



NAV of Equity Fund Forecasting Using ARIMA and RNN

Jennifer Alexandra¹

¹Information Systems Department, Bina Nusantara University, Jakarta, Indonesia
Email: 1jennifer.alexandra@binus.ac.id

Abstract

This study compares ARIMA and RNN methods to predict the Net Asset Value (NAV) of equity mutual funds using the Sucorinvest Equity Fund data from 2017 to 2019. The CRISP-DM framework is applied throughout the process. ARIMA uses only NAV data, while RNN includes stock indexes as additional variables. Prediction accuracy is evaluated using MSE and RMSE. The results show that although both methods yield similar error rates, RNN demonstrates slightly better performance.

Keywords: Equity Fund, NAV, ARIMA, RNN, CRIPS-DM

1. INTRODUCTION

Along the economic growth, currently investment can be made by anyone with a minimum of 100 thousand Rupiah in funds required, namely by investing in mutual funds. According to the Capital Market Law No. 8 of 1995, article 1 paragraph is defined that an Investment Fund is a container used to collect funds from the community of investors to subsequently be invested in securities portfolios by investment managers. Mutual funds themselves consist of various types, namely the money market, fixed income, stocks and mixed [1]. The rise of mutual fund investment is also supported by technological developments. Previously to buy mutual funds, customers need to come to the bank to buy products that are available, while now there are many applications that become online mutual fund sales agents, such as Bareksa, IPOTFUND, POEMS, and others. This application is certainly very useful and makes it easy for potential investors to find out and also makes it easier for investors to monitor their mutual fund portfolios [2].



According to CNBC Indonesia, total funds under management of the mutual fund industry in 2018 amounted to Rp 505.39 trillion. That number increased by 10.44% compared to the previous year. The increase in total managed funds in the mutual fund industry is supported by equity mutual fund products. The total funds under management of the equity mutual fund industry increased by 19% to Rp 143.8 trillion [3]. While the number of equity fund products also increased to 277 mutual fund products. The Investment Coordinating Board (BKPM) of the Republic of Indonesia also strengthened the data by stating that up to the first half of this year Indonesia's investment realization was showing an increasing trend. BKPM is optimistic that the investment target in Indonesia of Rp765 trillion this year will be achieved [4]. The Deregulation Director of the Investment Coordinating Board (BKPM) Yuliot said investment realization in the first quarter (January-March) was recorded at a quarter of the 2018 investment target of Rp765 trillion. This means that investment realization up to the first quarter amounted to Rp185.3 trillion [5] [6].

Each type of mutual fund has its own characteristics based on fund allocation, risk profile and various other factors [7]. When viewed from the level of risk, equity funds are a type of mutual fund with the characteristics of high risk-high return because the funds managed are largely invested in the stock market [8]. A mutual fund unit has a value called Net Asset Value (NAV). This NAV is calculated from total assets minus operating costs divided by unit participation. NAV continues to change every day depending on market conditions. For equity funds, NAV will be strongly influenced by the stock index [9].

Despite numerous studies on stock price forecasting using deep learning, there is a lack of research specifically focusing on predicting the daily NAV of equity mutual funds in the Indonesian market using advanced deep learning architectures [10] [11]. Most existing studies concentrate on stock price movements or general market indices, while NAV forecasting, especially for equity mutual funds, remains underexplored [12] [13]. Therefore, this study proposes a predictive model based on Recurrent Neural Network (RNN), particularly Long Short-Term Memory (LSTM), to forecast the daily NAV of equity mutual funds, addressing a crucial research gap and contributing to more accurate investment decision-making tools [14].

Therefore, this study will provide a proposed method that can be used to predict NAVs of stock mutual funds, because equity funds are the most aggressive type of mutual funds at risk. In addition, equity funds are also closely related to shares. Previously there have been many studies to predict stocks by analyzing the capital market. Some of them are using the Integrated Moving Average Autoregressive method (ARIMA) which is commonly used for the analysis of interconnected time series data. In addition, forecasting can also be done with an Artificial Neural Network (ANN) based approach. ANN's approach is based on artificial neural networks that pinpoint human intelligence to solve complex problems. At ANN, data components will be analyzed interrelated. The development of ANN that can be used for time series data analysis is Recurrent Neural Network (RNN) [15] [16].

2. METHODS

Business Understanding

There are many types of mutual funds found on the market [17]. This research will focus on equity funds only, but equity funds themselves also have a large variety of products. Therefore, a mutual fund product will be taken as a sample to conduct this research. The mutual fund product that will be used as a sample is certainly chosen because it is considered to be able to describe the characteristics of other equity funds.

The following are the criteria used in selecting stock mutual fund products as research objects:

- More than 5 years old
- Has assets of > 500 billion - 100 trillion
- Ever won the best ranked mutual fund award
- Funds are allocated to LQ45 shares
- Aggressive risk profile

These criteria are determined so that the object under study is a mutual fund product that is considered to be good and trusted by the community because it has a long age and has valuable assets [18]. It also won an award illustrating that the product is a quality product and can be a recommendation for investors who want to invest. Regarding the allocation of funds, the selected shares are LQ45 shares which are considered good and stable based on their liquidity and market capitalization [19]. An aggressive risk profile is a challenge for investors to

be able to choose the right time to conduct the mutual fund transaction. Based on this data, the preferred equity fund is the Sucorinvest Equity Fund, which is a product of Sucorinvest Asset Management [20]

As mentioned earlier, the stock mutual fund product that will be used is the Sucorinvest Equity Fund and obtained data based on the mutual fund's fund fact sheet, the allocation of 99.25% shares is on the stock market and only 0.75% on the money market. The following are the sectors that are allocated funds:



Figure 1. Fund Allocation

Table 1. Data understanding

Variable	Data Type	Information
<i>Tanggal (T)</i>	<i>Datetime</i>	The specific trading day used as the reference point for forecasting the NAV.
<i>IHSG (T-1)</i>	<i>Numeric/decimal</i>	The value of the Jakarta Composite Index (IHSG) on the previous day, which reflects the overall market movement influencing NAV.
<i>Mining Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's mining sector index, representing stock performance in the mining industry as part of equity fund allocation.
<i>Finance Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's financial sector index, reflecting the performance of banking,

Variable	Data Type	Information
		insurance, and financial institutions that impact NAV.
<i>Trade Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's trading sector index, indicating the movement of stocks related to wholesale and retail trade activities.
<i>Consumer Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's consumer goods sector index, which reflects the performance of companies producing consumer products
<i>Property Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's property sector index, representing the development of property-related stocks that may affect fund value
<i>Infrastructure Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's infrastructure sector index, reflecting stock performance in infrastructure development and utilities.
<i>Miscellaneous Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's miscellaneous sector index, capturing companies not included in other defined sectors
<i>Manufacture Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's manufacturing sector index, representing industrial production companies.
<i>Agriculture Idx (T-1)</i>	<i>Numeric/decimal</i>	Previous day's agriculture sector index, covering companies in the farming, plantation, and related industries

Variable	Data Type	Information
<i>Foreign net buy (T-1)</i>	<i>Numeric/decimal</i>	The value of net foreign investor purchases on the previous day, showing foreign capital inflow or outflow trends
<i>Gold price (T-1)</i>	<i>Numeric/decimal</i>	Previous day's international gold price, considered as a safe-haven asset that can influence investment sentiment
<i>Oil price (T-1)</i>	<i>Numeric/decimal</i>	Previous day's crude oil price, which reflects global commodity trends potentially impacting equity markets
<i>Coal price (T-1)</i>	<i>Numeric/decimal</i>	Previous day's coal price, relevant due to Indonesia's strong coal export sector and its impact on mining stocks
<i>Exchange rate (T-1)</i>	<i>Numeric/decimal</i>	Rupiah exchange rate against USD on the previous day, reflecting currency fluctuations that can affect investment performance.

- This research uses different tools for both methods. Research with ARIMA will use SPSS software. For ARIMA, it only uses 60 days of data, namely 10 June 2019 to 30 August 2019. Whereas RNN will use MatLab tools that use larger data sets. The data set will also be divided into 3 parts, namely training, validation and testing.
- For ARIMA, SPSS software with expert modeler is used and for RNN, MatLab software is used. The ARIMA model is derived from the integration between the AR and MA models so that it forms the ARIMA. The steps were identification, estimation, diagnostic check, and forecasting.

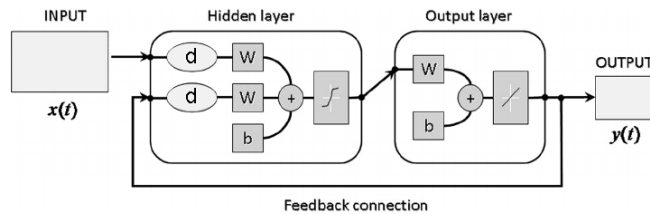


Figure 2. NARX Architecture

First of all, data for RNN was being normalized first. Even though the input set has different parameters, it will be used as a set of inputs that give weight to the hidden layer which will be processed at the hidden layer [21] [22]. The hidden layer in this RNN model will do a feedback loop based on time delay d which will later be determined through trial and error. The number of hidden layers is also determined through trial and error. After calculation, the data was transformed into the previous amount by denormalization [23] [24].

- After getting the forecasting results based on each method, the results will be compared with the actual value. Measuring instruments that can be used to compare the level of accuracy are:
 - Nilai Mean Square Error (MSE)

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

Figure 3. Mean Square Error (MSE)

- Nilai Mean Root Square Error (MRSE)

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$$

Figure 4. Mean Root Square Error (MRSE)

- If the results have been evaluated, then the method can be determined which is considered more appropriate and also how the application can be used as a proposal.

3. RESULTS AND DISCUSSION

- **ARIMA**

In ARIMA NAV data are used from July 10 to August 8 2019. In this method, the data used is only the NAV itself. The data is processed using SPSS 25 tools, using Expert Modeler, then the model that can be considered the best can be determined. The model obtained is ARIMA (1,1,2) whose art has a value of $p = 1$, a value of $d = 1$, and a value of $q = 2$.

Table 2. Model Description

Model ID	NAV(T)	Model_1	ARIMA(1,1,2)
----------	--------	---------	--------------

Following are the results of the analysis based on a predetermined model, the RMSE value is obtained 15. The RMSE value of 15 indicates the average magnitude of forecasting error, which will later be compared with the results obtained from the RNN model to evaluate which approach performs better. The Stationary R-squared value of 0.070 suggests that the model explains only a small portion of the variance in the differenced (stationary) series. Meanwhile, the R-squared value of 0.735 indicates that the model is able to explain approximately 73.5% of the variance in the observed data, reflecting a reasonably good fit. The RMSE value will later be compared with the results of forecasting using RNN

- **RNN**

The data used in the RNN are from 3 January 2017 to 30 August 2019, so that 638 data will be obtained which will be divided into 3 parts. The first stage is 70% (446 data) for training, 10% (64 data) for validation, and 20% (128 data) for testing. Previous data through the data normalization process first.

There are 6 inputs which are stock indexes on the previous day (T-1), consisting of: IHSG (T-1), Mining Index (T-1), Consumer Index (T-1), Finance Index (T-1), Trade Index (T-1), and LQ45 Index (T-1). For the output produced is the value of NAV on that day, which is NAV (T).

What needs to be found in the formula is the time delay by trial and error. To determine the architecture, it needs to be tested several times to trial and error get the smallest MSE value based on the number of hidden nodes and time delay. In this study three experiments were conducted for each combination of the number of hidden nodes 6 to 10 and time delays 1 and 2. There are 3 stages that will be recorded by the MSE, namely training, validation, and testing. Because in all experiments conducted the MSE value in the validation process is too small (almost 0), therefore only MSE is recorded for training and testing. Here are the results of the experiment:

Table 3. MSE

Hidden		1st Trial		2nd Trial		3rd Trial	
Node	Delay	Training	Testing	Training	Testing	Training	Testing
6	1	4.0200	5.0643	4.1923	5.2623	4.1700	5.0027
7	1	4.1506	5.1502	4.1306	5.0123	4.1743	4.9530
8	1	3.9345	5.4383	4.1797	4.5339	4.0173	4.4000
9	1	4.2762	4.6627	4.1912	4.7716	4.0128	5.7300
10	1	4.0681	5.3529	4.1599	4.3242	4.1471	5.0685
6	2	4.1849	4.6000	3.8918	5.1961	4.2199	4.0751
7	2	4.0289	2.7000	3.9246	5.1785	3.9753	4.8200
8	2	3.5109	5.3724	4.3501	4.7035	3.6456	5.0539
9	2	3.9102	5.4598	4.2618	4.1890	4.1002	4.9078
10	2	2.9312	7.4293	4.1917	4.3826	4.0675	5.6713

Obtained the best MSE value is the number of hidden nodes 6 and time delay 2. Because the architecture obtained from RNN is as follows:

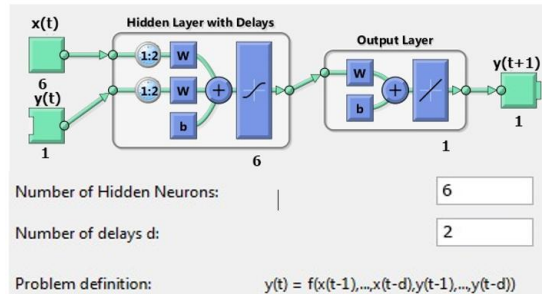


Figure 5. Research Model

After finding the right architecture, the results of testing will be returned by denormalization. The amount of testing data to be used is 59 data to adjust to ARIMA, also based on the same date so that it can be compared directly. MSE calculations are also adjusted to the data used for comparison. The MSE obtained was 201,785, so the RMSE was 14.21

• Comparison

If the visualization for evaluation is made into a graph, the comparison is:

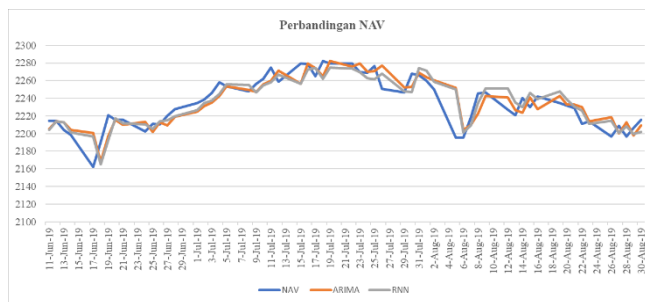


Figure 6. Comparison

To compare which method is better, judging from the value of Mean Square Error (MSE) and Root Mean Square Error (RMSE), a smaller MSE or RMSE value indicates a method that is considered better. Here is the comparison:

Table 4. Error Comparison

Method	MSE	RMSE
ARIMA	255	15

Method	MSE	RMSE
RNN	201.785	14.21

Based on the results of the Reliability Test above, it is known that each construct are reliable because it has a value greater than 0.7.

4. CONCLUSION

Based on the results of this study, both ARIMA and RNN methods are capable of forecasting time series data using historical inputs, although they apply different approaches. ARIMA relies on recognizing historical patterns and requires minimal data, using only NAV values without additional variables. In contrast, RNN, which mimics biological neural networks, incorporates various input variables, such as stock indices alongside NAV to predict future values, thus requiring more complex and extensive datasets. Using NAV data from the Sucorinvest Equity Fund for the 2017–2019 period, the study found that RNN outperformed ARIMA in terms of forecasting accuracy, as indicated by lower MSE and RMSE values.

This suggests that RNN's ability to process richer and more diverse data contributes to its higher precision. The implementation of RNN-based forecasting can be beneficial for both securities companies and investors, particularly through digital platforms that offer investment recommendations. Moreover, this approach can be applied to other equity mutual funds, and its accuracy may be further enhanced by incorporating more historical data and additional variables such as Asset Under Management (AUM), economic trends, and other relevant factors.

REFERENCES

- [1] A. A. Adebisi, A. Adewumi, and C. Ayo, "Stock Price Prediction Using the ARIMA Model," in *Proc. Int. Conf. on Computer Modelling and Simulation*, 2014.
- [2] R. Adhikari and R. Agrawal, *An Introductory Study on Time Series Modeling and Forecasting*. Lambert Academic Publishing, 2013.
- [3] I. A. Arief, "Wow! Dana kelolaan Reksa Dana Tembus Rp 500 T," *Indonesia*, Dec. 14, 2018.

- [4] D. Banerjee, "Forecasting of Indian Stock Market using Time-series ARIMA Model," in *Proc. Int. Conf. on Business Information Management (ICBIM)*, 2014.
- [5] Bareksa, "Apa dan Bagaimana Menghitung NAB Per Unit Reksa Dana?" *Bareksa*, Jul. 13, 2016.
- [6] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, "Recurrent Neural Networks for Short-Term Load Forecasting: An Overview and Comparative Analysis," *Springer*, Norway, 2017.
- [7] F. M. Bianchi, E. Maiorino, M. Kampffmeyer, A. Rizzi, and R. Jenssen, "Recurrent Neural Networks for Short-Term Load Forecasting: An Overview and Comparative Analysis," *Springer*, 2017.
- [8] O. A. Bullinaria, "Recurrent Neural Network," 2005.
- [9] Bursa Efek Indonesia, *Buku Panduan Indeks Harga Saham Bursa Efek Indonesia*, Jakarta, 2010.
- [10] T. D. Chaudhuri and I. Ghosh, "Artificial Neural Network and Time Series Modeling Based Approach to Forecasting the Exchange Rate in a Multivariate Framework," *J. of Insurance and Financial Management*, pp. 92–123, 2016.
- [11] J. Guo, "BackPropagation Through Time," 2013.
- [12] S. Haykin, *Neural Networks and Learning Machines*, 3rd ed. Pearson, 2009.
- [13] R. D. Hastuti, A. Prasetyo, and T. Wijaya, "The Effect of Fintech Application Use on Investment Behavior," *Jurnal Ekonomi dan Keuangan*, vol. 24, no. 3, pp. 135–142, 2020.
- [14] N. U. Hidayatullah and I. MT, "Estimasi Radiasi Matahari Perjam pada Permukaan Horizontal dengan Extreme Learning Machine (Studi Kasus di Surabaya)," *Jurnal Teknik POMITS*, vol. 2, 2014.
- [15] P. Hushani, "Using Autoregressive Modelling and Machine Learning for Stock Market Prediction and Trading," in *Proc. 3rd Int. Congr. on Information and Communication Technology, Advances in Intelligent Systems and Computing*, pp. 767–774, 2018.
- [16] L. Hussain, "Forecasting Time Series Stock Data using Deep Learning Technique in a Distributed Computing Environment," in *Int. Conf. on Computing, Power and Communication Technologies (GUCON)*, 2018.

- [18] Kumparan, "Tren Investasi 2018 di Indonesia Menunjukkan Peningkatan," *Sumatra Barat, Indonesia*, Jul. 19, 2018.
- [19] Mandiri Investasi, *Jenis-jenis Reksa Dana*, Indonesia, 2018.
- [20] MathWorks, *Neural Network Toolbox User's Guide*, MathWorks, 2002.
- [21] OJK, *Seri Literasi Keuangan Tingkat Perguruan Tinggi*, Jakarta, 2016.
- [22] A. Pankratz, *Forecasting with Univariate Box-Jenkins Models: Concepts and Cases*, John Wiley & Sons, Inc., 1983.
- [23] R. A. Putra, R. Hidayat, and A. Alamsyah, "Stock Price Forecasting Using LSTM: A Case Study of IDX Stocks," *Procedia Computer Science*, vol. 179, pp. 699–706, 2021.
- [24] R. Wirth and J. Hipp, "CRISP-DM: Towards a Standard Process Model for Data Mining," *ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA*, 2020.