

# Perbandingan Algoritma SARIMA dan Linear Regression dalam Memprediksi Indeks Harga Saham Gabungan

Luvi Nur Rupaiah\*, Muhammad Himi Syarif, Ultach Enri  
Universitas Singaperbangsa Karawang, Indonesia  
\*E-mail koresponden: [luvi.nur18050@student.unsika.ac.id](mailto:luvi.nur18050@student.unsika.ac.id)

*Diserahkan 1 November 2021; Direview 20 November 2021; Dipublikasikan 30 November 2021*

## Abstrak

*Prediksi harga saham adalah suatu hal yang menjadi daya tarik investor terhadap pasar saham. Pada perdagangan saham, pengamatan mobilitas Indeks Harga Saham Gabungan (IHSG) sangat dibutuhkan. Berubahnya nilai IHSG yang akan datang merupakan dasar dibutuhkannya metode prediksi yang dapat menentukan acuan untuk melakukan pengambilan keputusan oleh para investor. Dalam memproses data time series dapat dilakukan dengan berbagai macam metode analisis yang bertujuan untuk memprediksi kejadian yang akan datang. Penelitian ini dilakukan untuk membandingkan algoritma Seasonal Autoregressive Integrated Moving Average (SARIMA) dan Linear Regression dengan dataset publik untuk mendapatkan performa terbaik dari kedua algoritma tersebut. Dataset yang diuji bersumber dari Jakarta Composite Index (JKSE) dengan dataset time series dari tanggal 1 Januari 2019 hingga 23 Desember 2021. Hasil penelitian ini menyimpulkan bahwa indeks harga saham gabungan dengan performa terbaik adalah menggunakan Linear Regression.*

**Kata kunci:** *IHSG, Linear Regression, Prediksi, SARIMA, Time Series*

## Abstract

*Prediction of stock prices is something that attracts investors to the stock market. In stock trading, it is very important to observe the mobility of the Composite Stock Price Index (IHSG). Changes in the value of the JCI in the future is the basis for the need for a prediction method that can determine the reference for making decisions by investors. Processing time series data can be done with various analytical methods that aim to predict future events. This research was conducted to compare the SARIMA and Linear Regression algorithms with public datasets to get the best performance of the two algorithms. The dataset tested is sourced from the Jakarta Composite Index (JKSE) with a time series dataset from January 1, 2019 to December 23, 2021. The results of this study conclude that the composite stock price index with the best performance is using Linear Regression.*

**Keywords:** *JCI, Linear Regression, Prediction, SARIMA, Time Series*

## PENDAHULUAN

Pesatnya perkembangan teknologi saat ini berpengaruh besar terhadap pekerjaan masyarakat [1]. Pada waktu ini saham adalah jenis produk investasi yang terkenal dan diminati oleh perusahaan juga dikalangan warga. Saham merupakan salah satu pasar keuangan yang sedang naik daun [2]. Bergeraknya harga pasar modal secara acak, naik turunnya harga saham dipengaruhi oleh beberapa faktor seperti kondisi perekonomian, laju inflasi, penawaran dan permintaan serta masih banyak lagi [3]. Dengan demikian, proses peramalan Indeks Harga Saham campuran (IHSG) adalah hal yang selalu menarik bagi para investor di pasar saham dalam perencanaan serta pengambilan keputusan. Hasil prediksi dapat dijadikan referensi bagi para investor untuk menunjang aktivitas mereka diperdagangan serta investasi di pasar saham [4]. Salah satu model data mining sudah digunakan serta dikembangkan untuk memprediksi harga saham [14]. Data mining merupakan proses mengolah data mentah menjadi informasi yang berguna, dan juga untuk menggali nilai tambahan berupa informasi yang selama ini tidak diketahui secara manual dari suatu basis data [13]. Proses data mining berjalan secara semi otomatis serta dipergunakan untuk membantu pengambilan keputusan [5]. Hasil pengolahan data dengan metode data mining dapat digunakan untuk membuat keputusan pada masa mendatang [15].

Harapan dari setiap investor ialah menerima suatu keuntungan, bergantung di setiap harga pada masing-masing harga saham yang selalu berubah dari waktu ke waktu, atau biasa dikenal menjadi data runtut waktu (*time series*) [6]. *Time series* adalah rangkaian data berupa nilai pengamatan yang diukur jangka waktu tertentu, berdasarkan waktu dengan interval yang sama [7]. Prediksi data *time series* yang cukup pesat menyebabkan banyaknya pilihan metode untuk digunakan dalam memprediksi data disesuaikan berdasarkan kebutuhan, sebagai akibatnya perlu dibandingkannya antar metode untuk mendapatkan hasil performa terbaiknya dan dengan *error* yang terkecil. Penelitian ini prosedur pemecahan *Seasonal Autoregressive Integrated Moving Average* (SARIMA) dan *Linear Regression* yang menjadi pilihan pada prediksi.

Algoritma SARIMA atau *Seasonal ARIMA* adalah salah satu model dari *Box-Jenkins* yang memanfaatkan sikap dari data yang diamati serta adanya faktor musiman di data [8]. Untuk mendapatkan hasil yang sempurna dan seksama, banyak dikembangkan metode prediksi yaitu model SARIMA yang didasarkan beberapa aspek diantaranya aspek waktu, aspek taraf keakuratan prediksi yang diinginkan, serta aspek model yang ingin diamati [9]. *Seasonal ARIMA* menggunakan nilai lampau serta nilai saat ini yang berasal dari variabel *dependent* untuk menghasilkan peramalan jangka pendek yang akurat, model yang secara penuh mengabaikan variabel *independent* dalam membuat peramalan. Salah satu syarat utama dalam melakukan analisis ARIMA adalah data yang digunakan harus bersifat stasioner baik terhadap rata-rata maupun varians [10].

*Linear Regression* yaitu pemodelan serta analisis data numerik yang terdiri dari satu atau lebih variabel *independent* dan nilai variabel *dependent* [12]. Analisis *regression* umumnya dipergunakan untuk *forecasting* dan *prediction*, dimana aplikasi yang banyak dikembangkan menggunakan *machine learning*. Analisis *regression* juga dipergunakan dalam beberapa kasus untuk menentukan hubungan sebab akibat antara *independent* dan *dependent variable*. *Regression* sendiri hanya menunjukkan hubungan antara variabel *dependent* serta kumpulan dataset tetap dari variabel yang berbeda. *Linear Regression* terdiri dari *Simple Linear Regression*, *Multivariate Linear Regression*, dan *Polynomial Regression*.

Penelitian mengenai prediksi dengan menggunakan model algoritma yang mendukung penelitian ini telah banyak dikembangkan. Contohnya aplikasi peramalan penjualan menggunakan algoritma *Linear Regression* [17]. Hasil yang diperoleh pada penelitian tersebut menunjukkan hasil yang didapatkan ( $y$ ) pada hari ke  $x$  adalah sama, artinya peramalan penjualan

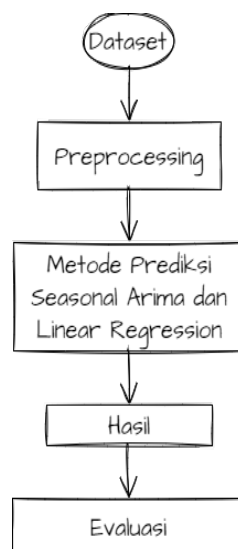
pada sistem sudah bisa dianggap valid karena hasilnya sama dengan yang diuji coba pada *Microsoft Excel*.

Penelitian dalam mengkomparasi *Artificial Neural Network* dan *Linear Regression* dalam memprediksi harga saham dengan pertimbangan faktor fundamental pada sektor industri [18]. Memperoleh hasil yang menunjukkan bahwa nilai *measure square error* dari hasil prediksi menggunakan *Artificial Neural Network* diperoleh nilai lebih kecil dibandingkan menggunakan *Linear Regression*. Penelitian metode SARIMA digunakan untuk membantu memprediksi harga saham dalam 3 tahun ke depan [19].

Penelitian mengembangkan dari penelitian yang ada dengan tujuan untuk membandingkan metode SARIMA dan *Linear Regression* berdasarkan nilai akurasi prediksi harga saham gabungan. Metode dengan performa yang baik dalam memprediksi diharapkan dapat menjadi acuan untuk memberikan referensi para investor mengenai harga saham selanjutnya.

## METODE PENELITIAN

Metode Penelitian yang digunakan pada penelitian kali ini dapat dilihat pada Gambar 1, *dataset* yang didapat akan melalui tahap *preprocessing* yaitu *cleaning*, *transformasi* dan *reduction*. Hasil dari tahap *preprocessing* akan digunakan untuk ke tahap selanjutnya yaitu membuat model masing-masing dari metode SARIMA dan *Linear Regression* yang kemudian dilakukan evaluasi.



Gambar 1. Metode Penelitian

### Dataset

Dataset yang dipakai adalah data Indeks Harga Saham Gabungan (IHSG) yang bersumber dari *Jakarta Composite Index* (JKSE) (<https://finance.yahoo.com/quote/%5EJKSE/history/>) dengan dataset *time series* dari tanggal 1 Januari 2019 hingga 23 Desember 2021 dan jumlah data 740 *record* seperti yang terdapat pada Tabel 1. Penelitian ini menggunakan data *training* dan *testing* dengan perbandingan 6:4 atau data training berjumlah 399 data dan *testing* berjumlah 267 data.

### Preprocessing

Pada tahap *preprocessing* ini data yang sudah diekstrak lalu diproses pada setiap anomali [11]. *Preprocessing* dilakukan proses *cleaning* buat menghilangkan data *missing value*, melakukan transformasi data dari *non-stationary* ke *stationary* data agar mampu menggunakan prosedur pemecahan SARIMA. Data dibagi menjadi dua yaitu data pelatihan dan data pengujian.

Tabel 1. Dataset *Timeseries*

	Date	Open	High	Low	Close	Adj Close	Volume
0	2019-01-02	6197.871094	6205.895020	6164.833984	6181.174805	6181.174805	52797800.0
1	2019-01-03	6176.151855	6221.009766	6176.151855	6221.009766	6221.009766	72166700.0
2	2019-01-04	6211.096191	6274.540039	6200.854004	6274.540039	6274.540039	80858100.0
3	2019-01-07	6317.625977	6354.757813	6287.224121	6287.224121	6287.224121	90278300.0
4	2019-01-08	6292.263184	6316.240234	6262.847168	6262.847168	6262.847168	90537400.0
...	...	...	...	...	...	...	...
735	2021-12-16	6648.288086	6661.944824	6594.797852	6594.797852	6594.797852	185929900.0
736	2021-12-17	6584.887207	6609.044922	6601.932129	6601.932129	6601.932129	210214300.0
737	2021-12-20	6571.345215	6579.128906	6547.111816	6547.111816	6547.111816	181341400.0
738	2021-12-21	6545.537109	6579.892090	6554.309082	6554.309082	6554.309082	186470700.0
739	2021-12-22	6572.535156	6529.592773	6529.592773	6529.592773	6529.592773	187213900.0

### Metode Prediksi/*forecasting*

Metode yang digunakan ialah SARIMA dan Linear Regression untuk prediksi. *Seasonal ARIMA* menggunakan nilai lampau serta nilai saat ini yang berasal dari variabel *dependent* untuk menghasilkan peramalan jangka pendek yang akurat, model yang secara penuh mengabaikan variabel *independent* dalam membuat peramalan. Model SARIMA yang didasarkan beberapa aspek diantaranya aspek waktu, aspek taraf keakuratan prediksi yang diinginkan, serta aspek model yang ingin diamati. *Linear Regression* yaitu pemodelan serta analisis data numerik yang terdiri dari satu atau lebih variabel *independent* dan nilai variabel *dependent*. Analisis *regression* umumnya dipergunakan untuk *forecasting* dan *prediction*, dimana aplikasi yang banyak dikembangkan menggunakan *machine learning*. *Linear Regression* terdiri dari *Simple Linear Regression*, *Multivariate Linear Regression*, dan *Polynomial Regression*.

### Evaluasi

Pada tahap ini, dilakukan pengukuran tingkat akurasi menggunakan contoh penilaian *Coefficient of Determination (R<sup>2</sup>)*, *Mean Absolute Error (MAE)*, *Mean Square Error (MSE)*, dan *Mean Absolute Percentage Error (MAPE)*. Semakin kecil nilai contoh evaluasi maka semakin baik tingkat akurasi prediksinya. Keakuratan sebuah model prediksi ketika melakukan peramalan dipengaruhi nilai awal terkecil oleh masing-masing metode akurasi data.

## HASIL DAN PEMBAHASAN

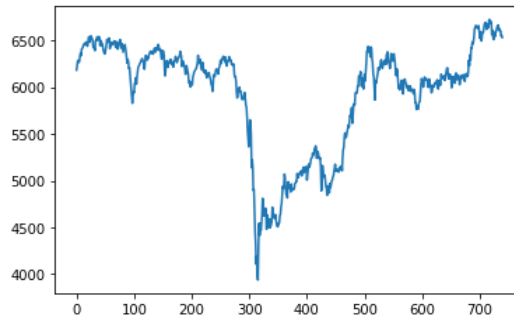
### Pengelolaan Data Awal

Data harga saham JKSE yang didapatkan dari *Yahoo Finance* dilakukan proses *preprocessing* terlebih dahulu agar data bermakna dan mudah dalam proses pemodelan yang akan dilakukan.

Tahapan dalam *preprocessing* ada tiga yaitu pertama data *cleaning* dilakukan untuk membersihkan data yang bernilai kosong atau *null value*. Pada dataset yang diperoleh sebelumnya terdapat *null value* yang dapat menghambat proses prediksi, sehingga harus dilakukan pembersihan, dengan hasil dalam bentuk grafik seperti pada Gambar 2.

Kedua yaitu data *transformation* dimana pada tahap ini dilakukan untuk memilih atribut mana yang akan dilakukan proses prediksi dan akan ada penambahan atribut pada dataset, yang pada kasus ini adalah atribut 'Close' merupakan atribut utama dalam proses prediksi, lalu atribut tersebut akan diubah menjadi 'JKSE' dan ditambahkan atribut 'JKSE\_sdff'. Atribut tersebut didalamnya terdapat data yang telah dilakukan proses transformasi.

Tahap Ketiga data *reduction* data dilakukan pemeriksaan jenis data yaitu *stationary* atau bukan. Cara pemeriksaan dataset menggunakan metode *statistic Augmented Dickey-Fuller*, yang memperoleh hasil seperti pada Gambar 3.



Gambar 2. Grafik *cleaning* data harga saham JKSE

Nilai statistik pada Gambar 3 terdapat *Test Statistic* sebesar -1,679 dan pada *Critical Value* (5%) sebesar -2,865, serta *p-value* sebesar 0,044 sehingga data tersebut bisa dikatakan data yang *non-stationary* dan perlu adanya tindak lanjut agar bisa dilakukan ke proses pemodelan. Data yang *non-stationary* akan diubah menjadi *stationary* dengan cara *seasonal differencing* (Persamaan 1).

$$y_t' = y_t - y_{(t-n)} \tag{1}$$

Setelah dilakukan proses *seasonal differencing* didapatkan nilai statistik seperti pada Gambar 4.

```

Results of Dickey-Fuller Test:
Test Statistic          -1.679020
p-value                 0.441953
#Lags Used              5.000000
Number of Observations Used  660.000000
Critical Value (1%)     -3.440297
Critical Value (5%)     -2.865929
Critical Value (10%)    -2.569107
    
```

Gambar 3. Uji *stationary* dengan menggunakan *augmented dickey-fuller*

```

Results of Dickey-Fuller Test:
Test Statistic          -2.905508
p-value                 0.044711
#Lags Used              16.000000
Number of Observations Used  649.000000
Critical Value (1%)     -3.440466
Critical Value (5%)     -2.866004
Critical Value (10%)    -2.569147
    
```

Gambar 4. Nilai statistik *seasonal differencing*

Pada nilai statistik yang baru (Gambar 4), didapatkan nilai *Test Statistic* sebesar -2,905. Nilai *Critical Value* (5%) sebesar -2,866, dan *p-value* sebesar 0,044 sehingga data yang telah dilakukan reduksi tersebut berubah menjadi *stationary* dan dapat dilanjutkan ke proses pemodelan. Gambar 5 adalah visualisasi data setelah *preprocessing*.

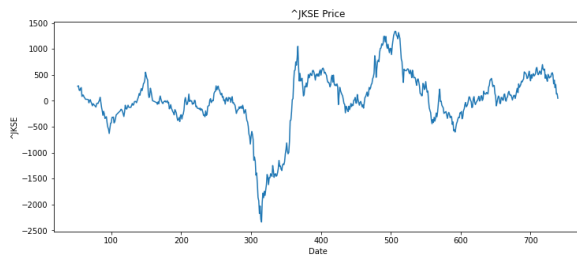
Adapun hasil dari *preprocessing* data indeks harga saham gabungan ditunjukkan pada Tabel 2.

Tabel 2. Hasil *preprocessing* data harga saham

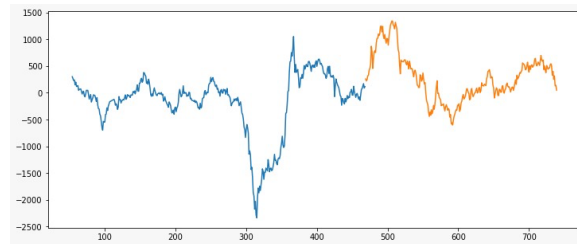
	Data	JKSE	JKSE_sdiff
52	2019-03-15	6461.183105	280.008300
53	2019-03-18	6509.446777	288.437011
54	2019-03-19	6480.275879	205.735840
55	2019-03-20	6482.729961	195.485840
56	2019-03-21	6501.775879	238.928711
...	...	...	...
735	2021-12-16	6594.797852	252.111817
736	2021-12-17	6601.932129	313.883301
737	2021-12-20	6547.111816	129.788574
738	2021-12-21	6554.309082	137.913086
739	2021-12-22	6529.592773	47.823730

## Pemodelan dan Evaluasi

Pemodelan dilakukan dengan *tools Python* dan algoritma SARIMA serta *Linear Regression*. Sebelum dilakukan pemodelan dataset dibagi menjadi 6:4. Perbandingan grafik ditunjukkan seperti Gambar 6.



Gambar 5. Visualisasi Data Setelah Preprocessing



Gambar 6. Training dan Testing

Setelah data dibagi selanjutnya dilakukannya proses pemodelan dan dilakukan proses evaluasi menggunakan *Mean Square Error (MSE)*, *Coefficient of Determination (R2)*, *Mean Absolute Error (MAE)*, dan *Mean Absolute Percentage Error (MAPE)*. Hasil akurasi prediksi dari model setiap algoritma ditunjukkan seperti Tabel 3.

Tabel 3. Hasil Evaluasi

Algoritma	Model Evaluasi			
	R2	MAE	MSE	MAPE
SARIMA	-0,972035	777,913481	697519,256554	0,134033
Linear Regression	0,961652	74,284642	13364,207136	0,013279

Berdasarkan hasil evaluasi perbandingan algoritma *Linear Regression* lebih baik karena nilai R2, MAE, MSE dan MAPE lebih kecil dari SARIMA dengan tingkat *error* yang lebih tinggi. Hal itu membuktikan bahwa prediksi *Linear Regression* dapat membantu peramalan harga saham dengan tingkat *error* yang lebih rendah.

## KESIMPULAN

Penelitian ini berhasil membandingkan algoritma SARIMA dan *Linear Regression* pada data harga saham JKSE yang berbentuk *time series* dari tanggal 1 Januari 2019 hingga 23 Desember 2021 dengan mengukur nilai R2, MAE, MSE, dan MAPE. Setelah dilakukan proses training dan testing, hasil yang diperoleh adalah algoritma *Linear Regression* memiliki performa terbaik dibandingkan SARIMA dalam memprediksi harga saham JKSE, dibuktikan dengan nilai uji coba lebih kecil asal SARIMA yaitu R2 0.961652, MAE 74.284642, MSE 13364.207136 serta MAPE 0.013279, sehingga dengan prediksi *Linear Regression* dapat membantu dalam peramalan harga saham JKSE di pasar modal.

## DAFTAR PUSTAKA

1. Sugianto, Castaka Agus, and Faishal Fachruddin. "Prediksi Pergerakan Harga Valas Menggunakan Algoritma Neural Network." *Jurnal Informatika: Jurnal Pengembangan IT* 3.1 (2018): 20-25.
2. Sofi, Khalis, et al. "Perbandingan Algoritma Linear Regression, LSTM, dan GRU dalam Memprediksi Harga Saham dengan Model Time Series." *SEMINASTIKA* 3.1 (2021): 39-46.

3. Maulana, Reza, and Devy Kumalasari. "Analisis dan Perbandingan Algoritma Data Mining dalam Prediksi Harga Saham GGRM." (2019): 22-28.
4. Susanti, Riana, and Askardiya Radmoyo Adji. "Analisis Peramalan IHSG Dengan Time Series Modeling ARIMA." *Jurnal Manajemen Kewirausahaan* 17.1 (2020): 97-106.
5. Valiant, Kevin, Yuan Lukito, and R. Gunawan Santosa. "Sistem Prediksi Harga Saham LQ45 Dengan Random Forest Classifier." *Jurnal Terapan Teknologi Informasi* 3.2 (2019): 127-136.
6. Ramdhani, Yudi, and Ade Mubarak. "Analisis Time Series Prediksi Penutupan Harga Saham Antm. Jk Dengan Algoritma SVM Model Regresi." *Jurnal Responsif: Riset Sains & Informatika* 1.1 (2019): 77-82.
7. Pandji, Bagas Yafitra, Indwiarti Indwiarti, and Aniq Atiqi Rohmawati. "Perbandingan Prediksi Harga Saham dengan model ARIMA dan Artificial Neural Network." *Indonesian Journal on Computing (Indo-JC)* 4.2 (2019): 189-198.
8. Hendayanti, Ni Putu Nanik, and Maulida Nurhidayati. "Perbandingan metode Seasonal Autoregressive Integrated Moving Average (SARIMA) dengan Support Vector Regression (SVR) dalam memprediksi jumlah kunjungan wisatawan mancanegara ke Bali." *Jurnal Varian* 3.2 (2020): 149-162.
9. Soekendro, Cesar Aini, Edi Sutoyo, and Faqih Hamami. "Prediksi Curah Hujan Di Kab. Bandung Dengan Analisis Time Series, Menggunakan Model SARIMA (seasonal Autoregressive Integrated Moving Average)." *eProceedings of Engineering* 8.2 (2021).
10. Dewi, Dyah Makutaning, M. Zaky Nafi, and Nasrudin Nasrudin. "Analisis Peramalan Harga Emas di Indonesia pada Masa Pandemi Covid-19 untuk Investasi." *Jurnal Litbang Sukowati: Media Penelitian dan Pengembangan* 5.2 (2022): 38-50.
11. Halimi, Imam, and Wahyu Andhyka Kusuma. "Prediksi Indeks Harga Saham Gabungan (IHSG) Menggunakan Algoritma Neural Network." *Jurnal Edukasi dan Penelitian Informatika (JEPIN)* 4.1 (2018): 24-29.
12. Bode, Andi. "Perbandingan Metode Prediksi Support Vector Machine Dan Linear Regression Menggunakan Backward Elimination Pada Produksi Minyak Kelapa." *Simtek: jurnal sistem informasi dan teknik komputer* 4.2 (2019): 104-107.
13. Boy, Ahmad Fitri. "Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara)." *JOURNAL OF SCIENCE AND SOCIAL RESEARCH* 3.2 (2020): 78-85.
14. Nurachim, Rusma Insan. "Pemilihan Model Prediksi Indeks Harga Saham Yang Dikembangkan Berdasarkan Algoritma Support Vector Machine (SVM) Atau Multilayer Perceptron (MLP) Studi Kasus: Saham Pt Telekomunikasi Indonesia Tbk." *Jurnal Teknologi Informatika dan Komputer* 5.1 (2019): 29-35.
15. Firnando, Irvan, et al. "Implementasi Algoritma Apriori Dan Forecasting pada Transaksi Penjualan." *Jurnal Mantik Penusa*. Vol. 3 No.3. 2019.
16. Lussa, M. O., & Marie, I. A. (2019). Pemanfaatan Artificial Neural Network dan Fuzzy Inventory Model untuk Penentuan Persediaan Pengaman. *Krea-TIF*, 7(2), 60-71.
17. Rival, Z., Wahyu S.J.S., dan Ni Ketut S. Aplikasi Peramalan Penjualan Menggunakan Metode Regresi Linier. *SCAN: Jurnal Teknologi Informasi dan Komunikasi*. Vol. 7 No. 3. 2014. Hal. 41-45.

18. Achmad Solechan dan Qorinta Shinta. Kajian Komparasi Artificial Neural Network dan Regresi Linier dalam Memprediksi Harga Saham dengan Mempertimbangkan Faktor Fundamental pada Sektor Industri. Seminar Nasional Teknologi Informasi & Komunikasi Terapan (SEMANTIK) . 2012. Hal 404-410.
19. Bayu Dwi P., Fajar Sodik P., dan Iqbal K. Pemodelan dan Peramalan Data Saham dengan Analisis Time Series menggunakan Python. Seminar Nasional PRISMA 2020. Vol 3 Hal 714-718.