

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 menggabungkan 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

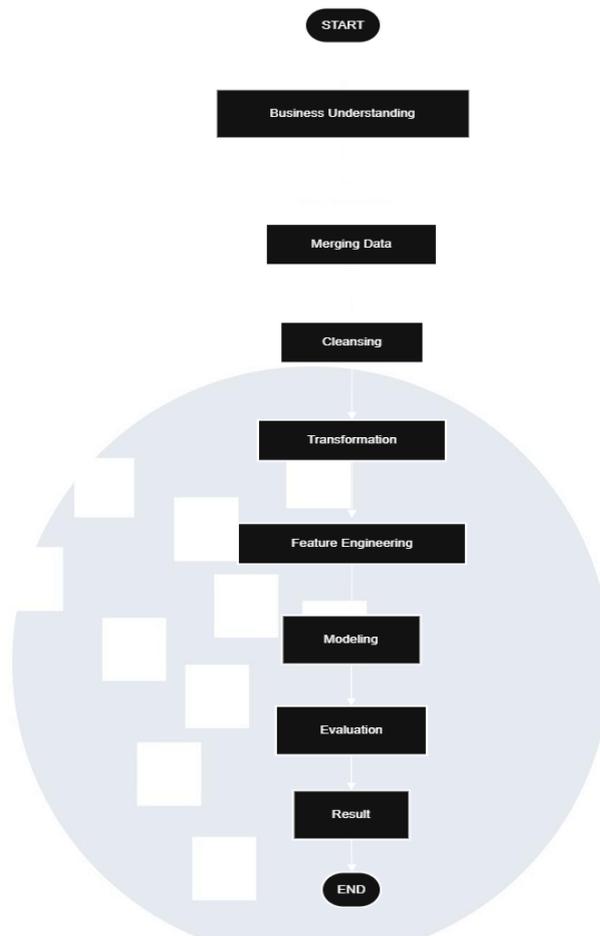
3.2.1 Alur Penelitian

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 = '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 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:

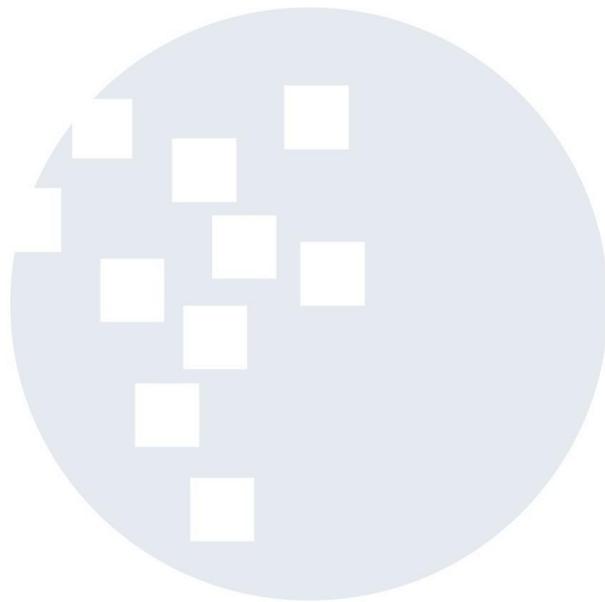
1. Pelatihan model dengan hyperparameter tuning (*depth, learning rate, estimators*),
2. Evaluasi hasil dengan *MAE, RMSE, dan R-squared (R²)*.

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