Journal of Theoretical and Applied Information Technology
28th February 2022. Vol.100. No 4
2022 Little Lion Scientific

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

COMPARISON STOCK PRICE PREDICTION BETWEEN ARIMA, MULTIPLE LINEAR REGRESSION AND LSTM MODELS BY ADDING STOCK SENTIMENT ANALYSIS AND USD/IDR FLUCTUATION

[1]CHRYSMIEN, [2]RIYANTO JAYADI

[1,2] Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management,

Bina Nusantara University, Jakarta, Indonesia

E-mail: [1]chrysmien@binus.ac.id, [2]riyanto.jayadi@binus.edu

ABSTRACT

Since the COVID-19 pandemic, the number of new investors entering the Indonesia capital market has significantly increased. Director of Index Harga Saham Gabungan (IHSG) said the pandemic COVID-19 also brought several new achievements in the capital market, including in terms of the number of investors, market capitalization, to volume, frequency, and investment value. Autoregressive Integrated Moving Average (ARIMA) is a model that is commonly used to predict the movement of a stock. However, there are also some limitations of ARIMA, ARIMA model can no longer accommodate when there is a sharp spike or drops in prices and if ARIMA model used for a long time, the forecast results will be constant. This research was conduct on FREN's stock price which recently has a high trend and the price is strongly influenced by public sentiment. In this study, researchers will compare the Multiple Linear Regression, LSTM model using proposed model with ARIMA to give new insight from business side and stock investors to give a better decisions in stock investment strategies. The results show that ARIMA which predicts stock movements based on historical data alone cannot predict FREN stocks when there is a sharp spikes of stock prices.

Keywords: *Stock Price Prediction, ARIMA, Multiple Linear Regression, LSTM, FREN*

1. INTRODUCTION

Stocks have always been an investment instrument with high returns in Indonesia. The average profit of each investor was about 12 to 14 percent per year, depending on the performance of Index Harga Saham Gabungan (IHSG) However, in fact, there are many investor also only get little profit or even not get the maximum return from stock shares. Most of them have to incur huge losses due to careless investment strategies and only capital, based on survey. [1]

The COVID-19 pandemic has accelerated digital transformation in all lines of life, including the financial services industry in Indonesia. This digital acceleration has changed people's behavior in everyday life, including in terms of financial transactions and investing. Before the pandemic, offline financial transactions (face to face physically) [2]. It is believed that the growth of the mutual fund industry will continue to grow positively in 2021 in line with the condition of the national economy which is starting to recover from the crisis due to the COVID -19 pandemic[3].

COVID-19 pandemic also giving impact to the world by investors entering the capital market has significantly increased. The capital market is an alternative investment that is widely eyed. The growth in the number of investors will further increase capital market transactions. On the other hand, the Director of IHSG said the pandemic has also brought several new achievements in the capital market, including in terms of the number of investors, market capitalization, to the volume, frequency, and value of stock exchange transactions, in fact, the growth trend from year to year is quite significant according to survey [4], Figure 1 shows a graph of the growth of stock investors from 2012 until 2021.

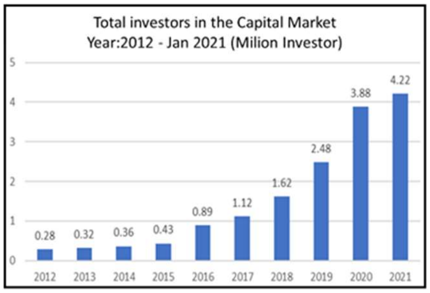

Figure 1 Total investors in the capital market

Autoregressive Integrated Moving Average (ARIMA) model is one model that is widely used to predict the movement of a stock. ARIMA model is the combination of two methods, namely Autoregressive and Moving Average. ARIMA looks for the most suitable pattern in the unit of time by using historical values and predicts future patterns but in the short term. ARIMA is suitable if the variable in the time series statistically related to each other (dependent). However, there are also some limitations of ARIMA, Majumder [5] conduct a research that shown First ARIMA model can no longer accommodate the occurrence of sharp price spikes or declines (in this case was a collapse market). The second one is if ARIMA model used for a long time, then the results of the forecast will be constant.

On the other hand, public sentiment is known as one factors that can affect stock prices, the negative sentiment will be reducing stock prices and also positive sentiment will raise the stock price. Research conducted by Valle-Cruz [6] shows that Financial sentiment analysis allows us to understand the effect of social media reactions and emotions on the stock market price.

Exchange rate movements are also considered as one of the factors that can affect stock prices Lee and Ryu [7], research says that there is an empirical relationship that foreign investors play a positive role in emerging markets as liquidity-providers. Research conducted by Bagh [8] shows that exchange rate fluctuation / volatility is considered to be the most important and persuasive variable that affects the performance of stock price, The results of the study found that there is positive and statistically significant relationship between Exchange Rate Volatility and Stock Index but in Pakistan Index.

In current era of technology, Artificial intelligence-based stock price prediction can identify relationships and patterns in the variables, offering better results than traditional statistics according to Maqsood et al [9], Machine Learning (ML) and Deep Learning (DL) methods are developed to evaluate the prediction power in the stock markets. The ML algorithms that are implemented to figure out patterns of data, measure the investment risk, or predict the investment future, the power of ML strategies in addressing the stock market prediction problem. For the DL technology is an artificial simulation of the human brain's abstract and figurative capabilities. It has distributed storage, self-organization, parallel processing and self-tuning capabilities, based on previous characteristics of the neural network make it more suitable for dealing with complex nonlinear problems with multiple influencing factors, instability and random types.

Based on the discussion before, it can be concluded that the most popular model that used to predicts stock price was ARIMA and Artificial Intelligence, even though both of them have their advantages and disadvantages according to Ma et al research [10]. Predicting the stock prices always been a hot topic of research, There is plenty of techniques that forecast the stock price the first technique used the historical data only to forecast the stock price, this techniques was well known as technical analysis [11], the second technique is known as fundamental analysis, this technique uses qualitative and quantitative measurement based on current conditions of the Initial Public Offering (IPO) company [12].

Almost all the previous works about stock prediction that only using the historical data, commonly compares model of ARIMA, Linear Regression, ANN, LSTM [10], [13]–[15] the overall result of the research shows sometimes ARIMA, Linear Regression and LSTM give better result. There also some research that combine stock prediction with sentiment analysis or Exchange Rate but not combined both of them [11], [16], [17].

Based on introduction section, the problems that can be research question in this study are represented in the following questions. First, can the

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

proposed prediction model overcome the limitation by the ARIMA model? And the second question is there a relationship between public sentiment and the dollar exchange rate on stock prices? The hypothesis based on some of literature review in introduction section, our results show that by propose a new model by adding the sentiment analysis and fluctuation of forex exchange rate will giving better result of prediction than using historical of stock time series data only.

Many studies have been conducted on the topic of stock prediction. However, not many look the topic of stock predictions with the help of other variables such as analysis of public sentiment and the movement of the USD/IDR exchange rate. But the focus of this research is to get a model with the best accuracy for stock price predictions with the help of user comments on sentiment analysis and USD/IDR exchange rate movements.

The remain of this paper is organized as follow. Section 2 that will provide overview of literature review that giving insight of model that will be proposed for sentiment analysis and predicting stock price. In Section 3 this paper will introduce the methodology of the research and step by step to implement the proposed model. Section 4 will show the result of the experiment of the proposed model and discuss the result of the experiment. Finally, Section 5 will conclude the finding of this research.

2. THEORETICAL FRAMEWORK AND RELATED REVIEW OF LITERATURE

2.1 Stock

Shares are one of the securities trading business fields in the capital market. The capital market or what is commonly referred to as the stock exchange is an activity of a private company in the form of investment. The main objective lies in the problem of capital requirements for companies that want to further advance their business by selling their shares to money owners or investors, both groups and business institutions.

The movement of stock prices can be observed from the amount of demand and supply for these shares [18]. If the supply is greater than the demand for shares, the stock price will fall, otherwise, if the demand is greater than the supply, the stock price will rise. This stock price will change at any time, namely in seconds, due to instantaneous assessments by sellers and buyers who are influenced by many factors.

2.2 ARIMA

The ARIMA model is a time series analysis that has the effective ability to generate short-term forecasts. It constantly outperformed complex structural models, Based to Box-Jenkins, there is an iterative approach consisting of the following 3 steps: ***Identification***: Use the data and all related information to help select the model that is deemed suitable. ***Estimates***: Use data to train model parameters, ***Diagnostic Checks:*** Evaluate the appropriate model in the context of the available data and examine areas where the model can be improved. The above process is an iterative process, so if new information is obtained after being diagnosed, then the process can be repeated from the beginning to adjust the best model.

2.3 Multiple Linear Regression

Basically, Multiple Linear Regression is a model that show the relationship between the dependent and independent variables (must more than one). The main purpose of using Multiple Linear Regression was to do a predictive analysis to find the business need and optimize it. [19].

2.4 Artificial Neural Network (ANN)

ANN model was a model widely used in approximating functions and predictions, one of the most significant advantages of ANN is that it is a general approximation model, which means that it can approximate many functions. Its ability comes from the parallel processing of data rather than the pre-set model [20]

2.5 Recurrent Neural Network (RNN)

The fundamental feature of a Recurrent Neural Network (RNN) is that it contains at least one feedback loop, so that activation can flow in a loop. It supports the network to do temporary things and learn sequences, such as forecasting. RNN architecture can take many forms

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

2.6 Long Short Term Memory (LSTM)

LSTM is an advance version of RNN capable of learning long term dependencies. The problem of vanishing gradient does not exist in LSTM they are capable of handling long sequences of sentences easily. Suppose a language model trying to predict the next word in a sentence based on the previous ones. If we are trying to predict the last word in the sentence say “The clouds are in the sky”. The context here was pretty simple and the last word ends up being sky all the time. In such cases, the gap between the past information and the current requirement are often bridged really easily by using recurrent neural networks. So, problems like Vanishing and Exploding Gradients which are present in don’t exist and this makes LSTM networks handle long-term dependencies easily. LSTM have chain-like neural network layer. In a standard recurrent neural network, the repeating module consists of one single function [21]

2.7 Sentiment Analysis

Sentiment analysis is an approach in looking at the expressions (emotions, attitudes, opinions and sentiments) of the text so that trends and classifications can then be analyzed. Sentiment analysis is useful in understanding sentiment in the form of large amounts of text, and in conducting sentiment analysis, text mining techniques are needed to extract information from unstructured text.

Sentiment analysis is the process of using text analytics to obtain various data sources from the internet and various social media platforms such as Twitter and Facebook. Every human being in general will express a response related to a thing or event to the input received by them. These responses can be categorized into three major groups, namely positive responses, negative responses, and neutral responses [22]. Public response or opinion plays an important role as feedback on products, services, and other topics. Social media is the right platform to express human response in this digital era.

2.8 Related Works

The authors of [11] aimed to predict the stock price movement in Indonesia based on sentiment analysis and currency exchange rate using SVM model. The reason of using these variable due stock price movement usualy influenced by a variety factor and this model is combined the technical analysis and fundamental analysis. the result of experiment giving the prediciton accuracy rate of 65.35% on average this is 11.78% higher without using the two variables.

Research in [16] tried to applying a hybrid model by applying LSTM and Sentiment Analysis, taking the technical indicators derived from the stock historical data and investor sentiment factor as fundamental input, the result of the research shows that the hybrid model based on LSTM and sentiment analysis is giving better performance in predicting the stock prices compared to the single model without using the sentiment analysis.

Meanwhile, the studiy at [13] tried to leverage stock price prediction with Artificial Intelligence (AI) strategies by using technical analysis approach of Machine Learning Regression algorithm and LSTM based on the end of the business day price, the result shows the linear regression model predict the closing price remarkably with a shallow error value in the technical analysis. Furthermore, while the linear regression can predict the closing price with a sensible range of error, it cannot precisely predict the same value for the next business day.

Ma (2020) [14] conduct to predict financial transaction data of Shanghai using LSTM Deep Neural Network model and compared it with Back Propagation (BP) Neural Network, Traditional RNN and RNN improved LSTM Deep Neural Network. The result of this research shows that LSTM deep neural network giving higher prediction accuracy and can effectively predict the stock market time series.

Another study by Ma (2020) [10],compared 3 model analysis principles of prediction (ARIMA, ANN and LSTM) in the end. It is believed that LSTM model have the best predictive ability although in this case introduces other variables not like the other 2 model and it’s really affected by the data processing, for the ANN model perform better than ARIMA model, hopefully the further development and use of LSTM model in stock price prediction can be more advance

Shi (2021) [17] proposed a new sentiment analysis system with deep neural network to forecast the stock movement price, the empirical result of this research showed that the proposed model of

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

sentiment analysis classification with deep neural network giving 9% improvement over logistic regression. The conclusion of this research tells us that emotion features extracted from comment are indeed effective to forecast stock price movement for china index.

Meanwhile, the study by [15] tried to compared the performance of time series model in prediction of price movement for 30 listed stock from Dow Johns Industrial Average (DIJA). The results of Comparison suggest that LSTM can give good estimates fit the price movement pattern well with relatively low latency, the conventional time-series models, such as ARIMA, tend to be more explainable but don't capture the trend well. Long short-term memory (LSTM) proves itself to be reliable than ANN and RNN.

Last, Majumder (2019) [5] forecasts the stock price but due finding limitation of ARIMA mode, the result of the research shows that ARIMA model not well performed when there is a sharp spikes or collapsed stock market, this research also proposed a model but not in artificial intelligence approach but c Compute Absolute Percentage Change (APC). the result of this research shows that the proposed model giving higher accuracy.

To sum up all of the previous study before related to stock prediction, there is a lot of stock prediction research like [11] use the combination of stock price and exchange rate and by using SVM approach the model gives 11.78% higher without using the two variables, while [17] research tell us that prediction with deep neural network giving much better result than using SVM alone, in other hand [10], [14]–[16] also tells that LSTM or neural network is giving better result by comparing with many models as in the research to predict the stock with or without any additional variables such as public sentiment or the exhchange rate. The contribution and the novelty of this research was using both of stock, public sentiment and exchange rates variables and this research want to compared 3 model that consist of ARIMA, Multiple Linear Regression and LSTM with combined of historical stock data. sentiment analysis and exchange rate data to see is it any correlation of data and the proposed model can outcome the limitation of ARIMA forecasting,

3. METHODOLOGY

The CRISP-DM (Cross-Industry Process for Data Mining) methodology is a methodology that used for this research, this methodology provides approaches and processes to data mining projects its application CRISP-DM goes through 6 processes that will be described in this section.

3.1 Business Understanding

The focus in this stage is to understand the business goal to from the project. At this stage, the project objectives and constraints must be understood.

The type of stock that will be selected is the stock in the telecommunications sector because since 5G technology arrived in Indonesia, based on the picture below, as in the public news FREN stock has been increasing continuously (presumably affected by the issue of public sentiment). Figure 2 and Figure 3 show that FREN's stock has increased considerably starting in June 2021, during which 5G can begin to available Indonesia.

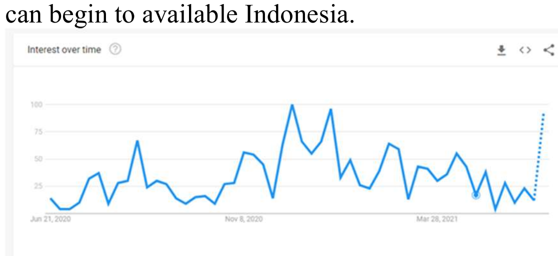

Figure 2 FREN keywords on google trends

Figure 3 FREN Stock fluctuation on google finance

To bound the scope of the problem in this research, it is necessary to limit the problem, namely as follows:

1. Stock data collection is done by downloading daily stock price data from the website www.finance.yahoo.com.
2. Exchange rate data is collected by downloading daily stock price data from the website www.bi.go.id.

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

3. The scope of this research is to examine the data on FREN stocks listed on the Indonesia Stock Exchange (IDX). For data with artificial intelligence based will be processed using Jupyter Notebook and for the ARIMA will be using Rapid Miner due faster and optimized processing.

3.2 Data Understanding

This stage consists of preparation with collecting the data and define the source of data. There are 3 categories of data shown for Table 1 for stock data, Table 2 for sentiment data, and Table 3 for forex data. The output of this stage was raw data and will be preprocessed in next step for this research, the data was collected with several technique with two types of variables.

Table 1 Stock Data Variable

	Variable Name	Description
Independent Variable	Stock Date (stock_dt)	stock trading date
	Stock Code (stock_cd)	stock code that consists of four letters.
	Open Price (open_prc)	the opening price of shares in one trading day.
	High Price (high_prc)	the highest stock price in one trading day.
	Close Price (close_prc)	Closing price of a stock in one trading day.
	Adjusted Close Price (adj_close_prc)	The price of shares on the end of trading day by adding distributions and corporate actions that occurred before the next day.
	Volume (volume)	Number of assets/shares that have been traded in a certain period.
Dependent Variable	Low Price (low_prc)	lowest stock price during the trading day.

Table 2 Sentiment Data Variable

	Variable Name	Description
Independent Variable	Sentiment Source Date (source_dt)	Sentiment date published
	Sentiment Source Type Code (sentiment_src_tp_cd)	Sentiment data source
	Sentiment Keyword Type Code (keyword_tp_cd)	keyword that used for sentiment.
	Sentiment Text (sentiment_text)	Sentiment content.
Dependent Variable	Sentiment Score (sentiment_score)	sentiment scoring with 1 = positive, 0 = neutral, -1 = negative.

Table 3 Forex Data Variable

	Variable Name	Description
Independent Variable	Forex Date (forex_dt)	exchange rate date
	Forex Type Code (forex_tp_cd)	exchange code
	Forex Description (forex_desc)	forex description
Dependent Variable	Forex Sell (forex_sell)	forex selling price.
	Forex Buy (forex_buy)	forex buying price.

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

3.3 Data Preprocessing

This stage consists of data selection, data integration, data transformation, visualize Data and data correllation check. Data preprocessing is done when we finalized the selection of data by observing the characteristic, we use the jupyter notebook to preprocess the data, also we propose all the data will be in numeric form, while the outlier data must be deleted or discarded because it will interfere with the research process. The purpose of the initial data processing is to eliminate missing values. In this study, the data (Stock, Forex, and Sentiment) are combined into a single unit using a query in PostgreSQL, The period of data was used for this research was 1 year (6/23/2020 - 6/23/2021), the total of sentiment data 988 row, for the stock data having 241 row and the exchange rate having 256 row of data,

3.3.1 Data Selection

Table 4 Variables from All Raw Data

No	Stock Variable	No	Sentiment Variable	No	Forex Variable
1	Stock Date (stock_dt)	9	Sentiment Source Date (source_dt)	14	Forex Date (forex_dt)
2	Stock Code (stock_cd)	10	Sentiment Source Type Code (sentiment_src_tp_cd)	15	Forex Type Code (forex_tp_cd)
3	Open Price (open_prc)	11	Sentiment Keyword Code (keyword_tp_cd)	16	Forex Sell (forex_sell)
4	High Price (high_prc)	12	Sentiment Text (sentiment_text)	17	Forex Buy (forex_buy)
5	Close Price (close_prc)	13	Sentiment score (1 = positive, 0 = neutral, -1 = negative)	18	Forex Description (forex_desc)
6	Volume (volume)				
7	Adjusted Close Price (adj_close_prc)				
8	Low Price (low_prc)				

From Table 4, variables 1, 9 and 14 can be represented as X1. The variables 2, 10, 11, 15 have no relevance to the research we choose to remove it for being dependent data. While variable 13 can be categorized into 3 columns: positive (X7), negative (X8), and neutral (X9) which each of the column represent count of sentiment for each day.

Table 5 Final Variables for Row Data

No	Variable	No	Variable	No	Variable
X1	Trans Date (joiner of stock_dt, sentiment_dt, forex_dt)	X7	Sentiment negative (Grouping of sentiment negative for each day)	X12	Forex Sell (forex_sell)
X2	Open Price (open_prc)	X8	Sentiment neutral (Grouping of sentiment neutral for each day)	X13	Forex Buy (forex_buy)
X3	High Price (high_prc)	X9	Sentiment positive (Grouping of sentiment positive for each day)		
X4	Close Price (close_prc)				
X5	Volume (volume)				
X6	Adjusted Close Price (adj_close_prc)				

3.3.2 Data Integration

From all the selected data, then the data is combined into a single data unit in one table named masterdataset using the PostgreSQL database. The reason of data processing with tables of database because the data is still dynamic and when there will be additional data. There will reduce the negligence that will cause data loss or corruption.

3.3.3 Data Transformation

In this step the data will be normalized to tune up the data processing, there will be 2 kind of normalization of data the first one for the numerical data that consist for the stock and exchange rate data and the second one for the text data that will consist of sentiment data.

1. **Numerical Data**
   - **Eliminate NULL Value and Change with Mean.**

Journal of Theoretical and Applied Information Technology
28th February 2022. Vol.100. No 4
2022 Little Lion Scientific

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

If in the data having a null or NaN value, we will change it to mean of the column.

- **Standardize data**
  Standardize the data at same range same with the target label that using the min-max scaler.

2. **Text Data**

- **Case Folding Data**
  The text that collected will be processed into this step:
  - Change all the words into lower case.
  - Remove all number in the text.
  - Remove all whitespaces split the text
  - Filtering all the words with sastrawi library.

3.3.4 Visualize Stock Data

As shown in Figure 4, the conclusion that can we see is when the price of stock is getting lower, the volume of transaction is increasing which mean it has negative correlation.

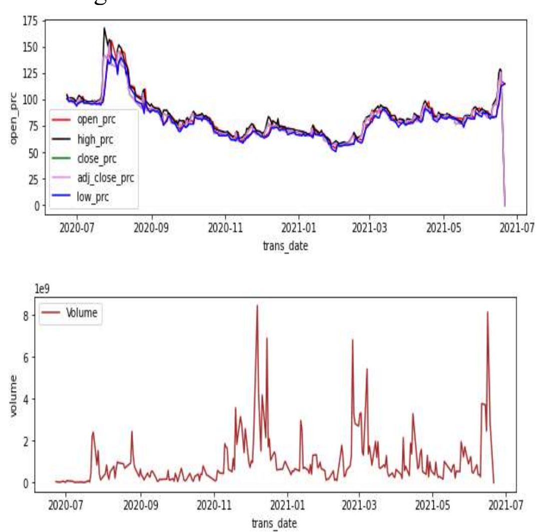

Figure 4 Visualization of Stock Data

3.3.5 Visualize Sentiment Data

As shown in Figure 5, as we can see the distribution of FREN sentiment data is quite a lot for each month due the high trend of FREN. The total data of Sentiment was 925 data with range of 1 year period.

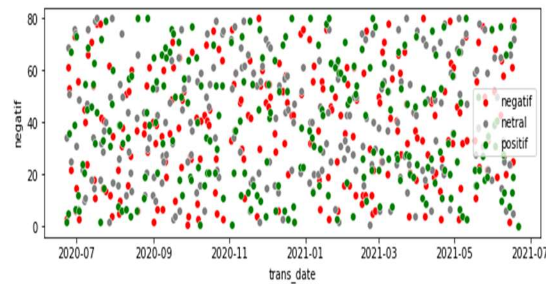

Figure 5 Visualization of Sentiment Data

3.3.6 Visualize Exchange Rate Data

As shown in Figure 6, as we can see the exchange rate data it's quite similar the same too as stock data. When the stock price is up, then the exchange rate is up also.

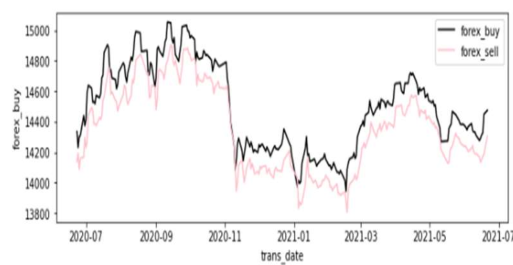

Figure 6 Visualization of Exchange Rate

3.3.7 Matrix Correlation using Heatmap

Matrix Correlation with Heatmap, as shown in Figure 6, visualize the correlation of chosen variable, the figure below uses Pearson correlation and then visualize it. The result shows that dependent variables in this research (low_prc) giving perfect correlation (0.81-1.00) to stock data except the volume, the exchange rate data giving moderate correlation (0.41-0.60) and sentiment data is giving no correlation (0.00-0.20) even minus. This could mean that the sentiment positive sentiment can increase lowprice value and the negative sentiment giving low_price a lower value.

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

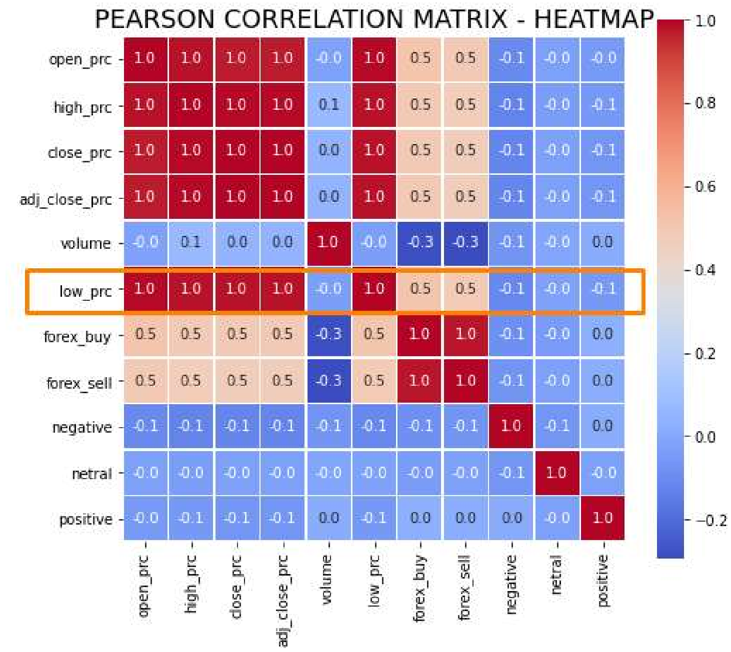

Figure 6 Visualization of Exchange Rate

3.4 Data Modeling

This section of data modeling consists of selecting modeling techniques, making test designs, making models, and assessing models, the model that we proposed on this project was separated into 2 step the first one was artificial intelligence approach model that consist of sentiment analysis and artificial intelligence approach, the result of sentiment analysis will be combined with historical data such as stock data and exchange rate data. The second was the conventional technique ARIMA.

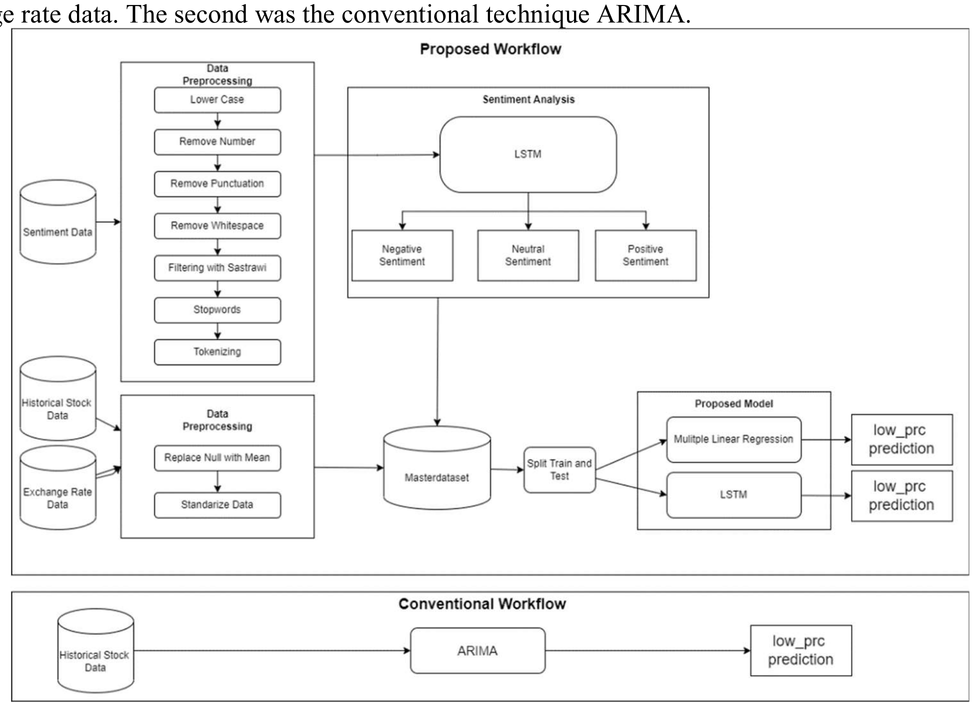

Figure 7 Proposed workflow and conventional workflow

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

3.4.1 Sentiment Analysis

The Sentiment Analysis in this research using LSTM model, the result of this process will be used in next process, this step following step several step:

1. Using LSTM library for word embedding.
2. Split the data with 8:2 between the Train and Test.
3. Run LSTM model using the sigmoid activation to find the binary_crossentropy, Adam optimizer to find the accuracy. The result of this sentiment analysis will be used for forecasting the stock price.

3.4.2 Stock Prediction

This step uses 3 kind of model which is Multiple Linear Regression, LSTM and ARIMA. The proposed workflow and conventional workflow in this study presented at Figure 7.

I. Multiple Linear Regression

The configuration in this steps was splitting the data between train and test with 8:2 ratio to forecast the stock with variable stock, exchange rate and sentiment data.

II. **LSTM**

The configuration for LSTM model using several step by using input shape with 1 loop back, 1 dense for the LSTM model, by finding the loss of mean_squared_error with Adam optimizer with 100 epochs.

III. ARIMA

For the ARIMA model still using the Rapid Miner with p=2,d=0 q=1 due the rapid miner can optimize the processing better, the variable used only historical of low_prc only.

4. RESULT AND DISCUSSION

4.1 Result

This section takes part of fifth part of CRISP-DM Evaluation, this section will show the result of the model from previous section for the sentiment analysis and stock price prediction, the metric that evaluate the sentiment will be the accuracy and for the stock price prediction will be Mean Square Error (MSE) and RMSE (Root Mean Square Error) for each of the train and the test data.

I. Sentiment Analysis – LSTM

The accuracy of the model was 72% which is quite high enough and the output of this sentiment analysis will be combined to predict the stock price.

II. Stock Prediction Multiple Linear Regression

The MSE and RMSE for train model was 2.664 and 1.632, the MSE and RMSE for the test model was little bit higher for 15.450 and 3.930.

Figure 8 shows the graph difference of prediction and actual result, the difference just a little between actual and prediction

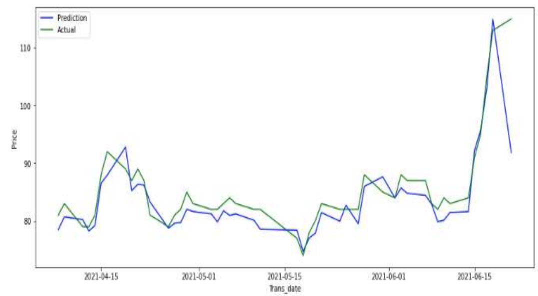

Figure 8 Comparison of Actual and Predict using Multi Linear Regression

III. Stock Prediction -LSTM

The MSE and RMSE for train model was 473.875 and 21.768, the MSE and RMSE for the test model was little bit higher for 74.181 and 8.612. The gap between actual and prediction is higher than using the Multi Linear Regression as shown in Figure 9.

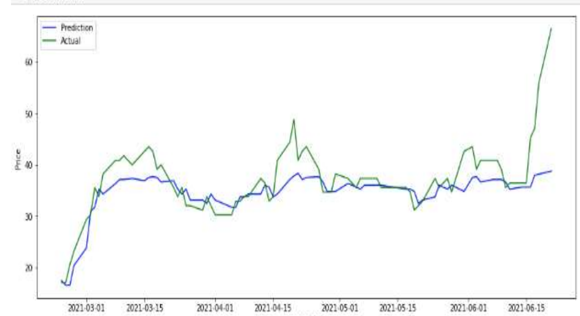

Figure 9 Comparison Actual and Predict using LSTM

IV. Stock Prediction -ARIMA

The MSE and RMSE for the ARIMA model was 24.770529 and 4.977. The value of ARIMA model is low enough, as shown in Figure 10. As we see the figure below at the end of the prediction, we can conclude that ARIMA forecast cannot predict when there is a high edge at the end of forecast.

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

Figure 20 ARIMA Comparison of actual and prediction

4.2 Discussion

Based on our research, even though RMSE and MSE from ARIMA model is low enough, but for FREN stock that having a trend conditions, it cause the big difference of the actual and the prediction. This conclude that the theory limitation of ARIMA was approved on this research.

Even though the result of Multi Linear Regression and LSTM giving quite high accuracy than ARIMA, the prediction almost always close to the actual price but the prediction almost always lower than the actual, hopefully future works can improve with parameter tuning so the line of prediction is equal or higher than the actual price.

5. CONCLUSION

In this paper, the focus was proposing a new approach to combine the stock price prediction with sentiment analysis and forex for Indonesia FREN Stock Market. The major contribution of this paper is in the integration of artificial intelligence to see the pattern of the data behavior using sentiment analysis to see the public opinion about FREN market and exchange rate fluctuation in order to help predicting stock price, the conclusion that the proposed model can overcome the limitation of ARIMA and bring more stable prediction.

The result of this research tells us that by using artificial intelligence approach, the forecasting can elaborate with other variable such as sentiment analysis and exchange rate fluctuation to improve the accuracy, the conclusion of the research tell us that stock prediction with Multiple Linear Regression Model give the better result with MSE and RMSE for train model was 473.875 and 21.768, the MSE and RMSE for the test model was little bit higher for 74.181 and 8.612 but unfortunately the prediction price usually lower than the actual price, that will cause the investor can miss the opportunity to buy the real lower stock price.

REFERENCES

[1] Y. Sukmana, “Stock Investment Is Not ‘Scary’ As Long As You Use This Way (Investasi Saham Tidak ‘Menakutkan’ asal Pakai Cara Ini),” 2021. .

[2] A. Malik, “Pandemic Pushes Super Apps Investments to Become a Trend,” 2021. .

[3] Aris Cahyadi, “The Trend of Mutual Fund Investment among Millennials Continues to Grow (Tren Investasi Reksa Dana di Kalangan Milenial Terus Tumbuh),” 2021. .

[4] A. S. Perwitasari, “The Covid-19 pandemic triggers a growth in the number of capital market and mutual fund investors (Pandemi Covid-19 picu pertumbuhan jumlah investor pasar modal dan reksadana),” 2021. .

[5] M. M. R. Majumder and M. I. Hossain, “Limitation of ARIMA in extremely collapsed market: A proposed method,” *2nd Int. Conf. Electr. Comput. Commun. Eng. ECCE 2019*, pp. 1–5, 2019.

[6] D. Valle-Cruz, V. Fernandez-Cortez, A. López-Chau, and R. Sandoval-Almazán, “Does Twitter Affect Stock Market Decisions? Financial Sentiment Analysis During Pandemics: A Comparative Study of the H1N1 and the COVID-19 Periods,” *Cognit. Comput.*, no. 0123456789, 2021.

[7] J. Lee and D. Ryu, “How does FX liquidity affect the relationship between foreign ownership and stock liquidity?,” *Emerg. Mark. Rev.*, vol. 39, no. March, pp. 101–119, 2019.

[8] T. BAGH, T. AZAD, S. RAZZAQ, I. LIAQAT, and M. A. KHAN, “The Impact of Exchange Rate Volatility on Stock Index: Evidence from Pakistan Stock Exchange (PSX),” *Int. J. Acad. Res. Accounting, Financ. Manag. Sci.*, vol. 7, no. 3, 2017.

[9] H. Maqsood *et al.*, “A local and global event sentiment based efficient stock exchange forecasting using deep learning,” *Int. J. Inf. Manage.*, vol. 50, no. July, pp. 432–451, 2020.

[10] Q. Ma, “Comparison of ARIMA, ANN and LSTM for Stock Price Prediction,” in *E3S Web of Conferences*, 2020, vol. 218, pp. 1–5.

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

[11] E. Ardyanta, H. S.-T. J. of A. Finance, and U. 2021, "A Prediction of Stock Price Movements Using Support Vector Machines in Indonesia," *Koreascience.or.Kr*, vol. 8, no. 8, pp. 399–407, 2021.

[12] J. Hur, M. Raj, and Y. E. Riyanto, "Finance and trade: A cross-country empirical analysis on the impact of financial development and asset tangibility on international trade," *World Dev.*, vol. 34, no. 10, pp. 1728–1741, 2006.

[13] S. Mokhtari, K. K. Yen, and J. Liu, "Effectiveness of Artificial Intelligence in Stock Market Prediction based on Machine Learning," *Int. J. Comput. Appl.*, vol. 183, no. 7, pp. 1–8, 2021.

[14] X. Yan, W. Weihan, and M. Chang, "Research on financial assets transaction prediction model based on LSTM neural network," *Neural Comput. Appl.*, vol. 33, no. 1, pp. 257–270, 2021.

[15] Z. Gao, "Stock Price Prediction with ARIMA and Deep Learning Models," *2021 IEEE 6th Int. Conf. Big Data Anal. ICBDA 2021*, pp. 61–68, 2021.

[16] N. Jing, Z. Wu, and H. Wang, "A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction," *Expert Syst. Appl.*, vol. 178, no. May 2020, p. 115019, 2021.

[17] Y. Shi, Y. Zheng, K. Guo, and X. Ren, "Stock movement prediction with sentiment analysis based on deep learning networks," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 6, pp. 1–16, 2021.

[18] A. K. Zyen and R. H. Kusumodestoni, "Pengembangan Model Prediksi Harga Saham Berbasis Neural Network," *J. DISPROTEK*, vol. 7, no. 1, pp. 74–83, 2016.

[19] B. Panwar, G. Dhuriya, P. Johri, S. Singh Yadav, and N. Gaur, "Stock Market Prediction Using Linear Regression and SVM," *2021 Int. Conf. Adv. Comput. Innov. Technol. Eng. ICACITE 2021*, vol. 7, pp. 629–631, 2021.

[20] G. Zhang, B. Eddy Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art," *Int. J. Forecast.*, vol. 14, no. 1, pp. 35–62, 1998.

[21] S. Mittal and A. Chauhan, "A RNN-LSTM-Based Predictive Modelling Framework for Stock Market Prediction Using Technical Indicators," *Int. J. Rough Sets Data Anal.*, vol. 7, no. 1, pp. 1–13, 2021.

[22] R. Batra and S. M. Daudpota, "Integrating StockTwits with sentiment analysis for better prediction of stock price movement," *2018 Int. Conf. Comput. Math. Eng. Technol. Inven. Innov. Integr. Socioecon. Dev. iCoMET 2018 - Proc.*, vol. 2018-Janua, no. December, pp. 1–5, 2018.