
PREDIKSI HARGA SAHAM PFIZER INC (FPE) MENGGUNAKAN METODE REGRESI LINEAR BERGANDA

PFIZER INC (FPE) STOCK PRICE PREDICTION USING MULTIPLE LINEAR REGRESSION

Taufan Adriansyah Lubis¹, Mulyati²

¹²Progam Studi Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Pakuan
065122001@student.unpak.ac.id, mulyati@unpak.ac.id

ABSTRAK

Investasi dalam pasar modal terutama saham saat ini sangat berkembang pesat. Salah satu Saham Pfizer Inc (FPE). Saham ini bergerak dalam bidang Kesehatan yaitu dalam memproduksi vaksin COVID-19 diklaim memiliki tingkat efektivitas yang tinggi sehingga harga sahamnya meningkat selama pandemi dan telah menunjukkan ketahanan yang luar biasa dalam beberapa dekade terakhir. Namun Investasi saham memiliki resiko yang cukup tinggi disebabkan pergerakan harga saham mengalami fluktuasi. Oleh karena itu tujuan penelitian ini adalah untuk Menganalisis prediksi harga saham salah satunya menggunakan Regresi Linear Berganda. Tahapan Penelitian yang dilakukan meliputi pengumpulan data, prepossessing, analisis dengan Regresi linier dan Evaluasi. Hasil penelitian dengan menggunakan data latih 80% dan data Uji 20% diperoleh nilai RMSE sebesar 0.074 yang berarti tingkat kesalahan dalam melakukan prediksi kecil. Dilakukan visualisasi antara nilai aktual dan nilai prediksi didapatkan bahwa ada hubungan yang positif diantara kedua nilai tersebut, hal ini berarti model dapat merepresentasikan pergerakan harga saham dengan baik.

Kata kunci : Pfizer Inc (FPE). Prediksi, Regresi Linear Berganda, Saham.

ABSTRACT

The Investment in the capital market, especially stocks, is currently growing rapidly, one of which is Pfizer Inc (FPE). Pfizer Inc (FPE). This stock is engaged in the Health sector, namely in producing the COVID-19 vaccine is claimed to have a high level of effectiveness, so that its share price has increased during the pandemic and has shown a high level of effectiveness. so that its share price has increased during the pandemic and has shown extraordinary resilience in recent decades. remarkable resilience in recent decades. However, stock investment has a fairly high risk due to stock price movements experiencing fluctuations. Therefore, the purpose of this study is to analyze the prediction of stock prices using Multiple Linear Regression. one of them using Multiple Linear Regression. The stages of research carried out include data collection, prepossessing, analysis with linear regression and evaluation. The results of the study using 80% training data and 20% test data obtained an RMSE value of 0.074 which means the error rate in making small predictions. Visualization between visualization between the actual value and the predicted value found that there is a positive relationship between the two values, this means that the model can represent stock price movements well. represent stock price movements well.

Keywords: Multiple Linear Regression, Pfizer Inc (FPE), Prediction, Stocks.

PENDAHULUAN

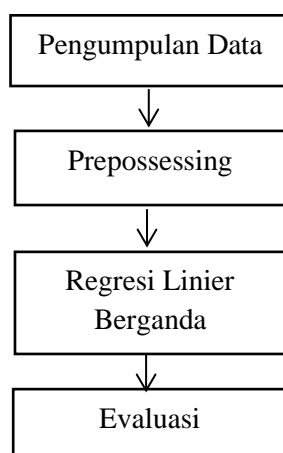
Investasi dalam pasar modal saat ini sangat berkembang pesat. Salah satu yang diperjual belikan dalam pasar modal adalah saham[1]. Saham Pfizer Inc (FPE) merupakan salah satu perusahaan dalam industri perawatan kesehatan yang memproduksi vaksin COVID-19 dan diklaim memiliki tingkat efektivitas yang

tinggi sehingga harga sahamnya meningkat selama pandemi dan telah menunjukkan ketahanan yang luar biasa dalam beberapa dekade terakhir[2]. Namun investasi saham memiliki resiko yang tinggi dikarena sulit meramalkan pergerakan pasar karena fluktuasi harga saham. oleh karena itu penting bagi perusahaan maupun investor untuk memahami dinamika saham dengan baik sebelum membuat Keputusan. Analisis prediksi dapat dilakukan untuk mengurangi resiko kerugian dan membantu dalam menilai investasi di masa depan [3]. Salah satu metode yang digunakan dalam memprediksi harga saham adalah Regresi Linear yaitu linier Berganda. Metode ini menggunakan lebih dari satu variabel bebas[4]. Kelebihan metode tersebut yaitu mampu diterapkan pada data tren seperti pergerakan saham[5], mampu mengumpulkan data dalam situasi ketidakpastian, dan mampu melakukan perhitungan secara paralel, sehingga menghasilkan proses yang lebih singkat[6], serta kemampuannya dalam menangkap hubungan linier antara Variabel bebas dan Variabel terikat [7]

Beberapa penelitian telah dilakukan mengenai prediksi saham dengan menggunakan regresi linier diantaranya prediksi harga saham Bank BCA dengan menggunakan Regresi Linear Berganda[8]. Prediksi harga saham menggunakan Improved Multiple Linear Regression untuk pencegahan data outlier[9], prediksi harga Emas dengan Menggunakan Algoritma Regresi Linear Berganda dan ARIMA[10], prediksi Pergerakan harga Saham BRI Dengan Menggunakan Metode Regresi Linear Berganda[11], dan prediksi prospek harga saham perusahaan perbankan Menggunakan Regresi Linear (Study Kasus Bank BCA Tahun 2015-2017)[12]. Dari beberapa penelitian tersebut dapat disimpulkan bahwa algoritma regresi linear berganda mampu memprediksi harga saham dengan tingkat akurasi yang baik. Oleh karena itu dalam penelitian dilakukanlah prediksi harga saham Pfizer Inc (FPE) dengan menggunakan metode Regresi Linear Berganda.

METODE PENELITIAN

Metode Penelitian dalam analisis data dan mengembangkan model meliputi tahap pengumpulan data, preprocessing, Regresi linier dan Evaluasi. Tahapannya dapat dilihat pada Gambar 1 berikut



Gambar 1. Tahapan Penelitian

Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini berupa data saham Pfizer Inc (FPE) diperoleh melalui *website finance.yahoo.com* dari tahun 2019 sampai dengan Tahun 2023 yang berbentuk *time series* dengan jumlah data sebanyak 1258. Dalam dataset ini ada 6 variabel, yang terdiri dari *Open* (harga pembuka pada periode tertentu), *High* (harga tertinggi yang dicapai pada periode tertentu), *Low* (harga terendah yang dicapai pada periode tertentu), *Close* (harga penutup pada periode tertentu), dan *Adj Close* (harga penutupan yang disesuaikan untuk mempertimbangkan faktor-faktor seperti dividen, perpecahan, dan atau perubahan lain yang mempengaruhi aset) dan *volume* (jumlah transaksi yang dilakukan selama bursa perdagangan buka). Dalam penelitian ini hanya menggunakan 5 variabel yaitu *Open*, *High*, *Low* dan *Close*, sedangkan Variabel *Adj Close* tidak digunakan karena harga tersebut sudah disesuaikan dengan faktor-faktor yang mempengaruhi asset.

Prepossessing

Tahap ini dilakukan untuk mengecek dataset yang sudah dikumpulkan terdapat data missing value. Ditahap ini juga dilakukan normalisasi data supaya data seragam berkisar antara 0 sampai 1[13]. Adapun metode normalisasi yang digunakan adalah *Min Max Scaling*, berikut rumus persamaan (1) dimana X berupa nilai sampel;

$$X_{sc} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

Regresi Linear Berganda

Metode yang digunakan dalam penelitian ini adalah Regresi Linear Berganda. Metode ini dibangun untuk model hubungan antara 1 variabel terikat dan 2 atau lebih variable bebas[14]. Pada umumnya persamaan regresi linear berganda dapat dituliskan seperti persamaan (2), bila terdapat 2 variabel bebas yaitu X_1 dan X_2 maka rumusnya dapat dilihat pada persamaan (3) selanjutnya untuk menghitung nilai a menggunakan persamaan (4) dan untuk menghitung nilai b menggunakan persamaan (5)(6), dimana a merupakan konstanta regresi linear, b adalah garis kemiringan, $X_1..X_n$ adalah Variabel Bebas, dan Y adalah variable terikat.

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (2)$$

$$Y = a + b_1X_1 + b_2X_2 \quad (3)$$

$$a = \frac{\sum Y - (b_1 \sum X_1) - b_2 \sum X_2}{n} \quad (4)$$

$$b_1 = \frac{[(\sum X_2^2 \times \sum X_1 Y) - (\sum X_2 Y \times \sum X_1 X_2)]}{[\sum X_1^2 \sum X_2^2 - (\sum X_1 X_2)^2]} \quad (5)$$

$$b_1 = \frac{[(\sum X_1^2 \times \sum X_2 Y) - (\sum X_1 Y \times \sum X_1 X_2)]}{[\sum X_1^2 \sum X_2^2 - (\sum X_1 X_2)^2]} \quad (6)$$

Evaluasi

Untuk melihat performa dari model Regresi Linear Berganda dapat diukur dengan menggunakan pengukuran nilai koefisien determinan (*R square*), dan *Root Mean Square Error* (RMSE). RMSE menghitung selisih antara nilai sebenarnya dan nilai prediksi. Nilai RMSE yang lebih kecil maka lebih akurat dibandingkan dengan nilai RMSE lebih besar[15]. Rumus RMSE dapat dilihat pada persamaan (7), dimana Y_i : Nilai aktual, \bar{Y}_i : Nilai hasil prediksi, dan n adalah Jumlah data

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2}{n}} \quad (7)$$

HASIL DAN PEMBAHASAN

Dataset yang sudah dikumpulkan berupa saham Pfizer Inc (FPE) diperoleh melalui *website finance.yahoo.com* dari tahun 2019 sampai dengan Tahun 2023 dengan jumlah data sebanyak 1258 yang terdiri dari atribut *Open, Low, High, Close, dan Volume*. Contoh dataset dapat dilihat seperti Gambar 2 berikut

	Date	Open	High	Low	Close	Volume
0	1/2/2019	40.910816	41.271347	40.455406	41.034157	26430315
1	1/3/2019	41.024670	41.119545	39.800758	39.886147	28503533
2	1/4/2019	40.265656	41.129032	40.094875	40.796963	27145348
3	1/7/2019	40.863377	41.385201	40.664139	41.015179	20995469
4	1/8/2019	41.328274	41.461102	40.939278	41.204933	19677231
...
1253	12/22/2023	28.440001	28.790001	28.180000	28.400000	35087200
1254	12/26/2023	28.430000	28.530001	28.250000	28.410000	30119100
1255	12/27/2023	28.469999	28.730000	28.309999	28.610001	35000700
1256	12/28/2023	28.610001	29.030001	28.580000	28.790001	30472900
1257	12/29/2023	28.780001	28.870001	28.540001	28.790001	30342500

1258 rows × 6 columns

Gambar 2. Dataset saham Pfizer Inc (FPE)

Setelah dilakukan pengumpulan data selanjutnya dilakukan preprocessing dengan menggunakan bahasa pemrograman Python, yaitu pengecekan *missing value*. Berdasarkan Gambar 3 dapat dilihat bahwa semua variable berbentuk int64 dan tidak terdapat *missing value*.

```
df.isna().sum()
Date      0
Open      0
High      0
Low       0
Close     0
Volume    0
dtype: int64
```

Gambar 3. Hasil pengecekan *missing Value*

Selanjutnya dilakukan normalisasi data hal ini dilakukan untuk meminimalisir *error* dan menyeragamkan data karena pada Variabel Volume nilainya terlalu jauh dibandingkan variable lainnya. Salah satu metode normalisasi adalah *Min Max Scaling*. Metode ini melakukan transformasi linier terhadap data asli dan cocok diterapkan pada kasus forecasting. Hasil normalisasi dengan metode tersebut akan menghasilkan nilai berada pada rentang 0 sampai 1 pada setiap variable. Menggunakan pemrograman Python hasil normalisasi data dapat dilihat pada Gambar 4.

```
#preprocessing data
from sklearn import preprocessing
import pandas as pd
normal=preprocessing.MinMaxScaler().fit_transform(df1)
normal_data=pd.DataFrame(normal,columns=df1.columns)
normal_data.head()
```

	Open	High	Low	Close	Volume
0	0.428802	0.415705	0.431330	0.424378	0.094426
1	0.432105	0.411365	0.412115	0.391690	0.103642
2	0.410086	0.411636	0.420748	0.417624	0.097604
3	0.427426	0.418959	0.437456	0.423838	0.070267
4	0.440913	0.421129	0.445532	0.429241	0.064408

Gambar 4. Hasil preprocessing menggunakan *Min Max Scaling*

Setelah dilakukan normalisasi selanjutnya dicari model dengan menggunakan metode Regresi Linear Berganda. Terlebih dahulu ditentukan Variabel bebas yaitu *Open*, *High*, *Low*, *Volume*, dan Variabel terikat *Close*. Selanjutnya membagi data latih data Uji dengan perbandingan 80:20, diperoleh hasil seperti pada Gambar 5 berikut

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
x_train, x_test, y_train, y_test = train_test_split(x,y,test_size=0.20, random_state=0)
Reg = LinearRegression()
Reg.fit(x_train, y_train)

#Koefisien
print ('Coefficients:',Reg.coef_)
print ('Intercept:',Reg.intercept_)

Coefficients: [-0.50051082  0.77545796  0.70929249  0.00520193]
Intercept: 0.0047989393700165195
```

Gambar 5. Koefisien regresi dan nilai *intercept* pada pengujian data

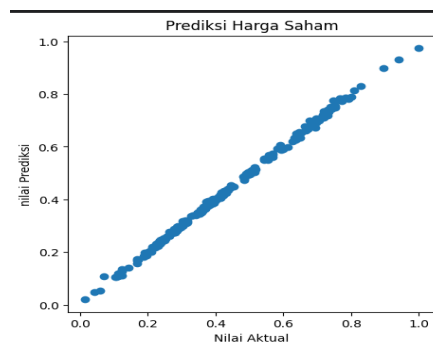
Berdasarkan Gambar 5 nilai koefisien regresi untuk masing-masing variabel bebas terhadap terhadap variable terikat. Variabel *Open* memiliki nilai koefisien negatif yaitu -0,50051082. Hal ini menunjukkan bahwa penurunan harga pembukaan (*Open*) berkorelasi dengan kenaikan harga penutupan(*Close*), begitu juga sebaliknya. Nilai koefisien tertinggi diperoleh pada variabel *High*, yang artinya bahwa Kenaikan harga *High* berkorelasi dengan kenaikan Harga *Close*.

```
#evaluasi model prediksi
from sklearn.metrics import r2_score
import math
R2score = r2_score(y_test,y_pred)
MSE = mean_squared_error (y_test,y_pred)
RMSE = math.sqrt(mean_absolute_error(y_test,y_pred))
print ("MSE :", MSE)
print ("RMSE :", RMSE)
print ("R2Score :", R2score)

MSE : 5.523461661900324e-05
RMSE : 0.0740444654032875
R2Score : 0.9985077238143099
```

Gambar 6. Evaluasi Model

Selanjutnya dilakukan evaluasi model dengan menggunakan *R square* (Koefisien determinan), MSE dan RMSE. Berdasarkan Gambar 6 didapatkan bahwa nilai koefisien determinan sebesar 0,9985 yang diartikan bahwa ada pengaruh antara harga *Open*, *High*, *Low* dan *Volume* terhadap harga *Close* yang sangat kuat. Berdasarkan Persamaan (7) diperoleh nilai RMSE sebesar 0,07, yang berarti tingkat kesalahan dalam prediksi kecil dan hasil prediksi sangat akurat. Nilai prediksi dengan nilai aktual memperoleh selisih yang tidak terlalu jauh seperti yang terlihat pada Gambar 7



Gambar 7. Diagram Regresi

Berdasarkan Gambar 7 dapat disimpulkan bahwa terdapat hubungan yang positif antara nilai prediksi dengan nilai aktual yang berarti model dapat merepresentasikan pergerakan harga saham dengan baik. Dengan demikian keberhasilan model dalam memprediksi harga saham tersebut dapat membantu investor atau para pemangku kepentingan di pasar dalam pengambilan keputusan investasi.

KESIMPULAN DAN SARAN

Penelitian dilakukan untuk memprediksi harga saham Pfizer Inc (FPE) dengan menggunakan metode Regresi Linier Berganda. Data saham diperoleh dari *website yahoo.finance.com* berjumlah 1258 data. selanjutnya dilakukan Preprocessing data yaitu dengan mengecek missing value dan juga dilakukan normalisasi data sehingga nilai data berkisar antara 0 sampai 1. Dataset yang sudah dilakukan preprocessing kemudian dibagi menjadi data latih 80% dan data Uji sebesar 20%. Selanjutnya dilakukan analisis regresi diperoleh nilai *intercept* sebesar 0,00479 dan nilai koefisien regresi untuk masing-masing variabel bebas terhadap variabel terikat. Variabel *Open* memiliki nilai koefisien negatif yaitu -0,50051082. Hal ini menunjukkan bahwa penurunan harga pembukaan (*Open*) berkorelasi dengan kenaikan harga penutupan (*Close*), begitu juga sebaliknya. Nilai koefisien tertinggi diperoleh pada variabel *High*, yang artinya bahwa Kenaikan harga *High* berkorelasi dengan kenaikan Harga *Close*

Berdasarkan hasil prediksi dari model menghasilkan tingkat *error* dengan nilai *error* RMSE yaitu 0.074 yang berarti tingkat kesalahan dalam prediksi kecil. Disamping itu dilakukan visualisasi antara nilai aktual dan nilai prediksi didapatkan bahwa ada hubungan yang positif diantara kedua nilai tersebut, hal ini berarti model dapat merepresentasikan pergerakan harga saham dengan baik.

DAFTAR PUSTAKA

- [1] A. S. Putri, "Analisis Indikasi Asimetris Informasi yang Terjadi pada Saham-Saham Sektor Perbankan di Pasar Modal Indonesia," *J. Ilm. Mhs. FEB*, vol. 10, no. 2, 2022, [Online]. Available: <https://jimfeb.uib.ac.id/index.php/jimfeb/article/view/8130>
- [2] I. Santoso, "Mengenal Dekat Saham Raksasa Farmasi, PFE Pfizer." [Online]. Available: <https://blog.hsb.co.id/saham/saham-pfe-pfizer/>
- [3] L. E. NADHIRA, "Prediksi Harga Saham Pfizer Menggunakan Long Short-Term Memory Dan Bidirectional Long Short-Term Memory (Studi Kasus: Close Price Saham Pfizer Inc. Pada Januari 2018-Januari 2022)," UNIVERSITAS ISLAM INDONESIA, 2022. [Online]. Available: <https://dspace.uin.ac.id/handle/123456789/40365?show=full>
- [4] A. B. K. Bruce L. Bowerman, Richard T. O'Connell, *Forecasting, Time Series, and Regression: An Applied Approach*, 4th, illustr ed. Thomson Brooks/Cole, 2005.
- [5] R. Zapar *et al.*, "PENERAPAN MODEL REGRESI LINIER UNTUK PREDIKSI HARGA SAHAM," vol. 8, no. 1, pp. 196–202, 2024.
- [6] A. Amrin, "Peramalan Tingkat Inflasi Indonesia Menggunakan Neural Network Backpropagation Berbasis Metode Time Series," *AMIK Bina Sarana Inform. Jakarta*, vol. XI, no. 2, pp. 129–136, 2014, [Online]. Available: <http://ejournal.nusamandiri.ac.id/ejournal/index.php/techno/article/view/87>
- [7] R. Zapar, D. Pratama, Kaslani, C. L. Rohmat, Faturrohman, "PENERAPAN MODEL REGRESI LINIER UNTUK PREDIKSI HARGA SAHAM BANK BCA PADA BURSA EFEK INDONESIA", *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol 8, no. 1, pp, 196-202, 2024 [Online]. Available: <https://ejournal.itn.ac.id/index.php/jati/article/download/8215/4967/>
- [8] A. S. Nuriyah, R. Saladini, K. N. Fujiah, and F. Indrayatna, "Prediksi Harga Saham Bank BCA Menggunakan Metode Regresi Linier Berganda," *Semin. Nas. Stat. Aktuaria Ii*, 2023, [Online]. Available: <https://prosidingnsa.statistics.unpad.ac.id/?journal=prosidingnsa&page=article&op=view&path%5B%5D=323>
- [9] A. Izzah and R. Widyastuti, "Prediksi Harga Saham Menggunakan Improved Multiple Linear Regression untuk Pencegahan Data Outlier," *Kinet. Game Technol. Inf. Syst. Comput. Network, Comput. Electron. Control*, vol. 2, no. 3, pp. 141–150, 2017, doi: 10.22219/kinetik.v2i3.268.
- [10] Y. F. Wijaya and A. Triayudi, "Penerapan Data Mining Pada Prediksi Harga Emas dengan

- Menggunakan Algoritma Regresi Linear Berganda dan ARIMA,” *J. Comput. Syst. ...*, vol. 5, no. 1, pp. 73–81, 2023, doi: 10.47065/josyc.v5i1.4615.
- [11] J. Teknologi *et al.*, “Implementasi Data Mining Untuk Memprediksi Pergerakan Harga Saham Bri Dengan Menggunakan Metode Regresi Linier Berganda Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD,” vol. 7, no. September 2022, pp. 151–159, 2024.
- [12] M. Keren and R. S. Oetama, “Prediksi prospek harga saham perusahaan perbankan Menggunakan Regresi Linear (Study Kasus Bank BCA Tahun 2015-2017),” *JSI J. Sist. Inf.*, vol. 11, no. 1, pp. 1703–1710, 2019, doi: 10.36706/jsi.v11i1.8105.
- [13] T. Tasmalaila Hanifa and S. Al-Faraby, “Analisis Churn Prediction pada Data Pelanggan PT. Telekomunikasi dengan Logistic Regression dan Underbagging,” *e-Proceeding Eng.*, vol. 4, no. 2, pp. 3210–3225, 2017.
- [14] M. P. Prof. Dr. A. Muri Yusuf, *Metode Penelitian Kuantitatif, Kualitatif & Penelitian Gabungan*, Cetakan ke. Jakarta: Kencana, 2017.
- [15] N. Salwa, N. Tatsara, R. Amalia, and A. F. Zohra, “Model Prediksi Liku Kalibrasi Menggunakan Pendekatan Jaringan Saraf Tiruan (JST) (Studi Kasus: Sub DAS Siak Hulu),” *J. Data Anal.*, vol. 1, no. 2011, pp. 21–31, 2018, [Online]. Available: <http://ce.unri.ac.id>