

Penerapan CRISP-DM untuk Prediksi Harga Saham NVIDIA Menggunakan Time Series Analysis

Juan Hubert Liem¹ dan Calvin Christian Tjong²

¹Program Studi Sistem Informasi, Universitas Ciputra Surabaya
Jalan CitraLand CBD Boulevard, Surabaya, Indonesia, 60219

Korespondensi: Juan Hubert Liem (jhubertliem@student.ciputra.ac.id)

Received: 24 Juli 2024 – *Revised:* 31 Agustus 2024 - *Accepted:* 05 Sept 2024 - *Published:* 10 Sept 2024

Abstrak. Saham memiliki pasar dengan volatilitas tinggi yang menuntut investor untuk memiliki kemampuan prediksi yang akurat dalam pengambilan keputusan. Penelitian ini mengkaji penerapan metode time series untuk memprediksi harga saham NVIDIA dengan data dari awal 2017 hingga akhir 2023. Metode yang digunakan adalah ARIMA (*AutoRegressive Integrated Moving Average*) dan SARIMA (*Seasonal AutoRegressive Integrated Moving Average*) berdasarkan pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang melibatkan enam fase: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan deployment. Data awal dianalisis untuk mengeksplorasi tren dan pola, kemudian diolah untuk menghilangkan noise dan mengisi data yang hilang. Model ARIMA dan SARIMA diterapkan untuk memprediksi harga saham, namun analisis menunjukkan bahwa saham NVIDIA tidak memiliki pola musiman yang signifikan. Evaluasi menggunakan Mean Absolute Percentage Error (MAPE) menunjukkan bahwa kedua model memiliki MAPE sebesar 39 persen, menandakan tingkat kesalahan prediksi yang sama. Hasil ini mengindikasikan bahwa volatilitas harga saham NVIDIA cukup tinggi dan tidak menunjukkan seasonality. Kedepannya perlu untuk dilakukan penelitian serupa dengan menggunakan metode prediksi selain ARIMA dan SARIMA.

Kata kunci: Saham NVIDIA, ARIMA, SARIMA, Model CRISP-DM

Citation Format: Liem, J.H., & Tjong, C.C. (2024). Penerapan CRISP-DM untuk Prediksi Harga Saham NVIDIA Menggunakan Time Series Analysis. *Prosiding SENAM 2024: Seminar Nasional Sistem Informasi & Informatika Universitas Ma Chung*. 4, 204-213. Malang: Ma Chung Press.

PENDAHULUAN

NVIDIA merupakan salah satu perusahaan teknologi yang memiliki peran besar di pasar saham. NVIDIA terkenal sebagai produsen GPU (*Graphics Processing Units*) terkemuka di dunia. Proyek yang besar di bidang teknologi membuatnya menjadi opsi bagi investor. Tingginya minat investor terhadap saham NVIDIA menjadikan harganya bergerak fluktuatif.

Terdapat beberapa metode *time series analysis* yang cocok dan tepat untuk memprediksi pergerakan harga saham ke depannya diantaranya *Autoregressive Integrated*

Moving Average (ARIMA) dan *Seasonal Autoregressive Integrated Moving Average* (SARIMA). Berdasarkan Montgomery *et al.* (2007). Kedua model ini memiliki kemampuan yang cukup baik dalam menganalisis data yang bersifat deret waktu (*time series*) dan prediksi sehingga hasil dihasilkan bisa menjadi lebih akurat. Prediksi harga saham menggunakan metode ARIMA telah dilakukan (Ningrum, 2022). Saham yang diprediksi adalah JKSE dengan data yang digunakan dari Januari 2017 - Desember 2021. *Mean Absolute Percentage Error* yang dihasilkan dari penelitian tersebut adalah 2 persen.

Berdasarkan penelitian terdahulu tersebut, maka Penulis meyakini masih terdapat peluang untuk meneliti saham dengan entitas yang berbeda dengan menggunakan metode ARIMA dan SARIMA yang mengacu pada pendekatan CRISP-DM (*Cross-Industry Standard Process for Data Mining*). Kedua metode tersebut diharapkan bisa memprediksi pergerakan harga dan tren pasar saham NVIDIA ke depannya. Penulis menggunakan dataset publik yang berjudul Pergerakan Harga Saham NVIDIA dari Tahun 2000 - 2024. Diharapkan penelitian ini bisa meyakinkan investor untuk berinvestasi di perusahaan NVIDIA.

MASALAH

Tantangan utama yang dihadapi oleh para investor adalah volatilitas tinggi di pasar saham, yang dapat mempengaruhi keputusan mereka dalam mengalokasikan dana ke berbagai perusahaan. Kondisi ini menyoroti pentingnya prediksi yang akurat dan tepat dalam membuat keputusan investasi saham. Prediksi yang baik dapat membantu investor mengantisipasi pergerakan harga saham, sehingga mereka dapat mengambil langkah yang lebih bijaksana dan mengurangi risiko kerugian. Selain itu, dengan memiliki prediksi yang handal, investor juga dapat memaksimalkan potensi keuntungan dengan mengidentifikasi waktu yang tepat untuk membeli atau menjual saham.

METODE PELAKSANAAN

Data Penelitian

Data yang digunakan untuk melakukan prediksi atau *forecasting* merupakan sebuah data saham bersumber dari *datasource* kaggle.com. Data ini memuat data saham NVIDIA dari tahun 2000 - 2024. Jumlah data yang disediakan sebanyak 6116 data. Variabel yang digunakan pada dataset tersebut antara lain variabel *date*, variabel *open*, variabel *high*, variabel *low*, variabel *close*, variabel *adj. close* dan variabel *volume*. Dalam penelitian ini

Penulis menggunakan data dari tahun 2017 - akhir tahun 2023 sebagai bahan uji penelitian ini. Data harga saham NVIDIA yang digunakan menggunakan satuan USD. Gambar 1 adalah data yang digunakan.

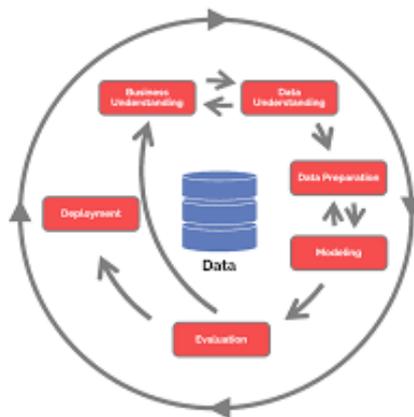
	Date	Open	High	Low	Close	Adj Close	Volume
0	2000-01-03	0.984375	0.992188	0.919271	0.975260	0.894608	30091200
1	2000-01-04	0.958333	0.960938	0.901042	0.949219	0.870721	30048000
2	2000-01-05	0.921875	0.937500	0.904948	0.917969	0.842055	18835200
3	2000-01-06	0.917969	0.917969	0.822917	0.858073	0.787112	12048000
4	2000-01-07	0.854167	0.881510	0.841146	0.872396	0.800251	7118400
...
6111	2024-04-18	849.700012	861.900024	824.020020	846.710022	846.710022	44726000
6112	2024-04-19	831.500000	843.239990	756.059998	762.000000	762.000000	87190500
6113	2024-04-22	781.039978	800.729980	764.000000	795.179993	795.179993	59634100
6114	2024-04-23	807.690002	827.690002	802.640015	824.229980	824.229980	43855900
6115	2024-04-24	839.500000	840.820007	791.830017	796.770020	796.770020	50961600

6116 rows x 7 columns

Gambar 1. Dataset Saham NVIDIA Tahun 2000-2024 (Kaggle.com, 2024)

Cross-Industry Standard Process for Data Mining

Pada penelitian ini Penulis menggunakan pendekatan berupa CRISP-DM. CRISP-DM adalah suatu standarisasi pemrosesan data mining yang sudah dikembangkan dimana data melewati setiap fase terstruktur dan terdefinisi dengan jelas (Hasanah *et al.*, 2021). Model CRISP-DM memiliki enam fase yakni *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modelling*, *Evaluation*, dan *Deployment*.



Gambar 2 . Standard Process CRISP-DM (datascience-pm.com, 2024)

Fase Business Understanding

Investor di pasar saham sering menghadapi tantangan besar dalam mengatasi volatilitas tinggi yang mempengaruhi keputusan investasi mereka. Saham, sebagai instrumen investasi, memiliki siklus bisnis yang bersifat fluktuatif dan pendapatannya

sangat tergantung pada kondisi pasar. Hal ini menekankan pentingnya prediksi yang akurat dan tepat untuk membuat keputusan investasi yang informatif.

NVIDIA, sebagai salah satu perusahaan teknologi terkemuka di dunia, dikenal sebagai produsen utama GPU (*Graphics Processing Units*). Keberhasilan dan proyek besar di bidang teknologi menjadikan NVIDIA salah satu pilihan utama bagi para investor. Minat yang tinggi terhadap saham NVIDIA menyebabkan pergerakan harganya sangat fluktuatif.

Fase Data Understanding

Data Understanding adalah tahap pengumpulan data dengan mengumpulkan data yang relevan terhadap pencapaian tujuan bisnis (Damayanti & Kuswayati, 2018). Pada penelitian ini Penulis menggunakan dataset harga saham NVIDIA dari awal tahun 2017 - akhir tahun 2023 dengan total data sebanyak 1760 data. Dataset tersebut memiliki 7 atribut yang dijelaskan pada Tabel 1.

Tabel 1. Variabel dalam Dataset

Variabel	Keterangan
<i>Date</i>	Tanggal perdagangan saham
<i>Open</i>	Pembukaan harga saham
<i>High</i>	Harga saham tertinggi
<i>Low</i>	Harga saham terendah
<i>Close</i>	Penutupan harga saham
<i>Adj Close</i>	Harga penutupan yang disesuaikan aksi korporasi
<i>Volume</i>	Jumlah lembar yang diperdagangkan

Fase Data Preparation

Menurut Purnama *et al.* (2012) Data Preparation adalah fase pemilihan dan pengolahan data yang nanti akan diperlukan dalam tahap pemodelan. Pada tahap ini Penulis melakukan pengecekan data, pembersihan data dan memastikan data-data yang digunakan valid dan akurat. Pada fase data preparation ini Penulis menggunakan bahasa pemrograman Python yang dijalankan pada aplikasi jupyter notebook.

Dataset dilakukan uji *missing value* melalui *syntax code* `data.isna().sum()`. Berdasarkan hasil uji *missing value* yang dilakukan, dataset tersebut tidak menunjukkan keberadaan *missing value*. Kemudian, dataset dilakukan pengecekan data bernilai 0 melalui *syntax code* `zero_values = (data == 0).any().any()`. Pengecekan saham bernilai 0 ini cukup penting untuk menghindari adanya data *outlier value*. Berdasarkan pengujian yang dilakukan, pada dataset ini tidak terdapat data yang bernilai 0.

Fase Modelling

Pada tahap pemodelan Penulis membuat model prediksi harga saham NVIDIA dari dataset saham NVIDIA pada tahun 2017 - 2024. Pada penelitian ini Penulis menggunakan metode SARIMA dan ARIMA yang akan memprediksi saham dari NVIDIA ke depannya dan membandingkan tingkat akurasi yang dipakai dalam melakukan *forecasting* saham dari NVIDIA.

Fase Evaluation

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|Y_t - \hat{Y}_t|}{Y_t}$$

Y_t = nilai aktual pada periode t

\hat{Y}_t = nilai *forecast* pada periode t

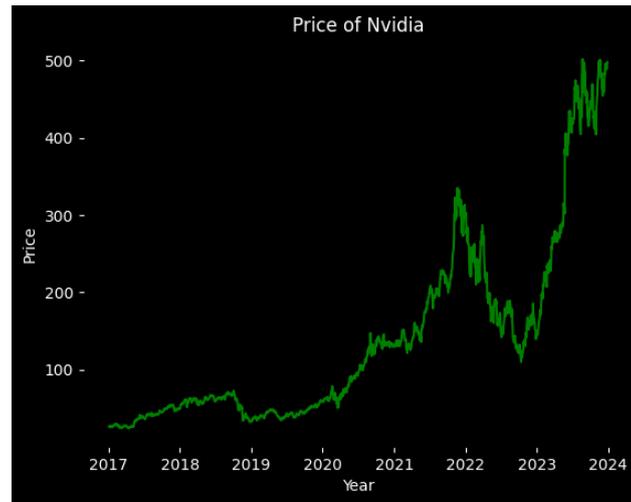
MAPE (*Mean Absolute Percentage Error*) adalah sebuah matrix evaluasi yang bisa digunakan untuk mengukur rata-rata kesalahan antara nilai aktual dan nilai prediksi (Widiyanto *et al.*, 2023). Semakin rendah MAPEnya maka semakin akurat hasil prediksi ARIMA ataupun SARIMA yang dihasilkan.

Fase Deployment

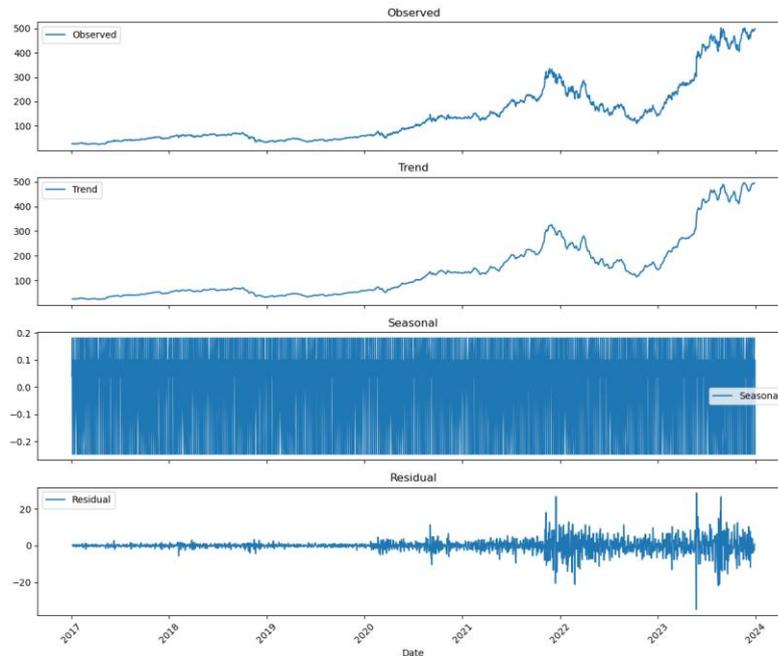
Setelah melakukan tahap evaluasi, maka dilakukan tahap *deployment* yang dimana pada tahap ini model yang dibuat perlu disesuaikan sehingga memberikan hasil yang sesuai dengan tujuan awal penelitian. Supaya hasil model prediksi/*forecasting* saham NVIDIA ini bisa lebih dipahami maka perlu divisualisasi menggunakan grafik.

HASIL DAN PEMBAHASAN

Plot data saham NVIDIA pada awal tahun 2017 hingga akhir tahun 2023 yang terdapat pada Gambar 3. terlihat mengalami peningkatan yang cukup signifikan. Pada gambar tersebut penurunan terjadi pada tahun 2022 dan mengalami kenaikan ekstrem pada tahun 2023. Penulis menggunakan atribut tahun sebagai variabel X dan atribut *open*/harga saham saat pasar saham dibuka sebagai variabel Y .



Gambar 3. Grafik Harga Saham NVIDIA



Gambar 4. Grafik Dekomposisi dari Atribut Open Saham NVIDIA tahun 2017 - 2023

Pada plot asli dapat dilihat pergerakan harga saham NVIDIA terus menerus mengalami kenaikan, begitu pula pada plot trend dapat dilihat juga secara jangka panjang harga cenderung naik dengan beberapa fluktuasi tertentu. Kemudian untuk plot musiman dapat dilihat bahwa polanya berulang dan terletak di sekitar nol yang berarti data saham NVIDIA yang dipakai tidak memiliki pola seasonal. Pada plot residu, dapat terlihat adanya fluktuasi yang cukup besar pada harga saham NVIDIA pada tahun 2023. Hal ini sejalan dengan kenaikan harga saham NVIDIA yang cukup tajam pada tahun 2023.

Dalam mengolah dataset saham NVIDIA dari awal tahun 2017 - akhir tahun 2023 Penulis menggunakan 80 persen data sebagai train data dengan tujuan agar model memiliki data yang cukup untuk dipelajari, semakin banyak data yang digunakan untuk melatih model, semakin baik performanya serta ini membantu model dalam memahami pola dan tren dalam data dengan lebih baik. 20 persen sisanya sebagai testing data yang bertujuan untuk memvalidasi dan memberi gambaran mengenai perbedaan hasil dari model dengan data yang sesungguhnya. Data yang digunakan sebagai training data berjumlah 1408 data. Seluruh data yang digunakan untuk train data dan testing data adalah atribut open.

Tabel 2. Pembacaan Notasi ARIMA

Hasil ARIMA	Interpretasi
(0,1,1)	Tidak ada komponen <i>Autoregressive</i> (AR) dalam model, deret waktu didiferencing satu kali dan memiliki satu komponen <i>Moving Average</i> (MA)
(0,0,0)	Tidak ada komponen musiman dalam model
[0]	Tidak ada periode musiman yang ditentukan

Pada saat dilakukan pemodelan ARIMA didapati hasil sebagai berikut ARIMA(0,1,1)(0,0,0)[0]. Hasil tersebut diinterpretasikan detailnya pada Tabel 2. Hasil ini menunjukkan bahwa data saham NVIDIA pada tahun 2017-2023 tidak mempunyai karakteristik musiman.

Tabel 3. Pembacaan Notasi SARIMA

Hasil ARIMA	Interpretasi
(0,1,1)	Tidak ada komponen <i>Autoregressive</i> (AR) dalam model, deret waktu <i>diferencing</i> satu kali dan memiliki satu komponen <i>Moving Average</i> (MA)
(0,0,0)	Tidak ada komponen musiman dalam model
[5]	Menggunakan siklus atau pola berulang setiap 5 periode

Pada saat dilakukan pemodelan SARIMA Penulis menggunakan $m=5$. Hal ini karena perdagangan saham hanya berlangsung lima hari (Senin-Jumat). Hasil pengujian didapati sebagai berikut ARIMA(0,1,1)(0,0,0)[5]. Hasil tersebut diinterpretasikan detailnya pada Tabel 3. Hasil ini menunjukkan bahwa data saham NVIDIA memiliki pola berulang setiap 5 periode meskipun tidak memiliki komponen musiman yang signifikan.

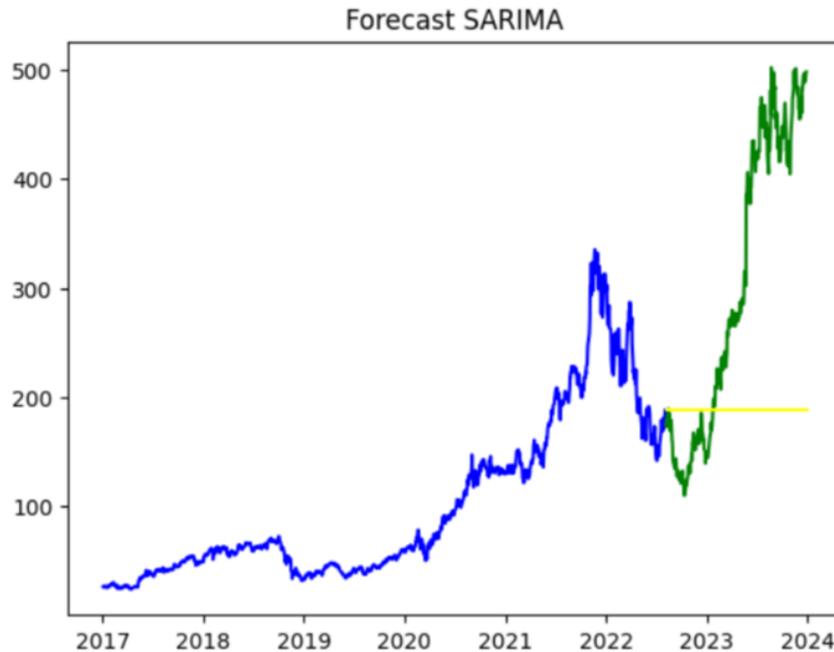


Gambar 5. Forecasting ARIMA

Pada Gambar 5, terlihat ada 3 warna yaitu biru, hijau, dan kuning. Warna ini memiliki arti masing-masing, biru menandakan data-data yang di *training* yaitu sebanyak 80%, hijau merupakan data untuk pengujian sebanyak 20% dan kuning adalah hasil dari *forecasting* ARIMA. Dari gambar tersebut dapat dilihat bahwa ada perbedaan yang cukup signifikan pada hasil *forecast* (kuning) dengan data yang sebenarnya (hijau). Dari hasil ARIMA, *Mean Squared Error* (MSE) memiliki nilai 31172,916, hal ini menunjukkan bahwa nilai perbedaan antara hasil *forecast* dengan data aslinya ketika dikuadratkan sebesar 31172.916, sedangkan untuk nilai *Mean Absolute Percentage Error* (MAPE) nya adalah 39% yang berarti ada penyimpangan sebanyak 39% dari nilai yang seharusnya. Berdasarkan prediksi ARIMA dapat dilihat, harga saham NVIDIA tidak mengalami kenaikan yang signifikan walaupun didata yang sebenarnya mengalami kenaikan. Salah satu penyebab adanya perbedaan hasil prediksi ARIMA dengan data yang sebenarnya dikarenakan adanya volatilitas pasar yang cukup tinggi sehingga membuat pergerakan harga saham NVIDIA menjadi sulit diprediksi.

Sama seperti pada Gambar 5, *forecasting* SARIMA pada gambar 6 juga memiliki 3 warna dan memiliki fungsi yang sama juga yaitu biru merupakan train data, hijau adalah test data, dan kuning adalah hasil *forecast*. Hasil dari MSE untuk SARIMA memiliki kesamaan dengan ARIMA yaitu 31172.916 dan MAPE nya adalah 39%. Berdasarkan prediksi SARIMA dapat dilihat, harga saham NVIDIA tidak mengalami kenaikan yang signifikan walaupun didata sebenarnya mengalami kenaikan. Penyebab terjadinya

perbedaan antara hasil SARIMA dengan data yang sebenarnya karena adanya volatilitas pasar yang cukup tinggi sehingga menyebabkan prediksi SARIMA menjadi tidak akurat.



Gambar 6. Forecasting SARIMA

KESIMPULAN

Penelitian ini menunjukkan data saham NVIDIA yang digunakan dari awal tahun 2017 - akhir tahun 2023 tidak memiliki musiman. Hal ini sesuai dengan hasil pemodelan baik ARIMA dan SARIMA yang tidak menunjukkan adanya komponen musiman dalam tabel sehingga akan lebih baik jika hanya menggunakan pemodelan ARIMA. Selain itu, hasil pemodelan SARIMA dan ARIMA terbilang baik untuk prediksi/forecasting harga saham NVIDIA kedepan. Dari hasil prediksi (*forecast*) tersebut dapat diketahui bahwa harga saham NVIDIA tidak mengalami kenaikan, walaupun pada data sebenarnya harga saham NVIDIA mengalami kenaikan yang signifikan. Saran untuk pengembangan penelitian ke depan bisa menggunakan metode prediksi yang berbeda untuk membandingkan keakuratan prediksi antara metode SARIMA, ARIMA atau metode prediksi yang lain.

DAFTAR PUSTAKA

- Damayanti, S. E., & Kuswayati, S. (2018). *Analisis dan Implementasi Framework CRISP-DM (Cross Industri Standard Process for Data Mining) untuk Clustering Perguruan Tinggi Swasta*. Sekolah Tinggi Teknologi Bandung, Indonesia.
- Hasanah, A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model

Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing (JAIC)*, vol.5, no.2, 103-108.

Montgomery, D. C., Jennings, C. D., & Kulahci, M. (2007). *Introduction to time series analysis and forecasting*. Wiley.

Ningrum. (2022). *Prediksi pergerakan saham JKSE dengan metode ARIMA menggunakan software R guna keputusan investasi pada masa pandemi Omicron* (Undergraduate thesis). Universitas PGRI Adi Buana Surabaya, Indonesia.

Purnama, I., Saputra, R., & Wibowo, A. (2012). Implementasi Data Mining Menggunakan CRISP-DM Pada Sistem Informasi Eksekutif Dinas Kelautan dan Perikanan Provinsi Jawa Tengah. Universitas Diponegoro Semarang, Indonesia.

Widiyanto, M. H., Rini, M., & Garno. (2023). Implementasi time series pada data penjualan di GAIKINDO menggunakan algoritma seasonal ARIMA. *JATI: Jurnal Mahasiswa Teknik Informatika*, vol.7, no.3, 1501-1506.



© 2024 by authors. Content on this article is licensed under a Creative Commons Attribution 4.0 International license. (<http://creativecommons.org/licenses/by/4.0/>).