

# Implementasi *Genetic Algorithm* dalam Model ARIMA untuk Memprediksi Observasi Time Series

Rangga Arya Pamungkas <sup>#1</sup>, Indwiarti <sup>\*2</sup>, Aniq Atiqi Rohmawati <sup>#3</sup>

*Program Studi Informatika, Fakultas Informatika, Universitas Telkom  
Jl. Telekomunikasi No. 1 Terusan Buah Batu, Bandung 40257, Indonesia*

<sup>1</sup> ranggaaryaap@student.telkomuniversity.ac.id

<sup>2</sup> indwiarti@telkomuniversity.ac.id

<sup>3</sup> aniqatigi@telkomuniversity.ac.id

## Abstract

The value of stock prices is always changing and fluctuate every day. To deal with problems regarding uncertainty in stock prices, a time series forecasting is needed to predict stock prices in the future. In this study, the method used to predict stock prices is the Autoregressive Moving Average (ARIMA) method. To improve the accuracy of stock price predictions, the Genetic Algorithm (GA) will be implemented in the best ARIMA model obtained from the ARIMA process. The results of this study indicate that stock price predictions of PT Bank Central Asia Tbk using the ARIMA (1,1,1) model have a Root Mean Square Error (RMSE) value of 418.1314. Whereas the results of stock price prediction of PT Bank Central Asia Tbk by implementing GA on ARIMA (1,1,1) model with 600 generations, 1200 generations, 1800 generations, 2400 generations, and 3000 generations each have RMSE values of 5827.378, 1319.903, 1080.704, 563.7984, and 371.0107 respectively. The results obtained indicate that the implementation of GA on ARIMA with 3000 generations can improve the accuracy of stock price predictions of PT Bank Central Asia Tbk, that is having a RMSE value of 371.0107.

**Keywords:** ARIMA Model, GA, Prediction, RMSE, Stock Price

## Abstrak

Nilai harga saham selalu berubah-ubah dan berfluktuasi setiap harinya. Untuk menghadapi masalah mengenai ketidakpastian harga saham, perlu dilakukan suatu peramalan *time series* untuk memprediksi harga saham di masa mendatang. Pada penelitian ini, metode yang digunakan untuk memprediksi harga saham adalah metode *Autoregressive Moving Average* (ARIMA). Untuk meningkatkan akurasi dari prediksi harga saham, akan diimplementasikan *Genetic Algorithm* (GA) pada model ARIMA terbaik yang didapatkan dari proses ARIMA. Hasil dari penelitian ini menunjukkan bahwa prediksi harga saham PT Bank Central Asia Tbk dengan menggunakan model ARIMA (1,1,1) memiliki nilai *Root Mean Square Error* (RMSE) sebesar 418.1314. Sedangkan hasil prediksi harga saham PT Bank Central Asia Tbk dengan mengimplementasikan GA pada model ARIMA (1,1,1) dengan 600 generasi, 1200 generasi, 1800 generasi, 2400 generasi, dan 3000 generasi masing-masing memiliki nilai RMSE berturut-turut sebesar 5827.738, 1319.903, 1080.704, 563.7984, dan 371.0107. Hasil yang didapat menunjukkan bahwa pengimplementasian GA pada ARIMA dengan 3000 generasi dapat meningkatkan akurasi prediksi harga saham PT Bank Central Asia Tbk, yaitu dengan memiliki nilai RMSE sebesar 371.0107.

**Kata Kunci:** GA, Harga Saham, Model ARIMA, Prediksi, RMSE

## I. PENDAHULUAN

**B**ELAKANGAN ini perkembangan bisnis saham di Indonesia mengalami peningkatan yang cukup pesat. Hal ini dikarenakan saham memiliki nilai bisnis yang tinggi untuk mendapatkan keuntungan baik bagi perusahaan maupun investor yang menanam modal di saham. Dalam melakukan investasi saham, ada risiko investasi yang diakibatkan oleh fluktuasi naik turunnya harga saham yang tidak menentu dan dapat mengakibatkan ketidakpastian tingkat keuntungan yang didapat.

Untuk mendapatkan tingkat keuntungan yang diinginkan, maka investor harus bisa memperkirakan dan menganalisis fluktuasi naik turunnya harga saham, sehingga investor dapat mengetahui kapan harus membeli dan kapan harus menjual saham. Oleh karena itu, perlu adanya peramalan akurat agar investor dapat memprediksi harga saham yang tentunya dapat diandalkan untuk dijadikan dasar keputusan dalam melakukan investasi saham.

Model ARIMA merupakan salah satu metode *time series* yang dapat digunakan untuk memprediksi harga saham. Model ARIMA dikenal efisien dalam memprediksi data *time series*, terutama untuk prediksi jangka pendek [1]. Model ARIMA melibatkan model regresi dan korelasi dari historis data pada masa lalu. Secara umum, model ARIMA dirumuskan sebagai ARIMA ( $p, d, q$ ) dimana  $p$  merupakan orde dari *Autoregressive* (AR),  $d$  merupakan orde *differencing*, dan  $q$  merupakan orde *Moving Average* (MA). Pada penelitian sebelumnya, model ARIMA pernah digunakan untuk melakukan prediksi harga saham *automobile* [2]. Penelitian tersebut menghasilkan model ARIMA yang dapat digunakan untuk melakukan prediksi harga saham dengan akurasi prediksi diatas 75%. Selain itu, pernah dilakukan penelitian mengenai perbandingan prediksi untuk penyakit *hepatitis B* menggunakan metode ARIMA dan *Grey Model* (GM). Hasil pada penelitian tersebut menunjukkan bahwa ARIMA mempunyai akurasi yang lebih tinggi jika dibandingkan dengan GM [3].

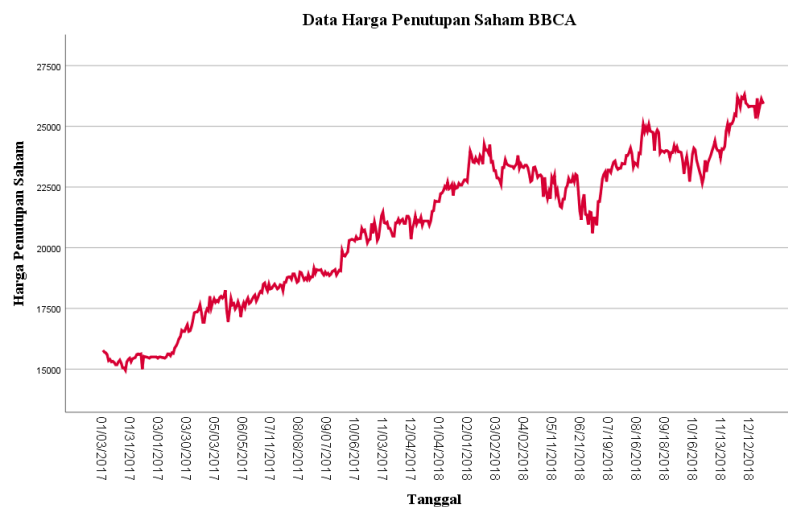
GA merupakan algoritma pencarian dan optimasi yang berdasarkan pada prinsip evolusi alam [4]. GA meniru perumpamaan evolusi biologis dengan tahapan-tahapan tertentu. Pada penelitian sebelumnya, pernah dilakukan prediksi harga saham dengan menggunakan model regresi linier yang dibangun dengan algoritma GA [5]. Hasil yang didapatkan adalah bahwa pembentukan model regresi linier dengan algoritma GA dapat memberikan akurasi yang baik dalam memprediksi harga saham. Selain itu, pernah dilakukan penelitian prediksi harga saham dengan menggunakan GA dan *Evolution Strategies* [6]. Pada penelitian tersebut, akurasi prediksi harga saham yang didapatkan dengan metode GA mempunyai akurasi yang lebih baik dibanding dengan *Evolution Strategies*, dengan akurasi prediksi GA sebesar 73.87%. Selanjutnya, pernah juga dilakukan penelitian prediksi curah hujan dengan menggunakan metode *hybrid Tsukamoto Fuzzy Inference System* (FIS) dan GA [7] dengan hasil yang menunjukkan bahwa metode *hybrid* mempunyai akurasi prediksi lebih baik jika dibandingkan dengan prediksi yang menggunakan FIS saja.

Pada penelitian ini, akan digunakan metode yang pernah digunakan pada penelitian sebelumnya, yaitu menggunakan metode ARIMA dan GA. Merujuk pada penelitian sebelumnya, metode-metode tersebut terbukti dapat diaplikasikan untuk memprediksi harga saham dengan baik. Metode ARIMA akan digunakan untuk memprediksi harga saham dari PT Bank Central Asia Tbk. Setelah itu, akan digunakan metode GA yang diimplementasikan pada model ARIMA ( $p, d, q$ ) untuk memprediksi harga saham dari PT Bank Central Asia Tbk. Metode GA bekerja dengan mencari nilai koefisien terbaik dari model ARIMA yang didapat. Dengan mengimplementasikan GA pada model ARIMA, diharapkan akurasi dari prediksi agar saham dapat ditingkatkan menjadi lebih baik.

## II. LITERATURE REVIEW

### A. Analisis Time Series

Data *time series* merupakan data yang terdiri dari rangkaian nilai suatu variabel yang berurutan dan memiliki beberapa periode waktu seperti harian, mingguan, bulanan, dan tahunan. Data *time series* yang dianalisis dapat membantu dalam memprediksi kejadian di masa yang akan datang. Pada data *time series*, nilai variabel-variabelnya memiliki ketergantungan terhadap waktu. Dalam penelitian ini, data *time series* yang digunakan adalah data harga penutupan saham dari PT Bank Central Asia Tbk (BBCA.JK).



Gambar 1. Data Time Series dari Harga Penutupan Saham BBCA

Analisis *time series* merupakan suatu teknik statistik yang digunakan untuk menganalisis data *time series*. Analisis *time series* bertujuan untuk mendapatkan informasi yang dapat membantu dalam memahami karakteristik data dan memprediksi nilai di masa mendatang [8]. Oleh karena itu, untuk memilih suatu metode *time series* yang tepat harus mempertimbangkan pola data agar metode yang digunakan dengan pola tersebut dapat diuji dan mendapatkan hasil berupa akurasi yang baik.

*Autoregressive Moving Average* (ARIMA) merupakan salah satu metode dalam analisis *time series* yang biasa digunakan untuk melakukan prediksi di masa mendatang. Model ARIMA dinotasikan dengan bentuk ARIMA ( $p,d,q$ ) dimana  $p$  menyatakan orde *Autoregressive* (AR),  $d$  menyatakan orde *differencing* dan  $q$  menyatakan orde *Moving Average* (MA). Secara matematis, model ARIMA dalam prediksi dapat dirumuskan sebagai berikut [9]:

$$y_t = \mu + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

dimana  $\mu$  merupakan konstanta,  $\phi_p$  dan  $\theta_q$  secara berturut-turut merupakan koefisien AR dan MA pada lag  $p$  dan  $q$ , dan  $y_{t-p}$  dan  $\varepsilon_{t-q}$  secara berturut-turut merupakan nilai variabel dependen pada lag  $p$  dan error prediksi pada lag  $q$ .

Dalam proses ARIMA, data *time series* yang digunakan dapat berbentuk data stasioner maupun non-stasioner. Data dikatakan stasioner apabila data tidak mengalami perubahan secara signifikan seiring dengan waktu yang berubah, dengan kata lain nilai rata-rata dan variansnya konstan terhadap waktu. Salah satu cara untuk mengidentifikasi bahwa data stasioner atau tidak adalah dengan menggunakan analisis *Autocorrelation Function* (ACF) dan *Partial Autocorrelation Function* (PACF). ACF merupakan korelasi atau hubungan antar

data pengamatan suatu data *time series*. Selain untuk mengecek kestasioneran data, ACF juga digunakan untuk menentukan orde  $q$  pada model ARIMA  $(p,d,q)$ . Perumusan fungsi ACF dapat dituliskan sebagai berikut [10]:

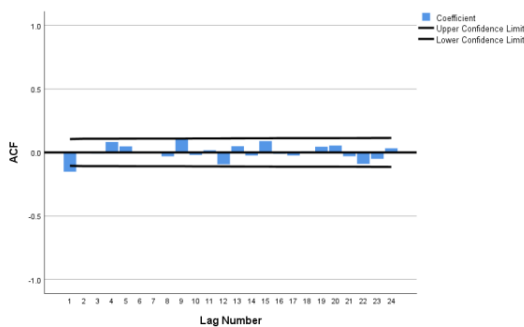
$$r_k = \frac{\sum_t^n (Y_t - \bar{Y})(Y_{t+k} - \bar{Y})}{\sum_t^n (Y_t - \bar{Y})^2} \quad (2)$$

dimana  $r_k$  merupakan koefisien autokorelasi lag ke- $k$ ,  $n$  merupakan jumlah data,  $Y_t$  merupakan nilai data dari *time series* yang sudah stasioner, dan  $\bar{Y}$  merupakan rata-rata peubah acak  $Y$ .

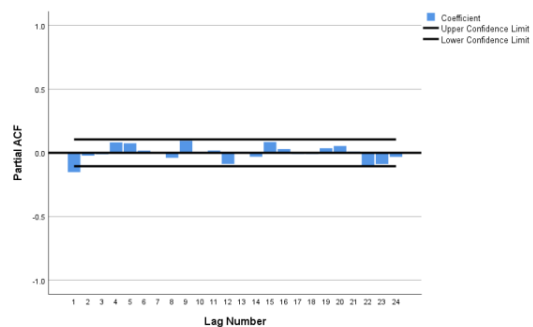
PACF merupakan korelasi antar deret pengamatan suatu deret waktu yang digunakan untuk mengetahui tingkat keeratan antara  $Y_t$  dan  $Y_{t-k}$ , apabila pengaruh dari waktu lag  $k$  dianggap terpisah. Selain untuk mengecek kestasioneran data, PACF juga digunakan untuk menentukan orde  $p$  pada model ARIMA  $(p,d,q)$ . Perumusan fungsi PACF dapat dituliskan sebagai berikut [10]:

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_p \rho_{k-p} \quad (3)$$

dimana  $\rho$  merupakan nilai fungsi dari ACF,  $\phi$  merupakan nilai fungsi dari PACF, dan  $k$  merupakan lag pada periode tertentu.



Gambar 2. Plot ACF



Gambar 3. Plot PACF

Gambar 2 dan 3 diatas merupakan contoh plot dari ACF dan PACF. Sumbu  $x$  pada gambar menyatakan lag pada periode tertentu. Sumbu  $y$  pada gambar menyatakan nilai dari ACF dan PACF pada lag tertentu. Kestasioneran suatu data *time series* bisa dilihat berdasarkan *cut off* pada plot ACF dan PACF. *Cut off* terjadi apabila data tiba-tiba mendekati nilai 0 pada lag tertentu. *Cut off* pada lag tertentu dalam plot ACF dan PACF juga menentukan orde  $q$  dan  $p$  secara berturut-turut pada model ARIMA  $(p,d,q)$ .

Jika data *time series* yang digunakan untuk diproses dengan ARIMA tidak stasioner, maka akan dilakukan *differencing*. *Differencing* dilakukan untuk mentransformasi data yang tak stasioner menjadi stasioner. *Differencing* menentukan orde  $d$  pada ARIMA  $(p,d,q)$ , yang nilainya sesuai dengan seberapa banyak *differencing* dilakukan. Proses *differencing* dilakukan dengan cara mengurangi nilai data pada suatu periode dengan nilai data periode sebelumnya yang dapat dirumuskan sebagai berikut [12]:

$$\Delta Y_t = Y_t - Y_{t-1} \quad (4)$$

dimana  $Y_t$  merupakan data pada waktu ke- $t$ , dan  $Y_{t-1}$  adalah data pada waktu  $t - 1$ . Jika data belum stasioner, maka akan dilakukan *differencing* kedua dan seterusnya hingga data menjadi stasioner.

## B. Genetic Algorithm

*Genetic Algorithm* (GA) merupakan salah satu cabang dari *Evolutionary Algorithm* yang berdasar pada teori seleksi alam. GA biasa digunakan untuk menyelesaikan masalah optimasi [13]. GA meniru perumpamaan evolusi biologis untuk menentukan kromosom atau individu berkualitas tinggi dalam suatu kawasan potensial yang disebut populasi. Populasi terdiri dari beberapa individu dimana setiap individu merepresentasikan sebuah solusi. Setiap populasi berisi beberapa kromosom, kromosom terdiri dari beberapa gen. Adapun langkah-langkah dari GA adalah sebagai berikut:

### 1. Pengkodean Kromosom

Pengkodean kromosom merupakan cara untuk mengkodekan gen dari suatu kromosom atau individu. Satu gen biasanya mewakili satu variabel. Gen dapat direpresentasikan dalam bentuk bilangan biner, bilangan *real*, elemen permutasi atau representasi lainnya yang dapat diimplementasikan untuk operator [14]. Pemilihan representasi gen bergantung pada permasalahan yang dihadapi. Pada penelitian ini, pengkodean dilakukan dengan menggunakan bilangan *real* sebagai representasi gen.

### 2. Inisialisasi Populasi

Tahap inisialisasi populasi merupakan tahap dimana dilakukan pembangkitan populasi awal berupa sekumpulan individu melalui prosedur tertentu. Individu yang dibangkitkan mewakili solusi dari permasalahan tertentu. Banyaknya populasi yang dibangkitkan bergantung pada masalah yang akan diselesaikan dan jenis genetika yang diimplementasikan.

### 3. Evaluasi Fitness

Tahap evaluasi *fitness* merupakan tahap dimana setiap individu dalam populasi akan dihitung nilai *fitness*-nya. *Fitness* merupakan nilai yang menyatakan baik tidaknya suatu individu. Semakin besar nilai *fitness* suatu individu maka semakin baik pula solusinya. Pada penelitian ini, fungsi *fitness* yang digunakan untuk mengevaluasi tiap individu adalah dengan menggunakan penghitungan *Root Mean Square Error* (RMSE). Fungsi *fitness* dapat dirumuskan sebagai berikut [5]:

$$f = \frac{1}{\varepsilon} \quad (5)$$

dimana  $f$  merupakan nilai *fitness* dan  $\varepsilon$  merupakan penghitungan RMSE.

### 4. Seleksi Induk

Pada tahap seleksi induk, individu yang sudah dievaluasi nilai *fitness*-nya akan dipilih untuk menjadi induk. Semakin tinggi nilai *fitness* suatu individu, maka probabilitas individu untuk dipilih akan semakin tinggi. Pada penelitian ini, seleksi yang dilakukan adalah dengan memilih dua individu dengan *fitness* tertinggi pada tiap populasi. Individu yang terpilih akan melakukan reproduksi untuk menghasilkan individu baru berupa anak pada populasi.

### 5. Pindah silang (crossover)

*Crossover* merupakan operator dalam GA yang melibatkan dua induk terpilih pada proses seleksi untuk menghasilkan individu yang baru. *Crossover* akan melakukan pertukaran nilai gen dari dua induk tergantung dari metode *crossover* yang digunakan. Pada penelitian ini, metode *crossover* yang digunakan adalah *extended intermediate crossover*. *Extended intermediate crossover* menghasilkan individu anak dari kombinasi nilai dua induk. *Extended intermediate crossover* dapat dirumuskan sebagai berikut [7]:

$$\begin{aligned} C_1 &= P_1 + \alpha(P_2 - P_1) \\ C_2 &= P_2 + \alpha(P_1 - P_2) \end{aligned} \quad (6)$$

dimana  $C_1$  dan  $C_2$  merupakan anak yang dihasilkan,  $P_1$  dan  $P_2$  merupakan individu yang terpilih menjadi induk, dan  $\alpha$  merupakan nilai acak dari interval 0 sampai 1. Banyak anak yang dihasilkan pada proses *crossover* bergantung pada probabilitas *crossover* ( $cr$ ) yang digunakan

#### 6. Mutasi

Mutasi merupakan operator dalam GA yang digunakan untuk mengubah nilai pada gen dalam suatu individu. Gen dalam suatu individu yang terpilih dalam operator mutasi akan menghasilkan individu yang baru. Perubahan nilai gen pada mutasi bergantung dari metode mutasi yang dilakukan. Pada penelitian ini, metode mutasi yang digunakan adalah *random mutation*. Pada *random mutation*, individu pada suatu populasi akan dipilih secara acak untuk dilakukan proses mutasi. Pada tiap individu terpilih, tiap gen pada individu akan mengalami mutasi. *Random mutation* dapat dirumuskan sebagai berikut [7]:

$$C_1 = P_1 + \alpha(max_i - min_i) \quad (7)$$

dimana  $C_1$  merupakan anak yang dihasilkan,  $P_1$  merupakan individu yang terpilih,  $\alpha$  merupakan nilai acak dari interval 0 sampai 1, dan  $max_i$  dan  $min_i$  secara berturut-turut merupakan nilai batas atas dan batas bawah *range* pada gen ke- $i$ . Banyak anak yang dihasilkan pada proses mutasi bergantung pada probabilitas mutasi ( $mr$ ) yang digunakan

#### 7. Penggantian Populasi

Pada tahap penggantian populasi, individu lama dalam populasi akan digantikan dengan individu yang baru untuk diproses ke generasi selanjutnya. Metode yang digunakan untuk menggantikan populasi adalah dengan membandingkan nilai *fitness* induk dan anak yang dihasilkan dari proses *crossover* dan mutasi. Apabila nilai *fitness* anak lebih baik, maka anak akan menggantikan induk pada populasi. Apabila nilai *fitness* induk lebih baik, maka induk akan dipertahankan pada populasi.

#### 8. Kriteria Terminasi

Apabila kriteria terminasi tercapai, maka proses GA akan berhenti. Dalam penelitian ini, kriteria terminasi yang digunakan adalah maksimum iterasi yang dilakukan oleh GA.

### C. Root Mean Square Error (RMSE)

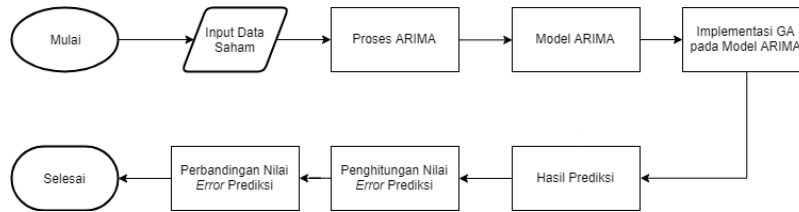
RMSE merupakan salah satu metode yang digunakan untuk mengukur tingkat akurasi hasil prediksi suatu model. RMSE dapat menyatakan ukuran besarnya kesalahan yang dihasilkan oleh suatu model prediksi. RMSE dapat dirumuskan sebagai berikut [11]:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (8)$$

dimana  $n$  merupakan banyak data,  $\hat{y}_t$  merupakan data hasil peramalan saat ke- $t$  dan  $y_t$  merupakan data aktual saat ke  $t$ .

## III. METODE PENELITIAN

Pemilihan metode yang digunakan pada penelitian ini adalah dengan melakukan prediksi harga saham menggunakan model ARIMA. Model ARIMA terbaik yang didapat akan diimplementasi dengan GA untuk menghasilkan prediksi harga saham. Hasil prediksi yang didapat dari dua metode yang dilakukan akan dihitung nilai errornya dan akan dibandingkan untuk mengetahui apakah implementasi GA pada model ARIMA dapat meningkatkan akurasi prediksi harga saham.



Gambar 4. Diagram Proses Penelitian

#### A. Input Data Saham

Data yang digunakan dalam penelitian ini adalah data harian harga penutupan saham dari PT Bank Central Asia Tbk (BBCA.JK) pada periode 1 Januari 2017 - 1 Januari 2019, dengan total sebanyak 471 data yang dibagi menjadi *training set* dan *testing set*. *Training set* yang digunakan adalah 65% dari jumlah data atau sekitar 302 *dataset*. *Testing set* yang digunakan adalah 35% dari jumlah data atau sekitar 169 *dataset*.

#### B. Proses ARIMA

Data yang digunakan dalam proses dan pembentukan model ARIMA adalah data *training set*.

##### 1. Cek Kestasioneran

Data yang digunakan akan diuji kestasionerannya. Stasioner atau tidaknya data dapat dilihat dari plot ACF dan PACF dengan melibatkan kondisi *cut off* grafik tersebut. Jika data belum stasioner, akan dilakukan *differencing*.

##### 2. Identifikasi Model Sementara

Identifikasi model dengan mengidentifikasi orde ARIMA  $(p,d,q)$  dimana untuk mencari orde  $(p,d,q)$  berturut-turut adalah dengan menggunakan PACF, *differencing*, dan ACF. Untuk mengetahui orde  $p$ , dapat dilihat pada grafik PACF yang *cut off*. Untuk mengetahui orde  $q$ , dapat dilihat pada grafik ACF yang *cut off*. Nilai orde  $d$  bergantung dari seberapa banyak *differencing* dilakukan.

##### 3. Estimasi Parameter

Estimasi parameter dilakukan dengan menguji semua kemungkinan model ARIMA yang diperoleh dari identifikasi model sementara. Pengujian dilakukan untuk mencari signifikansi dari model yang didapat.

##### 4. Uji Diagnostik

Pengujian diagnostik dilakukan dengan cara pengujian model yang sudah di estimasi yaitu terhadap nilai *error* pada model dengan statistik *Box-Pierce Q*.

#### C. Model ARIMA

Dari semua model yang didapatkan pada proses ARIMA, akan dipilih model ARIMA  $(p,d,q)$  yang terbaik berdasarkan *error* yang terkecil.

#### D. Implementasi GA pada model ARIMA

##### 1. Pengkodean Kromosom

Pengkodean kromosom yang digunakan adalah bilangan *real* sebagai representasi gen. Pembentukan gen pada kromosom dilakukan dengan membangkitkan nilai random antara -100 hingga 100. Nilai

random tersebut menyatakan koefisien pada tiap variabel dari model ARIMA. Banyak gen pada kromosom sesuai dengan banyak variabel dari model ARIMA.

2. *Inisiasi Populasi*

Populasi akan dibangkitkan dan banyak populasi yang akan dibangkitkan sesuai dengan jumlah yang ditentukan.

3. *Evaluasi Fitness*

Evaluasi dengan menghitung nilai *fitness* dari tiap individu pada populasi. Evaluasi *fitness* tiap individu dilakukan dengan menggunakan persamaan (5).

4. *Seleksi Induk*

Dua individu pada tiap populasi dipilih menjadi induk untuk melakukan *crossover*. Induk yang dipilih merupakan dua individu dengan *fitness* terbaik pada tiap populasi.

5. *Crossover*

Dua individu dengan *fitness* terbaik yang sudah dipilih pada tahap seleksi induk akan di *crossover* dengan menggunakan persamaan (6).

6. *Mutasi*

Individu dalam populasi secara acak dipilih untuk dilakukan mutasi. Tiap gen dalam individu yang terpilih akan dimutasi dengan menggunakan persamaan (7).

7. *Penggantian Populasi*

Penggantian populasi dilakukan dengan membandingkan nilai *fitness* induk dan anak yang dihasilkan dari proses *crossover* dan mutasi. Apabila nilai *fitness* anak lebih baik, maka anak akan menggantikan induk pada populasi. Apabila nilai *fitness* induk lebih baik, maka induk akan dipertahankan pada populasi.

8. *Terminasi*

Terminasi dilakukan ketika kriteria terminasi tercapai. Kriteria terminasi yang digunakan bergantung pada maksimum generasi yang ditentukan.

E. *Hasil Prediksi*

Pada tahap ini, akan didapatkan hasil prediksi *testing set* dari model ARIMA terbaik dan model ARIMA yang diimplementasikan dengan GA.

F. *Penghitungan Nilai Error Prediksi*

Hasil prediksi yang didapat dari model ARIMA terbaik dan model ARIMA yang diimplementasikan dengan GA, masing-masing akan dihitung nilai *error*nya dengan menggunakan RMSE.

G. *Perbandingan Nilai Error Prediksi*

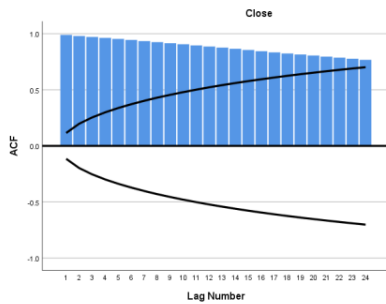
Nilai RMSE yang didapatkan dari model ARIMA dan model ARIMA yang diimplementasikan dengan GA akan dibandingkan untuk mengetahui apakah model ARIMA yang diimplementasikan GA dapat meningkatkan akurasi dari model ARIMA.



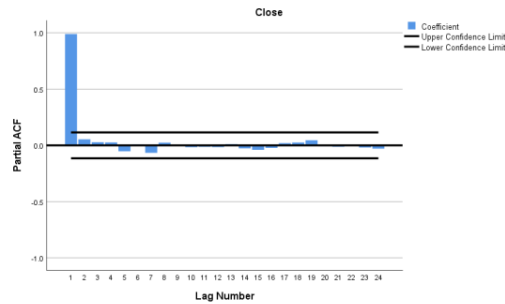
IV. HASIL DAN DISKUSI

A. Proses ARIMA

Data saham yang digunakan untuk mencari model ARIMA adalah data *training set* sebanyak 302 data.

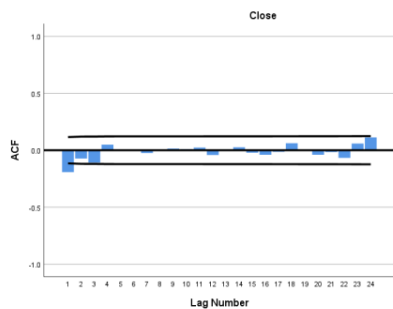


Gambar 5. Plot ACF Training Set

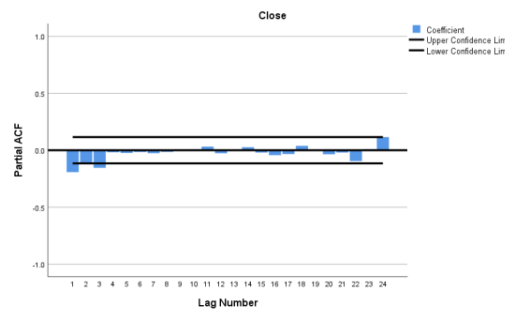


Gambar 6. Plot PACF Training Set

Dilihat dari Gambar 5 dan Gambar 6, data yang digunakan tidak stasioner karena plot ACF memperlihatkan trend linier yang jelas mengalami penurunan nilai autokorelasi secara perlahan-lahan. Karena data tidak stasioner, akan dilakukan *differencing* pertama.



Gambar 7. Plot ACF setelah differencing



Gambar 8. Plot PACF setelah differencing

Gambar 7 dan Gambar 8 menunjukkan ACF dan PACF dari data yang sudah diproses *differencing*. Dapat dilihat bahwa data sudah stasioner karena sudah terjadi *cut off*. Pada Gambar 7, plot ACF *cut off* di lag ke-1. Pada Gambar 8, plot PACF *cut off* di lag ke-3. Hal ini menunjukkan bahwa kemungkinan model ARIMA yang akan digunakan adalah model ARIMA ( $p,d,q$ ). Dilihat dari ACF dan PACF yang didapat, terdapat beberapa model ARIMA yang mungkin yaitu ARIMA dengan orde (1,0,0), (0,1,1), (3,1,1), (2,1,1), (1,1,1), (3,1,0), (2,1,0), dan (1,1,0). Setelah didapatkan model ARIMA yang memungkinkan, semua model akan diuji dengan estimasi parameter dan uji diagnostik. Hasil uji tersebut menghasilkan orde ARIMA yang memenuhi syarat dalam pembentukan model ARIMA, yaitu ARIMA dengan orde (1,0,0), (0,1,1), (2,1,1), (1,1,1), (3,1,0), (2,1,0), dan (1,1,0). Model ARIMA yang memenuhi syarat akan dilihat nilai RMSEnya untuk menentukan model mana yang akan digunakan dalam memprediksi harga saham.

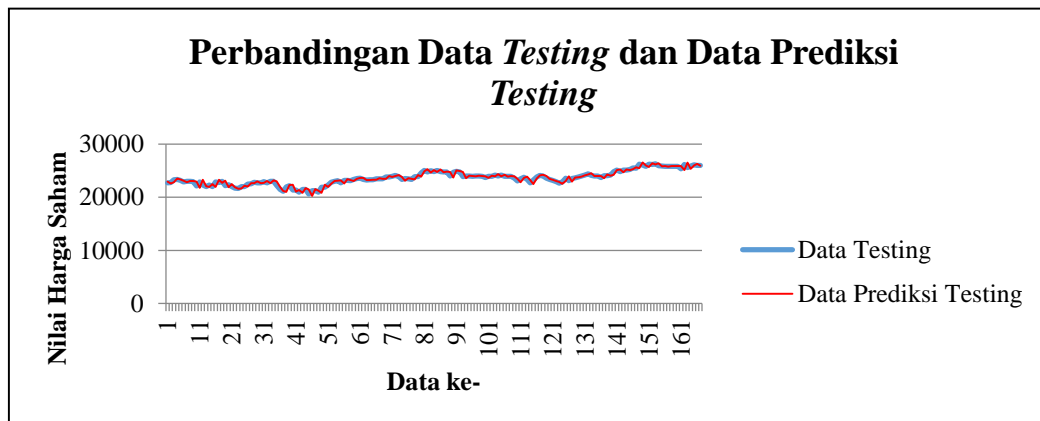
Tabel I  
 NILAI RMSE MODEL ARIMA

Model ARIMA	RMSE (Data Training)
(1,0,0)	315.614
(0,1,1)	228.017
(2,1,1)	226.025
<b>(1,1,1)</b>	<b>225.153</b>
(3,1,0)	225.972
(2,1,0)	228.334
(1,1,0)	229.448

Dapat dilihat pada Tabel I diatas bahwa nilai RMSE terkecil dari data *training* pada model ARIMA yang memungkinkan adalah model ARIMA dengan orde (1,1,1). Berdasarkan hasil tersebut, model ARIMA dengan orde (1,1,1) merupakan model yang terbaik yang akan digunakan untuk melakukan prediksi harga saham. Persamaan model ARIMA (1,1,1) dirumuskan sebagai berikut:

$$y_t = \mu + (1 + \phi_1)y_{t-1} - \phi_1y_{t-2} + \varepsilon_t - \theta_1\varepsilon_{t-1} \quad (9)$$

Persamaan model ARIMA (1,1,1) diatas didapat dari data *training* dan akan digunakan untuk memprediksi harga saham dari data *testing* yang berjumlah 169 *dataset*. Nilai parameter yang didapat untuk variabel  $\mu$ ,  $\phi_1$ ,  $\theta_1$  secara berturut-turut adalah 25.825, 0.432, dan 0.675.



Gambar 9. Perbandingan Data *Testing* dan Data Prediksi ARIMA

Berdasarkan Gambar 9, dapat dilihat bahwa data prediksi *testing* menggunakan persamaan (9) sudah mendekati data *testing* asli yang digunakan.

Tabel II  
 NILAI RMSE HASIL PREDIKSI ARIMA

Model ARIMA	RMSE (Data Prediksi Testing)
(1,1,1)	418.1314

Dapat dilihat dari Tabel II diatas bahwa nilai RMSE yang didapatkan dari prediksi data *testing* adalah sebesar 418.1314. Nilai tersebut menunjukkan bahwa model ARIMA (1,1,1) dapat digunakan untuk melakukan prediksi harga saham karena jika dilihat dari Gambar 9, hasil prediksi yang didapat sudah

mendekati data asli yang digunakan. RMSE yang didapat dari prediksi harga saham menggunakan model ARIMA masih cukup besar, oleh karena itu akan digunakan metode GA yang akan diimplementasikan pada model ARIMA terbaik untuk meningkatkan akurasi prediksi harga saham.

#### B. Proses GA - ARIMA

Model ARIMA terpilih dari proses ARIMA yang telah dilakukan, yaitu ARIMA dengan orde (1,1,1) akan diimplementasikan dengan proses GA. Proses GA akan mencari nilai koefisien tiap variabel pada model ARIMA (1,1,1). Data yang digunakan pada proses GA adalah data *testing* sebanyak 169 data.

Tabel III  
PARAMETER GA

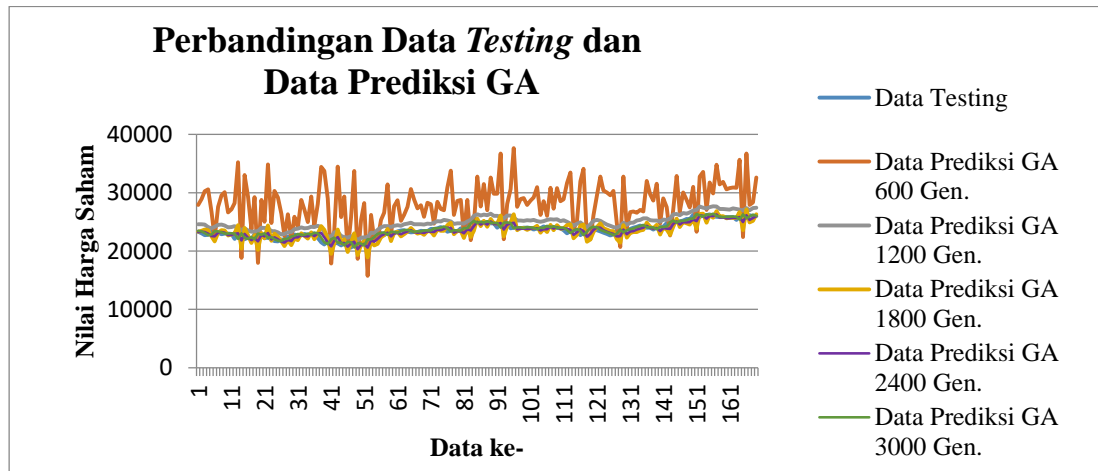
Parameter GA	Nilai
Jumlah Individu pada Populasi	200
Maksimum Generasi	600, 1200, 1800, 2400, dan 3000
Pengkodean Kromosom	Bilangan <i>Real</i>
Fungsi <i>Fitness</i>	1/RMSE
Seleksi Induk	Dua individu terbaik pada tiap populasi
<i>Crossover</i>	<i>Extended Intermediate Crossover</i>
Peluang <i>Crossover</i>	0.7
Mutasi	<i>Random Mutation</i>
Peluang Mutasi	0.3
Penggantian Populasi	Perbandingan <i>fitness</i> anak dan induk

Tabel III diatas menunjukkan parameter-parameter yang digunakan saat melakukan proses GA. Proses GA dilakukan dengan lima skenario yang berbeda, yaitu dengan menggunakan maksimum generasi sebanyak 600, 1200, 1800, 2400, dan 3000.

Tabel IV  
HASIL SKENARIO PROSES GA

Maksimum Generasi	Jumlah Individu pada Populasi	Peluang Crossover	Peluang Mutasi	Nilai <i>Fitness</i> Individu Terbaik
600	200	0.7	0.3	0.000042682
1200	200	0.7	0.3	0.00017609
1800	200	0.7	0.3	0.00025884
2400	200	0.7	0.3	0.00048815
3000	200	0.7	0.3	0.00076126

Tabel IV diatas menunjukkan nilai *fitness* individu terbaik tiap skenario yang digunakan saat melakukan proses GA. Dapat dilihat bahwa semakin besar maksimum generasi, maka nilai *fitness* individu terbaik akan semakin besar. Semakin tinggi nilai *fitness*, maka data prediksi harga saham dengan implementasi GA akan semakin baik.



Gambar 10. Perbandingan Data *Testing* dan Data Prediksi GA

Gambar 10 diatas menunjukkan perbandingan data *testing* asli dengan hasil prediksi GA pada data *testing*. Dapat dilihat dari grafik bahwa implementasi GA pada model ARIMA dengan skenario 3000 generasi sudah mendekati data *testing* asli yang digunakan. Gambar 10 juga menunjukkan bahwa semakin besar maksimum generasi yang digunakan, maka prediksi harga saham dengan implementasi GA akan semakin mendekati data *testing* asli dan berlaku sebaliknya. Hal ini juga berhubungan dengan nilai *fitness* yang didapatkan pada Tabel IV, yang membuktikan pernyataan bahwa semakin tinggi nilai *fitness*, maka prediksi data harga saham akan semakin baik.

Tabel V  
 NILAI RMSE HASIL PREDIKSI GA

Maksimum Generasi	RMSE (Data Testing)
600	5827.738
1200	1319.903
1800	1080.704
2400	563.7984
3000	<b>371.0107</b>

Dari Tabel V dapat dilihat bahwa dari proses GA dengan lima skenario yang dilakukan, GA dengan 3000 generasi memberikan hasil *error* terkecil yaitu dengan nilai RMSE sebesar 371.0107. Selain itu dapat dilihat dari Tabel V bahwa semakin besar maksimum generasi, maka hasil *error* akan semakin kecil.

### C. Perbandingan Nilai Error

Tabel VI  
PERBANDINGAN NILAI ERROR ARIMA DAN ARIMA-GA

Metode	RMSE (Data Testing)
ARIMA (1,1,1)	418.1314
ARIMA – GA 600 Generasi	5827.738
ARIMA – GA 1200 Generasi	1319.903
ARIMA – GA 1800 Generasi	1080.704
ARIMA – GA 2400 Generasi	563.7984
ARIMA – GA 3000 Generasi	<b>371.0107</b>

Tabel diatas menunjukkan nilai *error* yang didapatkan dari pengujian metode-metode yang dilakukan pada penelitian ini. Berdasarkan tabel diatas, nilai RMSE dari prediksi harga saham dengan menggunakan model ARIMA (1,1,1) adalah 418.1314. Sedangkan untuk hasil prediksi dengan mengimplementasikan GA pada model ARIMA (1,1,1) dengan 600 generasi, 1200 generasi, 1800 generasi, 2400 generasi, dan 3000 generasi masing-masing memiliki nilai RMSE dari prediksi harga saham sebesar 5827.738, 1319.903, 1080.704, 563.7984, dan 371.0107. Hasil yang didapat menunjukkan bahwa pengimplementasian GA pada ARIMA dengan 3000 generasi memiliki nilai *error* paling minimum yaitu dengan nilai RMSE sebesar 371.0107. Metode ARIMA – GA lebih efektif dari model ARIMA jika dilihat dari hasil RMSEnya yang mengalami penurunan. Efisiensi metode ARIMA – GA sedikit berkurang jika dibandingkan dengan metode ARIMA dikarenakan hasil yang didapatkan untuk mencapai peningkatan akurasi memakan waktu yang cukup lama dalam pengaplikasiannya.

### V. KESIMPULAN

Berdasarkan hasil pengujian dan analisis yang dilakukan mengenai “Implementasi Genetic Algorithm dalam Model ARIMA untuk Memprediksi Observasi Time Series”, dapat diambil kesimpulan bahwa dari hasil perbandingan *error* semua metode yang digunakan, metode GA yang diimplementasikan pada model ARIMA terbaik dengan skenario maksimum 3000 generasi menghasilkan nilai *error* terkecil yaitu dengan nilai RMSE sebesar 371.0107. Hal ini menunjukkan bahwa implementasi GA pada model ARIMA dapat meningkatkan akurasi prediksi harga saham jika parameter yang digunakan sesuai dengan permasalahan. Nilai RMSE yang didapat dengan menggunakan metode pada penelitian ini dapat diperkecil nilainya jika diperbanyak skenario pengujiannya, dengan merubah parameter-parameter GA yang digunakan.

### DAFTAR PUSTAKA

- [1] A. A. Adebisi, A. O. Adewumi & C. K. Ayo. Stock price prediction using the ARIMA model. *Proceedings – UKSim-AMSS 16<sup>th</sup> International Conference on Computer Modelling and Simulation, UKSim 2014*, 106-112, 2014.
- [2] A. Edward. Forecast Model Using ARIMA for Stock Prices of Automobile Sector. *International Journal of Research in Finance and Marketing*, 6(4), 1-9, 2016.
- [3] Y. Wen Wang, Z. Zhou Shen, & Y. Jiang. Comparison of ARIMA and GM (1,1) models for prediction of hepatitis B in China. *PLoS ONE*, 13(9), 1–11, 2018.
- [4] Haldurai Lingaraj. A Study on Genetic Algorithm and its Applications. *International Journal of Computer Sciences and Engineering*, 4, 139-143, 2016.

- [5] A. Rahmi, W.F. Mahmudy & B.D. Setiawan. Prediksi Harga Saham Berdasarkan Data Historis Menggunakan Model Regresi Yang Dibangun Dengan Algoritma Genetika. *DORO: Repository Jurnal Mahasiswa PTIK Universitas Brawijaya*, 5(12), 1-9, 2015.
- [6] G. Bonde & Reham Khaled. Stock price prediction using genetic algorithms and evolution strategies. *Proceedings of the 2012 International Conference on Genetic and Evolutionary Methods*, 10-15, 2012.
- [7] I. Wahyuni, W.F. Mahmudy & A. Iriany. Rainfall Prediction in Tengger region Indonesia using Tsukamoto fuzzy inference system. *Proceedings – 2016 1<sup>st</sup> International Conference on Information Technology, Information Systems and Electrical Engineering, ICITISEE 2016*, 130-135, 2017.
- [8] S. Green. Time Series Analysis of Stock Prices Using the Box-Jenkins Approach. *Electronic Theses & Dissertations*, 668, 2011.
- [9] R. Stoean, C. Stoean & A. Sandita. Evolutionary Regressor Selection in ARIMA Model for Stock Price Time Series Forecasting. In: Czarnowski I., Howlett R., Jain L. (eds) *Intelligent Decision Technologies 2017, IDT 2017, Smart Innovation, Systems and Technologies*, vol 73, Springer, Cham, 2018.
- [10] S. Makridakis, S. C. Wheelwright & V.E. McGee. *Metode dan Aplikasi Peramalan Jilid 1 Edisi Kedua (Vol. 61 & 339)*. Jakarta: Erlangga, 1999.
- [11] G. James, D. Witten, T. Hastie & R. Tibshirani. *An Introduction to Statistical Learning: With Applications in R*. New York: Springer New York, 2013.
- [12] S. H. A. Salmon, N. Nainggolan & D. Hatidja. *Pemodelan ARIMA Dalam Prediksi Penumpang Pesawat Terbang Pada Bandara Internasional Sam Ratulangi Manado*, 2014.
- [13] H. A. Saputro, W.F. Mahmudy & C. Dewi. Implementasi Algoritma Genetika Untuk Optimasi Penggunaan Lahan Pertanian. *Jurnal Mahasiswa PTIK Universitas Brawijaya*, 5(12), 12, 2015.
- [14] Entin. 2011. Kecerdasan Buatan: Bab Algoritma Genetika. Tersedia pada: <http://entin.lecturer.pens.ac.id/Kecerdasan%20Buatan/Buku/Bab%207%20Algoritma%20Genetika.pdf>.