

TUGAS ANALISIS BIG DATA

KONSEP DATA MINING

Dosen Pengampu: Sevi Nurafni ST., M.Si., M.Sc.



IKOPIN University

Disusun Oleh Ryan Fadhilah Faizal Hakim

Program Studi Sains Data

Fakultas Sains dan Teknologi

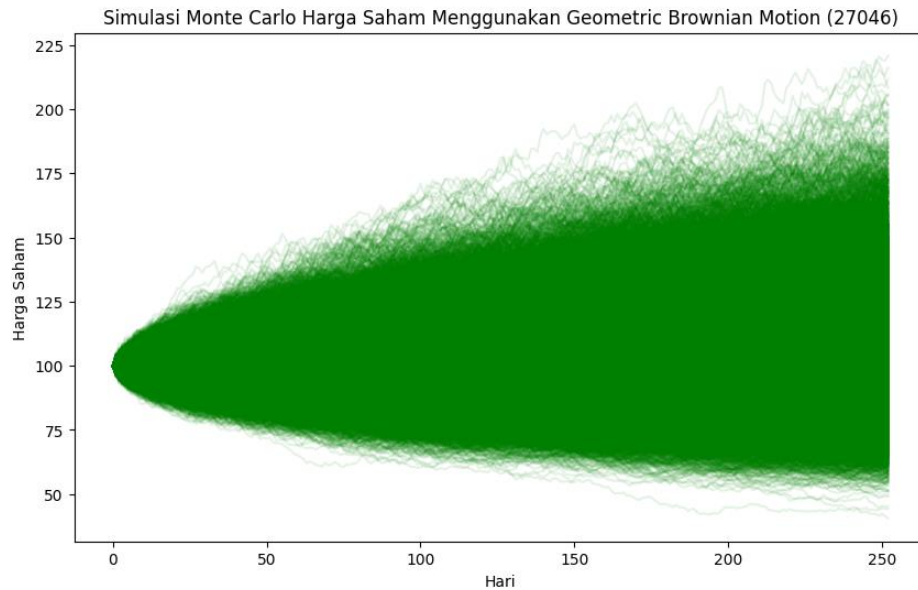
©2024 Ryan Hakim All Rights Reserved

Kasus:

1. Jalankan ulang program Monte Carlo untuk harga saham sebanyak 5 kali dengan jumlah simulasi yang berbeda-beda. Apa yang anda pahami dari output tersebut?
2. Dalam konteks Data Mining, jelaskan perbedaan antara metodologi, model, dan algoritma. Bagaimana ketiga konsep ini saling berhubungan dalam proses analisis data?

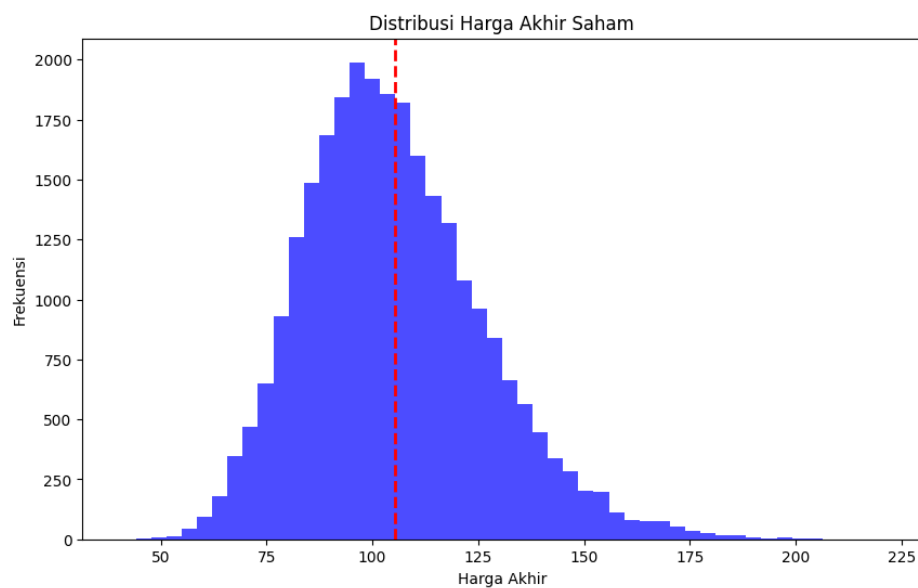
Penyelesaian:

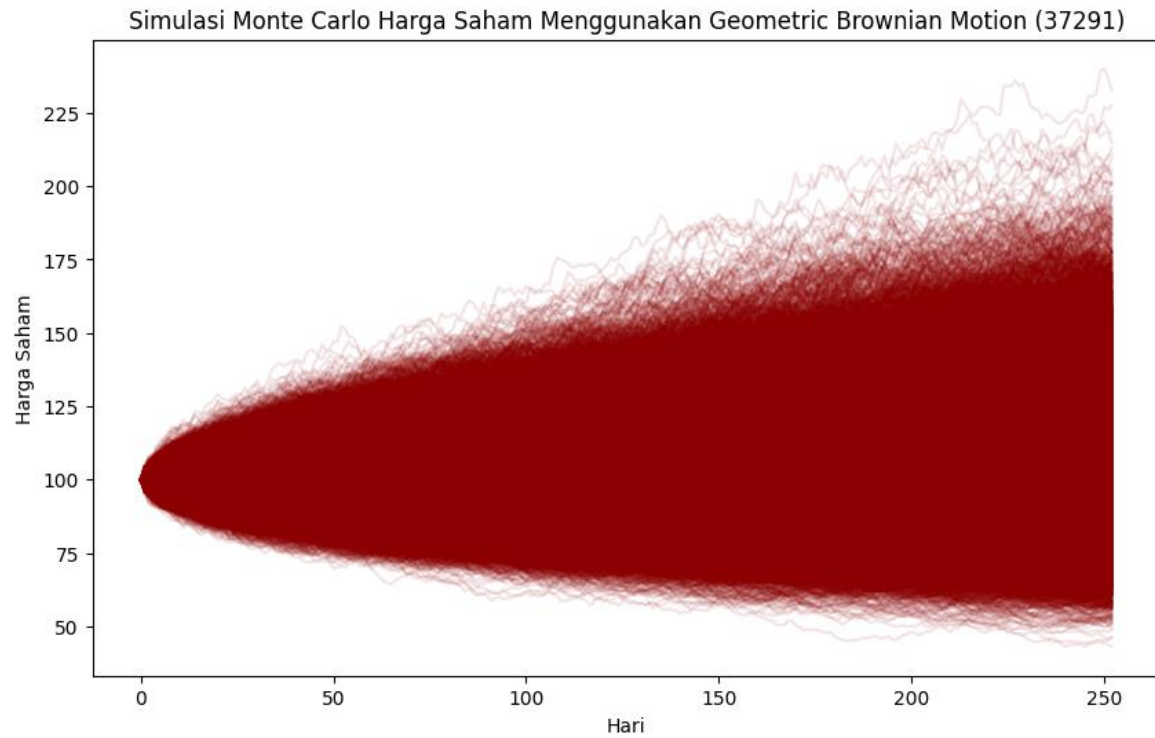
1. Simulasi Monte Carlo (Monte Carlo Simulation/MCS)



Estimasi harga akhir rata-rata: 105.22

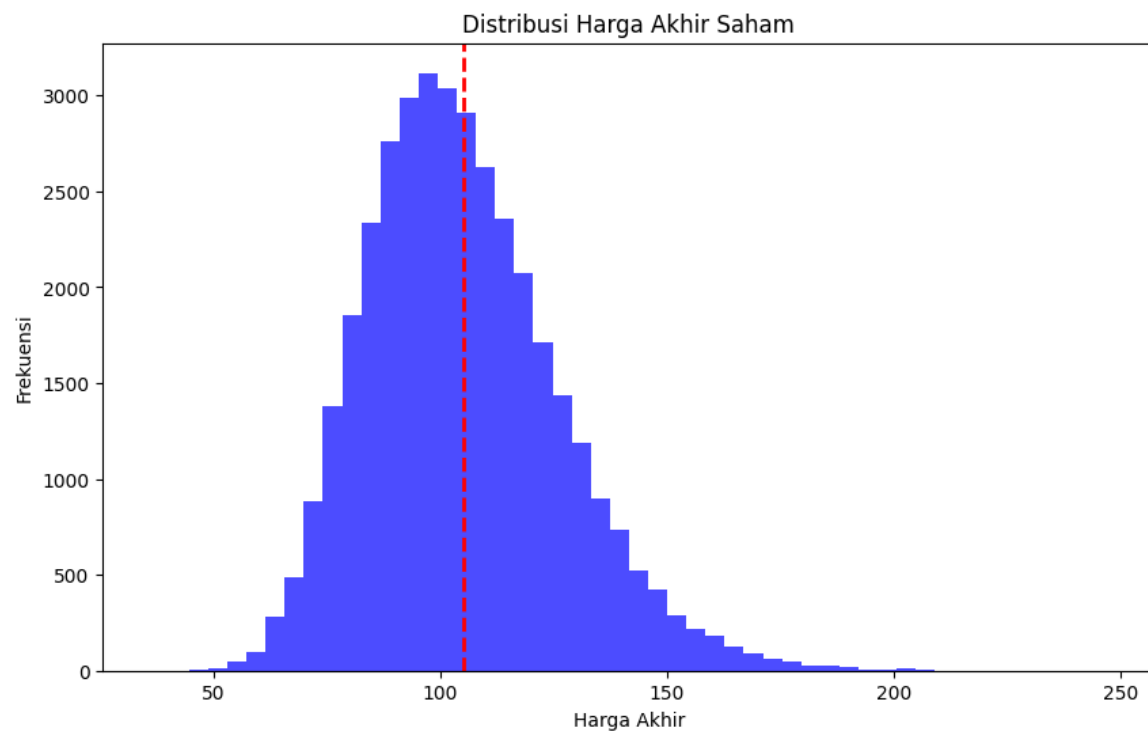
Deviasi standar harga akhir: 21.29

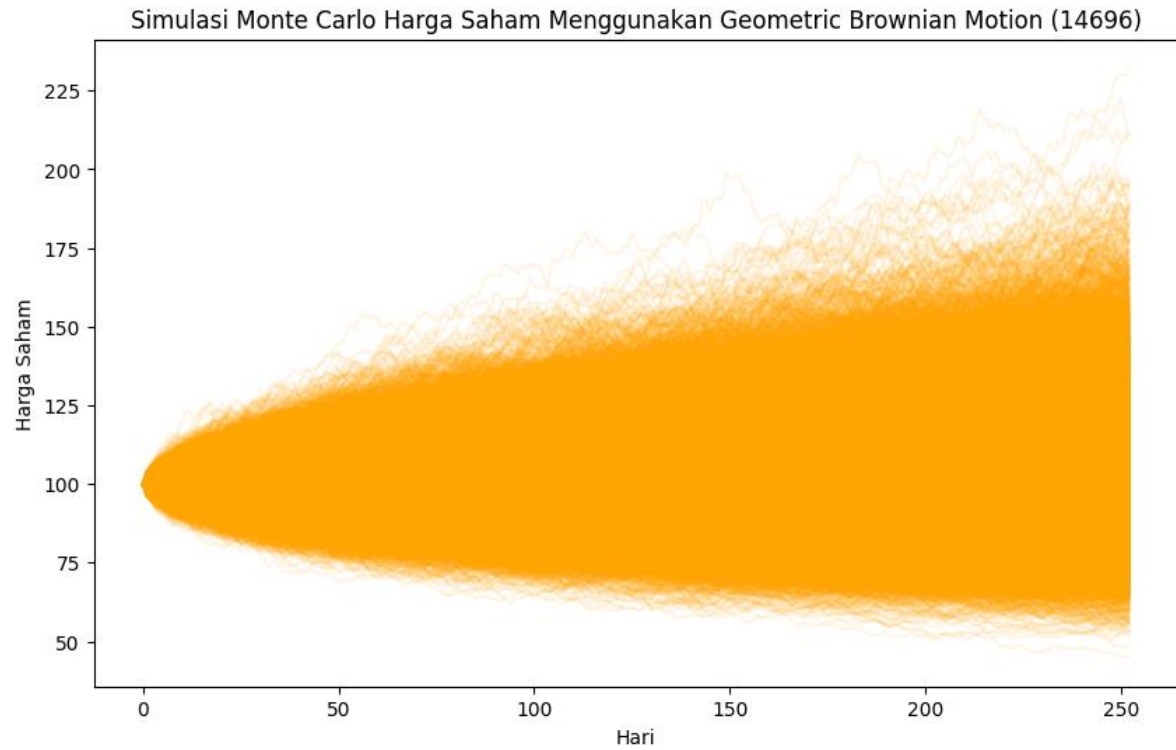




Estimasi harga akhir rata-rata: 105.00

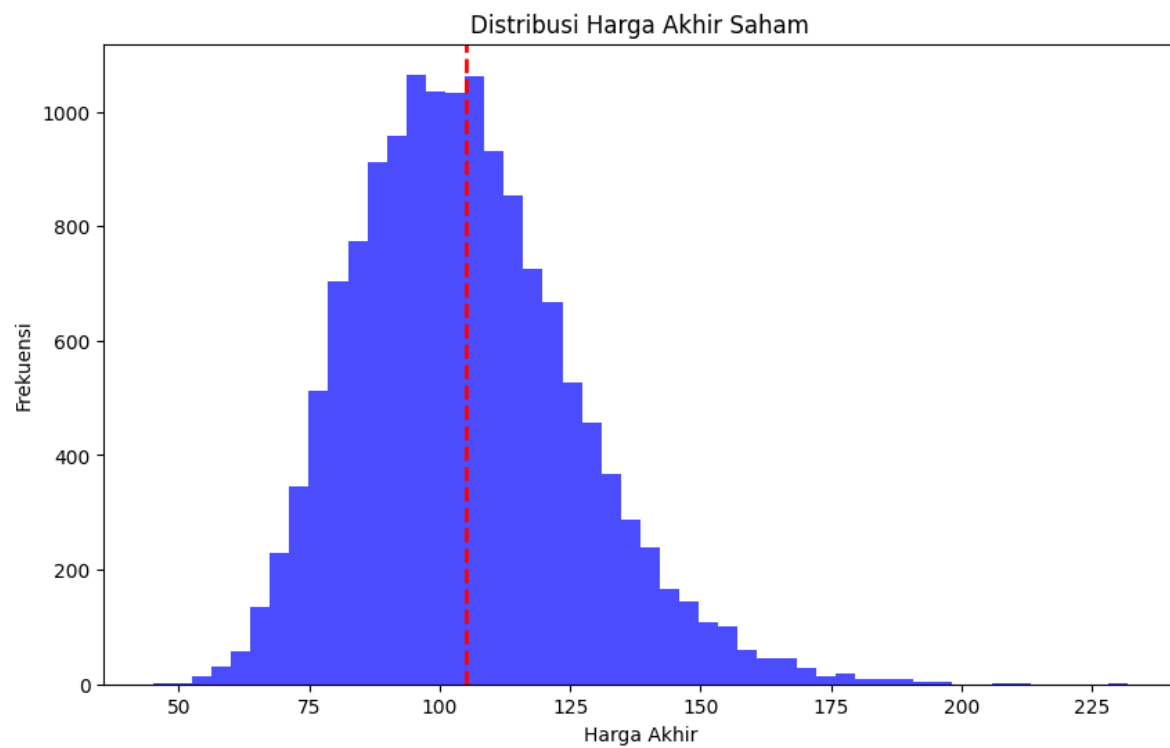
Deviasi standar harga akhir: 21.30

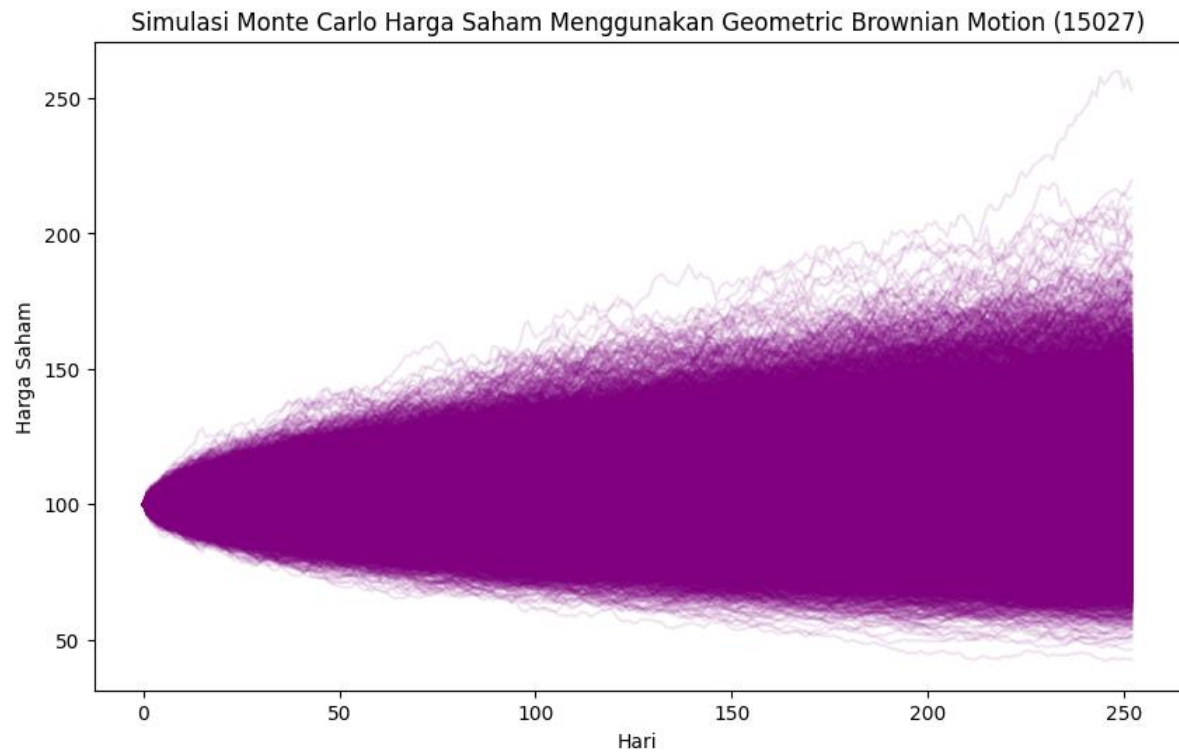




Estimasi harga akhir rata-rata: 105.00

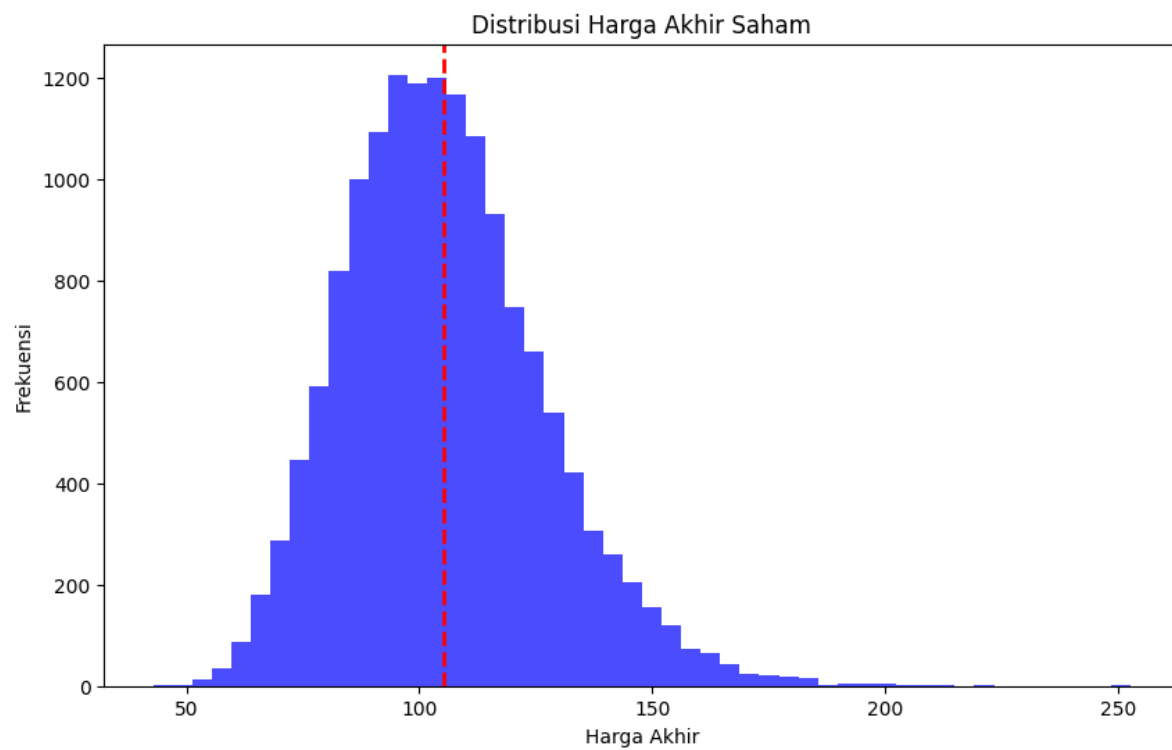
Deviasi standar harga akhir: 21.25

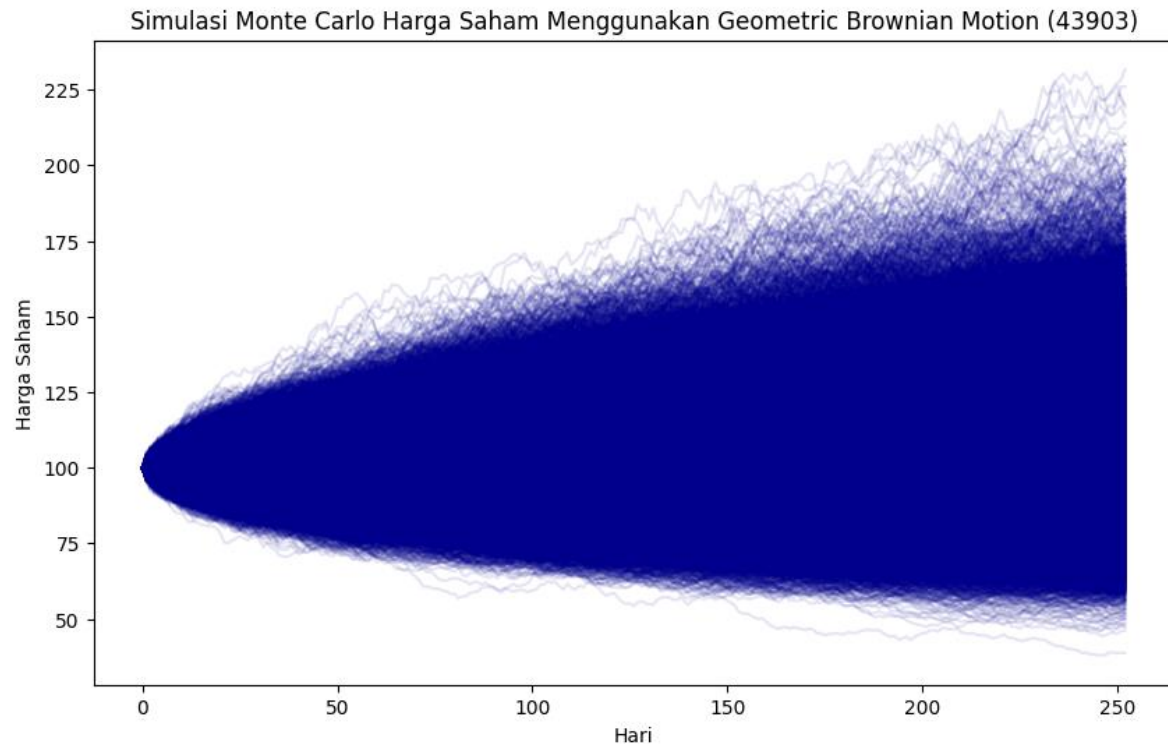




Estimasi harga akhir rata-rata: 105.35

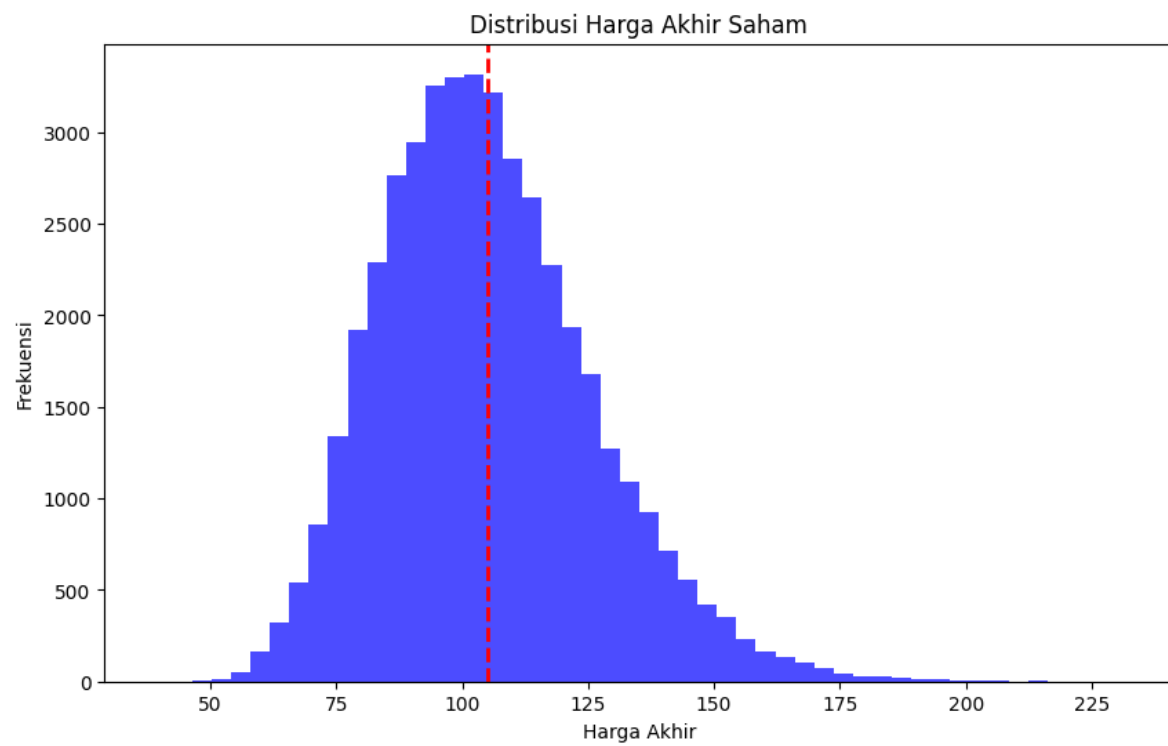
Deviasi standar harga akhir: 21.41





Estimasi harga akhir rata-rata: 104.96

Deviasi standar harga akhir: 21.19

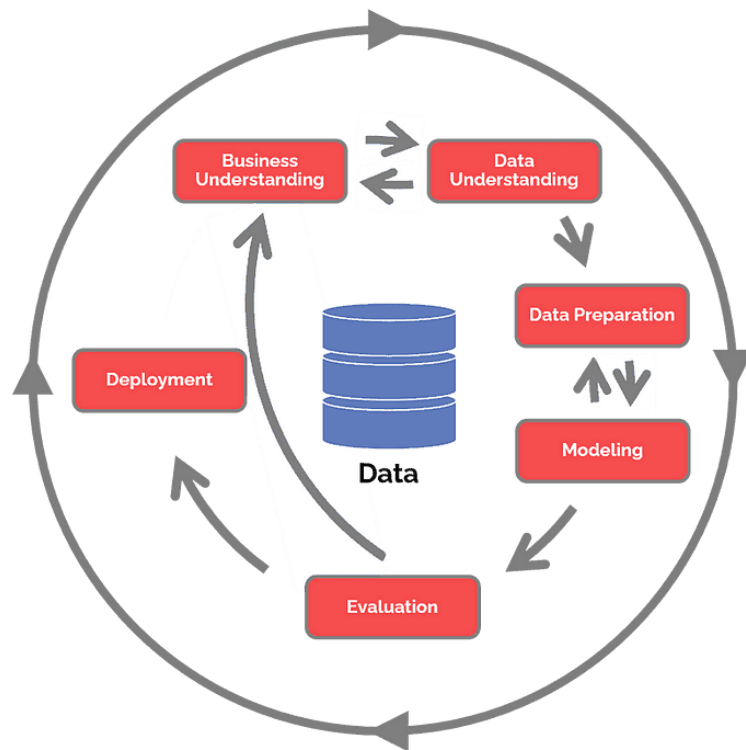


Pada plot Geometric Brownian Motion menyimulasikan jalur harga saham. Setiap jalur mewakili “masa depan” yang mungkin untuk harga saham. Dan selanjutnya terdapat analisis harga saham akhir dengan menghitung rata-rata dan standar deviasi dari harga akhir tersebut. Dan pada plot terakhir bahwa garis putus-putus merah dalam histogram mewakili harga akhir rata-rata.

2. Perbedaan Metodologi, Model, dan Algoritma dalam konteks data mining. Dan bagaimana ketiganya saling berhubungan:

Dalam konteks Data Mining, ketiga konsep ini memiliki peran yang berbeda namun saling terkait erat:

- a. Metodologi merujuk pada serangkaian langkah atau prosedur sistematis yang digunakan untuk melakukan analisis data. Ini mencakup rencana kerja yang lebih luas untuk mengelola dan menganalisis data dengan tujuan tertentu. Dalam konteks data mining, terdapat suatu metodologi bernama CRISP-DM (Cross-Industry Standards Process for Data Mining). Berikut merupakan penjelasan lebih lanjutnya:



- Understanding the business problem
Tahap pertama dalam proses CRISP-DM adalah memahami permasalahan bisnis yang ingin diselesaikan. Pada tahap ini, pertanyaan yang harus dijawab adalah mengenai apa saja yang dibutuhkan oleh bisnis atau masalah apa yang sedang dicoba diselesaikan oleh bisnis.

- Data understanding
Tahap kedua adalah memahami data yang tersedia. Pada tahap ini pertanyaan yang harus dijawab adalah mengenai apa saja data yang dimiliki dan dibutuhkan untuk menjawab pertanyaan bisnis dengan memahami sumber data, tipe data, dan kualitas data yang tersedia.
 - Data preparation
Tahap ketiga adalah mempersiapkan data untuk analisis. Pada tahap ini pertanyaan yang harus dijawab adalah bagaimana data tersebut disiapkan. Hal ini dapat dilakukan dengan menghapus data yang tidak relevan atau buruk, dan menyiapkan data yang dibutuhkan untuk analisis.
 - Modeling
Tahap keempat adalah mengembangkan model analisis. Pada tahap ini pertanyaan yang harus dijawab adalah mengenai model apa saja yang bisa diterapkan dan model analisis apa yang tepat untuk menjawab tujuan bisnis.
 - Evaluation
Tahap kelima adalah mengevaluasi kinerja model analisis dan memastikan model dapat memecahkan permasalahan bisnis yang ditargetkan.
 - Deployment
Tahap terakhir adalah mengimplementasikan hasil analisis dan model ke dalam lingkungan bisnis. Pada tahap ini perlu dipastikan hasil analisis dapat diintegrasikan dengan sistem bisnis yang ada dan model analisis dapat digunakan secara efektif.
- b. Model adalah representasi matematis atau statistik dari suatu fenomena dalam data. Ini adalah hasil dari proses analisis yang mencoba untuk menemukan pola, hubungan, atau struktur yang tersembunyi dalam data. Model dalam Data Mining dapat berupa model prediktif (misalnya, model regresi atau model klasifikasi) atau model deskriptif (seperti decision tree atau cluster). Dalam data mining, model deskriptif digunakan untuk merangkum atau mengeksplorasi data dan dapat digunakan untuk menentukan pola atau hubungan yang paling umum dalam data. Penggunaan model lain seperti model prediktif digunakan untuk memprediksi perilaku atau hasil dimasa depan berdasarkan data historis.
- c. Algoritma adalah prosedur komputasional yang digunakan untuk mengekstrak informasi dari data. Algoritma dalam Data Mining adalah instruksi langkah demi langkah yang diterapkan pada data untuk membangun model atau menemukan pola tertentu. Contoh algoritma dalam Data Mining termasuk k-means untuk clustering, Naive Bayes untuk klasifikasi, dan Apriori untuk analisis asosiasi.

Ketiga konsep ini saling terkait dalam proses analisis data sebagai berikut:

Metodologi memberikan kerangka kerja umum untuk melakukan analisis data. Ini menentukan langkah-langkah yang harus diambil, termasuk penggunaan model dan algoritma tertentu untuk mencapai tujuan analisis.

Model merupakan hasil dari penerapan metodologi dan algoritma pada data. Model mewakili penemuan atau pembelajaran yang terjadi dalam proses Data Mining. Model ini kemudian digunakan untuk membuat prediksi atau memberikan wawasan tentang data.

Algoritma adalah komponen teknis dari metodologi yang digunakan untuk menghasilkan model. Algoritma menentukan bagaimana data diproses dan diolah untuk mengidentifikasi pola atau struktur tertentu. Pemilihan algoritma yang tepat dalam metodologi sangat penting karena akan mempengaruhi kualitas dan kesuksesan analisis data.

Jadi, metodologi memberikan kerangka kerja untuk keseluruhan proses, algoritma digunakan untuk menerapkan langkah-langkah dalam metodologi, dan model adalah hasil akhir dari penerapan algoritma pada data dalam kerangka metodologi tersebut.