

Forecasting Saham Dengan Metode CRISP-DM Dan Algoritma Vector Autoregression Studi Kasus Saham ICBP.JK

*Stock Forecasting Using CRISP-DM Method and Vector Autoregression Algorithm Stock Case
Study ICBP.JK*

Muhamad Nurul Khaqim^{a,1}, Agung Triayudi^{*b,2}, Ira Diana Sholihati^{c,3}

^aMahasiswa Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika

^{b,c}Dosen Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika

^{a,b,c}Universitas Nasional, Jl. Sawo Manila, Pejaten Barat, Pasar Minggu Jakarta 12520 Indonesia

¹muhamadnurulkhaqim2018@student.unas.ac.id, ^{*2}agungtriayudi@civitas.unas.ac.id,

³iradiana2803@gmail.com

ABSTRAK

Memperkirakan data saham telah menjadi salah satu isu penting terutama bagi orang yang memilih investasi saham sebagai salah satu sarana manajemen keuangan. Banyak penelitian yang sudah dilakukan untuk membantu memperkirakan data saham dimasa depan menggunakan machine learning. Namun kebanyakan model machine learning yang digunakan kompleks dan hanya memprediksi satu variabel deret waktu. Pada penelitian ini berfokus pada membuat model machine learning menggunakan algoritma VAR untuk memprediksi beberapa variabel sekaligus dengan 1 model dan memberikan rekomendasi dan menggunakan kerangka kerja CRISP-DM dalam mengadakan penelitiannya. Variabel-variabel tersebut adalah harga pembukaan, tertinggi, terendah dan penutupan. Pada akhir penelitian diketahui model VAR mampu menghasilkan model yang mampu memprediksi 3 variabel sekaligus yaitu harga tertinggi, terendah dan penutupan dengan skor R2 masing-masing yaitu 0.60, 0.51, 0.54 dan menggunakan optimal lag 273 namun untuk variabel harga pembukaan dibuatkan model terpisah dengan beda lag yaitu 2 lag dan skor R2 0.63..

Kata Kunci : CRISP-DM, Forecasting, Machine Learning, Saham, Vector Autoregression

ABSTRACT

Estimating stock data has become an important issue, especially for people who choose stock investment as a means of financial management. Many studies have been done to help predict future stock data using machine learning. However, most of the machine learning models used are complex and predict only one time series variable. This research focuses on making machine learning models using the VAR algorithm to predict several variables at once with 1 model and provide recommendations and use the CRISP-DM framework in conducting research. These variables are the opening, highest, lowest and closing prices. At the end of the study, it is known that the VAR model is able to produce a model that is able to predict 3 variables at once, namely the highest, lowest and closing prices with R2 scores of 0.60, 0.51, 0.54 respectively and using an optimal lag of 273 but for the opening price variable a separate model is made with a different lag i.e. 2 lags and an R2 score of 0.63.

Keywords : CRISP-DM, Forecasting, Machine Learning, Stock, Vector Autoregression

Info Artikel :

Disubmit: 24 December 2021

Direview: 04 June 2022

Diterima : 10 June 2022

Copyright © 2022 – CSRID Journal. All rights reserved.

1. PENDAHULUAN

Investasi adalah komitmen individu atau kelompok untuk memberikan sebagian uang yang dimilikinya pada suatu aset dengan tujuan untuk mendapatkan keuntungan di masa depan [1]. Return atau keuntungan investasi saham dihasilkan dari penjualan saham tersebut. Menjual dan membeli saham membutuhkan literasi keuangan yang baik karena literasi yang baik berpengaruh pada nilai return atau keuntungan yang didapat di kemudian hari [2]. Ada beberapa cara untuk mendapatkan

keuntungan dari investasi saham, salah satunya yaitu dengan *forecasting* dengan bantuan *machine learning*.

Sudah banyak penelitian *forecasting* menggunakan *machine learning* mulai dari ARIMA, XGBoost, LSTM, SVR, kombinasi beberapa algoritma dan masih banyak lagi. Permasalahannya kebanyakan model yang digunakan kompleks dan hanya memprediksi satu variabel saja. Kemudian peneliti ingin membangun model *forecasting* yang dapat memprediksi lebih dari satu variabel sekaligus dengan model yang lebih sederhana.

Tujuan penelitian ini adalah untuk menganalisa apakah variabel harga pembukaan, tertinggi, terendah dan penutupan saham dapat diprediksi sekaligus berdasarkan data masa lalu dari masing-masing variabel. Kemudian membangun model *forecasting* menggunakan teknologi *machine learning* algoritma *Vector Autoregression* (VAR) dan metode penelitian CRISP-DM. Sehingga dapat memberikan rekomendasi harga menggunakan algoritma VAR. Adapun manfaat dari penelitian ini adalah untuk mengetahui apakah algoritma VAR dapat menghasilkan model dengan akurasi yang baik khususnya menggunakan variabel-variabel yang ada sehingga dapat digunakan untuk melakukan *forecasting* data saham.

Berdasarkan latar belakang masalah diatas, maka dapat dirumuskan beberapa masalah dalam penelitian ini adalah bagaimana penerapan metode CRISP-DM dalam melakukan penelitian ini dan bagaimana membuat model machine learning dengan algoritma VAR untuk memprediksi lebih dari satu variabel deret waktu.

Batasan masalah dari penelitian ini yaitu hanya menggunakan data harga pembukaan, tertinggi, terendah dan penutupan saham. Data yang digunakan merupakan data harian dari rentang tanggal 29 Juli 2019 - 26 November 2021. Batasan selanjutnya yaitu hasil model yang dibuat dalam penelitian ini hanya untuk memprediksi harga di masa depan sejauh 80 hari kerja sejak tanggal 26 November 2021 dan menampilkan hasilnya pada halaman *web* sederhana dengan memasukkan tanggal yang akan diprediksi.

Berinvestasi dalam saham merupakan sarana penting untuk manajemen keuangan masyarakat modern, dan bagaimana memperkirakan data saham telah menjadi salah satu isu penting [3]. Saham sendiri merupakan bagian kepemilikan suatu perusahaan dimana setiap satu lembarnya memberikan hak suara kepada pemiliknya [4]. ICBP.JK adalah kode saham perusahaan PT Indofood CBP Sukses Makmur Tbk yang merupakan produsen makanan ringan, bumbu makanan, susu, aneka minuman, biskuit dan makanan khusus tertentu.

Metode yang digunakan pada penelitian ini adalah CRISP-DM. *Cross Industry Standart Process for Data Mining* (CRISP-DM) secara luas diterima sebagai kerangka kerja untuk *data mining* industri dan *data analytic* untuk meneliti atau mencari pengetahuan berbasis data [5]. *Multivariate Time Series* (MTS) adalah serangkaian nilai dengan indikator statistik yang sama yang disusun dalam urutan kronologis kejadian. Fungsi dari data masa lalu dapat digunakan untuk memprediksi nilai masa depan variabel jika perubahan variabel tersebut terkandung dalam informasi dari data masa lalu [6]. Algoritma yang digunakan yaitu *Vector Autoregression* (VAR). *Vector Autoregression* (VAR) adalah perhitungan prediksi yang digunakan ketika dua deret waktu atau lebih berdampak satu dengan yang lain. [7].

Pada penelitian sebelumnya pada jurnal yang berjudul *Stock-Price Forecasting Based on XGBoost and LSTM*, peneliti membandingkan model algoritma *Autoregressive Integrated Moving Average* (ARIMA) atau disebut juga kombinasi antara algoritma *Autoregressive* dengan *Moving Average* dan kombinasi XGBoost dengan LSTM untuk memprediksi harga saham. Dataset yang digunakan diambil dari forex market dari 01/01/2008 sampai 19/03/2018 dengan total data 709.314. Kemudian peneliti tersebut menggunakan harga sebagai target. Hasilnya, model ARIMA skor *Mean Absolute Error* (MAE) sebesar $4,149 \times 10^{-4}$ dan *Root Mean Squared Error* (RMSE) 7.819×10^{-4} ,

sedangkan kombinasi XGBoost dengan LSTM menghasilkan model dengan skoe MAE sebesar $3,825 \times 10^{-4}$ dan RMSE $5,887 \times 10^{-4}$ namun tidak dicantumkan skor R2 pada penelitian ini [8].

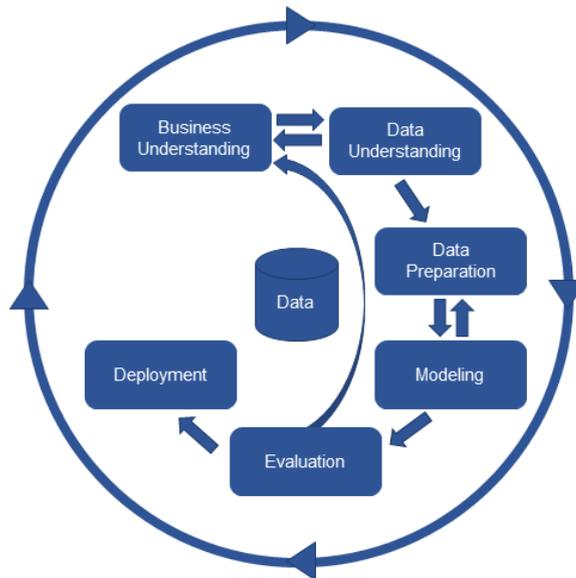
Pada penelitian lain pada jurnal yang berjudul *Stock Closing Price Prediction using Machine Learning Techniques* para peneliti membandingkan dua algoritma yaitu *Artificial Neural Network* (ANN), dan *Random Forest* (RF) untuk memprediksi harga penutupan saham. Variabel yang digunakan diantaranya harga tertinggi dikurangi harga terendah (H-L), Harga saham penutupan dikurangi harga pembukaan (O-C), *moving average* 7 hari ke belakang, 14 hari ke belakang, 21 hari ke belakang, dan standar deviasi 7 hari ke belakang. Dataset yang digunakan merupakan data harian dalam rentang waktu 10 tahun antara 4 Mei 2009 hingga 4 Mei 2019 dari beberapa perusahaan diantaranya Nike, JP Morgan and Co., Johnson & Johnson and Pfizer Inc. companies. Hasilnya, algoritma ANN memiliki skor RMSE terkecil yaitu antara 0.42 hingga 3.30 sedangkan algoritma RF memiliki skor RMSE antara 0.43 hingga 3.40 [9].

Kemudian ada penelitian serupa mengenai prediksi data saham dalam jurnal yang berjudul *Prediksi Harga Saham Menggunakan Metode Brown's Weighted Exponential Moving Average dengan Optimasi Levenberg-Marquardt*. Penelitian ini menggunakan algoritma *Brown's Weighted Exponential Moving Average* (B-WEMA) kemudian dikombinasikan dengan algoritma optimasi *Levenberg-Marquardt* dan membandingkannya dengan tanpa optimasi. Metode penelitiannya yaitu dimulai dari melakukan studi literature, kemudian mengumpulkan data, merancang sistem, mengimplementasikan sistem, melakukan analisis hasil dan menarik kesimpulan. Hasil penelitian mengemukakan bahwa algoritma B-WEMA dengan optimasi terbukti lebih baik dibandingkan tanpa optimasi yaitu dengan *error* hampir mencapai 0% dan skor MSE sebesar 719,56 [10].

Berdasarkan beberapa penelitian sebelumnya diatas, hanya memprediksi satu variabel walaupun ada yang menggunakan variabel lain. Pemilihan kombinasi metode atau kerangka kerja penelitian CRISP-DM dikarenakan alurnya mudah dipahami dan diimplementasikan sedangkan pemilihan algoritma VAR karena diakhir penelitian diharapkan dengan algoritma tersebut dapat menghasilkan model yang mampu memprediksi beberapa variabel deret waktu sekaligus. Kemudian dapat memberikan rekomendasi untuk menjual atau membeli saham ICBP.JK berdasarkan hasil prediksi tersebut.

2. METODE PENELITIAN

Penelitian ini menggunakan data sekunder yaitu kumpulan saham yang terdaftar di *Indonesia Stock Exchange* (IDX) dari berbagai emiten [11], namun yang digunakan hanya data saham ICBP.JK. Dataset saham ini didapat dari situs kaggle.com yang mana metode pengumpulannya dengan cara *scrapping* data dari situs resmi IDX idx.co.id. Penelitian ini menggunakan pendekatan metode CRISP-DM. Metode ini terdiri dari 6 fase yaitu *business understanding*, *understanding data*, *data preparation*, *modeling*, *evaluation* dan *deployment*. Gambaran dan deskripsi dari siklus fase tersebut dapat ditunjukkan pada Gambar 1 sebagai berikut:



Gambar 1. Siklus CRISP-DM.

Pada siklus diatas diawali dengan *business understanding* kemudian diikuti dengan *data understanding*. Terdapat dua panah bolak-balik antara dua fase tersebut yang mana menandakan bahwa *data understanding* harus berdasarkan *business understanding*. Fase selanjutnya yaitu *data preparation* lalu *modeling*. Panah bolak-balik di antara dua fase tersebut menandakan ketika *modeling* seringkali harus kembali ke fase *data preparation* untuk mendapatkan hasil yang lebih optimal. Setelah itu, melakukan *evaluation* untuk mengevaluasi hasil *modeling* dan mencocokkannya dengan *business understanding* dan fase terakhir yaitu *deployment* yaitu membuat model dapat digunakan atau didistribusikan [12].

A. *Bussiness Understanding*

Pada penelitian ini, pendekatan *machine learning* bertujuan untuk membangun model untuk *forecasting* data saham untuk memberikan rekomendasi harga untuk tanggal tertentu berdasarkan data di masa lalu. Sehingga diharapkan dapat membantu dalam pengambilan keputusan dalam bertransaksi saham khususnya saham ICBP.JK.

B. *Data Understanding*

Kumpulan data saham ini didapat dari situs kaggle.com yang mana metode pengumpulannya yaitu dengan *scrapping* data dari situs resmi IDX idx.co.id. Dataset tersebut terdiri dari 565 baris yang berarti terdapat 565 *lag*. Berikut keterangan dari atribut yang ada pada dataset yang digunakan:

Tabel 1. Keterangan Nama Kolom

Nama Kolom	Keterangan
Date	Tanggal jalannya perdagangan
Open Price	Harga pembukaan pada hari tersebut
High	Harga tertinggi pada hari tersebut
Low	Harga terendah pada hari tersebut
Close	Harga penutupan pada hari tersebut

Tabel diatas merupakan kamus data dari dataset yang digunakan dalam penelitian ini.

Tabel 2. Data yang Digunakan

Date	Open Price	High	Low	Close
2019-07-29	10600	10600	10150	10550
2019-07-30	10350	10625	10325	10625

2019-07-31	10700	10725	10600	10700
2019-08-01	10700	10825	10575	10700
2019-08-02	10600	10875	10600	10850
...
2021-11-22	9025	9100	9000	9075
2021-11-23	9050	9050	8900	8950
2021-11-24	8975	9000	8950	9000
2021-11-25	9000	9025	8950	8950
2021-11-26	8950	8975	8850	8875

Tabel diatas merupakan dataset yang digunakan dalam penelitian ini berisi 575 data harian saham ICBP.JK.

C. Data Preparation

Pada tahap ini, data akan dipersiapkan untuk diproses sebagai data latih maupun data uji. Berikut langkah-langkah yang dilakukan pada tahap ini:

1) Membuat Data Latih dan Uji

Langkah pertama pada tahap data preparation yaitu membuat data latih dan data uji dari dataset dengan dengan jumlah data uji sebanyak lima lag atau waktu sedangkan sisanya dijadikan data latih.

2) Uji Stasionaritas

Pada dasarnya VAR menggunakan regresi dalam implementasinya. Dalam regresi jika data yang digunakan non-stationary maka akan menghasilkan apa yang disebut dengan spurious regression [13]. Uji stasionaritas pada penelitian ini dilakukan menggunakan Augmented-Dickey Fuller (ADF).

3) Differencing

Jika terdapat data time series yang non-stationary maka dilakukan differencing menggunakan first difference pada data latih kemudian melakukan uji stasionaritas kembali hingga semua time series stasioner.

4) Uji Granger-Casuality

Uji Granger menguji antara dua variabel yang mana variabel pertama dikatakan granger-causes variabel kedua jika nilai variabel kedua tersebut dapat diprediksi dengan bantuan informasi yang terkandung dalam data masa lalu variabel pertama [14]. Uji granger-casuality dilakukan pada data latih.

D. Modeling

Pada penelitian ini yang dilakukan adalah membuat estimasi VAR. VAR adalah algoritma untuk menghitung prediksi yang digunakan ketika setidaknya dua deret waktu berdampak satu sama lain, yaitu, hubungan dua arah antara deret waktu yang disertakan. Algoritma VAR pada dasarnya memiliki bentuk formula yang sama dari *Autoregressive* AR(p) seperti berikut:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + \varepsilon_t \quad (1)$$

Dimana α adalah intercept atau konstanta dan β_1, β_2 hingga β_p adalah koefisien lag atau selang waktu sampai order p sedangkan Y adalah variabel prediktor dan t adalah waktu [7].

E. Evaluation

Pada tahap evaluasi, model dievaluasi menggunakan metrik Mean Absolute Error (MAE) yaitu menghitung rata-rata kesalahan (selisih nilai aktual dengan hasil prediksi) absolut, Root Mean Square Error (RMSE) menghitung total dari selisih hasil prediksi dengan data actual yang dikuadratkan kemudian dihitung rata-ratanya dan diakarkan dan yang terakhir yaitu R2 untuk melihat performa model yang dibangun [15].

$$MAE(y-\hat{y}) = \frac{1}{N} \sum_{j=1}^N |y - \hat{y}| \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{j=1}^N (y-\hat{y})^2}{N}} \quad (3)$$

$$R2 = 1 - \frac{\sum_{j=1}^N (y_j - \hat{y})^2}{\sum_{j=1}^N (y_j - \bar{y})^2} \quad (4)$$

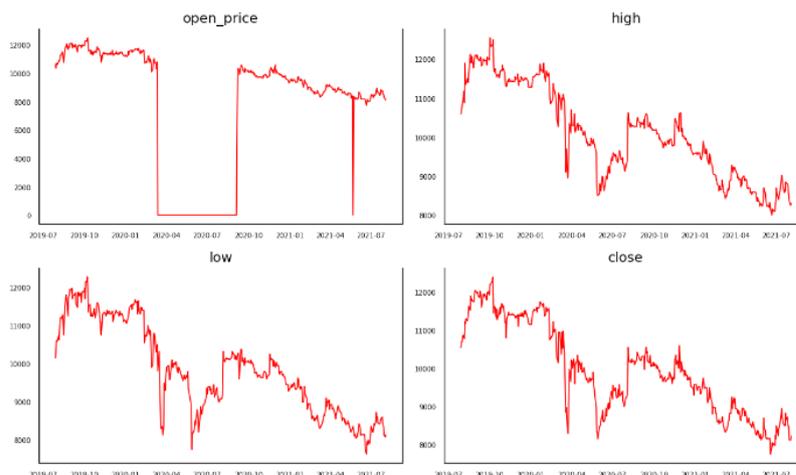
F. Deployment

Setelah melakukan semua tahapan diatas, tahap selanjutnya yaitu membuat model supaya lebih mudah diakses. Untuk itu dibuat halaman *web* sederhana dengan parameter tanggal kemudian menampilkan grafik dan tabel hasil prediksi yang berisi rekomendasi sehingga dapat membantu pengambilan keputusan dalam bertransaksi.

3. HASIL DAN PEMBAHASAN

A. Uji Stasioneritas

Data dibagi menjadi data latih dan data uji dengan proporsi 495:80 atau 495 *lag* untuk data latih dan 80 *lag* untuk data uji. kemudian data latih terlebih dahulu dilakukan uji ADF untuk mengetahui apakah semua variabel stasioner atau non-stasioner pada level dengan signifikansi 0.05 atau 5%.

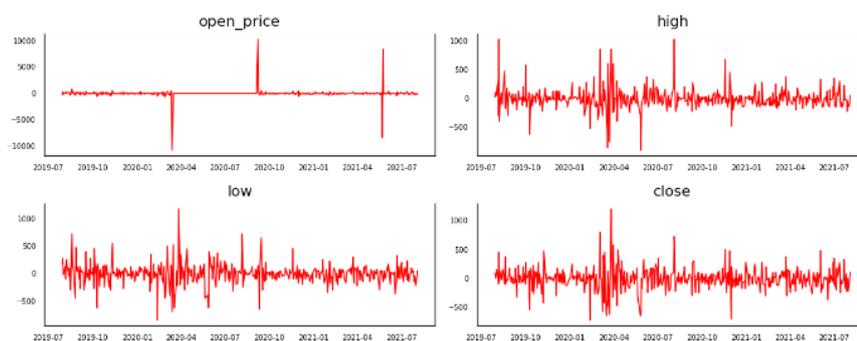


Gambar 2. Plot data latih pada level

Tabel 3. Hasil pengujian ADF data latih pada level

Nama Variabel	P-Value	Keterangan
Open Price	0.36	Non-stasioner
High	0.8	Non-stasioner
Low	0.61	Non-stasioner
Close	0.73	Non-stasioner

Pada gambar 2 terlihat bahwa data latih memiliki kecenderungan trend yang menurun dan tidak konsisten berada disekitar rata-rata. Berdasarkan hasil uji stasioneritas pertama (tabel 3) didapati bahwa hanya terdapat beberapa variabel yang belum stasioner. Oleh karena itu perlu dilakukan *differencing* dengan *first difference* yaitu menguji stasioneritasnya kembali. Hasilnya sebagai berikut:



Gambar 3. Plot data latih pada *first difference*

Tabel 4. Hasil pengujian ADF data latih pada *first difference*

Nama Variabel	P-Value	Keterangan
Open Price	0.0	Stasioner
High	0.0	Stasioner
Low	0.0	Stasioner
Close	0.0	Stasioner

Pada gambar 3, fluktuasi data latih pada *first difference* terlihat tidak memiliki kecenderungan terhadap trend tertentu dan berada disekitar nilai rata-rata yang konstan atau dengan kata lain sudah stasioner. Ini juga dibuktikan dengan *P-Value* hasil uji stasioneritas pada tabel 4.

B. Uji Granger-Casuality

Setelah semua data stasioner, langkah selanjutnya yaitu melakukan uji *granger-casuality* untuk melihat hubungan antar data tiap variabel. Fokus uji *granger-casuality* adalah untuk melihat apakah data masa lalu setiap variabel tersebut saling memberikan pengaruh untuk memprediksi nilai variabel yang lain.

Tabel 5. Hasil Uji *Granger-Casuality*

	Open Price	High	Low	Close
Open price	1.00	0.45	0.23	0.00
High	0.00	1.00	0.00	0.00
Low	0.00	0.00	1.00	0.00
Close	0.00	0.02	0.01	1.00

Tabel 5 menampilkan hasil uji *granger-casuality* dari semua kemungkinan kombinasi dari data deret waktu setiap variabel terhadap variabel yang lain dengan H_0 yang berarti variabel prediktor tidak mempunyai informasi yang dapat membantu memprediksi respons. Variabel pada baris merupakan variabel respons sedangkan yang berada pada kolom adalah prediktor dan nilai dalam tabel adalah *P-Values*. Hampir semua kombinasi variabel diatas menolak H_0 yang artinya data di masa lalu masing-masing memiliki pengaruh antar variabel untuk memprediksi nilai variabel yang lain. Terkecuali pada variabel respons *open price* dengan prediktor *low* dan *high* yang artinya tidak cukup bukti bahwa kedua variabel tersebut memberi pengaruh untuk memprediksi variabel *open price*.

C. Pemilihan Model

Pemilihan model, yang pertama kali dilakukan adalah mencari optimal *lag* menggunakan beberapa *information criterion* diantaranya *Akaike Information Criteria (AIC)*, *Bayesian Information Criteria (BIC)*, *Final Prediction Error (FPE)* dan *Hannan-Quinn information criterion (HQIC)*.

Namun pada prosesnya diketahui bahwa optimal lag yang didapatkan tidak konsisten jika diubah maksimum lag yang diberikan. Oleh karena itu pencarian optimal lag dilakukan dengan iteratif dengan maksimum lag antara 1 hingga 400 kemudian mengumpulkan semua optimal lag yang ada. Lalu mengestimasi untuk membuat 4 model dengan mengutamakan RMSE terendah untuk setiap variabel. Hasil model yang ditemukan dengan RMSE terendah untuk masing-masing variabel sebagai berikut:

Tabel 6. Komparasi RMSE untuk semua model

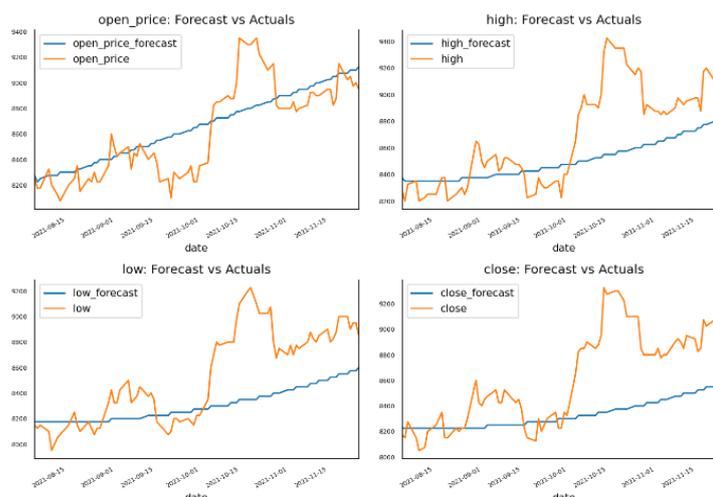
Nama Variabel	RMSE	Optimal Lag
Open Price	217.16	2
High	228.29	273
Low	247.71	273
Close	236.78	291

Berdasarkan hasil optimal lag setiap model diatas, perlu dilakukan pengecekan adakah autokorelasi pada model tersebut menggunakan *Durbin-Watson*. Uji *Durbin-Watson* (DW) umumnya digunakan untuk mendeteksi autokorelasi dalam deret waktu. Statistik DW dapat berkisar antara nol dan empat, dengan nilai mendekati dua menunjukkan tidak ada autokorelasi [16] Untuk mempermudah memahami tabel berikut, untuk model yang mengutamakan variabel *Open Price* akan disebut model OP, kemudian model HL untuk variabel *High* dan *Low* dan model C untuk variabel *Close*.

Tabel 7. Pengecekan autokorelasi menggunakan DW

Variabel	Open Price	High	Low	Close
Model				
OP	2.0	2.04	2.07	1.99
HL	1.97	2.22	1.95	2.03
C	1.7	2.05	1.88	1.87

Terlihat bahwa semua variabel pada mode-model tersebut independen atau tidak memiliki autokorelasi. Untuk memilih model yang tepat peneliti melakukan pengujian model diatas terhadap data uji lalu melihat plot data aktual vs hasil prediksi dan membandingkan skor R2 nya pada setiap variabel.

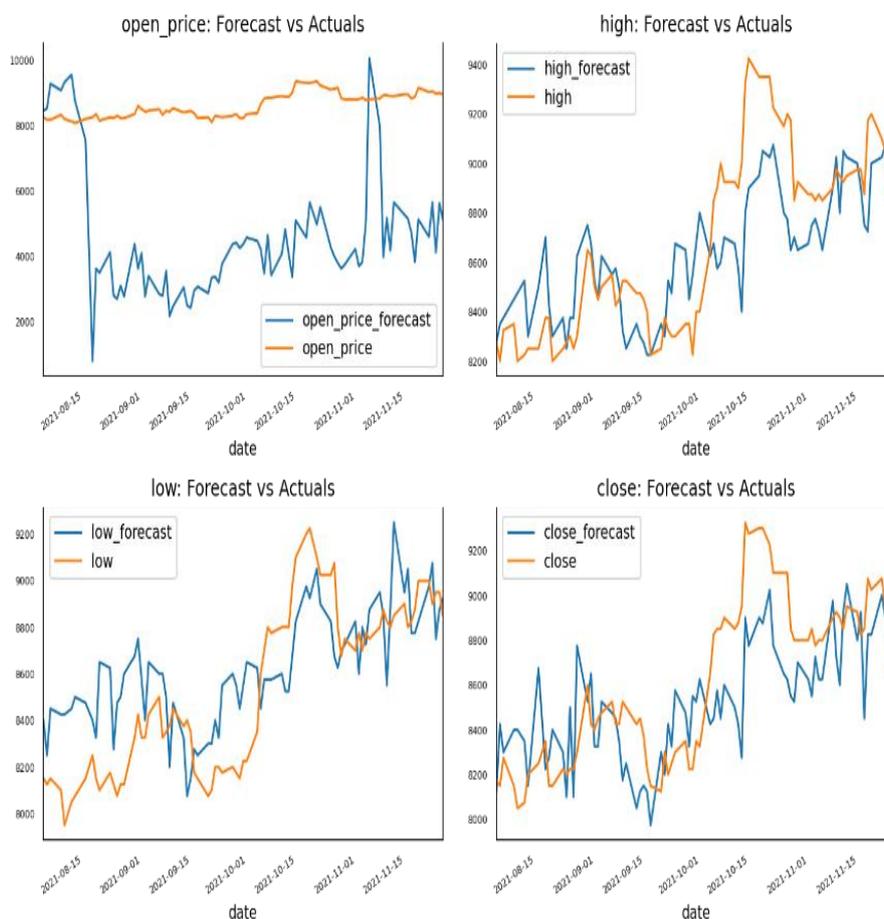


Gambar 4. Plot perbandingan antara data aktual dengan hasil prediksi model OP

Tabel 8. Skor hasil pengujian model OP

Variabel	RMSE	MAE	R2
Open Price	217.16	169.38	0.63
High	313.95	238.75	0.24
Low	339.47	265.94	0.07
Close	396.16	308.44	-0.22

Garis biru merupakan hasil prediksi sedangkan garis orange merupakan data aktual. Pada gambar 4, plot model OP menunjukkan hasil prediksi variable *Open Price* terlihat mengikuti trend data aktual walaupun tidak mengikuti polanya sedangkan pada variabel lainnya tidak terlalu mengikuti trend maupun pola data aktual. Lalu pada tabel 8 skor R2 untuk variabel *Open Price* cukup besar yaitu 63% namun tidak pada variabel yang lain terutama variabel *Close* dimana Skor R2 nya negatif yang menunjukkan bahwa model belum dapat dikatakan sebagai model terutama untuk memprediksi variabel *Close*.

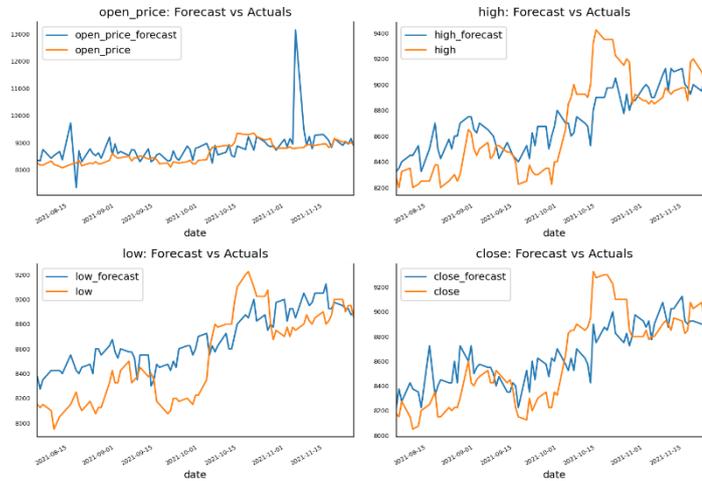


Gambar 5. Plot perbandingan antara data aktual dengan hasil prediksi model HL

Tabel 9. Skor hasil pengujian model HL

Variabel	RMSE	MAE	R2
Open Price	4171.07	3940.63	-134.21
High	228.29	184.69	0.60
Low	247.71	212.5	0.51
Close	242.79	199.06	0.54

Pada gambar 5, pola hasil prediksi model HL untuk variabel *Open Price* tidak mengikuti data aktual sedangkan untuk variabel *High*, *Low* dan *Close* terlihat mengikuti pola data aktual. Lalu hasil pengujian model HL (tabel 9) terlihat baik skor RMSE, MAE dan R2 lebih baik dari model OP kecuali untuk variabel *Open Price*.



Gambar 6. Plot perbandingan antara data aktual dengan hasil prediksi model C

Tabel 10. Skor hasil pengujian model C

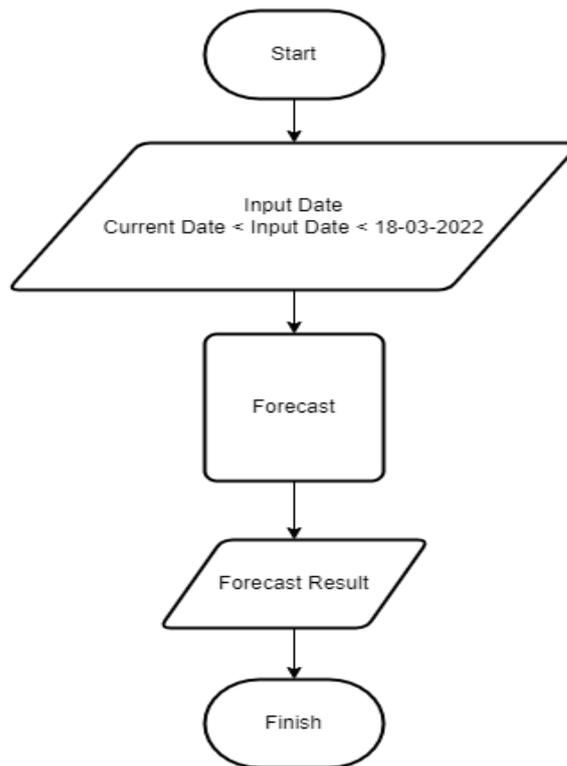
Variabel	RMSE	MAE	R2
Open Price	615.30	338.125	-1.94
High	235.85	195.63	0.57
Low	261.93	225.63	0.45
Close	236.78	193.13	0.56

Hasil pengujian model C pada tabel 10, memperlihatkan bahwa model tersebut lebih baik dari model OP namun tidak lebih baik dari model HL kecuali pada variabel *Close* dengan selisih yang tidak terlalu signifikan. Pola hasil prediksi model C (gambar 6) semuanya hampir mendekati atau sama dengan data aktual tetapi tidak lebih baik dari model HL kecuali untuk prediksi variabel *close*.

Berdasarkan hasil pengujian pada tabel 8 - 10, Peneliti memutuskan untuk menggunakan model HL untuk memprediksi variabel *High*, *Low*, dan *Close* dan model OP untuk memprediksi variabel *Open Price*.

D. Implementasi

Pada bagian ini akan menjelaskan tentang bagaimana implementasi dari model yang sudah dibuat. Implementasi ini merupakan bagian dari siklus CRISP-DM *deployment* yang akan menggunakan halaman *web* sederhana untuk menampilkan hasil prediksi. Berikut *flowchart* tentang alur kerja halaman *web* tersebut:



Gambar 7. Flowchart alur kerja halaman web

Gambar 7 menunjukkan bagaimana aplikasi *web* akan bekerja. Pada input tanggal, batas maksimal prediksi adalah 18 Maret 2022 lalu hanya bisa memilih tanggal lebih dari hari saat membuka halaman tersebut.



Gambar 8. Tampilan halaman web hasil prediksi

Pada gambar 8 terdapat *chart* yang menampilkan hasil prediksi semua variabel dari tanggal keesokan hari (hari saat mengakses *web* tersebut) hingga target tanggal yang dipilih. Kemudian dibawahnya terdapat tabel yang hanya menampilkan angka semua variabel beserta kolom tambahan rekomendasi yang berisi rekomendasi jual dan beli ± skor MAE masing-masing variabel.

4. KESIMPULAN

Diakhir penelitian, dapat diketahui bahwa metode kerangka kerja CRISP-DM untuk melakukan penelitian tentang *Data Mining* terbukti cukup rapih dan mudah dipahami siklusnya. Berdasarkan hasil pengujian dan evaluasi menggunakan skor R2, MAE dan RMSE pada model yang berhasil dibuat, dapat disimpulkan bahwa VAR dapat digunakan untuk melakukan prediksi harga saham tertinggi, terendah dan penutupan sekaligus dan memiliki akurasi yang cukup baik walaupun variabel harga pembukaan harus dibuatkan model terpisah namun tetap memiliki akurasi yang cukup baik. Hasil model yang dibuat juga dapat diimplementasikan dalam bentuk halaman *web* sederhana yang memperlihatkan *chart* dan tabel hasil prediksi serta rekomendasi (dengan \pm MAE masing-masing variabel) dengan inputan tanggal yang ingin diprediksi.

Pada penelitian sebelumnya [8,9,10], memang memiliki skor RMSE yang lebih rendah dibandingkan dengan hasil model pada penelitian ini walaupun tidak tercantum skor R2-nya. Namun model pada penelitian ini mampu memprediksi lebih dari satu variabel deret waktu sekaligus.

Berdasarkan kesimpulan hasil penelitian ini, diketahui bahwa untuk memprediksi harga pembukaan, tertinggi, terendah dan penutupan saham menggunakan algoritma VAR membutuhkan dua model yaitu satu model untuk memprediksi harga pembukaan saham dan yang lainnya untuk memprediksi harga terendah, tertinggi dan penutupan saham. Adapun saran untuk penelitian lebih lanjut yaitu dapat dengan menambahkan beberapa variabel lainnya yang memberikan pengaruh terhadap fluktuasi variabel-variabel yang diprediksi sehingga mampu menghasilkan model yang lebih baik.

UCAPAN TERIMA KASIH

Terima kasih kepada **Dr. Agung Triayudi, M.Kom** sebagai dosen pembimbing satu sekaligus penulis korespondensi dan **Ira Diana Solihati S.Si.,MMSI** sebagai dosen pembimbing dua yang telah membimbing penelitian tugas akhir ini hingga selesainya dan terbit jurnal ini di *publisher*.

REFERENSI

- [1] Faidah, F, Pengaruh Literasi Keuangan dan Faktor Demografi Terhadap Minat Investasi Mahasiswa, *Journal of Applied Business and Economic.*, vol. 5, no. 3, pp. 251-263, March 2019.
- [2] Fauziah, A., Purwanto, B., & Ermawati, W. J., Myopic loss aversion and financial literacy in Indonesian beginner investor Decision Making, *Jurnal Manajemen dan Organisasi (JMO).*, vol. 11, no. 1, pp. 1-7, April 2020.
- [3] Ko, C. & Chang, H., LSTM-based sentiment analysis for stock price forecast, *PeerJ Computer Science.*, vol. 7, no. 408, pp. 1-23, March 2021.
- [4] Egam, G. E. Y., Ilat, V., & Pengarepan, S., Pengaruh Return on Asset (ROA), Return on Equity (ROE), Net Profit Margin (NPM), dan Earning Per Share (EPS) terhadap harga saham perusahaan yang tergabung dalam Indeks LQ45 di Bursa Efek Indonesia Periode Tahun 2013-2015, *Jurnal EMBA.*, vol. 5, no. 1, pp. 105-114, March 2017.
- [5] Tripathi, S., Muhr, D., Brunner, M., Jodlbauer, H., Dehmer, M. & Emmert-Streib, F., Ensuring the robustness and reliability of data-driven knowledge discovery models in production and manufacturing, *PeerJ Computer Science.*, vol. 4, no. 576892, pp. 1-20, June 2021.
- [6] Zhang, R., & Jia, H., Production performance forecasting method based on multivariate time series and vector autoregressive machine learning model for water flooding reservoirs, *Petroleum Exploration and Development.*, vol. 48, no. 1, pp. 201-211, February 2021.
- [7] Sujat, R., Chatterjee, J. M., & Hassanien, A. E., A machine learning forecasting model for COVID-19

- pandemic in India, *Stochastic Environmental Research and Risk Assessment.*, vol. 34, pp. 959–972, July 2020.
- [8] Vuong, P. H., Dat, T. T., Mai, T. K., Uyen, P. H. & Bao, P. T. Stock-Price Forecasting Based on XGBoost and LSTM, *Computer Systems Science & Engineering.*, vol. 40, no.1, pp. 238-246, May 2021.
- [9] Vijn, M., Chandola, D., Tikkiwal, V. A., & Kumar, A., Stock Closing Price Prediction using Machine Learning Techniques, *Procedia Computer Science.*, vol. 167, no.2020, pp. 599-606, March 2020.
- [10] Putri, D. I., BudiPrasetijo, A., & Rochim, A. F., Stock Price Prediction Using Brown's Weighted Exponential Moving Average with Levenberg-Marquardt Optimization, *Jurnal Nasional Teknik Elektro dan Teknologi Informasi.*, vol. 10, no.1, pp. 11-18, February 2021.
- [11] Dataset Saham IDX. “Pergerakan Saham di Bursa Efek Indonesia”. Available: <https://www.kaggle.com/tiwill/saham-idx>.
- [12] Mauritsius, T., Braza, A. S., & Fransisca, Bank Marketing Data Mining using CRISP-DM Approach, *International Journal of Advanced Trends in Computer Science and Engineering.*, vol. 8, no.5, pp. 2322-2329, October 2019.
- [13] Hoang, N., & Grieb, T., Hedging positions in US wheat markets: a disaggregated data analysis?, *Studies in Economics and Finance* © Emerald Publishing Limited., vol. 37, no.3, pp. 429–455, January 2020.
- [14] Moaniba, I. M., Su, H., & Lee, P., On the drivers of innovation: Does the co-evolution of technological diversification and international collaboration matter?, *Technological Forecasting & Social Change.*, vol. 148, pp. 1–16, July 2019.
- [15] Majumder, A., Rahman, M., Biswas, A., Zulfiker, M. S., & Basak, S., Stock Market Prediction: A Time Series Analysis, *Smart Systems: Innovations in Computing.*, vol. 235, pp. 389-401, September 2021.
- [16] Turner, S. L., Forbes, A. B., Karahalios, A., Taljaard, M., & McKenzie, J. E., Evaluation of statistical methods used in the analysis of interrupted time series studies: a simulation study, *BMC Medical Research Methodology.*, vol. 21, pp. 1-18, August 2021