**MAKALAH TUGAS 5**

**“KONSEP DATA MINING”**

Makalah Ini disusun untuk Memenuhi Penilaian Mata Kuliah

Big Data Analytics dan Sebagai Hasil Pembelajaran Pribadi

**Dosen Pengampu :**

**Sevi Nurafni ST., M.Si., M.Sc**

****

**Oleh:**

**Catherine Vanya Pangemanan**

**2C2220008**

**PROGRAM STUDI S1 SAINS DATA**

**FAKULTAS SAINS DAN TEKNOLOGI**

**© Mei, 2024**

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?

Simulasi pertama sebanyak 2000. Simulasi kedua sebanyak 10.000. Simulasi ketiga sebanyak 100. Simulasi keempat sebanyak 10. Simulasi kelima sebanyak 2.

**Sintaks:** ```python

import numpy as np

import matplotlib.pyplot as plt

def simulate\_gbm(S0, mu, sigma, T, N, num\_simulations):

dt = T / N

prices = np.zeros((num\_simulations, N + 1))

prices[:, 0] = S0

for t in range(1, N + 1):

Z = np.random.standard\_normal(num\_simulations)

prices[:, t] = prices[:, t - 1] \* np.exp((mu - 0.5 \* sigma\*\*2) \* dt + sigma \* np.sqrt(dt) \* Z)

return prices

# Parameter model

S0 = 100 # Harga awal saham

mu = 0.05 # Tingkat pengembalian

sigma = 0.2 # Volatilitas

T = 1.0 # Waktu total (misalnya 1 tahun)

N = 252 # Jumlah langkah waktu (misalnya 252 hari perdagangan dalam setahun)

# Jumlah simulasi yang berbeda

simulations = [100, 50, 10, 2, 10000]

for num\_simulations in simulations:

prices = simulate\_gbm(S0, mu, sigma, T, N, num\_simulations)

plt.figure(figsize=(10, 6))

for i in range(min(num\_simulations, 10)): # Plot max 10 lines for visibility

plt.plot(prices[i])

plt.title(f'Stock Price Simulations ({num\_simulations} Simulations)')

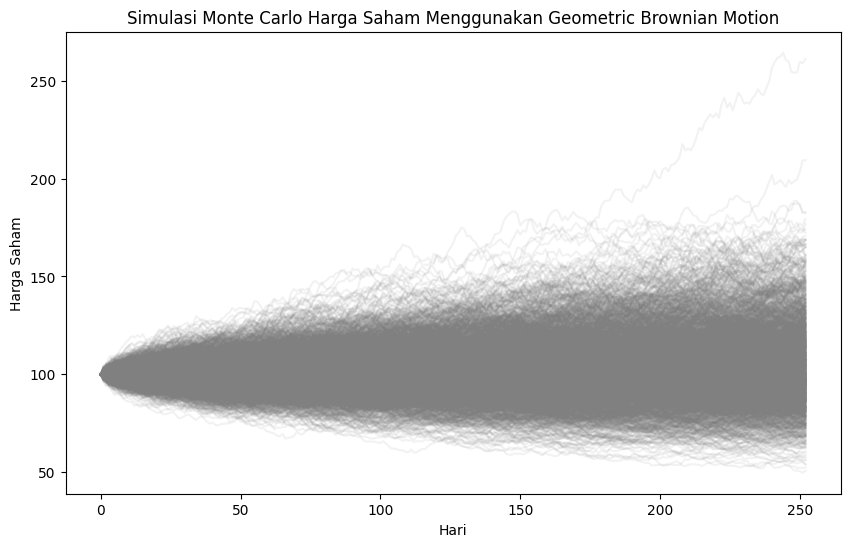
plt.xlabel('Time Steps')

plt.ylabel('Stock Price')

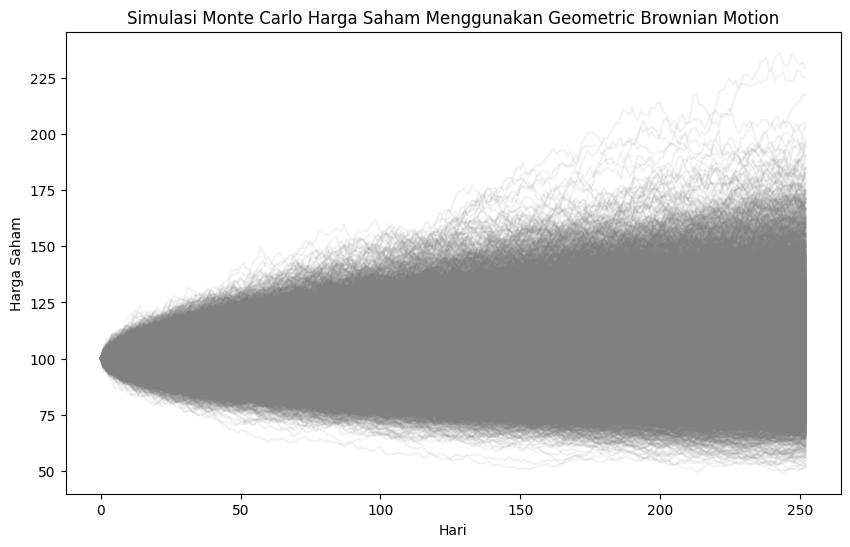
plt.show()

```

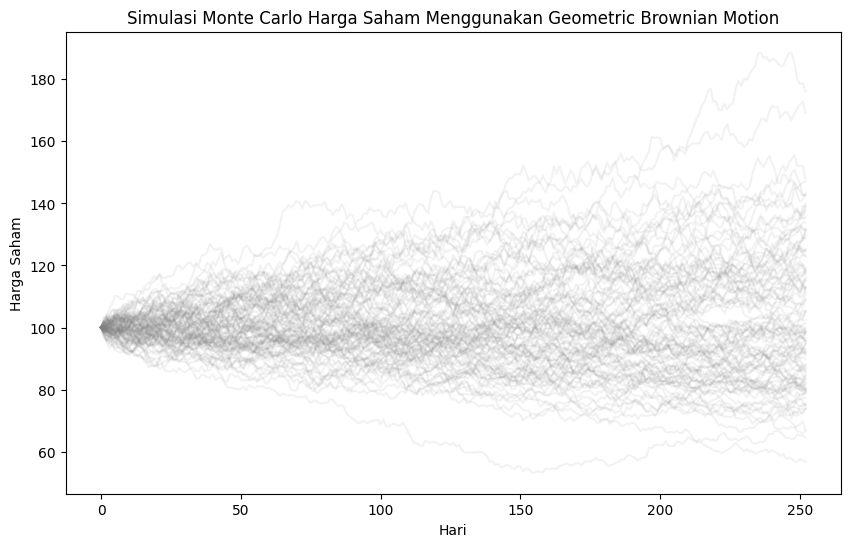
**Output**

****

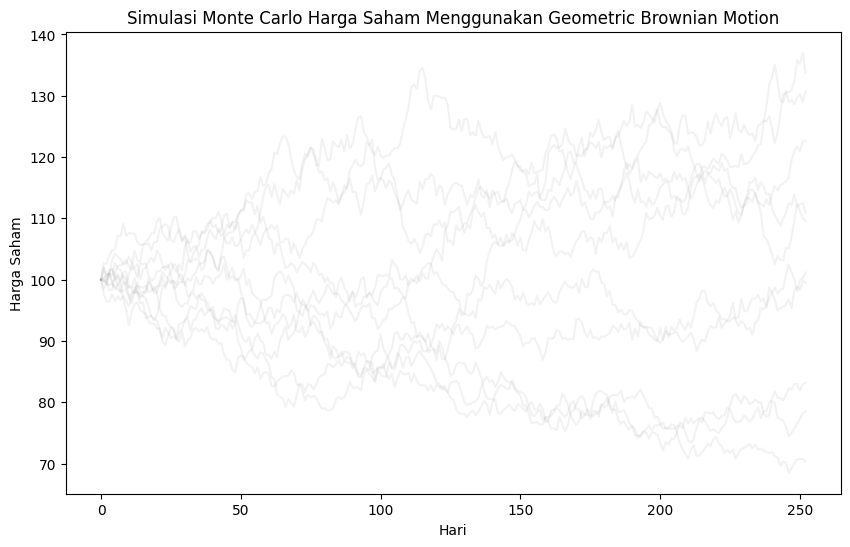
Simulasi 2.000



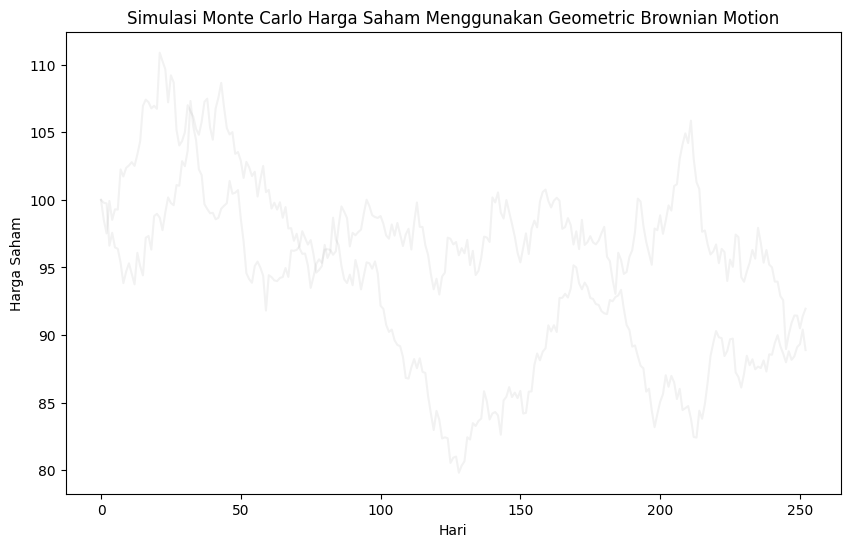
Simulasi 10.000



Simulasi 100



Simulasi 10



Simulasi 2

**Analisis Output**

Data time series harga saham yang dihasilkan dari sintaks Python untuk simulasi Monte Carlo dapat dikategorikan sebagai stochastic time series. Berikut adalah beberapa karakteristik yang menjelaskan jenis data time series ini:

1. Stochastic Nature: Data harga saham yang dihasilkan dari simulasi Monte Carlo berbasis Geometric Brownian Motion (GBM) bersifat stokastik, karena harga saham pada setiap langkah waktu ditentukan oleh proses stokastik yang melibatkan komponen acak (random walk).
2. Trend and Volatility: Data ini memiliki tren yang ditentukan oleh parameter tingkat pengembalian rata-rata (𝜇), serta volatilitas (σ) yang menentukan tingkat fluktuasi harga saham.
3. Autocorrelation: Harga saham pada suatu titik waktu sangat bergantung pada harga sebelumnya, yang merupakan karakteristik utama dari data time series.
4. Non-stationarity: Harga saham umumnya tidak stasioner, artinya statistiknya (misalnya mean, variance) dapat berubah seiring waktu.
5. Noise: Adanya komponen acak (noise) dalam model GBM menambahkan variabilitas pada data yang dihasilkan, yang sesuai dengan perilaku harga saham nyata di pasar.

Gambaran Output Sintaks:

* Jumlah Simulasi Besar (2,000 dan 10,000): Memberikan hasil yang lebih representatif dan rinci, membantu dalam analisis risiko dan prediksi yang lebih akurat.
* Jumlah Simulasi Sedang (100): Cukup untuk analisis dasar tetapi mungkin kurang lengkap.
* Jumlah Simulasi Kecil (10 dan 2): Tidak memadai untuk analisis yang dapat diandalkan, berisiko tinggi memberikan hasil yang bias dan tidak representatif.

Dari output tersebut, kita akan melihat berbagai jalur harga saham yang berfluktuasi secara stokastik, menunjukkan berbagai kemungkinan yang dapat terjadi dalam periode waktu yang ditentukan.

Kesimpulan

Data time series harga saham yang dihasilkan dari sintaks Python untuk simulasi Monte Carlo berbasis Geometric Brownian Motion dapat dikategorikan sebagai data stochastic time series yang mencerminkan karakteristik volatilitas, tren, dan komponen acak dari harga saham. Dengan meningkatkan jumlah simulasi, kita mendapatkan hasil yang lebih representatif dan konsisten dengan distribusi probabilitas teoritis. Namun, ada titik di mana peningkatan jumlah simulasi memberikan pengembalian yang semakin kecil dalam hal akurasi hasil.

B. Dalam konteks Data Mining, jelaskan perbedaan antara metodologi, model,dan algoritma. Bagaimana ketiga konsep ini saling berhubungan dalam prosesanalisis data?

1. Perbedaan antara Metodologi, Model, dan Algoritma dalam Data Mining

1.1. Metodologi

Metodologi adalah kerangka kerja atau serangkaian prosedur dan teknik yang digunakan untuk mengumpulkan, memproses, dan menganalisis data. Contoh: CRISP-DM (Cross-Industry Standard Process for Data Mining), SEMMA (Sample, Explore, Modify, Model, Assess).

1.2. Model

Model adalah representasi matematika dari fenomena yang diobservasi, digunakan untuk melakukan prediksi atau menjelaskan hubungan dalam data. Contoh: Model regresi linear, model decision tree, model neural network.

1.3. Algoritma

Algoritma adalah serangkaian langkah atau prosedur yang diimplementasikan untuk mengembangkan atau melatih model. Contoh: Algoritma K-Nearest Neighbors, algoritma Gradient Descent, algoritma Support Vector Machine.

1. Hubungan antara Ketiga Konsep

2.1. Metodologi: Memberikan panduan tentang bagaimana proses analisis data harus dilakukan, termasuk pemilihan model dan algoritma yang tepat.

2.2. Model: Adalah produk dari penerapan metodologi dan algoritma pada data, berfungsi untuk membuat prediksi atau memberikan wawasan.

2.3. Algoritma: Digunakan dalam berbagai langkah dalam metodologi untuk membangun, melatih, dan mengoptimalkan model.

Dengan demikian, metodologi, model, dan algoritma saling terkait dalam keseluruhan proses analisis data. Metodologi menetapkan kerangka kerja, algoritma mengembangkan model, dan model digunakan untuk menghasilkan wawasan atau prediksi dari data.

| Aspek | Algoritma | Metodologi | Model |
| --- | --- | --- | --- |
| Definisi | Serangkaian langkah atau prosedur untuk menyelesaikan masalah tertentu atau mengoptimalkan suatu proses | Kerangka kerja atau serangkaian prosedur dan teknik untuk mengumpulkan, memproses, dan menganalisis data. | Representasi matematis atau statistik dari fenomena yang diobservasi, digunakan untuk prediksi atau menjelaskan hubungan dalam data. |
| Contoh | K-Nearest Neighbors, Gradient Descent, Support Vector Machine. CRISP-DM (Cross-Industry Standard Process for | CRISP-DM (Cross-Industry Standard Process for Data Mining), SEMMA (Sample, Explore, Modify, Model, Assess). | Model regresi linear, decision tree, neural network. |
| Fungsi Utama | Mengembangkan dan mengoptimalkan model dengan mengolah data sesuai dengan langkah-langkah yang telah ditentukan. | Menyediakan panduan dan struktur untuk seluruh proses analisis data, termasuk pemilihan dan penggunaan algoritma serta model. | Digunakan untuk membuat prediksi, mengklasifikasikan data, atau menemukan hubungan dalam data berdasarkan input yang diberikan. |
| Penggunaan | Digunakan dalam langkah-langkah spesifik dalam proses metodologi untuk memproses data dan membangun model. | Mengarahkan keseluruhan proses data mining dari awal hingga akhir, termasuk pemilihan data, pre-processing, modelling, dan evaluasi. | Hasil akhir dari proses yang menggunakan algoritma dalam kerangka metodologi, digunakan untuk memberikan wawasan atau prediksi dari data. |
| Interaksi | Diterapkan sesuai panduan yang diberikan oleh metodologi untuk melatih atau mengoptimalkan model. | Mengarahkan pemilihan dan penerapan algoritma serta interpretasi dan evaluasi model. | Dibentuk dan dioptimalkan menggunakan algoritma dalam konteks metodologi yang dipilih, dan hasilnya dievaluasi berdasarkan metodologi. |
| Hasil | Proses atau langkah-langkah untuk mencapai tujuan spesifik seperti klasifikasi, regresi, clustering, dll. | Kerangka kerja untuk seluruh proses data mining yang memastikan kualitas dan konsistensi analisis data. | Alat prediksi atau analisis yang digunakan untuk memberikan wawasan atau melakukan keputusan berdasarkan data yang dianalisis. |

Tabel ini memberikan ringkasan tentang bagaimana algoritma, metodologi, dan model berbeda dalam definisi, fungsi, penggunaan, dan interaksinya dalam proses data mining.