Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 8, No. 4, December 2022, pp. 660-669
ISSN: 2338-3070, DOI:10.26555/jiteki.v8i4.25242

Export Commodity Price Forecasting in Indonesia Using Decision Tree, Random Forest, and Long Short-Term Memory

Shadifa Auliatama Harjanto, Siti Sa'adah, Gia Septiana Wulandari
Informatics, Telkom University, Bandung, Indonesia

ARTICLE INFO

Article history:

Received November 25, 2022
Revised January 02, 2023
Published January 04, 2023

Keywords:

GDP;
Export;
Commodity;
Decision Tree;
Random Forest;
LSTM

ABSTRACT

Gross Domestic Product (GDP) is an indicator that becomes a benchmark for a country's economic performance. One of the factors that significantly affect GDP is export activity. However, the problem that occurs is that the export value is relatively fluctuating, this is because commodity prices are always changing every time. Therefore, we need a system that can predict commodity prices accurately. The research contributions are to compare performance of several methods in commodity forecasting and build a system using artificial intelligence approach based on compared methods that has ability to forecast export commodities prices. In this study performance of several methods such as Decision Tree, Random Forest, and Long Short-Term Memory (LSTM) are compared to determine the best method in forecasting several export commodities in Indonesia. The commodities that were forecasted are the main commodities from each sector that dominates exports in Indonesia, namely palm oil from the manufacturing sector, coffee from the agricultural sector, and coal from the mining sector. The experiments in this study were conducted by testing several hyperparameters to determine the best model. To evaluate models, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) are used. The results show that LSTM has the lowest error among other methods with MAPE of 0.121, 0.494, and 0.282 in forecasting coal, coffee, and palm oil price respectively. Therefore, LSTM has proven to be the best method among Random Forest and Decision Tree in forecasting export commodity prices in Indonesia.

Corresponding Author:

Shadifa Auliatama Harjanto, Informatics, Telkom University, Bandung, Indonesia
Email: shadif@student.telkomuniversity.ac.id

1. INTRODUCTION

The growth of the goods production and services provision in the country's economic area can show the economic development of a country [1]-[3]. Production is considered to grow if, within a certain time interval, it can provide added value to the economy in the area concerned for production [4]. One of the indicators used to measure the added value is Gross Domestic Product (GDP) [1]. GDP is an indicator that becomes a benchmark for a country's economic performance [1], [4], [5]. One of the factors that significantly affect GDP is export activities [6]-[8]. In Indonesia, export activities are dominated by several commodities such as palm oil, coffee, and coal [9]-[11]. However, the problem that occurs is that the export value is relatively fluctuating. This is because commodity prices are always changing every time [12]. Based on this, we need a system that can predict commodity prices accurately so that it is hoped that the government can make appropriate export policies based on predictions of commodity prices in the future.

Several research related to commodity price forecasting has been done previously [13]-[18]. In 2020, Liu et al. conducted research on copper price prediction using Decision Tree [16]. In this study, the prediction of copper commodity prices on the Dow Jones Index was carried out using the Decision Tree. This study shows that the Decision Tree has a good performance with a mean absolute percentage error below 4% [16]. In 2019, Manjula and Karthikeyan also conducted research on predicting gold price using Machine Learning techniques [17]. The results of this study indicate that Random Forest has the best accuracy compared to other algorithms [17]. Another study was conducted by Sabu and Kumar in 2020 on predicting Arecanuts prices [18]. The result indicates that LSTM has the best performance among other methods [18].

However, forecasting that was conducted in these studies are limited to one commodity hence the best method from each of these studies has not been tested and compared on forecasting other commodities yet. As a result, performance of aforementioned methods in forecasting other commodities have not been analyzed yet. Therefore, this study aims to propose a forecasting system for export commodities prices in Indonesia made using an artificial intelligence approach by comparing several methods based on previous studies, namely Decision Tree, Random Forest, and LSTM. Data used in this study are from several export commodities price in Indonesia. Experiments also conducted using various hyperparameters of each method to determine method with best performance. Commodities forecasted in this study are the main commodities that dominate exports in Indonesia, namely palm oil from the manufacturing sector, coffee from the agricultural sector, and coal from the mining sector [9]-[11]. The research contributions are to compare performance of several methods in commodity forecasting and build a system using artificial intelligence approach based on compared methods that has ability to forecast export commodities prices.

2. METHODS

This research was conducted in several steps that started from collecting the data and preparing it to make sure that the data is feasible to be processed by Decision Tree, Random Forest, and LSTM, then the algorithms are trained using several hyperparameters to determine the best models. After that, the models are tested and evaluated. The flowchart of research methodology is visualized in Fig. 1 and the details are elaborated in this section.

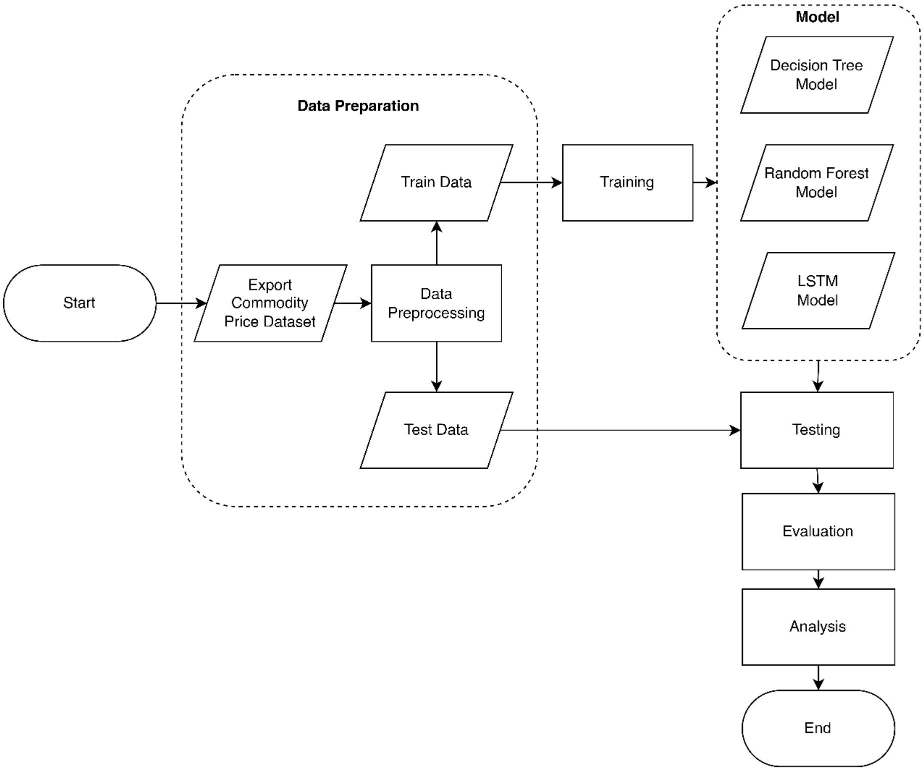

Fig. 1. Research Method Flowchart

2.1. Export Commodity Price Dataset

This study used data on commodities prices for coffee, palm oil, and coal taken from the www.investing.com website to train and test models. Due to limited data on palm oil prices, the period of data used in this study is from January 2011 to July 2022. The overview of this data is presented in Table 1.

Table 1. Data Overview

Commodity	Date	Close	Open	Highest	Lowest	Volume	Change%
Coal	07/01/2022	388.00	388.00	388.00	388.00	0.01K	0.53%
	06/30/2022	385.95	380.00	390.00	380.00	0.01K	1.57%
	06/29/2022	380.00	389.00	390.00	380.00	0.04K	-3.43%
	…	…	…	…	…	…	…
Coffee	07/01/2022	228.45	232.85	232.85	227.90	0.07K	-0.72%
	06/30/2022	230.10	227.85	231.65	227.85	14.05K	0.81%
	06/29/2022	228.25	218.00	230.20	217.55	26.70K	4.82%
	…	…	…	…	…	…	…
Palm Oil	07/01/2022	108.43	106.01	109.34	104.56	305.34K	2.52%
	06/30/2022	105.76	109.70	110.45	105.10	362.89K	-3.66%
	06/29/2022	109.78	111.86	114.05	109.22	322.06K	-1.77%
	…	…	…	…	…	…	…

Two different time frames were used in this study, daily and monthly. A total of more than 2000 daily data and 100 monthly data from each commodity are processed using three different methods with various hyperparameters that are used in this study.

2.2. Data Preparation

At this stage, the data was prepared before entering the processing stage. Some of the steps that are carried out at this stage include adjusting the data type, identifying the presence of empty values and outliers, and handling them [19]. After that, due to the significant range difference in the value scale of the data in the "Volume" and "Change%" columns, a scaling feature was carried out so that the data scale in each column is the same. This is done because the difference in scale in the data will make the prediction model less than optimal [20]. Furthermore, the data were divided into train data for model training process and test data for model evaluation with a split percentage of 80 and 20 respectively.

2.3. Decision Tree

As shown in its name, this algorithm resembles a data structure which is Tree [21], [22]. In general, there are two types of decision trees, namely the classification decision tree which is used to make predictions with discrete outputs and the regression decision tree which is used to make predictions with continuous outputs [21], [23]. The algorithm uses a hierarchical structure whose process starts from the root node towards the leaf [24]. Each branch in the algorithm states the conditions that must be met [25]. To determine the root node, it is necessary to calculate the entropy of the data and the gain of each attribute contained in the data [23]. The attribute that has the highest gain will be set as the root node.

$$Entropy(S) = \sum_{i=1}^{n} -p_i \times log_2 p_i \tag{1}$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^{n} \frac{|S_i|}{|S|} \times Entropy(S_i) \tag{2}$$

The formula to calculate $Entropy$ and $Gain$ are shown in (1) and (2) respectively, where S is the dataset case, p_i is the proportion of samples, n is number of S partitions, A is the feature, $|S_i|$ is number of instances with value i of attribute A, and $|S|$ is number of instances in dataset S [25].

2.4. Random Forest

This is an algorithm that includes an ensemble method that can perform classification and regression using bagging techniques [17], [26]. This algorithm was first introduced by Breiman in 2001 by combining Bagging theory with classification decision trees and regression decision trees [27], [28]. This algorithm is formed by combining several decision trees to lower the variance [29], [30]. All trees that are constructed in this algorithm are individual and the average output of each tree was defined as the final result of this algorithm [27].

This algorithm is efficient because every tree that was constructed is a binary tree with a maximum distance from the root to the terminal node, this distance is also known as the maximum depth [26]. One of the advantages of this algorithm is that it has feature engineering capabilities that function to distinguish the most vital features inside data [17], [31]. This feature selection process is conducted randomly while the tree is growing [31].

2.5. Long Short-Term Memory

In general, the architecture of this algorithm is the advance form of a conventional Recurrent Neural Network (RNN) [32]-[35]. This algorithm is superior to RNN because it can perform better in remembering long periods of information, this is because LSTM has memory cells that can be manipulated [33]. LSTM has four main components, namely memory cells, forget gates, input gates, and output gates [36]. In the forget gate, the data entered will be determined whether to be used or discarded [37]. The determination will be made by the sigmoid activation function contained in the gate with an output of 1 if the data will be used and 0 if the data will be discarded.

$$f_t = \sigma\left(W_f . [h_{t-1}, x_t] + b_f\right) \tag{3}$$

The determination of this formula is shown in (3). Where f_t symbolize the forget gate of an LSTM unit, σ symbolize activation function inside the unit, which is sigmoid, W_f symbolize weight that applied in forget gate, h_{t-1} symbolize output resulted from previous LSTM unit, x_t symbolize input from present LSTM unit, and b_f symbolize bias in forget gate. Then at the input gate, there is a sigmoid activation function which will normalize the calculation of weight, bias, and input multiplication. In a memory cell, information stored in vector form, the formulas are shown in (4) and (5).

$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \tag{4}$$

$$\check{c}_t = tanh(W_c . [h_{t-1}, x_t] + b_c) \tag{5}$$

In (4) and (5), i_t symbolize forget gate, σ symbolize sigmoid activation function, W_i symbolize weight in input gate, h_{t-1} symbolize output resulted from earlier unit of LSTM, while x_t symbolize input from present LSTM unit, b_i symbolize bias in input gate, $\check{c}_t$ symbolize a candidate for memory cell state, $tanh$ symbolize the applied activation function, W_c symbolize applied weight in cell state, and b_c symbolize bias in cell state. Then at the output gate, the final output value from current unit of LSTM will be determined, the formulas are shown in (6) and (7).

$$o_t = \sigma(W_o . [h_{t-1}, x_t] + b_o) \tag{6}$$

$$h_t = o_t tanh(c_t) \tag{7}$$

In (6) and (7), o_t symbolize forget gate, σ symbolize activation function, W_o symbolize weight in output gate, h_{t-1} symbolize previous LSTM output, x_t is input from current LSTM, b_o symbolize output gate bias, h_t symbolize output from present LSTM unit, tanh symbolize the applied activation function, and c_t is the cell state.

2.6. Training using Decision Tree, Random Forest, and LSTM

In the training process, experiments were conducted using various hyperparameters. Experiments using Decision Tree tested splitter and maximum depth as hyperparameters to determine the best model. In these experiments, there are two types of splitting strategies, “best” and “random”. The maximum depth that was tested in these experiments is 10, 20, 30, and 40. Experiments using Random Forest also use maximum depth as one of the hyperparameters since Random Forest are a combination of Decision Trees. The difference is Random Forest used several trees of 100, 200, 300, and 400. In experiments using LSTM, the hyperparameters used are the LSTM unit and dropout rate. In these experiments, several LSTM model tests using LSTM unit variations of 100 and 200 and dropout rates of 0, 0.2, 0.5, and 0.7.

2.7. Testing using Decision Tree, Random Forest, and LSTM

After the training process was completed, each model was tested by processing test data to forecast commodities prices. The performance of every model was measured using evaluation metrics used in this study. Performance comparison of every model was also conducted using both daily data and monthly data. The testing results will be evaluated and analyzed in the next stage.

2.8. Evaluation

At this stage, model evaluation is carried out individually using Root Mean Squared Error (RMSE), Mean Average Error (MAE), and Mean Absolute Percentage Error (MAPE). The measurement conducted using these metrics was carried out to assess each method, thereby the best method was obtained. The formula of RMSE, MAE, and MAPE are shown in (8), (9), and (10) respectively [38]-[40].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2} \tag{8}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y_i}| \tag{9}$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y_i}|}{y_i} \tag{10}$$

Where n is total data points, y_i symbolize actual value in $i - th$, and $\hat{y_i}$ symbolize predicted value in i-th predicted value.

2.9. Analysis

The evaluation results are analyzed at this stage. Every experiment conducted in this study was compared to determine which model has the best performance alongside the hyperparameters that the model used. The best model is considered by the overall performance calculated using the average result of the model in forecasting each commodity price that was evaluated in the previous stage.

3. RESULTS AND DISCUSSION

The result of experiments conducted in this study is given in this section. There are two different data time frames for each commodity price data used in this study, those time frames are daily and monthly. Each time frame is trained and tested using a aforementioned methods. The details of the results are given in the following section.

3.1. Results

The first experiment conducted is using the Decision Tree, and the results are shown in Table 2. From the experiments using the Decision Tree, the overall average error of the model is lower when a greater number of maximum depth of the tree is used and these results are similar when using both the best splitter and random splitter. Decision Tree models that applied maximum depth with value of 20 had the lowest error among other models. This indicates that increasing the depth of the tree significantly improves the model's performance, which means trees with a maximum depth of 20 forecasts was approaching the actual data.

On the other hand, the use of different splitters caused different results for each commodity. From the results, there were two models with a maximum depth of 5 that has the best performance and this applies to all commodities. However, in experiment on forecasting coal price, model that applied maximum depth with value of 20 and using a random splitter are better than the model that applied maximum depth with value of 20 and using the best splitter but in coffee and palm oil result are the opposite. As shown in Table 2, in the daily time frame, the MAPE range of coal, coffee, and palm oil was 0.1-1.3, these results indicate that the best splitter did not always work best in every data.

Using the same maximum depth variation as the Decision Tree and combined with total trees as hyperparameters, the result of experiments using Random Forest as the second model are shown in Table 3. Random Forest with 100 trees and a maximum depth of 20 achieved the lowest error with a MAPE of 0.189. The combination of multiple trees makes Random Forest performance is superior to a conventional Decision Tree. Nevertheless, according to Table 3, the performance of models with 100 trees is similar to the model with 400 trees. This indicates that raising tree quantity does not improve the Random Forest performance significantly, because the results MAPE was still in the range of 0.18-0.19 in daily data.

The forecasting results of LSTM are presented detailly in Table 4. RMSE, MAE, and MAPE are lower than the test results using previous methods. The cell memory of LSTM is playing important role in achieving these results as cell memory could remember a large amount of information. This gives LSTM an advantage in forecasting time series data that has a long period. Both results in daily and monthly time frames show that models with 200 LSTM unit have better performance than models with 100 LSTM units. These results also indicate that increasing dropout did not improve nor affect the model's performance significantly as shown that the model with 0 dropout rate achieved the best performance among others.

Table 2. Commodities Price Forecasting Result Using Decision Tree

Hyperparameter		**Commodities**								
		Coal			**Coffee**			**Palm Oil**		
Splitter	**Maximum Depth**	**RMSE (%)**	**MAE (%)**	**MAPE (%)**	**RMSE (%)**	**MAE (%)**	**MAPE (%)**	**RMSE (%)**	**MAE (%)**	**MAPE (%)**
Daily Data										
best	10	0.396	0.240	0.235	5.179	2.318	1.106	0.851	0.620	0.665
	20	0.355	0.183	0.175	5.158	2.300	1.102	0.867	0.638	0.684
	30	0.355	0.183	0.175	5.046	2.275	1.097	0.867	0.638	0.684
	40	0.355	0.183	0.175	5.046	2.275	1.097	0.867	0.638	0.684
random	10	0.395	0.246	0.241	5.892	2.594	1.242	0.980	0.696	0.745
	20	0.332	0.167	0.154	5.990	2.735	1.309	0.933	0.675	0.720
	30	0.332	0.167	0.154	6.206	2.736	1.291	1.034	0.709	0.758
	40	0.332	0.167	0.154	6.206	2.736	1.291	1.034	0.709	0.758
Monthly Data										
best	10	9.707	5.400	5.147	16.493	11.229	5.988	6.855	5.657	6.144
	20	9.501	4.982	4.768	13.230	9.212	4.998	6.855	5.657	6.144
	30	9.501	4.982	4.768	13.230	9.212	4.998	6.855	5.657	6.144
	40	9.501	4.982	4.768	13.230	9.212	4.998	6.855	5.657	6.144
random	10	7.784	3.676	3.609	14.586	9.719	5.240	9.701	8.333	8.889
	20	8.165	4.285	4.283	13.924	9.285	5.065	7.485	5.872	6.292
	30	8.165	4.285	4.283	13.924	9.285	5.065	7.485	5.872	6.292
	40	8.165	4.285	4.283	13.924	9.285	5.065	7.485	5.872	6.292

Table 3. Commodities Price Forecasting Result Using Random Forest

Hyperparameter		**Commodities**								
		Coal			**Coffee**			**Palm Oil**		
Trees	**Maximum Depth**	**RMSE (%)**	**MAE (%)**	**MAPE (%)**	**RMSE (%)**	**MAE (%)**	**MAPE (%)**	**RMSE (%)**	**MAE (%)**	**MAPE (%)**
Daily Data										
100	10	0.373	0.248	0.241	4.846	2.215	1.064	0.807	0.606	0.652
	20	0.332	0.199	0.189	4.841	2.210	1.063	0.813	0.610	0.656
	30	0.346	0.210	0.199	4.840	2.210	1.063	0.812	0.609	0.655
200	10	0.375	0.248	0.242	4.847	2.216	1.065	0.807	0.605	0.651
	20	0.344	0.208	0.197	4.840	2.210	1.063	0.813	0.610	0.657
	30	0.357	0.215	0.202	4.842	2.212	1.064	0.813	0.610	0.656
300	10	0.374	0.248	0.241	4.847	2.216	1.065	0.808	0.607	0.653
	20	0.35	0.211	0.199	4.841	2.211	1.063	0.813	0.610	0.655
	30	0.348	0.210	0.199	4.841	2.211	1.063	0.813	0.610	0.656
400	10	0.378	0.251	0.243	4.847	2.216	1.065	0.807	0.606	0.652
	20	0.339	0.205	0.195	4.842	2.211	1.063	0.812	0.609	0.654
	30	0.344	0.208	0.197	4.841	2.212	1.064	0.813	0.610	0.656
Monthly Data										
100	10	2.278	1.818	1.901	12.939	8.290	4.413	6.125	5.240	5.639
	20	2.345	1.870	1.944	12.854	8.165	4.353	6.062	5.184	5.583
	30	2.345	1.870	1.944	12.854	8.165	4.353	6.062	5.184	5.583
200	10	2.133	1.743	1.852	13.157	8.443	4.492	6.032	5.140	5.550
	20	2.193	1.794	1.894	13.089	8.372	4.455	5.869	4.982	5.379
	30	2.193	1.794	1.894	13.089	8.372	4.455	5.869	4.982	5.379
300	10	2.068	1.684	1.798	13.159	8.513	4.542	6.117	5.194	5.619
	20	2.104	1.718	1.825	13.115	8.461	4.513	5.985	5.072	5.485
	30	2.104	1.718	1.825	13.115	8.461	4.513	5.985	5.072	5.485
400	10	2.085	1.703	1.814	13.132	8.459	4.506	5.937	4.962	5.377
	20	2.112	1.729	1.835	13.144	8.485	4.519	5.840	4.854	5.260
	30	2.112	1.729	1.835	13.144	8.485	4.519	5.840	4.854	5.260

Table 4. Commodities Price Forecasting Result Using LSTM

Hyperparameter		Commodity								
		Coal			Coffee			Palm Oil		
LSTM Unit	Dropout Rate	RMSE (%)	MAE (%)	MAPE (%)	RMSE (%)	MAE (%)	MAPE (%)	RMSE (%)	MAE (%)	MAPE (%)
Daily Data										
100	0	0.149	0.120	0.121	1.841	1.005	0.517	0.359	0.263	0.282
	0.2	0.284	0.191	0.204	4.229	1.794	0.858	1.148	0.881	0.961
	0.5	0.375	0.302	0.320	5.375	2.641	1.329	1.398	1.073	1.143
	0.7	0.962	0.828	0.905	6.76	3.322	1.630	1.603	1.37	1.446
200	0	0.158	0.123	0.122	1.631	0.955	0.494	0.352	0.271	0.288
	0.2	0.452	0.407	0.410	3.29	1.383	0.658	0.662	0.564	0.587
	0.5	0.354	0.244	0.258	3.737	1.969	0.996	1.920	1.706	1.802
	0.7	1.853	1.659	1.607	4.608	2.304	1.125	1.555	1.303	1.392
Monthly Data										
100	0	3.857	3.637	4.198	11.801	7.400	4.041	4.669	3.787	4.073
	0.2	3.461	3.099	3.612	11.544	7.362	4.008	4.784	3.888	4.180
	0.5	4.275	3.956	4.626	11.539	7.770	4.278	4.770	3.824	4.127
	0.7	4.259	3.389	3.817	11.554	8.012	4.364	5.997	5.061	5.261
200	0	2.429	2.256	2.728	6.646	3.411	1.717	1.117	0.772	0.821
	0.2	1.325	1.253	1.458	7.940	3.351	1.606	1.180	0.788	0.835
	0.5	3.363	3.175	3.591	9.393	4.164	2.115	1.266	1.037	1.085
	0.7	2.950	2.540	2.756	5.249	3.075	1.613	3.230	3.036	3.194

3.2. Discussion

The best results from the Decision Tree, Random Forest, and LSTM are summarized in Table 5. In this table, the model with the lowest error from each method was considered the best in this study. The results of monthly data forecasting show significantly higher errors than daily data. This was caused by fewer monthly data than daily data in the same period, therefore these methods are trained with fewer data in forecasting the monthly price of each commodity. In general, LSTM has the best performance among other methods. In daily time frames, the best LSTM model in this study achieved MAPE of 0.121, 0.494, and 0.282 in forecasting coal, coffee, and palm oil prices respectively.

Table 5. Best Results Comparison

Commodity	Time Frame	Model								
		Decision Tree			Random Forest			LSTM		
		RMSE (%)	MAE (%)	MAP E (%)	RMSE (%)	MAE (%)	MAPE (%)	RMSE (%)	MAE (%)	MAPE (%)
Coal	Daily	0.332	0.167	0.154	0.332	0.199	0.189	0.149	0.120	0.121
	Monthly	7.784	3.676	3.609	2.068	1.684	1.798	1.462	1.077	1.135
Coffee	Daily	5.046	2.275	1.097	4.841	2.211	1.063	1.631	0.955	0.494
	Monthly	13.230	9.212	4.998	12.854	8.165	4.353	4.407	2.917	1.554
Palm Oil	Daily	0.851	0.620	0.665	0.807	0.605	0.651	0.359	0.263	0.282
	Monthly	6.855	5.657	6.144	5.840	4.854	5.260	1.376	1.126	1.205

The best result from this study is superior compared to previous studies in forecasting commodities prices. Based on research that was conducted by Herrera et al. in 2019, their Random Forest model has the best performance with a MAPE of 8.148 [41]. In 2020, Novanda et al. experiments on forecasting coffee prices, and the result was their ARIMA model had the best performance with a MAPE of 3.760 [42]. Another study was also conducted by Nhita et al. on forecasting agricultural commodities and their FLNN model optimized using the Artificial Intelligence Bee Colony algorithm has the best performance with MAPE of 7.68 [43].

4. CONCLUSION

From the experiments conducted in this study show that LSTM has the best performance in forecasting Indonesia's export commodities compared to other methods. In experiments using the Decision Tree and Random Forest, higher value of maximum depth seems to be improved the model's performance. However, the overall result from these two methods shows that the performance of Random Forest in forecasting export commodities in Indonesia is marginally superior to Decision Tree. Significant results are achieved by LSTM models from this study, supported by the outcomes based on the best LSTM model which have MAPE of

0.462, 0.483, and 0.348 in forecasting coal, coffee, and palm oil prices in this study respectively. For future work, more data commodities can be used to give comprehensive performance comparisons in commodity price forecasting. Various methods (particularly for the deep learning approach) and hyperparameters also can be implemented to determine the most fitting model for commodity price forecasting.

REFERENCES

[1] Y. Chen, G. Wu, Y. Ge, and Z. Xu, "Mapping Gridded Gross Domestic Product Distribution of China Using Deep Learning with Multiple Geospatial Big Data," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 15, pp. 1791-1802, 2022, https://doi.org/10.1109/JSTARS.2022.3148448.

[2] S. C. Agu, F. U. Onu, U. K. Ezemagu, and D. Oden, "Predicting gross domestic product to macroeconomic indicators," *Intelligent Systems with Applications*, vol. 14, 2022, https://doi.org/10.1016/j.iswa.2022.200082.

[3] Y. Xu, L. He, Y. Liang, J. Si, and Y. Bao, "Enterprise Power Consumption Data and GDP Forecasting Based on Ensemble Algorithms," *E3S Web of Conferences*, vol. 30, pp. 3-6, 2021, https://doi.org/10.1051/e3sconf/202123301030.

[4] X. Wu, Z. Zhang, H. Chang, and Q. Huang, "A Data-Driven Gross Domestic Product Forecasting Model Based on Multi-Indicator Assessment," *IEEE Access*, vol. 9, pp. 99495-99503, 2021, https://doi.org/10.1109/ACCESS.2021.3062671.

[5] Z. Fannoun and I. Hassouneh, "The causal relationship between exports, imports and economic growth in Palestine," *Journal of Reviews on Global Economics*, vol. 8, pp. 258-268, 2019, https://doi.org/10.6000/1929-7092.2019.08.22.

[6] M. R. Islam and M. Haque, "The Trends of Export and Its Consequences to the GDP of Bangladesh," *Journal of Social Sciences and Humanities*, vol. 1, no. 1, pp. 63-67, 2018, [Online]. Available: http://www.aascit.org/journal/archive2?journalId=931&paperId=6460

[7] O. O. Awe, D. M. Akinlana, O. O. S. Yaya, and O. Aromolaran, "Time series analysis of the behaviour of import and export of agricultural and non-agricultural goods in West Africa: A case study of Nigeria," *Agris On-line Papers in Economics and Informatics*, vol. 10, no. 2, pp. 15-22, 2018, https://doi.org/10.7160/aol.2018.100202.

[8] A. M. Khan and U. Khan, "The stimulus of export and import performance on economic growth in oman," *Montenegrin Journal of Economics*, vol. 17, no. 3, pp. 71-86, 2021, https://doi.org/10.14254/1800-5845/2021.17-3.6.

[9] B. Rahardjo, B. Mukhammad, B. Akbar, and A. Shalehah, "Analysis and strategy for improving Indonesian coffee competitiveness in the international market," *BISMA*, vol. 12, no. 2, pp. 154-167, 2020, https://doi.org/10.26740/bisma.v12n2.p154-167.

[10] A. Rifin, Feryanto, Herawati, and Harianto, "Assessing the impact of limiting Indonesian palm oil exports to the European Union," *J Econ Struct*, 2020, https://doi.org/10.1186/s40008-020-00202-8.

[11] R. Meyliawati, D. Saputera, G. N. Alam, D. F. Moenardi, R. Muttaqin, and R. A. Dewi, "Determinants of Indonesian Coal Commodity Export Before and Post the Spread of the Covid-19 Policy," *East Asian Journal of Multidisciplinary Research*, vol. 1, no. 5, pp. 801-812, 2022, https://doi.org/10.55927/eajmr.v1i5.567.

[12] S. T. Wahyudi and I. Maipita, "Comparative Analysis on The Market Share of Indonesian Export Commoties: Opportunities and Challenges," *Jurnal Ekonomi Pembangunan: Kajian Masalah Ekonomi dan Pembangunan*, vol. 19, no. 2, pp. 163-171, 2019, https://doi.org/10.23917/jep.v19i2.5708.

[13] B. Shao, M. Li, Y. Zhao, and G. Bian, "Nickel Price Forecast Based on the LSTM Neural Network Optimized by the Improved PSO Algorithm," *Math Probl Eng*, 2019, https://doi.org/10.1155/2019/1934796.

[14] H. Lin and Q. Sun, "Crude oil prices forecasting: An approach of using Ceemdan-based multi-layer gated recurrent unit networks," *Energies (Basel)*, vol. 13, no. 7, pp. 1-21, 2020, https://doi.org/10.3390/en13071543.

[15] I. E. Livieris, E. Pintelas, N. Kiriakidou, and S. Stavroyiannis, "An Advanced Deep Learning Model for Short-Term Forecasting U.S. Natural Gas Price and Movement," *IFIP, Springer International Publishing*, pp. 165-176, 2020, https://doi.org/10.1007/978-3-030-49190-1_15.

[16] C. Liu, Z. Hu, Y. Li, and S. Liu, "Forecasting copper prices by decision tree learning," *Resources Policy*, vol. 52, pp. 427-434, 2017, https://doi.org/10.1016/j.resourpol.2017.05.007.

[17] K. A. Manjula and P. Karthikeyan, "Gold Price Prediction using Ensemble based Machine Learning Techniques," *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, pp. 1360-1364, 2020, https://doi.org/10.1109/ICOEI.2019.8862557.

[18] K. M. Sabu and T. K. M. Kumar, "Predictive analytics in Agriculture: Forecasting prices of Arecanuts in Kerala," *Procedia Comput Sci*, vol. 171, pp. 699-708, 2020, https://doi.org/10.1016/j.procs.2020.04.076.

[19] A. Andiojaya and H. Demirhan, "A bagging algorithm for the imputation of missing values in time series," *Expert Syst Appl*, vol. 129, pp. 10-26, 2019, https://doi.org/10.1016/j.eswa.2019.03.044.

[20] J. C. Pena, G. Nápoles, and Y. Salgueiro, "Normalization method for quantitative and qualitative attributes in multiple attribute decision-making problems," *Expert Syst Appl*, vol. 198, p. 116821, 2022, https://doi.org/10.1016/j.eswa.2022.116821.

[21] E. Chen and X. J. He, "Crude Oil Price Prediction with Decision Tree Based Regression Approach Crude Oil Price Prediction with Decision Tree Based Regression Approach," *Journal of International Technology and Information Management*, vol. 27, no. 4, 2019, [Online]. Available: https://scholarworks.lib.csusb.edu/jitim/vol27/iss4/1

[22] T. Brabenec, P. Suler, J. Horak, and M. Petras, "Prediction of the Future Development of Gold Price," *Acta Montanistica Slovaca*, vol. 25, 2021, https://doi.org/10.46544/AMS.v25i2.11.
[23] H. Zhou, J. Zhang, Y. Zhou, X. Guo, and Y. Ma, "A feature selection algorithm of decision tree based on feature weight," *Expert Syst Appl*, vol. 164, p. 113842, 2021, https://doi.org/10.1016/j.eswa.2020.113842.
[24] K. M. Hindrayani, T. M. Fahrudin, R. Prismahardi Aji, and E. M. Safitri, "Indonesian Stock Price Prediction including Covid19 Era Using Decision Tree Regression," *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems, ISRITI 2020*, pp. 344-347, 2020, https://doi.org/10.1109/ISRITI51436.2020.9315484.
[25] L. D. Yulianto, A. Triayudi, and I. D. Sholihati, "Implementation Educational Data Mining for Analysis of Student Performance Prediction with Comparison of K-Nearest Neighbor Data Mining Method and Decision Tree C4.5," *Jurnal Mantik*, vol. 4, no. 1, pp. 441-451, 2020, [Online]. Available: https://iocscience.org/ejournal/index.php/mantik/article/view/770
[26] R. K. Paul *et al.*, "Machine learning techniques for forecasting agricultural prices: A case of brinjal in Odisha, India," *PLoS One*, vol. 17, pp. 1-17, 2022, http://dx.doi.org/10.1371/journal.pone.0270553.
[27] W. Deng, Y. Guo, J. Liu, Y. Li, D. Liu, and L. Zhu, "A missing power data filling method based on improved random forest algorithm," *Chinese Journal of Electrical Engineering*, vol. 5, no. 4, pp. 33-39, 2019, https://doi.org/10.23919/CJEE.2019.000025.
[28] L. Tapak, H. Abbasi, and H. Mirhashemi, "Assessment of factors affecting tourism satisfaction using K-nearest neighborhood and random forest models," *BMC Res Notes*, vol. 12, no. 1, pp. 1-5, 2019, https://doi.org/10.1186/s13104-019-4799-6.
[29] S. R. Polamuri, K. Srinivas, and A. Krishna Mohan, "Stock market prices prediction using random forest and extra tree regression," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 1224-1228, 2019, https://doi.org/10.35940/ijrte.C4314.098319.
[30] C. Pierdzioch and M. Risse, "Forecasting precious metal returns with multivariate random forests," *Empir Econ*, vol. 58, no. 3, pp. 1167-1184, 2020, https://doi.org/10.1007/s00181-018-1558-9.
[31] K. Bhatia, R. Mittal, J. Varanasi, and M. M. Tripathi, "An ensemble approach for electricity price forecasting in markets with renewable energy resources," *Util Policy*, vol. 70, p. 101185, 2021, https://doi.org/10.1016/j.jup.2021.101185de.
[32] Z. Xu, H. Deng, and Q. Wu, "Prediction of soybean price trend via a synthesis method with multistage model," *International Journal of Agricultural and Environmental Information Systems*, vol. 12, no. 4, pp. 1-13, 2021, https://doi.org/10.4018/IJAEIS.20211001.oa1.
[33] L. Boongasame, P. Viriyaphol, K. Tassanavipas, and P. Temdee, "Gold-Price Forecasting Method Using Long Short-Term Memory and the Association Rule," *Journal of Mobile Multimedia*, vol. 19, no. 1, pp. 165-186, 2022, https://doi.org/10.13052/jmm1550-4646.1919.
[34] P. H. Kuo and C. J. Huang, "An electricity price forecasting model by hybrid structured deep neural networks," *Sustainability (Switzerland)*, vol. 10, no. 4, pp. 1-17, 2018, https://doi.org/10.3390/su10041280.
[35] M. F. Maulana, S. Sa, and P. E. Yunanto, "Crude Oil Price Forecasting Using Long Short-Term Memory," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika*, vol. 7, no. 2, pp. 286-295, 2021, https://doi.org/10.26555/jiteki.v7i2.21086.
[36] T. B. Shahi, A. Shrestha, A. Neupane, and W. Guo, "Stock price forecasting with deep learning: A comparative study," *Mathematics*, vol. 8, no. 9, pp. 1-15, 2020, https://doi.org/10.3390/math8091441.
[37] D. J. V. Lopes, G. D. S. Bobadilha, and A. P. V. Bedette, "Analysis of lumber prices time series using long short-term memory artificial neural networks," *Forests*, vol. 12, no. 4, 2021, https://doi.org/10.3390/f12040428.
[38] T. O. Hodson, "Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not," *Geosci Model Dev*, vol. 15, no. 14, pp. 5481-5487, 2022, https://doi.org/10.5194/gmd-15-5481-2022.
[39] J. Wełeszczuk, B. Kosińska-Selbi, and P. Cholewińska, "Prediction of Polish Holstein's economical index and calving interval using machine learning," *Livest Sci*, vol. 264, 2022, https://doi.org/10.1016/j.livsci.2022.105039.
[40] M. U. Ahmed and I. Hussain, "Prediction of Wheat Production Using Machine Learning Algorithms in northern areas of Pakistan," *Telecomm Policy*, vol. 46, no. 6, p. 102370, 2022, https://doi.org/10.1016/j.telpol.2022.102370.
[41] G. P. Herrera, M. Constantino, B. M. Tabak, H. Pistori, J. J. Su, and A. Naranpanawa, "Long-term forecast of energy commodities price using machine learning," *Energy*, vol. 179, pp. 214-221, 2019, https://doi.org/10.1016/j.energy.2019.04.077.
[42] R. R. Novanda *et al.*, "A Comparison of Various Forecasting Techniques for Coffee Prices," *J Phys Conf Ser*, vol. 1114, no. 1, 2018, https://doi.org/10.1088/1742-6596/1114/1/012119.
[43] F. Nhita, D. Saepudin, A. Paramita, S. Marliani, and U. N. Wisesty, "Price prediction for agricultural commodities in Bandung regency based on Functional Link Neural Network and artifical bee colony algorithms," *Journal of Computer Science*, vol. 15, no. 10, pp. 1390-1395, 2019, https://doi.org/10.3844/jcssp.2019.1390.1395.

BIOGRAPHY OF AUTHORS

Shadifa Auliatama Harjanto presently pursuing his bachelor's degree in informatics at Telkom University. His research interests focus on Artificial Intelligence especially in Machine Learning and Deep Learning. Email: shadif@student.telkomuniversity.ac.id

Siti Sa'adah pursued Bachelor's and Master's from Informatic Faculty at Telkom University. since 2009 until now, she has joined as a lecturer at her almameter, and then interested in the research topics of AI/ML, Data Mining, and Financial Computing. Email: sitisaadah@telkomuniversity.ac.id

Gia Septiana Wulandari completed her doctoral study in 2021 at the University of York with a dissertation topic on formal verification of graph programming. In addition to continuing research from his dissertation, Gia is currently actively teaching courses on Algorithmic Complexity Analysis, Algorithmic Strategy, Knowledge Representation, and Introduction to Competitive Programming. Email: giaseptiana@telkomuniversity.ac.id