

Analisis Harga Bitcoin Berbasis Regresi Linier

Rian Kenji¹

¹ Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara,
Jln. Letjen S. Parman No. 1, Jakarta, 11440, Indonesia
E-mail: rian.825210099@stu.untar.ac.id

Abstrak

Perkembangan teknologi sekarang ini yang makin canggih menyebabkan berbagai bidang mengalami perobakan salah satunya ada dalam bidang investasi lebih tepatnya crypto. Trader dan investor menggunakan banyak cara untuk memperdagangkan Bitcoin, aset digital di dunia kripto. Indodax adalah platform lokal Indonesia yang memungkinkan perdagangan aset digital. Analisis statistik dilakukan untuk memprediksi pergerakan harga bitcoin, namun volatilitas yang tinggi membuat sulit untuk memprediksi pergerakan bitcoin. Menggunakan analisis statistik dan pembelajaran mesin adalah salah satu cara untuk memprediksi bitcoin .. metode yang digunakan dalam menganalisis data adalah regresi linier dengan data bitcoin dan saham gabungan IHSG, yang diambil dari website yahoo finance. Datapun diolah dan dianalisis menggunakan Bahasa pemrograman python dan dilakukan dalam google colab notebook. Hasil yang didapat dari analisis adalah prediksi bitcoin dan bisa melihat visualisasi data dalam bentuk line plot harga bitcoin dari tahun 2020-2023. Dianalisis juga dalam bentuk regresi linier data saham IHSG dan BTC dan divisualisasikan dalam scatter plot.

Kata kunci—analisis statistik, Bitcoin, regresi linier

Abstract

Numerous industries, including the one of investing in digital assets, particularly cryptocurrency, are changing as a result of technological advancements that are happening more quickly. One of the digital assets in the cryptocurrency world, Bitcoin, is traded in a variety of methods by investors and traders. An Indonesian-made platform called Indodax facilitates trades in digital assets. Although bitcoin price movements are predicted using technical and fundamental analysis, excessive volatility makes this task challenging. One technique to anticipate the future of bitcoin is to employ statistical analysis and machine learning. The Python programming language was used to process and analyze the data in the Google Colab notebook. The analysis's findings include bitcoin forecasts and For the price of bitcoin from 2020 to 2023, you may view data visualization in the form of line charts. Data from the JCI and BTC stock prices were also evaluated using linear regression and presented as a scatter plot.

Keywords—statistic analysis, bitcoin, linear regression

1. PENDAHULUAN

Latar Belakang

Di masa digital sekarang ini teknologi mengalami perkembangan yang sangat pesat dan peranannya penting dalam mendukung aktivitas manusia supaya bisa memanfaatkan waktu dengan lebih baik [7]. Machine Learning merupakan contoh teknologi yang sangat membantu manusia. ML (Machine Learning) adalah cabang dari ilmu Kecerdasan Buatan (Artificial Intelligence) yang bertumbuh dengan cepat dan sudah menghasilkan manfaat berupa penanganan

masalah di berbagai bidang menjadi lebih efisien seperti masalah klasifikasi, regresi, klastering, dan anomaly detection [1].

Meliput dari data Coinmarketcap.com yang dibuka pada tanggal 3 Juli 2023, Bitcoin mencatat marketcap 603.03B dollar. Hal inilah yang menetapkan Bitcoin sebagai pemegang marketcap crypto terbesar didunia. Adapun yang membuat harga naik dan turunnya crypto salah satunya bitcoin yaitu ada 7, yaitu ketersediaan pasokan dan permintaan, adopsi bitcoin, berita dan komentar tokoh ternama, ketidakpastian nilai bitcoin di masa depan, regulasi pemerintah, karakter investor dan tingkat penjualan bitcoin di pasar.

Harga bitcoin naik tajam pada tahun 2016 dan turun pada tahun 2018 [2]. Bitcoin dapat diperdagangkan kapan saja disebabkan tidak adanya batas waktu untuk melakukan penjualan yang berbeda dengan pasar tradisional dimana produk fisik dijual dan ada waktu kadaluarsanya. Cryptocurrency berfluktuasi dengan sangat cepat karena tidak memiliki nilai tetap seperti mata uang biasa. Jadi, investor tidak ingin menderita kerugian karena mereka tidak dapat melihat perkembangan nilai bitcoin yang benar, menjual pada waktu yang salah, atau dipendam investor bahkan pada saat yang tepat untuk menjual [3].

Berdasarkan penelitian literatur, untuk menghindari kerugian investasi pada cryptocurrency bisa dilakukan dengan melakukan prediksi harga cryptocurrency menggunakan metode komputasi yaitu machine learning [4][5].

Makalah penelitian ini bermaksud untuk mencoba memprediksi arah harga Bitcoin melalui penggunaan metode analisis data dalam data mining. Ada beberapa upaya untuk memprediksi arah harga Bitcoin tetapi dengan berbagai akurasi dan metodologi. Sebagai contoh, penelitian oleh [6] mencapai akurasi 52%, hal ini dapat terjadi karena kurangnya eksplorasi opsi algoritma lainnya. Karena itu, keakuratan model mungkin masih dianggap rendah dibandingkan dengan pasar perdagangan tradisional. Penelitian ini mencoba untuk menguji menggunakan model berbasis regresi karena kebutuhan untuk memprediksi nilai numerik kontinu yang merupakan harga Bitcoin (BTC) terhadap Dolar Amerika Serikat (USD). algoritma yang akan diteliti seperti Regresi Linier (LR).

Permasalahan

1. Bagaimana menganalisis data harga bitcoin 2022 - 2023 dengan analisis statistik

Tujuan

1. Menganalisis data harga bitcoin 2020 - 2023 dengan memanfaatkan analisis statistik

2. METODE PENELITIAN

2.1. Pengumpulan data:

Pengumpulan Data merupakan hal pertama yang diambil untuk memulai proyek apa pun. Ini didefinisikan sebagai prosedur pengumpulan, pengukuran, dan analisa informasi yang akurat untuk penelitian menggunakan standar teknik validasi [10]. Seorang analis kemudian akan dapat

menilai teori mereka berdasarkan pengumpulan informasi. Pada umumnya, pengumpulan informasi adalah kemajuan yang penting dan paling signifikan untuk penelitian. Metodologi bermacam-macam informasi adalah beragam untuk berbagai bidang studi, bergantung pada data yang diperlukan. Tujuan yang paling penting pengumpulan data memastikan bahwa informasi yang dikumpulkan kaya akan konten dan dapat dimanfaatkan untuk statistic analisis sehingga keputusan dapat dibuat secara efisien dan efektif. Kumpulan data berisi hari transaksi dari 15 Juni 2020 hingga 23 Juni 2023. Dataset ini sendiri diambil dari website yahoo finance dengan judul file BTC-USD dan sampel data yang ada sendiri mencapai 1095 baris.

	Date	Open	High	Low	Close	Adj Close	Volume
0	2020-06-15	9386.035156	9504.860352	8990.175781	9450.702148	9450.702148	26699704768
1	2020-06-16	9454.266602	9579.430664	9400.445313	9538.024414	9538.024414	21565537209
2	2020-06-17	9533.784180	9540.422852	9327.339844	9480.254883	9480.254883	20177709879
3	2020-06-18	9481.567383	9482.782227	9328.395508	9411.840820	9411.840820	17770083003
4	2020-06-19	9410.293945	9440.875977	9274.295898	9288.018555	9288.018555	19632223107

Gambar 1. Data Harga Bitcoin 2020-2023

2.2. Pra-pemrosesan data :

Data yang sudah kita dapat diproses terlebih dahulu dengan python melalui google colab, untuk menghilangkan missing values yang ada untuk menjamin data yang lebih akurat dan mendapatkan hasil yang lebih baik.

```
df.isna().sum()/len(df)*100
```

Date	0.0
Open	0.0
High	0.0
Low	0.0
Close	0.0
Adj Close	0.0
Volume	0.0
dtype:	float64

Gambar 2. Kode Untuk Cek Missing Values

Gambar diatas merupakan kode pada python untuk mengecek missing values dalam suatu file, dan gambar dibawah adalah cara untuk menangani missing values yaitu dengan mengisi value yang null atau kosong dengan value baris sebelumnya atau sesudahnya. bfill() merupakan singkatan dari backward fill yang berguna untuk mengisi value yang kosong dengan value data sebelumnya, kebalikannya dengan ffill() yang berarti forward fill yang dimana fungsi ini akan mengisi value yang kosong dengan data setelahnya.

```
#menangani missing values  
df = df.ffill()  
df = df.bfill()
```

Gambar 4. Kode Menangani Missing Values

2.3. Analisis Data


Analisis Data dilakukan secara kuantitatif yang berarti data akan ditampilkan dalam bentuk angka atau numerik dan berfokus pada statistic , matematik dan numerik dari suatu kumpulan data dengan pengukuran data yang berbentuk angka

Regresi Linier digunakan untuk mengidentifikasi hubungan antara dependen dan independen variabel dan leveraged untuk memprediksi hasil masa depan. Ketika hanya menggunakan satu dependen dan satu variabel independen maka disebut regresi linier sederhana [8]. Ketika jumlah independen dan dependen bertambah, maka disebut sebagai regresi berganda. Grafik diplot menggunakan garis lurus melintasi grafik scatter[9].

Tujuan umum analisis regresi linier sederhana adalah untuk mengetahui pengaruh variabel bebas terhadap variabel terikat. Rumus regresi linier sederhana mengisi formulir persamaan umum: $Y = \beta_0 + \beta_1 X$. Ketika Y adalah variabel dependen, X adalah variabel independen dan β_0 dan β_1 adalah koefisien regresi. Rumus untuk menentukan nilai β_0 dan β_1 menurut persamaan berikut. .

R u m u s H i t u n g

Persamaan Regresi Linear Sederhana

$$Y = \beta_0 + \beta_1 X$$


$$\beta_0 = \frac{\sum y \cdot \sum x^2 - \sum x \cdot \sum xy}{n \cdot \sum x^2 - (\sum x)^2}$$
$$\beta_1 = \frac{n \cdot \sum xy - \sum x \cdot \sum y}{n \cdot \sum x^2 - (\sum x)^2}$$

Gambar 5. Rumus Regresi Linier

3. HASIL DAN PEMBAHASAN

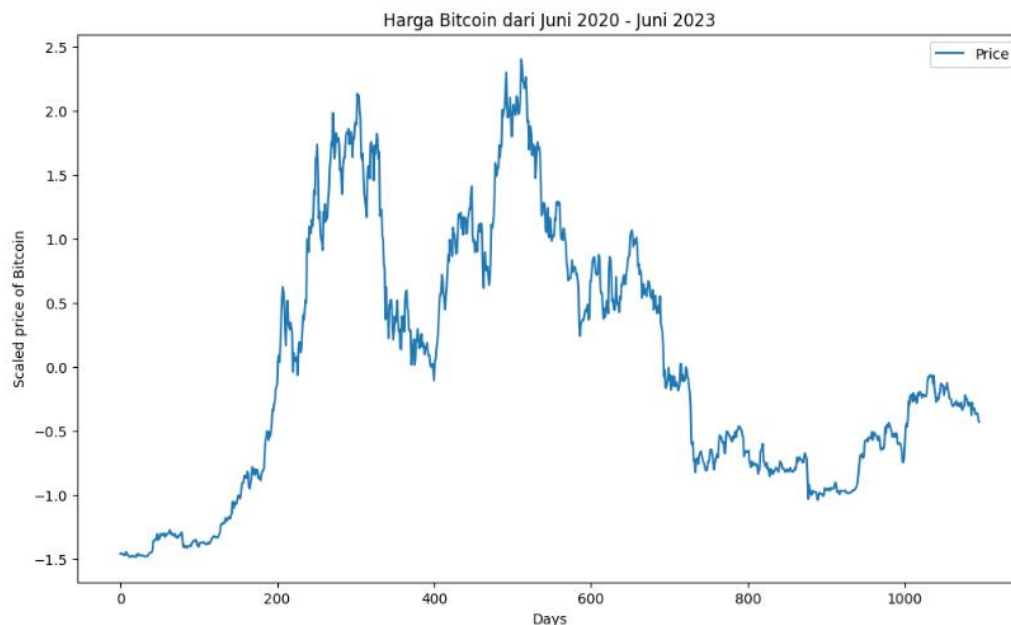
Analisis data proses merupakan penjabaran data yang masuk ke dalam sistem dan data bisa menjadi bahan untuk diproses. Data yang sudah di proses pada tahap preprocessing dataset membuat data menjadi bersih dan dapat digunakan untuk analisis kuantitatif dimana metode tersebut merupakan salah satu metode yang dapat digunakan untuk dalam menganalisis harga crypto. Proses regresi diawali dengan pembagian dataset menjadi dua bagian, yaitu training data dan testing data

```
import numpy as np
import pandas as pd
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score
from sklearn.linear_model import LinearRegression
import matplotlib.cm as cm
import matplotlib.pyplot as plt
from scipy.spatial import distance
from sklearn import preprocessing
import time
import seaborn as sns

[4] #Import library untuk Linear Regression
    from sklearn.linear_model import LinearRegression
    #Import library numpy untuk pemodelan input array dari Linear Regression
    import numpy as np
```

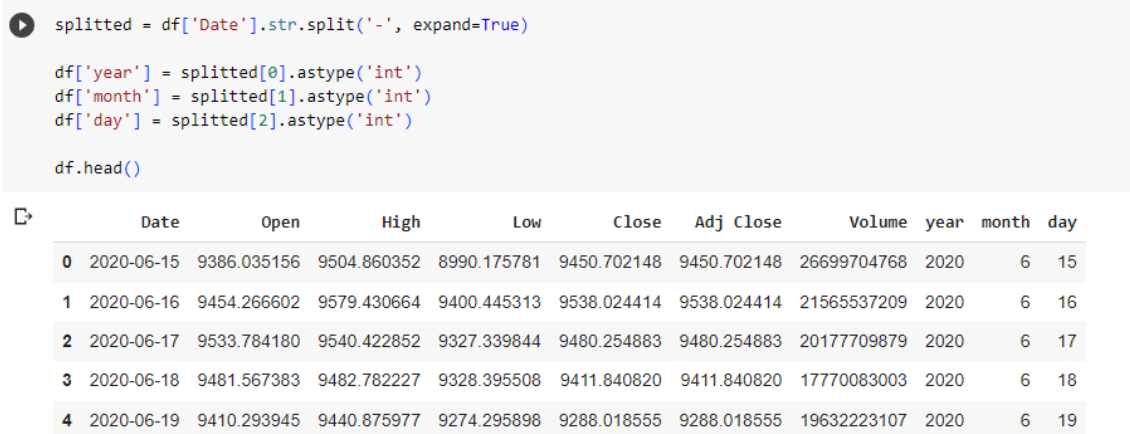
Gambar 6. Kode Import Library

Gambar diatas merupakan langkah pertama yang harus dilakukan saat kita mau melakukan analisis data di google colab yaitu melakukan import library-library yang diperlukan yaitu pandas untuk manipulasi data, matplotlib untuk visualisasi data dan juga sklearn untuk melakukan linear regression dan analisis statistic lainnya



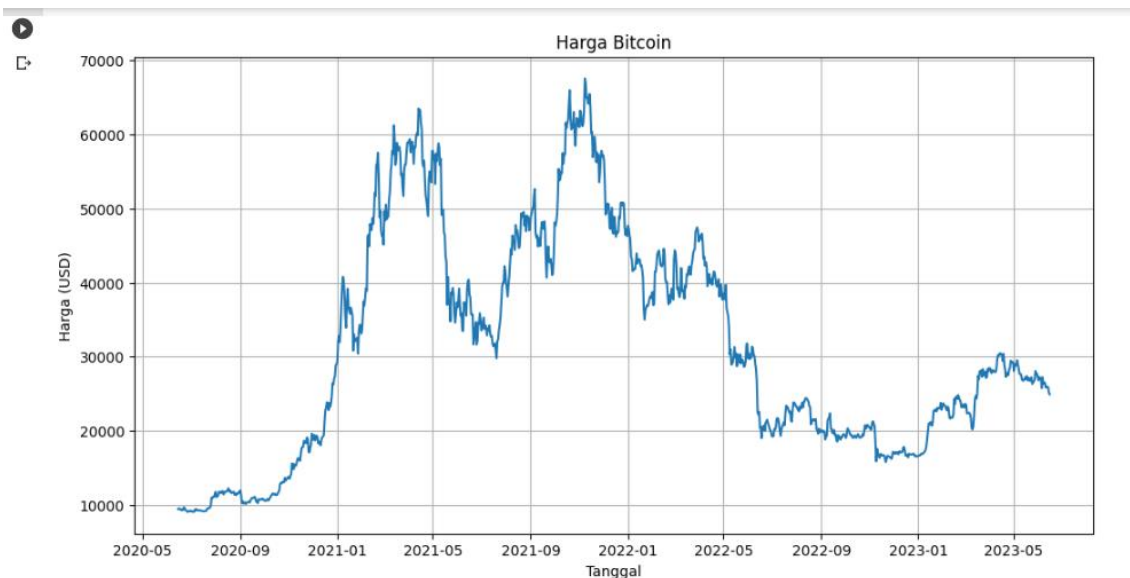
Gambar 7. Line Plot Skala Harga Bitcoin

Gambar diatas ini merupakan bentuk line plot yang didapat dari mengimport matplotlib, di gambar diatas adalah line plot untuk melihat skala data harga bitcoin dari juni 2020 – juni 2023



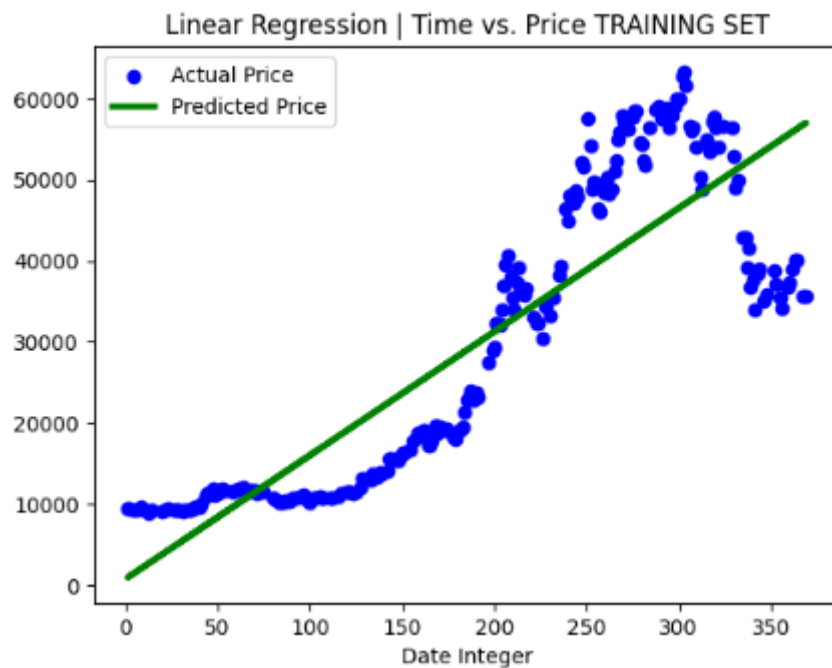
Gambar 8. Kode Membuat Tahun, Bulan dan Hari

Gambar diatas adalah kode untuk menambahkan kolom year, month dan day pada kumpulan data.



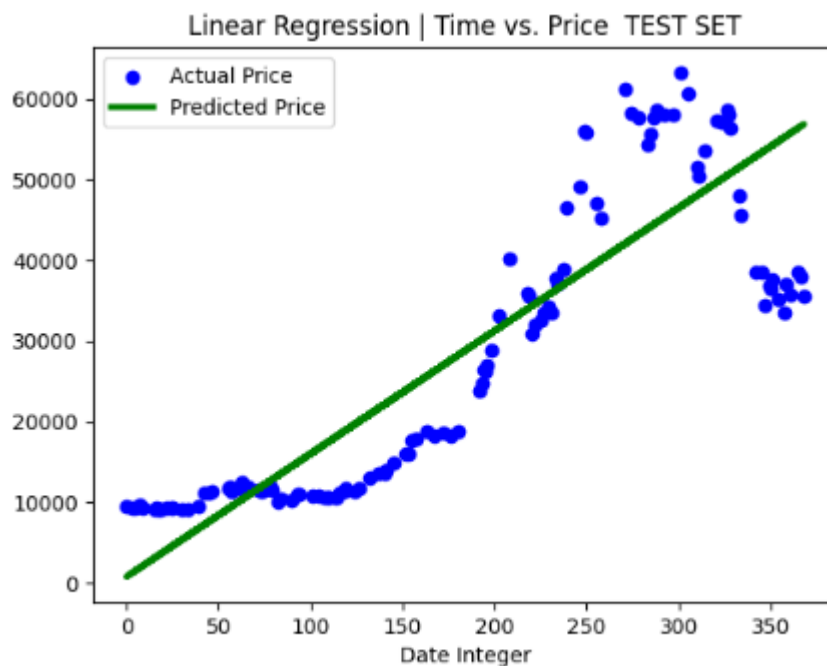
Gambar 9. Line Plot Harga Bitcoin 2020-2023 Berdasarkan Tanggal

Data pada gambar diatas adalah gambaran harga bitcoin dari 2020 – 2023 dengan harga dalam bentuk USD dimana bisa dilihat bitcoin mengalami peningkatan pesat dari tahun 2020 menuju ke tahun 2021 dan harga puncak tertinggi ada pada tahun 2021 mendekati tahun 2022 dan terlihat mulai menurun di tahun 2022.



Gambar 10. Scatter Plot Training Data

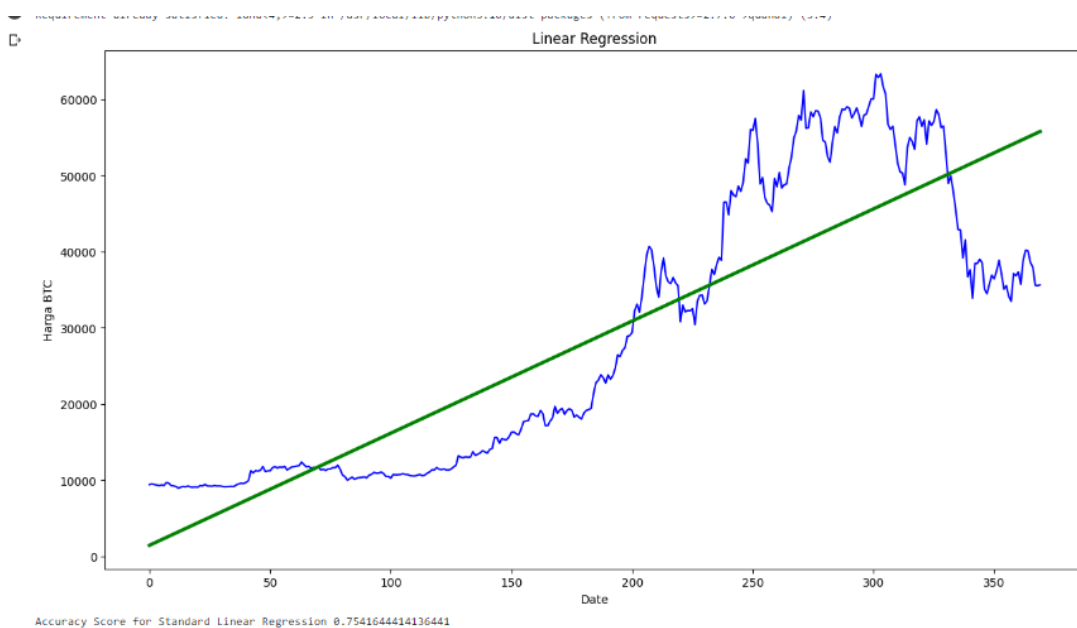
Gambar diatas adalah training set data dengan garis merah merupakan harga yang diprediksi dan scatter berwarna biru adalah harga yang sebenarnya.



Accuracy score for test set 0.7381015741585265

Gambar 11. Scatter Plot Test Data

Gambar 11 adalah test set data dengan garis merah merupakan harga yang diprediksi dan scatter berwarna biru adalah harga yang sebenarnya.



Gambar 12. Regresi Linier Bitcoin

Gambar 12 adalah linear regression yang merupakan prediksi dari saham bitcoin dengan skor akurasi 0,75.

```
import yfinance as yahooFinance

GetInformationJKSE = yahooFinance.Ticker("^JKSE")
datasetJKSE = GetInformationJKSE.history(
    start="2018-07-01", end="2023-7-01",
    interval = "1mo")

GetInformationBTC = yahooFinance.Ticker ("BTC-USD")
datasetBTC = GetInformationBTC.history(
    start="2018-07-01", end="2023-07-01",
    interval = "1mo")

print(datasetJKSE)
print(datasetBTC)
```

Gambar 13. Kode Import Data BTC dan JKSE

Bukan hanya bitcoin saja tapi peneliti juga menganalisis data secara statistic pada data IHSG dengan Bitcoin. Dilakukan import yfinance as yahooFinance berguna untuk mengambil data secara realtime.


```
#Format data saham BTC dan JKSE ke bentuk array
cJKSE = []
cBTC = []
for item in datasetJKSE ["Close"]:
    cJKSE.append(float(item))
for item in datasetBTC ["Close"]:
    cBTC.append(float(item))
```

Gambar 14. Kode Format Data ke bentuk Array

Langkah Selanjutnya adalah melakukan format data saham yang sudah diperoleh dari yahoo finance ke dalam bentuk format data array.

```
#Cari nilai x ( harga close saham JKSE) dan y ( harga close BTC)
x = []
y = []
n_data = len (datasetBTC)
for i in range (1,n_data):
    x.append(cJKSE [i]/ cJKSE[i-1]-1)
    y.append(cBTC [i]/ cBTC [i-1]-1)
```

Gambar 15. Mencari nilai x dan y

Selanjutnya bisa dilihat pada gambar, adalah mencari nilai x yang berupa harga akhir (close) saham JKSE dan y yang berupa harga akhir (harga close)

```
#Rubah array dari kedua saham ke numpy array
#x dibentuk menjadi matriks kolom dari
x = np.array(x).reshape((-1,1))
#y dibentuk menjadi vektor
y = np.array(y)
```

Gambar 16. Kode Ubah Array ke numpy array

Lalu peneliti akan merubah array dari kedua saham menjadi numpy array dan x dibentuk menjadi matriks kolom dan y akan menjadi vector.

```
#Panggil fungsi Linear Regression
model = LinearRegression().fit(x,y)

#Hitung R Square dan tampilkan
r_sq = model.score(x,y)
print(f"coefficient of determination:{r_sq}")

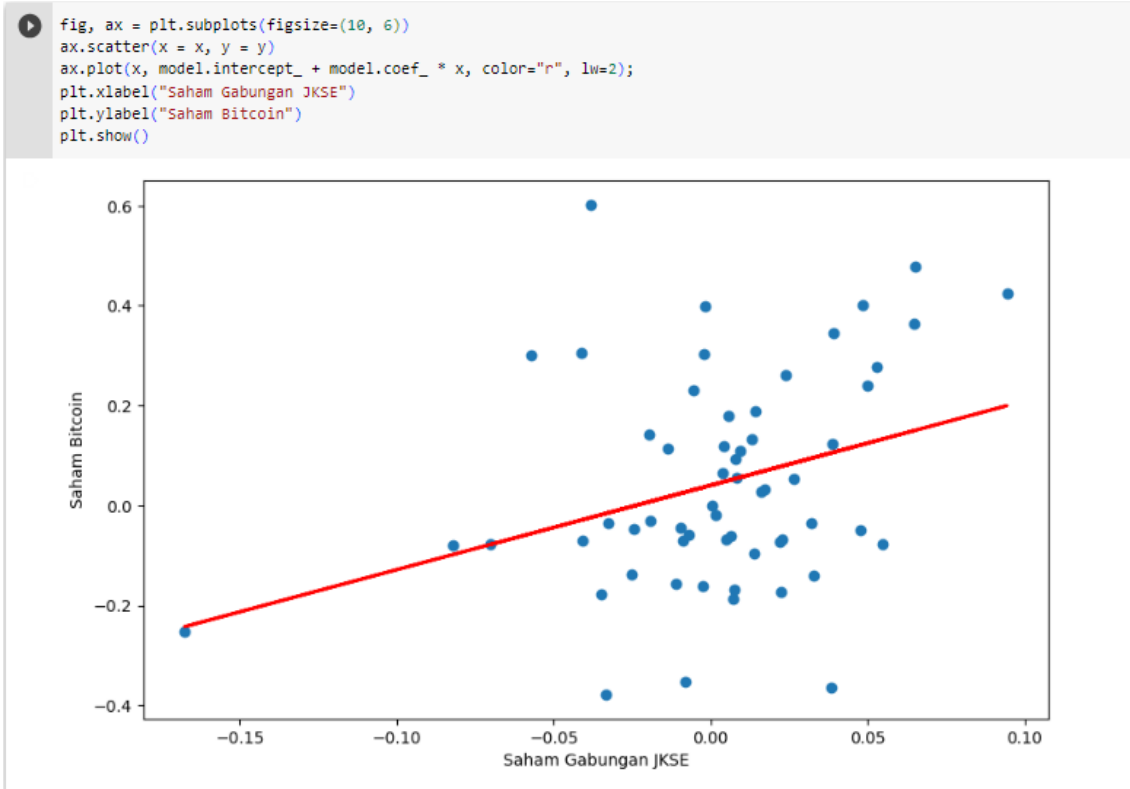
#Hitung interception (  $\beta_0$ ) dan tampilkan
print(f"intercept : {model.intercept_}")

# Hitung slope (  $\beta_1$ ) dan tampilkan
print(f"slope : {model.coef_}")

coefficient of determination:0.10067473299486063
intercept : 0.040262765571296676
slope : [1.68856105]
```

Gambar 17. Linear Regression BTC & JKSE

Langkah Selanjutnya adalah memanggil fungsi regresi linier dan menghitung koefisien, interception dan slope lalu ditampilkan seperti bisa dilihat pada gambar. Lalu langkah terakhir yang dilakukan adalah melakukan visualisasi data dengan scatter plot seperti bisa dilihat pada gambar 18.



Gambar 18. Scatter Plot Linear Regression BTC & JKSE

4. KESIMPULAN

Kesimpulan yang bisa diambil dari analisis ini adalah harga saham bitcoin pada tahun 2020 – 2023 telah mengalami kenaikan pada tahun 2021 dan mulai menurun di pertengahan tahun 2022. dan dengan menggunakan regresi linier didapatkan nilai koefisien : 0,1006, intercept : 0,04 Dan slope : [1,68]. Untuk kenaikan dan penurunan ini dipengaruhi oleh beberapa factor, seperti influensi dari para influencer. dan berdasarkan hasil regresi linier bisa diprediksi dengan akurasi 0.75.

UCAPAN TERIMA KASIH

Terima kasih sebesar-besarnya diucapkan peneliti kepada Bapak Tony Lie selaku dosen mata kuliah big data telah memberikan penjelasan dan pembelajaran saat proses belajar mengajar big data di kelas sehingga dapat terbuatnya laporan ini

DAFTAR PUSTAKA

- [1] Heryadi, Y., & Wahyono, T. (2020). Machine Learning Konsep dan Implementasi. Gava Media.
- [2] K. H. Al-Yahyaee, M. U. Rehman, W. Mensi, and I. M. W. Al-Jarrah, "Can uncertainty indices predict Bitcoin prices? A revisited analysis using partial and multivariate wavelet approaches," *North Am. J. Econ. Financ.*, vol. 49, no. December 2018, pp. 47–56, 2019, doi: 10.1016/j.najef.2019.03.019
- [3] R. Faizal, B. D. Setiawan, and I. Cholissodin, "Prediksi Nilai Cryptocurrency Bitcoin menggunakan Algoritme Extreme Learning Machine (ELM)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. Univ. Brawijaya*, vol. 3, no. 9, pp. 4226–4233, 2019.
- [4] C. H. Wu, C. C. Lu, Y. F. Ma, and R. S. Lu, "A new forecasting framework for bitcoin price with LSTM," *IEEE Int. Conf. Data Min. Work. ICDMW*, vol. 2018-November, pp. 168–175, Feb. 2019, doi: 10.1109/ICDMW.2018.00032.
- [5] P. Mogilev, A. Boldyreva, M. Alexandrov, and J. Cardiff, "GMDH-based Models for Mid-term Forecast of Cryptocurrencies (on example of Waves)," *Int. Sci. Tech. Conf. Comput. Sci. Inf. Technol.*, vol. 2,
- [6] A. Muhammad Fahmi, N. Azah Samsudin, A. Mustapha, N. Razali, and S. Kamal Ahmad Khalid, "Regression based Analysis for Bitcoin Price Prediction," *Int. J. Eng. Technol.*, vol. 7, no. 4.38, p. 1070, 2018, doi: 10.14419/ijet.v7i4.38.27642.
- [7] Jordan, M.I. and Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), pp.255-260.
- [8] Yuliara, I.M., 2016. Regresi linier sederhana. *Regresi Linier Sederhana*, 13.
- [9] Ningsih, S. and Dukalang, H.H., 2019. Penerapan metode suksesif interval pada analisis regresi linier berganda. *Jambura Journal of Mathematics*, 1(1), pp.43-53.
- [10] Anufia, B. and Alhamid, T., 2019. Instrumen pengumpulan data.