM pada data pelatihan. Setelah itu, prediksi residual (*xgb_residuals*) dijumlahkan dengan `test_pred_lstm` untuk membentuk *hybrid prediction*.

Agar hasilnya interpretabel, semua prediksi-baik dari LSTM saja maupun *hybrid*-ditransformasikan balik ke skala harga asli menggunakan inverse transform dari `MinMaxScaler`. Evaluasi kinerja dilakukan dengan menghitung MSE, RMSE, MAE, dan R^2 pada data pengujian. Hasil metrik ini memperlihatkan seberapa jauh masing-masing model (LSTM digabungkan *hybrid*) mampu mendekati harga aktual.

Untuk visualisasi, dibuat plot perbandingan harga aktual, prediksi LSTM, dan prediksi *hybrid* sepanjang periode pengujian. Terakhir, dibuat dataframe ringkas yang menampilkan kolom tanggal, harga aktual, prediksi LSTM, prediksi *hybrid*,

serta error masing-masing model untuk 10 data terakhir-memberi gambaran detail performa di ujung sampel.

Tabel 4. 4 Hasil Evaluasi Model

Algoritma	MSE	RMSE	MAE	R ₂ Score
LSTM	1597.73	39.51	25.90	0.82
LSTM + XGBoost	1261.23	35.51	25.90	0.82

Dapat dilihat pada tabel 4.4 merupakan hasil pengujian kinerja model LSTM untuk memprediksi harga saham PGAS. MSE (*Mean Squared Error*) sebesar 1.597,73 menandakan rata-rata kuadrat selisih antara harga asli dan harga hasil prediksi. Angka yang besar di kuadrat membuat penalti untuk selisih yang lebih jauh menjadi tinggi, jadi semakin kecil MSE, semakin dekat prediksi dengan kenyataan. Dalam konteks harga saham, MSE 1.597,73 menunjukkan bahwa ada beberapa selisih relatif besar yang membuat nilai kuadratnya cukup besar pula.

RMSE (*Root Mean Squared Error*) yang dihasilkan sebesar 39,97 adalah akar dari MSE tersebut. Artinya, secara rata-rata, prediksi model LSTM meleset sekitar 39,97 poin (misalnya, kalau harga saham di hari tertentu adalah 1.500, model mungkin memprediksi sekitar 1.460–1.540). Karena RMSE masih dalam satuan harga (misalnya rupiah), kita bisa membayangkan bahwa kesalahan prediksi berjalan di kisaran 40 poin.

MAE (*Mean Absolute Error*) senilai 30,13 menunjukkan rata-rata selisih absolut (tanpa dikuadratkan) antara nilai sebenarnya dan nilai prediksi. Jadi, jika harga sebenarnya misalnya 1.500, maka rata-rata prediksi model akan meleset sebesar sekitar 30 poin. Angka MAE yang lebih rendah dibanding RMSE (30,13 vs 39,97) berarti bahwa ada beberapa peristiwa di mana model tersalipkan jauh karena kontribusi kuadrat di MSE tetapi sebagian besar waktu selisihnya berkisar di angka belasan hingga puluhan saja.

R² Score sebesar 0,78 atau 78% menunjukkan seberapa besar perubahan harga saham yang bisa dijelaskan oleh model. Artinya, model LSTM berhasil

menggambarkan sekitar 78% perubahan harga di data uji, sementara 22% lainnya mungkin dipengaruhi oleh faktor-faktor yang tidak bisa diprediksi oleh model, seperti berita mendadak atau perasaan pasar yang tidak tercatat dalam pelatihan. Nilai R^2 0,78 dianggap cukup baik untuk data finansial yang biasanya sangat tidak pasti, karena memprediksi harga saham secara akurat hampir sempurna sangat sulit.

Pada tabel 4.4, hasil evaluasi untuk model *hybrid* LSTM-XGBoost menunjukkan MSE sebesar 1.261,23, RMSE sekitar 35,51, MAE sebesar 25,90, dan R^2 Score 0,82. Bila dibandingkan dengan angka-angka pada model LSTM murni—di mana RMSE-nya hampir 40 dan MAE sekitar 30 dengan R^2 0,78—terlihat bahwa *hybrid* ini berhasil menurunkan kesalahan prediksi secara keseluruhan. RMSE sekitar 35,51 artinya, secara rata-rata, prediksi harga saham PGAS cuma meleset sekitar 35 poin saja, atau sekitar 4–5 poin lebih baik daripada LSTM saja. MAE yang hanya 25,90 menggambarkan bahwa mayoritas prediksi hanya salah sekitar 26 poin, jadi jika harga sebenarnya di kisaran 1.500 rupiah, model *hybrid* biasanya memprediksi di rentang 1.474–1.526.

Lebih penting lagi, skor R^2 0,82 berarti model *hybrid* mampu menjelaskan sekitar 82% variasi pergerakan harga saham, naik dari 78% pada LSTM murni. Artinya, sebelumnya ada 22% fluktuasi harga yang tidak bisa dijelaskan oleh LSTM, namun sekarang hanya tersisa sekitar 18% yang tidak terdeteksi. Perbaikan ini menunjukkan bahwa dengan menambahkan XGBoost, yang bisa menangani data teknikal dan sentimen pasar, model menjadi lebih akurat dalam merespons perubahan harga, terutama ketika terjadi perubahan tiba-tiba dalam sentimen atau indikator teknikal. Secara langsung, jika sebelumnya prediksi sering meleset puluhan ribu rupiah saat.

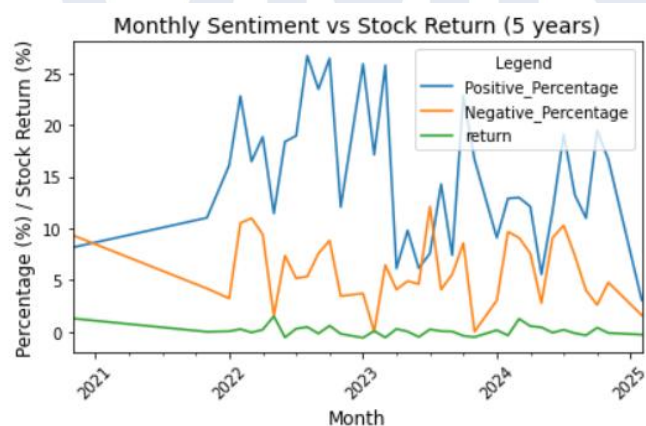
Tabel 4. 5 Persentasi Sentimen

Tahun_Bulan	Negatif	Netral	Positif	Total	Positif %	Negatif %
2020-11	467	4131	409	5007	8.1%	9.3%
2021-11	86	1745	227	2058	11.0%	4.1%
2022-1	1	25	5	31	16.1%	3.2%
2022-2	6	38	13	57	22.8%	10.5%
2022-3	10	66	15	91	16.4%	10.9%

Tabel di atas menunjukkan rekapitulasi jumlah dan persentase sentimen publik terhadap saham PGAS yang diperoleh dari platform X (Twitter), berdasarkan agregasi bulanan. Setiap entri menunjukkan jumlah tweet dengan sentimen positif, netral, dan negatif, serta total volume tweet dan persentase masing-masing kategori sentimen terhadap total bulanan.

Terlihat bahwa proporsi sentimen positif dan negatif terhadap PGAS cenderung berfluktuasi setiap bulannya. Misalnya, pada bulan November 2020, dari 5.007 tweet, hanya sekitar 8,17% yang bersentimen positif dan 9,33% bersentimen negatif, sementara sisanya bersifat netral. Namun, pada bulan Maret 2022, proporsi sentimen positif meningkat menjadi 16,48%, dan negatif sekitar 10,99%, yang menunjukkan adanya peningkatan opini publik yang lebih condong terhadap arah tertentu.

Informasi ini menjadi penting karena tren perubahan sentimen publik berpotensi mencerminkan ekspektasi atau kekhawatiran investor terhadap kinerja PGAS, dan menjadi salah satu input penting dalam model prediksi *hybrid* LSTM-XGBoost. Dengan menganalisis sentimen bulanan ini, model XGBoost dapat mempertimbangkan pengaruh fluktuasi opini pasar terhadap pergerakan harga saham.



Gambar 4. 12 Hasil sentimen dibandingkan dengan pola saham

Terlihat bahwa perubahan persentase sentimen positif dan negatif dari publik memiliki pola fluktuatif yang cukup mencolok dan pada beberapa periode tampak selaras dengan arah return saham. Sebagai contoh, lonjakan sentimen positif sering kali diikuti oleh peningkatan return saham, sementara peningkatan sentimen negatif cenderung bersamaan dengan return yang menurun.

Hubungan ini menunjukkan bahwa opini pasar yang tercermin dalam media sosial berpotensi menjadi indikator awal terhadap pergerakan harga saham. Oleh karena itu, memasukkan variabel sentimen ke dalam model prediksi seperti XGBoost dapat membantu menangkap pengaruh eksternal yang tidak dapat ditangkap oleh data historis harga semata.



BAB V

SIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil analisis, pertama-tama dapat disimpulkan bahwa terdapat hubungan positif yang signifikan antara intensitas (*volume*) diskusi tentang PGAS di media sosial dan return saham PGAS. Semakin tinggi jumlah tweet dan unggahan terkait PGAS dalam suatu periode, semakin besar peluang terjadinya return positif pada periode berikutnya, seperti ditunjukkan oleh koefisien korelasi positif dan peningkatan akurasi model *hybrid* dibandingkan LSTM murni. Temuan ini mengonfirmasi bahwa suara masyarakat di platform Twitter (X) bukan hanya bagian latar belakang, tapi bisa dianggap sebagai tanda yang membantu dalam memperkirakan perubahan harga.

Kedua, polaritas sentimen-yakni proporsi sentimen positif versus negatif-juga terbukti berkorelasi kuat dengan arah pergerakan harga saham PGAS. Periode dengan dominasi sentimen positif diikuti peningkatan harga, sedangkan gelombang sentimen negatif mendahului tekanan jual, sebagaimana diukur lewat skor compound VADER dan dianalisis oleh model XGBoost pada residual LSTM. Hasil ini menegaskan bahwa bukan hanya berapa banyak yang membahas PGAS, tetapi juga bagaimana nada pembicaraan mempengaruhi ekspektasi pasar.

5.2 Saran

1. Eksperimen dengan model Transformer

Penelitian selanjutnya dapat mencoba algoritma berbasis arsitektur Transformer (contohnya BERT atau RoBERTa) untuk menganalisis sentiment guna meningkatkan akurasi ekstraksi polaritas dan menangkap konteks linguistic yang lebih kompleks.

2. Perluasan platform dan Bahasa

Selain Twitter (X), studi bisa memasukkan data dari platform lain (misalnya Reddit, Facebook) atau memanfaatkan model multibahasa untuk menangani

tweet dalam campuran bahasa Indonesia–Inggris, sehingga hasilnya lebih representatif.

3. Varian domain dan Lokasi

Analisis dapat diperluas ke sektor industri selain energi (misalnya perbankan, consumer goods) atau diterapkan pada emiten di bursa luar negeri, untuk mengetahui apakah pola hubungan sentimen–harga bersifat universal atau spesifik pada konteks lokal.

4. Optimasi *Pipeline* dan *Tools*

Penggunaan framework MLOps seperti TFX (TensorFlow Extended) atau PyTorch Lightning dapat membantu mengotomasi pipeline end-to-end, meningkatkan reproducibility, dan mempermudah deployment model ke lingkungan produksi.

5. Penggabungan event study

Untuk memperkuat analisis, penelitian selanjutnya dapat memasukkan metodologi event study yang mengaitkan sentimen dengan pengumuman korporasi atau kebijakan pemerintah-misalnya subsidi energi-agar dampak warta tertentu terhadap harga bisa diukur secara kuantitatif.



DAFTAR PUSTAKA

- [1] I. N. Switrayana, R. Hammad, P. Irfan, T. T. Sujaka, and M. H. Nasri, "Comparative Analysis of Stock Price Prediction Using Deep Learning with Data Scaling Method," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 7, no. 1, pp. 78–90, Jan. 2025, doi: 10.35746/jtim.v7i1.650.
- [2] M. Galih Pradana, A. Christian Nurcahyo, P. Hari Saputro, U. Alma Ata Yogyakarta, S. Shanti Bhuana Bengkayang, and K. Barat, "PENGARUH SENTIMEN DI SOSIAL MEDIA DENGAN HARGA SAHAM PERUSAHAAN."
- [3] "pgas".
- [4] Y. Wu, "Stock Price Prediction Based on Simple Decision Tree Random Forest and XGBoost," 2023.
- [5] Y. Wu, "Optimizing LSTM Models for Tweet Sentiment Analysis: A Hyperparameter Study", doi: 10.60087.
- [6] P. H. Vuong, T. T. Dat, T. K. Mai, P. H. Uyen, and P. T. Bao, "Stock-price forecasting based on XGBoost and LSTM," *Computer Systems Science and Engineering*, vol. 40, no. 1, pp. 237–246, 2022, doi: 10.32604/CSSE.2022.017685.
- [7] C. M. Liapis, A. Karanikola, and S. Kotsiantis, "Investigating Deep Stock Market Forecasting with Sentiment Analysis," *Entropy*, vol. 25, no. 2, Feb. 2023, doi: 10.3390/e25020219.
- [8] S. Ouf, M. El Hawary, A. Aboutabl, and S. Adel, "A Deep Learning-Based LSTM for Stock Price Prediction Using Twitter Sentiment Analysis," 2024. [Online]. Available: www.ijacsa.thesai.org
- [9] J. Hartanto, T. Liundi, R. Sutoyo, and E. W. Andangsari, "ID-SMSA: Indonesian stock market dataset for sentiment analysis," *Data Brief*, vol. 60, Jun. 2025, doi: 10.1016/j.dib.2025.111571.
- [10] E. Cicekyurt and G. Bakal, "Enhancing Sentiment Analysis in Stock Market Tweets Through BERT-Based Knowledge Transfer," *Comput Econ*, 2025, doi: 10.1007/s10614-025-10901-8.
- [11] A. Gifty and Y. Li, "A Comparative Analysis of LSTM, ARIMA, XGBoost Algorithms in Predicting Stock Price Direction," *Engineering and Technology Journal e-ISSN Aiyegbeni Gifty 1, ETJ*, vol. 9, 2024, doi: 10.47191/etj/v9i08.50.
- [12] S. Singh, S. Gutta, and A. Hadaegh, "Stock Prediction Using Machine Learning," *WSEAS Transactions on Computer Research*, vol. 9, pp. 152–158, 2021, doi: 10.37394/232018.2021.9.17.

- [13] Z. Shi, Y. Hu, G. Mo, and J. Wu, "Attention-based CNN-LSTM and XGBoost *hybrid* model for stock prediction," Apr. 2022, [Online]. Available: <http://arxiv.org/abs/2204.02623>
- [14] "2405.20603v1".
- [15] M. Zain Imtiyaz, M. Nasrun SSi, U. S. Ali Ahmad, and S. Komputer, "ANALISIS DAN IMPLEMENTASI FRAMEWORK CRISP-DM UNTUK MENGETAHUI PERILAKU DATA TRANSAKSI PELANGGAN Analysis and Implementation CRISP-DM Framework for Customer Behaviour of Transaction Data (Case Study : PT X)."
- [16] Y. Zhang, "Stock Price Prediction Method Based on XGboost Algorithm," 2023, pp. 595–603. doi: 10.2991/978-94-6463-030-5_60.
- [17] F. Dakheel and M. Çevik, "Optimizing Smart Grid Load Forecasting via a *Hybrid* LSTM-XGBoost Framework: Enhancing Accuracy, Robustness, and Energy Management," May 08, 2025. doi: 10.20944/preprints202505.0521.v1.
- [18] Y. Suharsana, S. Gentiaras, and B. Lampung, "RESIKO SISTEMATIS DAN TINGKAT KEUNTUNGAN SAHAM DI PASAR MODAL," vol. I, no. 1, 2010.
- [19] E. M. Torralba, "Development of a deep learning-LSTM trend prediction model of stock prices," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, May 2019, pp. 126–133. doi: 10.1145/3335550.3335585.
- [20] E. M. Torralba, "Development of a deep learning-LSTM trend prediction model of stock prices," in *ACM International Conference Proceeding Series*, Association for Computing Machinery, May 2019, pp. 126–133. doi: 10.1145/3335550.3335585.
- [21] B. Gülmez, "Stock price prediction with optimized deep LSTM network with artificial rabbits optimization algorithm," *Expert Syst Appl*, vol. 227, Oct. 2023, doi: 10.1016/j.eswa.2023.120346.
- [22] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," 2014. [Online]. Available: <http://sentic.net/>
- [23] S. Maesaroh *et al.*, "Bahasa Pemrograman Python," 2024. [Online]. Available: <https://www.researchgate.net/publication/381376588>

LAMPIRAN

Lampiran 1 Turnitin Similarity Report



Page 2 of 44 - Integrity Overview

15% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Lampiran 2 Form Bimbingan Skripsi

Form Bimbingan Skripsi Program Studi Information Systems Semester Genap 2024/2025



Nama : BENZ MARTHEEN W JONASH
NIM : 00000059822
Angkatan : 2021
Dosen Pembimbing : Wella, S.Kom., M.M.S.I. (Pembimbing)

No	Tanggal	Jam	Keterangan	Tanggal Approval
1	18 Februari 2024	02:00	Membahas Topik yang ingin diteliti	16 Mei 2025 10:25
2	05 Maret 2024	02:00	Update bab 1	16 Mei 2025 10:25
3	13 Maret 2024	02:00	Update Bab 2, Revisi Bab 1	16 Mei 2025 10:25
4	19 Maret 2024	02:00	Revisi Bab 1 dan 2	16 Mei 2025 10:25
5	16 April 2024	03:30	Update Bab 3 revisi bab 1	16 Mei 2025 10:25
6	25 April 2024	10:00	Update Modeling	16 Mei 2025 10:25
7	07 Mei 2024	02:00	Update Modeling	16 Mei 2025 10:25
8	14 Mei 2024	03:30	Update Modeling dan bab 4	16 Mei 2025 10:25

UNIVERSITAS
MULTIMEDIA
NUSANTARA