

## Prediksi Pergerakan Harga Saham PT. Astra Internasional Tbk Menggunakan Vector Auto Regressive (VAR) Stasioner dan Logistic Regression

Ayu Wulandari<sup>1</sup>, Aniq Atiqi Rohmawati<sup>2</sup>

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung

<sup>1</sup>ayuwulandari@students.telkomuniversity.ac.id, <sup>2</sup>aniqatiqi@telkomuniversity.ac.id,

---

### Abstrak

Saham merupakan satuan nilai dalam berbagai instrumen keuangan yang kini menjadi metode utama untuk meningkatkan modal bisnis. Harga saham yang bersifat fluktuatif membuat diperlukannya metode yang tepat untuk memprediksi pergerakan harga saham untuk mengurangi kerugian yang terjadi. Dalam penelitian kali ini, dilakukan prediksi terhadap pergerakan harga saham menggunakan data harga saham harian pada PT Astra Internasional Tbk dengan metode Vector Autoregressive (VAR) Stasioner dan Logistic Regression. Metode VAR merupakan metode deret waktu yang digunakan untuk memodelkan dan meramalkan secara simultan lebih dari satu variabel yang saling berkaitan. Sedangkan metode Logistic Regression merupakan metode analisis yang mendeskripsikan hubungan antara dua variabel dependen yang memiliki dua kategori atau lebih dengan satu atau lebih variabel independent yang berskala kategori atau kontinu. Penelitian ini menghasilkan kesimpulan bahwa metode Logistic Regression memberikan hasil yang cenderung lebih akurat yaitu nilai akurasi sebesar 69.60%, presisi 78.57% dan recall 46.80%, dibandingkan dengan metode Vector Autoregressive (VAR) yang menghasilkan nilai akurasi sebesar 50.98%, presisi 28.57%, dan recall 4.25%.

**Kata kunci : Saham, VAR Stasioner, Logistic Regression**

---

### Abstract

Stock is a unit of value in various financial instruments which is now the main method for increasing business capital. Fluctuating stock prices make the need for appropriate methods to predict stock price movements to reduce losses that occur. In this research, a prediction of stock price movements is carried out using daily stock price data on PT Astra International Tbk using Stationary Vector Autoregressive (VAR) and Logistic Regression methods. The VAR method is a time series method that is used to model and predict simultaneously more than one interrelated variable. While the Logistic Regression method is an analytical method that describes the relationship between two dependent variables that have two or more categories with one or more independent variables. This research concludes that the Logistic Regression method gives results that tend to be more accurate, namely an accuracy value of 69.60%, a precision of 78.57% and a recall of 46.80%, compared to the Vector Autoregressive (VAR) method which produces an accuracy value of 50.98%, 28.57% precision, and 4.25% recall.

**Keywords: Stock Price, VAR Stationary, Logistic Regression**

---

### 1. Pendahuluan

#### Latar Belakang

Saham merupakan satuan nilai dalam berbagai *instrument financial* yang mengacu pada kepemilikan sebuah perusahaan. Saham menjadi salah satu metode utama untuk meningkatkan modal bisnis karena memungkinkan perusahaan yang membutuhkan pendanaan jangka panjang untuk menjualnya kepada investor dengan imbalan berupa uang tunai. Saham dapat dijual melalui pasar primer maupun pasar sekunder. Dewasa ini investasi saham menjadi pilihan investasi yang diminati karena memberikan keuntungan yang cepat dan besar. Namun terdapat beberapa faktor yang mampu menyebabkan kerugian dalam investasi saham ini [1].

Analisis deret waktu banyak digunakan sebagai alat yang digunakan untuk memprediksi secara efektif nilai variabel dimasa yang akan mendatang berdasarkan tren sebelumnya [3]. Beberapa penelitian menggunakan analisis deret waktu digunakan untuk memprediksi pergerakan harga saham pada studi kasus suatu perusahaan. Misalnya sebuah penelitian yang dilakukan oleh Agri Pratomo, Rian Febrian Umbara dan Aniq Atiqi Rohmawati memprediksi pergerakan harga saham yang mengadaptasi metode *Random Forest* Menggunakan *Trend Deterministic Data Preparation* pada studi kasus 3 perusahaan, yaitu PT Astra Internasional Tbk, PT Garuda Indonesia Tbk dan PT Indosat Tbk. Studi kasus ini berhasil mengembangkan alat yang dapat digunakan untuk memprediksi pergerakan harga saham pada tahun 2011 – 2017. Bahkan dalam studi kasus ini menyatakan bahwa metode yang digunakan menunjukkan hasil akurasi yang lebih baik dibandingkan dengan menggunakan metode

*Support Vector Machine* (SVM) [19]. Studi kasus lain yang berhasil dilakukan yaitu Ukuran Resiko Model *Vector Autoregressive* pada Harga Saham di PT Unilever Indonesia, yang dilakukan oleh Fitriani Amalia, Deni Saepudin, dan Aniq Atiqi Rohmawati. Studi kasus ini berhasil mengembangkan alat yang digunakan untuk memprediksi variabel-variabel harga saham yang cenderung mengalami perubahan secara signifikan dengan menghasilkan nilai error yang cukup kecil yaitu sebesar 26.09 [20].

Prediksi pergerakan harga saham menjadi hal yang penting bagi para investor untuk menjadi bahan pertimbangan dalam keputusan kapan harus membeli saham milik perusahaan lain atau menjual saham yang dimiliki. Pada tugas akhir ini dilakukan penelitian tentang prediksi pergerakan harga saham PT Astra Internasional Tbk dengan menggunakan data harian saham yaitu data harga penutupan dan data harga terendah menggunakan metode klasifikasi *Vector Autoregressive* (VAR) Stasioner dan *Logistic Regression*. Banyak penelitian telah dilakukan sebelumnya dengan mengadaptasi metode *Vector Autoregressive* (VAR) dan *Logistic Regression* sebagai metode penelitian dalam memprediksi nilai variabel yang akan mendatang pada bidang ekonomi.

### Topik dan Batasannya

Topik dan batasan masalah dalam penelitian tugas akhir ini adalah untuk memastikan apakah metode *Vector Autoregressive* (VAR) Stasioner dan *Logistic Regression* dapat digunakan untuk memprediksi pergerakan harga saham dan membandingkan hasil kinerja prediksi pergerakan saham kedua metode tersebut dengan menggunakan data harga saham PT Astra Internasional Tbk. Data yang digunakan berupa data harga saham harian yaitu data harga penutupan dan data harga terendah pada periode Januari 2013 – Desember 2018 yang diperoleh dari portal web [www.id.investing.com](http://www.id.investing.com). Prediksi ini dilakukan hanya menggunakan data historis saham dari PT Astra Internasional Tbk sehingga faktor eksternal yang dapat mempengaruhi perubahan nilai harga saham diabaikan.

### Tujuan

Tujuan yang ingin dicapai dalam penelitian tugas akhir ini adalah memprediksi pergerakan harga saham pada perusahaan PT Astra Internasional Tbk dan membandingkan hasil klasifikasi kinerja prediksi menggunakan metode *Vector Autoregressive* (VAR) Stasioner dan *Logistic Regression*.

## 2. Studi Terkait

### 2.1.1 Harga Saham

Saham merupakan surat berharga yang dikeluarkan oleh sebuah perusahaan berbentuk Perseoran Terbatas (PT) yang merupakan tanda kepemilikan seseorang atau badan terhadap suatu perusahaan. Harga saham adalah harga per lembar saham yang terbentuk karena adanya interaksi antara penjual dan pembeli saham. Dimana, dalam melakukan proses pembelian dan penjualan saham ini, investor sebagai pembeli saham memerlukan adanya analisis terhadap harga saham menggunakan pendekatan analisa fundamental dan analisa teknikal. Analisa fundamental digunakan untuk mengidentifikasi faktor-faktor fundamental yang menitikberatkan pada kinerja perusahaan yang mengeluarkan saham yang nantinya diperkirakan akan mempengaruhi harga saham. Sedangkan analisa teknikal digunakan untuk memperkirakan harga saham dengan mengamati perubahan harga saham diperiode lalu yang dihubungkan dengan kejadian pada saat itu (pengaruh ekonomi, pengaruh politik, pengaruh perdagangan, pengaruh psikologis) dengan menggunakan indikator-indikator teknis [2, 3].

### 2.1.2 Vector Autoregressive (VAR)

*Vector Autoregressive* (VAR), model ini pertama kali dikembangkan oleh Sims (1980) sebagai alternatif yang lebih baik daripada model persamaan simultan dinamis tradisional untuk menguji interaksi dinamis antara data deret waktu yang saling terkait. Model ini banyak digunakan untuk memodelkan ketergantungan temporal dalam rangkaian waktu multivarian yang menjelaskan atau memprediksi nilai-nilai seperangkat variabel pada suatu titik waktu tertentu yang banyak digunakan dalam peramalan dan tes kausalitas [7, 8]. Model ini merupakan ‘vektor’ karena mengasumsikan vektor dengan keadaan multivariate  $X$  dan ‘autoregressive’ karena dimana keadaan selanjutnya ditentukan oleh regresi terhadap tingkat  $L$  kondisi saat ini dan kondisi lalu [9]. Model ini dilambangkan sebagai [9],

$$X_t = \sum_{i=1}^L A_i X_{t-i} + \varepsilon_t \quad (1)$$

Keterangan:

- $X_t$  : nilai observasi saat t
- $A_i$  : parameter autoregressive ke-n
- $\varepsilon$  : nilai error saat t

Bentuk umum model VAR orde- $p$  dinotasikan dengan VAR( $p$ ), dimana orde- $p$  diperoleh melalui plot korelasi menggunakan analisis *Partial Autocorrelation Function* (PACF). Kemudian secara umum, persamaan model VAR( $p$ ) dapat ditulis sebagai berikut.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \dots + \alpha_p Y_{t-p} + \varepsilon_t \tag{2}$$

Dengan bentuk matriks sebagai berikut [15,17].

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \\ \vdots \\ Y_{mt} \end{bmatrix} = \begin{bmatrix} \alpha_{111} & \alpha_{112} & \dots & \alpha_{11m} \\ \alpha_{121} & \alpha_{122} & \dots & \alpha_{12m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{1m1} & \alpha_{1m2} & \dots & \alpha_{1mm} \end{bmatrix} \begin{bmatrix} Y_{1(t-1)} \\ Y_{2(t-1)} \\ \vdots \\ Y_{m(t-1)} \end{bmatrix} + \dots + \begin{bmatrix} \alpha_{p11} & \alpha_{p12} & \dots & \alpha_{p1m} \\ \alpha_{p21} & \alpha_{p22} & \dots & \alpha_{p2m} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{pm1} & \alpha_{pm2} & \dots & \alpha_{pmm} \end{bmatrix} \begin{bmatrix} Y_{1(t-p)} \\ Y_{2(t-p)} \\ \vdots \\ Y_{m(t-p)} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{mt} \end{bmatrix}$$

Keterangan:

- $m$  : jumlah variabel yang digunakan
- $Y_t$  : vektor variabel berukuran  $m \times 1$  pada waktu  $t$  dan  $t-i, i = 1, 2, \dots, p$
- $\alpha$  : matrix parameter variable berukuran  $m \times m$  untuk setiap  $i=1, 2, \dots, p$
- $\varepsilon_t$  : vektor residu berukuran  $m \times 1$

### 2.1.3 Logistic Regression

Logistik Regresi ini merupakan metode yang paling umum digunakan dalam pendekatan untuk membuat model prediksi probabilitas kejadian suatu peristiwa. Logistik Regresi hanya digunakan jika variabel output dari model yang digunakan didefinisikan sebagai kategori biner. Metode ini memprediksi variable terikat yang berskala dikotomi, yaitu skala nominal yang memiliki dua ketegori, misalnya: Ya dan Tidak, atau Tinggi dan Rendah [6]. Metode ini digunakan untuk mencari hubungan antara satu variabel biner independen dan satu atau lebih variabel independen lainnya. Setiap variabel independen dikalikan dengan bobot dan disimpulkan. Hasilnya akan menambah fungsi sigmoid untuk menemukan hasil antara 0 dan 1. Penting untuk menemukan bobot terbaik atau koefisien regresi. Jadi teknik optimasi digunakan untuk menemukan koefisien dan bobot regresi terbaik [10]. Secara umum, persamaan Regresi Logistik dinyatakan dengan persamaan berikut [18].

$$\ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_i x_i \tag{3}$$

Dengan  $x_i$  adalah variabel prediktor ke- $i$ ,  $\pi(x)$  adalah peluang yang bergantung pada variabel  $x_i$  dan  $\beta_i$  adalah parameter ke- $i$  dengan  $i=1, 2, 3, \dots, n$ . Parameter  $\beta_i$  dapat diestimasi dengan menggunakan fungsi pendugaan maksimum likelihood yang dinyatakan dalam persamaan berikut.

$$L(\beta) = \prod_{i=1}^n \pi(x_i)^{y_i} (1 - \pi(x_i))^{1-y_i} \tag{4}$$

dengan,

$$\pi(x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k)}$$

Keterangan:

- $\Pi(x)$  : probabilitas variabel respon
- exp : fungsi eksponensial (notasi lain  $e^x$ , dimana  $e=2.718\dots$ )
- $\beta$  : koefisien regresi
- $x$  : nilai variabel prediktor

### 2.1.4 Granger Causality

Shochrul (2011) menyatakan bahwa *Granger Causality* atau uji kausalitas Granger merupakan metode untuk menganalisis hubungan antar variabel yang diamati. Apakah variabel tersebut memiliki hubungan dua arah atau hanya mempunyai hubungan satu arah saja [5].

2.1.5 Augmented Dicky Fuller (ADF)

Uji stasioner atau dapat juga disebut sebagai uji akar unit (*Unit Root Test*) merupakan uji yang dilakukan untuk mengetahui data yang diamati merupakan data runtut waktu stasioner atau tidak. Data stasioner merupakan data runtut waktu yang tidak mengandung akar-akar unit. Uji stasioner yang dilakukan pada data yang diamati ini menggunakan metode *Augmented Dickey Fuller* (ADF) yang ditunjukkan dalam persamaan berikut [15].

$$\Delta Y_t = \beta_0 + \phi Y_{t-1} + \sum_{i=1}^p \alpha_i \Delta Y_{t-i} + e_t \tag{5}$$

Keterangan:

- $Y_t$  : data series pada waktu ke- $t$
- $\beta_0$  : konstanta
- $\alpha$  : koefisien signifikansi
- $\phi$  : koefisien regresi  $Y_t$  dan  $Y_{t-1}$ ,  $\rho - 1$
- $e_t$  : komponen residual, dimana  $(e_t \sim N(0, \sigma^2))$

dengan hipotesis,

- $H_0$  :  $p\text{-value} \geq \alpha$  (terdapat unit root, data tidak stasioner)
- $H_1$  :  $p\text{-value} < \alpha$  (tidak terdapat unit root, data stasioner)

Setelah dilakukan pengujian, didapatkan nilai  $p\text{-value}$  yang kemudian dilakukan evaluasi menggunakan hipotesis diatas untuk menyatakan bahwa  $Y_t$  bersifat stasioner atau tidak stasioner [15].

2.1.6 Akaike Information Criterion (AIC)

*Akaike Information Criterion* (AIC) merupakan salah satu metode yang umum digunakan untuk menghitung panjang lag. Dimana jika panjang lag yang digunakan terlalu kecil akan membuat model tidak dapat digunakan karena kurang mampu menjelaskan hubungannya, namun jika panjang lag yang digunakan terlalu besar maka akan menghasilkan estimasi yang kurang efisien karena berkurangnya *degree of freedom*. *Akaike Information Criterion* (AIC) dirumuskan dalam persamaan berikut [17].

$$AIC = T \log|\Sigma| + 2 N \tag{6}$$

Dimana  $|\Sigma|$  adalah determinan matriks residual varians atau kovarian dan  $N$  adalah jumlah total dari parameter yang diestimasi.

2.1.7 Ordinary Least Square (OLS)

Metode *Ordinary Least Square* (OLS) merupakan metode yang digunakan dengan menduga koefisien regresi ( $\beta$ ) dengan meminimumkan kesalahan. Metode ini juga digunakan untuk mengestimasi parameter dengan cara mencari nilai minimal untuk jumlah kuadrat kesalahan antara nilai prediksi dengan nilai aktualnya. Metode ini ditunjukkan dalam persamaan berikut [16].

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{7}$$

Dimana  $\hat{\beta}$  merupakan vektor dari parameter yang diestimasi berukuran  $(p + 1) \times 1$ , untuk  $X$  merupakan matriks variabel prediktor berukuran  $x (p + 1)$  dan  $y$  merupakan vektor observasi dari variabel respon berukuran  $n \times 1$ .

2.1.8 RMSE

Keakuratan metode menjadi ukuran untuk menunjukkan ketepatan metode yang digunakan untuk mereproduksi data. Semakin kecil nilai kesalahan maka semakin baik dan tepat metode yang digunakan [12]. RMSE merupakan salah satu aturan penilaian yang mengukur besarnya rata-rata kesalahan dalam memprediksi suatu kejadian. RMSE adalah akar

kuadrat dari rata-rata perbedaan kuadrat antara prediksi dan aktual pengamatan. RMSE dapat dituliskan dalam persamaan sebagai berikut [4, 21].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \tag{8}$$

Keterangan:  
 RMSE : *Root Mean Square Error*  
 n : jumlah total data  
 Pi : nilai prediksi pada saat *i*  
 Oi : nilai aktual pada saat *i*

2.1.9 Confusion Matrix

*Confusion matrix* merupakan konsep dari pembelajaran mesin yang dapat digunakan untuk menganalisis, mengevaluasi atau memvisualisasikan perilaku model dalam konteks klasifikasi yang diawasi. Matriks ini adalah matriks kuadrat dimana baris mewakili kelas aktual yang sedangkan kolom mewakili kelas prediksi. Untuk menyelesaikan tugas dengan klasifikasi biner, maka *Confusion Matrix* adalah matriks 2 x 2 yang direpresentasikan sebagai berikut [11, 13].

Tabel 1. Confusion Matrix

Actual Class	Assigned Class	
	Positive	Negative
	Positive	True Positives (TP)
Negative	False Positives (FP)	True Negatives (TN)

Sejumlah ukuran kinerja klasifikasi juga dapat didefinisikan sebagai berikut.

$$Recall = \frac{TP}{(TP+FN)} \tag{9}$$

$$Precision = \frac{TP}{(TP+FP)} \tag{10}$$

$$Akurasi = \frac{TP+TN}{TP+TN+FN+FP} \tag{11}$$

3. Sistem yang Dibangun

3.1 Dataset

Dalam penelitian tugas akhir ini, digunakan data harian harga saham dari PT Astra Internasional Tbk pada Januari 2013 – Desember 2018 yang diperoleh dari [www.id.investing.com](http://www.id.investing.com). Pemilihan data harga saham ini didasarkan pada adanya kecenderungan kestabilan nilai dari setiap atributnya.

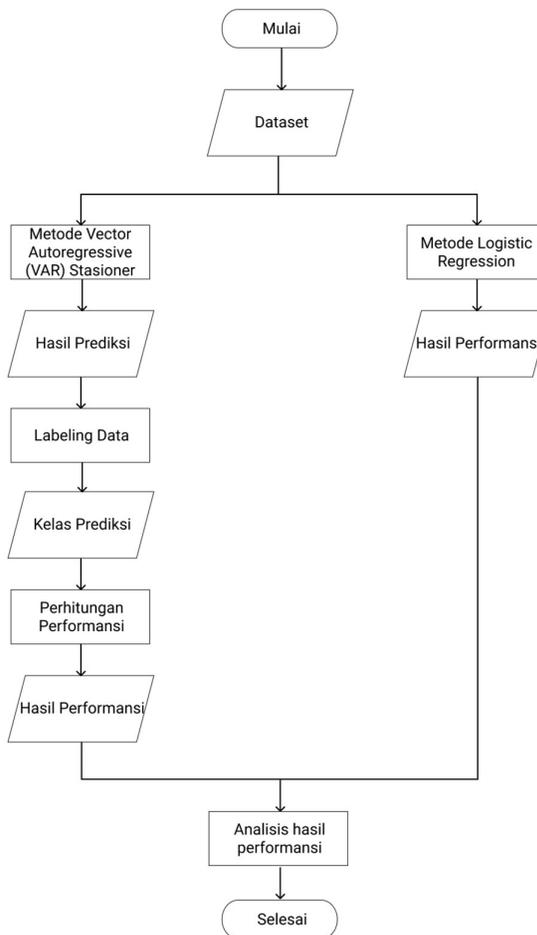
Tabel 2. Sampel Data Harga Saham PT Astra Internasional Tbk

No	Tanggal	Data Harga Penutupan	Data Harga Pembukaan	Data Harga Tertinggi	Data Harga Terendah
1	2/1/2013	7.500	7.600	7.700	7.450
2	24/2/2014	6.775	6.975	7.000	6.750
3	15/10/2015	6.400	6.300	6.550	6.300
4	6/4/2016	7.225	7.250	7.300	7.200
5	28/7/2017	7.950	8.075	8.075	7.900
6	28/12/2018	8.225	8.275	8.400	8.225

3.2 Deskripsi Sistem

3.2.1 Secara Umum

Secara umum, sistem dimulai dengan dataset yang berupa data harian harga penutupan dan harga terendah saham PT Astra Internasional Tbk periode Januari 2013 – Desember 2018. Kemudian dataset diproses ke metode *Vector Autoregressive (VAR) Stasioner* dan *Logistic Regression*. Metode *Vector Autoregressive (VAR) Stasioner* akan menghasilkan nilai prediksi yang selanjutnya dilakukan labeling data sehingga mendapatkan nilai prediksi berupa kelas prediksi. Kelas prediksi akan digunakan sebagai inputan untuk mendapatkan hasil performansi yang direpresentasikan ke dalam bentuk *confusion matrix*. Kemudian akan dilakukan analisis terhadap hasil performansi kedua metode.

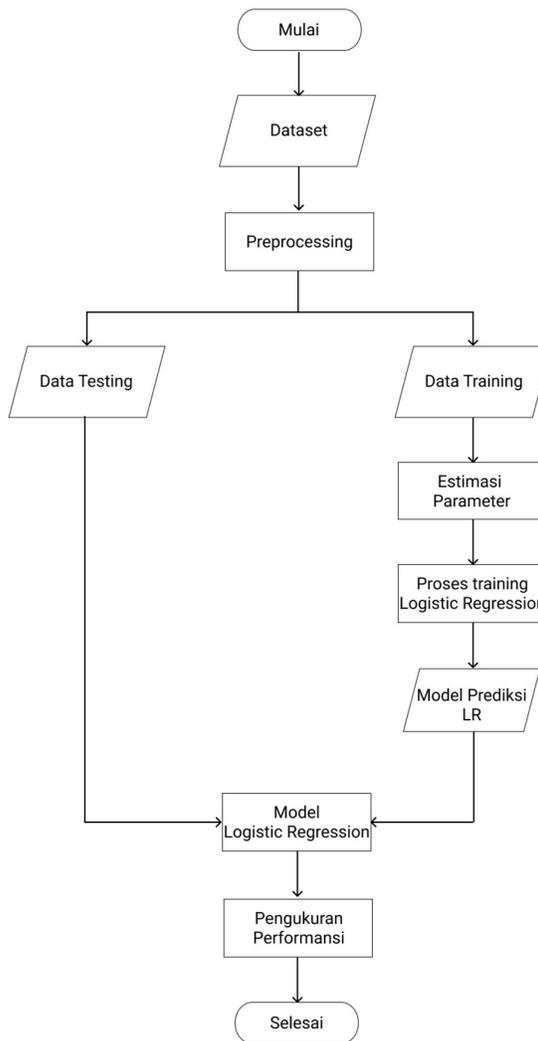


Gambar 1. Deskripsi Sistem Secara Umum

3.2.2 Logistic Regression

Pelaksanaan sistem dimulai dengan data harian saham yaitu data harga penutupan dan data harga terendah yang masing-masing terdiri dari 1456 data yang diambil dari <https://www.id.investing.com/>. Kemudian dilakukan proses preprocessing berupa labeling data menjadi kelas 0 dan 1. Selanjutnya data ini dibagi menjadi 1354 data training dan 102 data testing. Data training digunakan untuk proses training menggunakan *Logistic Regression* yang setelah selesai maka akan didapatkan model prediksi *Logistic Regression*. Selanjutnya data testing digunakan untuk mencoba model yang sudah dibangun sebelumnya. Setelah pemodelan

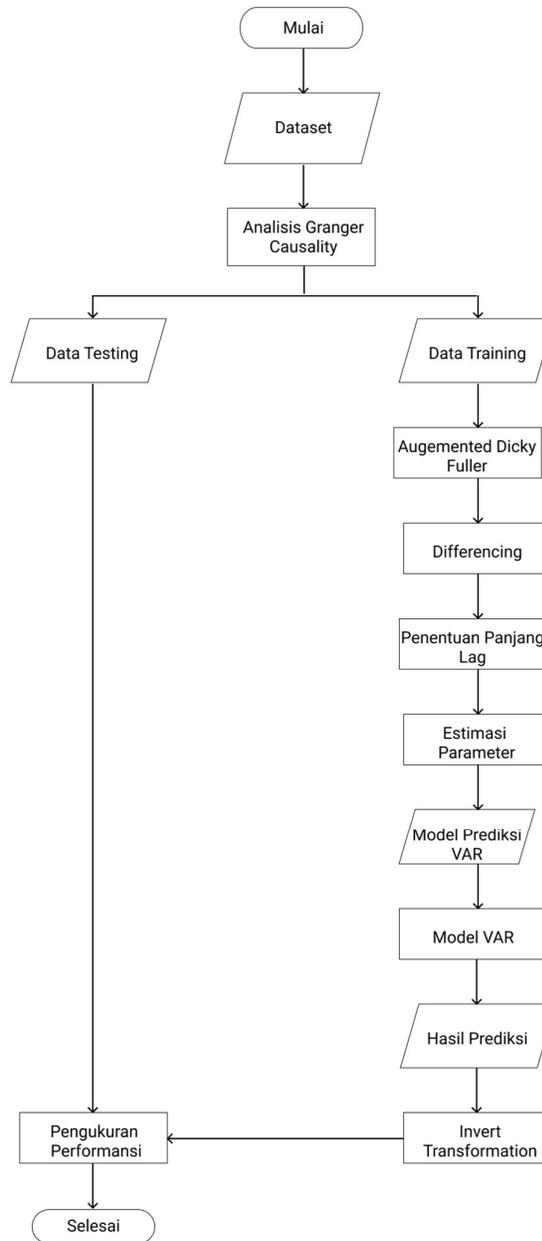
menggunakan *Logistic Regression* ini selesai, maka model ini akan mengeluarkan hasil prediksi yang kemudian akan dihitung nilai performansinya.



**Gambar 2. Deskripsi Sistem Metode Logistic Regression**

3.2.3 Vector Autoregressive (VAR) Stasioner

Sistem dimulai dengan data yang digunakan menggunakan metode *Logistic Regression*. Kemudian dilakukan analisis *Granger Causality* terhadap atribut pentupan dan atribut terendah untuk mengetahui asosiasi atribut dengan pergerakan harga saham. Dari hasil analisis, selanjutnya data dibagi menjadi 1353 data training dan 103 data testing. Selanjutnya pada data training dilakukan uji *Augmented Dicky Fuller* (ADF) untuk mengetahui kestasioneran data tersebut dan melakukan *differencing* pada data yang tidak stasioner. Tahap selanjutnya menentukan panjang lag menggunakan AIC. Selanjutnya dilakukan proses training untuk mendapatkan hasil pemodelan VAR. Setelah pemodelan menggunakan VAR ini selesai, maka model ini akan mengeluarkan hasil prediksi yang kemudian akan dihitung nilai performansinya.



Gambar 3. Deskripsi Sistem Vector Autoregressive (VAR) Stasioner

4. Evaluasi

4.1 Hasil Pengujian Data Saham PT Astra Internasional Tbk menggunakan metode Logistic Regression

4.1.1 Hasil Labeling Data

Tabel 3. Sampel Data Hasil Labeling Metode Logistic Regression

Tanggal	Penutupan	Terendah	Kelas
1/2/2013	7,500	7,450	-
1/3/2013	7,850	7,500	1
1/4/2013	7,850	7,750	0
1/7/2013	7,750	7,700	0
1/8/2013	7,650	7,600	0

1/9/2013	7,700	7,650	1
1/10/2013	7,350	7,300	0
1/11/2013	7,300	7,300	0
1/14/2013	7,500	7,350	1
1/15/2013	7,600	7,500	1

Tabel 3 menunjukkan sampel data yang telah diberi label. Label data direpresentasikan dengan kelas 1 atau 0, dimana 1 menunjukkan adanya kenaikan nilai atribut penutupan yang dibandingkan pada nilai tepat di hari sebelumnya berdasarkan nilai plot korelasi yang diperoleh menggunakan metode *Vector Autoregressive* (VAR) dan kelas 0 menunjukkan keadaan sebaliknya. Selanjutnya, dari 1456 data, dibagi menjadi 1354 data training dan 102 data testing yang kemudian digunakan untuk inputan prediksi harga saham menggunakan metode Logistic Regression.

4.1.2 Hasil Prediksi dan Performansi

**Tabel 4. Sampel Data Perbandingan Hasil Kelas Aktual dan Kelas Prediksi Metode Logistic Regression**

Kelas Aktual	Kelas Prediksi
0	0
1	1
0	0
0	1
0	0
1	1
1	1

4.2 Hasil Pengujian Data Saham PT Astra Internasional Tbk menggunakan metode Vector Autoregressive (VAR) Stasioner

4.2.1 Hasil Granger Causality

**Tabel 5. Hasil Uji Granger Causality**

Variabel Respon	Variabel Prediktor		
		Penutupan	Terendah
	Penutupan	1.0000	0.0545
Terendah	0.0000	1.0000	

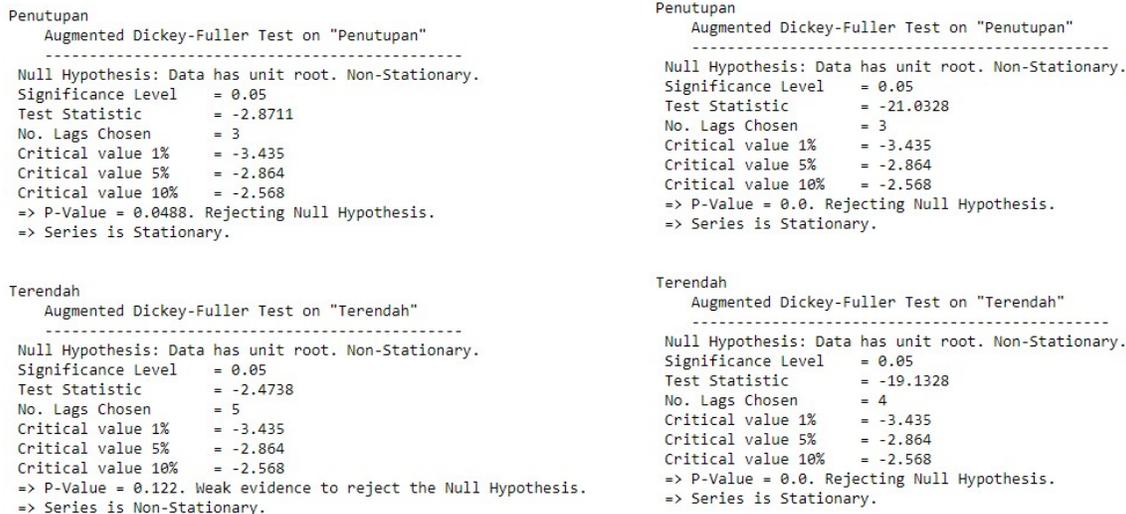
Hipotesis untuk Granger Causality :

H<sub>0</sub> : Penutupan tidak mempengaruhi terendah,  $p\text{-value} < \beta$ , tolak H<sub>0</sub>  
 H<sub>1</sub> : Penutupan mempengaruhi terendah,  $p\text{-value} > \beta$ , terima H<sub>1</sub>

H<sub>0</sub> : Terendah tidak mempengaruhi penutupan,  $p\text{-value} < \beta$ , tolak H<sub>0</sub>  
 H<sub>1</sub> : Terendah mempengaruhi penutupan,  $p\text{-value} > \beta$ , terima H<sub>1</sub>

Granger Causality test digunakan untuk menganalisis asosiasi antar atribut yang diamati. Dari tabel diatas, dapat disimpulkan bahwa nilai  $p\text{-value}$  H<sub>1</sub> lebih besar dibandingkan nilai  $p\text{-value}$  untuk H<sub>0</sub> (0.0545 > 0), maka kesimpulan yang diambil adalah terima H<sub>1</sub> dimana terendah mempengaruhi penutupan.

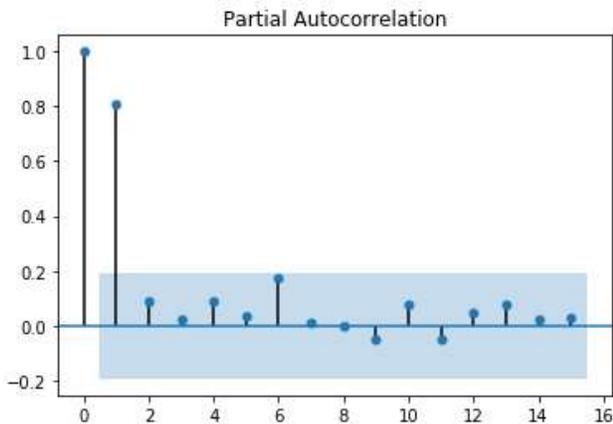
4.2.2 Hasil Transformasi



Gambar 4. Hasil Transformasi

Gambar 4 menunjukkan Augmented Dicky-Fuller Test yang dilakukan pada atribut penutupan dan atribut terendah. Hasil ADF Test menunjukkan bahwa atribut terendah bersifat tidak stasioner, dimana harus dilakukan Differencing test untuk menjadikan atribut terendah bersifat stasioner agar dapat dilanjutkan pada proses selanjutnya. Data yang telah bersifat stasioner ditunjukkan pada gambar sebelah kanan.

4.2.3 Hasil Plot Korelasi



Gambar 5. Hasil Plot Korelasi

Berdasarkan hasil plot korelasi pada data yang sudah stasioner diatas menunjukkan bahwa, terjadi *cut off* dilag ke-2 dengan kondisi dimana nilai batas *lag* tidak berkorelasi berada diantara -0.02 dan 0.02. Artinya, data yang berada di *lag* ke-2 dan seterusnya tidak memiliki korelasi dengan data yang berada di *lag* ke-1 dan antar data dengan jarak 2 *lag* tidak terdapat korelasi. Maka, pada penelitian ini model VAR(*p*) yang digunakan yaitu VAR(1).

## 4.2.4 Hasil Pengujian Panjang Lag

Tabel 6. Hasil Pengujian Panjang Lag

Lag Order	AIC
1	18.475403244418636
2	18.358442808131333
3	18.268449760403676
4	18.240582711148328
5	18.220321720057328
6	18.19370306335493
7	18.17860605733136
8	18.180067758929688
9	18.16596647857317
10	18.161577831534625
11	18.15778580629275
12	18.154946938300256
13	18.152490363685544
14	18.154754083485443
15	18.158879202350057

Tabel 6 menunjukkan hasil pengujian panjang lag menggunakan metode AIC berdasarkan persamaan (6). Panjang lag yang menunjukkan hasil optimum dipilih berdasarkan nilai AIC yang paling minimum. Dengan memperhatikan nilai AIC pada tabel, nilai AIC terendah terjadi ketika ada dilag ke-13. Maka dalam penelitian ini, panjang lag yang digunakan yaitu lag ke-13.

## 4.2.5 Hasil Prediksi

Hasil prediksi menunjukkan keadaan nilai harga saham prediksi yang cenderung lebih rendah dibandingkan nilai aktualnya. Hal ini ditunjukkan dengan nilai performansi RMSE sebesar 908.557 untuk atribut penutupan dan 902.1946 untuk atribut terendah.

## 4.2.6 Hasil Labeling Data

Tabel 7. Sampel Data Hasil Labeling Data Metode Vector Autoregressive (VAR)

Penutupan	Kelas
7114	-
7080	0
7056	0
7015	0
7004	0
7011	1
6992	0
7018	1
7011	0
7004	0
7114	1

Tabel 7 menunjukkan sampel hasil labeling data berdasarkan atribut penutupan yang diperoleh dari hasil prediksi menggunakan metode *Vector Autoregressive* (VAR). Labeling data yang digunakan sama seperti pada labeling data pada metode *Logistic Regression* dimana, label data direpresentasikan sebagai kelas 1 atau 0. Kelas 1 menunjukkan adanya kenaikan nilai jika dibandingkan nilai pada hari sebelumnya dan kelas 0 menunjukkan sebaliknya. Data

ini selanjutnya akan digunakan sebagai inputan nilai prediksi untuk menghitung hasil performansi.

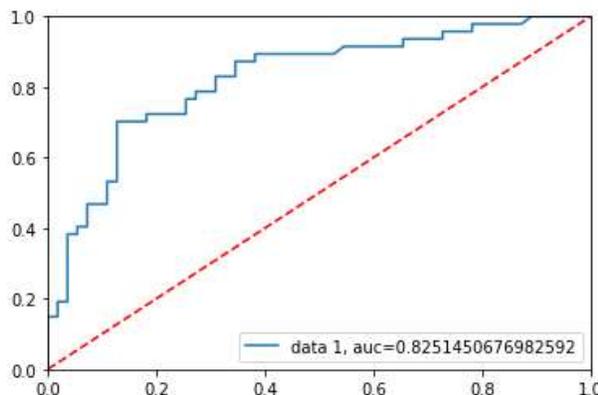
4.3 Hasil Performansi Klasifikasi metode Vector Autoregressive (VAR) Stasioner dan Logistic Regression

4.3.1 Logistic Regression

Tabel 8. Nilai Pengujian Confusion Matrix Metode Logistic Regression

Actual Class	Assigned Class	
	Positive	Negative
Positive	49	6
Negative	25	22

Tabel 8 menunjukkan nilai pengujian *confusion matrix* metode *Logistic Regression*. Model ini menghasilkan nilai *True Positive* (TP) sebesar 49, *False Positive* (FP) sebesar 25, *True Negative* (TN) sebesar 6, dan *False Negative* (FN) sebesar 22. Dari nilai tersebut, didapatkan nilai akurasi sebesar 69.60% , nilai precision 78.57% dan nilai recall sebesar 46.80%. Serta menghasilkan nilai AUC sebesar 0.8251, ditunjukkan pada gambar 7. Nilai ini menunjukkan hasil yang baik karena semakin mendekati nilai 1 merepresentasikan klasifikasi yang baik.



Gambar 6. Grafik ROC

4.3.2 Vector Autoregressive (VAR) Stasioner

Tabel 9. Nilai Pengujian Confusion Matrix Metode Logistic Regression

Actual Class	Assigned Class	
	Positive	Negative
Positive	50	5
Negative	45	2

Tabel 9 menunjukkan nilai pengujian *confusion matrix* metode *Vector Autoregressive* (VAR) Stasioner. Model ini menghasilkan nilai *True Positive* (TP) sebesar 50, *False Positive* (FP) sebesar 45, *True Negative* (TN) sebesar 5, dan *False Negative* (FN) sebesar 2. Dari nilai tersebut, didapatkan nilai akurasi sebesar 50.98% , nilai precision 28.57% dan nilai recall sebesar 4.25%.

5. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, kedua metode ini dapat digunakan sebagai alat untuk memprediksi pergerakan harga saham. Didapatkan hasil akurasi sebesar 70%, precision 78.5% dan recall 47% menggunakan metode Logistic Regression. Sedangkan untuk metode Vector Autoregressive (VAR) Stasioner didapatkan hasil akurasi sebesar 50.98%, precision sebesar 28.57% dan recall 4.25%. Jika dibandingkan hasil

performansinya, baik dari hasil akurasi, presisi dan recall metode Logistic Regression cenderung menghasilkan nilai performansi yang lebih baik dibandingkan metode Vector Autoregressive (VAR) Stasioner pada PT Astra Internasional Tbk yang memiliki perubahan harga tidak signifikan atau cenderung stabil pada tahun 2013 – 2018.