ICIC Express Letters
Volume **13**, Number **11**, November **2019**
 ISSN 1881-803X
pp. **987–993**

FINANCIAL REVENUE PREDICTION MODEL IN MINING INDUSTRY USING DEEP LEARNING

Venta Adrian Ahnaf and Suharjito

Computer Science Department, BINUS Graduate Program – Master of Computer Science
Bina Nusantara University
Jl. K. H. Syahdan No. 9, Kemanggisan, Palmerah, Jakarta 11480, Indonesia
venta.ahnaf@binus.ac.id; suharjito@binus.edu

Received March 2019; accepted June 2019

Abstract. *The financial revenue prediction model for the mining industry has been challenged due to the many factors that influence commodity price fluctuations. Revenue in the mining industry from year to year has continued to decline, and this has an impact on large companies that have an impact on the company's performance and profits. Through this problem the researcher aims to propose a prediction model in the mining industry using deep learning (DL), which has not been widely applied in financial revenue especially in the mining sector. This study uses a model of neural network LSTM learning approaches. The implementation of financial revenue prediction is done by comparing the LSTM model with an ARIMA. Data obtained for the last 6 years 2012-2017 from data set of mining industry processed by the Ministry of Energy and Mineral Resources and the World Bank. The evaluation model is compared between LSTM and ARIMA used by root mean squared error (RMSE). In the ARIMA prediction with model (5, 1, 0) a ratio of 0.1 coefficients and a standard error below 1, p value below 0, this indicates that the LSTM model has better accuracy compared to ARIMA with an average error of 0.24 on revenue indicators and 0.32 on the price indicator commodity.*
Keywords: Deep learning, Financial revenue, Predictive model, LSTM, ARIMA

1. **Introduction.** Financial prediction models in the mining industry have been carried out as much as in the analysis of predictions and estimates of future results of commodity prices, one of which is by predicting future stocks, stock prices and even exchange rates [1]. The financial revenue prediction model for the mining industry has been challenged due to the many factors that influence commodity price fluctuations [2].

Revenue in the mining industry from year to year has continued to decline, and this has an impact on large companies that have an impact on the company's performance and profits. Based on data obtained from the Organization of the Petroleum Exporting Countries (OPEC) at the end of 2014 world oil prices fell by around 45%, which previously averaged 100 dollars per barrel at the end of 2014 down to 53 dollars per barrel [3].

Through this problem the researcher aims to propose a prediction model in the mining industry using deep learning (DL). DL is one part of machine learning to predict the model that is currently developing and is widely used as a model for predicting the performance of DL-based companies [4]. This prediction model has not been widely applied to financial revenue in the mining sector. The neural network that will be developed using the long short-term memory (LSTM) model is based on supervised learning data representation [5]. In addition, this study compares with the univariate autoregressive integrated moving average model (ARIMA).

The ARIMA model is the most popular model for exponentially analyzing time series predictions. Univariate auto regressive ARIMA models can be developed more than one variable, such as time series on ISHG stocks [6]. These models are very close to predicting

DOI: 10.24507/icicel.13.11.987

time periods and time series in accordance with the trend of needs in an organization, and this application was developed and carried out in mining sector agencies for decision making and policies related to time series prediction models.

The objective of this paper is forecasting of the time series model of revenue and commodity price in mining industry, using long short-term memory (LSTM) compared to the auto-regressive integrated moving average (ARIMA). Also, the advantage of this paper enables management or head practices to make decisions and guidelines regarding time series prediction models in the mining industry. Revenue forecast in the mining industry is maximally usable.

The structure of this manuscript is as follows. Section 2 gives literature review related to the paper. Section 3 presents research methodology. Section 4 shows experiment and results. Finally, in Section 5, conclusion and future research are given.

2. **Literature Review.** Implementation of the financial revenue prediction model in the mining industry is not a new thing applied to deep learning (DL) performance. So, there is a lot of literature that discusses the results of predictions with various kinds of algorithms and methods applied to DL.

Lee et al. [4] argued that prediction models are developed to predict company performance using technical data and indicator data. The method applied uses deep belief network (DBN) with tuning training using the restricted Boltzmann machine (RBM). The input variables used are revenue growth, operating profit, net profit, gross profit, operating expense, shareholder's equity, total assets and capital adequacy ratio. The evaluation model used is root mean square error (RMSE) by evaluating the value of actual data with predictive data for 3 years. The results of studies show pre-training of RBM models and backpropagation algorithms for fine tuning similar to the results of the DBN model, and the proposed model shows 1.3-1.5 times a good decline. In addition, the results of [7] were carried out by comparing methods with 4 prediction models, namely ANN, SVM, Naïve Bayes and random forest. Relevant research on the results of this study uses research from Chen et al. [8] and Siami-Namini and Namin [6] the same characteristics also use the basis of the deep learning model, one by comparing between ARIMA and LSTM, while [8] used a hybrid crude oil price based comparison with deep learning, namely ARMA, DBN and LSTM. These results indicate the performance of the model is very sensitive from the parameters used, as in the statistical model is $MSE_{x10\text{-}4}$, CW_{RW}, CW_{ARMA}.

3. **Research Method.** In the research methodology, it is carried out through a step of research methodology for building deep learning schemes in revenue prediction models. The step of research methodology is shown in Figure 1.

The stages in the roadmap are divided into 3 parts, namely, preprocessing, learning, evaluation and final prediction results. In the preprocessing process, it explains the initial steps at the stage of the research methodology by determining, tidying random and messy

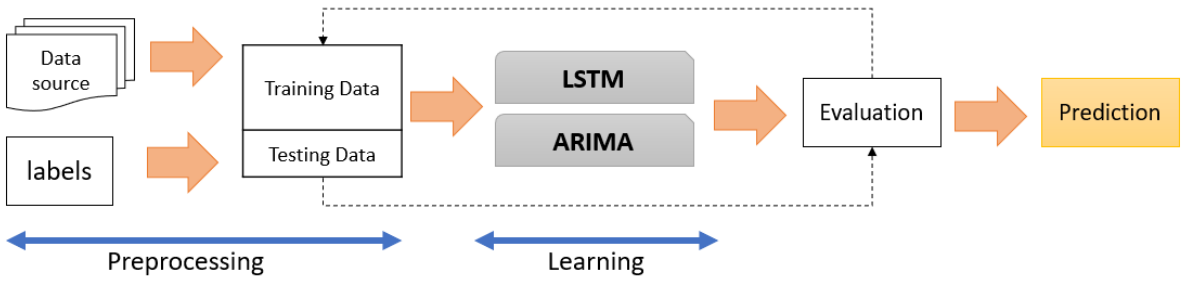


FIGURE 1. Step of research methodology: preprocessing, learning, evaluation and prediction

data through several processes including feature selection, feature extraction and scaling and data sampling. Second process of learning/learning at this stage the process of training and testing the learning algorithm begins with the model selection, cross-validation, performance metrics or hyperparameter optimization using two models, LSTM and ARIMA. In final process of evaluating, it uses the root mean square error (RMSE) and mean absolute error (MAE) models to determine which results are best accurate.

3.1. **Data source.** In this study the data are obtained during the last 6 years each month, starting from January 2012 to December 2017. The data source is divided into 2 parts, first the data comes from 3 samples of companies engaged in the mining industry processed by the Ministry of Energy and Mineral Resources and the World Bank, both of data sourced from data market and WTI crude oil market for commodity prices. The observation results have a flat on commodity prices totaling 72, with 3 types of commodity prices ranging from 216 data sets. In financial revenue, there are 216 data sets divided into 3 mining industry sample companies. On revenue it is classified into 3 parameters: revenue royalties, sales of mining revenue and revenue tax.

This study shows the parameters of the revenue and indicators on commodity prices divided into 3 categories, namely revenue for coal, gold and petroleum mining. Likewise, commodity prices are divided by the same 3 categories. The algorithm model used uses one-step forecast, that is, with the model estimating a training data set and then one step forecast calculated on the remaining data sets.

3.2. **Pre-processing with normalization.** Normalization or what is called standardization is a predictive model step in deep learning through the process of pre-processing. Normalizing the conversion of the original data into 0 and 1, in practice using MinMaxScaler by using the learn scikit library to scale fit 0 and 1.

The results of the normalization scale in Table 1 show that between normalization variables in commodity prices is data that was previously done to be produced into learning outcomes of the neural network. This min-max scale is widely used in addition to this case it is also used for stock prices, such as the denormalization results, namely: prediction = output (data range) + minimum. Here output represents a neural network with a scale (0.1), the data range represents the range of values of the number of original attributes and the minimum is the range of the smallest attribute value [1].

Table 1. Description of MinMaxScaler's normalization of commodity prices

Coal Price	Gold Price	Oil Price
0.610236	0.939821	0.846503
0.580709	0.918054	0.984499
0.513780	0.930858	0.842344
0.470472	0.935980	0.805293
0.446850	1.000000	0.762571

4. Experiment and Results.

4.1. **LSTM model.** Long short-term memory (LSTM) is a type of RNN proposed by [9]. LSTM is popular among other types of RNN because it does not experience exponential loss or errors that grow through long time serial data, unlike other types of RNN [10]. Explanation of deep learning architecture is an analysis of mining revenue predictions which are sequential data. Sequential data means here is $T + 1$, the current time (t) by predicting the value at the next time in the order ($t + 1$), and using the current time (t), and twice before ($t - 1$ and $t - 2$) as an input variable.

Table 2 shows the characteristics and specifications of the proposed method, namely the long short-term memory (LSTM). The LSTM method has an architecture with a hidden layer 3 input layer of revenue and commodity price neurons in each $t-1$ variable in the revenue prediction time series. The characteristic an optimization is Adam specification, and weight initiation of this characteristic is Randomize. The data is subdivided into epoch 100 in 80% Training (2012-2016) and 20% Test (2016-2017). This study tries to do an in-depth comparison of the neural network in LSTM by performing the training value with 3 hidden layers for revenue and commodity prices of neurons. These neurons have a block component that contains gate elements to set status and output of blocks. These blocks operate based on the input sequence, and each gate in the block uses a sigmoid activation unit to control whether they are triggered or not.

Table 2. Characteristics and specifications of the LSTM method

Characteristics	**Specifications**
Architecture	1 input layer
Neuron hidden	3 input layer (revenue) 3 input layer (price)
Optimization	Adam
Weight initiation	Random
Training data set	80%
Test data set	20%
Activation output	Linear
Epoch	100

Therefore, this paper must justify this result [6] by comparing only 4 hidden neurons, whether the accuracy is best compared to other neurons. These results can then be explained in the following Table 3.

Table 3. Hyperparameter comparison of RMSE evaluations on LSTM

Indicator	**Neuron Layer**	**RMSE**	
		Train	**Test**
Mining Coal	2	0.17	0.10
	4	**0.17**	**0.09**
	6	0.18	0.14
	8	0.18	0.14
Mining Gold	2	0.12	0.11
	4	**0.12**	**0.11**
	6	0.13	0.11
	8	0.13	0.11
Oil	2	0.18	0.10
	4	**0.18**	**0.09**
	6	0.18	0.10
	8	0.18	0.10

Table 3 shows a hyperparameter comparison of the evaluation of several hidden neurons by training and testing using the LSTM model. The hyperparameters used are hidden neurons and the same batch size, 2, 4, 6, and 8 neurons. These results show that the difference in determining the RMSE score is convergent. The best results are indicated by convergent values of hidden 4 and batch size 4. The results of the RMSE assessment summary of the LSTM method are also shown in Table 4.

TABLE 4. Performance evaluation predictions of revenue using LSTM method

Revenue	Epoch	RMSE		MAE	
		Train	Test	Train	Test
Mining Coal	100	0.17	0.09	0.18	0.19
Mining Gold	100	0.12	0.11	0.27	0.29
Oil	100	0.18	0.09	0.31	0.32
Average		**0.16**	**0.10**	**0.25**	**0.27**

TABLE 5. Performance evaluation predictions of commodity price using LSTM method

Commodity Price	Epoch	RMSE		MAE	
		Train	Test	Train	Test
Coal Price	100	0.04	0.14	0.18	0.33
Gold Price	100	0.07	0.10	0.24	0.27
Oil Price	100	0.06	0.05	0.22	0.19
Average		**0.06**	**0.10**	**0.21**	**0.26**

In Table 4, it can be seen the results of performance evaluations for the mining industry revenue, indicating that the evaluation of RMSE and MAE is through the results of training and testing data with an average comparison on RMSE 0.16 and 0.10, while at MAE 0.25 and 0.32. the parameters used are 4 neurons, epoch 100 and batch size 4.

Table 5 shows the results of the performance evaluation for the commodity prices of the mining industry, indicating the evaluation of RMSE and MAE through the results of training and testing data with an average comparison on RMSE 0.06 and 0.10, while at MAE 0.21 and 0.26. The parameters used are 4 neurons, epoch 100 and batch size 4. From both Tables 4 and 5, RMSE can be proven with standard errors smaller than MAE, and these results can be seen in revenue predictions and commodity prices with the LSTM method in Figure 2.

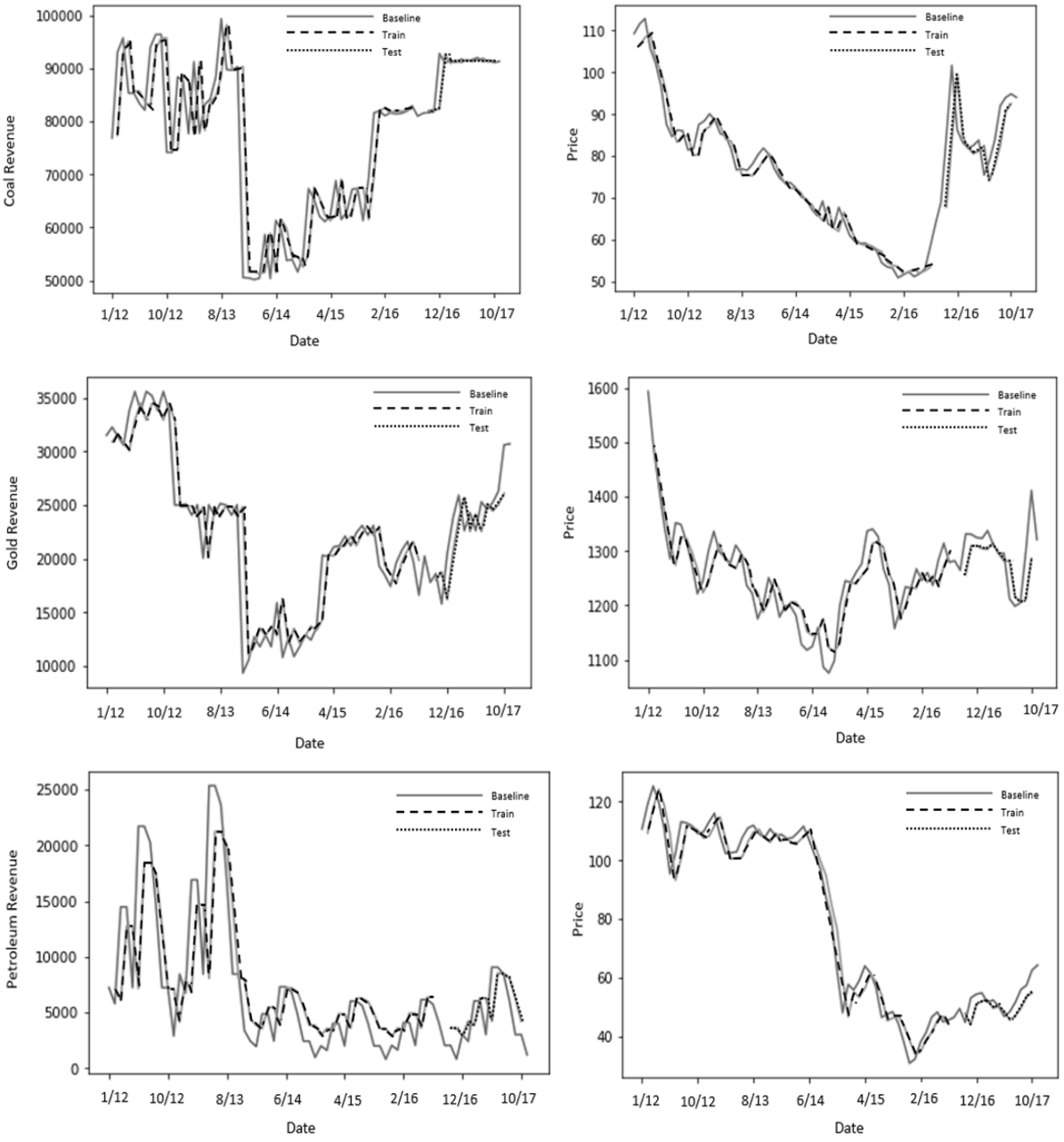
4.2. **ARIMA model.** Autoregressive integrated moving average (ARIMA) model is one of the common models of autoregressive moving average/ARMA (p, q) by combining the autoregressive/AR (p) and moving average/MA (q) processes by building a combined model for the time model series [11]. Then if combined the ARIMA model is captured with initial elements (p, d, q) or $(p, 0, q)$ as in the following formula:

$$x_t = c + \sum_{i=1}^{p} \emptyset_i x_{t-i} + \varepsilon_t + \sum_{i=0}^{q} \emptyset_i \varepsilon_{t-i} \tag{1}$$

Parameter of Equation (1) from the formula of p and q is part AR and MA command. The ARIMA prediction model in the formula is described, x_t is a general average, c a constant, $\emptyset_i x_{t-i}$ autoregressive parameter to p, $\emptyset_i \varepsilon_{t-i}$ moving average parameter to q [6].

The fit model in the ARIMA algorithm in this study used $(5, 1, 0)$ fitting model. This sets the lag value to 5 for autoregression, uses a difference order of 1 to make the time series stationary and uses a moving average model of 0. Data split is the same as that used by LSTM model which is 80% training data and 20% testing data tested on that data. The prediction model on ARIMA is the fit forecase model. The evaluation of the results of training and testing also uses the root mean squared error (RMSE) and mean absolute error (MAE) evaluation models.

The results are obtained from the ARIMA fit model $(5, 1, 0)$ with a coefficient of 0.1278, standard error 0.190 and invert AR root 0.50 MA root 1.00. Then calculate the RMSE error score in the mining industry revenue prediction model by configuring the comparison



(a) (b)

FIGURE 2. (a) Prediction of coal, gold and petroleum mining revenue; (b) prediction of coal, gold and petroleum prices using LSTM

between the ARIMA and LSTM methods to find out how much the difference in accuracy is significant to compare.

Table 6 shows the comparison of testing results using the RMSE evaluation model for revenue predictions in the mining industry between the LSTM and ARIMA methods. It can be seen that the RMSE evaluation on LSTM shows that the average standard error of LSTM is lower than ARIMA.

5. **Conclusion and Future Research.** The combination model between the results of RMSE evaluation with hidden layer 4 using epoch 100 parameters, the number of 4 batch size and 4 layer neurons produces the value of RMSE and MAE with the LSTM neural network architecture in predicting performance evaluations for coal mining, gold and petroleum mining with evaluation RMSE test 0.09, 0.11 and 0.09 as well as for

TABLE 6. Comparison of the results of testing the best evaluation model with RMSE between the LSTM and ARIMA methods

Indicator of Revenue	**RMSE**		**Indicator of Commodity Price**	**RMSE**	
	LSTM	**ARIMA**		**LSTM**	**ARIMA**
Coal Price	0.09	0.23	Coal Price	0.14	0.64
Gold Price	0.11	0.55	Gold Price	0.10	0.45
Oil Price	0.09	0.24	Oil Price	0.05	0.16
Average	**0.10**	**0.34**	**Average**	**0.10**	**0.42**

evaluation performance on commodity performance by evaluating the RMSE test 0.14, 0.10 and 0.05. Hyperparameter shows the results of RMSE evaluation on LSTM with hidden layer 4, batch size 4, epoch 100, Adam Optimizer is convergent value. ARIMA prediction model $(5, 1, 0)$ with a ratio of 0.1 coefficients and a standard error below 1, p value is below 0, this indicates that the LSTM model has better accuracy compared to ARIMA with an average error of 0.24 on revenue indicators and 0.32 on the price indicator commodity. The results of the RMSE evaluation on LSTM show that the average LSTM error standard is lower than ARIMA with the average comparison on RMSE 0.06 and 0.10, while at MAE 0.21 and 0.26. In future work, we will continue the research to explore another deep neural network method, also exploring with any variables of this paper in mining industry with enrich predictive of regression and sequence learning model.

Acknowledgment. This research was supported by Bina Nusantara University AI Research Center for the computing facility.

REFERENCES

[1] D. T. Larose and C. D. Larose, *Data Mining and Predictive Analytics*, 2015.
[2] D. Liang, C.-F. Tsai and H.-T. Wu, The effect of feature selection on financial distress prediction, *Knowledge-Based Syst.*, vol.73, pp.289-297, 2015.
[3] C. Baumeister and L. Kilian, Understanding the decline in the price of oil since June 2014, *J. Assoc. Environ. Resour. Econ.*, vol.3, no.1, pp.131-158, 2016.
[4] J. Lee, D. Jang and S. Park, Deep learning-based corporate performance prediction model considering technical capability, *Sustain.*, vol.9, no.6, 2017.
[5] H. Sak, A. Senior and F. Beaufays, Long short-term memory recurrent neural network architectures for large scale acoustic modeling, *Interspeech*, pp.338-342, 2014.
[6] S. Siami-Namini and A. S. Namin, Forecasting economic and financial time series: ARIMA vs. LSTM, *arXiv:1803.06386*, 2018.
[7] J. Patel, S. Shah, P. Thakkar and K. Kotecha, Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques, *Expert Syst. Appl.*, vol.42, no.1, pp.259-268, 2015.
[8] Y. Chen, K. He and G. K. F. Tso, Forecasting crude oil prices: A deep learning based model, *Procedia Comput. Sci.*, vol.122, pp.300-307, 2017.
[9] S. Hochreiter and J. Schmidhuber, Long short-term memory, *Neural Comput.*, vol.9, no.8, pp.1735-1780, 1997.
[10] F. A. Gers, N. N. Schraudolph and J. Schmidhuber, Learning precise timing with LSTM recurrent networks, *J. Mach. Learn. Res.*, vol.3, pp.115-143, 2002.
[11] M. H. Pesaran, *Time Series and Panel Data Econometrics*, Oxford University Press, 2015.