

Perbandingan Metode Regresi Linier Berganda dan Autoregressive Integrated Moving Average (ARIMA) Untuk Prediksi Saham PT. BSI, Tbk.

Hilman Winnos

Tehnik Informatika, Universitas Gaja Putih

Email: Himanaulia33@gmail.com

Richashanty Septima

Tehnik Informatika, Universitas Gaja Putih

Email: richaseptima@gmail.com

Husna Gemasih

Tehnik Informatika, Universitas Gaja Putih

Email: gemasihhusna@gmail.com

Korespondensi penulis: Himanaulia33@gmail.com

Abstract. *In the period from 2018 to 2021 the shares of PT. BSI Tbk (BRIS) tends to experience price fluctuations every day, so from the time-series data collection of BRIS stocks, data mining is needed to find patterns of predictive models that can find useful information. The data mining method, also known as knowledge discovery, is extracting information stored in a large database through studies of collecting, cleaning, processing, and analyzing, so as to get useful things from BRIS stock data. The model used is the multiple linear regression method, and the ARIMA method where both methods have advantages in numerical data analysis which is quite accurate. The purpose of this research is to apply and produce an accurate equation model between the two methods in predicting the stock price of PT. BSI Tbk. The results obtained are multiple linear regression models with a MAPE value of 1.1%, which is 98.9% more accurate than the ARIMA model, namely a MAPE value of 2.36% and an accuracy of 98.9%.*

Keywords: *Minimum Data; Multiple Linear Regression; ARIMA; MAPE.*

Abstrak. Pada periode tahun 2018 sampai dengan tahun 2021 saham PT. BSI Tbk (BRIS) cenderung mengalami fluktuasi harga setiap hari, sehingga dari kumpulan data time-series saham BRIS dibutuhkan penggalian data untuk menemukan pola model prediksi yang dapat menemukan informasi yang bermanfaat. Metode Data mining atau disebut juga dengan knowledge discovery merupakan penggalian informasi yang tersimpan dalam basis data yang besar melalui studi mengumpulkan, membersihkan, memproses, dan menganalisis, sehingga mendapatkan hal yang berguna dari data saham BRIS. Adapun model yang digunakan adalah metode regresi linier berganda, dan metode ARIMA dimana kedua metode tersebut memiliki keunggulan dalam analisis data numerik yang cukup akurat. Tujuan penelitian ini adalah untuk menerapkan dan menghasilkan model persamaan yang akurat antara kedua metode tersebut dalam memprediksi harga saham PT.

Received Februari 07, 2022; Revised Maret 2, 2022; April 22, 2022

* Hilman Winnos, Himanaulia33@gmail.com

BSI Tbk. Hasil yang didapat adalah model regresi linier berganda dengan hasil nilai MAPE 1,1% yaitu 98,9% lebih akurat dibandingkan dengan model ARIMA yaitu dengan hasil nilai MAPE 2,36% dan akurasinya 98,9%.

Kata kunci: Data Mining; Regresi Linier Berganda; ARIMA; MAPE.

LATAR BELAKANG

Data mining atau disebut juga dengan *knowledge discovery* merupakan penggalian informasi yang tersimpan dalam basis data yang besar melalui studi mengumpulkan, membersihkan, memproses, dan menganalisis, sehingga mendapatkan hal yang berguna dari data (Angarwal, 2015). Adapun fokus *data mining* sendiri adalah pada bagaimana manusia dapat memecahkan masalah dari pola hasil identifikasi pembelajaran komputer atau disebut juga *machine learning* (Faisal dan Nugrahadi, 2019).

Dalam bidang *data mining* ada banyak algoritma atau metode yang dapat digunakan untuk melakukan penggalian data. Salah satu metode penggalian data yang terkait dengan data *time-series* adalah *regression* yang secara umum, metode ini digunakan untuk memprediksi nilai variabel respon (dependen) dari satu atau lebih variabel prediktor (independen), dimana variabelnya numerik dengan bentuk regresi, seperti linier, ganda, dan lain-lain (Han *et al.*, 2012). Untuk perbandingan metode tersebut penulis membandingkan dengan model statistika yaitu metode *autoregressive integrated moving average* (ARIMA) dimana metode ini biasa digunakan untuk analisis data deret berkala (*time-series*), peramalan dan pengendalian (Angarwal, 2015).

ANALISA MASALAH DAN PERANCANGAN PROGRAM

1. Tahap Statistik Deskriptif

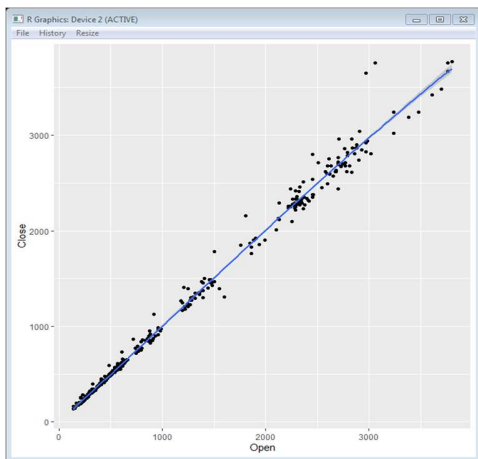
Tahap ini bertujuan untuk menganalisis data statistik terhadap variabel target atau dependent yaitu harga *close* dengan menggunakan fungsi *summary()*, dengan hasil analisis dapat dilihat pada tabel 3.5 berikut:

Tabel 3.5 Statistik variable target *close*

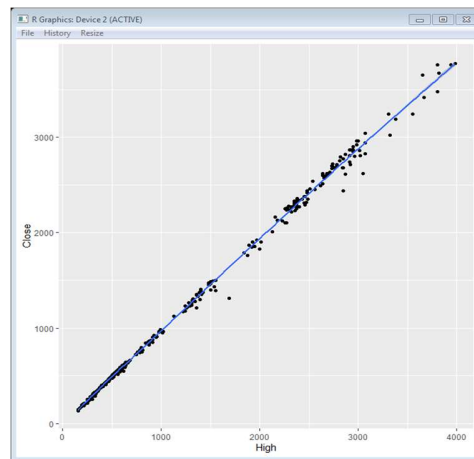
Min.	1st Qu	Median	Mean	3rd Qu	Max.
135.0	398.0	540.0	842.8	750.0	3770.0

Dari tabel 3.5 dapat diketahui bahwa harga *close* minimal berada pada nilai 135 dan maksimal pada nilai 3770, kemudian nilai kuartal pertama adalah 398 dan kuartal ketiga 750, selanjutnya nilai median pada 540 dan mean 842 sehingga dapat diasumsikan bahwa data harga *close* tidak terdistribusi secara simetris, yaitu ada kemungkinan peningkatan harga *close* setiap harinya.

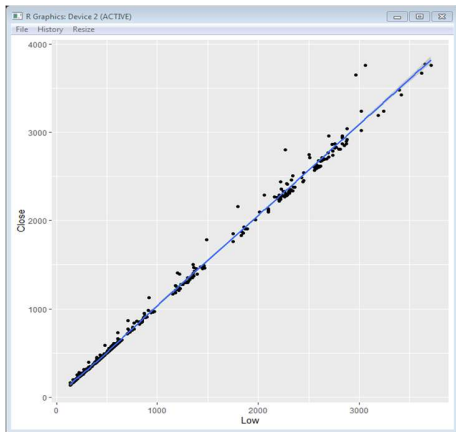
Sehingga untuk membuktikannya dibutuhkan *scatter plot* untuk harga *close* terhadap variabel bebas/ independent yaitu terhadap variabel harga *open*, *low*, *high*, dan *volume* dengan menggunakan fungsi *ggplot()*, dengan hasil seperti pada gambar 3.5, 3.6, 3.7 dan 3.8.



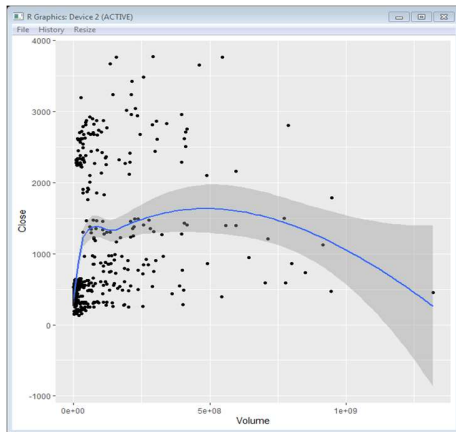
Gambar 3.5 Hubungan variable *Open* terhadap *Close*
 Hubungan variable *High* terhadap *Close*



Gambar 3.6



Gambar 3.7 Hubungan variable *Low* terhadap *Close*



Gambar 3.8 Hubungan variable *Volume* terhadap *Close*

2. Tahapan estimasi

Pada tahap ini bertujuan untuk mengestimasi model regresi yang akan digunakan dengan melihat nilai *t value* apakah variabel bebas/ independent signifikan pada taraf 5% yaitu dengan cara menggunakan fungsi *lm()* yaitu fungsi *linier modeling*, dengan hasil analisis estimasi dapat dilihat pada gambar 3.9 berikut:

```
Call:
lm(formula = Close ~ Open + High + Low + Volume, data = DatasetBRIS_JK)

Residuals:
    Min       1Q   Median       3Q      Max
-230.340  -3.317    0.296    4.207   156.867

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.471e+00  1.286e+00   1.144   0.253
Open        -7.625e-01  2.877e-02 -26.504 <2e-16 ***
High         8.584e-01  2.039e-02  42.096 <2e-16 ***
Low          8.980e-01  3.048e-02  29.459 <2e-16 ***
Volume      -5.832e-09  7.662e-09  -0.761   0.447
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 22.04 on 717 degrees of freedom
Multiple R-squared:  0.9992,    Adjusted R-squared:  0.9992
F-statistic: 2.281e+05 on 4 and 717 DF, p-value: < 2.2e-16
```

3. Tahap Interpretasi

Berdasarkan gambar 3.9 pada tahap estimasi pemodelan maka diperoleh koefisien untuk $b_0, b_1, b_2, b_3,$ dan b_4 yaitu:

$$b_0 = 1.471$$

$$b_1 = -0.7625$$

$$b_2 = 0.8584$$

$$b_3 = 0.898$$

$$b_4 = -0.000000005832$$

Untuk mendapatkan nilai koefisien regresi untuk $b_0, b_1, b_2, b_3,$ dan b_4 pada variable-variabel $x_1, x_2, x_3,$ dan x_4 pada model persamaan (1) dapat juga dilakukan dengan perhitungan manual yaitu dengan metode matriks (metode kuadrat terkecil).

Metode Penelitian

Dalam penelitian ini penulis menggunakan 3 (tiga) metode yaitu:

1. Metode Pengumpulan Data melalui observasi langsung pada data saham PT BSI Tbk dari *website* Yahoo Finance dan studi pustaka dalam menggali teori-teori dan penelitian terkait dengan penerapan *data mining* menggunakan metode regresi linier berganda dan analisis data *time-series* menggunakan ARIMA.

2. Metode analisis *data mining* melalui 5 (lima) tahapan yaitu: Seleksi Data, *Preprocessing Data*, Transformasi Data, *Mining Data*, dan *Evaluation*.
3. Metode Evaluasi perbandingan hasil model menggunakan metode *mean absolute percentage error* (MAPE).

HASIL DAN PEMBAHASAN

Dalam pembahasan ini akan dilakukan tahap akhir yaitu *forecasting* atau prediksi terhadap *dataset* yang digunakan. Adapun prediksi dibagi dalam dua bagian yang pertama prediksi dengan persamaan (1) yaitu model persamaan regresi linier berganda, kemudian yang kedua menggunakan model persamaan (2) yaitu model persamaan ARIMA.

Berdasarkan pada tabel 3.3 bahwa variabel independen ada 4 (empat) yaitu variabel *open*, *high*, *low*, dan *volume*. Sedangkan untuk variabel dependennya adalah variabel *close*.

1. Model persamaan regresi linier berganda

No	Open	High	Low	Volume	Close
1	625	635	605	19430400	610
2	610	620	605	7919500	605
3	610	630	605	24427900	625
4	625	630	615	6796800	615
5	620	620	605	4444900	610
6	610	610	590	7244900	595
7	605	605	575	6415800	585
8	585	590	575	3595000	585
9	585	585	550	16082600	555
10	555	575	555	17860900	570
...
...
...
720	1860	1995	1830	111361900	1830
721	1850	1890	1835	49657900	1870
722	1880	1925	1870	34389500	1905

Adapun untuk metode ARIMA variabel yang digunakan adalah variable *date* dan variable *close*, seperti yang dapat dilihat pada tabel 3.4 berikut:

Tabel 3.4 *Dataset* saham PT. BSI Tbk (BRIS) untuk ARIMA

2. Model persamaan ARIMA(1,1,1)

No.	Date	Close
1	6/4/2018	610
2	6/5/2018	605
3	6/6/2018	625
4	6/7/2018	615
5	6/8/2018	610
6	6/20/2018	595
7	6/21/2018	585
8	6/22/2018	585
9	6/25/2018	555
10	6/26/2018	570
...
...
...
720	5/27/2021	1830
721	5/28/2021	1870
722	5/31/2021	1905

Kesimpulan

Adapun kesimpulan yang dapat diambil dari penelitian ini adalah:

- 1 Penerapan metode regresi linier berganda dalam memprediksi harga saham PT BSI, Tbk menghasilkan model persamaan regresi linier berganda dengan akurasi yang tinggi yaitu 98,9% dengan nilai MAPE hanya 1,1%.
- 2 Penerapan metode ARIMA dalam memprediksi harga saham PT BSI, Tbk menghasilkan model persamaan ARIMA dengan akurasi yang tinggi yaitu 97,64% dengan nilai MAPE hanya 2,36%.
- 3 Hasil perbandingan model regresi linier berganda dan ARIMA menunjukkan bahwa nalia akurasi tertinggi didapat oleh model persamaan dengan metode regresi linier berganda yaitu 98,9%.

Daftar Pustaka

- Aggarwal, C. C. *Data Mining*. New York: Springer, 2015.
- Faisal, M Reza dan Dodon T. Nugrahadi. *Belajar Data Science: Klasifikasi dengan Bahasa Pemrograman R*. Banjar Baru: Scripta Cendekia, 2019.
- Han, Jiawei dkk. *Data Mining: Concepts And Techniques 3rd Edition*. United States: Morgan Kaufmann. 2012.
- Salim, Joko. *Cara Gampang Bermain Saham*. Jakarta: Visimedia. 2010
- Wardhono, Adhitya dkk. *Analisis Data Time Series Dalam Model Makroekonomi*. Jawa Timur: CV. Pustaka Abadi. 2019.
- Larose, T. Daniel. *Discovering Knowledge In Data: An Introduction to Data Mining*. Canada: Jhon Wiley & Sons, Inc. 2005.
- Widoatmodjo, Sawidji. *Cara Sehat Investasi di Pasar Modal*. Jakarta: Media Komputindo. 2005.
- Akhmad, Ekka Pujo Ariesanto. "Data Mining Menggunakan Regresi Linear untuk Prediksi Harga Saham Perusahaan Pelayaran". *Jurnal Aplikasi Pelayaran dan Kepelabuhanan* 10 (2) (2020): 120-131.
- Maharadja, Alfanda Novebrian dkk. "Penerapan Metode Regresi Linier Berganda untuk Prediksi Kerugian Negara Berdasarkan Kasus Tindak Pidana Korupsi". *Journal of Applied Informatics and Computing* Vol.5 No.1 (2021): 95-102.
- Prasetyo, Vicentius Riandaru dkk. "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier". *Jurnal Nasional Teknologi dan Sistem Informasi* Vol.7 No.1 (2021): 8-17.
- Ayuni, Najla Ayuni dan Devi Fitrihanah. "Penerapan Metode Regresi Linier Untuk Prediksi Penjualan Properti pada PT XYZ". *Jurnal Telematika* Vol.14 No.2 (2019): 79-85.

Maulana, Reza dan Devy Kumalasari. “Analisis Dan Perbandingan Algoritma *Data Mining* Dalam Prediksi Harga Saham GGRM”. Jurnal Informatika Kaputama Vol.3 No.1 (2019): 22-28.

Nur Rohmi Aida. Merger Bank Syariah Indonesia. [online] <https://www.kompas.com/tren/read/2021/05/02/080000865/merger-bank-syariah-indonesia-perlukah-nasabah-migrasi-rekening-?page=all/>. (diakses Februari 2022)

Admin. Sejarah Perusahaan. [online] <https://www.bankbsi.co.id/company-information/tentang-kami/>. (diakses Februari 2022)