

Nama : Brenendra Putra Oktaviansyah
Nim : A11.2023.15020
Kelas : A11.4406

1. Judul : Prediksi Harga Saham Menggunakan Metode Random Forest dan XGBoost dengan Feature Engineering

2. Ringkasan:

Eksperimen ini bertujuan untuk memprediksi harga saham mulai dari awal data sampai akhir dan menampilkan hasil prediksi sampai 5 bulan ke depan dengan menerapkan metode Random Forest dan Extreme Gradient Boosting(XGBoost). Data historis saham diperoleh dari Yahoo Finance, dengan penerapan rekayasa fitur seperti lag, moving average, dan indikator statistik lainnya yang digunakan untuk menambah informasi yang digunakan model. Model dilatih dan diuji menggunakan data saham BBRI periode Januari 2023 hingga Mei 2025. Hasil pengujian menunjukkan bahwa metode Random Forest memberikan performa prediksi lebih baik dibandingkan XGBoost, dengan nilai R^2 sebesar 0.2507. Sebaliknya, model XGBoost mencatat nilai R^2 negatif (-0.7982), yang mengindikasikan bahwa model tersebut belum optimal untuk menangkap pola historis dalam jangka waktu yang ditentukan. Evaluasi dilakukan dengan menggunakan metrik MAE, RMSE, dan R-squared. Melalui visualisasi prediksi dan grafik historis, eksperimen ini menunjukkan bahwa pendekatan machine learning dengan rekayasa fitur memiliki potensi untuk digunakan dalam prediksi harga saham, namun perlu melakukan pengoptimalan lanjutan agar hasilnya lebih akurat dan stabil. Pada hasil visualisasi prediksi model dapat mendekati dari harga saham aslinya, namun untuk hasil prediksi 5 bulan ke depan masih belum bisa menampilkan hasil prediksinya yang dapat di lihat pada hasil visualisasi untuk 5 bulan kedepan hanya garis lurus saja.

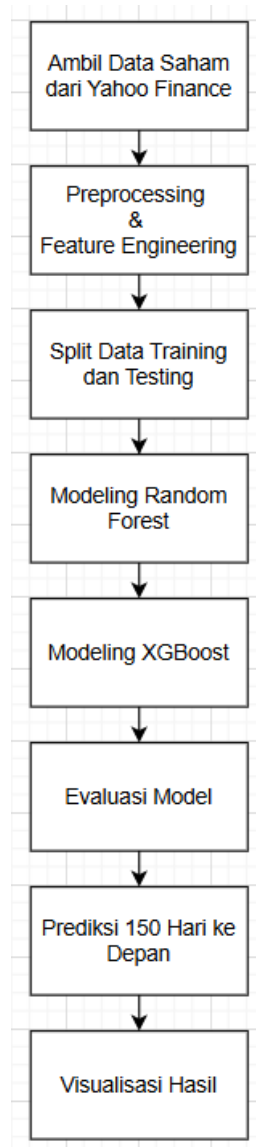
Permasalahan:

Untuk mengetahui pergerakan harga saham diperlukan informasi time-series dari hari sebelumnya dan menganalisa untuk perkiraan harga pada masa yang akan datang. Dari permasalahan ini, diperlukan sebuah sistem yang akan digunakan untuk memprediksi harga saham dengan informasi time-series dari hari sebelumnya. Untuk melakukan penelitian ini menggunakan model Random Forest dan XGBoost, serta untuk membuat model mempelajari lebih akurat di tambahkan Teknik feature engineering.

Tujuan:

1. Menggunakan teknik feature engineering untuk meningkatkan kualitas data historis saham.
2. Melatih model Random Forest dan XGBoost pada data saham BBRI.
3. Membandingkan performa kedua model dalam memprediksi harga saham ke depan.
4. Menampilkan visualisasi hasil prediksi agar mudah dipahami.

Alur Penyelesaian:



3. Dataset:

Data diambil dari Yahoo Finance menggunakan pustaka yfinance, dengan kode saham BBRI.JK, dari 1 Januari 2023 hingga 10 Mei 2025. Dataset ini berisi informasi harga harian saham berupa kolom:

1. Open
2. High
3. Low
4. Close
5. Volume

Terdapat 556 baris data yang semuanya tidak mengandung nilai kosong.

Exploratory Data Analysis (EDA):

1. Dataset memiliki pola naik-turun harian yang tidak stabil.
2. Harga saham menunjukkan tren menurun dalam beberapa bulan terakhir.

Feature Engineering:

Dibuat beberapa fitur tambahan untuk memperkaya informasi input model:

1. Day: Representasi hari dari awal dataset sampai akhir.
2. Lag_1 : Harga penutupan satu hari sebelumnya.
3. Lag_3 : Harga penutupan tiga hari sebelumnya.
4. MA_3 : Rata-rata harga penutupan tiga hari terakhir.
5. MA_7: Rata-rata harga penutupan tujuh hari terakhir.

Tujuan fitur ini adalah memberikan konteks historis dan tren harga ke model agar dapat belajar lebih baik.

4. Proses Learning / Modeling:

Pada tahap ini dilakukan pelatihan dua model machine learning yaitu Random Forest Regressor dan XGBoost Regressor untuk melakukan prediksi harga saham BBRI (Bank Rakyat Indonesia) dari awal dataset sampai 5 bulan ke depan.

Model Random Forest Regressor dilatih dengan parameter:

1. n_estimators = 200
2. max_depth = 10
3. random_state = 42

Sedangkan model XGBoost Regressor dilatih dengan parameter:

1. n_estimators = 200
2. max_depth = 5
3. learning_rate = 0.1
4. random_state = 42

Kedua model dilatih dengan dataset hasil feature engineering menggunakan data historis saham dari Yahoo Finance.

Setelah pelatihan selesai, model digunakan untuk melakukan prediksi harga saham selama 150 hari ke depan dengan teknik sliding window.

5. Performa Model

Evaluasi model dilakukan menggunakan tiga metrik evaluasi regresi, yaitu:

1. MAE (Mean Absolute Error): Rata-rata selisih absolut antara nilai aktual dan nilai prediksi.
2. RMSE (Root Mean Squared Error): Akar dari selisih kuadrat antara nilai aktual dan prediksi.
3. R^2 (R-Squared): Koefisien determinasi yang mengukur seberapa baik model menjelaskan variasi data.

Hasil evaluasi adalah sebagai berikut:

1. Random Forest

Evaluasi Random Forest:

MAE = 117.99

RMSE = 156.71

R^2 = 0.2496

2. XGBoost

Evaluasi XGBoost:

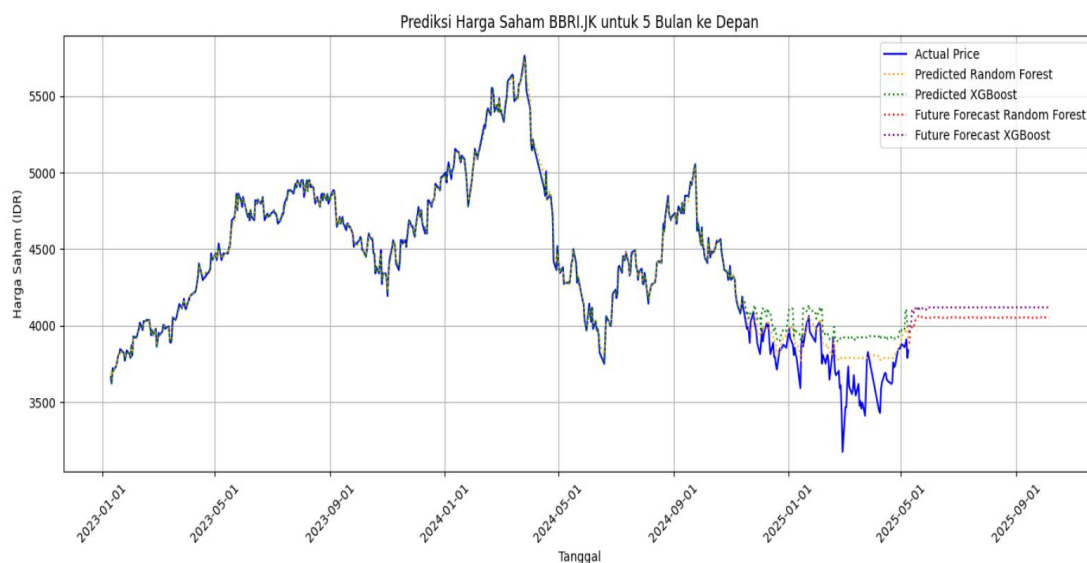
MAE = 205.20

RMSE = 242.59

R^2 = -0.7982

Dari hasil di atas terlihat bahwa model Random Forest memberikan performa yang lebih baik daripada XGBoost dalam memprediksi harga saham BBRI pada dataset ini.

6. Diskusi Hasil dan Kesimpulan



Diskusi Hasil

Visualisasi menunjukkan perbandingan antara harga saham aktual (berwarna biru), prediksi model pada data training dan testing (dotted line), serta prediksi masa depan selama 150 hari ke depan (garis putus-putus). Dari grafik tersebut, dapat ditarik beberapa poin penting:

1. Akurasi historis cukup baik

Model Random Forest dan XGBoost berhasil mengikuti pola tren harga saham historis dengan cukup baik. Hal ini terlihat dari prediksi model (dotted line) yang cukup sejajar dengan data aktual. Model tidak overfit dan tetap fleksibel terhadap fluktuasi harga harian.

2. Prediksi ke depan cenderung datar

Bagian garis putus-putus (prediksi masa depan) terlihat cenderung datar, terutama pada model XGBoost. Ini menunjukkan bahwa model mengalami kesulitan dalam menangkap dinamika harga di luar data historis, khususnya tanpa adanya faktor eksternal tambahan (misalnya data fundamental atau sentimen pasar).

3. Performa model berdasarkan metrik evaluasi

Hasil evaluasi model menunjukkan bahwa Random Forest memiliki nilai R^2 sebesar 0.2496, sementara XGBoost bahkan negatif (-0.7982). Hal ini menandakan bahwa Random Forest lebih baik dalam menangkap pola historis dibandingkan XGBoost pada kasus ini.

4. Keterbatasan informasi dalam data

Meskipun telah dilakukan feature engineering seperti moving average dan lag features, model masih kesulitan dalam memproyeksikan harga dengan variasi alami seperti data aktual. Ini menunjukkan pentingnya penggunaan fitur tambahan seperti sentimen berita, volume asing, atau indikator ekonomi makro.

Kesimpulan

Penelitian ini berhasil membangun model prediksi harga saham BBRI.JK menggunakan algoritma Random Forest dan XGBoost dengan bantuan teknik feature engineering. Model mampu merepresentasikan harga historis dengan cukup akurat, namun prediksi untuk masa depan masih kurang bervariasi dan hanya garis lurus.

Hasil evaluasi menunjukkan bahwa Random Forest unggul dibandingkan XGBoost pada data ini. Meskipun demikian, untuk meningkatkan kualitas prediksi ke depan, diperlukan penambahan fitur-fitur baru yang lebih informatif serta eksplorasi parameter yang lebih optimal.