Journal of Computer Science

Original Research Paper

Comparative Study of Garuda Indonesia Stock Price Prediction Using SVM, LSTM and Multiple Linear Regression

[1]Muhammad Naufal Luthfi and [2]Evaristus Didik Madyatmadja

[1]*BINUS Graduate Program Master of Information Systems Management, Bina Nusantara University, Jakarta, Indonesia*
[2]*Department of Information Systems, School, Bina Nusantara University, Jakarta, Indonesia*

Article history
Received: 05-08-2024
Revised: 30-10-2024
Accepted: 18-12-2024

Corresponding Author:
Muhammad Naufal Luthfi
BINUS Graduate Program
Master of Information Systems Management, Bina Nusantara University, Jakarta, Indonesia
Email: muhammad.luthfi008@binus.ac.id

Abstract: Stock shares are one of the investment products or tools that have been used by many people. Shares have interesting options for saving or investment and are able to provide attractive returns based on corporate or company growth. Many factors can affect the share prices, whether internal or external companies. This research was conducted on Garuda Indonesia's stock price with the shares code (GIAA); Garuda Indonesia is one of the big airlines in Indonesia. Machine learning and deep learning are popular topics that give insight and recommendations for stock price movement and prediction. In this study, the researcher will compare the multiple linear regression, support vector machine, and long short-term memory model to give new insight to other researchers and investors using stock price data and exchange rate between IDR and USD data for a better decision in stock investment strategies. The results show that multiple linear regression gave the best result in predicting the stock price movement of Garuda Indonesia company with exchange rate currency between IDR and USD, with the best value result of R-Squared, MAPE, MSE, and RMSE. Showing that the exchange rate between IDR and USD is influenced by stock price movement.

Keywords: Machine Learning, Artificial Intelligence, Data Mining, Stock Price Prediction, Garuda Indonesia

Introduction

Shares are an investment tool or instrument that is often used by many people because they can provide attractive returns. Shares can be interpreted as the participation of a person or corporate body in a company or limited liability company (PT). The average surplus obtained by each person who invests capital reaches between 11-13% per year, supported by the performance of the IHSG. However, many investors in Indonesia want to take part in purchasing shares but get results that do not match expectations or even suffer losses. The advantages of buying shares include having a dividend, which gives a percentage of profits to shareholders; the amount given depends on the results of the shareholders' general meeting.

The advantage of buying shares is having a dividend, which gives a percentage of profits to shareowners; the amount given depends on the results. You get capital gains when shareholders sell their shares at a more profitable price than the usual price.

Apart from the advantages, there are also disadvantages in buying shares, such as not getting dividends because the company we invested in does not have good performance or performance declines. The second is the opposite of capital gain, namely capital loss, where shareholders sell shares at a condition that is lower than the purchase price. The last one is liquidity risk, which is where the issuer goes bankrupt or is liquidated so that the worst probability is that shareholders will not make a profit.

According to CNBC Indonesia news on January 25, 2024, written by (Puspadini, 2024), the number of investors in Indonesia reached 12.2 million Single Investor Identification (SID) as stated by the development director of the Indonesian stock exchange (Fig. 1), Mr. Jeffrey Hendrick, who said that the number of investors in Indonesia only covers 5% of the total population in Indonesia. This figure is very small compared to Singapore, which has reached 30-40% of its total population. Below is a graph showing the development of the number of stock investors in Indonesia from March 2023-December 2023.

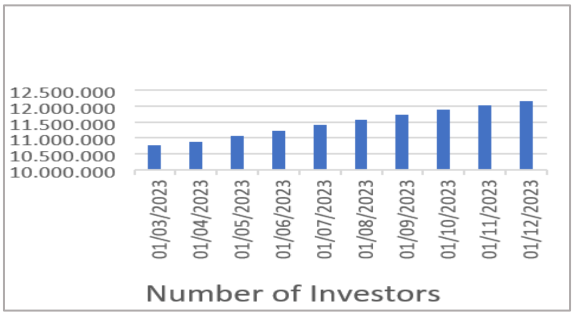

Fig. 1: Number of investors in Indonesia

In the current technological era, artificial intelligence can help predict stock prices by studying patterns in each variable that are better than simple statistical methods. Machine Learning (ML) and Deep Learning (DL) can make predictions based on time series data. ML is implemented to read data patterns to determine stock data movements and reduce investment risk when making decisions. Meanwhile, DL is a technology used to simulate human-like neural networks and solve complex non-linear problems.

Predicting time series data is generally very difficult due to the unprecedented changes caused by changing economic trends. Therefore, an assessment of forecasting accuracy is very necessary when using various forms of machine learning models, as we know that each model has several limitations. Some examples of models used to analyze time series data include Support Vector Machine (SVM), Long Short-Term Memory (LSTM) and multiple linear regression.

Many previous studies have used stock predictions, machine learning and artificial intelligence, among others. Analysis and forecasting of time-series data using S-ARIMA, CNN and LSTM (Dwivedi *et al*., 2021). Stock Pred: A framework for stock price prediction (Sharaf *et al*., 2021). A Comparison of ARIMA and LSTM in forecasting time series (Siami-Namini *et al*., 2018).

We can formulate the problem in this research. The first question compares three machine learning models that have been proposed and the second is how USD and IDR exchange rates affect stock price. The goal of this research is to know the best model proposed and the answer to the USD and IDR exchange rates that can affect the stock price.

The research benefits are separated into two perspectives; the first is for the researcher; we expect that this research can help with machine learning and deep learning knowledge to predict stock price movement. So, other researchers will hopefully use this research as a reference for studying machine learning and deep learning used to predict stock price movement. The second benefit is for investors. Hopefully, this research can help investors in making decisions for buying and selling and make this research a reference on stock price trading.

Ethical Considerations

This study does not involve any student and organization questionnaires. The stock price data and the Indonesia Rupiah and US dollar exchange were downloaded on public space in each organization.

Literature Review

Stock

Shares are trading activities in securities on the stock exchange. The stock exchange or capital market is a place where private company activities take place in the form of investment. Shares are one of the ways for companies to fund company capital. By issuing shares in 2 classes, you can get attractive profits both from the side of the company, which gets capital income for company expansion and investors who invest in the form of shares and gain profits from the expansion or development of the business.

Stock price always has different movements, and there are two factors that can influence share prices: the external factors and internal factors condition of the company influence the movement of the company's share price; the external factors are macroeconomic fundamental conditions and fluctuation in the rupiah exchange rate. Related to foreign currencies, government policies, panic factors and market manipulation, these factors occur outside the company's internal control. The following are internal factors that can influence share prices and the impact of company activities, including company fundamental factors, company corporate actions, and company performance projections in the future (Otoritas Jasa Keuangan, 2019).

Data Mining

Data Mining is a method used to extract large amounts of data and look for patterns or insights from the data collection. The techniques used in data mining are statistical techniques to summarize and analyze patterns and trends in data, then machine learning to build an algorithm that learns from data and makes predictions and, finally, a database system to store, manage and retrieve data automatically and efficiently.

There are several methodologies used in carrying out data mining processes, including CRISP-DM (Cross-Industry Standard Process for Data Mining), KDD (Knowledge Discovery in Database) and SEMMA (Sample, Explore, Modify, Model, Assess). However, in this research, we will use the CRISP-DM framework as a data mining process.

Machine Learning

Machine learning is a technique for studying patterns and shapes using data and statistics. Machine learning models work by providing input in the form of data so that the model created can provide output in the form of decisions, recommendations, and predictions. There are three categories of how machine learning learns:

- Supervised learning is where the model created learns a pattern from data that has been labeled. This model can map input and output
- Unsupervised learning is a model used on data that does not have labels. Unsupervised learning is tasked with reviewing data so that hidden patterns or data groupings can be explored. Usually used in clustering or grouping analysis
- Reinforcement learning is a method used for users who must make decisions and actions in certain circumstances with the aim of maximum rewards. The three categories above are used in different conditions and the form of data presented to study patterns

Linear Model

The linear model represents a fundamental algorithm that can produce a linear relationship between the input feature (Independent Variable) and the target output (Dependent Variable). Has a simple character, ability to interpret, and efficiency in training data and predictions? Linear models have parameters in the form of coefficients that determine the relationship between the independent variable and the dependent variable. Training is carried out by a linear model by learning optimal parameters, minimizing the loss function and calculating or measuring the error between the predicted value and the actual value.

Some linear models used in machine learning include linear regression, multiple linear regression, logistic regression, support vector machine, and linear discriminant analysis. According to (Nunno, 2014), stock price movements can be predicted using several linear regression models such as support vector machine, linear regression, multiple linear regression and neural network-based regression.

Multiple Linear Regression

Multiple Linear Regression is a model that can show the relationship between one variable and another variable. Usually, multiple linear regression is used to carry out predictive analysis with the aim of making decisions when taking steps in business (Panwar *et al*., 2021).

Support Vector Machine

SVM is an algorithm used for classification and regression in supervised learning. SVM can find patterns in complex data. SVM works by looking for a hyperplane that can maximize the margin between 2 classes. The hyperplane itself is the boundary that SVM uses to separate two classes of data.

Recurrent Neural Network

Recurrent Neural Network (RNN) has the basic feature that this model works in at least one feedback loop so that the selected label function can work in at least one loop. RNN is also designed to process sequential data like text data, speech, or time series data. The RNN module will perform at least one repetition. Mittal and Chauhan (2021).

Long Short-Term Memory

Long Short-Term Memory (LSTM) is a type of RNN that is designed to overcome the weakness of standard RNNs. LSTM introduces a special portal that can control the flow of information in and out of the hidden layer, which is the storage unit of the network. This portal can determine which information can be stored and discarded and how much can be input.

Cross Industry Process for Data Mining (CRISP-DM)

Cross Industry Process for Data Mining (CRISP-DM) is a methodology with an approach and process description for data mining projects. There are six implementation processes in the CRISP-DM methodology, which are business understanding, data understanding, data preparation, modeling, evaluation, and deployment. Pete *et al*. (2000).

Brzozowska *et al*. (2023), stated that CRISP-DM is able to achieve good-quality model results; the indicated model can be used for analytical support, planning, and making decisions. CRISP-DM also facilitates the smooth operation of an organization by improving the accuracy of decision making and identifying the decision-making process.

Related Work

Irzky Ardianta and Sari's research discusses two traditional techniques used by investors to predict stock prices, with forecasting carried out using previous data and fundamental analysis. This study also uses sentiment analysis, technical analysis, and fundamental analysis to predict stock price movements in Indonesia. This research uses a Support Vector Machine (SVM) as a prediction model. This research also uses data from news originating from the macroeconomic and microeconomic spheres, foreign stock price movements, and currency movements between the Rupiah and the Dollar for analysis. The results obtained by the author are that the average percentage of accuracy produced by SVM is 65.33%. It can be concluded that this research is useful for analyzing stock value movements in Indonesia. Irzky Ardyanta and Sari (2021).

Dwivedi *et al*. (2021), research analyzed and forecasted time-series data, namely Nifty-500 index data on the stock market using S ARIMA, CNN, and LSTM. And carried out comparisons and evaluations of the performance of each model, demonstrating promising results. The tools used in this research are Tensor Flow and Keras for implementation. MSE to measure the accuracy of the forecasting made by the model. The results of this research show that deep learning models outperform traditional machine learning models. Dwivedi *et al*. (2021).

Gao (2021), created this journal to discuss the use of Recurrent Neural Network (RNN) and Sequence to Sequence (Seq2Seq) models in machine learning, specifically for tasks such as machine translation and stock index prediction. This journal explains the function and structure of RNN, Long Short-Term Memory (LSTM) and attention layers in this model, and it provides code examples for implementing Seq2Seq in predicting stock price indices. This research also compares the performance of differentiated time-series data models to predict stock price movements from the Dow Jones Industrial Average (DIJA) and finds that the LSTM and Seq2Seq models outperform other models in calculating the mean squared error. Features low latency relativity and sequential prediction. Gao (2021).

Yan *et al*. (2021), This journal discusses the use of LSTM deep neural networks to predict stock prices based on data from the previous N days. Comparing LSTM with other neural networks and traditional statistical models, the aim is to improve prediction accuracy in financial market time series data. This research also focuses on optimizing the training process, comparing differences in optimization methods, and exploring the consequences of the input provided. The results of the research obtained from this research are that LSTM deep neural networks are very effective in predicting stock prices and in solving the problem of vanishing gradients in traditional RNNs. Yan *et al*. (2021).

Siami-Namini *et al.* (2018) this journal compares the performance of ARIMA models with deep learning based on LSTM models in predicting time series data with economic relations and financial variables. This journal discusses the ARIMA and LSTM algorithms and evaluates the level of accuracy using RMSE as a measuring tool. This journal also shows that LSTM outperforms ARIMA with an average error reduction in the range of 84-87%. This research also emphasizes the influence of parameters such as epochs and neurons in LSTM training models and supports the use of deep learning based on algorithms in the economic and financial sectors. Siami-Namini *et al*. (2018).

Febrilia *et al*. (2021), this research discusses the implementation of a Support Vector Machine (SVM) to predict stock movements at Garuda Indonesia Tbk. The data used is Garuda Indonesia stock data with a time period of March 18, 2019-April 23, 2021. The SVM algorithm achieved a prediction accuracy score of 0.545. It was concluded that SVM was able to help investors in making decisions about buying and selling shares. Febrilia *et al*. (2021).

Bansal *et al*. (2022), this research predicts stock prices using five machine learning models, including K-Nearest Neighbors, linear regression, support vector regression, decision tree regression, and long short-term memory. The data was taken from 12 Indian companies for more than 7 years for analysis. The results of this research show that LSTM outperforms all algorithms in terms of accuracy level, then followed by SVR, which is second in terms of performance, Linear Regression, and Decision Tree Regression show the same results, and the last algorithm, K-NN, shows poor results. Satisfactory in predicting stock prices. Bansal *et al*. (2022).

In this research by Akhtar *et al*. (2022), predictions were made of stock price movements that focused on preprocessing raw data and using machine learning algorithms. The machine learning models used are random forest and support vector machine. The method proposed by the researchers obtained an accuracy score of 78.7% for Support Vector Machine and an accuracy score for random forest of 80.8%. Akhtar *et al*. (2022).

Ketut *et al.* (2023) this journal compare optimization models (Adam, SGD, RMSprop) on LSTM aimed at predicting the share price of Telkom Indonesia, Tbk from January 1, 2019, to January 11, 2023. LSTM shows results with very good prediction accuracy with low values using Mean Absolute Percentage Error (MAPE). And Adam's optimization shows an. 45%. Ketut *et al*. (2023) accuracy of 98.

Chrysmien and Jayadi (2022) this research compare LSTM MLR and ARIMA machine learning on stock price movements with additional sentiment data and Rupiah and USD exchange rate movements. The stock data analyzed is FREN stock data, namely, a telecommunications company in Indonesia. The results of this research show that Multiple Linear Regression is the best model for predicting stock prices, with figures of 473,875 in MSE and 21,768 in RMSE in training data analysis; for testing data, it achieved figures of 74,181 in MSE and 8,612 in RMSE. Chrysmien and Jayadi (2022).

It can be concluded from previous research that Long Short-Term Memory (LSTM) shows good figures in predicting stock price movements (Bansal *et al*., 2022; Dwivedi *et al*., 2021; Gao, 2021; Ketut *et al*., 2023; Siami -Namini *et al*., 2018; Yan *et al*., 2021) and several studies show that SVM is the best model in predicting stock prices (Akhtar *et al*., 2022; Febrilia *et al*., 2021; Ardyanta and Sari, 2021). Other researchers have also stated that multiple linear regression is the best model for predicting stock prices (Chrysmien and Jayadi, 2022).

Materials and Methods

Researchers use the CRISP-DM framework to predict stock price movements. This chapter will explain it into six stages consisting of Business understanding, data understanding, data preparation, modeling and deployment.

Business Understanding

Garuda Indonesia is the first civil flight in Indonesia that was born on January 26, 1949, under the initiative of the Republic Indonesia Air Force. At this moment, Garuda Indonesia serves more than 90 destinations in local and international, with 600 flights in one day. Garuda Indonesia group also operates around 210 fleet; 142 planes operate as the main brand of Garuda Indonesia and 68 planes operate as the main brand of Citilink.

Garuda Indonesia is a member of SkyTeam and the 2nd largest airline in Indonesia after Lion Air. Garuda Indonesia shares were chosen as the topic to discuss stock price movements due to the release of Unusual Market Activity (UMA) from the Indonesia Stock Exchange (BEI) in January 2023 because the share movement looks very unusual.

Garuda Indonesia (GIIA) shares that operate in the aviation transportation sector or another name for commercial air transportation services, which operate in the transportation and logistics sector, including in the airline industry. In this sector, there are also stocks such as PT AirAsia Indonesia, Tbk (CMPP) and PT Jaya Trishindo, Tbk (HELI).

Data was collected from 2 website portals, where first we obtained data from finance.yahoo.com to collect data on Garuda Indonesia shares as seen (Fig. 2) and bi.go.id to collect data on changes in the Indonesian currency exchange rate (Rp) to the United States (USD) on (Fig. 3). In the time span from January 2023-January 2024.

Data Understanding

Data understanding: The author describes each attribute contained in the data and defines the data that will be used in this research. Because the research obtained two different types of data, the types of data will be explained in Table (1) for the stock movement data table and Tables (2-3) for the Rupiah and USD exchange rate movements.

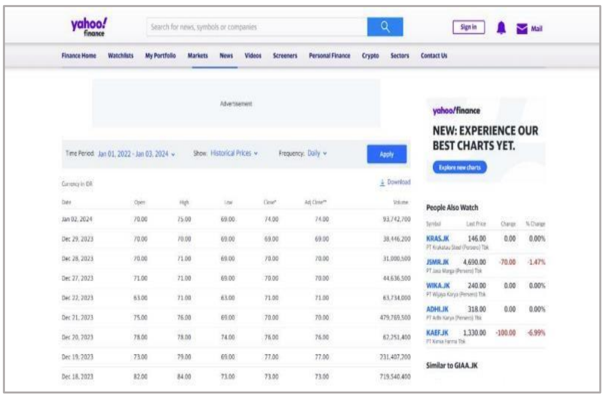

Fig. 2: Aruda Indonesia stock price on yahoo finance website

Fig. 3: Bank Indonesia website page

Table 1: Stock data variable attribute

Attribute name	Description
Date	Share the sale and purchase date
Open	The opening price of shares in one trading day
High	Highest price of the stock in one trading day
Close	Closing stock price in one trading day
Adjusted close	The share price at the end of the buying and selling day has been changed by additional distributions and corporate actions that occur before the next day opens
Low	Lowest price of the stock in one-day trading
Volume	Number of purchased stocks that have been trading in a certain period

Table 2: Currency exchange data variable

Variable name	Description
Date	Exchange rate date
Forex Sell	Forex selling price
Forex Buy	Forex buying price
Value	The exchange value between the currency

Table 3: Selection data attribute

No.	Variable
1	Date
2	Open
3	High
4	Low
5	Close
6	Adj Close
7	Volume
8	Forex Sell
9	Forex Buy

Data Preprocessing

Here, the process carried out consists of data visualization, data integration, data cleaning and data testing. In this research, data preprocessing was carried out in Google Collab using Python. The goal of data preprocessing is to create the best data quality so that it can be continued to the next stage. The data in the research will be combined under the name "master data," which is a combination of stock movement data and IDR to USD currency movement data.

Data Selection

By referring to the data above, we can determine which attributes we will use in this research. Below are the variables that will be used in the research.

Data Visualization

Stock Data Visualization

It can be seen in the graph (Fig. 4) that when the stock price is at a low position or going towards a low position, transaction volume increases. It can be concluded that there is an inverse relationship between stock price and sales volume.

Exchange Rate Visualization

It can be seen in the picture (Fig. 5) that when stock prices rise on (Fig. 4), the currency exchange rate also rises (Fig. 5).

DOI: 10.3844/jcssp.2025.817.826

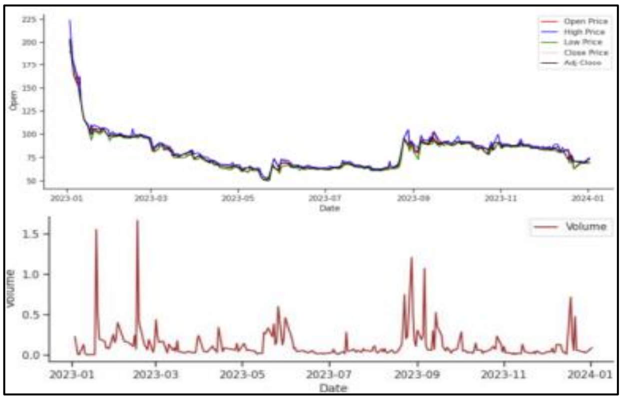

Fig. 4: Visualization of stock price exchange

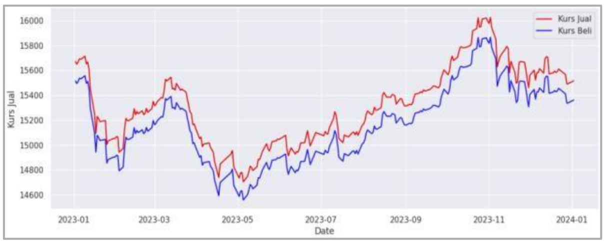

Fig. 5: IDR/USD exchange rate visualization

```
main_df = saham_df.set_index('Date').join(kurs_df.set_index('Date'))
main_df = main_df.reset_index()
main_df.head()
```

	Date	Open	High	Low	Close	Adj Close	Volume	NO	Nilai	Kurs Jual	Kurs Beli
0	2023-01-02	203.554123	203.554123	203.554123	203.554123	203.554123	0	243	1	15669.96	15514.04
1	2023-01-03	204.000000	224.000000	190.000000	202.000000	202.000000	231656400	242	1	15649.86	15494.14
2	2023-01-04	194.000000	204.000000	188.000000	188.000000	188.000000	112231400	241	1	15667.95	15512.05
3	2023-01-05	177.000000	183.000000	175.000000	175.000000	175.000000	13209400	240	1	15693.08	15536.92
4	2023-01-06	175.000000	175.000000	163.000000	163.000000	163.000000	8413200	239	1	15688.05	15531.95

Fig. 6: Combining stock data and exchange rate data

```
main_df.isnull()
```

	Date	Open	High	Low	Close	Adj Close	Volume	NO	Nilai	Kurs Jual	Kurs Beli
0	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...
235	False	False	False	False	False	False	False	False	False	False	False
236	False	False	False	False	False	False	False	False	False	False	False
237	False	False	False	False	False	False	False	False	False	False	False
238	False	False	False	False	False	False	False	False	False	False	False
239	False	False	False	False	False	False	False	False	False	False	False

240 rows × 11 columns

Fig. 7: Checking missing value

```
main_df.head()
```

	Date	Open	High	Low	Close	Adj Close	Volume	Kurs Jual	Kurs Beli
0	2023-01-02	203.554123	203.554123	203.554123	203.554123	203.554123	0	15669.96	15514.04
1	2023-01-03	204.000000	224.000000	190.000000	202.000000	202.000000	231656400	15649.86	15494.14
2	2023-01-04	194.000000	204.000000	188.000000	188.000000	188.000000	112231400	15667.95	15512.05
3	2023-01-05	177.000000	183.000000	175.000000	175.000000	175.000000	13209400	15693.08	15536.92
4	2023-01-06	175.000000	175.000000	163.000000	163.000000	163.000000	8413200	15688.05	15531.95

Fig. 8: Remove column "NO" and "Value"

Data Integration

At this stage, we combine the stock data and exchange rate data that have been used in Google Collab as seen on (Fig. 6) and able to see all attribute for the dataset.

Cleaning Data

In this process, we check whether there are missing values in the dataset and remove columns that are not needed to be run in the model so that the model can run optimally.

Checking for Missing Value

At this stage, we check whether there are any missing values in the data we have. It can be seen (Fig. 7) that the dataset we have does not have missing values for all of its attributes.

Remove the "NO" and "Value" Column

Because the “sequence number” column is not used in the model and the exchanged value column is not used in the model that will be implemented, we perform the drop function on these two columns; (Fig. 8) are the results after dropping the column in the dataset.

Splitting the Data

Before implementing the model with the dataset, the dataset should have split into 2 parts with a comparison of 80% for the training data and 20% for the testing data. According to (Bichri *et al*., 2024), increasing the size of train data to more than 70% of the dataset is required in the training step to achieve better performance.

Testing the Data

Testing the data is a critical component of the CRISP-DM process. Testing the data is needed before we use the data in our models. We use Correlation metrics to be carried out in order to determine whether the relationship between each variable is a linear relationship or is mutually correlated.

The second test is the Anderson Darling Test. This testing is used to check whether the data comes from a specific distribution; this test is very useful for small data sizes. The last test uses Durbin-Watson testing for detecting autocorrelation in the residuals of regression models.

Modelling

At this stage, the model selection is carried out by predicting stock prices using a machine learning and artificial intelligence approach by implementing data preprocessing by combining stock data and the exchange rate between Rupiah and USD data and checking if any missing value after the data combined and do standardization the data including testing the data before the data is used for the model to train. And model implementation by using 3 models proposed (LSTM, MLR and SVM). And expected the result by being able to predict stock price prediction. Is the workflow that will be used.

DOI: 10.3844/jcssp.2025.817.826

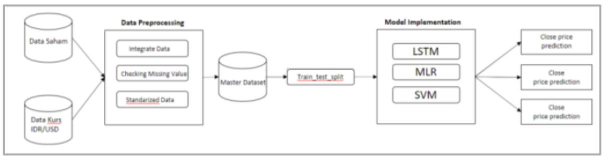

Fig. 9: Research workflow

As seen on (Fig. 9) the research workflow of stock data and exchange rate between IDR and USD is run through data preprocessing including integrate and combine the data, checking missing value and standardized the data and we have master dataset. Next, we split the data for train split data, after that using machine learning (LSTM, MLR and SVM) for machine learning to train and learn the data pattern, and finally given the output for stock price prediction.

Long Short-Term Memory

The LSTM model uses shape for one loop back, one dense for the LSTM model by finding the mean-square-error loss with Adam optimizer with 100 epochs.

Multiple Linear Regression

Multiple linear regression is a model that will be used to learn the relationship between features and targets. Train the model with training data so that the model can learn the patterns in the data. And make a prediction using a model that has been trained to predict the target value on the test data.

Support Vector Machine

The SVM model has three steps. First, we need to determine the grid parameters for grid search, which determines the set of parameters that will be tested to determine the best combination in the SVR model. Second, we create a basic SVR model and grid search. Third, we create a function to determine the best parameters and train the model. Finally, make predictions on testing data.

Results

This section is the fifth part of the CRISP-DM Evaluation, this section will show the data analyst result and test data result using the correlation matrix Anderson Darling test and Durbin Watson test. Elaborate on the result for each model in the previous section, the metric that evaluates stock price prediction will be R-Squared, Mape (Mean Absolute Percentage Error), MSE (Mean Square Error) and RMSE (Root Mean Square Error) for each data.

Data Testing Results

Using histogram of residuals, boxplot and Q and Q with result and diagram or graphs.

Histogram of Residuals

Shown on the (Fig. 10), the residuals are distributed symmetrically around zero, suggesting normality in the data that we use for our model.

Q-Q Plot of Residual Results

With the Q-Q plot of residuals on (Fig. 11), it is confirmed that the residuals closely follow a normal distribution, with most points lying on or near the 45-degree line, with only minor deviations at the tails. However, slight deviations are normal and acceptable. Q-Q Plot of residuals graph.

Correlation Metrics

The image shows on (Fig. 12) that the dependent variable in this study (Close) provides a very good correlation (0.99-1) for all stock data except volume at (0.075), which is almost uncorrelated and is followed by a correlation with a sufficient level (0.44).

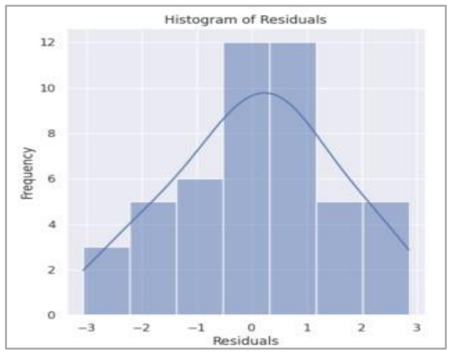

Fig. 10: Histogram of residual

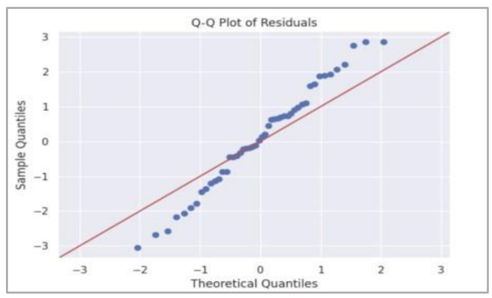

Fig. 11: Q-Q Plot of residuals

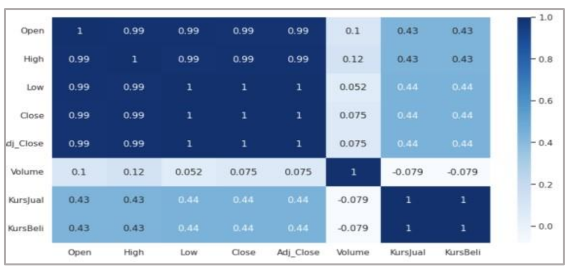

Fig. 12: Data correlation

DOI: 10.3844/jcssp.2025.817.826

```
from statsmodels.stats.diagnostic import normal_ad

# Performing the test on the residuals
p_value = normal_ad(residuals)[1]
print('p-value from the test Anderson-Darling test below 0.05 generally means non-normal: ', p_value)

if p_value < 0.05:
    print("The residuals are not normally distributed.")
else:
    print("The residuals are normally distributed.")

p-value from the test Anderson-Darling test below 0.05 generally means non-normal:  0.8728666510020506
The residuals are normally distributed.
```

Fig. 13: Anderson darling test

```
from statsmodels.stats.stattools import durbin_watson

durbin_watson = durbin_watson(residuals)
print('Durbin-Watson: ', durbin_watson)

if durbin_watson < 1.5:
  print("The residuals are positively autocorrelated.")
  print("Assumption not satisfied")
elif durbin_watson > 2.5:
  print("The residuals are negatively autocorrelated.")
  print("Assumption not satisfied")
else:
  print("The residuals are not significantly autocorrelated.")
  print("Assumption satisfied")

Durbin-Watson:  1.9847762084118328
The residuals are not significantly autocorrelated.
Assumption satisfied
```

Fig. 14: Durbin watson test

Table 4: Model evaluation matrix

Model evaluation	SVM		LSTM		MLR	
	Data train	Data test	Data train	Data test	Data train	Data test
R2	0.995	0.9816	0.9421	0.823	0.9962	0.992
Mape	0.013	0.0174	11.090	0.021	0.0125	0.014
MSE	2.534	5.1069	0.0005	6.460	1.8992	2.206
RMSE	1.591	2.2598	0.0229	2.541	1.3781	1.485

Anderson Darling Test

The image on Fig. (13), showing the result of Anderson Darling's results with a P-Value of 0.8728666510020506 indicates that residuals from the regression model are normally distributed. This is a desirable outcome as it can validate one of the key assumptions of linear regression, supporting the reliability and validation of the model inference and predictions.

Durbin Watson Test

With the Durbin Watson shows on Fig. (14), statistic of approximately 1.98, the residuals are not significantly autocorrelated. The assumption of independence of residuals in the regression model is satisfied, which is important for the validity of the regression results. Are the images Durbin Watson test?

Model Performance Results

In this section, we will view the predicted value and actual value in each model on the graph and the model performance based on Metric evaluation of R-squared, Mean Absolute Percentage Error, Mean Squared Error, and Root Mean Squared Error shown on (Table 4).

R^2 result shows that MLR has the highest R^2 values for both training (0,996) and testing data (0,992), which indicate MLR has a very strong fit. SVM is slightly lower than MLR with R^2 values of training (0,995) and testing (0,982). LSTM shows a lower value of R^2, especially on the test data (0,824) train (0,942).

Mean Absolute Percentage Error shows that MLR has the lowest score on both training (1,26%) and test (1,46%), indicating that MLR has the highest accuracy in percentage terms. Followed by SVM with slightly higher MAPE value for train (1,35%) and test (1,74%). And LSTM has the highest MAPE result on the train (11,09%) and test (2,12%), suggesting larger deviations from the actual value.

Mean Squared Error and Root Mean Squared Error results show that MLR has the lowest MSE and RMSE on both train and test data, indicating that MLR has high accuracy. SVM has slightly higher MSE and RMSE than MLR. LSTM shows the highest MSE and RMSE results on the test data, meaning that its predictions are less accurate compared to MLR and SVM.

Discussion

For justification for model selection, given the metrics, MLR appears to be the most effective model for this dataset due to its simplicity and great performance. It not only captures the underlying pattern effectively but also generalizes it well, as evidenced by the low error metrics both on train data and test data. SVM could be an alternative, but it does not outperform MLR significantly. LSTM shows higher error and appears to be less effective in capturing the data pattern here. LSTM may not be suitable in this case because its performance metrics are considerably poorer, especially on the test data.

Below is the graph on each model performance with predicted value and actual value (Fig. 15).

SVM and MLR perform quite well in following the actual values. Meaning that these models capture the pattern effectively. This aligns with their low error metrics such as MSE, RMSE, and MAPE. The LSTM model does not capture the peaks and valleys, nor do SVM and MLR. This model is smoother, with an almost flat trend showing that the LSTM might not be well suited for the dataset. This analysis supports selecting MLR as the primary model, with SVM as a potential secondary option.

Based on the research of Chrysmien and Jayadi (2022), showing that multiple linear regression is the best model for predicting the stock price same what we achieved her in the research showing the multiple linear regression is the best model for predicting stock price combined with exchange rate between IDR and USD.

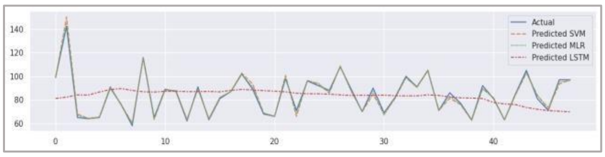

Fig. 15: Actual value and predicted value in each model

Conclusion

In this study, the focus was proposing an approach of using two combined datasets, which are the exchange rate between IDR and USD data and stock price data. Using artificial intelligence and machine learning to learn the data and predict them also gives investors insight into making decisions. With the result of the correlation metric, we can conclude that Exchange Rate data and Stock Price data are able to correlate with each other and have an effect on each variable.

Based on provided results on each model, we can rank the models according to the different metrics. The conclusion is that multiple linear regression performs best across all metrics for both test data and train data. The second rank goes to Support Vector Machine, with consistently better performance than LSTM but not as good as Multiple Linear Regression. The last rank is for Long Short-Term Memory because it lags behind significantly, with lower R-Square and higher error rates, and may not be the best fit for this particular problem of the dataset.

This research, by combining the stock price of the Garuda Indonesia dataset and the exchange rate currency between the IDR and USD dataset, hopefully can bring a new perspective on how using machine learning and artificial intelligence to analysis also help in decision-making in stock price execution. This research is also able to help investors buy, sell, hold and make decisions to understand the stock price movement.

Acknowledgment

We acknowledge Prof. Dr. Evaristus Didik Madyatmadja, ST., M.Kom., M.T., Lecturer of Information Systems, Department of Information Systems, School, Bina Nusantara University for his advice, motivation, input and support during the creation of the manuscript.

Funding Information

The authors have not received any financial support or funding to report.

Author’s Contributions

Muhammad Naufal Luthfi: Written the research, specifically analyzing and understanding the data, writing and executing the code, following the research framework, found results, and making a conclusion based on research purposes.

Evaristus Didik Madyatmadja: Designing the research framework, specifically providing guidance on creating the research framework based on previous research and offering recommendations for data analysis based on research findings.

Ethics

This article is original by the first and second authors and has not been previously published.

References

Akhtar, M. M., Zamani, A. S., Khan, S., Shatat, A. S. A., Dilshad, S., & Samdani, F. (2022). Stock market prediction based on statistical data using machine learning algorithms. Journal of King Saud University-Science, 34(4). https://doi.org/10.1016/j.jksus.2022.101940

Bansal, M., Goyal, A., & Choudhary, A. (2022). Stock Market Prediction with High Accuracy using Machine Learning Techniques. Procedia Computer Science, 215, 247–265. https://doi.org/10.1016/j.procs.2022.12.028

Bichri, H., Chergui, A., & Hain, M. (2024). Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets. In IJACSA) International Journal of Advanced Computer Science and Applications (Vol. 15, Issue 2). www.ijacsa.thesai.org

Brzozowska, J., Pizoń, J., Baytikenova, G., Gola, A., Zakimova, A., & Piotrowska, K. (2023). Data Engineering in Crisp-Dm Process Production Data – Case Study. Applied Computer Science, 19(3), 83–95. https://doi.org/10.35784/acs-2023-26

Chrysmien, & Jayadi, R. (2022). Comparison Stock Price Prediction Between Arima, Multiple Linear Regression and Lstm Models by Adding Stock Sentiment Analysis and Usd/Idr Fluctuation. Journal of Theoretical and Applied Information Technology, 100(4), 1158–1169.

Dwivedi, S. A., Attry, A., Parekh, D., & Singla, K. (2021). Analysis and forecasting of Time-Series data using S-ARIMA, CNN and LSTM. Proceedings - IEEE 2021 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2021, 131–136. https://doi.org/10.1109/ICCCIS51004.2021.9397134

Febrilia, R., Wulandari, T., & Anubhakti, D. (2021). Implementasi algoritma support vector machine (svm) dalam memprediksi harga saham pt. Garuda indonesia tbk. In indonesia journal information system (idealis) (vol. 4, Issue 2). http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/indexRatihFebriliaTriWulandari|http://jom.fti.budiluhur.ac.id/index.php/IDEALIS/index|

Gao, Z. (2021). Stock Price Prediction with ARIMA and Deep Learning Models. 2021 IEEE 6th International Conference on Big Data Analytics, ICBDA 2021, 61–68. https://doi.org/10.1109/ICBDA51983.2021.9403037

Irzky ARDYANTA, E., & Sari, H. (2021). A Prediction of Stock Price Movements Using Support Vector Machines in Indonesia. Journal of Asian Finance, 8(8), 399–0407. https://doi.org/10.13106/jafeb.2021.vol8.no8.0399

Ketut, I., Enriko, A., Nizar Gustiyana, F., Putra, R. H., & Kunci, K. (2023). JURNAL MEDIA INFORMATIKA BUDIDARMA Komparasi Hasil Optimasi Pada Prediksi Harga Saham PT. Telkom Indonesia Menggunakan Algoritma Long Short-Term Memory. https://doi.org/10.30865/mib.v7i2.5822

Mittal, S., & Chauhan, A. (2021). A RNN-LSTM-Based Predictive Modelling Framework for Stock Market Prediction Using Technical Indicators. International Journal of Rough Sets and Data Analysis, 7(1), 1–13. https://doi.org/10.4018/ijrsda.288521

Nunno, L. (2014). Stock Market Price Prediction Using Linear and Polynomial Regression Models.

Otoritas Jasa Keuangan. (2019, November 25). Penyebab naik turun harga saham suatu perusahaan. Https://Sikapiuangmu.Ojk.Go.Id/FrontEnd/CMS/Article/10507.

Panwar, B., Dhuriya, G., Johri, P., Yadav, S. S., & Gaur, N. (2021). Stock Market Prediction Using Linear Regression and SVM. 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), 629–631. https://doi.org/10.1109/ICACITE51222.2021.9404733

Pete, C., Julian, C., Randy, K., Thomas, K., Thomas, R., Colin, S., & Wirth, R. (2000). Crisp-Dm 1.0. CRISP-DM Consortium, 76.

Puspadini, M. (2024, January 25). BEI Sebut Investor RI Dua Kali Lipat dari Singapura, Tapi. CNBC Indonesia.

Siami-Namini, S., Tavakoli, N., & Siami Namin, A. (2018). A Comparison of ARIMA and LSTM in Forecasting Time Series. Proceedings - 17th IEEE International Conference on Machine Learning and Applications, ICMLA 2018, 1394–1401. https://doi.org/10.1109/ICMLA.2018.00227

Yan, X., Weihan, W., & Chang, M. (2021). Research on financial assets transaction prediction model based on LSTM neural network. Neural Computing and Applications, 33(1), 257–270. https://doi.org/10.1007/s00521-020-04992-7