

# Metode *Long Short Term Memory* dan *Generalized Autoregressive Conditional Heteroscedasticity* untuk Pemodelan Data Saham

Jannatun Khustia Lubis<sup>a</sup>, Iqbal Kharisudin<sup>b</sup>

<sup>a, b</sup> Universitas Negeri Semarang, Sekaran, Gunungpati, Semarang, 50229, Indonesia

\*Alamat Surel: [jannatunkl@gmail.com](mailto:jannatunkl@gmail.com)

## Abstrak

Salah satu investasi yang diminati masyarakat saat ini adalah investasi pada saham. Terdapat risiko dalam berinvestasi, sehingga diperlukan peramalan untuk meminimalisir kerugian. Banyak metode peramalan yang dapat digunakan, baik linier maupun nonlinier. Artikel ini membahas tentang metode *Long Short Term Memory* (LSTM) dan *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) untuk pemodelan data saham. Data yang digunakan adalah data harga penutupan saham *Alphabet Inc* dengan kode GOOGL. Tujuan dari penelitian ini adalah mendapatkan metode terbaik dalam peramalan harga saham, sehingga memberikan peramalan yang lebih efektif dan hasil yang lebih akurat.

## Kata kunci:

*Long Short Term Memory*, *Generalized Autoregressive Conditional Heteroscedasticity*, pemodelan time series, peramalan.

© 2021 Dipublikasikan oleh Jurusan Matematika, Universitas Negeri Semarang

## 1. Pendahuluan

Kesadaran akan pentingnya investasi pada masyarakat akhir-akhir ini semakin meningkat. Salah satu investasi yang diminati masyarakat saat ini adalah investasi pada saham. Saham adalah tanda kepemilikan investor atau trader atas investasi sejumlah dana yang diinvestasikan dalam suatu perusahaan (Azis *et al.*, 2015).

Data harga penutupan saham dapat ditemukan di *website yahoo.finance.com* ataupun pada sekuritas yang terdaftar. Investasi saham dapat menimbulkan kerugian maupun keuntungan. Untuk meminimalkan risiko kerugian dan memaksimalkan keuntungan yang didapat para investor membutuhkan peramalan harga saham.

Harga saham merupakan kategori data *time series*. Peramalan *time series* merepresentasikan banyak tantangan nyata yang memengaruhi secara langsung atau tidak langsung kehidupan manusia, seperti pemrosesan bahasa, peramalan harga saham, atau peramalan cuaca. Langkah dasar untuk menyelesaikan masalah *time series* adalah mengumpulkan dan memilih variabel yang tepat, lalu memilih model terbaik yang memberikan akurasi terbaik (Berradi & Lazaar, 2019). Terdapat banyak cara untuk memodelkan data *time series*, diantaranya model linier, model nonlinier, dan model multivariat. Beberapa model nonlinier adalah *Long Short Term Memory* (LSTM) dan *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH).

Berdasarkan hasil pengujian pada penelitian yang telah dilakukan pada tahun 2017-2019, dengan performa yang baik dan tingkat kesalahan yang relatif kecil LSTM dapat memprediksi harga saham. Dibandingkan dengan penggunaan metode *Support Vector Regression* (SVR), LSTM memiliki nilai *loss* lebih baik. LSTM mampu menanggulangi ketergantungan jangka panjang dan mampu memprediksi harga saham dengan hasil lebih akurat (Arfan & Lussiana, 2020).

## To cite this article:

Lubis, J. K., & Kharisudin, I. (2021). Metode *Long Short Term Memory* dan *Generalized Autoregressive Conditional Heteroscedasticity* untuk Pemodelan Data Saham. *PRISMA, Prosiding Seminar Nasional Matematika 4*, 652-658

Pada penelitian yang telah dilakukan dalam meramalkan harga saham bank BRI menggunakan model ARIMA-GARCH diperoleh unsur heteroskedastik. Kemudian model terbaik yang didapat yaitu ARIMA(2,1,1)-GARCH(2,2) dengan nilai koefisien determinasi atau (*R-squared*) sebesar 99,91% (Yolanda *et al.*, 2017).

Dalam penelitian ini terdapat beberapa rumusan masalah diantaranya, (1) bagaimana penerapan metode LSTM dalam meramalkan harga saham? (2) bagaimana penerapan metode GARCH dalam meramalkan harga saham? dan (3) bagaimana hasil peramalan saham dengan metode LSTM dan GARCH untuk tiga periode berikutnya berdasarkan pada model terbaik?

Tujuan dari penelitian ini adalah, (1) mengetahui penerapan metode LSTM dalam meramalkan harga saham, (2) mengetahui penerapan metode GARCH dalam meramalkan harga saham, dan (3) mengetahui hasil peramalan saham dengan metode LSTM dan GARCH untuk tiga periode berikutnya berdasarkan pada model terbaik

Terdapat beberapa manfaat dalam penelitian ini diantaranya, (1) peneliti dapat mengembangkan dan mengaplikasikan ilmu pengetahuan di bidang matematika, mengetahui dan menambah wawasan dalam penerapan metode LSTM dan GARCH dalam meramalkan harga penutupan saham, dan (2) pembaca dapat menambah pengetahuan terhadap metode LSTM dan GARCH dalam meramalkan harga penutupan saham, mengetahui metode yang terbaik untuk meramalkan harga penutupan saham yang akan datang.

---

## 2. Metode

Metode yang digunakan adalah metode studi pustaka dengan mengumpulkan sumber pustaka yang berupa buku, jurnal, skripsi, dan literatur lainnya. Data yang digunakan adalah data harga penutupan saham perusahaan *Alphabet Inc* mulai tanggal 01 Juli 2015-08 September 2020. Perusahaan tersebut merupakan induk perusahaan *google*.

### 2.1. Harga Penutupan

Harga penutupan saham adalah harga yang diminta penjual dan pembeli pada akhir hari bursa. Pada saat akhir bursa, kemungkinan dapat terjadi transaksi suatu saham secara tiba-tiba, karena ada kesepakatan antar penjual dan pembeli. Sehingga, harga penutupan tersebut telah menjadi harga pasar. Namun, harga tersebut tetap menjadi harga penutupan saham pada hari tersebut (Huda, 2019).

### 2.2. Perangkat

Perangkat yang digunakan untuk peramalan data harga penutupan saham perusahaan *Alphabet Inc* menggunakan LSTM dan GARCH dikembangkan pada *hardware* dengan spesifikasi yang terdiri dari Komputer ASUS PRO series P2430U, Processor Intel (R) Core (TM) i3-6006U CPU @ 2.00GHz, 1992Mz, 2 Core(s), 4 Logical Processor, RAM 4,00 GB, HDD 500 GB. *Software* terdiri dari OS Windows 10 Pro, 64-bit Operating System, Python 3.9 Manuals (64-bit), RStudio.

### 2.3. Long Short Term Memory (LSTM)

*Long Short Term Memory* (LSTM) adalah metode yang diusulkan oleh Sepp Hochreiter dan Jurgen Schmidhuber pada tahun 1997. Metode ini merupakan pengembangan dari *Recurrent Neural Network* (RNN). RNN tidak dapat belajar menghubungkan informasi karena memori lama yang tersimpan akan semakin tidak berguna dan tertimpa dengan memori baru. Sedangkan, LSTM tidak memiliki kekurangan tersebut karena dapat mengatur memori pada setiap masukannya dengan menggunakan *memory cells* dan *gate units* (Arfan & Lussiana, 2019).

Normalisasi data dilakukan karena *range* nilai *input* tidak sama. Nilai normalisasi yang kecil diperoleh dari *input* yang diproses ke nilai *output* yang kecil. Denormalisasi data berfungsi untuk mengembalikan nilai yang telah dinormalisasi menjadi nilai asli (Huda, 2019).

### 2.4. Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

*Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) merupakan perluasan dari model *Autoregressive Conditional Heteroscedasticity* (ARCH) yang diperkenalkan oleh Bollerslev pada tahun 1986. GARCH menggunakan lebih sedikit parameter sehingga mengurangi tingkat kesalahan perhitungan, sehingga memberikan hasil yang lebih sederhana (Hendikawati, 2015).

*Time series* adalah serangkaian pengamatan terhadap suatu peristiwa, gejala atau variabel yang diambil dari waktu ke waktu, kemudian dicatat secara teliti sesuai urutan terjadinya lalu disusun sebagai data statistik. Sehingga dari *time series* dapat diketahui apakah peristiwa atau variabel yang diamati berkembang mengikuti pola-pola perkembangan yang teratur (Hanke & Wichern, 2014).

Model Box-Jenkins atau *Autoregressive Integrated Moving Average* (ARIMA) merupakan gabungan antara model *autoregressive* (AR) dan *moving average* (MA). Kedua model tersebut mensyaratkan data yang dianalisis stasioner. Dalam memilih model yang cocok untuk peramalan data runtut waktu univariat, metode yang digunakan terdiri dari tiga tahap yaitu identifikasi model, estimasi parameter, dan peramalan (Enders, 2004).

### 2.5. Pemecahan masalah

Dalam penelitian ini, secara umum langkah-langkah pemecahan masalah terdiri menjadi tiga tahap yaitu melakukan peramalan dengan metode *Long Short Term Memory* (LSTM) dibantu dengan *software* Python, melakukan peramalan dengan metode *Generalized Autoregressive Conditional Heteroscedasticity* (GARCH) dibantu dengan *software* Rstudio, dan membandingkan hasil peramalan dengan data aktual.

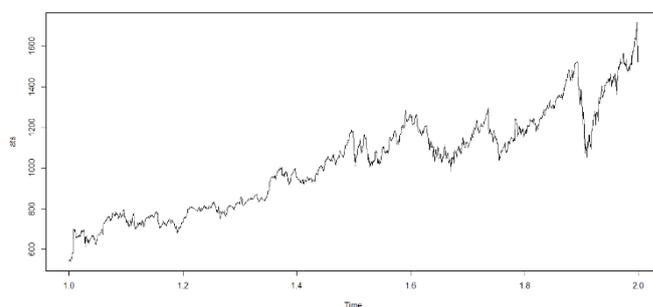
Langkah-langkah pemecahan masalah dalam peramalan harga penutupan saham menggunakan metode LSTM sebagai berikut.

- Menginput data harga penutupan saham.
- Melakukan analisis deskriptif.
- Melakukan *preprocessing* data.
- Menentukan *optimizer* dan jumlah *hidden neuron*.
- Pemodelan dengan LSTM.
- Menentukan model terbaik.
- Melakukan peramalan.

Sedangkan peramalan harga penutupan saham menggunakan metode GARCH sebagai berikut.

- Menginput data harga penutupan saham.
- Membuat data *time series* dan plot data.
- Melakukan uji stasioner data.
- Mengidentifikasi model ARIMA.
- Melakukan uji efek heteroskedastik.
- Mengidentifikasi model GARCH.
- Memilih model GARCH terbaik.
- Melakukan peramalan.

## 3. Pembahasan



**Gambar 1.** plot data *time series* harga saham

Dapat diketahui dari plot data di atas bahwa pergerakan harga penutupan saham *Alphabet Inc* memiliki komponen pola siklus disertai pola *trend* baik kenaikan maupun penurunan dan sangat fluktuatif.

### 3.1. LSTM

Merujuk pada penelitian sebelumnya (Huda, 2019), digunakan analisis deskriptif untuk melihat gambaran umum dari harga penutupan saham *Alphabet Inc*. Dari analisis deskriptif akan diperoleh *mean* yang

menunjukkan harga rata-rata saham, standar deviasi yang menunjukkan penyimpangan harga saham, *min* yang menunjukkan harga terendah dari data saham, dan *max* yang menunjukkan harga tertinggi dari saham.

*Preprocessing* data bertujuan untuk membuat data mentah lebih mudah diterima oleh *neural network* (Chollet, 2018). Tahap-tahap *preprocessing* sebagai berikut.

- Penentuan pola *input* dan *output*

Dalam penelitian ini menerapkan pembelajaran *supervised learning* yaitu membutuhkan pasangan data *input* dan *output* aktual untuk dipelajari. Data *input* yang digunakan adalah harga saham tiga periode sebelumnya untuk memprediksi harga saham satu hari berikutnya. Hal ini diasumsikan bahwa harga saham dipengaruhi oleh harga saham tiga periode sebelumnya.

- Normalisasi data

Pada tahap ini data *input* dan *output* dinormalisasi menjadi *range* 0-1 menggunakan rumus *min-max normalization* sebagai berikut.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

dimana

$x'$  = Data setelah dinormalisasikan

$x$  = Data aktual

$x_{min}$  = Data minimum dari keseluruhan data

$x_{max}$  = Data maksimum dari keseluruhan data

- Pembagian data *training* dan *testing*

Pembuatan data *training* (pelatihan) dilakukan dengan tujuan meningkatkan kinerja dari LSTM terhadap data *testing* dalam memilih parameter terbaik untuk model. Data *training* akan digunakan untuk proses pelatihan dengan menggunakan metode LSTM sehingga terbentuk suatu model yang akan diuji performasinya terhadap data *testing*. Pembagian data yang digunakan yaitu 80% data *training* dan 20% data *testing*. Jumlah data *training* lebih besar dikarenakan agar mesin pembelajaran lebih terlatih untuk mempelajari model. Sehingga model yang dihasilkan dapat memberikan peramalan data *testing* yang lebih optimal.

Secara teori, tidak ada aturan baku dalam menentukan arsitektur LSTM yang optimal untuk diterapkan ke dalam sistem. Sehingga perlu dilakukan secara *trial* dan *error* dalam pencarian arsitektur dan parameter *training* LSTM.

Arsitektur jaringan LSTM yang didapat terdiri dari *input layer*, *LSTM layer*, dan *output layer*. Sedangkan, parameter-parameter terdiri dari sejumlah *epoch* dan sejumlah *batch size*. *Epoch* adalah langkah yang dilakukan pada proses pembelajaran *neural network*, besarnya *epoch* mempengaruhi besaran proses pembelajaran dan berhenti tepat pada nilai *epoch* yang telah ditentukan. *Epoch* sendiri prosesnya serupa dengan iterasi yang berupa rambatan balik (Rahmadya, 2017). Sedangkan *batch size* adalah jumlah sampel data yang disebarkan ke *neural network*.

Adapun parameter-parameter yang dijadikan fokus dalam penelitian ini adalah parameter *optimizer* dan jumlah *hidden neuron*. *Optimizer* yang dapat dipilih diantaranya adalah Adam, Adamax, dan RMSProp. Sedangkan jumlah *hidden neuron* terdiri dari 10, 20, 30, 40, 50, 60, 70, 80, 90, dan 100. Perfoma model diukur menggunakan metrik RMSE, semakin kecil nilai RMSE pada data *testing* maka model itulah yang terbaik. Sebelum menghitung RMSE, data-data yang sebelumnya dinormalisasi diubah kembali menjadi data sebelumnya menggunakan rumus denormalisasi sebagai berikut.

$$x = x'(x_{max} - x_{min}) + x_{min} \quad (2)$$

dimana

$x'$  = Data setelah dinormalisasikan

$x$  = Data aktual

$x_{min}$  = Nilai minimum dari keseluruhan data

$x_{max}$  = Nilai maksimum dari keseluruhan data

Setelah menentukan dan jumlah *hidden neuron* yang terbaik, dilakukan peramalan LSTM menggunakan model terbaik LSTM. Hasil yang didapatkan berupa plot grafik, perbandingan antara hasil peramalan harga penutupan saham dan data aktual harga penutupan saham, serta nilai RSME.

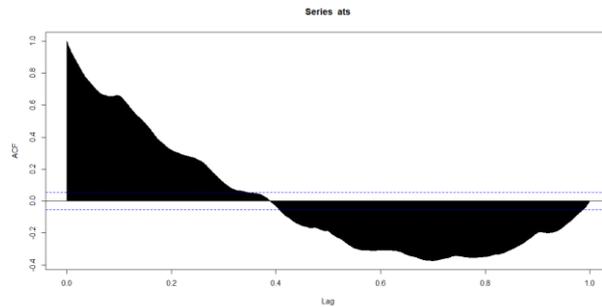
### 3.2. GARCH

Merujuk pada penelitian sebelumnya (Yolanda *et al.*, 2017), akan diperoleh hasil uji stasioneritas menggunakan uji akar unit (*unit root test*) yang didapat berdasarkan uji hipotesis sebagai berikut.

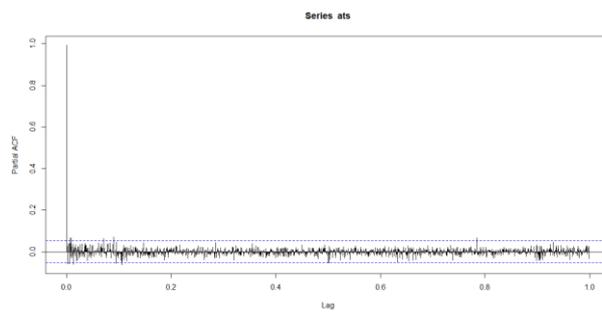
$H_0$  = data memiliki akar unit (data tidak stasioner)

$H_1$  = data tidak memiliki akar unit (data stasioner)

Hasil output nilai  $p\text{-value} = 0,01368 < 0,05$ , sehingga  $H_0$  ditolak sehingga data tidak memiliki akar unit atau data sudah stasioner.

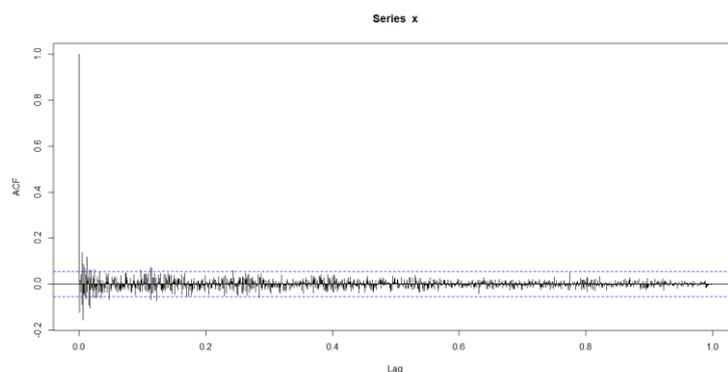


**Gambar 2.** plot ACF

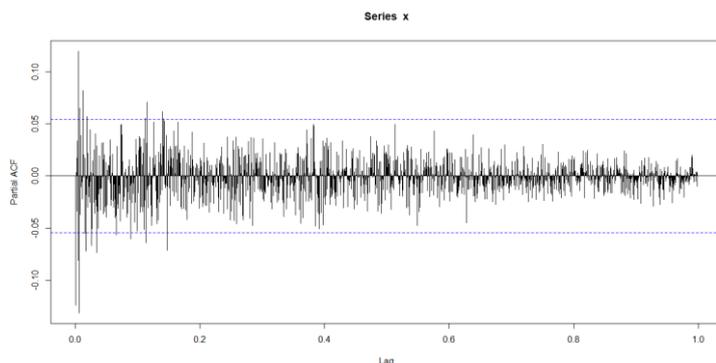


**Gambar 3.** plot PACF

Berdasarkan pada uji akar unit, diketahui bahwa data sudah stasioner tetapi pada grafik ACF mengalami penurunan secara perlahan, sehingga disimpulkan bahwa data belum stasioner dan perlu dilakukan proses *differencing*. Setelah proses *differencing* dilakukan, diperoleh data sudah stasioner yang dapat dilihat dengan melakukan pengecekan kembali, plot ACF, plot PACF, dan nilai ADF 0,01 pada data hasil proses *differencing*.



**Gambar 4.** plot ACF hasil *differencing*



**Gambar 5.** plot PACF hasil *differencing*

Dari gambar di atas dapat diperoleh bahwa mengalami *differencing* pada tingkat 1, *cuts off* ACF pada lag 1 dan PACF pada lag 1. Sehingga diperoleh model ARIMA yang diestimasi, yaitu model ARIMA(1,1,1), ARIMA(1,1,0), dan ARIMA(0,1,1).

Model terbaik yang dipilih dilihat dari nilai AIC dan MLE dengan ketentuan mempunyai nilai AIC terendah dan MLE tertinggi. Untuk mendapatkan model yang mampu menjelaskan data dengan baik, model terbaik tersebut diuji apakah akan menghasilkan residual yang bersifat random (*white noise*) atau tidak. Jika hasil dari *output* didapat memiliki nilai  $p\text{-value} < \alpha = 0,05$ , maka dapat disimpulkan bahwa model tidak memenuhi asumsi *white noise*. Sehingga, diperlukan estimasi model ARIMA lagi. Tetapi, jika hasil dari *output* didapat memiliki nilai  $p\text{-value}$  lebih dari  $\alpha$ , maka dapat disimpulkan bahwa model memenuhi asumsi *white noise*. Sehingga, dilanjutkan ke tahap berikutnya yaitu identifikasi efek heterokedastik.

Dalam pemodelan GARCH dilakukan identifikasi data yang diamati mengalami heteroskedastik atau tidak. Untuk hasil ARCH-LM keputusan yang diambil berdasarkan uji hipotesis sebagai berikut.

$H_0$  = data tidak mempunyai efek heterokedastik

$H_1$  = data mempunyai efek heterokedastik

Berdasarkan uji ARCH-LM jika data menunjukkan bahwa nilai  $p\text{-value} < 0,05$  artinya tolak  $H_0$  dan terima  $H_1$ . Hal ini berarti terdapat efek ARCH-GARCH (heteroskedastik). Selanjutnya akan diestimasi model ARCH-GARCH yang ditentukan berdasarkan plot ACF dan plot PACF. Model GARCH terbaik dipilih dengan ketentuan nilai AIC yang lebih rendah dari sebelumnya pada model ARIMA sebelumnya.

Model terbaik GARCH akan diuji kembali ARCH-LM. Jika hasil model GARCH dari pengujian kembali diperoleh nilai  $p\text{-value} > 0,05$ , maka diasumsikan model GARCH tersebut sudah tidak memiliki unsur heteroskedastik dan dapat dilanjutkan ke tahap berikutnya yaitu peramalan.

#### 4. Simpulan

Model LSTM terbaik dipilih dengan menentukan *optimizer* dan *hidden neural* serta melihat nilai RSME terkecil. Untuk model dengan metode GARCH untuk meramalkan harga penutupan saham Alphabet Inc dipilih dengan melihat uji efek heteroskedastik dan nilai AIC terendah. Hasil peramalan dari metode LSTM dan GARCH dibandingkan dengan data aktual, sehingga akan diperoleh metode mana yang lebih efektif dan akurat.

#### Daftar Pustaka

- Arfan, A., & Lussiana, E. T. P. (2020). Perbandingan Algoritma Long Short-Term Memory dengan SVR Pada Prediksi Harga Saham di Indonesia. *Petir: Jurnal Pengkajian dan Penerapan Teknik Informatika*, 13(1), 33-43.
- Arfan, A., & Lussiana, E. T. P. (2019). Prediksi Harga Saham Di Indonesia Menggunakan Algoritma Long Short-Term Memory. *SenTIK*, 3(1), 225–230.
- Azis, M., Mintarti, S., & Nadir, M. (2015). *Manajemen Investasi Fundamental, Teknikal, Perilaku Investor dan Return Saham*. Sleman: Deepublish.

- Berradi, Z., & Lazaar, M. (2019). Integration of Principal Component Analysis and Recurrent Neural Network to Forecast the Stock Price of Casablanca Stock Exchange. *Procedia Computer Science*, 148, 55–61.
- Chollet, F. (2018). *Deep Learning with Python*. Shelter Island: Manning.
- Enders, W. (2004). Applied econometrics time series. In *Applied Time Series Econometrics* (4th ed.).
- Hanke, J. E., & Wichern, D. (2014). *Business Forecasting* (9th ed.). London: Pearson Education Limited.
- Hendikawati, P. (2015). *Peramalan Data Runtut Waktu Metode dan Aplikasinya dengan Minitab & Eviews* (D. Mubarak, Ed.). Semarang: FMIPA Unnes.
- Huda, F. F. El. (2019). Penerapan Long Short-Term Memory Untuk Peramalan Harga Saham Perusahaan Perkebunan Di Indonesia. (*Doctoral Disertation*). Universitas Islam Indonesia.
- Rahmadya. (2017). Epoch, Iteration, dan Stop Condition. (*Online*). (<https://rahmadya.com/2017/10/23/epoch-iteration-dan-stop>)
- Yolanda, N. B., Nainggolan, N., & Komalig, H. A. H. (2017). Penerapan Model ARIMA-GARCH Untuk Memprediksi Harga Saham Bank BRI. *Jurnal MIPA*, 6(2), 92.