

LAPORAN FINAL PROJEK

Analisis Prediksi Harga Saham NVIDIA Menggunakan Metode Linear Regression dan Evaluasi Kinerja Model Berdasarkan Error Residual dan Indikator Teknikal

Oleh: Reynald Aryansyah

LATAR BELAKANG:

Perkembangan teknologi khususnya di bidang kecerdasan buatan (AI) dan GPU telah mendorong popularitas saham NVIDIA secara signifikan. Tujuan dari analisa ini adalah untuk membangun pemahaman menyeluruh tentang perilaku harga saham NVIDIA, dengan memadukan teknik prediksi linear, evaluasi error residual, serta pendekatan teknikal melalui Moving Average (MA) dan Relative Strength Index (RSI).

DATASET DAN PEMROSESAN AWAL DAN SPEFIKASI DATASET

Dataset diambil dari sumber publik (kaggle) berisi harga historis harian saham NVIDIA, Apple, Google, Microsoft, dan Amazon selama 15 tahun. Fokus utama saya pada Close_NVDA sebagai target prediksi.

Spesifikasi Dataset

- Periode: Januari 2010 - Januari 2025
- Jumlah observasi (data) = 3774 data poin harian
- Fitur utama: Tanggal, harga buka, tinggi, rendah, tutup, volume (NVDA)
- Dataset bersifat time-series dan berskala besar yang cocok untuk eksplorasi dan prediksi

```
import pandas as pd
df = pd.read_csv(
    '/content/drive/MyDrive/Semester 4/BIG DATA /15 Years Stock Data of NVDA AAPL MSFT GOOGL and AMZN.csv')
```


EDA (EXPLORATORY DATA ANALYSIS)

EDA dilakukan untuk memahami pola tren harga, hubungan antar variabel, serta potensi sinyal teknikal sebelum pemodelan.

VISUALISASI TREN HARGA

Harga NVDA menunjukkan pola pertumbuhan eksponensial terutama sejak 2020. Lonjakan terbesar terjadi pada periode 2023-2024 yang berkorelasi dengan adopsi masif AI dan teknologi data center.



EDA STATISTIK DESKRIPTIF DAN KORELASI

```
df_nvda = df[['Open_NVDA', 'High_NVDA', 'Low_NVDA', 'Close_NVDA', 'Volume_NVDA']]
df_nvda.describe()
```

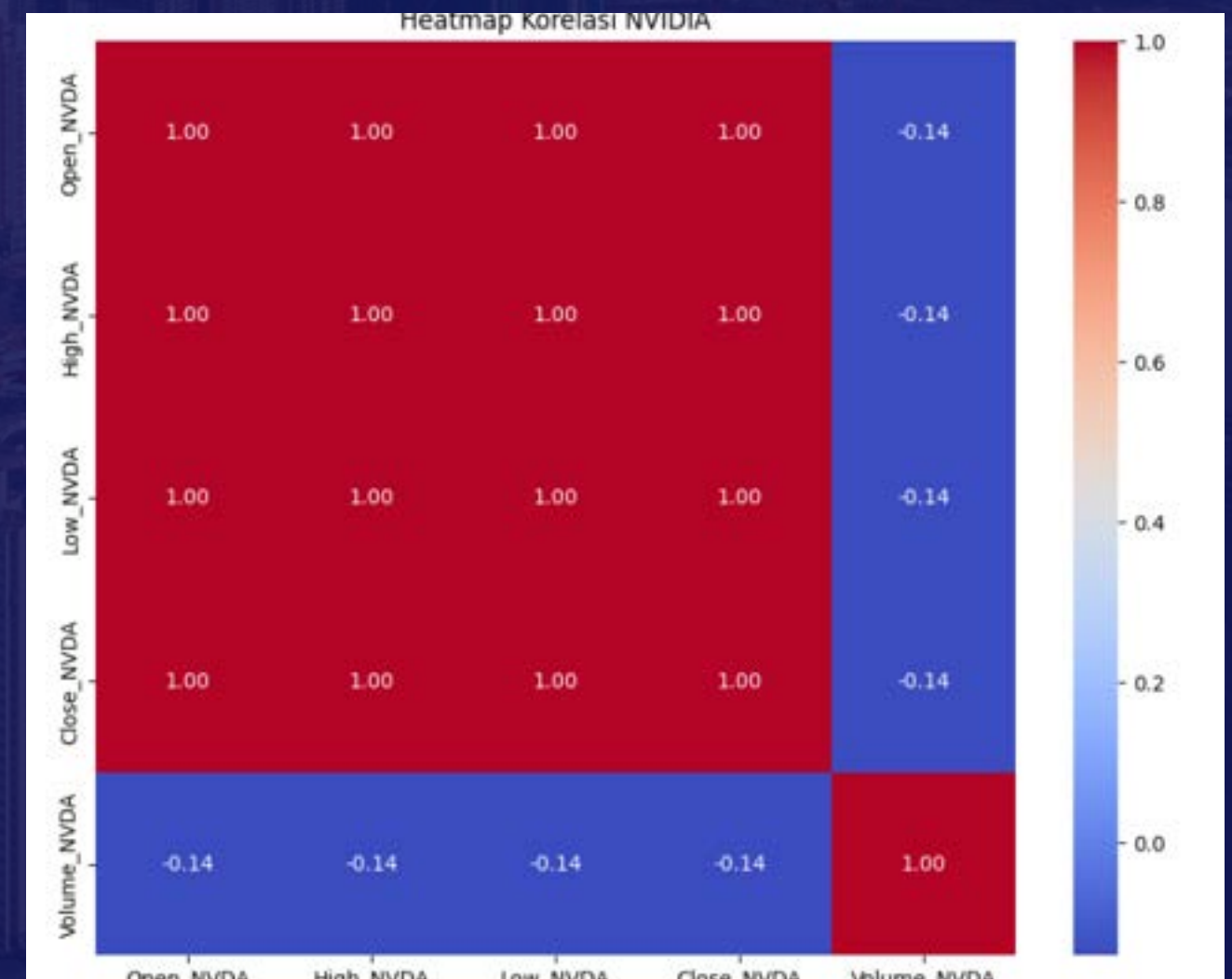
| | Open_NVDA | High_NVDA | Low_NVDA | Close_NVDA | Volume_NVDA |
|-------|-------------|-------------|-------------|-------------|--------------|
| count | 3774.000000 | 3774.000000 | 3774.000000 | 3774.000000 | 3.774000e+03 |
| mean | 13.979288 | 14.231440 | 13.701438 | 13.977600 | 4.969955e+08 |
| std | 28.117355 | 28.610675 | 27.532272 | 28.088261 | 3.064756e+08 |
| min | 0.199906 | 0.207472 | 0.198301 | 0.203574 | 4.564400e+07 |
| 25% | 0.412650 | 0.418210 | 0.403137 | 0.411332 | 3.034280e+08 |
| 50% | 3.515616 | 3.594330 | 3.449628 | 3.544295 | 4.283440e+08 |
| 75% | 13.597044 | 13.823134 | 13.332437 | 13.567127 | 6.025145e+08 |
| max | 149.325985 | 152.865409 | 146.236457 | 148.856064 | 3.692928e+09 |

Berdasarkan analisis deskriptif untuk saham NVIDIA, harga penutupan rata-rata selama 15 tahun adalah sekitar \$13.977600 dengan nilai maksimum mencapai hampir \$148.856064 Data menunjukkan adanya peningkatan signifikan dalam beberapa tahun terakhir. Volume transaksi rata-rata juga menunjukkan minat pasar yang stabil terhadap saham ini

```
selected_cols = ['Open_NVDA', 'High_NVDA', 'Low_NVDA', 'Close_NVDA', 'Volume_NVDA']
numeric_df = df[selected_cols]

# Tampilkan heatmap korelasi jika kolom mencukupi
if numeric_df.shape[1] >= 4:
    plt.figure(figsize=(10, 8))
    corr = numeric_df.corr()
    sns.heatmap(corr, annot=True, cmap='coolwarm', fmt=".2f")
    plt.title('Heatmap Korelasi NVIDIA')
    plt.show()
```

Korelasi antara Close , High , dan Low sangat tinggi (1.00), sedangkan Volume tidak memiliki pengaruh signifikan(-0,14).

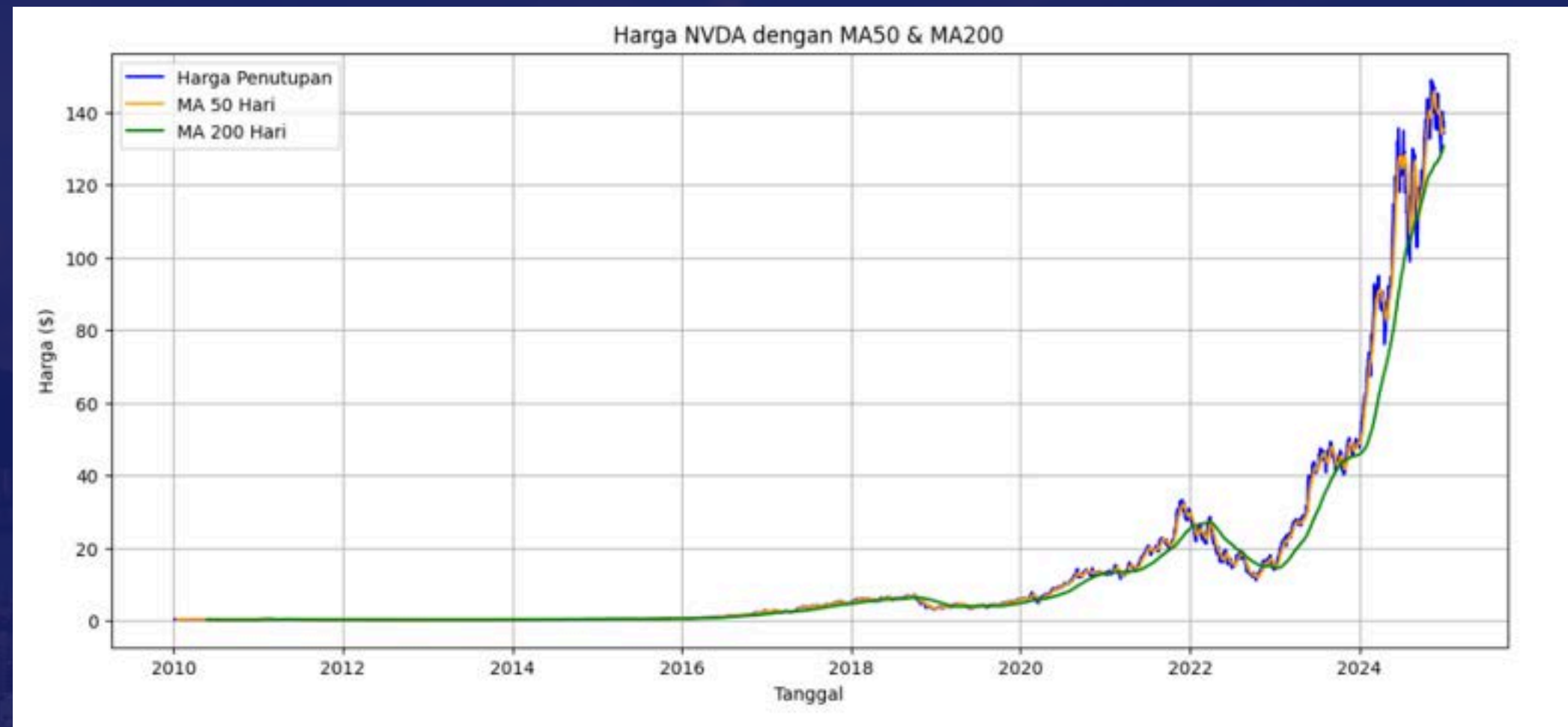


EDA MOVING AVERAGE

```
# Moving Averages
df['MA50'] = df['Close_NVDA'].rolling(window=10).mean()
df['MA200'] = df['Close_NVDA'].rolling(window=100).mean()
```

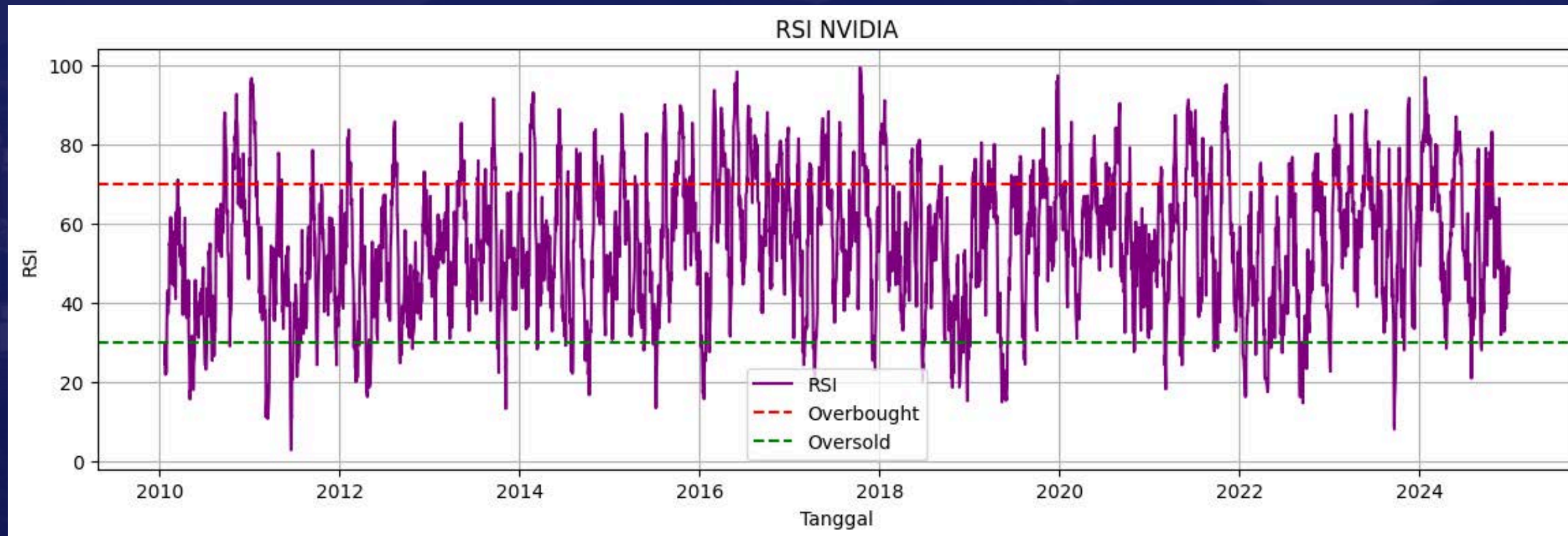
MA adalah alat yang digunakan dalam analisis teknikal untuk mengidentifikasi tren jangka panjang dan jangka pendek atau mengidentifikasi arah tren.

Berdasarkan grafik, harga saham NVIDIA cenderung bergerak di atas garis MA50 dan MA200, yang menunjukkan tren naik jangka panjang yang kuat. Terjadi momen ketika MA50 memotong MA200 dari bawah (golden cross), dan hal itu diikuti oleh kenaikan harga yang tajam. Selama periode analisis, tidak ditemukan death cross, yang berarti arah tren tidak pernah berubah ke penurunan besar. Selain itu, garis MA ini juga sering menjadi titik pantulan saat harga turun, sehingga bisa dianggap sebagai support dinamis yang dipercaya banyak pelaku pasar. Bisa dibilang NVIDIA sedang dalam fase naik daun secara teknikal.



EDA RELATIVE STRENGTH INDEX (RSI)

```
# RSI
delta = df['Close_NVDA'].diff()
gain = (delta.where(delta > 0, 0)).rolling(window=14).mean()
loss = (-delta.where(delta < 0, 0)).rolling(window=14).mean()
rs = gain / loss
df['RSI'] = 100 - (100 / (1 + rs))
```



RSI adalah indikator momentum yang mengukur kecepatan dan perubahan harga. RSI adalah indikator teknikal yang digunakan untuk mengukur kekuatan relatif tren harga dalam jangka pendek, biasanya 14 hari terakhir.

RSI menunjukkan kondisi jenuh beli saat harga berada jauh di atas MA50 dan MA200, mengindikasikan momentum kuat dalam tren naik. Namun, frekuensi RSI > 70 yang tinggi juga mengisyaratkan potensi koreksi jangka pendek meski tren jangka panjang tetap positif. Sebaliknya, ketika RSI turun di bawah 30 dan harga mendekati atau melewati MA, hal ini menandakan melemahnya tren dan potensi pembalikan arah. Dengan demikian, RSI melengkapi indikator MA dengan memberikan konteks kekuatan momentum terhadap arah tren yang sedang berlangsung.

PEMBANGUNAN MODEL LINEAR REGRESSION

```
[ ] from sklearn.model_selection import train_test_split
    from sklearn.linear_model import LinearRegression
```

```
features = ['Open_NVDA', 'High_NVDA', 'Low_NVDA']
target = 'Close_NVDA'
```

```
1 X = df[features]
  y = df[target]

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, shuffle=False)

model = LinearRegression()
model.fit(X_train, y_train)

y_pred = model.predict(X_test)

lr_model = LinearRegression()
lr_model.fit(X_train, y_train)
```

EVALUASI MODEL LINEAR REGRESSION

```
1 import numpy as np
  from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score, mean_absolute_percentage_error

print('Root Mean Squared Error (RMSE) :\t %.2f' % np.sqrt(mean_squared_error(y_pred, y_test)))
print('Mean Absolute Error (MAE)\t\t : %.2f' % mean_absolute_error(y_pred, y_test))
print('Mean Squared Error (MSE)\t\t : %.2f' % mean_squared_error(y_pred, y_test))
print('Mean Absolute Percentage Error (MAPE)\t : %.2f' % mean_absolute_percentage_error(y_test, y_pred))
print('R Square (Koefisien Determinasi)\t : %.2f' % r2_score(y_test, y_pred))
```

```
➦ Root Mean Squared Error (RMSE) :      0.71
  Mean Absolute Error (MAE)           : 0.47
  Mean Squared Error (MSE)            : 0.51
  Mean Absolute Percentage Error (MAPE) : 0.01
  R Square (Koefisien Determinasi)    : 1.00
```


PERBANDINGAN HARGA PREDIKSI DAN AKTUAL SERTA EVALUASI MODEL

```
dates_test = df.loc[y_test.index, 'Date']

y_pred = lr_model.predict(X_test)

print("Data Actual/Asli Variabel harga: \n", y_test[0:5])
print("\n")
print("Data Prediksi Variabel harga: \n", y_pred[0:5])

plt.figure(figsize=(12, 5))
plt.plot(dates_test, y_test, label='Actual', linewidth=2)
plt.plot(dates_test, y_pred, label='Predicted', linestyle='--', linewidth=2)
plt.title('Memprediksi NVIDIA dengan Fitur Historis')
plt.xlabel('Date')
plt.ylabel('Price (USD)')
plt.grid(True)
plt.legend()

# Format tanggal pada sumbu x
plt.xticks(rotation=45)
plt.tight_layout()
plt.show()
```

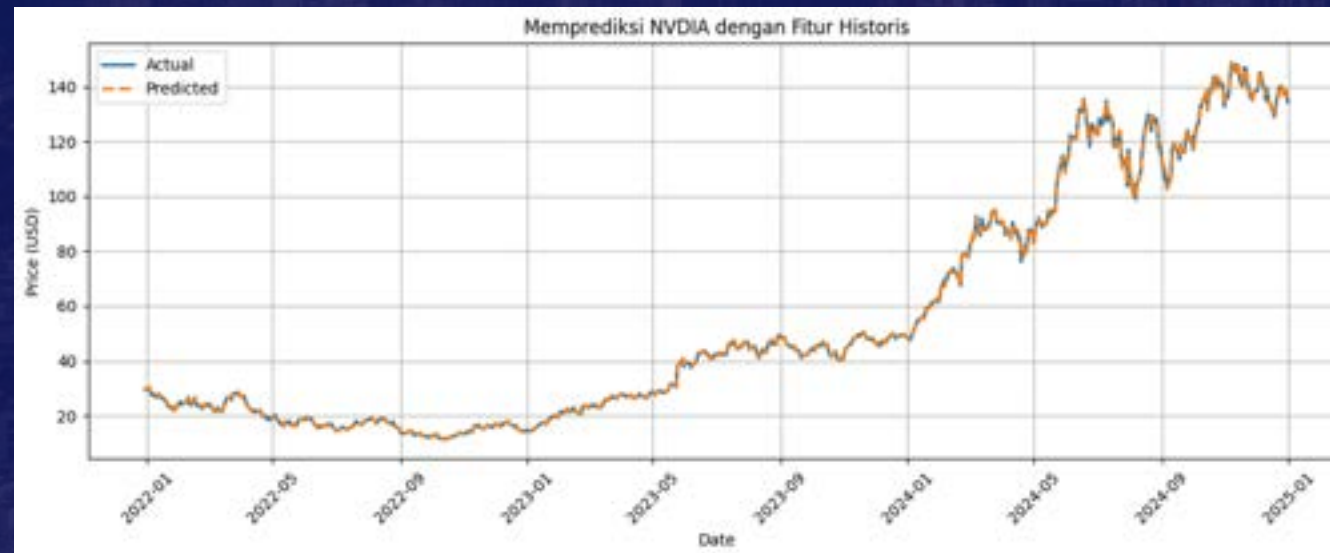
Data Actual/Asli Variabel harga:

| | |
|------|-----------|
| 3019 | 29.534159 |
| 3020 | 29.359468 |
| 3021 | 30.068226 |
| 3022 | 29.238682 |
| 3023 | 27.555637 |

Name: Close_NVDA, dtype: float64

Data Prediksi Variabel harga:

| | | | | |
|------------|------------|-------------|------------|--------------|
| [30.099236 | 29.6992564 | 30.49700842 | 28.8510981 | 28.15198629] |
|------------|------------|-------------|------------|--------------|

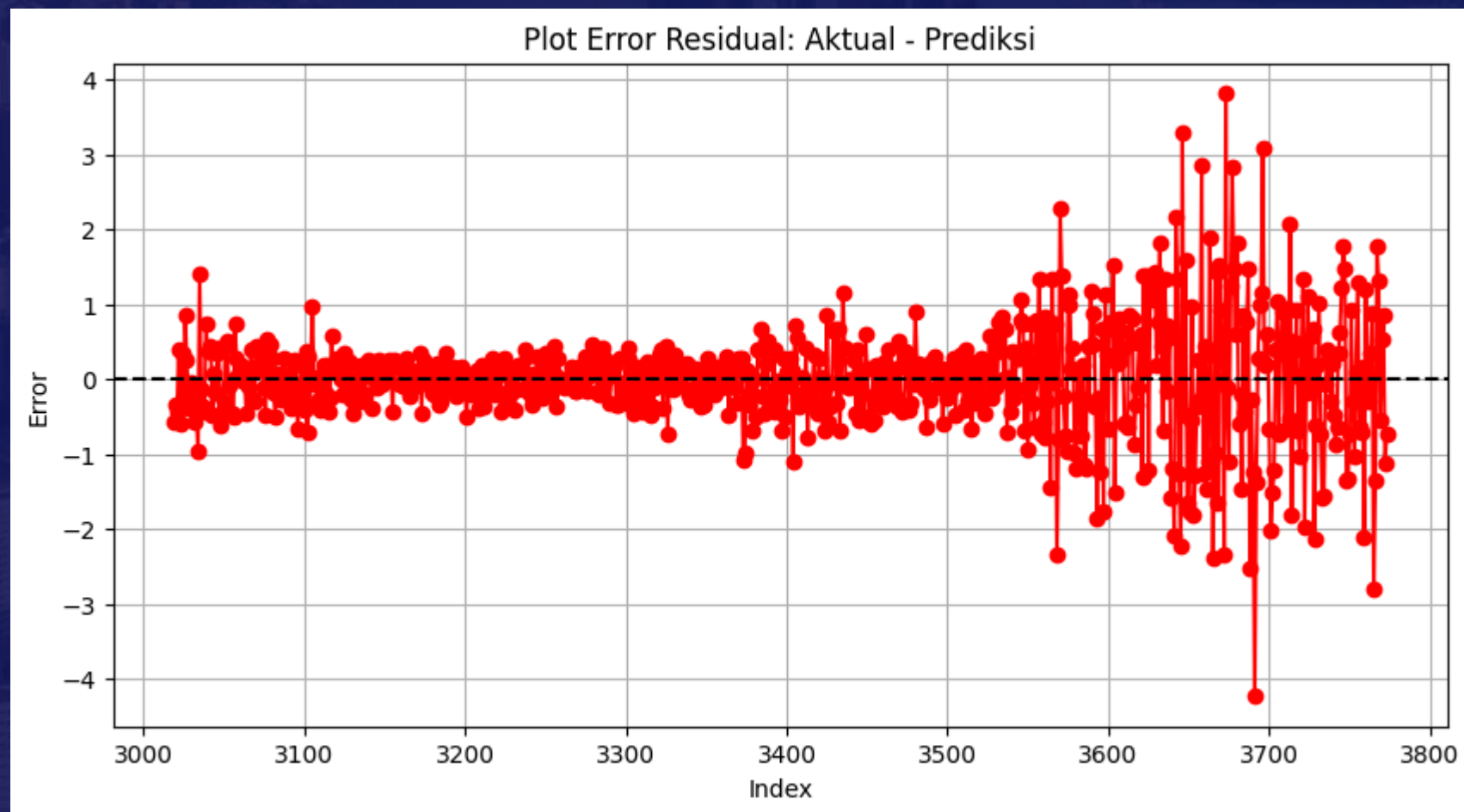


Perbandingan antara harga aktual dan hasil prediksi menunjukkan bahwa model linear regression memiliki akurasi yang cukup baik. Grafik memperlihatkan bahwa pola prediksi mengikuti pergerakan harga aktual dengan sangat rapat. Selisih antara keduanya rata-rata berada di bawah 1 dolar, yang menandakan bahwa model cukup responsif terhadap dinamika pasar dalam jangka pendek.

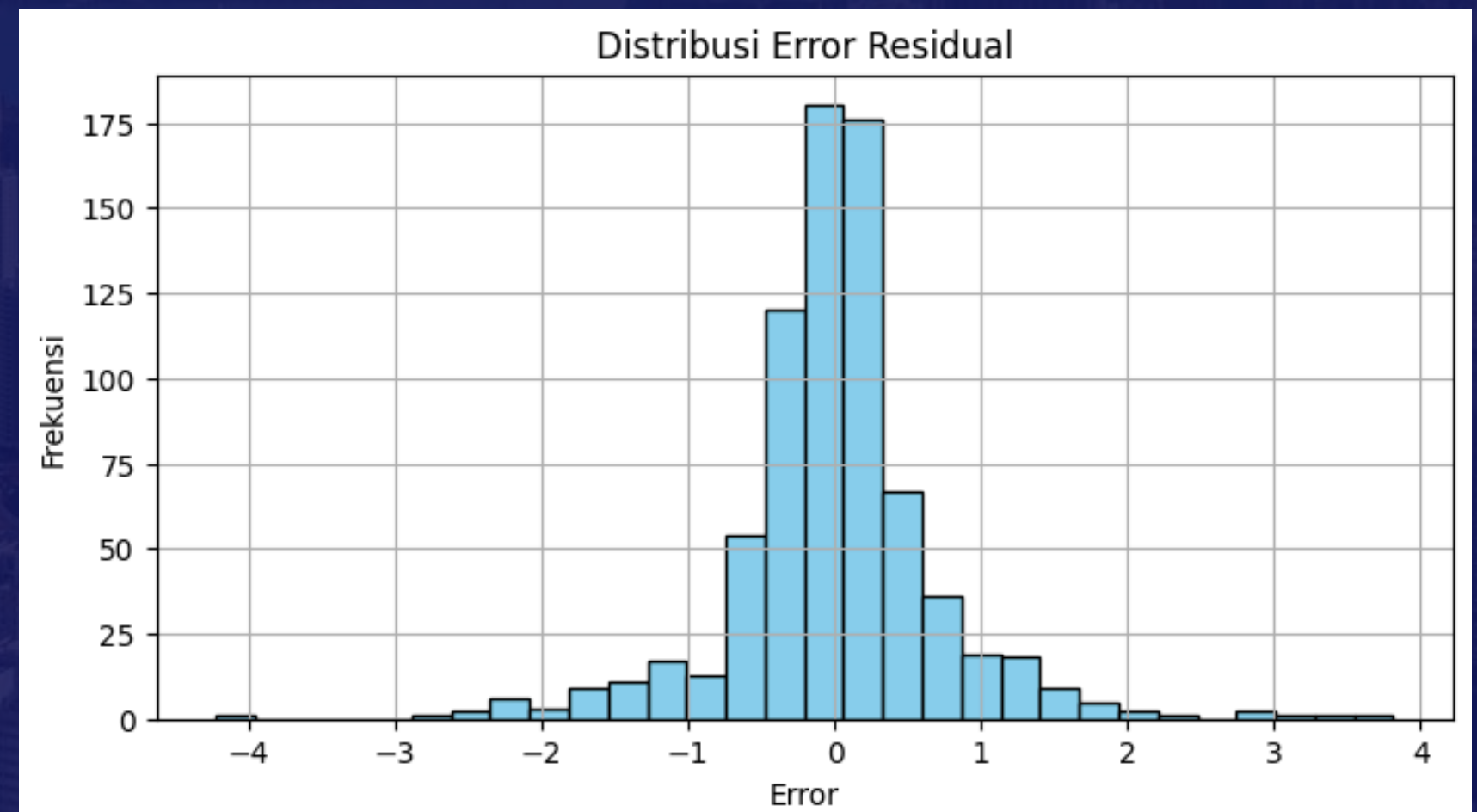
EVALUASI RESIDUAL DAN INTERPRETASI ERROR

```
# Evaluasi Error (Residual)
error = y_test - y_pred

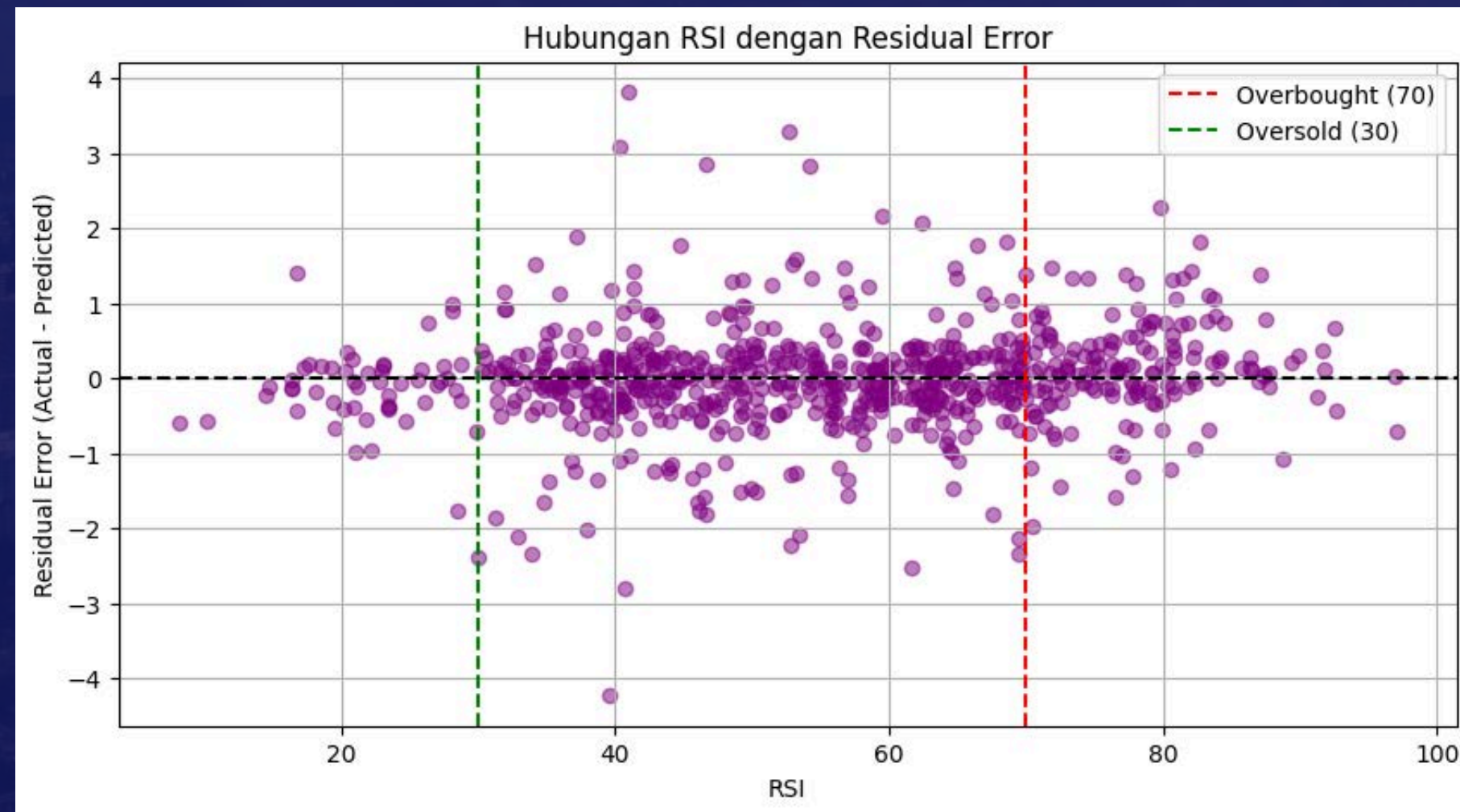
plt.figure(figsize=(10,5))
plt.plot(y_test.index, error, marker='o', linestyle='-', color='red')
plt.axhline(y=0, color='black', linestyle='--')
plt.title('Plot Error Residual: Aktual - Prediksi')
plt.xlabel('Index')
plt.ylabel('Error')
plt.grid(True)
plt.show()
```



```
plt.figure(figsize=(8,4))
plt.hist(error, bins=30, color='skyblue', edgecolor='black')
plt.title('Distribusi Error Residual')
plt.xlabel('Error')
plt.ylabel('Frekuensi')
plt.grid(True)
plt.show()
```



HUBUNGAN RSI DENGAN RESIDUAL ERROR



Berdasarkan grafik hubungan RSI dan residual error, terlihat bahwa model linear regression menghasilkan error yang lebih besar ketika RSI berada di luar rentang normal. Saat RSI tinggi (overbought), model cenderung underpredict karena tidak mampu mengikuti lonjakan harga yang cepat. Sebaliknya, saat RSI sangat rendah (oversold), model cenderung overpredict karena tidak menangkap tekanan jual yang ekstrem. Di sisi lain, akurasi model lebih baik dan stabil saat RSI berada di antara 30 hingga 70. Hal ini menunjukkan bahwa kondisi pasar yang terlalu ekstrem membuat model linier kesulitan mempertahankan presisinya.

KESIMPULAN

Berdasarkan hasil analisis, model linear regression mampu memprediksi harga saham NVIDIA dengan cukup akurat saat pasar dalam kondisi stabil, namun performanya menurun saat menghadapi pergerakan harga yang ekstrem. Indikator teknikal Moving Average (MA50 dan MA200) membantu memahami arah tren jangka panjang, sementara RSI memberikan sinyal kekuatan momentum yang juga berkaitan erat dengan tingkat error model. Ditemukan bahwa error cenderung meningkat saat RSI berada di luar zona normal (di atas 70 atau di bawah 30), menandakan bahwa pendekatan linier kurang mampu menangkap dinamika pasar yang ekstrem. Secara keseluruhan, kombinasi pendekatan statistik dan teknikal memberikan gambaran yang lebih menyeluruh dalam menganalisis dan memahami perilaku harga saham NVDA.

CLOSING STATEMENTS

Proyek ini bukan hanya tentang prediksi harga saham NVIDIA, tapi juga tentang perjalanan akademik saya yang penuh kode error, grafik naik turun, dan skrip Python yang kadang bisa ngambek. Dari sekadar belajar linear regression di kelas, saya akhirnya bisa memahami bahwa angka di pasar saham ternyata lebih lincah dari hati mahasiswa saat UTS. Melalui tugas ini, saya belajar bahwa data bukan cuma untuk dianalisis, tapi juga untuk dimengerti bahkan jika itu berarti harus bolak-balik split-train-test tengah malam. Semoga laporan ini bisa menjadi bukti bahwa kuliah tidak hanya mengajarkan teori, tapi juga bagaimana kita bertahan hidup bersama `plt.plot()` dan `mean_squared_error`.