Stock Price Forecasting on Telecommunication Sector Companies in Indonesia Stock Exchange Using Machine Learning Algorithms

Jimmy H. Moedjahedy
Computer Science Department
Universitas Klabat
Airmadidi, Indonesia
jimmy@unklab.ac.id

Reymon Rotikan
Computer Science Department
Universitas Klabat
Airmadidi, Indonesia
reymonr@unklab.ac.id

Wien Fitrian Roshandi
IT Department
STIK Muhammadiah Pontianak
Pontianak Indonesia
wien@stikmuhptk.ac.id

Joe Yuan Mambu
Computer Science Department
Universitas Klabat
Airmadidi, Indonesia
joeyuan.mambu@unklab.ac.id

2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS) | 978-1-7281-7257-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICORIS50180.2020.9320758

***Abstract*—Stock investment is a demand-driven and demanding monetary practice. Hence, the study of stock forecasts or, more precisely, the forecasting of stock prices, plays an essential role in the stock market. Mistakes in forecasting share prices have a significant impact on global finance; thus, they require an effective method of predicting changes in share prices. Machine learning is one of the methods that can be used to predict the stock price. To predict the stock price of five companies in the telecommunications sector, Bakrie Telecom Tbk (BTEL), PT. XL Axiata Tbk (EXCL), PT. Smartfren Telecom Tbk (FREN), PT. Telekomunikasi Indonesia Tbl (TLKM), and PT. Indosat Tbk (ISAT), two algorithms are used to predict the stock prices, which are the Gaussian Process and SMOreg and train dataset from January 1, 2017, to December 31, 2019. The result of this study is SMOreg has the best result than the Gaussian Process with an RMSE value of 0.00005, MAPE 1.88%, and MBE 0.00025.**

***Keywords*—Forecasting, Stock Price, Gaussian Process, SMOreg, Telecommunication Sector**

I. INTRODUCTION

The stock is an investment that develops year after year in Indonesia; due to the number of transaction frequencies from the previous year, there is an increase [1]. According to [2], stock investment is an in-demand and challenging monetary activity, so the analysis of predictions or forecasting of stocks is more specifically the share price, playing an essential role in the stock market. The inaccuracy in predicting the stock price has a massive impact on global finance; therefore, it is necessary to evaluate risk management in predicting the stock value to make the right decisions.

Because of stock market volatility, though, it is difficult to understand precisely the rule of changes in the stock market or how the share price shifts. So, this makes it hard for the government to effectively regulate the financial market at the right moment. Moreover, people usually rely on discretionary valuations to perform stock trading and ignore effective methods to assist in decision - making. The purpose of the research we did was to implement several machine learning algorithms to forecast stock price changes and then compare which algorithm has produced the best results. This is achieved so that equity traders or buyers have a guide for making stock market trades.

II. LITERATURE REVIEW

Some studies that have the same topic as this research are conducted by [4] Where four deep learning architectures are used, that is, MLP, RNN, LSTM, CNN to predict the share price based on existing stock history data from the company. The result is that CNN outperforms the other three deep learning algorithms in predicting stock prices. Furthermore, research conducted by [5] that use RNN in predicting the share price, but before using Principle Component Analysis (PCA) to reduce the number of existing features. The result is that if you have used PCA and run RNN, then the predicted results are better. Research by [6] also related to the topic of predicting time-series data to compare Bayesian Forecasting and ARIMA in predicting time series data; the result is Bayesian Forecasting outperforms ARIMA in predicting time series data.

Subsequently, the research conducted by [7] that use RNN-LSTM to predict future share prices produced satisfactory results by applying RNN-LSTM to forecast the two companies' shares. [8] Similar studies used ANN and Random Forest to forecast the trading price of shares on the next day for five companies, with ANN findings providing better predictive value than Random Forest. Furthermore, when [3] they propose a new method by combining methods or hybrid structures such as LSTM, PSO, EWT, ORELM, the consequence is that the proposed hybrid system has a better value than traditional methods.

Another study with similar topics is also covered by [9] by making stock pricing forecasts or when to sell and buy shares of WTI crude oil utilizing vector regression, random forest, XGboost, light gradient boosting computer, neural networks, and genetic algorithm methods. XGBoost and LightGBM are, thus, models that deliver the best results from the many techniques used. They predict stock value movement based on past and present data using Fuzzy Time-Series-based models coupled with PLR and SVR. The result is that the Fuzzy Time - Series model is more sensitive than PLR and SVR in predicting stock movements and estimating when to make the right decision to sell or buy shares faster.

Research conducted by [11] to model FRS and assess the sufficient interval length using the methods FTSF-DBN, FTSF-LSTM, and FTSF-SVM then compared them with literature papers from other scholars. As a result, FTSF-SVM is most effective in modelling FRS and determining the interval duration. [12] Similar research has also been conducted to predict stock market price shifts using Adaboost computing devices, XGboost, GBDT, decision tree, and public-opinion logistics regression. The consequence is that the GDBT algorithm is more accurate in predicting a 64.555 % accuracy shift in stock value. [13] also conducted research that tried to forecast stock prices using the Fuzzy Time Series and coupled with the ICA Algorithm (Imperialist Competitive Algorithm) with the effects of the implementation of the Fuzzy Time Series alone produced good results. But if the results obtained are more robust when combined with the ICA algorithm, then.

Research conducted by [14] to integrate adaptive network-based fuzzy inference system (ANFIS) with an Empirical Model of Decomposition (EMD) in predicting the value of shares in the Taiwan Stock Exchange Capitalization Weighted Stock Index (TAIEX) and The Hang Seng Stock Index (HSI) showed that EMD-ANFIS was more significant than other research results utilizing different methods. Finally, the Shanghai Stock Exchange Composite Index and Nasdaq Composite Index, and S&P 500 Composite Stock Price Index are research carried out by [15] to predict the stock market value of indices using the architectural simulation of the Logistic Regression (LR) model and gradient boosted decision trees (GBDT) models in two different stock indices. The analysis of LR, GBDT, SVM, NN, and TPOT models with LR2GBDT on stock indices forecasts reveals better accuracy for LR2GBDT

III. METHOD

In doing this research, we used two machine learning algorithms, namely Gaussian Process (GP) and SMOreg. We use these algorithms to predict stock price changes by comparing results after choosing which algorithm is best at predicting. For the implementation of these algorithms, we use the Weka software version 3.8.4.

Fig. 1. is the flow of work that we have performed on this research. The research process begins by conducting literature reviews to look for subjects, types of companies, and algorithms. The business is a telecommunications company based in Indonesia. There are six firms in this market, but we only use five companies, since only one company is established in 2019 because the historical data is already low. Bakrie Telecom Tbk (BTEL), PT are the five companies. Axiata Tbk XL (EXCL), PT. Smartfren Telecom Tbk, PT. Telekomunikasi Indonesia (TLKM) and PT. Tbk Indosat (ISAT). The historical data set of share price data collected from January 1, 2017, to December 31, 2019. The dataset that we use is obtained from the Yahoo Finance website by extracting the historical data given. The dataset is still in HTML format, to be used in the training process, the dataset is converted to a CSV file. After that, we do the data-cleaning process, then split the testing and check the data to 70:30. Table I shows the size of the company based on the number of employees in each telecommunication sector company in Indonesia

The next step is that each dataset is evaluated using two algorithms, GP and SMOreg. The implications of forecasting the share price for the next ten days will be borne by each algorithm. The above is focused on the results of the analysis, which decides which algorithm is better suited to predict the share price.

TABLE I. COMPANY SIZE IN TELECOMMUNICATION COMPANY

No	Code	Company	Size
1.	BTEL	Bakrie Telecom	67
2.	TLKM	Telkom	24.722
3.	ISAT	Indosat	3.406
4.	EXCL	Exel	1.616
5.	FREN	Fren	2.713

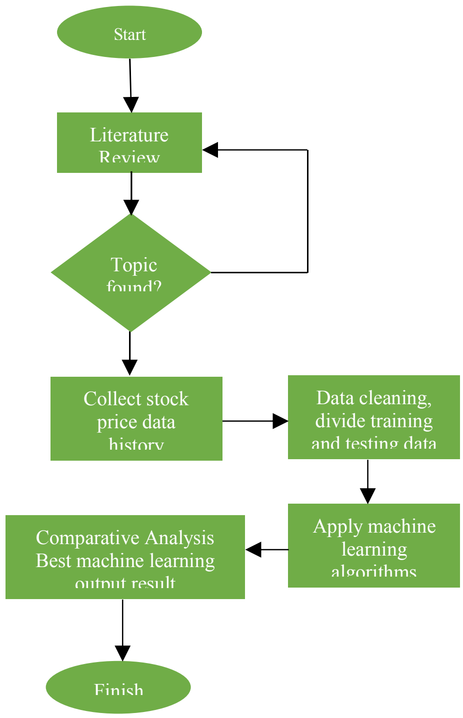

Fig. 1. Research flow of work that we have performed

IV. RESULT AND DISCUSSION

After applying two algorithms to five telecommunications sector companies with historical data from January 1, 2017, to December 31, 2019, the prediction of share prices over the next ten days can be seen in examples in Fig. 2. In this picture, the scenario used compares the price of the existing stock that has been withdrawn with the anticipated share price.

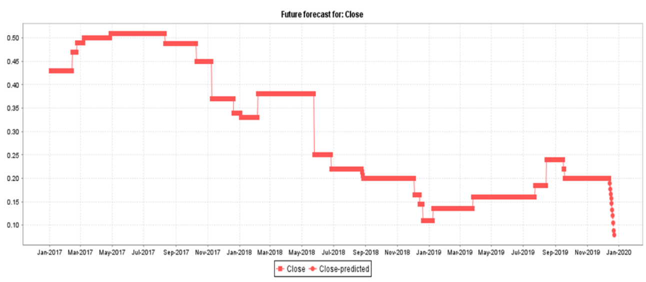

Fig. 2. Ten days forecast result

Fig.3 is a graph of the estimated value of the stock price and its original value using the Gaussian process. The graph's findings and the illustration of one of the reference values in Table II show that both the forecast results and the market price have a value that can be assumed to be similar. With the lowest forecast difference of 0.0024 on December 15, 2019, the highest difference was 0.3461 on December 23, 2019.

TABLE II. FORECAST RESULT DATA EXAMPLE

Date	Forecast	Original
12/14/19	4.6075	4.55
12/15/19	4.6303	4.64
12/16/19	4.6576	4.66
12/17/19	4.6635	4.61
2 12/18/19	4.7001	4.58
12/19/19	4.717	4.55
12/20/19	4.7371	4.705
12/21/19	4.7568	4.705
12/22/19	4.7783	4.51
12/23/19	4.8061	4.46

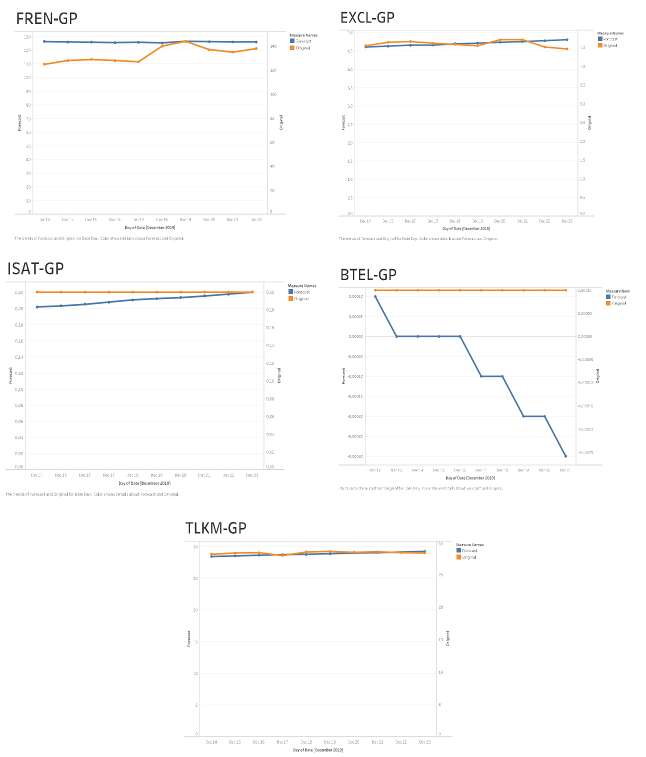

Fig. 3. Forecast vs. Original stock price using Gaussian Process

From Fig. 3. on the Gaussian Process as well, it can be seen that the product of the Gaussian Process method against the TLKM stock has a nearly perfect graph as well as an EXCL graph. In addition, the findings of FREN and ISAT also indicate results that are similar to. The outcome is slightly different for BTEL.

Fig. 4. is a comparison of the expected value of the share price and its original value utilizing SMOreg. The graph's findings and the illustration of one of the reference values in Table I indicate that the real predictive value is not very significant. There tends to be a difference from the map due to only ten days. Based on Fig. 4. it can be seen that TLKM and ISAT, as well as FREN and EXCL, have almost perfect graphics. BTEL also showed very similar results, but on the tenth day, there was a gap in results.

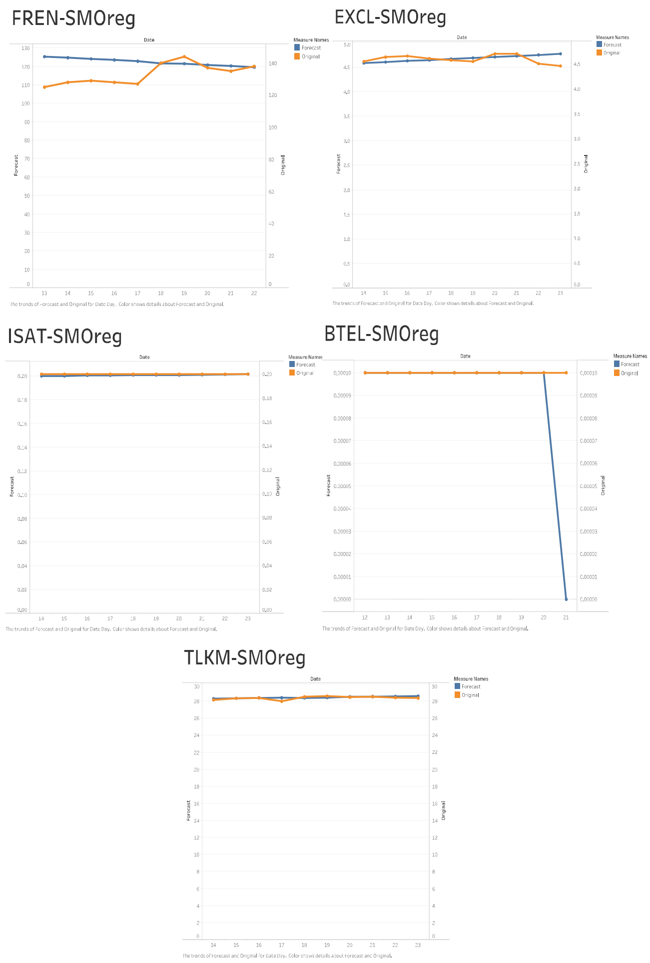

Fig. 4. Forecast vs. Original stock price using SMOreg

The comparative analysis of the Gaussian process and SMOreg reveals that SMOreg has the best results by having smaller RMSE, MAPE, and MBE values. Table III shows that the value of RMSE of all companies using SMOreg is the lowest value. Like MAPE, it can be demonstrated that SMOreg also has the lowest value. MBE also has a low value compared to the Gaussian process. Based on observations, the lowest values of RMSE, MAPE, and MBE are 0.00005, 1.88, and 0.0002.

TABLE III. RMSE, MAPE, MBE COMPARATIVE ANALYSIS

Company	Gaussian Process			SMOreg		
	RMSE	MAPE	MBE	RMSE	MAPE	MBE
BTEL	0.00007	39.95%	0.00025	0.00005	23.01%	0.00025
FREN	9.01867	7.48%	16.4078	7.43586	4.76%	15.91228
ISAT	0.01362	5.17%	0.01861	0.0047	2.008%	0.01398
EXCL	0.18403	4.61%	0.23886	0.16617	4.11%	0.21999
TLKM	0.66623	2.34%	0.84403	0.52836	1.88%	0.71132

V. CONCLUSION

Based on the calculation and comparative analysis, it can be concluded from this analysis that SMOreg has a higher and more precise value as it has the lowest values of 0.00005, 1.88 and 0.00025 from the comparative values of RMSE, MAPE and MBE for the telecommunications companies using the Gaussian Method and SMOreg.

ACKNOWLEDGMENT

The author thanked MTI AMIKOM Yogyakarta for the assistance provided by Dr. Arief Setyanto M.Kom in completing this research. It also offered moral and financial support to Universitas Klabat and STIK Muhammadiah Pontianak.

REFERENCES

[1] IDX, "IDX Annual Statistics," Jakarta, 2019.

[2] C. H. Cheng and J. H. Yang, "Fuzzy time-series model based on rough set rule induction for forecasting stock price," Neurocomputing, 2018, doi: 10.1016/j.neucom.2018.04.014.

[3] H. Liu and Z. Long, "An improved deep learning model for predicting stock market price time series," Digit. Signal Process. A Rev. J., vol. 102, p. 102741, 2020, doi: 10.1016/j.dsp.2020.102741.

[4] M. Hiransha, E. A. Gopalakrishnan, V. K. Menon, and K. P. Soman, "NSE Stock Market Prediction Using Deep-Learning Models," Procedia Comput. Sci., vol. 132, no. Iccids, pp. 1351–1362, 2018, doi: 10.1016/j.procs.2018.05.050.

[5] Z. Berradi and M. Lazaar, "Integration of Principal Component Analysis and Recurrent Neural Network to Forecast the Stock Price of Casablanca Stock Exchange," Procedia Comput. Sci., vol. 148, pp. 55–61, 2019, doi: 10.1016/j.procs.2019.01.008.

[6] R. Nariswari and H. Pudjihastuti, "Bayesian forecasting for time series of count data," Procedia Comput. Sci., vol. 157, pp. 427–435, 2019, doi: 10.1016/j.procs.2019.08.235.

[7] A. Moghar and M. Hamiche, "Stock Market Prediction Using LSTM Recurrent Neural Network," Procedia Comput. Sci., vol. 170, pp. 1168–1173, 2020, doi: 10.1016/j.procs.2020.03.049.

[8] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock Closing Price Prediction using Machine Learning Techniques," Procedia Comput. Sci., vol. 167, no. 2019, pp. 599–606, 2020, doi: 10.1016/j.procs.2020.03.326.

[9] C. F. Ivașcu, "Option pricing using Machine Learning," Expert Syst. Appl., vol. 163, p. 113799, 2021, doi: 10.1016/j.eswa.2020.113799.

[10] P. C. Chang, J. L. Wu, and J. J. Lin, "A Takagi-Sugeno fuzzy model combined with a support vector regression for stock trading forecasting," Appl. Soft Comput. J., vol. 38, pp. 831–842, 2016, doi: 10.1016/j.asoc.2015.10.030.

[11] S. Panigrahi and H. S. Behera, "A study on leading machine learning techniques for high order fuzzy time series forecasting," Eng. Appl. Artif. Intell., vol. 87, no. August 2019, p. 103245, 2020, doi: 10.1016/j.engappai.2019.103245.

[12] J. Yang, C. Zhao, H. Yu, and H. Chen, "Use GBDT to Predict the Stock Market," Procedia Comput. Sci., vol. 174, no. 2019, pp. 161–171, 2020, doi: 10.1016/j.procs.2020.06.071.

[13] H. J. Sadaei, R. Enayatifar, M. H. Lee, and M. Mahmud, "A hybrid model based on differential fuzzy logic relationships and imperialist competitive algorithm for stock market forecasting," Appl. Soft Comput. J., vol. 40, pp. 132–149, 2016, doi: 10.1016/j.asoc.2015.11.026.

[14] L. Y. Wei, "A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting," Appl. Soft Comput. J., vol. 42, pp. 368–376, 2016, doi: 10.1016/j.asoc.2016.01.027.

[15] F. Zhou, Q. Zhang, D. Sornette, and L. Jiang, "Cascading logistic regression onto gradient boosted decision trees for forecasting and trading stock indices," Appl. Soft Comput. J., vol. 84, pp. 1–13, 2019, doi: 10.1016/j.asoc.2019.105747.