

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 = 'pertaminaaNEW2025.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/MMetumbubastatus/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","1323467426
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
96961025,"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 negeh 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","Aet: -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 \$LINL \$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 klaster 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

Clea

nsing

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	@buruhharian_ 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 Sobat 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

Tran

sformation

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.

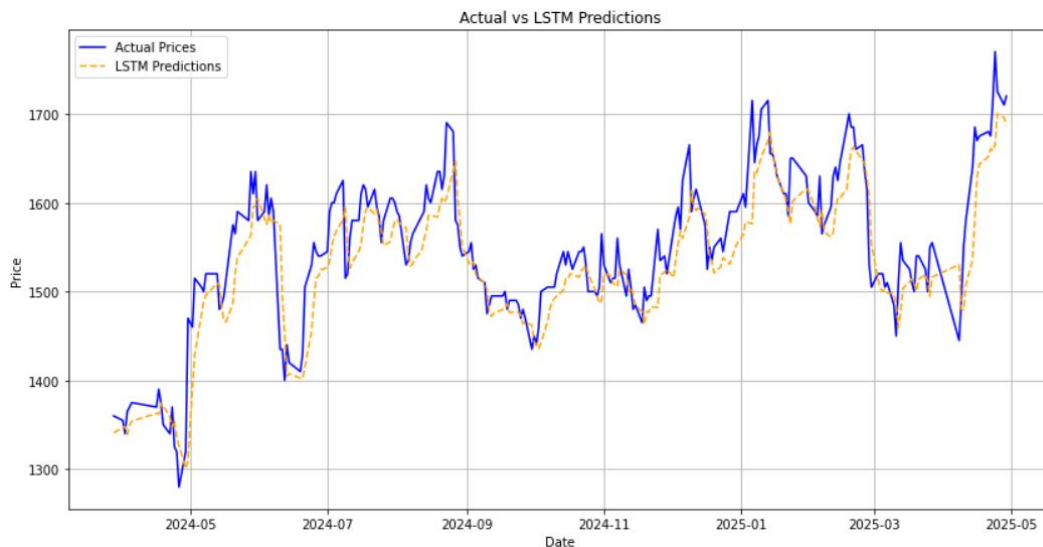
4.3 Modeling

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 teknis 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
)
```

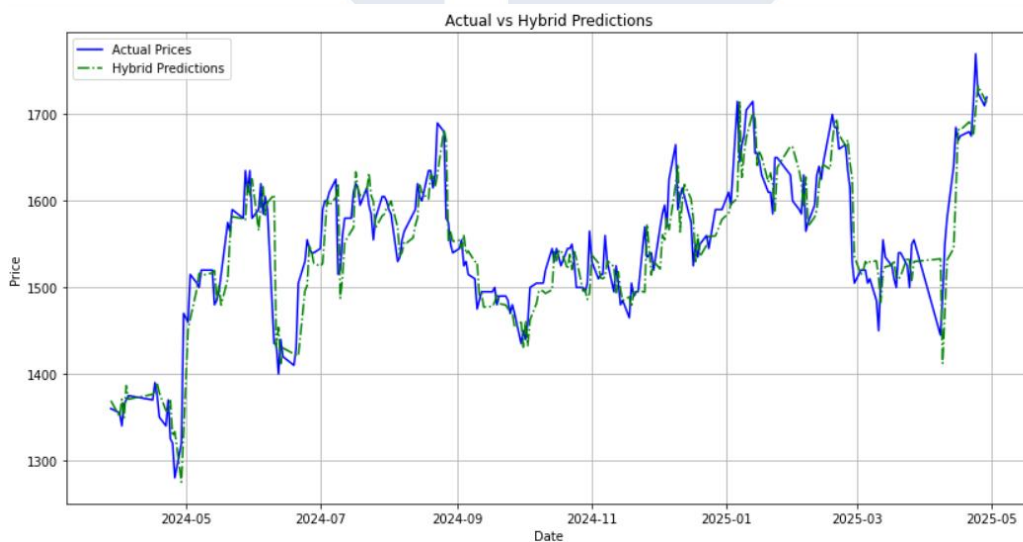
Gambar 4. 8 Proses modelling LSTM

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 actual 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.



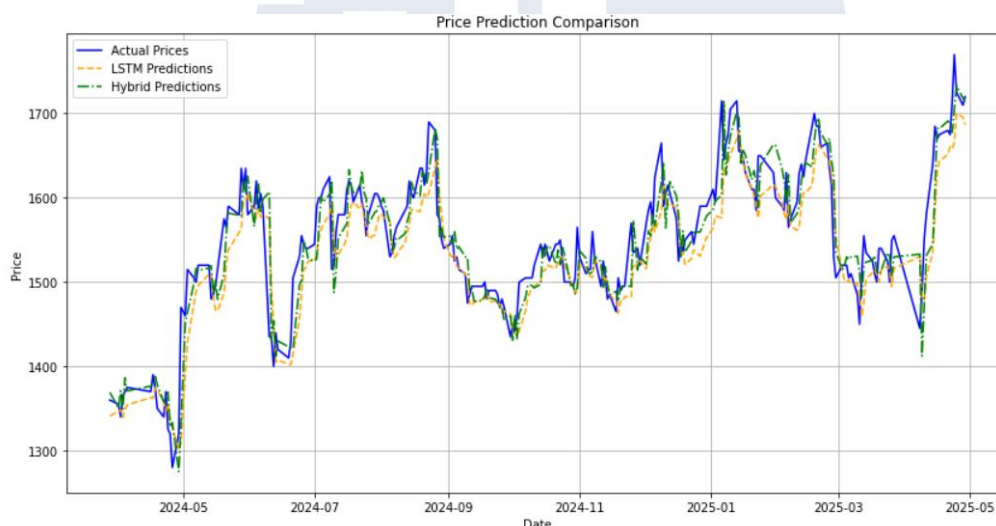
Gambar 4. 10 Perbandingan hasil actual dengan LSTM *Hybrid*

Pada gambar 4.8 memperlihatkan hasil perbandingan antara harga saham actual 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 terpankaskan 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 R² 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 R² Score 0,82. Bila dibandingkan dengan angka-angka pada model LSTM murni—di mana RMSE-nya hampir 40 dan MAE sekitar 30 dengan R² 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 R² 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

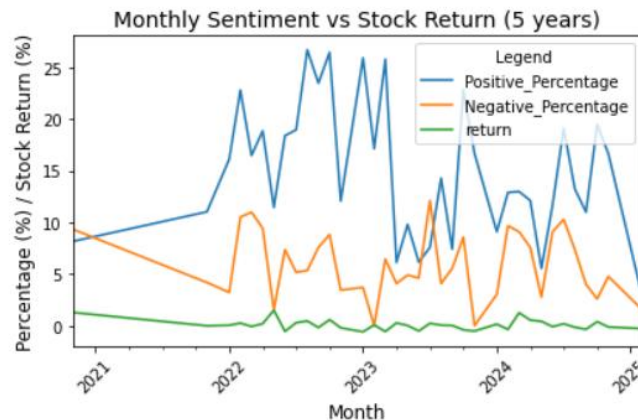
Tahun_Bulan	Negatif	Netral	Positif	Total	Positif %	Negatif %
2020-11	467	4131	409	5007	8.1%	9.3%
2021-11	86	1745	227	2058	11.0%	4.1%
2022-1	1	25	5	31	16.1%	3.2%
2022-2	6	38	13	57	22.8%	10.5%
2022-3	10	66	15	91	16.4%	10.9%

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.