

# **STOCK PRICE SENTIMENT PADA SAHAM BBRI**

---

## **TUGAS BESAR DATA MINING**

---



Disusun oleh:

Muhammad Azka Nuril Islami (714220001)

Gaizka Wisnu Prawira (714220011)

Muhammad Fathir (714220021)

Salwa Mutfia Indah Putri (714220026)

Dosen Pengampu:

Nisa Hanum Harani, S.T., M.T., CDSP.,SFPC

NIK. 117.89.223

**PROGRAM STUDI DIV TEKNIK INFORMATIKA  
UNIVERSITAS LOGISTIK & BISNIS INTERNASIONAL  
BANDUNG  
2025**

## HALAMAN PERNYATAAN ORISINALITAS

Laporan tugas besar ini adalah hasil karya kami sendiri dan semua sumber, baik yang dikutip maupun dirujuk telah kami nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan dengan etika dan kaidah penulisan karya ilmiah yang berlaku, maka kami bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Yang menyatakan,

Nama : Muhammad Azka Nuril Islami

NIM : 714220001

Tanda Tangan: .....

Tanggal: Kamis, 10 Juli 2025

Mengetahui,

Ketua : ..... (.....tanda tangan. ....)

Dosen Pengampu Mata Kuliah : ..... (.....tanda tangan. ....)

# KATA PENGANTAR

Puji syukur kami panjatkan kepada Tuhan Yang Maha Esa atas limpahan rahmat dan karunia-Nya sehingga kami dapat menyelesaikan Laporan Tugas Besar Data Mining ini yang berjudul "Stock Price Sentiment pada Saham BBRI".

Laporan ini disusun untuk memenuhi tugas akhir mata kuliah Data Mining pada Program Studi D4 Teknik Informatika, Universitas Logistik dan Bisnis Internasional.

Kami mengucapkan terima kasih kepada:

- Dosen pengampu mata kuliah Data Mining atas bimbingan dan ilmunya selama perkuliahan berlangsung.
- Orang tua dan keluarga yang selalu memberikan dukungan moril dan semangat.
- Rekan satu kelompok atas kerja sama dan komitmen dalam menyelesaikan tugas ini bersama.

Kami menyadari bahwa laporan ini masih memiliki kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat kami harapkan demi perbaikan di masa mendatang.

Bandung, 10 Juli 2025

Penyusun,

Muhammad Azka Nuril

Gaizka Wisnu Prawira

Muhammad Fathir

Salwa Mutfia Indah Putri

# **HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Logistik dan Bisnis Internasional, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Azka Nuril Islami

NIM : 714220001

Selaku ketua kelompok, menyatakan menyetujui untuk memberikan kepada Universitas Logistik dan Bisnis Internasional, hak bebas royalti noneksklusif (non-exclusive royalty free right) atas karya ilmiah kami yang berjudul, "STOCK PRICE SENTIMENT PADA SAHAM BBRI" beserta perangkat yang ada (jika diperlukan). Dengan hak ini, ULBI berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir kami selama tetap mencantumkan nama kami sebagai penulis/pemilik hak cipta.

Demikian pernyataan ini saya buat dengan sebenar-benarnya.

Dibuat di : Bandung

Pada tanggal : 10 Juli 2025

Yang menyatakan,

Muhammad Azka Nuril Islami  
Ketua Kelompok

## **ABSTRAK**

## **ABSTRACT**

# DAFTAR ISI

<b>HALAMAN PERNYATAAN ORISINALITAS.....</b>	<b>2</b>
<b>KATA PENGANTAR.....</b>	<b>3</b>
<b>HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS .....</b>	<b>4</b>
<b>ABSTRAK .....</b>	<b>5</b>
<b>ABSTRACT .....</b>	<b>6</b>
<b>DAFTAR ISI.....</b>	<b>7</b>
<b>DAFTAR TABEL .....</b>	<b>9</b>
<b>DAFTAR GAMBAR.....</b>	<b>10</b>
<b>DAFTAR RUMUS.....</b>	<b>11</b>
<b>DAFTAR NOTASI .....</b>	<b>12</b>
<b>BAB I PENDAHULUAN .....</b>	<b>13</b>
1.1    Latar Belakang.....	13
1.2    Rumusan Masalah .....	14
1.3    Tujuan penelitian .....	14
1.3.1.    Tujuan Umum .....	14
1.3.2.    Tujuan Khusus.....	14
1.4    Manfaat Penelitian.....	14
1.4.1    Manfaat Teoretis.....	14
1.4.2    Manfaat Praktis .....	14
1.5    Ruang Lingkup .....	15
<b>BAB II TINJAUAN PUSTAKA.....</b>	<b>16</b>
2.1    Kajian Teori dan Konsep Penting .....	16
2.1.1    Data Mining dan Machine Learning .....	16
2.1.2    Teknik yang Digunakan .....	16
2.2    Studi Terkait (Penelitian Sejenis) .....	17
2.3    Visualisasi (Diagram Alir Konsep).....	18
2.4    State of The Art.....	19
<b>BAB III METODOLOGI PENELITIAN .....</b>	<b>20</b>
3.1    Tahapan penelitian.....	20
3.2    Deskripsi Dataset.....	21
3.3    Algoritma / Data Mining Tools.....	21
3.4    Evaluasi Kinerja .....	22

DAFTAR PUSTAKA.....	23
LAMPIRAN.....	25



## **DAFTAR TABEL**

## **DAFTAR GAMBAR**

## **DAFTAR RUMUS**

## DAFTAR NOTASI

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Pasar saham adalah komponen utama dari sistem keuangan suatu negara yang mencerminkan kesejahteraan ekonomi [1]. Fluktuasi harga saham biasanya dipengaruhi oleh berbagai faktor, seperti kondisi perusahaan, kebijakan ekonomi, dan faktor psikologis investor. Dalam praktiknya, keputusan pembelian dan penjualan saham tidak semata-mata didasarkan pada pertimbangan rasional, tetapi juga pada persepsi dan emosi para pelaku pasar yang dapat berubah sewaktu-waktu [2]. Keadaan ini berarti bahwa informasi dan opini publik dapat menjadi penyebab utama perubahan harga saham, terutama dalam jangka pendek [3].

Perkembangan teknologi informasi, khususnya media sosial seperti Twitter, telah mengubah cara individu dalam memberi dan menerima informasi. Twitter saat ini menjadi forum yang sangat aktif untuk memberikan komentar, berita, dan reaksi terhadap berbagai isu ekonomi dan perusahaan [4]. Kecepatan dan volume informasi yang disebarkan di Twitter berpotensi memberikan gambaran langsung tentang opini publik mengenai suatu perusahaan atau area bisnis [5]. Oleh karena itu, analisis sentimen tweet berpotensi menjadi sumber sekunder yang berharga untuk data dalam prediksi pasar [4][6].

Meskipun harga saham historis telah menjadi dasar model prediksi yang dibangun selama bertahun-tahun, metode ini tidak lengkap. Model prediksi tradisional tidak dapat menangkap kekuatan psikologis dan pola pikir pasar yang berubah dengan cepat. Selama krisis atau pengumuman berita penting dalam situasi yang tidak stabil, data media sosial dapat memberikan peringatan lebih awal daripada indikator teknikal. Oleh karena itu, integrasi data sentimen dengan data historis sangat potensial untuk meningkatkan akurasi model prediksi saham.

Beberapa penelitian sebelumnya telah mengindikasikan bahwa sentimen media sosial berhubungan dengan pergerakan harga saham [7]. Hasil-hasil ini menunjukkan bahwa sentimen publik yang positif menyebabkan pergerakan harga saham naik, sementara sentimen negatif dapat menjadi pendahulu pergerakan harga turun. Namun, sebagian besar penelitian ini masih kurang dalam hal pengujian korelasi yang tepat tanpa mencoba meniru model prediksi yang lebih umum. Pendekatan yang lebih formal diperlukan untuk menggabungkan kedua jenis data di bawah kerangka model prediksi yang terjamin.

Dengan potensi besar dari sentimen media sosial, penelitian ini bertujuan untuk membuat dan memvalidasi sebuah model untuk prediksi harga saham yang menggabungkan data masa lalu dan sentimen Twitter secara real-time [5][8]. Tujuannya adalah untuk mengidentifikasi apakah dengan menggabungkan kedua sumber tersebut dapat memberikan model yang lebih baik dan responsif terhadap perubahan pasar. Penelitian ini diharapkan dapat memberikan kontribusi yang berarti bagi perkembangan teknologi keuangan modern dan menginformasikan pengambilan keputusan investasi yang lebih berwawasan.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang sebelumnya, dapat dirumuskan beberapa pertanyaan utama sebagaimana berikut:

- A. Apakah analisis sentimen real time dari Twitter memiliki korelasi signifikan dengan pergerakan harga saham?
- B. Bagaimana kinerja model prediksi harga saham yang hanya menggunakan data historis dibandingkan dengan model yang menggabungkan data historis dan sentimen Twitter?
- C. Sejauh mana penambahan fitur sentimen Twitter dapat meningkatkan akurasi dan ketahanan model terhadap volatilitas pasar?

## 1.3 Tujuan penelitian

### 1.3.1. Tujuan Umum

Menganalisis pengaruh integrasi data sentimen Twitter terhadap akurasi model prediksi harga saham dibandingkan dengan model prediksi yang hanya menggunakan data historis.

### 1.3.2. Tujuan Khusus

- A. Melakukan analisis sentimen pada data Twitter menggunakan model RoBERTa.
- B. Menggabungkan data hasil analisis sentiment dengan data historis berdasarkan tanggal data.
- C. Membangun dan mengevaluasi model klasifikasi sentimen dengan algoritma XGBoost, SVR, Random Forest, MLP, dan Logistic Regression.

## 1.4 Manfaat Penelitian

### 1.4.1 Manfaat Teoretis

- A. Memberikan kontribusi pada pengembangan ilmu pengetahuan di bidang keuangan dan data science, khususnya terkait integrasi analisis sentimen media sosial dalam prediksi harga saham.
- B. Menjadi referensi akademik mengenai penggunaan data sentimen Twitter untuk mendukung model prediksi harga saham yang lebih akurat.
- C. Memperkaya literatur terkait pemanfaatan big data dan text mining dalam analisis pasar modal.

### 1.4.2 Manfaat Praktis

- A. Membantu investor dalam mengambil keputusan investasi yang lebih tepat dengan mempertimbangkan informasi sentimen publik dari media sosial.
- B. Memberikan wawasan bagi perusahaan sekuritas dan analis pasar untuk mengembangkan sistem prediksi harga saham yang lebih responsif terhadap dinamika pasar.

- C. Menjadi dasar bagi pengembangan aplikasi atau sistem prediksi harga saham yang memanfaatkan integrasi data historis dan data sentimen secara real-time.

### 1.5 Ruang Lingkup

Penelitian ini memiliki ruang lingkup yang difokuskan pada pemanfaatan data historis harga saham dan data sentimen dari media sosial Twitter untuk membangun model prediksi harga saham. Adapun ruang lingkup penelitian ini dijabarkan sebagai berikut: Penelitian ini dibatasi pada:

- A. Penelitian menggunakan data historis saham dari perusahaan tertentu yang terdaftar di bursa, dalam periode waktu tertentu (misalnya satu tahun terakhir)
- B. Informasi diambil dari tweet publik yang relevan dengan saham perusahaan tersebut menggunakan kata kunci atau tagar tertentu.
- C. Sentimen akan diklasifikasikan ke dalam kategori positif, negatif, atau netral menggunakan metode pemrosesan bahasa alami (Natural Language Processing/NLP) seperti IndoBERT atau RoBERTa.
- D. Model prediksi harga saham akan dikembangkan dengan pendekatan machine learning (XGBoost, SVR, Random Forest, MLP, dan Logistic Regression) dan dibandingkan antara model berbasis data historis saja dan model yang juga mengintegrasikan sentimen Twitter.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Kajian Teori dan Konsep Penting**

##### **2.1.1 Data Mining dan Machine Learning**

- A. Data mining adalah proses ekstraksi informasi berharga, pola, dan pengetahuan yang tersembunyi dari kumpulan data yang besar dan kompleks. Tujuan utamanya adalah untuk mengidentifikasi hubungan dan tren yang dapat digunakan untuk mendukung pengambilan keputusan strategis. Dalam esensinya, data mining merupakan teknik analisis yang menggunakan metode statistik, matematika, dan kecerdasan buatan untuk menggali pengetahuan yang belum diketahui secara otomatis dari data [9].
- B. Machine learning merupakan bagian dari artificial intelligence/kecerdasan buatan yang membutuhkan data-data valid untuk proses belajarnya. Machine learning dapat membuat keputusan yang tepat dan cepat, serta dapat memberikan solusi terhadap berbagai permasalahan. Machine learning memiliki kemampuan untuk belajar sendiri dan memutuskan sesuatu tanpa harus diprogram berulang kali oleh manusia, hal ini dapat terjadi karena adanya pengalaman berbagai data yang dimiliki [10].

##### **2.1.2 Teknik yang Digunakan**

###### **A. Bi-LSTM dan RoBERTa**

**Bi-LSTM** pengembangan dari model LSTM yang memproses data sekuensial dari dua arah, yaitu forward dan backward. Dengan arsitektur ini, Bi-LSTM mampu menangkap konteks kata sebelum dan sesudah secara lebih baik dibanding LSTM biasa. Hal ini membuat Bi-LSTM banyak digunakan dalam analisis sentimen karena dapat memahami dependensi kata dalam kalimat secara menyeluruh, sehingga meningkatkan akurasi klasifikasi sentimen teks.

**RoBERTa** (Robustly Optimized BERT Pretraining Approach) adalah model transformer yang merupakan pengembangan dari BERT dengan optimasi pada jumlah data pre-training, ukuran batch, dan strategi masking yang lebih dinamis. RoBERTa terbukti memiliki performa yang lebih tinggi dibanding BERT pada berbagai tugas NLP, termasuk analisis sentimen media sosial. Model ini dapat



memahami makna kata dalam konteks kalimat secara lebih mendalam dan kompleks, sehingga menghasilkan klasifikasi sentimen yang lebih akurat.

#### B. Algoritma Klasifikasi Sentimen

- a. Support Vector Regression (SVR): Meskipun SVR umumnya digunakan untuk regresi, dalam penelitian analisis sentimen SVR dapat digunakan untuk memprediksi skor sentimen yang kemudian dipetakan menjadi kategori sentimen positif, negatif, atau netral.
- b. Multilayer Perceptron (MLP): Merupakan model jaringan saraf tiruan (artificial neural network) yang terdiri dari beberapa lapisan tersembunyi (hidden layers) dan dapat menangkap pola non-linear kompleks dalam data teks.
- c. Logistic Regression: Model regresi untuk klasifikasi biner atau multi-kelas yang banyak digunakan sebagai baseline pada analisis sentimen karena interpretasinya yang sederhana dan proses training yang cepat.
- d. Extreme Gradient Boosting (XGBoost): Merupakan algoritma boosting berbasis decision tree yang memiliki performa tinggi dan efisien. XGBoost sering digunakan pada kompetisi data science karena akurasi yang baik pada berbagai jenis dataset.
- e. Random Forest: Algoritma ensemble learning yang menggabungkan banyak decision tree untuk meningkatkan akurasi dan mengurangi overfitting, efektif pada data dengan banyak fitur seperti representasi TF-IDF.

#### C. Model Prediksi Harga Saham

Untuk prediksi harga saham, model yang digunakan antara lain:

- a. Long Short-Term Memory (LSTM): Jaringan saraf tiruan yang dirancang khusus untuk menangani data deret waktu dengan memperhatikan dependensi jangka panjang.
- b. Random Forest dan SVR: Dapat digunakan untuk regresi harga saham atau klasifikasi arah pergerakan harga saham.

### 2.2 Studi Terkait (Penelitian Sejenis)

Beberapa penelitian sebelumnya telah dilakukan untuk mengintegrasikan data sentimen dengan data historis dalam prediksi harga saham. Penelitian oleh Maharani et al. melakukan post-training IndoBERT dengan korpus finansial Indonesia untuk meningkatkan akurasi analisis sentimen dan

topik di domain keuangan, menunjukkan potensi pengembangan model RoBERTa domain-spesifik untuk analisis sentimen finansial [11].

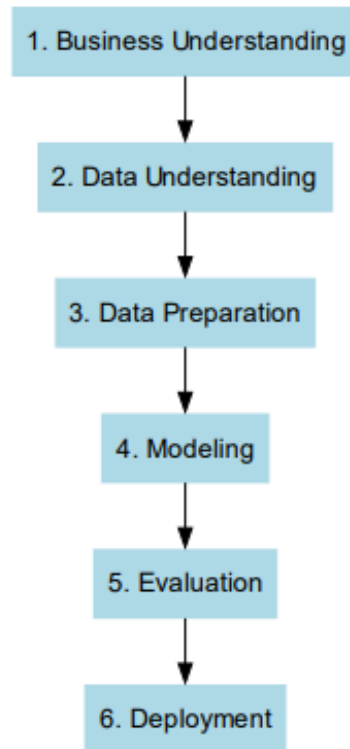
Penelitian lain menggabungkan data tweet dan berita dalam model prediksi harga saham menggunakan MLP dan LSTM, di mana hasilnya menunjukkan bahwa integrasi kedua sumber data tersebut dapat meningkatkan akurasi model dibandingkan hanya menggunakan data historis[12]. Selain itu, terdapat penelitian yang menggunakan analisis sentimen microblogging (Twitter) dan machine learning untuk prediksi harga saham, yang menunjukkan bahwa integrasi data sentimen dengan data historis mampu meningkatkan akurasi prediksi [13].

Penelitian lainnya juga menggabungkan data tweet dan berita dalam prediksi harga saham menggunakan metode machine learning seperti MLP dan LSTM, dan hasilnya menunjukkan bahwa kombinasi kedua sumber data tersebut dapat meningkatkan akurasi model prediksi dibandingkan hanya menggunakan data historis saham [14]. Studi lain yang menggunakan analisis sentimen microblogging (Twitter) dan machine learning untuk memprediksi pasar saham menunjukkan bahwa model yang mengintegrasikan data sentimen dari media sosial dengan data historis dapat memberikan hasil prediksi yang lebih akurat [15].

Selain itu, terdapat penelitian yang mengembangkan model prediksi harga saham dengan menggabungkan data historis dan sentimen Twitter menggunakan Bi-LSTM dan RoBERTa, dan hasil penelitian tersebut menunjukkan bahwa integrasi data sentimen dengan data historis secara signifikan meningkatkan akurasi prediksi dibandingkan model yang hanya menggunakan data historis [15]. Penelitian lainnya mengumpulkan lebih dari 12.000 komentar investor saham di China, dan menggunakan Bi-LSTM untuk prediksi harga saham dengan integrasi analisis sentimen. Hasilnya menunjukkan bahwa akurasi prediksi meningkat ketika sentimen dimasukkan ke dalam model [16]. Penelitian lain juga menggabungkan RoBERTa untuk analisis sentimen microblog (mirip Twitter) dan LSTM untuk prediksi harga saham, dan penelitian ini menunjukkan bahwa model integrasi mampu outperform model berbasis data historis saja [17].

### 2.3 Visualisasi (Diagram Alir Konsep)

Diagram alur berikut menggambarkan tahapan-tahapan utama yang dilakukan dalam penelitian ini, dimulai dari pemahaman permasalahan bisnis hingga tahap deployment. Penelitian ini mengadopsi model proses CRISP-DM (Cross Industry Standard Process for Data Mining) sebagai kerangka kerja utama karena model ini terbukti fleksibel dan banyak digunakan dalam proyek data mining. Setiap tahapan saling berkaitan secara iteratif, memungkinkan penyesuaian kembali terhadap proses sebelumnya bila ditemukan temuan baru dalam proses selanjutnya. Berikut adalah diagram alur dalam penelitian ini:



## 2.4 State of The Art

Perkembangan terkini dalam bidang prediksi harga saham menggunakan pendekatan data mining, khususnya yang menggabungkan data historis dan data sentimen media sosial. Dalam beberapa tahun terakhir, pemanfaatan media sosial seperti Twitter sebagai sumber data alternatif dalam prediksi pasar saham semakin berkembang.

Salah satu pendekatan mutakhir yang banyak digunakan adalah integrasi machine learning dengan data non-tradisional, seperti opini publik dari media sosial. Model seperti Long Short-Term Memory (LSTM) dan Random Forest telah digunakan secara luas karena kemampuannya dalam mengenali pola dari data waktu dan menangani ketidakpastian dalam data pasar yang fluktuatif.

Selain itu, teknologi Natural Language Processing (NLP) juga mengalami kemajuan pesat. Model analisis sentimen tidak lagi terbatas pada metode leksikal sederhana, tetapi mulai beralih ke model berbasis pembelajaran mendalam seperti BERT, XLNet, dan model pre-trained lainnya. Model ini dapat memahami konteks dan nuansa bahasa alami dengan lebih baik.

Secara umum, perkembangan terkini menunjukkan bahwa model prediksi yang menggabungkan data historis dan sentimen sosial dapat memberikan akurasi yang lebih tinggi dan respon yang lebih cepat terhadap perubahan pasar, terutama pada kondisi yang tidak stabil atau penuh ketidakpastian.

## BAB III

### METODOLOGI PENELITIAN

#### 3.1 Tahapan penelitian

Dalam upaya memperoleh hasil analisis data yang terarah dan dapat dipertanggungjawabkan secara metodologis, digunakan pendekatan berbasis kerangka kerja yang telah teruji seperti CRISP-DM, sebuah model proses standar independen dari industri untuk data mining yang terdiri dari enam fase iterative [18].

Tabel 1, deskripsi proses CRISP-DM [2]

Fase	Deskripsi
<i>Business Understanding</i>	Memahami situasi bisnis, menentukan tujuan data mining, seperti klasifikasi (dalam laporan ini) serta kriteria keberhasilan, dan menyusun rencana proyek.
<i>Data Understanding</i>	Mengumpulkan data dari sumber yang relevan, mengeksplorasi, mendeskripsikan, dan memeriksa kualitas data menggunakan analisis statistik.
<i>Data Preparation</i>	Melakukan seleksi data dengan kriteria inklusi-eksklusi, membersihkan data yang berkualitas buruk, serta membangun atribut turunan sesuai model yang akan digunakan.
<i>Modeling</i>	Memilih teknik pemodelan, menyusun kasus uji, membangun model, menetapkan parameter, lalu mengevaluasi model sesuai kriteria yang telah ditentukan.
<i>Evaluation</i>	Memeriksa hasil model terhadap tujuan bisnis awal, menginterpretasi hasil, menentukan tindakan selanjutnya, dan melakukan review keseluruhan proses.
<i>Deployment</i>	Menerapkan hasil melalui laporan akhir atau komponen perangkat lunak, serta merencanakan pemantauan dan pemeliharaan implementasi model.

Namun, dalam laporan kali ini, proses metodologi hanya akan dilaksanakan sampai pada tahap *Evaluation*, sehingga tahapan *Deployment* tidak menjadi fokus kajian.

### 3.2 Deskripsi Dataset

Tabel 2, Dataset

Tipe	Sumber	Ukuran	Attribut
sentimen	x/twitter	2232	15 (conversation_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)
historis	website (investing.com)	37 / day	7 (Tanggal, Terakhir, Pembukaan, Tertinggi, Terendah, Vol., Perubahan%)

### 3.3 Algoritma / Data Mining Tools

Tabel 3, Algoritma

Algoritma	Alasan
Random Forest	Kuat pada data dengan interaksi antar fitur & tahan noise. Cocok untuk regresi harga saham berdasarkan banyak lag features.
XGBoost	Sering outperform model lain dalam kompetisi prediksi harga karena fokus pada residual dan regularisasi kuat.
Logistic Regression	Baseline linear sederhana untuk memeriksa apakah data cukup dijelaskan oleh relasi linear, sangat interpretatif.
SVR	Menangkap hubungan non-linear dengan kernel (misalnya RBF), memfokuskan prediksi pada pola utama dengan mengabaikan outlier kecil.
MLP	Neural network dasar untuk mempelajari representasi kompleks antar waktu tanpa harus memprogram interaksi secara manual.

Tabel 4, Tools

Tools	Fungsi
pandas	Mengelola data frame, generate lag features, moving average, RSI.
numpy	Operasi numerik cepat (vectorized), misalnya menghitung return.
matplotlib / seaborn	Visualisasi distribusi harga, heatmap korelasi, plot prediksi.

scikit-learn	Pipeline training: Random Forest, Logistic Regression, SVR, MLP; preprocessing (StandardScaler/MinMaxScaler), metrics (accuracy, confusion matrix).
xgboost	Model XGBoostClassifier, powerful untuk data tabular.
pandas_ta / ta-lib	Menghitung indikator teknikal (SMA, RSI, MACD, Bollinger Band).
statsmodels	Uji stasionaritas (ADF Test) jika ingin mengecek pola.
yellowbrick	Visualisasi residual, ROC, class prediction error.
mlflow / wandb	Tracking experiment untuk tuning hyperparameter & logging metrics.
joblib / pickle	Menyimpan model hasil training.

### 3.4 Evaluasi Kinerja

Tabel 5, Evaluasi

Evaluasi	Penjelasan	Alasan
Accuracy	Proporsi prediksi arah benar (misalnya naik vs turun) dari seluruh prediksi.	Untuk baseline sederhana: seberapa sering model benar dalam memprediksi arah.
Precision	Dari semua yang diprediksi naik, berapa yang benar-benar naik.	Menghindari false signals, penting jika cost salah beli mahal.
Recall	Dari semua hari yang benar-benar naik, berapa yang terprediksi naik.	
F1-Score	Harmonik rata-rata precision dan recall.	Seimbang memerhatikan missed naik (false negative) dan false naik (false positive).
ROC AUC Score	Area under curve ROC yang membandingkan True Positive Rate vs False Positive Rate di semua threshold.	Untuk melihat kemampuan model membedakan naik vs turun terlepas threshold.
Confusion Matrix	Matriks jumlah prediksi naik/turun vs aktual naik/turun.	Memberi gambaran kesalahan model, bisa fokus memperbaiki misclassification.
Directional Accuracy	Persentase prediksi arah benar (mirip accuracy), tapi kadang dihitung dari return positif/negatif.	Sangat relevan dalam trading untuk memastikan prediksi arah lebih sering tepat.

## DAFTAR PUSTAKA

- [1] A. Bagheffar and C. Saous, "The Impact of Investor Sentiment on Stock Returns in the Indonesian Stock Market During the Period (2001-2022): An Econometric Study."
- [2] "Investor sentiment and stock prices," *Academic Journal of Business & Management*, vol. 5, no. 22, 2023, doi: 10.25236/ajbm.2023.052215.
- [3] G. Liu, Y. Yang, W. Mo, W. Gu, and R. Wang, "Private Placement, Investor Sentiment, and Stock Price Anomaly," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 27, no. 5, pp. 771–779, Sep. 2023, doi: 10.20965/jaciii.2023.p0771.
- [4] Z. Janková, "CRITICAL REVIEW OF TEXT MINING AND SENTIMENT ANALYSIS FOR STOCK MARKET PREDICTION," *Journal of Business Economics and Management*, vol. 24, no. 1, pp. 177–198, Jan. 2023, doi: 10.3846/jbem.2023.18805.
- [5] P. Patel, "Real-Time Sentiment Analysis of Twitter Streams for Stock Forecasting," *International Journal of Computer Trends and Technology*, vol. 72, no. 5, pp. 204–209, May 2024, doi: 10.14445/22312803/ijctt-v72i5p125.
- [6] M. Mokhtari, A. Seraj, N. Saeedi, and A. Karshenas, "The Impact of Twitter Sentiments on Stock Market Trends."
- [7] Z. Li, "The Impact of Social Media Sentiment on Stock Price Changes," *Advances in Economics, Management and Political Sciences*, vol. 170, no. 1, pp. 49–59, Jun. 2025, doi: 10.54254/2754-1169/2025.lh23972.
- [8] E. Arif, S. Suherman, and A. P. Widodo, "Predicting Stock Prices of Digital Banks: A Machine Learning Approach Combining Historical Data and Social Media Sentiment from X," *Ingenierie des Systemes d'Information*, vol. 30, no. 3, pp. 687–701, Mar. 2025, doi: 10.18280/isi.300313.
- [9] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024.
- [10] G. H. F. N. S. D. B. I. T. A. V. Y. P. R. Y. I. I. A. L. S. C. A. C. R. A. I. R. S. M. R. S. R. Maulani, *Machine Learning*. CV. Mega Press Nusantara, 2025.
- [11] N. P. I. Maharani, Y. Yustiawan, F. C. Rochim, and A. Purwarianti, "Domain-Specific Language Model Post-Training for Indonesian Financial NLP," Oct. 2023, [Online]. Available: <http://arxiv.org/abs/2310.09736>
- [12] H. Zolfagharinia, M. Najafi, S. Rizvi, and A. Haghighi, "Unleashing the Power of Tweets and News in Stock-Price Prediction Using Machine-Learning Techniques," *Algorithms*, vol. 17, no. 6, Jun. 2024, doi: 10.3390/a17060234.
- [13] P. Koukaras, C. Nousi, and C. Tjortjis, "Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning," *Telecom*, vol. 3, no. 2, pp. 358–378, Jun. 2022, doi: 10.3390/telecom3020019.
- [14] H. Zolfagharinia, M. Najafi, S. Rizvi, and A. Haghighi, "Unleashing the Power of Tweets and News in Stock-Price Prediction Using Machine-Learning Techniques," *Algorithms*, vol. 17, no. 6, Jun. 2024, doi: 10.3390/a17060234.

- [15] P. Koukaras, C. Nousi, and C. Tjortjis, "Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning," *Telecom*, vol. 3, no. 2, pp. 358–378, Jun. 2022, doi: 10.3390/telecom3020019.
- [16] Y. Luan, H. Zhang, C. Zhang, Y. Mu, and W. Wang, "Stock Price Prediction with Sentiment Analysis for Chinese Market," 2024. [Online]. Available: <http://www.data.csmar.com>
- [17] P. Koukaras, C. Nousi, and C. Tjortjis, "Stock Market Prediction Using Microblogging Sentiment Analysis and Machine Learning," *Telecom*, vol. 3, no. 2, pp. 358–378, Jun. 2022, doi: 10.3390/telecom3020019.
- [18] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," in *Procedia Computer Science*, Elsevier B.V., 2021, pp. 526–534. doi: 10.1016/j.procs.2021.01.199.



## LAMPIRAN