



INNOVATIVE: Journal Of Social Science Research

Volume 3 Nomor 6 Tahun 2023 Page 10238-10257

E-ISSN 2807-4238 and P-ISSN 2807-4246

Website: <https://j-innovative.org/index.php/Innovative>

## Perbandingan Metode Multivariatif GRU dan VAR Berdasarkan Sentimen Investor dan Nilai Kurs Dollar Untuk Prediksi Harga Saham

David Adiyaksa Aquinaldo

Program Studi Teknik Informatika Fakultas Teknologi Informasi

Universitas Kristen Satya Wacana Salatiga

Email : [davidaquinaldo@gmail.com](mailto:davidaquinaldo@gmail.com)

### Abstrak

Prediksi harga saham merupakan salah satu hal yang krusial untuk masyarakat supaya dapat berinvestasi dengan tepat. Dengan mengetahui prediksi harga saham, investasi dapat menghasilkan keuntungan yang maksimal. Prediksi harga saham perlu memperhatikan sentimen investor dan juga faktor eksternal yang dapat mempengaruhi harga saham. Sentimen investor dapat di representasikan dengan menggunakan technical indicator. Untuk meminimalisir sentimen investor dapat melakukan regresi LASSO dari variabel – variabel yang di bangkitkan technical indicator. Faktor lainnya juga ada kurs dolar AS yang secara aktif mempengaruhi nilai saham. Data yang digunakan untuk penelitian ini yaitu harga saham BMRI dari bulan April 2005 sampai Maret 2023. Prediksi menggunakan Multivariate GRU dan Multivariate VAR untuk membandingkan model prediksi mana yang paling akurat untuk memprediksi harga saham dengan faktor eksternal kurs dolar AS. Dari hasil kedua model, Multivariate VAR memiliki hasil yang lebih akurat dari pada model Multivariate GRU.

Kata Kunci : *Multivariate GRU, Multivariate VAR, LASSO, Investasi Saham*

## Abstract

Stock market prediction is one of the most crucial thing for public so they can invest properly. With knowing the price prediction of stock market, people could get the maximum profit from investing. Stock market prediction needs pay attention to investor sentiment and external factor that could influence stock market price. Investor sentiment can be represent as technical indicator. To minimize investor sentiment, LASSO regression could be used for minimalizing the variables. The other factor that can influence stock market price actively is US Dollar exchange rate. This research is using BMRI stock prices data from April 2005 until March 2023. Multivariate GRU and Multivariate VAR are used for predicting and each model are compared to look for the best model. Form this reasearch, Multivariate VAR is more accurate than its competitor.

Keyword ; *Multivariate GRU, Multivariate VAR, LASSO, Stock Investing.*

## PENDAHULUAN

Investasi merupakan hal yang biasa didengar di kalangan masyarakat. Ada banyak jenis investasi yang tersedia di tengah masyarakat, seperti emas antam, crypto coin, NFT, saham, dll. Diantara itu, investasi saham merupakan investasi yang populer ditengah masyarakat Indonesia. Investasi saham merupakan tanda penyertaan modal, baik badan usaha maupun perseorangan pada suatu perusahaan. Dengan menyertakan modal, maka yang bersangkutan dapat memiliki hak atas pendapatan perusahaan dan aset perusahaan. Investasi saham merupakan hal yang berbeda dengan trading saham. Investasi saham lebih mengarah kepada mendapatkan keuntungan jangka panjang dari pengelolaan perusahaan. Keuntungan yang dapat diambil saat investasi saham yaitu capital gain dan deviden. Capital gain adalah selisih dari harga jual dan harga beli saham akibat fluktuasi pasar. Deviden merupakan pembagian laba sesuai kuantitas kepemilikan saham. Nilai dari capital gain dan deviden bisa naik dan turun bergantung pada keuangan perusahaan tersebut. Saham Blue Chip merupakan saham dari perusahaan yang memiliki reputasi tinggi, memiliki reputasi di industri yang digeluti perusahaan tersebut, memiliki pendapatan stabil, dan konsisten dalam membayar deviden. Meskipun investasi saham Blue Chip memiliki reputasi yang baik, tetapi banyak juga masyarakat yang meragukannya dan memilih investasi dipilihan lain, termasuk juga investasi di platform yang bisa saja menipu.

Machine Learning dapat memiliki peran yang besar dalam membantu masyarakat berinvestasi saham. Machine Learning merupakan percabangan dari Artificial Intelligence dan computer science yang fokus kepada pemakaian data dan algoritma untuk mengimitasi cara manusia belajar dan semakin meningkatkan keakurasiannya. Forecasting atau peralaman merupakan bagian algoritma dari Machine Learning, yang berguna untuk memprediksi masa depan dengan menggunakan data historis. Ada berbagai macam

algoritma Forecasting yang dapat digunakan untuk memprediksi harga saham, diantaranya terdapat algoritma peramalan time series, seperti VAR dan GRU yang akan digunakan dalam penelitian. Time series forecasting menggunakan data historis dengan stempel waktu dan melibatkan pembentukan suatu model melalui analisa historis serta menggunakan model tersebut untuk melakukan suatu observasi dan mendorong pengambilan keputusan strategi di masa depan. Oleh karena banyaknya algoritma untuk time series, maka diperlukan penelitian untuk mengetahui algoritma time series mana yang paling mendekati nilai asli dari dataset harga saham.

VAR (Vector Autoregressive) merupakan metode prediksi time series yang memiliki beberapa variabel endogen secara bersamaan, tetapi masing – masing variabel endogen dijelaskan oleh lag dari nilainya sendiri dan variabel lainnya dalam endogen.

GRU (Gated Recurrent Neural Network) merupakan salah satu varian dari RNN (Recurrent Neural Network). Dengan menggunakan struktur gating, dapat menyelesaikan permasalahan RNN yang susah untuk mengatasi akuisisi informasi long-term.

Untuk membandingkan berbagai macam model time-series, bisa dilakukan dengan cara membandingkan nilai error. Metode untuk mengukur nilai error diantaranya terdapat MAD (Mean Absolute Deviation), MSD (Mean Squared Devbiation), MAPE (Mean Absolute Precantage Error).

Tujuan dari penelitian ini yaitu menguji kedua algoritma multivariate time series VAR dan GRU pada dataset harga saham, untuk mendapatkan algoritma terbaik dengan membandingkan nilai error. Prediksi dilakukan dengan mengintegrasikan beberapa variasi indikator teknikal, seperti sentimen investor, yang akan direduksi dimensinya menggunakan LASSO. Penelitian dapat menjadi acuan bagi masyarakat dalam mengambil keputusan untuk berinvestasi saham dengan cara menganalisa capital gain dimasa depan.

Tujuan dari penelitian ini yaitu menguji kedua algoritma time series VAR dan GRU pada dataset harga saham, untuk mendapatkan algoritma yang paling akurat.

Dapat menjadi acuan bagi masyarakat dalam menganalisa capital gain harga saham di masa depan, yang berfungsi untuk pengambilan keputusan berinvestasi saham. Serta dapat mengetahui algoritma forecasting time series yang tepat untuk prediksi harga saham.

## METODE PENELITIAN

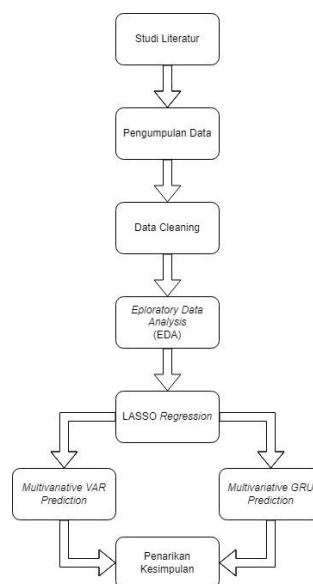
Metode penelitian yang digunakan pada penelitian ini yaitu penelitian kuantitatif. Metode penelitian kuantitatif merupakan penelitian yang sistematis, terencana dan terstruktur dengan jelas dari awal hingga desain penelitiannya (Siyoto dan Sodik, 2015). Metode ini biasanya teknik pengambilan sampel dilakukan secara acak, pengumpulan data

menggunakan instrumen penelitian, analisis data bersifat kuantitatif yang bertujuan untuk menguji hipotesis yang telah ditetapkan (Sugiyono, 2015).

Pada penelitian ini, dataset diambil dari situs yahoo finance, yang menyediakan setiap data harga saham dan kurs mata uang asing dengan akurat. Data yang diambil dalam rentang waktu April 2005 sampai Maret 2023. Data yang telah didapatkan akan dianalisis secara statistik dengan menggunakan algoritma Machine Learning untuk mendapatkan hasil prediksi dari dataset.

### 3.2 Tahap Penelitian

Tahap dari penelitian akan digambarkan dalam sebuah diagram alur untuk mempermudah menyampaikan informasi terhadap langkah – langkah yang akan dilakukan. Berikut diagram alur dari tahap penelitian.



Gambar 1 Diagram Alur Tahap Penelitian

#### *Exploratory Data Analysis*

EDA merupakan proses krusial dari melakukan investigasi awal di data untuk menemukan pola – pola, menemukan anomali dalam data, untuk menguji hipotesis dan mengecek asumsi dengan bantuan *summary* statistik dan representasi grafik.

Hasil dari EDA merupakan suatu grafik. Grafik – grafik dari EDA sebagai berikut:

- a. *Heatmap*, representasi grafik dari data yang menggunakan sistem *color-coding* untuk merepresentasikan nilai yang berbeda.
- b. *Box plot*, menampilkan 5 nomor *summary* dari dataset yaitu nilai minimum, kuartil pertama, median, kuartil ketiga, dan maksimum.
- c. *Distribution plot*, secara visual menilai distribusi dari data sampel dengan cara mengkomparasi distribusi *empirical* data.

### *Least Absolute Shrinkage and Selection Order (LASSO)*

LASSO merupakan metode regresi berganda yang digunakan untuk menyusutkan koefisien taksiran mendekati angka nol (*shrinkage*) dan menyeleksi variabel-variabel independen sehingga menghasilkan model dengan variabel terbaik (*selection operator*).

Penaksir koefisien pada  $(\hat{\beta}_j^{LASSO})$  regresi Lasso diperoleh dengan cara meminimumkan persamaan berikut:

$$\hat{\beta}^{LASSO} = \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2$$

Dengan fungsi  $\sum_{j=1}^p |\beta_j| \leq t$  koefisien kendala. Persamaan diatas dapat juga dituliskan dalam persamaan Lagrange berikut :

$$\hat{\beta}^{LASSO} = \operatorname{argmin} \left\{ \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

Dimana  $y_i$  menyatakan variabel dependen pengamatan ke- $i$ ,  $\beta_0$  merupakan konstanta,  $\beta_j$  merupakan koefisien dari variabel independen ke- $j$ ,  $x_{ij}$  merupakan variabel independen,  $N$  menyatakan banyaknya observasi dan  $p$  merupakan banyaknya variabel independen dalam model.

### *Vector Autoregressive (VAR)*

Metode Vector Autoregressive (VAR) menurut Gujarati (2012) merupakan pemodelan persamaan simultan yang memiliki beberapa variabel endogen secara bersamaan, namun masing-masing variabel endogen dijelaskan oleh lag dari nilainya sendiri dan variabel endogen lainnya dalam model. Model VAR digunakan jika data stasioner pada level.

Secara umum model Vector Autoregressive (VAR) dapat diformulasikan sebagai berikut:

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + A_3 x_{t-3} + \dots + A_p x_{t-p} + e_t$$

$X_t$  = vektor berukuran  $n \times 1$  yang berisi  $n$  peubah dalam model VAR

$A_0$  = vektor intersep berukuran  $n \times 1$

$A_1$  = matriks koefisien berukuran  $n \times n$

$e_t$  = vektor sisaan berukuran  $n \times 1$

### *GATED RECURRENT NEURAL NETWORK (GRU)*

GRU merupakan salah satu varian dari *Recurrent Neural Network* (RNN) yang diperkenalkan oleh Kyunghyun Cho. Pada GRU, pembaharuan (*input*) *Gate* menentukan berapa banyaknya *input* ( $x_t$ ) dan *output* sebelumnya ( $h_t - 1$ ) untuk di teruskan ke *cell*

selanjutnya dan *reset Gate* berguna untuk menentukan seberapa banyak informasi lampau yang akan dilupakan. Isi dari *current memory* memastikan hanya informasi yang relevan dapat lalu ke iterasi selanjutnya, yang ditentukan oleh berat dari  $W$ . Operasi utama di GRU diatur dalam persamaan berikut :

*Update Gate*

$$z_t = \sigma(W_z * [h_{t-1}, x_t])$$

*Reset Gate*

$$r_t = \sigma(W_r * [h_{t-1}, x_t])$$

Setelah *reset* dan *update Gate*, kandidat *status value* dari unit GRU adalah  $\tilde{h}_t$  dan output final *status value* yaitu  $h_t$ :

$$\tilde{h}_t = \tanh(W_{h\sim t} * [r_t * h_{t-1}, x_t])$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t$$

## HASIL DAN PEMBAHASAN

Pengujian dilakukan dengan menggunakan dataset dari data historis harga saham BMRI dan harga kurs dolar dengan rentang waktu dari 3 Juli 2005 sampai dengan 3 Mei 2023. Data diambil dari situs yahoo finance. Berikut merupakan bagian dari dataset.

Tabel 1 Data Historis BMRI – USD

	Date	Open	High	Low	Close	Adj Close	Volume	USD
0	3/7/2005	909.555725	924.305298	904.639221	919.388794	524.591980	89409830	9322.5
1	3/8/2005	919.388794	919.388794	904.639221	914.472290	521.786560	61565935	9376.8
2	3/9/2005	909.555725	909.555725	894.806213	899.722717	513.370544	52407013	9328.9
3	3/10/2005	899.722717	904.639221	889.889648	894.806213	510.565399	81817057	9358.2
4	3/14/2005	894.806213	894.806213	865.307068	875.140137	499.344147	104920809	9366.8
5	3/15/2005	875.140137	875.140137	830.891479	840.724487	479.706970	251535765	9348.2
6	3/16/2005	845.640991	855.474060	845.640991	850.557556	485.317627	149614029	9334.6
7	3/17/2005	850.557556	870.223633	850.557556	865.307068	493.733551	97272103	9360.0
8	3/18/2005	865.307068	865.307068	845.640991	850.557556	485.317627	52774143	9352.2
9	3/21/2005	855.474060	860.390564	835.807983	840.724487	479.706970	54803018	9405.0

### Persiapan Data

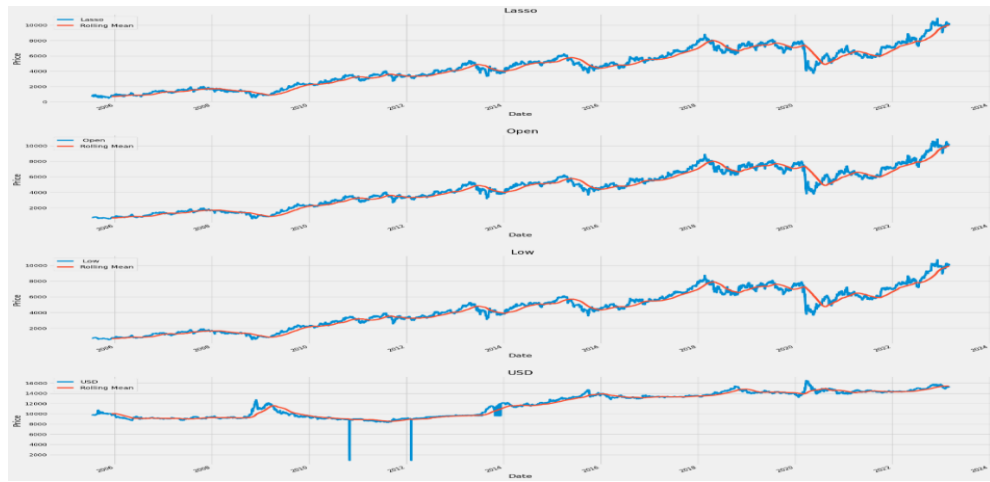
Persiapan data yang pertama kali dilakukan adalah pengecekan data untuk mengeliminasi data yang null dengan menggunakan library Pandas yang tersedia di Python.

Dimana `bmri_df` merupakan data dan perintah `isnull()` merupakan fitur dari library pandas untuk melihat apakah ada data yang null dalam dataset. Dari perjalanan kode tersebut, tidak terdapat data yang null di setiap kolom.

Selanjutnya, untuk melihat keterkaitan satu data dengan yang lain dapat menggunakan Exploratory Data Analysis.

#### 1. Rolling Mean

Berikut merupakan gambar dari setiap plot dari dataset dengan Rolling Mean.

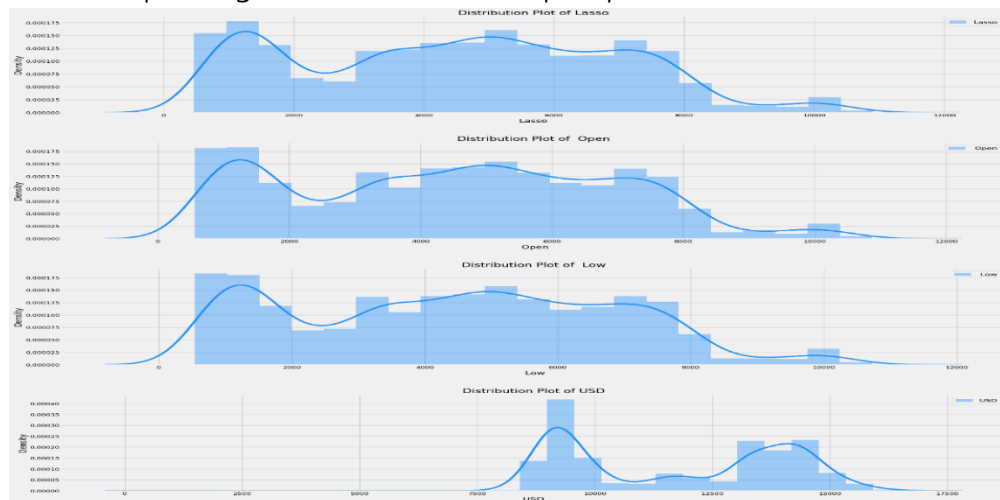


Gambar 2 Grafik Plot dari Setiap Kolom

Jika dilihat dari gambar 2, dapat dilihat bahwa setiap data memiliki pola yang hampir sama dengan pengecualian fluktuasi data short-term. Di beberapa titik terlihat memiliki fluktuasi yang sama.

## 2. Distribusi Plot

Berikut merupakan gambar dari distribusi plot pada dataset.

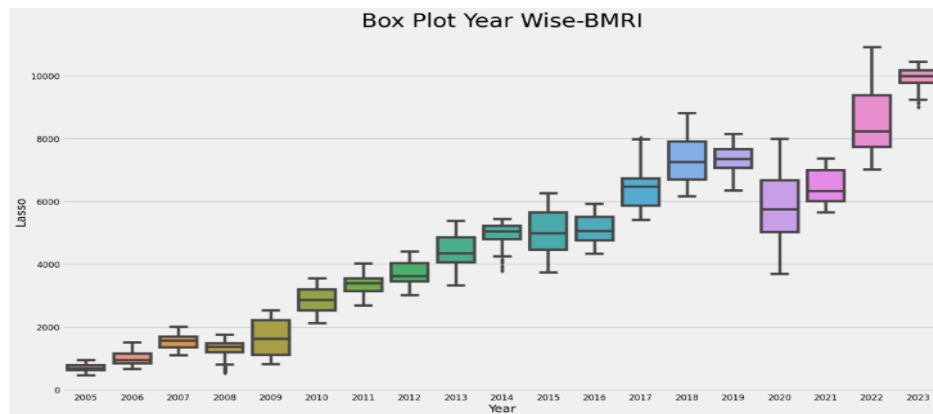


Gambar 3 Distribusi Plot dari Dataset

Dari gambar 3, terlihat dari distribusi setiap variabel terdapat dua puncak pada bentuk histogram tersebut, hal ini dapat mempengaruhi hasil dari prediksi karena distribusi plot tidak mengikuti distribusi normal. Pada Lasso, Low, dan Open terlihat bahwa variabel terbanyak terletak di harga 0 – 2000 dan di USD terdapat di 7500-10000.

## 3. Box Plot

Berikut merupakan gambar dari Box plot pada dataset.

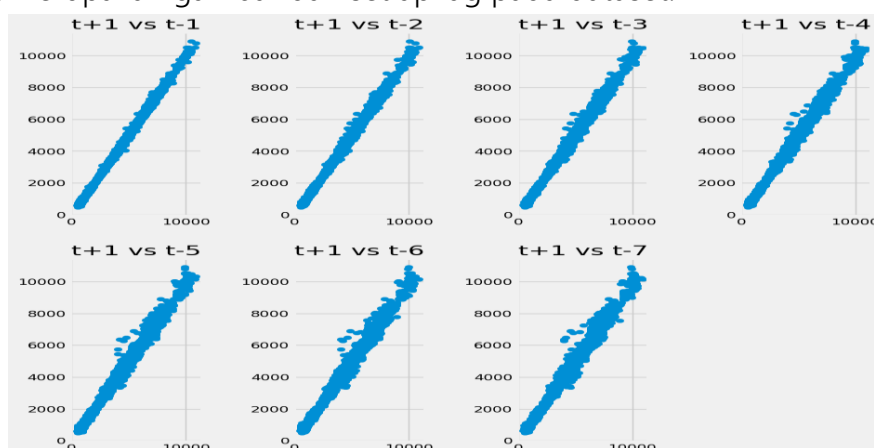


Gambar 4 Box Plot pada Dataset

Pada gambar 4 terlihat pada tahun 2020 dan 2022 memiliki nilai harga yang memiliki nilai minimum dan maksimum paling jauh nilainya, hal tersebut dapat mempengaruhi hasil prediksi oleh karena fluktuasi yang tinggi. Dari box plot diatas juga dapat dilihat persebaran data terbesar ada di harga 0-2000.

#### 4. Series Lag Relationship (ACF)

Berikut merupakan gambar dari setiap lag pada dataset.

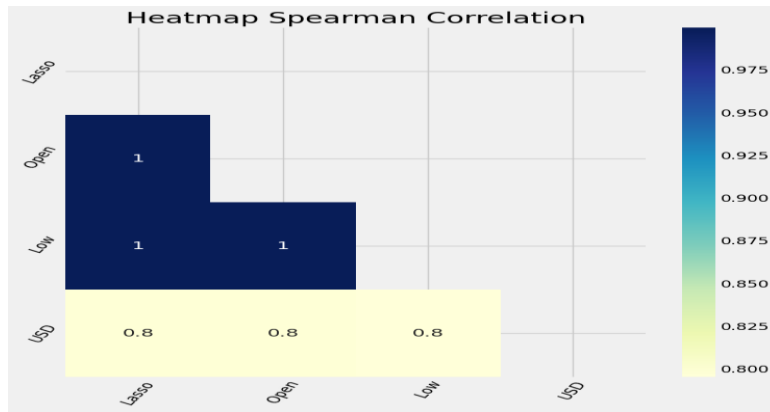


Gambar 5 Series Lag

Gambar 5 menjelaskan bahwa pada lag 1-4 hanya menunjukkan sedikit perubahan pada lag nya dan pada lag 5-7 mulai terlihat lag yang menjauhi angka 0, dimana lag = 0 merupakan korelasi yang sempurna.

#### 5. Heatmap Spearman Correlation

Berikut gambar Heatmap Spearman dari dataset.



Gambar 6 Heatmap Spearman

Pada gambar 6, setiap kotak pada axis menandakan korelasi antara variabel, korelasi tersebut berjarak antara -1 sampai +1. Jika nilai di kotak bernilai dekat dengan nol, menandakan tidak ada tren linier pada dua variabel tersebut, semakin dekat dengan 1 menandakan semakin memiliki korelasi yang positif dan jika mendekati -1 salah satu variabel akan berkurang dan yang lain akan meningkat. Heatmap diatas menandakan setiap variabel terhadap yang lain memiliki korelasi yang positif.

### LASSO Regression

Berikut merupakan gambar dari koefisien dari setiap model dan teknikal indikator yang dipakai :

```
{ 'Open': 0.1771780395008312,
  'Low': 0.5995296083652067,
  'Close': 0.22438601683524206,
  'Adj Close': 0.0028802483163845547,
  'Volume': 7.6129218362333e-10,
  'USD': -0.0001989953963697748,
  'MA': 0.026622489429413546,
  'EMA': 0.8416633269893552,
  'DEMA': 0.11579355555959058,
  'HT_TRENDLINE': -0.0161663373955634,
  'KAMA': -0.018641790283117875,
  'MIDPOINT': -0.019351597596548926,
  'MIDPRICE': -0.029201240569062148,
  'SAR': 9.316170733989696e-06,
  'SAREXT': -3.593388416503019e-05,
  'SMA': -0.06119280005347054,
  'T3': 0.7072004894437253,
  'TEMA': 0.1580815970608144,
  'TRIMA': 0.027568075818444653,
  'WMA': 0.2470905318023311,
  'RSI': -0.03074844667002254,
  'ADX': 0.11273425316364505,
  'ADXR': -0.031082049074192505,
  'BOP': 1.25755779491008,
  'CCI': 0.11491484249547172,
  'DX': 0.0153886037808402,
  'TRIX': -0.117090737882568,
  'MFI': 0.009840430862051841,
  'MMI': -0.01114170007027783,
  'PPO': 0.33529732728199746,
  'ROC': 0.19077621930870248,
  'WILLR': -0.5220522447226564,
  'EMA12': -1.494647477942559,
  'EMA26': -0.48081179277138003,
  'MACD': -1.013835684100957,
  'OBV': 9.222100800739602e-11,
  'ADOSC': -3.220326804225314e-08,
  'ATR': 0.06452170189940876,
  'NATR': -0.02990084555944089,
  'TRANGE': 0.660870638407893,
  'HT_DCPERIOD': -0.014863258482065642},
  { 'Open': 0.7773555126831194,
    'High': 0.14865745231769323,
    'Low': 0.00768803296753212,
    'Adj Close': 0.006010374274609148,
    'Volume': 0.0,
    'USD': 0.0,
    'MA': 0.0,
    'EMA': 0.0,
    'DEMA': 0.01016162162137534,
    'HT_TRENDLINE': 0.01647347869361419,
    'KAMA': 0.0011463689008964946,
    'MIDPOINT': 0.031507950928575786,
    'MIDPRICE': 0.0,
    'SAR': 0.0033722179215916587,
    'SAREXT': 0.00351382237875434,
    'SMA': 0.0,
    'T3': 0.0,
    'TEMA': 0.0,
    'TRIMA': 0.0,
    'WMA': 0.0,
    'RSI': 0.1312878159169711,
    'ADX': 0.0,
    'ADXR': 0.0,
    'BOP': 129.64754183087308,
    'CCI': 0.0,
    'DX': 0.0,
    'TRIX': 0.0,
    'MFI': 0.0,
    'MMI': 0.04290994703513481,
    'PPO': 0.0,
    'ROC': 0.0,
    'WILLR': 0.09133682998641954,
    'EMA12': 0.0,
    'EMA26': 0.0,
    'MACD': 0.0,
    'OBV': 0.0,
    'ADOSC': 0.0,
    'ATR': 0.0,
    'NATR': 0.0,
    'TRANGE': 0.0,
    'HT_DCPERIOD': 0.0}
```

Gambar 7 Coefficient Baseline Model

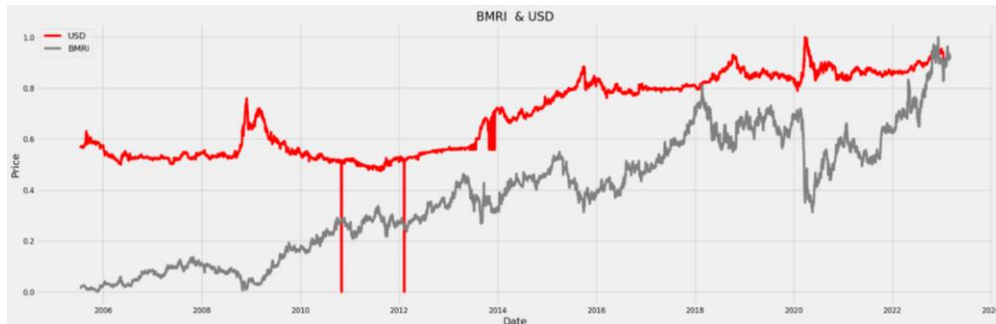
Gambar 8 Coefficient Lasso Model

Model regresi linier di *Baseline model* mendapat MSE 381 dan *Lasso model* mendapat MSE sebesar 949. Jika dilihat dari kedua koefisien, terlihat pada *Lasso model* meskipun nilai MSE lebih tinggi, tetapi koefisien lebih terminimalisir. Hal itu berpengaruh pada model prediksi oleh karena koefisien yang mempengaruhi model jauh lebih sedikit, sehingga lebih akurat.

## Multivariate Gated Recurrent Unit (GRU)

Model ini menggunakan Multivariate GRU dengan 4 tipe model yaitu Vanilla model, Zscore Outlier Removal, Financial Outlier Removal, dan Mahalanobis Outlier Removal. Model akan memprediksi harga saham yang sudah di regresi dengan harga USD di kurun waktu yang sama.

Library Keras digunakan untuk membangun model GRU. Berikut merupakan grafik perbandingan antara harga saham BMRI dan harga USD.



Gambar 9 Grafik BMRI dan USD

Dari gambar 9 terlihat dari pola USD dan BMRI memiliki tren naik dan di momen tertentu seperti di sekitar tahun 2008-2010 terlihat ketika USD naik BMRI menurun. Hal tersebut menunjukkan adanya keterkaitan antar kedua data.

Untuk model Multivariate GRU, data perlu diubah menjadi 3D array sehingga diperlukan untuk reshaping data train dan test. Reshaping Test dan Train Data akan membentuk data supaya dapat di proses oleh GRU. Data tersebut akan menjadi 3D array dengan 60 timesteps.

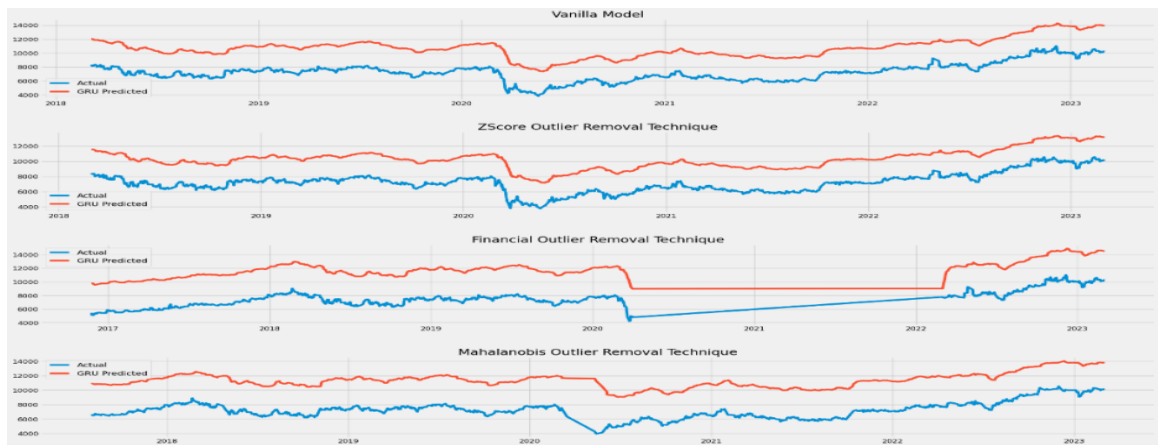
Test data akan diprediksi setelah data train sudah selesai dalam proses training nya, lalu data akan di inverse ke posisi semula.

Berikut hasil dari training data GRU dengan epoch = 50.

Tabel 2 Hasil Training Model GRU

Model	Layer	Neurons	MAE	R2 Score	Train Time
Financial Outlier GRU	4	32	4335,323	-12,296	582
Zscore Outlier GRU	4	32	2975,479	-4,504	588
Mahalanobis Outlier GRU	4	32	4334,318	-17,439	549
Vanilla Model	4	32	3365,413	-5,852	583

Berikut merupakan grafik dari setiap model.



Gambar 10 Grafik Hasil Prediksi GRU

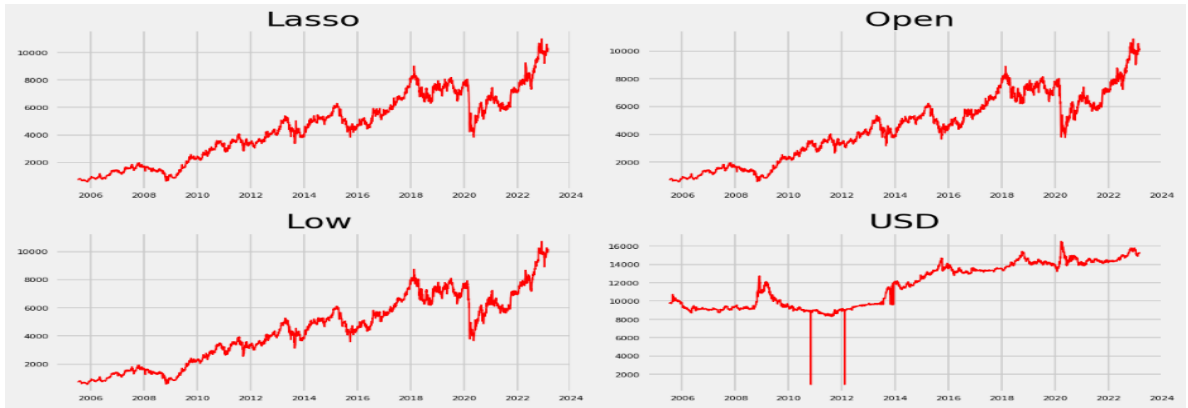
Dari gambar 10, terlihat oleh karena perbedaan nilai dari data, pola dari nilai asli dan prediksi memiliki arah dan fluktuasi yang tidak jauh berbeda. Akan tetapi, antara nilai harga dan prediksi memiliki nilai error yang tinggi oleh karena perbedaan dari harga dari BMRI yang sangat jauh. Meskipun telah menggunakan scaler, nilai error tetap besar. Dataset di tahun 2005 memiliki harga dari BMRI yang bernilai Rp 932,00 dan 2023 bernilai Rp5.250,00, yang berarti memiliki kenaikan harga 563,30% dari tahun 2005. Hal ini sangat berpengaruh ke dataset karena multivariatir tidak hanya memakai data BMRI, tapi juga memakai USD yang kenaikan harga tidak sebesar BMRI. Pada tahun 2005 nilai USD sebesar Rp 9.322,00 dan pada tahun 2023 memiliki nilai Rp 14.755,00, yang berarti hanya memiliki kenaikan harga sebesar 158,28% dari tahun 2005. Tetapi, terlihat dari pola bahwa USD berpengaruh terhadap BMRI oleh karena pola yang memiliki kesamaan tren dan fluktuasi di titik yang sama.

#### Multivariate Vector Autoregression (VAR)

Pada model prediksi Multivariate VAR menggunakan dataset yang sama dengan yang digunakan pada model Multivariate GRU. Dataset memiliki kolom Date, Lasso, Open, Low, USD.

Untuk model ini menggunakan library dari statsmodel. VARMAX dari statsmodel digunakan oleh karena dapat mengubah data yang non stasionary dan dipaksakan menjadi data yang stasionary. Berbeda dengan VAR biasa yang tidak dapat mengubah data menjadi stasionary.

Berikut merupakan grafik dari setiap data pada dataset.



Gambar 11 Grafik Data VAR

Model Multivariate VAR, data harus stasionary. Pengecekan stasionarity data, dapat dilakukan Augmented Dickey-Fuller (ADF) test di python.

Berikut hasil dari ADF test:

```

BMRI
ADF Statistic: -35.2208985142786
p-value: 0.0
-----

USD
ADF Statistic: -21.470710442528393
p-value: 0.0

```

Data dikatakan stasionary ketika p-value memiliki nilai kurang dari 5% atau 0,05. Dari hasil ADF test, data memiliki p-value kurang dari 0,05, dengan demikian data dapat dikatakan stasionary.

Data dalam perubahan nilainya dapat juga dipengaruhi oleh pergerakan data lainnya. Untuk mengetahui tentang apakah data terpengaruh oleh data yang lain, dapat menggunakan Granger causality test yang dapat digunakan untuk menguji data apakah saling mempengaruhi. Granger causality test dapat digunakan melalui library statsmodel.

Berikut merupakan hasil dari Granger causality test.

<pre> BMRI causes USD? ----- </pre>	<pre> number of lags (no zero) 1 ssr based F test:      F=38.6417 , p=0.0000 , df_denom=4371, df_num=1 ssr based chi2 test:  chi2=38.6682 , p=0.0000 , df=1 </pre>
<pre> Granger Causality </pre>	

likelihood ratio test:  $\chi^2=38.4983$  ,  
 $p=0.0000$  ,  $df=1$   
parameter F test:  $F=38.6417$  ,  
 $p=0.0000$  ,  $df_{denom}=4371$ ,  $df_{num}=1$

#### Granger Causality

number of lags (no zero) 2  
ssr based F test:  $F=13.7039$  ,  
 $p=0.0000$  ,  $df_{denom}=4368$ ,  
 $df_{num}=2$   
ssr based  $\chi^2$  test:  $\chi^2=27.4392$  ,  
 $p=0.0000$  ,  $df=2$   
likelihood ratio test:  $\chi^2=27.3534$  ,  
 $p=0.0000$  ,  $df=2$   
parameter F test:  $F=13.7039$  ,  
 $p=0.0000$  ,  $df_{denom}=4368$ ,  
 $df_{num}=2$

#### Granger Causality

number of lags (no zero) 3  
ssr based F test:  $F=9.8744$  ,  
 $p=0.0000$  ,  $df_{denom}=4365$ ,  
 $df_{num}=3$   
ssr based  $\chi^2$  test:  $\chi^2=29.6706$  ,  
 $p=0.0000$  ,  $df=3$   
likelihood ratio test:  $\chi^2=29.5704$  ,  
 $p=0.0000$  ,  $df=3$   
parameter F test:  $F=9.8744$  ,  
 $p=0.0000$  ,  $df_{denom}=4365$ ,  
 $df_{num}=3$

#### Granger Causality

number of lags (no zero) 4  
ssr based F test:  $F=9.6160$  ,  
 $p=0.0000$  ,  $df_{denom}=4362$ ,

$df_{num}=4$

ssr based  $\chi^2$  test:  $\chi^2=38.5433$  ,  
 $p=0.0000$  ,  $df=4$

likelihood ratio test:  $\chi^2=38.3743$  ,  
 $p=0.0000$  ,  $df=4$

parameter F test:  $F=9.6160$  ,  
 $p=0.0000$  ,  $df_{denom}=4362$ ,  
 $df_{num}=4$

USD causes BMRI?

-----

#### Granger Causality

number of lags (no zero) 1  
ssr based F test:  $F=2.8766$  ,  
 $p=0.0899$  ,  $df_{denom}=4371$ ,  $df_{num}=1$   
ssr based  $\chi^2$  test:  $\chi^2=2.8786$  ,  
 $p=0.0898$  ,  $df=1$   
likelihood ratio test:  $\chi^2=2.8776$  ,  
 $p=0.0898$  ,  $df=1$   
parameter F test:  $F=2.8766$  ,  
 $p=0.0899$  ,  $df_{denom}=4371$ ,  $df_{num}=1$

#### Granger Causality

number of lags (no zero) 2  
ssr based F test:  $F=2.3785$  ,  
 $p=0.0928$  ,  $df_{denom}=4368$ ,  
 $df_{num}=2$   
ssr based  $\chi^2$  test:  $\chi^2=4.7624$  ,  
 $p=0.0924$  ,  $df=2$   
likelihood ratio test:  $\chi^2=4.7598$  ,  
 $p=0.0926$  ,  $df=2$   
parameter F test:  $F=2.3785$  ,  
 $p=0.0928$  ,  $df_{denom}=4368$ ,

df\_num=2

Granger Causality

number of lags (no zero) 3

ssr based F test: F=1.6345 ,

p=0.1792 , df\_denom=4365, df\_num=3

ssr based chi2 test: chi2=4.9114 ,

p=0.1784 , df=3

likelihood ratio test: chi2=4.9086 ,

p=0.1786 , df=3

parameter F test: F=1.6345 ,

p=0.1792 , df\_denom=4365, df\_num=3

Granger Causality

number of lags (no zero) 4

ssr based F test: F=1.1977 ,

p=0.3096 , df\_denom=4362,

df\_num=4

ssr based chi2 test: chi2=4.8008 ,

p=0.3084 , df=4

likelihood ratio test: chi2=4.7981 ,

p=0.3086 , df=4

parameter F test: F=1.1977 ,

p=0.3096 , df\_denom=4362,

df\_num=4

Jika p-value memiliki nilai kurang dari 0,05, maka data terpengaruh oleh data satunya. Pada hasil test tersebut, dapat terlihat bahwa p-value pada granger 1 memiliki nilai kurang dari 0,05 yang berarti data BMRI dipengaruhi oleh USD dalam pergerakan datanya. Untuk hasil dari granger 2, terlihat bahwa p-value memiliki nilai diatas 0,05 yang berarti data USD tidak terpengaruh oleh data BMRI.

Selanjutnya, data dipisah menjadi train data dan test data, dengan train data memiliki jumlah 4275 baris dan test data memiliki 100 baris. Jumlah lag yang optimal untuk model VAR dapat dicari dengan menggunakan fungsi VAR yaitu select order.

Tabel 3 Hasil Lag Selection Criteria

VAR Order Selection (* highlights the minimums)				
=====				
	AIC	BIC	FPE	HQIC
-----				
0	20.31	20.32	6.645e+08	20.32
1	20.08	20.09	5.270e+08	20.09
2	19.99	20.01	4.810e+08	20.00
3	19.94	19.96	4.561e+08	19.95
4	19.90	19.93*	4.393e+08	19.91
5	19.90	19.93	4.375e+08	19.91*
6	19.90	19.94	4.380e+08	19.91

7	19.90	19.94	4.385e+08	19.91
8	19.90*	19.95	4.373e+08*	19.91
9	19.90	19.95	4.373e+08	19.92
10	19.90	19.96	4.379e+08	19.92
11	19.90	19.97	4.386e+08	19.92
12	19.90	19.97	4.391e+08	19.93
13	19.90	19.98	4.390e+08	19.93
14	19.90	19.99	4.396e+08	19.93
15	19.90	20.00	4.403e+08	19.94
16	19.90	20.00	4.406e+08	19.94
17	19.90	20.01	4.411e+08	19.94
18	19.91	20.02	4.418e+08	19.95
19	19.91	20.02	4.425e+08	19.95
20	19.91	20.03	4.431e+08	19.95

-----

Dari hasil diatas, lag order nomer 8 memiliki lag yang terbaik oleh karena nilai AIC dan FPE yang paling kecil. AIC, BIC, FPE dan HQIC merupakan lag selection criteria untuk menentukan mana lag yang terbaik.

Berikut hasil dari fitting model :

Tabel 4 Model Multivariate VAR

Statespace Model Results			Time:	19:05:45	HQIC
=====			628723.350		
=====			Sample:	0	
=====				- 4275	
====			Covariance Type:	opg	
Dep. Variable:	['USD', 'Lasso']	No.	=====		
Observations:	4275		=====		
Model:	VAR(8)	Log	=====		
Likelihood	-314283.105		=====		
	+ intercept	AIC	Ljung-Box (L1) (Q):	4161.84, 4157.37	
			Jarque-Bera (JB):	46.89, 21.12	
628640.209			Prob(Q):	0.00, 0.00	
Date:	Mon, 01 May 2023	BIC	Prob(JB):	0.00, 0.00	
628875.549					

Heteroskedasticity (H):	0.76, 0.91
Skew:	-0.01, 0.03
Prob(H) (two-sided):	0.00, 0.08
Kurtosis:	3.51, 2.66
Results for equation	
USD	
=====	
=====	
=====	
====	
	coef    std err    z    P> z
[0.025    0.975]	
-----	
-----	
intercept	1.166e+04    42.572    273.861
0.000	1.16e+04    1.17e+04
L1.USD	-0.0812    0.006    -13.700
0.000	-0.093    -0.070
L1.Lasso	-0.0362    0.013    -2.694
0.007	-0.062    -0.010
L2.USD	-0.1634    0.007    -22.639
0.000	-0.178    -0.149
L2.Lasso	0.4834    0.017    27.701
0.000	0.449    0.518
L3.USD	-0.1192    0.007    -17.706
0.000	-0.132    -0.106
L3.Lasso	0.4585    0.019    24.781
0.000	0.422    0.495
L4.USD	-0.0577    0.007    -8.247
0.000	-0.071    -0.044
L4.Lasso	0.3236    0.019    17.175
0.000	0.287    0.361
L5.USD	0.0029    0.006    0.473
0.636	-0.009    0.015
L5.Lasso	0.2041    0.019    10.989

0.000	0.168	0.241	
L6.USD	0.0330	0.007	4.644
0.000	0.019	0.047	
L6.Lasso	0.0506	0.018	2.845
0.004	0.016	0.085	
L7.USD	0.0328	0.008	4.079
0.000	0.017	0.049	
L7.Lasso	-0.1968	0.018	-11.138
0.000	-0.231	-0.162	
L8.USD	0.0441	0.007	6.075
0.000	0.030	0.058	
L8.Lasso	-0.4009	0.013	-31.813
0.000	-0.426	-0.376	
Results for equation			
Lasso			
=====			
=====			
=====			
====			
	coef    std err    z    P> z		
[0.025    0.975]			
-----			
-----			
intercept	4410.9624	22.817	193.319
0.000	4366.242	4455.683	
L1.USD	-0.1693	0.001	-115.437
0.000	-0.172	-0.166	
L1.Lasso	0.1357	0.005	26.373
0.000	0.126	0.146	
L2.USD	-0.0903	0.003	-33.840
0.000	-0.096	-0.085	
L2.Lasso	0.3715	0.007	53.758
0.000	0.358	0.385	
L3.USD	-0.0334	0.003	-11.716
0.000	-0.039	-0.028	

L3.Lasso	0.2366	0.007	32.092
0.000	0.222	0.251	
L4.USD	-0.0201	0.003	-7.212
0.000	-0.026	-0.015	
L4.Lasso	0.0676	0.008	8.951
0.000	0.053	0.082	
L5.USD	-0.0328	0.002	-16.039
0.000	-0.037	-0.029	
L5.Lasso	-0.0319	0.007	-4.379
0.000	-0.046	-0.018	
L6.USD	-0.0463	0.002	-22.294
0.000	-0.050	-0.042	
L6.Lasso	-0.0136	0.007	-1.956
0.050	-0.027	2.89e-05	
L7.USD	-0.0401	0.002	-16.159
0.000	-0.045	-0.035	
L7.Lasso	0.1602	0.007	23.228
0.000	0.147	0.174	
L8.USD	-0.0407	0.002	-18.336

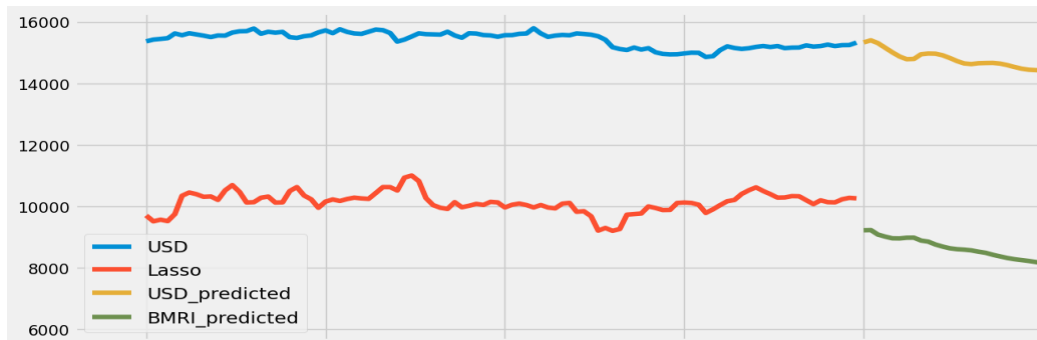
0.000	-0.045	-0.036	
L8.Lasso	0.3485	0.005	70.267
0.000	0.339	0.358	
		Error covariance	
matrix			
=====			
=====			
=====			
=====			
		coef	std err
z	P> z	[0.025	0.975]
-----			
-----			
sqrt.var.USD	230.8358		0.294
784.685	0.000	230.259	231.412
sqrt.cov.	-4.7286	0.442	-10.705
0.000	-5.594	-3.863	USD.Lasso
sqrt.var.Lasso	90.1903	0.280	321.763
0.000	89.641	90.740	

Berikut sebagian dari hasil prediksi :

Tabel 5 Hasil Prediksi Model Multivariate VAR

	USD_predicted	BMRI_predicted
4275	15328.479020	9221.598159
4276	15397.481943	9228.037881
4277	15303.629289	9080.983222
4278	15155.061790	9014.559004
4279	15005.486437	8961.339991
...	...	...
4370	13106.753864	6197.095953
4371	13093.850551	6178.051457
4372	13081.058086	6159.201588
4373	13068.384319	6140.546826
4374	13055.839947	6122.084173

Berikut grafik dari hasil prediksi.



Gambar 12 Grafik Hasil Prediksi Multivariate VAR

Pada grafik diatas, hasil dari prediksi memiliki pola yang menurun, hal ini dapat disebabkan oleh data dari tahun 2020 - 2022 yang memiliki banyak fluktuasi menurun dan hal ini memiliki kemungkinan besar untuk mempengaruhi hasil prediksi. USD predicted merupakan hasil prediksi USD terhadap BMRI dan BMRI predicted merupakan hasil prediksi BMRI terhadap USD. Untuk mengetahui keakuratan hasil prediksi, perlu dilakukan test untuk menilai keakuratan data dengan menggunakan metode MAE.

Keakuratan dari prediksi USD lebih tinggi jika dibandingkan oleh prediksi BMRI. Terlihat dari gambar 12, pada grafik hasil prediksi, USD predicted memiliki pola yang hampir sama dengan data test USD. Untuk BMRI, pola dari hasil prediksi lebih menjorong ke bawah dari data asli.

## SIMPULAN

Dari hasil penelitian, dapat disimpulkan bahwa untuk dataset harga saham BMRI dan nilai kurs USD, model *Multivariate VAR* lebih akurat daripada *Multivariate GRU*. Nilai *error MAE* pada hasil prediksi *Multivariate VAR* sebesar 2701,187 dan nilai *error MAE* hasil prediksi *Multivariate GRU* sebesar 2975,479. Terdapat banyak faktor yang mempengaruhi data, seperti dataset BMRI yang memiliki rentang nilai minimum dan maksimum yang tinggi, menyebabkan hasil prediksi dengan *multivariate GRU* memiliki nilai *error* tinggi. Untuk perbedaan *multivariate GRU* dan *VAR*, yaitu prediksi dalam *GRU* dapat memprediksi dalam beberapa dimensi, dalam penelitian ini 4 dimensi, sedangkan *VAR* hanya dapat memprediksi dalam 2 dimensi. Model *GRU* tidak memerlukan data untuk stasioner sehingga lebih fleksibel dalam menjalankan modelnya. Model *VAR* memerlukan data untuk stasioner, jadi jika data tidak stasioner memerlukan beberapa tahapan untuk mengubah data menjadi stasioner. Di dalam model *GRU* juga memiliki beberapa model untuk membuat data lebih bersih, seperti *financial outlier removal*, *zscore outlier removal*, dan *mahalanobis outlier removal*, sehingga dapat menyingkirkan data yang dapat merusak hasil prediksi. Secara model, *Multivariate GRU* lebih fleksibel dan banyak fitur tambahan untuk membuat hasil prediksi lebih akurat, sedangkan *Multivariate VAR* lebih sederhana.

## DAFTAR PUSTAKA

- Gao, Y., Wang, R., & Zhou, E. (2021). Stock Prediction Based on Optimized LSTM and GRU Models. *Scientific Programming*, 2021. <https://doi.org/10.1155/2021/4055281>
- Robbani, M., Agustiani, F., & Herrhyanto, N. (n.d.). REGRESI LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR (LASSO) PADA KASUS INFLASI DI INDONESIA TAHUN 2014-2017.
- Febrianti, D. R., Tiro, M. A., & Sudarmin, S. (2021). Metode Vector Autoregressive (VAR) dalam Menganalisis Pengaruh Kurs Mata Uang Terhadap Ekspor Dan Impor Di Indonesia. *VARIANSI: Journal of Statistics and Its Application on Teaching and Research*, 3(1), 23. <https://doi.org/10.35580/variansiunm14645>
- Cavalcante, L., Bessa, R. J., Reis, M., & Browell, J. (n.d.). LASSO Vector Autoregression Structures for Very Short-term Wind Power Forecasting. <https://doi.org/10.1002/we>
- Mateus, B. C., Mendes, M., Farinha, J. T., Assis, R., & Cardoso, A. M. (2021). Comparing LSTM and GRU models to predict the condition of a pulp paper press. *Energies*, 14(21). <https://doi.org/10.3390/en14216958>
- Che, Z., Purushotham, S., Cho, K., Sontag, D., & Liu, Y. (2018). Recurrent Neural Networks for Multivariate Time Series with Missing Values. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-24271-9>
- Efriyenty, D. (n.d.). PENGARUH INFLASI DAN KURS TERHADAP HARGA SAHAM DI INDUSTRI DASAR DAN KIMIA. In *Going Concern: Jurnal Riset Akuntansi* (Vol. 15, Issue 4). <http://www.bi.go.id>
- Maronrong, R., Nugrhoho, K. PENGARUH INFLASI, SUKU BUNGA DAN NILAI TUKAR TERHADAP HARGA SAHAM STUDI KASUS PADA PERUSAHAAN MANUFAKTUR OTOMOTIF TERDAFTAR DI BURSA EFEK INDONESIA TAHUN 2012 – 2017. *Jurnal STEI Ekonomi* (Vol 26, No. 2). <https://ejournal.stei.ac.id/index.php/JEMI/article/view/38>
- Rahayu, P. I., Famalika, A., & Sihombing, R. (2021). PENERAPAN MODEL VECTOR AUTOREGRESSIVE (VAR (2)) PADA DATA INFLASI DI PROVINSI JAWA TIMUR DAN BALI (Vol. 1, Issue 1). <http://bayesian.lppmbinabangsa.id/index.php/home>
- Saputra, A. (2020). VECTOR AUTOREGRESSIVE INTEGRATED (VARI) MENGGUNAKAN SOFTWARE R VECTOR AUTOREGRESSIVE INTEGRATED (VARI) USING SOFTWARE R. In *Jurnal Baut dan Manufaktur* (Vol. 02, Issue 01). [www.BPS.GO.ID](http://www.BPS.GO.ID)
- USMAN, H. H., DJAKARIA, I., & PAYU, M. R. F. (2020). PENDEKATAN MODEL VECTOR AUTOREGRESSIVE (VAR) UNTUK MERAMALKAN FAKTOR-FAKTOR YANG MEMPENGARUHI INFLASI DI PROVINSI GORONTALO. *Jambura Journal of Probability and Statistics*, 1(1), 13–23. <https://doi.org/10.34312/jjps.v1i1.5408>

- Putra Robertus, B.R., Hendry, 2022, Multivariate Time Series Forecasting pada Penjualan Barang Retail dengan Recurrent Neural Network, Jurnal Invotek Polbeng 7(1): 71-82, <https://media.neliti.com/media/publications/466725-multivariate-time-series-forecasting-pad-b2582a8c.pdf> Diakses pada tanggal 28 April 2023.
- Chatfield, C., 2011, Exploratory Data Analysis, European Journal of Operational Research 23(1): 5-13, <https://www.sciencedirect.com/science/article/abs/pii/0377221786902092> Diakses pada tanggal 28 April 2023.
- Wutsqa, D. U., 2008, THE VAR-NN MODEL FOR MULTIVARIATE TIME SERIES FORECASTING, Jurnal Mat Stat 8(1): 35-43, [http://research-dashboard.binus.ac.id/uploads/paper/document/publication/Journal/MatsTat/Vol.%2008%20No.%201%20Januari%202008/04\\_Dhoriva.pdf](http://research-dashboard.binus.ac.id/uploads/paper/document/publication/Journal/MatsTat/Vol.%2008%20No.%201%20Januari%202008/04_Dhoriva.pdf) Diakses pada tanggal 28 April 2023.