

Laporan Tugas Prediksi Harga Saham dengan Kombinasi Machine Learning dan Gerakan Brown Geometrik (GBM)

Tugas Besar Pemodelan dan Simulasi Kelompok 10



Anggota Kelompok:

Muhammad Hermawan Alghozy - 1301213473

Valens Christian Yosua - 1301213033

Zaky Zaidan – 1301213354

1. Pendahuluan

Dalam dunia keuangan, prediksi harga saham merupakan topik yang sangat penting dan menantang. Banyak metode telah dikembangkan untuk memodelkan pergerakan harga saham, termasuk metode statistik tradisional dan pendekatan machine learning. Salah satu model statistik yang populer adalah Gerakan Brown Geometrik (GBM), yang digunakan untuk memodelkan pergerakan harga saham dengan mempertimbangkan drift (perubahan harga) dan volatilitas. Dalam laporan ini, kami menggabungkan pendekatan GBM dengan machine learning untuk memprediksi harga saham dengan lebih akurat.

2. Tujuan

Tujuan dari tugas ini adalah:

- Mempelajari konsep Gerakan Brown Geometrik (GBM) untuk simulasi perubahan harga saham.
- Menggunakan teknik khusus untuk membuat fitur-fitur penting bagi model machine learning.
- Melatih model machine learning untuk memprediksi nilai drift dalam simulasi GBM.
- Merancang simulasi berbagai kemungkinan perubahan harga saham di masa depan menggunakan GBM dengan nilai drift yang diprediksi oleh machine learning.
- Mengimplementasikan hasil prediksi dalam web interface untuk aplikasi web.

3. Deskripsi Kumpulan Data dan Fitur yang Dipilih

Data yang digunakan dalam analisis ini adalah data historis harga saham dari Bank Jago Tbk. (ARTO.JK). Data ini mencakup harga pembukaan, harga penutupan (CLOSE), harga tertinggi, harga terendah, volume perdagangan, dan tanggal/datetime selama periode dari 12 Januari 2016 hingga 12 Januari 2024. Dataset ini mencakup kolom-kolom berikut:

	Open	High	Low	Close	Adj Close	Volume
Date						
2016-01-12	18.093210	24.831371	17.594088	21.587070	21.587070	201810759
2016-01-13	21.711851	22.335756	19.715361	20.089703	20.089703	56410138
2016-01-14	19.840139	19.840139	18.093210	18.093210	18.093210	19347534
2016-01-15	18.093210	18.093210	16.595840	16.720621	16.720621	6852819
2016-01-18	16.720621	21.711851	15.348033	15.971937	15.971937	90500330

- Date: Tanggal data yang bersangkutan.

- Open: Harga saham pada saat pasar dibuka pada hari tersebut.
- High: Harga tertinggi yang dicapai oleh saham pada hari tersebut.
- Low: Harga terendah yang dicapai oleh saham pada hari tersebut.
- Close: Harga saham pada saat pasar ditutup pada hari tersebut.
- Adj Close: Harga penutupan yang telah disesuaikan untuk semua tindakan korporat (seperti dividen, stock splits, dll).
- Volume: Jumlah saham yang diperdagangkan pada hari tersebut.

Fitur yang Dipilih

1. Log Return Harian: Return logaritmik harian dihitung untuk menghindari masalah pembagian dengan nol.
2. Technical Indicators: Fitur-fitur tambahan berupa indikator teknikal seperti:
 - Simple Moving Average (SMA): Memberikan rata-rata harga yang halus untuk mengidentifikasi tren.
 - Relative Strength Index (RSI): indikator momentum yang mengukur kecepatan dan perubahan pergerakan harga.
 - Exponential Moving Average (EMA): Memberikan bobot lebih besar pada harga terbaru untuk mengidentifikasi tren dengan lebih responsif.
 - Bollinger Bands (BB): Mengukur volatilitas dan mengidentifikasi kondisi overbought atau oversold dengan batas atas dan bawah yang dinamis.

4. Penjelasan Model Machine Learning dan Performanya

Untuk memprediksi nilai drift dalam simulasi GBM, kami menggunakan model machine learning Random Forest Regressor. Langkah-langkah yang dilakukan meliputi:

- Membagi data menjadi set pelatihan dan pengujian dengan proporsi 80% pelatihan dan 20% pengujian.
- Melatih model Random Forest pada data pelatihan menggunakan drift historis (dihitung dari return logaritmik) sebagai target variable.
- Mengevaluasi performa model pada set pengujian menggunakan metrik seperti Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan R-squared (R^2).

Hasil Evaluasi Model

```
Mean Squared Error (MSE): 3.734567747112622e-05
Root Mean Squared Error (RMSE): 0.006111110985011336
Mean Absolute Error (MAE): 0.0005705332877271776
R2 Score: 0.9861287253304377
```

Berdasarkan hasil evaluasi, model random forest ini memberikan hasil performa yang sangat baik. Nilai MSE, RMSE, dan MAE yang rendah menunjukkan bahwa model memiliki kesalahan prediksi yang kecil, sedangkan nilai R2 yang tinggi menunjukkan bahwa model cukup baik dalam menjelaskan variasi dalam data pengujian.

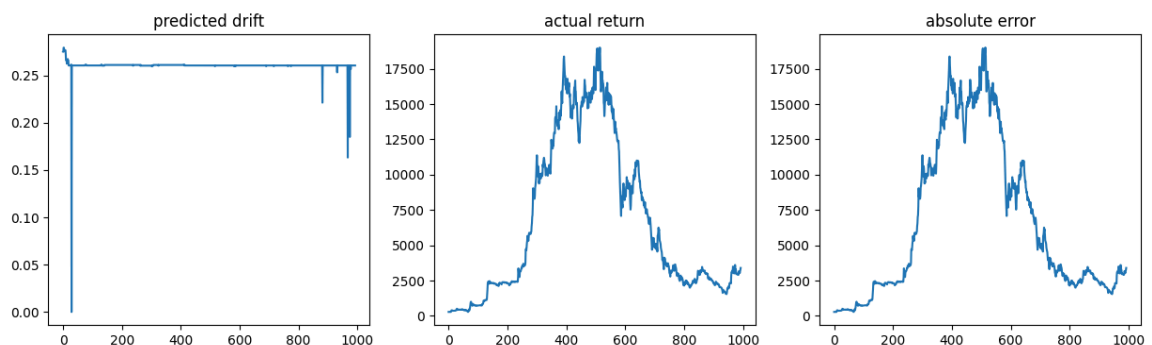
Oleh karena itu, model ini cocok untuk simulasi memprediksi pergerakan harga saham menggunakan teknik kombinasi antara machine learning dan Gerakan Brown Geometrik (GBM).

5. Visualisasi Distribusi Harga Simulasi di Masa Depan

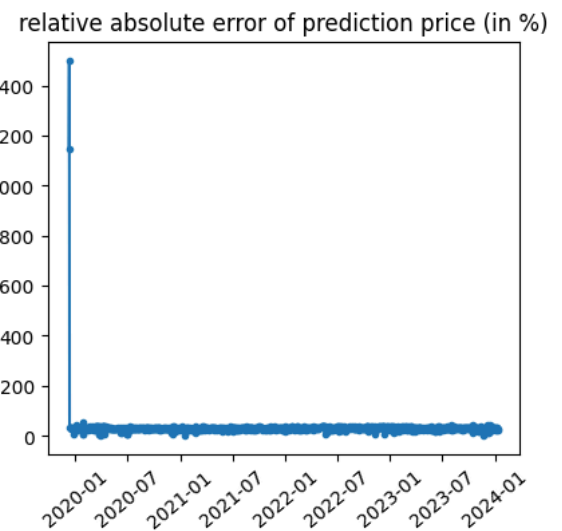
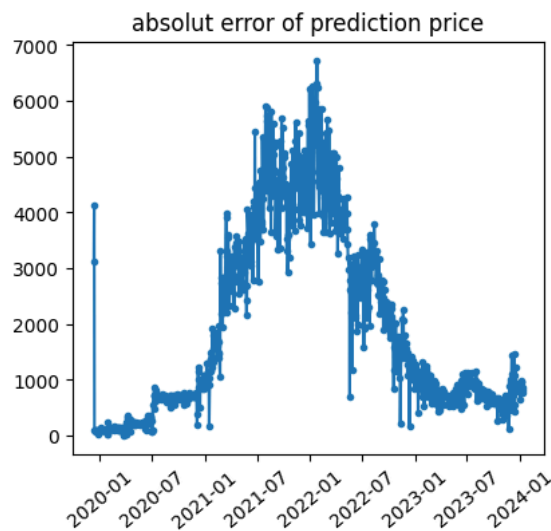
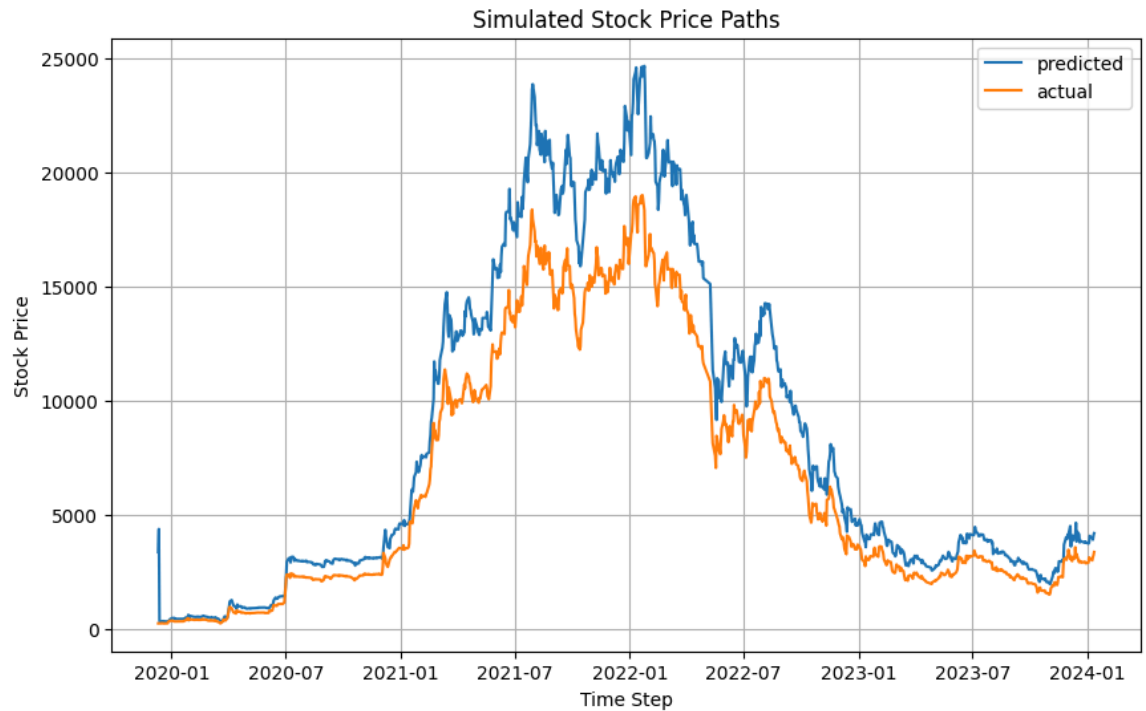
Kami melakukan simulasi GBM untuk harga saham dengan drift yang diprediksi oleh model machine learning. Hasil simulasi ini divisualisasikan dalam bentuk grafik yang menunjukkan jalur harga saham hasil simulasi dan harga saham aktual dalam periode waktu tertentu.

Grafik Prediksi dan Aktual

1. Predicted Drift vs Actual Return: Grafik ini menunjukkan prediksi drift yang dihasilkan oleh model Random Forest dibandingkan dengan return aktual dari harga saham berdasarkan data historis dan memberikan gambaran tentang seberapa besar kesalahan prediksi model pada setiap titik waktu dengan grafik absolute error.



2. Simulated Stock Price Path: Grafik ini menunjukkan jalur harga saham hasil simulasi dan harga saham aktual dalam periode waktu 2020 - 2024.



Penjelasan:

1. Simulated Stock Price Path = Menunjukkan jalur harga saham hasil simulasi (biru) dan harga saham aktual (oren) dalam periode 2020 - 2024. Meskipun hasil prediksi lebih tinggi dibanding aktualnya, hasil simulasi menunjukkan bahwa model Random Forest dan GBM mampu memprediksi jalur harga saham secara cukup akurat dalam beberapa kasus, kedua garis menunjukkan tren yang miridan cenderung sama. Ini menunjukkan bahwa model berhasil menangkap arah pergerakan harga saham.
2. Absolut error of prediction price = Absolute error dihitung sebagai nilai absolut dari selisih antara harga prediksi dan harga aktual.

3. Relative Absolute Error of Prediction Price (in %) = Grafik di panel kanan menunjukkan relative absolute error dalam persentase.

Distribusi Harga Simulasi di Masa Depan

Hasil percentile dari distribusi harga simulasi di masa depan menunjukkan kisaran nilai yang mungkin dan persentase harga yang diharapkan jatuh dalam kisaran tersebut:

- 5th Percentile: Rp550.54
- 25th Percentile: Rp2,998.93
- 50th Percentile (Median): Rp4,573.78
- 75th Percentile: Rp13,987.12
- 95th Percentile: Rp21,083.97

Hasil percentile tersebut memberikan gambaran tentang distribusi harga saham di masa depan, ini menunjukkan kisaran nilai yang mungkin dan persentase harga yang diharapkan jatuh dalam kisaran tersebut.

Contoh nya pada 25th Percentile, Harga di bawah persentil ini menunjukkan bahwa ada 25% kemungkinan harga saham akan berada di bawah nilai Rp2.998,93. Ini membantu memahami rentang potensial harga saham di masa depan dan tingkat risiko yang terkait dengan prediksi tersebut.

6. Diskusi Keterbatasan Pendekatan dan Potensi Peningkatan

Keterbatasan Pendekatan

1. Ketergantungan pada Data Historis: Pendekatan ini sangat bergantung pada data historis, yang mungkin tidak selalu mencerminkan kondisi pasar di masa depan.
2. Kompleksitas Model: Model ini menyederhanakan faktor-faktor yang mempengaruhi harga saham aslinya.

Potensi Peningkatan

1. Peningkatan Model GBM: Menggunakan model volatilitas yang lebih kompleks untuk menangkap perubahan dinamis dalam volatilitas saham.
2. Penggunaan Fitur Tambahan: Menambahkan fitur-fitur yang lebih kaya dan relevan, seperti data fundamental perusahaan, data sentimen dari media sosial atau berita, dan indikator makroekonomi.
3. Mencoba Model Machine Learning Lainnya: Mencoba model LSTM, GRU, dan menggunakan teknik ensemble untuk meningkatkan akurasi prediksi.

7. Kesimpulan

Penerapan kombinasi Random Forest dan Gerakan Brown Geometrik (GBM) menunjukkan potensi untuk memprediksi tren perubahan harga saham dengan cukup baik. Grafik prediksi drift dan actual return menunjukkan tren yang serupa, meskipun terdapat beberapa periode di mana prediksi model tidak sepenuhnya akurat. Oleh karena itu, perlu dilakukan penelitian lebih lanjut untuk meningkatkan akurasi prediksi dan mengevaluasi model secara menyeluruh.

Lampiran

Link Collab:

<https://colab.research.google.com/drive/13C7yz0y8ZfNCEHnSo7BfqABJvthk9H8?usp=sharing>