

PREDIKSI INDEKS HARGA SAHAM GABUNGAN (IHSG) MENGGUNAKAN ALGORITMA *AUTOREGRESSIVE INTEGRATED MOVING AVERAGE* (ARIMA)

Ulfa Khaira*, Pradita Eko Prasetyo Utomo, Tri Suratno, Pikir Claudia Septiani Gulo

Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Jambi

email: ulfa.ilkom@gmail.com

Abstract

There are various types of investment in Indonesia, one of which is the Indeks Harga Saham Gabungan (IHSG) or in English it is called the Indonesia Composite Index, ICI, or IDX Composite. IHSG is an important parameter to consider when making an investment considering that IHSG is a joint stock. This study aims to predict the price of the IHSG with data mining techniques using an algorithm that can be used as a reference for investors when making an investment. ARIMA is a model for generating estimates from historical data. Data in this study were collected from the monthly IHSG from January 4, 2010 - November 26, 2019. Based on the correlation plot, two autocorrelations (lag 1, lag 32) were found to be significant. This model can predict with an average percentage error of 0.004 so that this prediction is considered good enough to predict the stock price of the IHSG.

Keywords: IHSG, Forecasting, ARIMA

1. PENDAHULUAN

Perkembangan teknologi saat ini sangat berpengaruh terhadap mata pencaharian masyarakat. Mudah-mudahan dalam melakukan suatu aktivitas yang menguntungkan dalam aspek finansial misalnya dengan mendapatkan uang melalui online shop, investasi dan trading. Investasi adalah penanaman modal untuk satu atau lebih aktivitas yang dimiliki dan biasanya berjangka waktu lama dengan harapan mendapatkan keuntungan dimasa yang akan datang. Investasi dapat berupa tanah, mesin, bangunan maupun aset finansial seperti deposito, saham atau obligasi. Dari sekian banyak aset finansial yang diperjual belikan di pasar modal, saham adalah yang cukup dikenal pada masyarakat. Akhir ini baik di media elektronika dan media cetak mulai rutin membahas tentang saham, mulai dari pergerakan harga hingga isu yang beredar. Pergerakan harga saham dari waktu ke waktu dapat dilihat dari dampak simultan

dan kompleks atas berbagai faktor yang terjadi pada perekonomian suatu negara. Saat ini indeks harga indeks saham dapat dijadikan sebagai barometer landasan analisis statistik atas kondisi pasar terakhir dan sebagai kesehatan ekonomi suatu negara (Wijaya et al, 2015).

Pergerakan harga saham dapat diamati dari besarnya permintaan dan penawaran akan saham tersebut. Apabila penawaran lebih besar dari permintaan saham akan mengakibatkan harga saham turun sebaliknya apabila penawaran lebih besar jika dibandingkan dengan penawaran, maka mengakibatkan harga saham naik, demikian pula sebaliknya. Harga saham ini akan mengalami perubahan setiap saat yaitu dalam hitungan detik, dikarenakan penilaian sesaat oleh para penjual maupun pembeli. Faktor lain yang mempengaruhi harga saham antara lain adalah tingkat suku bunga deposito, laju inflasi, jumlah laba yang diperoleh perusahaan, strategi pemasaran, pengembalian dan tingkat

resiko (Hanas, 2007).

Ada berbagai macam saham di Indonesia, salah satunya adalah Indeks Harga Saham Gabungan (IHSG) atau dalam bahasa Inggris disebut *Indonesia Composite Index*, ICI, atau *IDX Composite*. IHSG merupakan parameter penting yang dipertimbangkan pada saat akan melakukan investasi mengingat IHSG adalah saham gabungan (Untari, 2009). Data IHSG memiliki fluktuasi yang sangat besar, serta pergerakan harga IHSG sangat cepat sehingga sulit untuk diprediksi kenaikan atau penurunannya.

Oleh karena itu untuk membantu mengatasi permasalahan di atas, maka penulis membuat alternatif untuk memanfaatkan teknik *data mining* untuk melakukan prediksi pada harga *open* menggunakan algoritma *ARIMA*. Penggunaan metode *ARIMA* dalam peramalan jangka pendek sangat tepat digunakan karena metode *ARIMA* memiliki ketepatan yang sangat akurat.

2. TINJAUAN PUSTAKA

2.1 Forecasting

Menurut Indrajit dan Djokopranoto (2003) *forecasting* adalah kegiatan yang berhubungan dengan meramalkan atau memproyeksikan hal-hal yang terjadi dimasa lampau ke masa depan. Pengertian lain oleh Heizer dan Render (2006) *forecasting* adalah seni dan ilmu untuk memperkirakan kejadian dimasa depan. Hal ini dapat dilakukan dengan melibatkan pengambilan data masa lalu dan menempatkannya ke masa yang akan datang dengan suatu bentuk model matematis. Bisa juga merupakan prediksi intuisi yang bersifat subjektif, atau bisa juga dengan menggunakan kombinasi.

Forecasting atau ramalan yang dilakukan pada umumnya akan berdasarkan data yang terdapat dimasa lampau yang dianalisis dengan menggunakan metode-metode tertentu. *Forecasting* diupayakan dibuat dapat meminimumkan pengaruh ketidakpastian tersebut, dengan kata lain bertujuan mendapatkan ramalan yang bisa meminimumkan kesalahan meramal (*forecast error*) yang biasanya diukur dengan *Mean Absolute Deviation*, *Absolute Error*, dan lain sebagainya (Subagyo, 1986). Menurut

Ruamiana, et al. (2018) dalam dunia usaha sangat penting diperlukan hal-hal yang terjadi dimasa depan sebagai dasar untuk pengambilan keputusan. Peramalan adalah seni dan ilmu memprediksi peristiwa-peristiwa masa depan.

Jenis peramalan dapat dibedakan menjadi beberapa tipe. Peramalan dibagi menjadi 3 macam yaitu (Ruamiana, et al., 2018):

1. Peramalan ekonomi Peramalan ekonomi menjelaskan siklus bisnis dengan memprediksi tingkat inflasi, ketersediaan uang, dana yang dibutuhkan untuk membangun perumahan dan indikator perencanaan lainnya.
2. Peramalan teknologi Peramalan teknologi memperhatikan tingkat kemajuan teknologi yang dapat meluncurkan produk baru yang menarik, yang membutuhkan pabrik dan peralatan baru.
3. Peramalan permintaan Peramalan permintaan adalah proyeksi permintaan untuk produk atau layanan suatu perusahaan.

Ruamiana, et al. (2018) juga mengutarakan bahwa peramalan dapat diklasifikasikan berdasarkan horizon waktu masa depan yang dicakupnya. Dalam hubungannya dengan horizon waktu peramalan terbagi atas beberapa kategori, yaitu:

1. Ramalan jangka pendek (*short-range forecast*) Mencakup masa depan yang dekat (*immediate future*) dan memperhatikan kegiatan harian suatu perusahaan bisnis, seperti permintaan harian atau kebutuhan sumber daya harian.
2. Ramalan jangka menengah (*medium-range forecast*) Mencakup jangka waktu satu atau dua bulan sampai satu tahun. Ramalan jangka waktu ini umumnya lebih berkaitan dengan rencana produksi tahunan dan akan mencerminkan hal-hal seperti puncak dan lembah dalam suatu permintaan dan kebutuhan untuk menjamin adanya tambahan untuk sumber daya untuk tahun berikutnya.

- Ramalan jangka panjang (*long-range forecast*) Mencakup periode yang lebih lama dari satu atau dua tahun. Ramalan ini berkaitan dengan usaha manajemen untuk merencanakan produk baru untuk pasar yang berubah, membangun fasilitas baru, atau menjamin adanya pembiayaan jangka panjang.

2.2 Time Series

Data time series adalah nilai-nilai suatu variabel yang berurutan menurut waktu (misal: hari, minggu, bulan, tahun). Salah satu permasalahan penting dalam analisis time series adalah peramalan nilai-nilai suatu variabel pada masa akan datang atau dikenal dengan istilah forecasting. Forecasting merupakan hal yang mendasar dalam sebuah perencanaan dan pengambilan kebijakan dalam suatu instansi karena adanya ketidakpastian dari nilai-nilai suatu variabel pada masa yang akan datang. Metode yang paling sering digunakan dalam pemodelan data time series untuk peramalan (forecasting) adalah metode Box-Jenkins (Aini, 2010).

2.3 Autoregressive Integrated Moving Average (Arima)

Metode Autoregressive Integrated Moving Average (ARIMA) yang biasa disebut dengan metode Box-Jenkins merupakan metode yang dikembangkan oleh George Box dan Gwilym Jenkins pada tahun 1970 (Iriawan, 2006). Metode ARIMA (Autoregressive Integrated Moving Average) adalah metode yang digunakan untuk peramalan jangka pendek. Penggunaan metode ARIMA dalam peramalan jangka pendek sangat tepat digunakan karena metode ARIMA memiliki ketepatan yang sangat akurat. Dan juga menentukan hubungan statistik yang baik antar variabel yang akan diramal dengan nilai yang digunakan untuk peramalan. Sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya nilai peramalan akan cenderung konstan untuk periode yang cukup panjang. Model Autoregressive Integrated Moving Average (ARIMA) adalah model yang secara penuh mengabaikan variabel independen dalam membuat peramalan. Nilai yang digunakan oleh ARIMA untuk

peramalan yaitu menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat (Razak, 2009). Kelompok model yang termasuk dalam metode Autoregressive Integrated Moving Average (ARIMA) yaitu:

- Autoregressive (AR)

Model Autoregressive (AR) diperkenalkan pertama kali oleh Yule pada tahun 1926 dan kemudian dikembangkan oleh Walker pada tahun 1931. Disebut model Autoregressive dikarenakan pada model ini diregresikan terhadap nilai-nilai sebelumnya dari variabel itu sendiri. Model Autoregressive dengan ordo p disingkat menjadi AR(p) atau ARIMA($p,0,0$)

Model :

$$Z_t = \mu + \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \dots + \phi_p Z_{t-p} + a_t$$

dimana, Z_t = deret waktu stasioner

$$\mu = \text{konstanta}$$

$$Z_{t-p} = \text{variabel bebas}$$

ϕ_p = koefisien parameter autoregressive ke- p

$$a_t = \text{sisaan pada saat ke-}t$$

Model diatas disebut sebagai model Autoregressive (regresi diri sendiri) karena model tersebut mirip dengan persamaan regresi pada umumnya, hanya saja yang menjadi variabel independen bukan variabel yang berbeda dengan variabel dependen melainkan nilai sebelumnya (lag) dari variabel dependen (Z_t) itu sendiri.

- Moving Average (MA)

Model Moving Average (MA) pertama kali diperkenalkan oleh Slutsky pada tahun 1973, dengan orde q ditulis MA (q) atau ARIMA ($0,0,q$) dan dikembangkan oleh Wadsworth pada tahun 1989 (Halim, 2006).

Model :

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}$$

dimana,

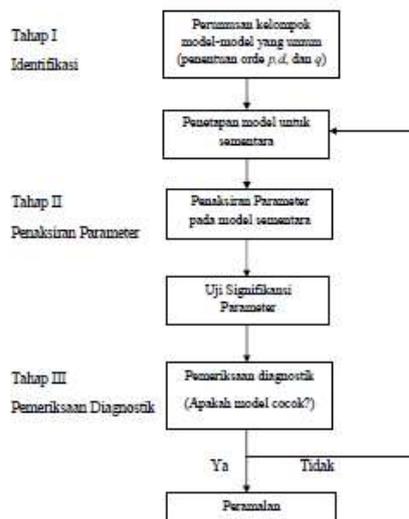
$$Z_t = \text{deret waktu stasioner}$$

$$\mu = \text{konstanta}$$

$$a_{t-1} = \text{variabel bebas}$$

θ_q = koefisien parameter moving average ke- q

$$a_t = \text{sisaan pada saat ke-}t$$



Gambar 1. Model Kerja ARIMA

• Autoregressive Moving Average (ARMA)
Model Autoregressive Moving Average (ARMA) merupakan model gabungan dari Autoregressive (AR) dan Moving Average (MA). Dan model ini memiliki asumsi bahwa data periode sekarang dipengaruhi oleh data periode sebelumnya dan nilai sisaan dari periode sebelumnya.

Model :

$$Z_t = \mu + \phi_1 Z_{t-1} + \dots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (3)$$

dimana,

Z_t = deret waktu stasioner

μ = konstanta

Z_{t-p} = variabel bebas

ϕ_p = koefisien parameter

autoregressive ke-p

a_{t-1} = variabel bebas

θ_q = koefisien parameter

moving average ke-q

a_t = sisaan pada saat ke-t

• Autoregressive Integrated Moving Average (ARIMA)

Model Autoregressive Integrated Moving Average (ARIMA) digunakan berdasarkan asumsi bahwa data deret waktu yang digunakan harus stasioner yang artinya rata-rata variasi dari data.

2.4 Model ARIMA

Yang dimaksud adalah konstan. Namun, ada beberapa hal yang terjadi ketika suatu data tidak stasioner. Dalam mengatasi ketidakstasioneran data ini dilakukan proses differencing agar data menjadi stasioner. Karena model Autoregressive (AR), Moving

Average (MA), Autoregressive Moving Average (ARMA) tidak mampu menjelaskan arti dari differencing, maka digunakan model campuran yang disebut Autoregressive Integrated Moving Average (ARIMA) atau ARIMA (p,d,q) sehingga menjadi lebih efektif dalam menjelaskan proses differencing. Pada model campuran ini series stasioner merupakan fungsi linier nilai lampau beserta nilai sekarang dan kesalahan lampainya.

Model :

$$\Phi_p(B) D^d Z_t = \mu + \theta_q(B) a_t \quad (4)$$

dimana,

Φ_p = koefisien parameter autoregressive ke-p

θ_q = koefisien parameter moving average ke-q

B = operator backshift

D = differencing

μ = konstanta

a_t = sisaan pada saat ke-t p = derajat autoregressive

d = tingkat proses differencing

q = derajat moving average

3. METODE PENELITIAN

A. Dataset Penelitian

Pada penelitian ini data adalah data Indeks Harga Saham Gabungan yang diambil terhitung dari tanggal 04 Januari 2010 hingga 29 November 2019. Data tersebut diperoleh dari website resmi Indeks Harga Saham Gabungan (IHSG) yang di ambil pada investing

(<https://finance.yahoo.com/quote/AALLJK/history?litr=1>) dengan *time frame* satu hari yang digunakan sebagai acuan penelitian dalam menganalisa. Tools yang digunakan dalam penelitian ini adalah : plugin *matplotlib*, *pandas*, *numpy*, *statsmodels.api* dan *seaborn* yang ditulis dalam lingkungan bahasa pemrograman *Python* melalui IDE *Jupyter*.

B. Model yang digunakan

Pada penelitian ini model yang digunakan adalah *ARIMA*. Pada model ini dataset sebelum dilakukan preprosesing di tahap ini ada dua hal, yaitu penentuan *time frame* dan repair data.

C. Prosedur Penelitian

Jalannya penelitian dilakukan dengan tahapan sebagai berikut :

1. Pembersihan dataset (*preprocessing*)

Data yang digunakan dalam penelitian ini memiliki 6 *field* yang meliputi : tanggal, pembukaan (harga saham awal di tanggal tersebut), tertinggi (harga saham tertinggi di tanggal tersebut), terendah (harga saham terendah di tanggal tersebut), penutupan (harga penutup di tanggal tersebut) dan volume (banyaknya nilai saham pada tanggal tersebut). Karena dalam penelitian ini terfokus pada prediksi harga saham dimasa yang akan datang, maka *field* yang digunakan adalah *field* tanggal dan pembukaan. Dalam dataset ini tidak terdapat *missing value* sehingga dapat dilanjutkan ke tahap normalisasi.

2. Normalisasi data

Normalisasi data dalam penelitian ini dilakukan dengan *Min-Max Scaler*. *Min-Max normalization* merupakan metode normalisasi dengan melakukan transformasi linier terhadap data asli sehingga menghasilkan keseimbangan nilai perbandingan antar data saat sebelum dan sesudah proses (Hanifa,2017).

Metode ini dapat menggunakan rumus sebagai berikut :

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Dimana :

X' : Nilai hasil (Normal)

X : Nilai saat ini

Xmin : Nilai terkecil dalam field

X

Xmax : Nilai terbesar dalam field X

3. Penentuan *training dataset* dan *testing dataset*

Dalam Penelitian ini percobaan di lakukan sebanyak dua kali . Percobaan pertama akan menggunakan data yang diambil adalah dari rentan tanggal pada 04 Januari 2010 hingga 29 November 2019 dengan jumlah 2.432 record. Penulis menggunakan data *training* berjumlah 80% atau 1.947 data dan data *testing* berjumlah 20% atau 485 data.

4. Analisis Data

Analisis data pada penelitian ini dilakukan dengan menggunakan metode ARIMA. Sebelum dilakukan perhitungan dengan menggunakan metode ARIMA, terlebih dahulu dilakukan serangkaian uji-uji seperti uji kestasioneran data, proses pembedaan dan pengujian correlogram untuk menentukan koefisien autoregresi.

4. HASIL DAN PEMBAHASAN

1. Normalisasi Data

Sebelum dilakukan proses *mining*, data terlebih dahulu dinormalisasi guna mempermudah proses *data mining*.

2. Data Analisis

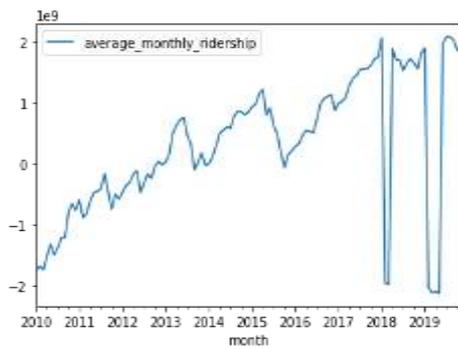
Data penelitian ini adalah data yang terdiri dari 2.432 record Indeks Harga Saham Gabungan dari tanggal 04 Januari 2010 hingga 26 November 2019.

Tabel 1. Data sebelum dan sesudah normalisasi

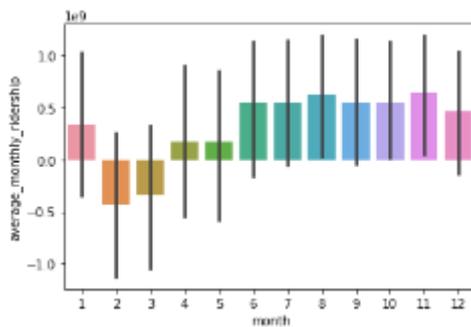
Tahun	Data sebelum	Data sesudah
2002	115.243	-0.96241062
2003	125.065	-0.64208376
2004	129.805	-0.74586313
2005	141.157	-0.99080695

Tabel 2. Hasil data analisis yang digunakan

Count	1.1900
Mean	4.82821
Std	1.03785
min	2.53394
25%	4.12581
50%	5.72664
75%	6.635354

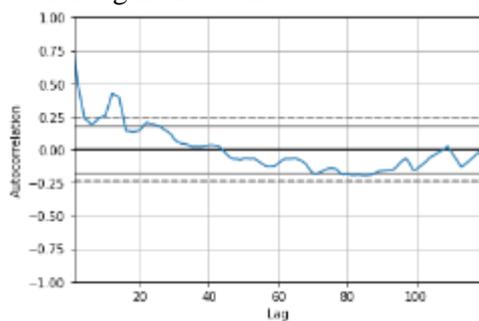


Gambar 2. Plot data yang digunakan



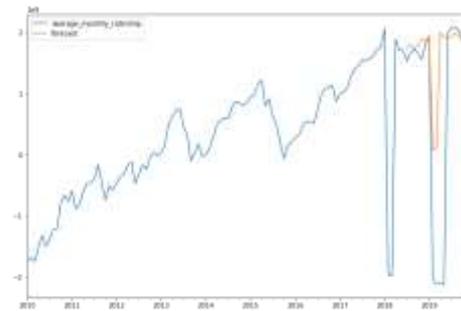
Gambar 3. Bar Plot data setiap bulan

Kestasioneran suatu time series dapat dilihat dari plot ACF yaitu koefisien autokorelasinya menurun menuju nol dengan cepat, biasanya setelah lag ke-2 atau ke-3.



Gambar 4. Plot Autokorelasi data

Setelah dilakukan proses mining menggunakan algoritma ARIMA diperoleh model predict yang dapat digunakan untuk melakukan prediksi. Dengan menggunakan model predict yang dihasilkan, adapun grafik perbandingan antara hasil prediksi training dataset dan testing dataset dengan data sebenarnya yakni:



Gambar 5. Grafik Perbandingan harga saham dengan harga saham asli

-Hasil prediksi harga saham

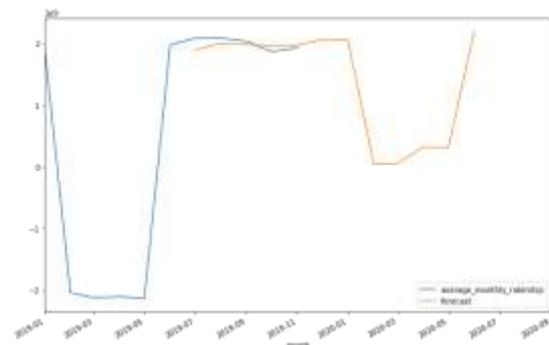
Setelah didapatkan *model predict* pada tahapan sebelumnya, didapatkan hasil prediksi untuk 5 bulan kedepan

Tabel 3. Hasil Prediksi IHSG 5 bulan kedepan

Bulan	Prediksi
2020-01-01	1.9547
2019-02-01	5.1867
2019-03-01	9.4052
2019-04-01	2.0075
2019-05-01	1.9267

- Grafik Hasil Prediksi

Berdasarkan hasil prediksi harga saham yang diperoleh untuk 5 bulan ke depan maka dapat tersebut dapat divisualisasikan sebagai berikut:



Gambar 6. Grafik hasil prediksi

Berdasarkan penelitian yang telah dilakukan akan terjadi kenaikan harga saham pada tanggal 01 Januari 2020 – 01 Maret 2020 Harga saham meningkat secara drastis namun pada tanggal 01

April 2020 – 01 Mei 2020 Harga saham menurun dengan cepat. Harga Saham tertinggi terjadi pada bulan Maret. Secara ekonomi, hal tersebut akan menyebabkan perubahan investasi di Indonesia. Ditinjau dari hasil prediksi dalam penelitian ini, aktivitas pembelian saham internasional dan investasi saham pada bulan Februari sampai dengan April kurang direkomendasikan. Sebaliknya, Investasi saham di bulan April dan Mei direkomendasikan.

Dalam penelitian ini model prediksi yang dihasilkan cukup baik, hal ini diutarakan karena hasil pengujian skor MSE menggunakan training dataset dan testing dataset menunjukkan angka 0.004 dan 0.003.

5. KESIMPULAN

Dari penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Akan terjadi kenaikan harga saham yang sangat drastis di bulan Maret 2020.
2. Penurunan harga saham gabungan terjadi di bulan April hingga Mei.
3. Dalam *ARIMA* banyaknya training dataset dan jumlah iterasi berbanding lurus dengan keakuratan prediksi yang dihasilkan oleh sistem.

6. REFERENSI

- Aini, Nur.,2010, “Analisis Statistik Terhadap Polutan Udara Partikulat (PM10) Berdasarkan Pengukuran Stasiun Pemantau Taman Prestasi Dengan Metode Bootstrap Time Series”, Surabaya: ITS.
- Halim, “Diktat Time Series”, Universitas Kristen Petra, Surabaya, 2006.
- Hanias, M., Curtis, P., & Thalassinis, J. (2007). Prediction with neural networks: The Athens stock exchange price indicator. *European Journal of Economics, Finance and Administrative Sciences*, 9, 21-27.
- Indrajit, R. E., & Djokopranoto, R. (2003). Manajemen Persediaan. *Jakarta: PT Gramedia Widiasarana Indonesia*.
- Iriawan, N., & Astuti, S. P. (2006). Mengolah data statistik dengan mudah menggunakan minitab 14. *Penerbit Andi, Yogyakarta*.
- Razak, F. A., Shitan, M., Hashim, A. H., & Abidin, I. Z. (2009). Load forecasting using time series models. *Jurnal Kejuruteraan*, 21, 53-62.
- Ruamiana, W. B., Nangi, J. & Tajidun, L., (2018). Aplikasi Forecasting Jumlah Frekuensi Penumpang Pesawat Terbang Lion Air pada Bandar Udara Halu Oleo dengan menggunakan Metode Least Square. *semanTIK*, 4(1).
- Subagyo, 1986. *Forecasting Konsep dan Aplikasi*. Yogyakarta: BPFE.
- Tilakaratne, C. D., et al. (2009). Modified Neural Network Algorithms for Predicting Trading Signals of Stock Market Indices. *Journal of Applied Mathematics and Decision Science Vol. 2009*.
- Untari, N., Mattjik, A. A., & Saefuddin, A. (2009, April). Analisis Deret Waktu Dengan Ragam Galat Heterogen Dan Asimetrik Studi Indeks Harga Saham Gabungan (Ihsg) Periode 1999-2008. In *Forum Statistika dan Komputasi* (Vol. 14, No. 1, pp. 22-33).
- Wijaya, T. S., & Agustin, S. (2015). Faktor-Faktor yang Mempengaruhi Nilai IHSG yang Terdaftar di Bursa Efek Indonesia. *Jurnal Ilmu dan Riset Manajemen*, 4(6), 1-16.