Journal of Theoretical and Applied Information Technology
31st January 2021. Vol.99. No 2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EVALUATING THE IMPACT OF MULTIPLE LEADING INDICATOR IN FORECASTING NEXT DAY STOCK PRICE WITH LSTM

[1]ADITYO HIDAYAT, [2]AGUS SIHABUDDIN
[1,2] Department of Computer Science and Electronics, Faculty of Mathematics and Natural Science, Universitas Gadjah Mada, Yogyakarta, Indonesia
Corresponding Author E-mail: a_sihabudin@ugm.ac.id

ABSTRACT

LSTM was the preferred choice for financial time series forecasting, whereas fundamental analysis and technical analysis were among the most favorable feature sets. Earlier studies had several suggestions to improve forecasting performance, by using features known to carry information about the future price movement, and features associated with substantial price movements: the foreign investors' trading volume. Overall trading volume and those volumes from foreign investors have been introduced as a leading indicator. However, empirical literatures which examines price-volume relationship using LSTM had not used foreign investors' trading volume. This study evaluates the use of multiple leading indicators as input, and optimum hyperparameters configurations using LSTM to next day prediction performance. Experiments are evaluated based on 88 stocks in Indonesia stock market, ranging from Jan. 2, 2015 to Dec. 30, 2019. Financial time series forecasting using simple LSTM architecture performs as good as baseline performance with the advantage of fewer computing requirements. Optimum hyperparameters are a single hidden layer, 50 nodes, and ten days of the input window. The highest winning stocks are achieved using feature sets consisting of a lagging indicator (price) and multiple leading indicators (overall trading volume and foreign investors' trading volume). The findings indicate that multiple leading indicators contain predictability factors which can be further explored to improve financial time series forecasting. This study contributes the use of foreign investors' which improves financial time series forecasting with LSTM.

Keywords: *Financial Time Series Forecasting, Foreign Trading Volume, Stock Market, LSTM*

1. INTRODUCTION

Studies on financial time series forecasting had attracted attention to its significance in the financial market industry. However, forecasting financial time series is challenging due to the inherently noisy and non-stationary nature of stock prices [1]. Dynamic changes in the relationship between independent and dependent variables frequently happen in financial time series. Therefore traditional statistical methods are not effectively applied to financial context [2]. Currently, performing technical forecasting based on a machine learning method is considered one of the most effective solutions to the challenge [3].

Most of the studies conducted in 2015 – 2019 on financial time series forecasting with Deep Learning uses one or more feature sets, such as historical stock prices (open, high, low, close, volume), technical analysis, fundamental analysis, macro-economic (unemployment rate, GDP, etc.) and text mining from financial news [4]. Studies in the field of economics and finance have shown a strong relationship between stock price and trading volume. Past information on trading volume helps predict the behavior of the stock price [5]. Trading volume has been introduced as a leading indicator, where trading volume dynamics and price impact subsequent corporate decisions [6]. The trading volume of foreign institutional investors has been known to transmit valuable signals, since the increase of foreign institutions at the previous day, would boost return further [7]. Earlier studies on financial time series forecasting with Deep Learning have not investigated the use of trading volume from foreign investors.

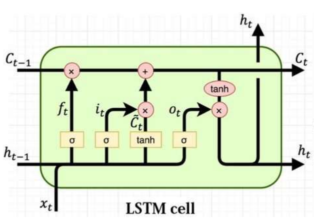
This paper aims to contribute to the empirical literature by evaluating the Indonesia stock market's price-volume relationship. This study evaluates the use of multiple leading indicators as input, and optimum hyperparameters configurations (feature set, input window sizes, number of layers and nodes) using LSTM to achieve a better and consistent performance of next day prediction. Leading indicators are our subject of interest in this study, they are: overall trading volume and foreign investors' trading volume. Other indicators such as fundamental, technical analysis, or text mining, have not regarded as leading indicators, therefore did not covered in this study.

The remainder of this paper is organized as follows. The next section reviews the related work. Section 3 introduces the methodology. Section 4 discusses the results. Section 5 presents the conclusions.

2. RELATED WORK

2.1 Financial Time Series Forecasting with LSTM

RNN-based models (particularly LSTM) has been the preferred choice for financial time series forecasting. Their applications include Natural Language Processing (NLP), language modeling, language translation, speech recognition, sentiment analysis, predictive analysis, and financial time series analysis [4]. LSTM network prevents the loss of essential features and whole sequences using long-term memory while retaining short-term memory (as with simple recurrent neural networks) [8].



$$i_t = \sigma(x_t U^i + h_{t-1} W^i)$$
$$f_t = \sigma(x_t U^f + h_{t-1} W^f)$$
$$o_t = \sigma(x_t U^o + h_{t-1} W^o)$$
$$\tilde{C}_t = \tanh(x_t U^g + h_{t-1} W^g)$$
$$C_t = \sigma(f_t * C_{t-1} + i_t * \tilde{C}_t)$$
$$h_t = \tanh(C_t) * o_t$$

Figure 1: Structure of the LSTM cell and equations [9]

LSTM networks are composed of an input layer, one or more hidden layers, and an output layer. The number of neurons in the input layer is equal to the number of explanatory variables (feature sets). The number of neurons in the output layer reflects the output space (next day prediction). The main characteristic of LSTM networks is contained in the hidden layer(s) consisting of so-called memory cells [10].

There are extra gates in an LSTM cell, namely the input, forget, and output gate used to decide which signals will be forwarded to another node. W is the recurrent connection between the previously hidden layer and the current hidden layer. U is the weight matrix that connects the inputs to the hidden layer. $\tilde{C}$ is a candidate hidden state that is computed based on the current input and the previous hidden state. C is the internal memory of the unit, which is a combination of the last memory, multiplied by the forget gate, and the newly computed hidden state, multiplied by the input gate. The equations that describe the behavior of all gates in the LSTM cell are illustrated in Figure 1.

Sezer et al. [4] have highlighted several studies that use LSTM to forecast stock price and market indices, particularly in one step prediction, with variations in input variables, preprocessing methods and architectures. Most of the studies use daily stock trading data (OHLC) [11][12][13][14][15][16][17][18][19] and trading volume [14][16][17] as raw input. In contrast, others are combined with technical analysis (TA) indicators (such as MACD, EMA, MA5, MA20, etc.), fundamental analysis (FA) [20] and macroeconomic (the exchange rate and interest rate) [14][19]. Most of the studies use normalization [15][16][17][12][18][13][19], and feature extraction (principal component analysis, autoencoders, and restricted Boltzmann machine) [15] as preprocessing methods. Some of the clearly stated LSTM architectures are: one hidden layer [17], two hidden layers [20], three hidden layers [16][11], four hidden layers [19]. Summary of recent studies as in Table 1.

Table 1: Summary of Recent Studies

Ref	Feature Set						LSTM
	Price	OHLC	Vol	TA	FA	Econ	Layers
[11]		✔					2
[12]	✔						N/A
[13]		✔					1-3
[14]		✔	✔	✔		✔	N/A
[15]	✔						N/A
[16]		✔	✔				3
[17]		✔	✔				1
[18]	✔						N/A
[19]	✔			✔			4
[20]	✔				✔		2

Vol = Overall trading volume; Econ = Exchange rate and interest rate

Several suggestions are arising from previous studies, which will further improve forecasting performance. Choosing appropriate features has been a very crucial factor in any prediction model [19]. Exploiting factors with strong evidence of predictability may likely give better performance than simply dumping a large raw dataset. Additional factors that are known to carry information about the future price movement, such as trading volume and the price of a derivative linked to the stock [15]. Directly including traditional technical indicators and oscillators into the prediction model is not useful for predicting the market trend one day ahead [21]. These suggestions imply opportunities for further elaboration on defining types of trading volume and its derivatives.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.2 Trading Volume

Earlier studies on economics and finance had shown a strong relationship between stock price and trading volume. One notable study by Gervais et al. in 2001 [22], which has been highly cited, showed that individual stocks experiencing extreme trading volume contains important information about subsequent stock returns. Periods of extremely high volume tend to be followed by positive excess returns, whereas periods of extremely low volume tend to be followed by negative excess returns. This characteristic of volume and return relationship is defined as a high-volume return premium. Decades later, Kaniel et al. in 2012 [23] discover that high-volume return premium is a global phenomenon found across 41 countries, significantly present in nearly all developed markets and several emerging markets. Most recent findings by Kaniel et al. in 2017 [6] showed a positive association between abnormally high volume and subsequent corporate investment. This association is not subsumed by information disclosure related to earnings announcements two weeks following extreme volume shocks. Positive shocks to trading volume lead to a reduction in the cost of capital and a concomitant price appreciation. In this study, trading volume has been introduced as a leading indicator.

Past information on trading volume helps predict the behavior of the stock price [5]. A significant part of the trading volume in financial markets is attributed to institutional investors, those to be professional asset managers which manages a portfolio for a mutual fund, a hedge fund, a pension fund, an endowment, an asset management team in a bank, or insurance company, etc. [24]. Price contributions generated by large orders of professional institutions are closely related to future stock performance [25]. Stocks with high foreign ownership outperform stocks with low foreign ownership [26]. The trading volume of foreign institutional investors has transmitted valuable signals since foreign institutions' increase on the previous day would boost return further [7]. There are substantial price movements associated with foreign investors' trade, where impacts are larger in emerging markets [27].

Earlier studies on financial time series forecasting with Deep Learning, particularly those who use LSTM, have not investigated the use of trading volume from foreign investors, as summarized in Table 1. Therefore, the trading volume of foreign investors is our subject of interest in this study.

3. METHODOLOGY

3.1 Dataset

We evaluate the forecasting performance based on the Indonesia stock market, ranging from Jan. 2, 2015 to Dec. 30, 2019. The chosen period includes bull and bear periods, reflecting the impact of significant world economic crisis events at the moment (the crisis of Greece and Turkey) in Indonesia, which is an important feature to test the robustness of the prediction model. **Error! Reference source not found.** clearly shows the high volatility in the Indonesia stock market during the period analyzed.

Stocks are selected conforming to several criteria: 1) Should be a constituent of the KOMPAS100 index, which contains 100 actively traded stocks with high liquidity and high market capitalizations; 2) Transaction data available during the selected period. These criteria were resulting in 88 stocks as our dataset.

Each trading day contains four feature sets with one lagging indicator (price), and three others are leading indicators, as described in Table 2.

Table 2: Feature Sets

#	Feature	Description
F_1	Price	Daily closing stock price
F_2	Volume	Number of shares traded daily
F_3	Foreign Volume	Daily net trading volume (in shares) of foreign investors [28]
F_4	Foreign Volume (accumulated)	Daily cumulative sum of F_3

We obtained all required data from the Indonesia Stock Exchange (IDX), available on the Stock Summary page. Price (F_1) is the closing price of stock for each share, represented in the currency amount (Rupiah). Volume (F_2) is expressed in the number of shares traded daily.

IDX provides daily data of Foreign Buy and Foreign Sell of every stock, represented in the number of shares. Foreign volume (F_3) is a daily net trading volume obtained by subtracting foreign Buy's value with Foreign Sell. The value of F_3 can be positive or negative, can also be repeated at any different point of time t. We can have a continuous pattern with similar characteristics as price by cumulatively summing this value each day. The value of time t is the sum of foreign volume at time t with the cumulative sum from $t-1$.

Dataset consists of 1.213 trading days (records) for each stock and then split with 80% as a training

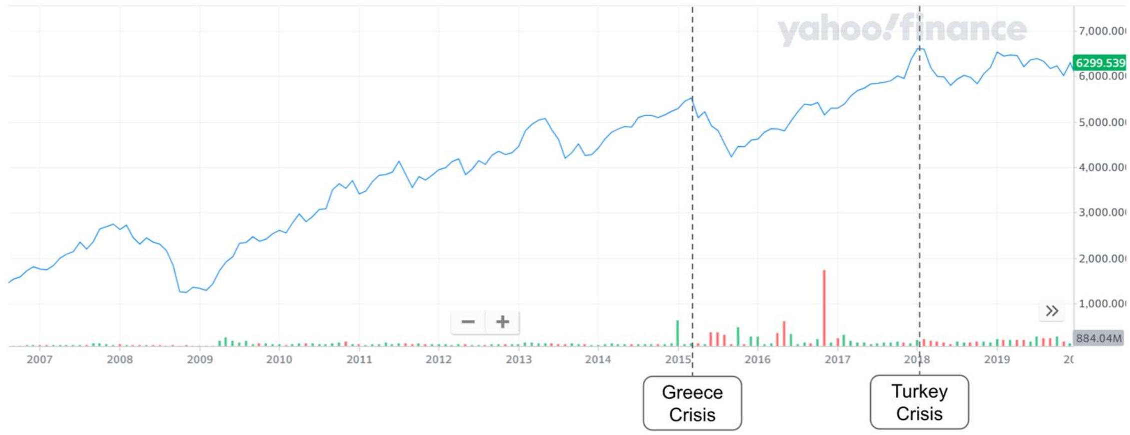


Figure 2 Dataset Period

set and 20% as a test set. The reason for choosing this 80:20 rule is twofold: increasing training data and representing both world economic crisis events in the training data. The training set starts from Jan. 2, 2015 to Jan. 2, 2019, and the test set starts from Jan. 3, 2019 to Dec. 30, 2019.

We are using the whole dataset (consisting of 88 stocks) as a unified source of the dataset; prior splitting is done. Therefore, each of the feature set from all stocks needs to be normalized into a unified range value. The dataset would have a total of 106.744 records, and after the split, it would generate 74.624 records as the training set and 32.120 records as the test set. Normalization ensures that the larger value input attributes do not overwhelm smaller value inputs, decreasing prediction errors [28].

Normalization transforms the dataset as a representation of overall market behavior, and the neural network model will be generated based on all stocks' behavior. Then, testing the prediction accuracy can be done from any stocks in the test set. Sequences of input (from any stocks) can be fed into the model to generate predictions.

3.2 Pre-processing

Each value in every feature set is normalized by scaling into the specified range of $[min, max]$, using the formula:

$$X_{const} = \left(\frac{X - X_{min}}{X_{max} - X_{min}}\right) \tag{1}$$

$$X_{scaled} = X_{const} * (max - min) + min \tag{2}$$

where X_{scaled} is the resulting new value within the range, X is the original value, X_{min} is the minimum value in the feature set, X_{max} is the maximum value in the feature set.

Each feature set is assigned its range. Range [0, 1] is set for features which are an independent variable (price) and for features with an accumulating value throughout time (F_3 and F_5). Range [0, 0.5] is set for features that represent daily values. Range assignments specified as in Table 3.

Table 3: Normalization Range

#	Feature	Range
F_1	Price	[0, 1]
F_2	Volume	[0, 0.5]
F_3	Foreign Volume	[0, 0.5]
F_4	Foreign Volume (accumulated)	[0, 1]

Training set then split into multiple samples of observations, which will be based on learning from the model. The model would generate a one-step prediction; therefore, each sample would have an input consisting of several prior observations (multiple features), and an output which is the actual next day price. Training set transformed into multiple samples using a sliding window approach, as in Figure 3.

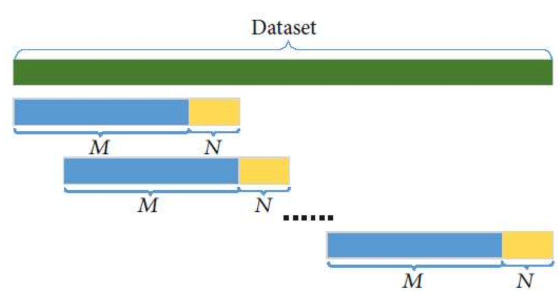


Figure 3: Multiple Samples Using Sliding Window [29]

Related studies mentioned by Shynkevich et al. [30], the optimum prior observations found for the short-term stock prediction is equal to three days,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and that for the long-term stock predictions is found to be 20 days. One of the aims of this study is to determine how far back in the past does each observation (across feature sets) has a better impact on next day prediction. Therefore we would evaluate several the optimum number of prior observations (or input window). Let M is the number of input window, and N is the next day prediction, where $M \in \{3, 5, 10\}$ and $N = 1$.

3.3 Network Architecture

We build LSTM architectures for a one-step prediction of normalized stock prices. In each experiment, we execute trials to iterate over hyperparameters to evaluate which architecture gives the best prediction performance. Evaluated hyperparameters are the number of hidden layers (1, 3, 5, 7, and 10) and the nodes on each layer (50, 100, 200). Network architecture consists of the input layer, multiple hidden layers, a dense layer, and lastly, the output layer. Other hyperparameters have been specified with 500 epochs, batch size 1024, and an Adam optimizer.

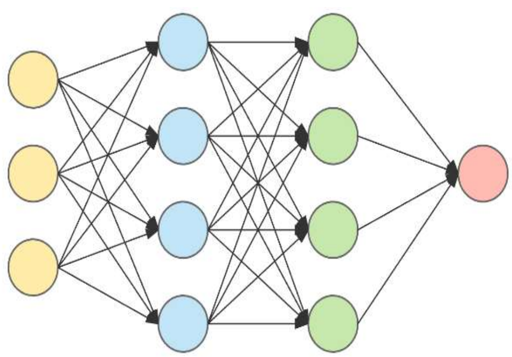


Figure 4: Network Architecture

3.4 Evaluation Metrics

We evaluate prediction performance using three measures: 1) Root Mean Squared Error (RMSE); 2) Hit Ratio; and 3) Winning Stocks. Choosing these measures has its own purpose, depending on the expected objective. RMSE is used to evaluate forecasting performance in terms of accuracy, when concerning which predictions closest to the actual values. For stock trading purposes, measuring the correct direction of prediction is more relevant than accuracy, which is measured with Hit Ratio and Winning Stocks.

We use RMSE because the predicted price variable is normalized, to facilitate uniform comparison of predicted prices across all stocks in the test set. RMSE is defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(y_{t+1}^{n} - \hat{y}_{t+1}^{n})^2} \tag{3}$$

where N is the total days in the test set, y_{t+1} is the next day actual price and $\hat{y}_{t+1}$ is the next day predicted price in the test set, respectively.

Hit Ratio denoted the percentage of trials when the predicted direction was correct. Hit Ratio measures the rate of accuracy relating to the series trend. A high Hit Ratio promises more winning trades. Hit Ratio is defined as:

$$Hit\ Ratio = \frac{1}{N}\sum_{n=1}^{N} P_i(i = 1, 2, \dots, N) \tag{4}$$

where P_i is the prediction result for the i^{th} day, defined as:

$$P_i = \begin{cases} 1, & if\ y_{t+1} \cdot \hat{y}_{t+1} > 0, \\ 0, & otherwise, \end{cases} \tag{5}$$

Predictions generated by the model is validated against all stocks in the test set. We are mainly interested in a model capable of generating high Winning Stocks, where each stock's prediction in the test set results in a Hit Ratio above 60%. Consistency of prediction performance is represented by Winning Stocks, which defined as:

$$Winning\ Stocks = \sum_{n=1}^{n} H_i(i = 1, 2, \dots, n) \tag{6}$$

where n is the number of stocks in the test set, and H_i is the Hit Ratio for the n^{th} stock, defined as:

$$H_i = \begin{cases} 1, & if\ Hit\ Ratio > 60\%, \\ 0, & otherwise, \end{cases} \tag{7}$$

3.5 Experiment Design

We design experiments with different feature sets to determine whether adding more leading indicators as input would impact in better prediction performance. Feature sets are assigned to experiments, as in Table 4.

Table 4: Experiment Design

#	F_1 Price	F_2 Volume	F_3 Foreign Volume	F_4 Foreign Volume (Accum)
E1	✔			
E2	✔	✔		
E3	✔		✔	

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

E4	✔		✔	✔
E5	✔	✔	✔	
E6	✔			✔
E7	✔	✔	✔	✔

The first experiment (E1) is univariate forecasting, which will be our baseline performance reference. Trials will be executed on each experiment, iterating to varying hyperparameters and input window M. Each trial is tested against 88 stocks in the test set; therefore, each trial will have an average RMSE, average Hit Ratio, and how much Winning Stocks. Since we are interested only in stocks with high Winning Stocks, we calculate average RMSE and average Hit Ratio based on those stocks capable of achieving a Hit Ratio above 60%. Due to the stochastic nature of neural networks, trials with high Winning Stocks are repeated at least three times to confirm consistent and reproducible results.

Each experiment will also have an average RMSE and average Winning Stocks, calculated based on trials with the best top-three result. An overall evaluation would be compared among experiment results.

4. RESULTS

Each experiment is conducted at a minimum of 12 trials by combining layers, input, and nodes. Trials achieving high Winning Stocks (comparable to baseline performance) are repeated at least three times. Therefore each experiment could have up to 36 trials. Experiment results are presented as in Table 5, showing only the best three trials.

Table 5: Experiment Results

Exp. #	l	M	n	Hit Ratio %	RMSE	WS
E1	3	3	50	63.42	0.0159	43
E1	3	3	50	63.45	0.0161	42
E1	3	3	50	63.44	0.0161	42
E1 Average: Input 3d				**63.44**	**0.0160**	**42.33**
E1	3	10	50	63.4	0.0159	43
E1	3	10	50	63.63	0.0156	43
E1	3	10	50	63.68	0.0158	43
E1 Average: Input 10d				**63.57**	**0.0158**	**43.00**
E2	1	3	50	63.51	0.0100	40
E2	1	3	50	63.48	0.0162	41
E2	1	3	50	63.45	0.0161	42
E2 Average				**63.48**	**0.0141**	**41.00**
E3	3	3	50	63.56	0.0162	41
E3	3	3	50	63.45	0.0161	42
E3	3	3	50	63.45	0.0161	42
E3 Average				**63.49**	**0.0161**	**41.67**
E4	1	3	100	63.46	0.0161	42
E4	1	3	100	63.32	0.0163	39
E4	1	3	100	63.37	0.0172	38
E4 Average				**63.38**	**0.0166**	**39.67**
E5	1	3	100	63.43	0.0161	42
E5	1	3	100	63.71	0.0167	40
E5	1	3	100	63.45	0.0162	42
E5 Average: Input 3d				**63.53**	**0.0163**	**41.33**
E5	1	10	50	63.69	0.0170	40
E5	1	10	50	63.62	0.0158	43
E5	1	10	50	63.62	0.0158	43
E5 Average: Input 10d				**63.64**	**0.0162**	**42.00**
E6	1	5	100	63.51	0.0162	42
E6	1	5	100	63.30	0.0118	5
E6	1	5	100	63.66	0.0139	24
E6 Average				**63.49**	**0.0140**	**23.67**
E7	3	3	50	63.97	0.0138	33
E7	3	3	50	62.89	0.0157	6
E7	3	3	50	62.81	0.0151	23
E7 Average				**63.22**	**0.0149**	**20.67**

l = layers; M = input; n = nodes; WS = Winning Ratio

Overall experiment results are shown in Figure 5, with summaries for each experiment, as in Table 6.

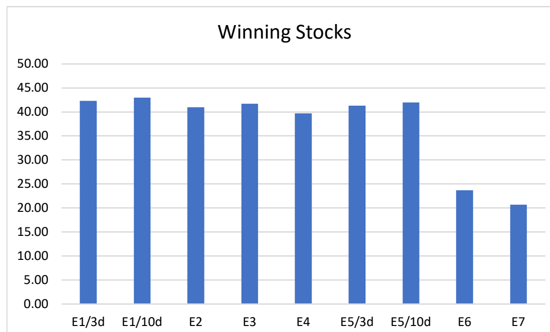


Figure 5: Plot of Experiment Results

Table 6: Summary of Experiment Results

Exp #	l	M	n	Hit Ratio %	RMSE	WS	Improvement (%)	
							RMSE	**WS**
E1	3	3	50	63.44	0.0160	42.3	0	0
E1	3	10	50	63.57	0.0158	43.0	1.63	1.57
E2	1	3	50	63.48	0.0141	41.0	12.29	-3.15
E3	3	3	50	63.49	0.0161	41.6	-0.72	-1.57
E4	1	3	100	63.38	0.0166	39.6	-3.16	-6.30
E5	1	3	100	63.53	0.0163	41.3	-1.76	-2.36
E5	1	10	50	63.64	0.0162	42.0	-0.89	-0.79
E6	1	5	100	63.49	0.0140	23.6	12.76	-44.09
E7	3	3	50	63.22	0.0149	20.6	8.35	-51.18

l = layers; M = input; n = nodes; WS = Winning Ratio

E1 generates consistent forecasting performance under three hidden layers with 50 nodes, either with 3 or 10 days of input. The use of ten days as input resulting in better RMSE and Winning Stocks than three days. Neural network hyperparameters are considered skillful if capable of achieving Winning Stocks higher than 42.33 or 43, which will be our

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

baseline performance. Figure 6 shows one of the learning curves with univariate forecasting from E1. Figure 7 shows the prediction plot from one of the stocks in the test set in E1, which achieves the best Hit Ratio 71.87%.

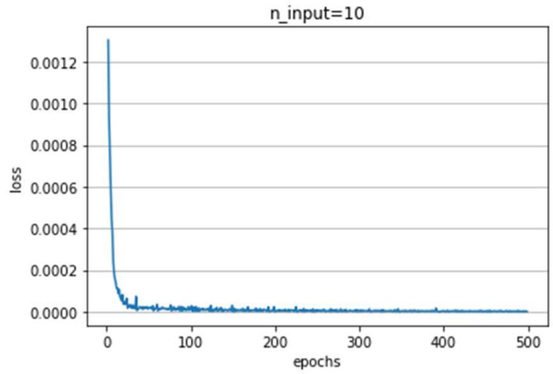


Figure 6: Loss Plot from E1

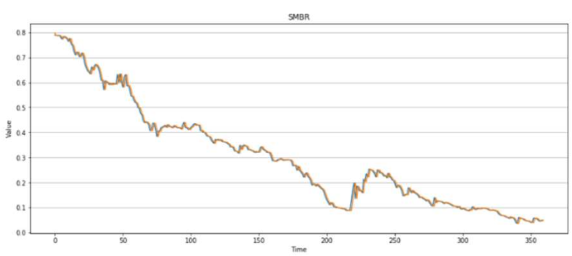


Figure 7: Prediction Plot from E1

The experiments aim to identify trials with consistent results. Throughout the trials, experiments E6 and E7 ultimately failed in generating consistent results. E6 had the best RMSE improvement, but achieved worst in Winning Stocks, only 23.67 stocks from the whole 88 stocks in the test set. E7 was the worst among other experiments. We will further discuss more on the findings from each experiment.

E2 incorporates trading volume (F_2) along with the price (F_1) as feature set, resulting in a comparable Winning Stocks as E1, noticeably substantial improvement in RMSE (12.29%) with less computation requirement. E2 uses a single hidden layer compared to E1 with three hidden layers. This result is aligned with the findings by Israeli et al. [6]. It also confirms the suggestion by Eunsuk et al. [15] that trading volume is one factor with strong predictability of future price movement. Figure 8 shows one of the learning curves from E2.

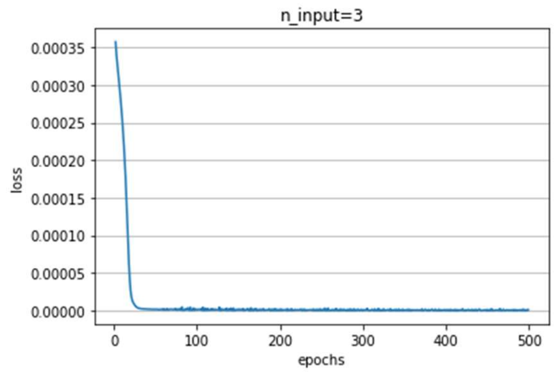


Figure 8: Loss Plot from E2

E3 focuses on trading volume by foreign investors (F_3) combined with the price (F_1) as feature set, resulting in a comparable Winning Stocks and RMSE, using three hidden layers and three days of input. This result amplifies the previous work by Richards [27] and Huang et al. [7], where foreign volume is strongly associated with substantial price movements. We could imply that foreign volume has similar predictability factors as overall trading volume. Figure 9 shows one of the learning curves from E3.

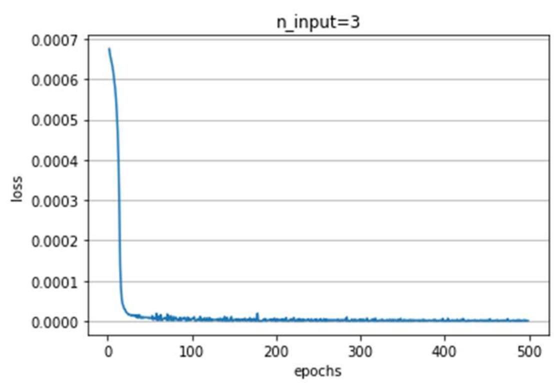


Figure 9: Loss Plot from E3

E6 attempts the neural network to learn the relationship of two feature sets with similar non-stationary characteristics: price (F_1) and daily cumulative sum of foreign volume (F_4). Surprisingly, among extensive repetitive trials in this experiment, we found no hyperparameters capable of generating consistent results beyond baseline performance E1. However, this experiment resulting in the most considerable RMSE improvement (12.76%) compared to the benchmark. It appears that 1) F_4 is the most suitable feature in terms of lowest RMSE, capable of producing predictions that are closest to the actual values; 2) In terms of Hit Ratio, F_4 does not have the same predictability factors as trading volume (F_3). Figure 10 shows one of the learning curves from E6.

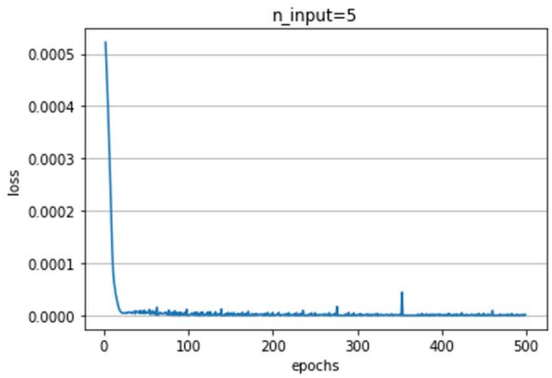


Figure 10: Loss Plot from E6

E4 attempts to evaluate whether incorporating factors related to foreign volume (F_3) and its derivatives (daily cumulative sum, F_4) can improve forecasting performance. We would typically expect forecasting performance to be better than E3, which uses only foreign volume (F_3) as input. Unfortunately, although a consistent result is achieved, the results are not better than E3. This result indicated that the use of features with less predictability factors would significantly affect neural networks' learning. Figure 11 shows one of the learning curves from E4.

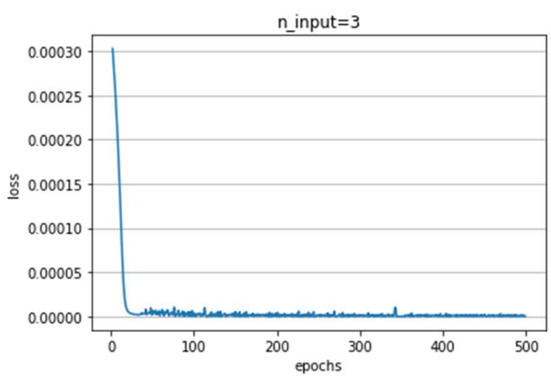


Figure 11: Loss Plot from E4

E7 is resulting in the worst performance among all experiments because of the same reason as E4. Incorporating all features (F_1, F_2, F_3, F_4) as input, including those with less predictability factor, did not contribute to a better prediction. This result agrees with Eunsuk et al. [15], where dumping a large raw dataset may not likely give better performance. Niu et al. [3] and Khare et al. [19] had underlined the necessity of performing feature selection for multivariate financial time series forecasting. This finding implies the importance of feature selection, where features need to be carefully selected to ensure their positive contribution in improving forecasting. Figure 12 shows one of the learning curves from E7.

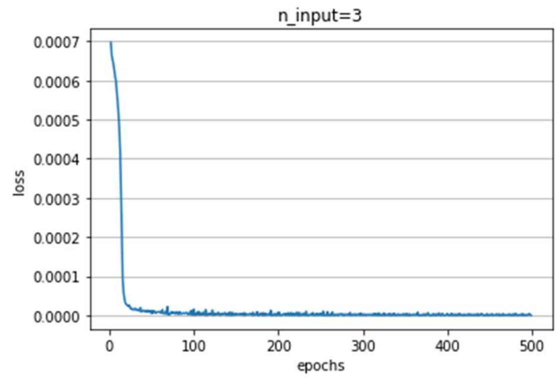


Figure 12: Loss Plot from E7

Lastly, E5 evaluates features that previously had been identified as factors with strong predictability factors: trading volume (F_3) and foreign volume (F_4). The result of E5 is comparable to E1, with the advantage of fewer computing requirements. E5 uses a single hidden layer and 50 nodes to achieve a similar result as E1 with ten days of input. Figure 13 shows one of the learning curves with multivariate forecasting from E5. Figure 14 shows the prediction plot from one of the stocks in the test set in E5, which achieves the best Hit Ratio 71.87%.

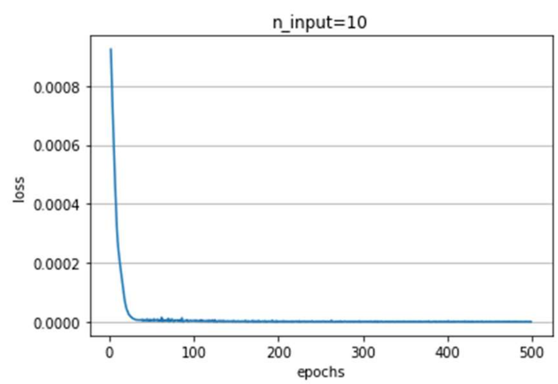


Figure 13: Loss Plot from E5

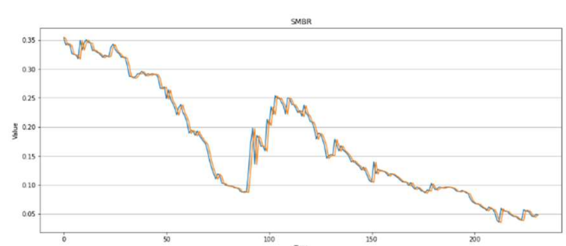


Figure 14: Prediction Plot from E5

The experiment shows that neural networks have their optimum hyperparameters depending on the dataset's characteristics and the intended metrics to be achieved. Implementing the most complex architectures with more hidden layers and more nodes did not contribute to better predictions. Consider an example from E2, one of the optimum hyperparameters with consistent results are: single hidden layer, 50 nodes and three days of input. These hyperparameters produce learning curves as in Figure 8. Changing the hyperparameters to 200 nodes with five days of input dramatically reduces the learning effectiveness, resulting in very noisy curves as in Figure 15. Figure 16 shows the stock's prediction plot in the test set from E5, with lowest Hit Ratio 47.88%.

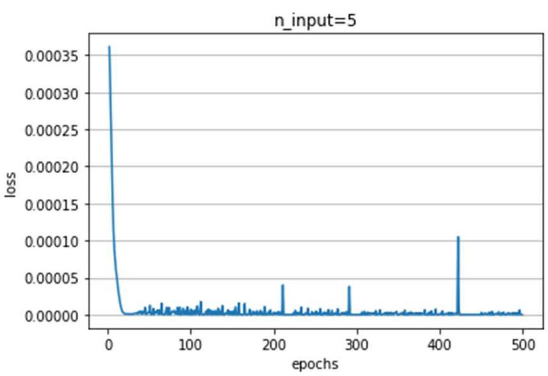


Figure 15: Noisy Loss Plot from E2

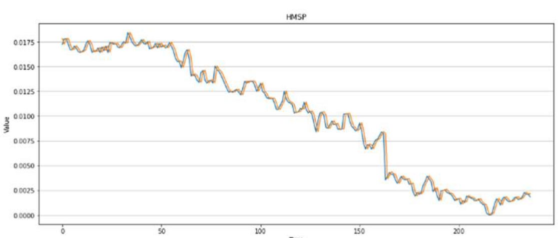


Figure 16: Prediction Plot with Lowest Hit Ratio from E2

So far, we had discussed the number of achieved Winning Stocks from the perspective of each experiment, as shown in Table 6, which would lead to conclude that E5 with ten days of input had the most optimum hyperparameters configuration. We observe each stock's Hit Ratio from all three trials, as presented earlier in Table 5, particularly from E1 and E5. We measure the standard deviation of each stock's Hit Ratio from all trials in each experiment. We are interested to see if there is consistency in Hit Ratio (any number of Hit Ratio) among all three trial results within an experiment. The standard deviation will be zero if all trials in each stock are resulting in the same Hit Ratio. Table 7 shows a summary of Hit Ratio consistency.

Table 7: Summary of Hit Ratio Consistency

	Experiments			
	E1/3d	E5/3d	E1/10d	E5/10d
# Stocks with Consistent Output	45	49	17	40
% Consistency	51%	56%	19%	45%

Experiments are grouped based on an input window period: 3 days and ten days. We see a substantial difference between the two groups. Three days of input delivers higher prediction consistency (above 50%) compared to 10 days of input. This finding implies that 1) Hit Ratio consistency could be an indication of optimum hyperparameters configuration, 2) model with the sub-optimum result (E5 with three days of input) could be acceptable if aiming to achieve Hit Ratio consistency.

5. CONCLUSIONS

In this paper, the impact of multiple leading indicators as feature sets to a next-day prediction model based on LSTM has been evaluated. This study provides empirical evidence on feature sets with strong predictability of future price movement.

We evaluate prediction performance against 88 stocks in the test set. Our primary evaluation metrics are Winning Stocks, which is the number of stocks from the test set capable of achieving a Hit Ratio above 60%. Baseline performance reference is RMSE 0.01609 and 42.33 Winning Stocks, performed using a univariate forecasting model with three hidden layers, 50 nodes, and three days of input.

A combination of feature sets produces different results, depending on objectives and the preferred evaluation metrics. In terms of RMSE, the best improvement is achieved by using two feature sets, either with trading volume (F_2) or daily cumulative sum of foreign volume (F_4). It is complementing this feature along with the price (F_1), capable of producing predictions that are closest to the actual values. This finding is in line with earlier studies on financial time series forecasting which mostly combines price (OHLC) and trading volume as input.

In contrast, achieving Hit Ratio above 60% across 88 stocks poses a higher challenge than simply improving RMSE. In terms of Winning Stocks, it is using a combination of 3 feature sets, namely: price (F_1), trading volume (F_2), and foreign volume (F_3), capable producing comparable results to benchmark with the additional advantage of fewer computing requirements. The result is achieved by using a simple LSTM architecture (a single hidden layer, 50 nodes, and ten days of the input window), contrasted with the benchmark architecture which uses 3 hidden layers. Adding more complexities, such as more hidden layers and more nodes, did not contribute to better predictions. To achieve these advantages, it is essential to 1) validate feature sets prior to incorporating as input to a model; 2) searching the optimum combination of hyperparameters which capable of producing the highest Hit Ratio.

This empirical evidence agrees with 1) Chen and Liu [5], that trading volume helps predict the behavior of the stock price, and 2) Richards [27], which confirms that price movement is associated with trades of foreign investors. This study contributes the use of foreign volume along with price and trading volume which would improve financial time series forecasting with LSTM.

There are two limitations in this study that could be addressed in future research. First, this study focuses on evaluating the impact of multiple feature sets; therefore, most of the chosen hyperparameters are not thoroughly assessed. The use of other optimizers, batch normalization, dropout layer could contribute to better performance. Second, this study limits the evaluated maximum input window to 10 days because the availability of the dataset is only five years. A study by Shynkevich et al. [30] evaluates the input window up to 30 days with ten years period of dataset.

This study has presented results which can be improved further. 1) Most recent advancement in recurrent neural network (RNN) can also enhance prediction performance. Attention-based RNN has achieved state of the art in processing long sequences in the field of natural language processing. Longer forecasting horizon could harness the advantages of attention-based RNN. 2) Hit Ratio consistency is a newly introduced metric complementing the commonly used Hit Ratio as directional prediction accuracy.

REFERENCES:

[1] J.-J. Wang, J.-Z. Wang, Z.-G. Zhang, and S.-P. Guo, "Stock index forecasting based on a hybrid model," *Omega*, vol. 40, no. 6, pp. 758–766, 2012, doi: https://doi.org/10.1016/j.omega.2011.07.008.

[2] R. C. Cavalcante, R. C. Brasileiro, V. L. F. Souza, J. P. Nobrega, and A. L. I. Oliveira, "Computational Intelligence and Financial Markets: A Survey and Future Directions," *Expert Systems with Applications*, vol. 55, pp. 194–211, Aug. 2016, doi: 10.1016/j.eswa.2016.02.006.

[3] T. Niu, J. Wang, H. Lu, W. Yang, and P. Du, "Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting," *Expert Systems with Applications*, vol. 148, p. 113237, 2020, doi: https://doi.org/10.1016/j.eswa.2020.113237.

[4] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning : A systematic literature review: 2005–2019," *Applied Soft Computing*, vol. 90, p. 106181, 2020, doi: https://doi.org/10.1016/j.asoc.2020.106181.

[5] S.-W. Chen and T. Liu, "On Asymmetric Causality Between Stock Prices and Trading Volume for Some Developed and Emerging Stock Markets : A Preliminary Analysis.", *International Review of Accounting, Banking and Finance*, vol 7 no. 1, 2015.

[6] D. Israeli, R. Kaniel, and S. A. Sridharan, "The Real Side of the High-Volume Return Premium," *SSRN Electronic Journal*, 2017, doi: 10.2139/ssrn.2909494.

[7] P.-Y. Huang, Y.-S. Ni, and C.-M. Yu, "The Microstructure of the Price-Volume Relationship of the Constituent Stocks of the Taiwan 50 Index," *Emerging Markets Finance and Trade*, vol. 48, no. sup2, pp. 153–168, Jul. 2012, doi: 10.2753/REE1540-496X48S209.

[8] S. Kim and M. Kang, "Financial series prediction using Attention LSTM," *CoRR*, vol. abs/1902.1, 2019.

[9] S. Varsamopoulos, K. Bertels, and C. Almudever, *Designing neural network based decoders for surface codes*. Preprint, 2018.

[10] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European Journal of Operational Research*, vol. 270, no. 2, pp. 654–669, Oct. 2018, doi: 10.1016/J.EJOR.2017.11.054.

[11] E. DEZSI and I. A. NISTOR, "Can Deep Machine Learning Outsmart The Market? A Comparison Between Econometric Modelling And Long- Short Term Memory," *Romanian Economic Business Review*, vol. 11, no. 4.1, pp. 54–73, Dec. 2016.

[12] H. M, G. E.A., V. K. Menon, and S. K.P., "NSE Stock Market Prediction Using Deep-Learning Models," *Procedia Computer Science*, vol. 132, pp. 1351–1362, 2018, doi: https://doi.org/10.1016/j.procs.2018.05.050.

[13] S. il Lee and S. J. Yoo, "Threshold-Based Portfolio: The Role of the Threshold and Its

Applications," *Journal of Supercomputing*, vol. 76, no. 10, pp. 8040-8057, 2018.

[14] W. Bao, J. Yue, and Y. Rao, "A deep learning framework for financial time series using stacked autoencoders and long-short term memory," *PLOS ONE*, vol. 12, no. 7, pp. 1-11. 2017.

[15] E. Chong, C. Han, and F. Park, "Deep Learning Networks for Stock Market Analysis and Prediction: Methodology, Data Representations, and Case Studies," *Expert Systems with Applications*, vol. 83, 2017, doi: 10.1016/j.eswa.2017.04.030.

[16] K. Chen, Y. Zhou, and F. Dai, "A LSTM-based method for stock returns prediction: A case study of China stock market," in *2015 IEEE International Conference on Big Data (Big Data)*, 2015, pp. 2823–2824, doi: 10.1109/BigData.2015.7364089.

[17] A. J. P. Samarawickrama and T. G. I. Fernando, "A recurrent neural network approach in predicting daily stock prices an application to the Sri Lankan stock market," in *2017 IEEE International Conference on Industrial and Information Systems (ICIIS)*, Dec. 2017, pp. 1–6, doi: 10.1109/ICIINFS.2017.8300345.

[18] S. Selvin, R. Vinayakumar, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "Stock price prediction using LSTM, RNN and CNN-sliding window model," in *2017 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2017, pp. 1643–1647, doi: 10.1109/ICACCI.2017.8126078.

[19] K. Khare, O. Darekar, P. Gupta, and V. Z. Attar, "Short term stock price prediction using deep learning," in *2017 2nd IEEE International Conference on Recent Trends in Electronics, Information Communication Technology (RTEICT)*, 2017, pp. 482–486, doi: 10.1109/RTEICT.2017.8256643.

[20] B. Zhou, "Deep Learning and the Cross-Section of Stock Returns: Neural Networks Combining Price and Fundamental Information," *SSRN Electronic Journal*, 2018, doi: 10.2139/ssrn.3179281.

[21] F. Garcia, F. Guijarro, J. Oliver, and R. Tamošiūnienė, "Hybrid fuzzy neural network to predict price direction in the German DAX-30 index," *Technological and Economic Development of Economy*, vol. 24, no. 6, pp. 2161–2178, Nov. 2018, doi: 10.3846/tede.2018.6394.

[22] S. Gervais, R. Kaniel, and D. H. Mingelgrin, "The High-Volume Return Premium," *The Journal of Finance*, vol. 56, no. 3, pp. 877–919, 2001.

[23] R. Kaniel, A. Ozoguz, and L. Starks, "The high volume return premium: Cross-country evidence," *Journal of Financial Economics*, vol. 103, no. 2, pp. 255–279, 2012, doi: https://doi.org/10.1016/j.jfineco.2011.08.012.

[24] S. Basak and A. Pavlova, "Asset Prices and Institutional Investors," *American Economic Review*, vol. 103, no. 5, pp. 1728–1758, Aug. 2013, doi: 10.1257/aer.103.5.1728.

[25] D. Lien, P.-H. Hung, and Z.-W. Lin, "Whose trades move stock prices? Evidence from the Taiwan Stock Exchange," *International Review of Economics & Finance*, vol. 66, no. 1, pp. 25–50, 2020, doi: https://doi.org/10.1016/j.iref.2019.10.011.

[26] R. D. Huang and C.-Y. Shiu, "Local Effects of Foreign Ownership in an Emerging Financial Market: Evidence from Qualified Foreign Institutional Investors in Taiwan," *Financial Management*, vol. 38, no. 3, pp. 567–602, Aug. 2009.

[27] A. Richards, "Big Fish in Small Ponds: The Trading Behavior and Price Impact of Foreign Investors in Asian Emerging Equity Markets," *The Journal of Financial and Quantitative Analysis*, vol. 40, no. 1, pp. 1–27, Aug. 2005.

[28] M. Qiu and Y. Song, "Predicting the Direction of Stock Market Index Movement Using an Optimized Artificial Neural Network Model," *PLOS ONE*, vol. 11, no. 5, p. e0155133, May 2016.

[29] X. Zhou, Z. Pan, G. Hu, S. Tang, and C. Zhao, "Stock Market Prediction on High-Frequency Data Using Generative Adversarial Nets," *Mathematical Problems in Engineering*, vol. 2018, p. 4907423, 2018, doi: 10.1155/2018/4907423.

[30] Y. Shynkevich, T. M. Mcginnity, S. Coleman, A. Belatreche, and Y. Li, "Forecasting Price Movements using Technical Indicators: Investigating the Impact of Varying Input Window Length," *Neurocomputing*, vol. 264, pp. 71-78, 2017, doi: 10.1016/j.neucom.2016.11.095.