# **IMPLEMENTASI ALGORITMA LINEAR REGRESSION UNTUK ANALISIS HUBUNGAN HARGA PEMBUKA DAN PENUTUPAN SAHAM YAHOO FINANCE**



|  |  |
| --- | --- |
| Nama | NIM |
| Amirul Al Arif | 221240001249 |
| Amirul Al Arif | 221240001249 |
| Amirul Al Arif | 221240001249 |

**UNIVERSITAS ISLAM NAHDATUL ULAMA JEPARA**

**FAKULTAS SAINS DAN TEKNOLOGI**

**PROGRAM STUDI TEKNIK INFORMATIKA**

# **BAB I PENDAHULUAN**

### 1.1 Latar Belakang

Pergerakan harga saham di pasar modal merupakan fenomena kompleks yang dipengaruhi oleh berbagai faktor, termasuk informasi fundamental perusahaan, kondisi makroekonomi, sentimen pasar, dan interaksi pelaku pasar. Harga saham pada setiap hari perdagangan biasa dicatat dalam beberapa titik penting, antara lain harga pembuka (Open) dan harga penutupan (Close). Harga pembuka merefleksikan keseimbangan awal antara penawaran dan permintaan pada pembukaan sesi perdagangan yang dipengaruhi oleh peristiwa yang terjadi di luar jam perdagangan (overnight news, laporan keuangan, keputusan kebijakan, dsb.). Harga penutupan mencerminkan keseimbangan akhir pada hari tersebut dan sering digunakan sebagai referensi untuk analisis historis dan pembuatan strategi investasi.

Analisis hubungan antara harga pembuka dan penutupan penting karena memberikan informasi, apakah kondisi awal suatu hari cenderung menentukan hasil akhir hari itu. Jika hubungan ini kuat, informasi harga pembuka saja dapat berguna untuk membuat estimasi cepat terhadap harga penutupan, yang bermanfaat untuk trading intraday, manajemen risiko, dan sistem peringatan dini. Regresi linear sederhana adalah metode statistik yang mudah diinterpretasikan untuk menilai seberapa besar pengaruh satu variabel (X) terhadap variabel lain (Y), sehingga cocok untuk menguji apakah harga Open memengaruhi Close.

Pada penelitian ini digunakan dataset harga saham harian dari Yahoo Finance periode 2018–2023. Dataset ini menyediakan observasi harian yang cukup panjang (multi-tahun) sehingga memungkinkan analisis hubungan statistik yang stabil dan pemeriksaan asumsi model. Untuk memastikan hasil yang dapat diandalkan, pendekatan yang digunakan meliputi perhitungan manual di Excel (menggunakan rumus Σ) serta implementasi komputasional di Python (Google Colab) dengan library statistik, sehingga kedua metode dapat saling memvalidasi.

### 1.2 Tujuan Penelitian

Tujuan utama penelitian ini adalah:

* Mengukur dan menguji hubungan antara harga pembuka (X) dan harga penutupan (Y) menggunakan regresi linear sederhana.
* Menghasilkan persamaan regresi Y^=a+bX\hat{Y} = a + bXY^=a+bX yang dapat dipakai untuk estimasi harga penutupan dari harga pembuka.
* Membandingkan hasil perhitungan manual (Excel) dengan implementasi di Python untuk memverifikasi konsistensi perhitungan dan mengidentifikasi potensi kesalahan prosedural.
* Memberikan interpretasi praktis terhadap koefisien regresi dan koefisien determinasi (R²) dalam konteks pasar saham.

### 1.3 Batasan dan Asumsi

* **Asumsi linearitas:** Diasumsikan hubungan antara Open dan Close mendekati linier. Bila ada non-linearitas, model sederhana akan kurang akurat.
* **Independensi error:** Diasumsikan residu bersifat independen. Namun, karena data bersifat time-series, ada kemungkinan autokorelasi residual yang perlu dicek.
* **Homoscedasticity:** Diasumsikan varians residual konstan; jika pelanggaran ditemukan, interpretasi SE/p-value dari OLS dapat menjadi tidak valid.
* **Tidak memasukkan faktor eksternal:** Analisis ini hanya menggunakan satu variabel independen; variabel penting lain (volume, berita, kondisi pasar) tidak dimasukkan dan dapat menyebabkan omitted variable bias.
* **Skala waktu:** Analisis dilakukan pada level harga (level series) bukan return; harga level biasanya non-stasioner, sehingga prediksi jangka panjang menggunakan level kurang disarankan tanpa de-trending atau differencing.

# **BAB II METODELOGI PENELITIAN DAN IMPLEMENTASI**

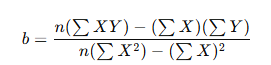
2.1 Data Set

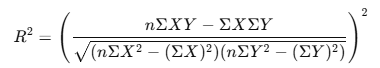
* Sumber Data: Dataset yang digunakan diambil dari Yahoo Finance dengan rentang waktu 2018–2023 dan berisi data harga saham harian. Kolom utama yang digunakan adalah 'Open' (harga pembuka) dan 'Close' (harga penutupan).
* Deskripsi Data: Data Yahoo Finance terdapat 1258 baris dan kolom 7. dimana terdapat Date, Open, Hight, Low, Close, Adj Close, dan Volume. Variabel Open menjadi variabel bebas (X) dan variabel Close menjadi variabel terikat (Y).
* Tools/Library: Penghitungan linier regresi pada data yah finance menggunkan 2 tools yaitu : exel dan python dengan google colab

# **BAB II HASIL DAN PEMBAHASAN**

Regresi linear sederhana merupakan model matematis yang menjelaskan hubungan antara satu variabel bebas (X) dan satu variabel terikat (Y). Model regresi linear dituliskan sebagai:  
  
Y = a + bX

Keterangan:  
• Y = variabel terikat (harga penutupan)  
• X = variabel bebas (harga pembuka)  
• a = konstanta (intercept)  
• b = koefisien regresi (slope)  
  
Koefisien regresi dihitung dengan rumus:  
b = (nΣXY - ΣXΣY) / (nΣX² - (ΣX)²)

  
a = (ΣY - bΣX) / n

  
  
Koefisien determinasi (R²) digunakan untuk mengetahui seberapa besar pengaruh variabel X terhadap variabel Y, dengan rumus:  


Dataset yang digunakan diambil dari Yahoo Finance dengan rentang waktu 2018–2023 dan berisi data harga saham harian. Kolom utama yang digunakan adalah 'Open' (harga pembuka) dan 'Close' (harga penutupan). Implementasi dilakukan menggunakan dua pendekatan berikut:

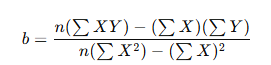
1. Perhitungan Manual di Excel

* Menggunakan rumus ΣX, ΣY, ΣXY, ΣX² (=SUM) , bonus n (=ROWS(B2:B1259))

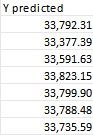


* Menghitung a(intercept) dan b(slope).



 = 

* Menghasilkan persamaan regresi: Y = 0.9979X + 64.56



* Nilai r (koefisien korelasi) = 0.997591895 menunjukkan hubungan yang sangat kuat dan positif.



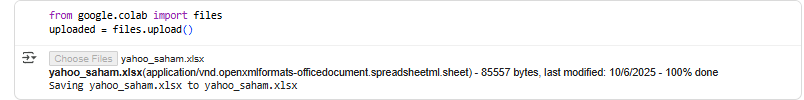
* Nilai R²(Koefisien Determinasi)= 0.995189589, menunjukan 99% hubungan nya sangat sangat berkaiatan.

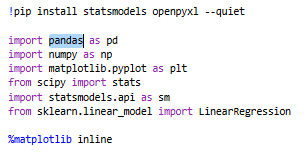
* Berikut adalah tampilan **Scatter Chart** hasil perhitungan di Excel:

2. Implementasi Menggunakan Python di Google Colab

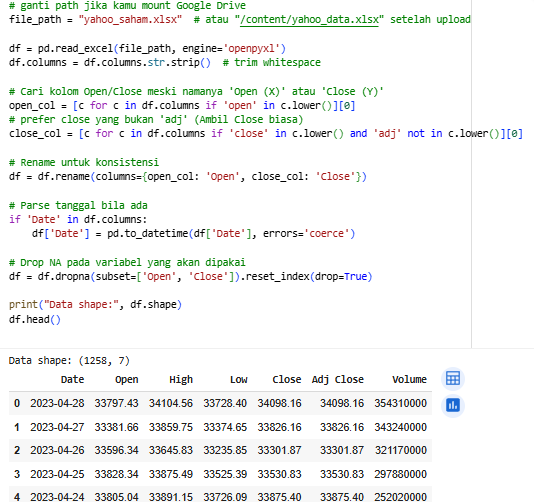
* Mengimport file exel (xlsx) ke dalam Google Colab-



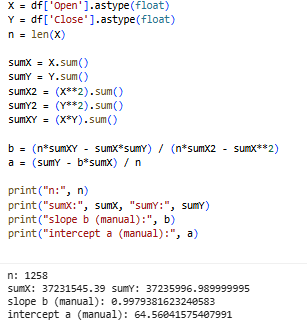
* Instal & import paket yang diperlukan



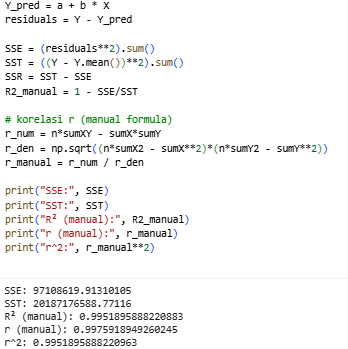
* Baca file dan persiapan data (robust terhadap nama kolom)



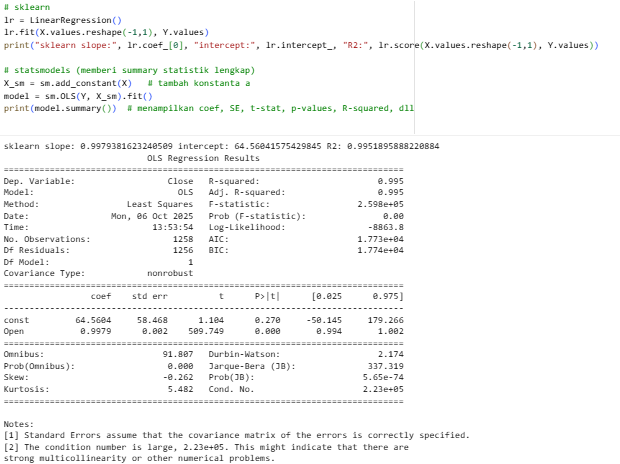
* Perhitungan regresi secara **manual**



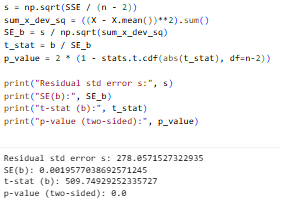
* Buat prediksi, residual, SSE, SST, R²



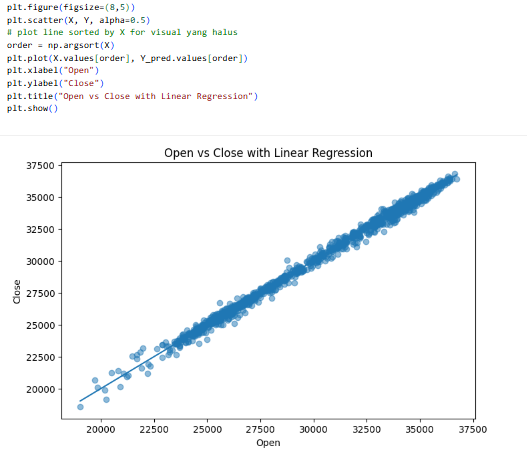
* Verifikasi dengan sklearn (LinearRegression) dan statsmodels (OLS)

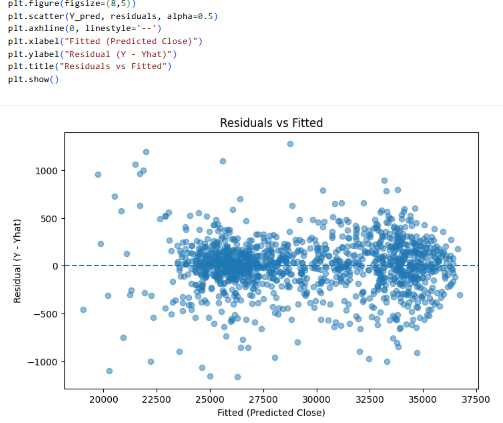


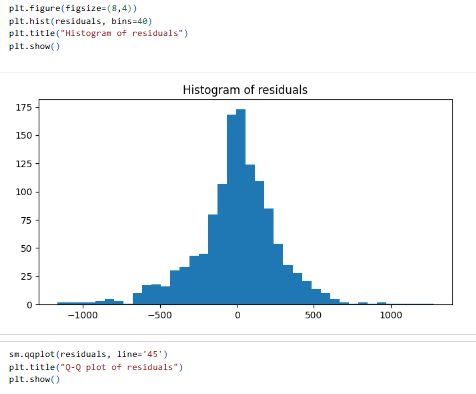
* Hitung standard error slope, t-stat, p-value secara manual

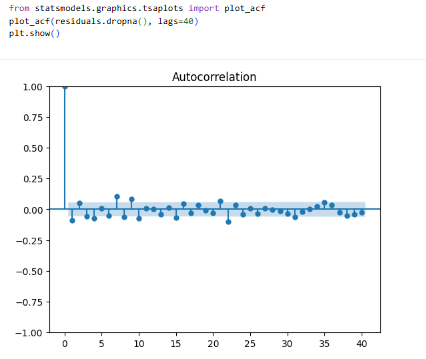


* Visualisasi (scatter + garis regresi, residual plot, histogram residual, Q-Q)

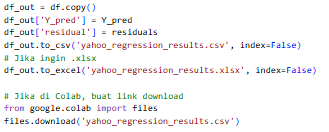








* Simpan hasil & unduh file



Hasil regresi linear dari kedua pendekatan menunjukkan bahwa harga pembuka (Open) memiliki pengaruh positif dan sangat kuat terhadap harga penutupan (Close). Nilai koefisien determinasi R² = 0.995 berarti 99,5% variasi harga penutupan dapat dijelaskan oleh harga pembuka. Koefisien korelasi r = 0.997 menunjukkan hubungan yang hampir sempurna.

Perbandingan hasil manual dan hasil komputasional menunjukkan nilai yang identik:  
• Slope (b) = 0.9979  
• Intercept (a) = 64.56  
• R² = 0.995  
  
Dengan demikian, model regresi linear sederhana dapat digunakan untuk memprediksi harga penutupan berdasarkan harga pembuka saham.

# **BAB V KESIMPULAN DAN SARAN**

Berdasarkan hasil implementasi dan analisis,semua tujuan penelitian telah tercapai. Analisis menunjukkan adanya hubungan linear positif yang signifikan antara harga pembuka (Open) dan harga penutupan (Close). Persamaan regresi yang dihitung (baik secara manual di Excel maupun menggunakan Python/statsmodels) adalah: Y^=64,56+0,99794 X\hat{Y} = 64{,}56 + 0{,}99794\,XY^=64,56+0,99794X. Koefisien determinasi R2≈0,995R^2 \approx 0{,}995R2≈0,995 menunjukkan bahwa sekitar 99,5% variasi harga penutupan dapat dijelaskan oleh harga pembuka pada data ini. Uji statistik pada koefisien slope menghasilkan nilai t sangat besar dan p-value ≈ 0, sehingga hubungan bersifat signifikan secara statistik. Validasi silang antara perhitungan manual dan hasil komputasi mengonfirmasi konsistensi numerik (tidak ada kesalahan prosedural). Namun, perlu dicatat bahwa R² yang sangat tinggi sebagian disebabkan oleh sifat data (Open dan Close pada hari yang sama) ,sehingga meskipun model menjelaskan variasi dengan baik, penggunaan model ini untuk prediksi jangka panjang memerlukan kehati-hatian dan langkah analitis lanjutan (mis. penggunaan returns, model ARIMA, atau termasuk variabel tambahan).

Saran untuk penelitian lanjutan adalah memperluas model dengan menambahkan variabel lain seperti Volume, High, dan Low, serta menggunakan model regresi berganda atau algoritma machine learning lainnya seperti Random Forest atau ARIMA untuk prediksi yang lebih kompleks.