Moderna Vaccine Close Price Stock Prediction using Linear Regression Algorithm, Extreme Gradient Boosting Regressor, and ARIMA

Christofer Miko Lee 00000034222
Information Systems
Universitas Multimedia Nusantara
Tangerang, Indonesia
christofer.miko@student.umn.ac.id

Fiona Tanadi 00000034054
Information Systems
Universitas Multimedia Nusantara
Tangerang, Indonesia
fiona.tanadi@student.umn.ac.id

Nicholas Fernando 00000034082
Information Systems
Universitas Multimedia Nusantara
Tangerang, Indonesia
nicholas.fernando@student.umn.ac.id

Willibrordus Bayu 00000034000
Information Systems
Universitas Multimedia Nusantara
Tangerang, Indonesia
willi.brordus@student.umn.ac.id

***Abstract*—Covid-19 Pandemic has infected most of the society around the world. The pandemic didn't just affecting social sector, but also culture, health and economy. Nowadays, most of the stock price dropped day by day, one of the many company affected is Moderna Company's stocks (MRNA), a biotechnology industry that produced vaccine for health care and body's immunity to avoid getting infected by Covid-19 Virus. As we know, every day the Stock price's may fluctuate and rarely gain it's stable condition. Hence, we need a methodology to analyze about the growth of Moderna's stock price for the next few periods. As for the Algorithm that will be used in this research are Linear Regression, Extreme Gradient Boosting Regresssor, and ARIMA. Those model will be used to evaluate the growth and predict the Moderna Vaccine Company Stock's Closing Price throughout the next 60 days. The evaluation parameters used in this research is determined by measurement using the score of MAE, MSE, RMSE, as well as R2 score from each of the models.**

***Index Terms*—Covid-19, Moderna, Linear Regression, Extreme Gradient Boosting Regressor, ARIMA**

I. Background

Covid-19 Pandemic mainly occurs starting from December 2019, in Wuhan, Hubei Province, China. Corona Virus Disease (Covid-19) has been spreading around the world and infecting millions of people in worldwide scales. The positive cases amount has also increasing day by day. Covid-19 didn't just effect socials and cultures but also effect the economic section, just like cryptocurrencies and the growth of Indonesian's economical movements. Stocks is a product in the capital market that has unpredictable growth. Stocks in pharmacy companies such as Moderna Vaccine Company (ticker name : MRNA), is a biotechnology that has been developing vaccine to cure Covid-19 viruses. It has been proved that their research is highly potential as a prevention towards the Covid-19 viruses in the last few weeks. Therefore, we needed a methods to analyze how the Moderna Vaccine Company Stock's growing in the future.

The methods in this research are implementing three kinds of Algorithms, that will be used to make the prediction. Those Algorithms i.e. Linear Regression, Extreme Gradient Boosting, and ARIMA. All of this models will evaluate the growth of Close Price for Moderna Vaccine company's stocks for the next 60 days. While doing the research, we've encountered some problem that will be solved during the discussion process. Those problems i.e. how does the Close price's growing? How about prediction results to predict the Closing price for the next 60 days? What is the finding and value of the accuracy that we get from implementing three different algorithm used for predicting the close price for the next 60 days? We have been merging those problem as one as our main target in doing this research.

When we wanted to solve the problem in this research, we met some limitation i.e. the data collected on this research is found within a large number of amount, but has some limitation because of the Covid-19 pandemic, we could only do the research online and couldn't do a real face-to-face interviews to get real data. Then, our scope of research only consist of Covid-19 pandemic and only used 3 kinds of algorithm (Linear Regression, Extreme Gradient Boosting Regressor, and ARIMA to build the model and further prediction. The prediction in this research is no more than predicting for 60 days ahead. The goals of this research is to find out the escalation of the Close Price from Moderna Vaccine Company's Stocks. From the result, we got the prediction result on Close Price for the next 60 days and finding how much accuracy does the model had while implementing 3 different type of Algorithm. The measurement that we use for as the evaluation parameters of the models are based on the MAE, MSE, RMSE, as well as the R2 score. The benefits

from this research is giving a good reference for the readers or the next potential researcher that has a will to do the same research in the future. Also to finish the final project as given by the Machine Learning's Lecturer in this semester. In the end, we could also get some documentation for what we have research on and grateful of every person with their own role in finishing the studies.

II. Literature Review

A. Linear Regression

Linear Regression is an algorithm in machine learning that has quite been used for predict the average value for Y's numerical value for a given value. For example, X could apply a linear regression model. Then, with this X, Y could be estimated in the calculation to stated as an estimated values. Through the model, Y is counted using the linear combination calculation for X as input value. The formula of Linear Regression will are given below.

$$h_\theta(x) = \theta_0\alpha_0 + \theta_1\alpha_1 + \theta_2\alpha_2 + \cdots + \theta_n\alpha_n \tag{1}$$

Where the value of h is the result we gain from counting a dependent variable that has been known from the independent variable. (x), θ_0 is an intercept, a prediction reault from h when α_1 is 0. The function θ will stores the coefficient or weight from the input feature x and has a similiar dimension with x. But, if we wanted to add some support to the model's constant condition, we will have to begin the vector x with the value of 1. While doing this input and output, then we could conclude that Linear Regression is quite simple and useful [1].

B. Extreme Gradient Boosting Regressor

In the former research, the Extreme Gradient Boosting Regressor (XGBRegressor) is very reccomended to use, from the accuracy, strength, balance and efficiency, XGBRegressor is better than the regular SVM (Support Vector Machine) methods. E.g., the research of the traffic in Bali using XGBRegressor in Python, the use of XGBRegressor yang is implemented well with Python as the programming language, it's ensured to give a nice reccomendation to declare a plan of reducing the traffic jams [2]. XGBRegressor is a model found by Tianqi Chen and Carlos Guestrin in 2011 and has been quite useful for some researcher to make it as a framework based on the Boosting Tree model. The formula of XGBRegressor are given below.

$$\hat{y}_i = \sum_{k=1}^{K} f_k\left(x_i\right), f_k \in F \tag{2}$$

With assumptions that K is the total of the tree, F to represents the tree basic model. Model (2) is used to find an objective function :

$$L = \sum_i l\left(\hat{y}_i, y_i\right) + \sum_k \Omega\left(f_k\right) \tag{3}$$

Where l is a loss function, that represents an error between prediction value and the true value. Ω is a function to do the regulation to prevent overfitting. If we don't do those methods, and if all of the sample weights on the leaf nodes is less than the threshold value, then the splitting will automatically be stopped. This will prevent the model to do the learning on the special training sample. Moreover, the sample will be randomed when helping each tree [3].

C. ARIMA

ARIMA (Auto Regresive Integrated Moving Average) is a time series predictor model to make a linear prediction as a model while gathering the data from the parameter that wants to be predicted. The uses of ARIMA is limited due its stationery behavior from the collection of data that will be used to catch the linearity trends of that data [4]. ARIMA is one of the common time series regression model that uses an assumption that time series resulted is stationary. But in facts, the time series data in general won't have stationer behavior at all [5]. Because of that, the methods that we are going to use is a differentiation method, to change the unstationary data in average and transformation process for unstationary data in a variant. The mathematics model from ARIMA is as follows.

$$\Phi_{\mathrm{p}}(\mathrm{B})\nabla^{\mathrm{d}}\mathrm{Y}_{\mathrm{t}} = \xi + \Theta_{\mathrm{q}}(\mathrm{B})\varepsilon_{\mathrm{t}} \tag{4}$$

Where the value of Φ_{p} is an autoregressive parameter, Y_{t} is the observation value when t occur, B is moving back-off operator, d is differentiation parameter, ξ as constant parameter, and Θ_{q} is a parameter moving average parameter. Then, ε_{t} will represents the remaining error values [6].

D. Evaluation score of MAE, MSE, RMSE, and R2 Score

To validate the model's effectiveness, the prediction accuracy is determined by the functions of measurement, i.e Mean Square Error (MSE), and Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R square (R2) between measured and predicted values. The following acceptable criteria are used for evaluation are low MSE, low MAE, low RMSE, and high R2 score. The higher value of R2 means the lesser value of the errors indicate an accurate model towards the prediction [7]. The mathematical expressions of these functions are given below,

$$\begin{aligned} MSE &= \frac{1}{n}\sum_{i=1}^{n}\left(y_i - y_i'\right)^2 \\ MAE &= \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i'\right| \\ RMSE &= \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\left|y_i - y_i'\right|\right)^2} \\ R^2 &= 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i'\right)^2}{\sum_{i=1}^{n}\left(y_i - \overline{y}_i\right)^2} \end{aligned} \tag{5}$$

Where y_i denotes the actual values at i^n samples and $\overline{y_i}$ denotes mean of absolute values at i^n samples. While y_i' denotes the predicted values at i^n samples where n is the number of samples [8].

III. Research Methodologies

The following methodologies that we are going to use in this research will be stated as follows.

A. Methods of Research

The methods used in organizing and doing this research will be explained as in step-by-step as follows.

1) Literature Review. The references that has been obtained from Journals and books that is related to the research topic is learned to manage research materials and to make it as a reference in doing this research.
2) Making the flowchart. After doing some literature review, we made a flowchart to picture the steps required to do the prediction. The flowchart mentioned is referred on Fig 1.

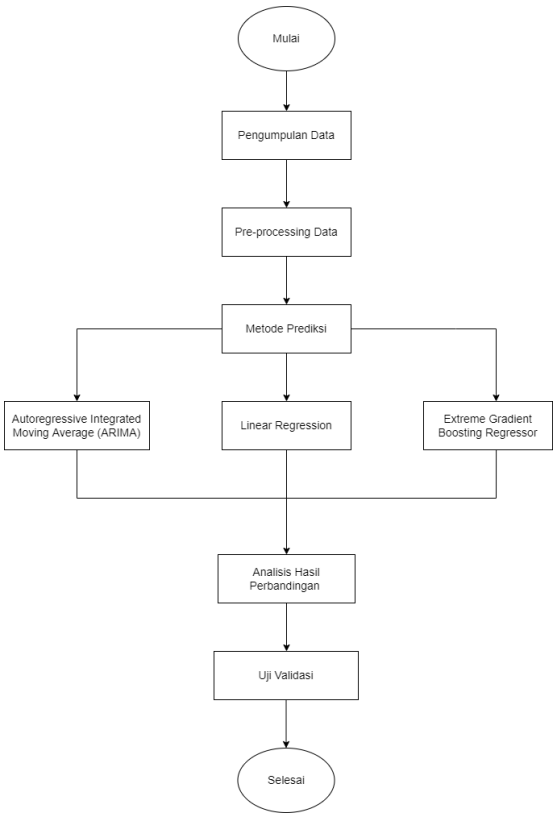

Fig. 1. Research Method's Flowchart

3) Data Collection. The Dataset used in this research is "COVID Vaccine Company Stock Analysis" that we obtained from a trusted dataset provider website, kaggle.com.
4) Import Library. All of the Library needed in this research is called, so we could ran all of the codes without getting any errors in Python.
5) Data Extraction. The Data obtained will be extracted in Jupyter Notebook will Python as it's programming language. The Data will be viewed also to see the amount and size, before continuing to next step.
6) Exploratory Data Analysis. First of all, we fixes the attributes and took some special attributes for this Moderna Dataset. Then, the changes of data type is required since some data has incompatibility data types. So, we changed the data type become "float". Next, we entered the 'Date' variable to an index to show an effective visualization. After that, before we started to build the model, we made a normalization and split the data into training and testing for the model's output to become more efficient.
7) Applying prediction methods for all chosen algorithm. After we explores the data as an initial steps to start the analytics, researcher then enters the step of building model. The goals of applying prediction and model construction methods using three different types of Algorithms is to see which one will outcome the best model to predict the Moderna Vaccine Stock's for the next 60 days.
8) Prediction with ARIMA. ARIMA's model focuses on deciding which is the best prediction on the Close Price by seeing the order value based on SARIMAX to decide the order value to be more efficient in predicting. ARIMA forecast the training data towards the previously used Close Price, before and after.
9) Prediction with Linear Regression Algorithm. The Linear Regression Model will connects the correlation value between the Close Price variable towards the other variable, like High, Low, Open, and Volume to predict the Close Price.
10) Prediction with Extreme Gradient Boosting Regressor Algorithm, which focuses on seeing the best hyper parameter on building the model to predict the Close Price. At the end of the result, only Linear Regression's model and Extreme Gradient Boosting Regressor's model that has the same visualization output.
11) Validation Testing and evaluating the model's result on all chosen algorithm by measuring it's errors value. All of the algorithm as some different elements, as in ARIMA we checked Mean Squarred Error, Mean Absolute Error, Root Mean Squared Error and Mean Absolute Percentage Error. For Linear Regression, we checked Mean Squarred Error, Mean Absolute Error, Root Mean Squared Error, Mean Absolute Percentage Error, and R2 Score. While on Extreme Gradient Boosting Regressor, we checked on Mean Squarred Error, Mean Absolute Error, Root Mean Squared Error, Mean Absolute Percentage Error, and R2 Score. This step will be included in the next chapter (result and discussion).

IV. Results and Discussion

A. Data Collection

The Dataset used in this research is "COVID Vaccine Company Stock Analysis" that we obtained from kaggle.com. The example of the dataset mentioned are referred in Fig 2. In the dataset, there were the other companies' name and some mixed variables, like the value of High, Low, Open,

Close, Volume, and Adj. Close. To narrow down our scope of research, we chose only 1 company as a research object, which is Moderna Vaccine Company.

Stock data of the following companies from Oct 2019

BioNTech	BioNTech	Moderna	Moderna	Moderna
Company Name	Company Name	Company Name	Company Name	Company Name
475 unique values	466 unique values	468 unique values	466 unique values	462 unique values
Volume	Adj Close	High	Low	Open
2805800	14.239999771118164	14.40999984741211	13.869999885559082	14.130000114440918
1364300	13.819999694824219	14.699999809265137	13.989999771118164	14.239999771118164
899100	12.75	14.270000457763672	13.670000076293945	14.1899995803833
484700	12.920000076293945	14.449999809265137	13.529999732971191	13.930000305175781
524200	12.9399995803833	14.319999694824219	14.020000457763672	14.270000457763672
413000	13.0	14.720000267028809	14.149999618530273	14.229999542236328
231900	14.140000343322754	14.850000381469727	14.069999694824219	14.5
290000	16.280000686645508	15.59000015258789	14.65999984741211	14.899999618530273

Fig. 2. Moderna Company Vaccine Stocks Dataset

B. Importing Library

Import Library is the step of calling the libraries needed to run several code in the research. Those code will be usedin learning and analyzing to predict the Stock's Close Price for Moderna Vaccine Company, using Linear Regression Algorithm, Extreme Gradient Boosting Regressor, and ARIMA. As for example of importing library, try to look at Fig 3.

```
In [1]: #import library
        import warnings
        warnings.filterwarnings('ignore')
        import math
        import numpy as np
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        plt.style.use('bmh')
        from sklearn.preprocessing import MinMaxScaler
        import statsmodels.api as sm
        from statsmodels.tsa.stattools import adfuller
        from statsmodels.tsa.seasonal import seasonal_decompose
        from statsmodels.tsa.arima_model import ARIMA
        from pmdarima.arima import auto_arima
        from sklearn.linear_model import LinearRegression
        from xgboost import XGBRegressor, plot_importance
        from sklearn.model_selection import train_test_split, GridSearchCV
        from sklearn import metrics
```

Fig. 3. Importing Library

C. Data Extraction

In Data Extraction, researcher loads the vaccine dataset "vaccine stocks.csv" into Python. We used two platform in analyzing the data, Google Collab and Jupyter Notebook. Then, we showed the 'Head' of the Dataset, to check the total rows and columns, which has 480 rows and 49 columns, as seen on Fig 4. We also did some checking on the name of columns on the dataset.

	Ticker	BioNTech	BioNTech.1	BioNTech.2	BioNTech.3	BioNTech.4	BioNTech.5	Moderna	M
0	Stock Info	High	Low	Open	Close	Volume	Adj Close	High	
1	Date	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
2	2019-10-10	16.698999404907227	13.010000228881836	16.5	14.239999771118164	2805800	14.239999771118164	14.409999847412111	13.8699998
3	2019-10-11	15.34000015258789	13.5	15.34000015258789	13.819999694824219	1364300	13.819999694824219	14.699999809265137	13.9899997
4	2019-10-14	13.909999847412111	12.524999618530273	13.829999923706055	12.75	899100	12.75	14.270000457763672	13.6700000

5 rows × 49 columns

Fig. 4. Data Extraction

D. Exploratory Data Analysis (EDA)

On the raw data of Moderna Vaccine Company Stock, there were a lot of other companies' name, but to narrow down the scope, we only picked one company to analyze, which is Moderna. To do so, we selected some attributes on the raw data, and deleted some unused data that has not related to Moderna company. Next, the unnamed columns in the raw data is renamed so it could be recognized easily. In Exploratory Data Analysis (EDA), we also made some election of choosing the ticker of each vaccine's stock, only the ticker with some attributes or feature that has related to Moderna will be chosen, as shown on Fig. 5. These Attributes has 478 rows and 6 column, i.e. Date, High, Low, Open, Close and Volume.

	Date	High	Low	Open	Close	Volume
2	2019-10-10	14.40999984741211	13.869999885559082	14.130000114440918	14.15999984741211	1343300
3	2019-10-11	14.699999809265137	13.989999771118164	14.239999771118164	14.055000305175781	934900
4	2019-10-14	14.270000457763672	13.670000076293945	14.1899995803833	13.930000305175781	954300
5	2019-10-15	14.449999809265137	13.529999732971191	13.930000305175781	14.329999923706055	1153500
6	2019-10-16	14.319999694824219	14.020000457763672	14.270000457763672	14.149999618530273	914100

Fig. 5. Moderna Company Vaccine Stocks

Then, we will describe the dataset using the describe function, the dataset consists of some feature, i.e. average, maximum, minimum and standard deviation as shown on Fig. 6.

	High	Low	Open	Close	Volume
count	478.000000	478.000000	478.000000	478.000000	4.780000e+02
mean	112.400511	104.757408	108.401628	108.941632	1.467463e+07
std	99.154501	91.851526	94.989140	95.999062	1.588032e+07
min	14.270000	13.530000	13.930000	13.930000	4.889000e+05
25%	32.949999	29.540000	32.090000	31.650000	5.487050e+06
50%	73.910000	70.160004	72.035000	71.575001	1.068410e+07
75%	160.707500	148.384995	154.108749	156.447498	1.695458e+07
max	497.489990	454.000000	485.500000	484.470001	1.251304e+08

Fig. 6. Dataset Distribution

After describing the dataset, we will check the outlier using a boxplot on prediction target, the feature of close. That feature has minimum value amount 13.9300, median 71.5750 and maximum value is 484.4700. Those amount were as show on Fig. 7.

After choosing the stock's ticker, we will choose the correlation of each feature on the Moderna Vaccine Company's stocks using Pearson Correlation method, as show in Fig. 8.

Besides showing the chosen data on Moderna's Vaccine Company Stock's Price. On Fig. 9, we will be checking the distribution of each feature, i.e. High, Low, Open, Close and Volume. Those 5 features resulted a skew distribution, especially on the Close Price as the prediction target. Then, we will transform the skew distribution into a normal distribution, just like as shown on Fig. 10.

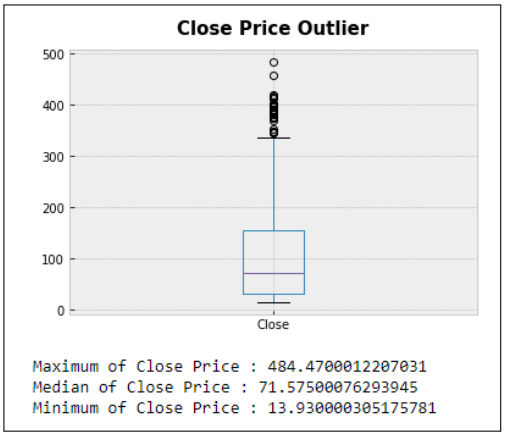

Fig. 7. Close Outlier

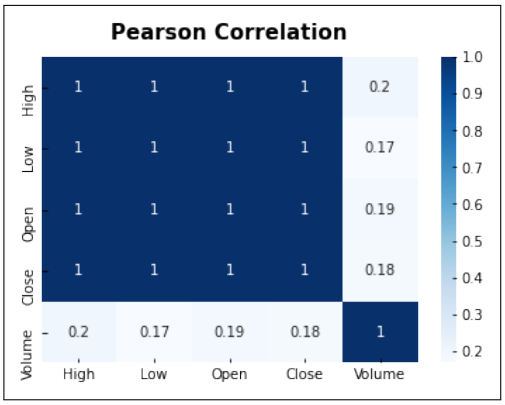

Fig. 8. Pearson Correlation

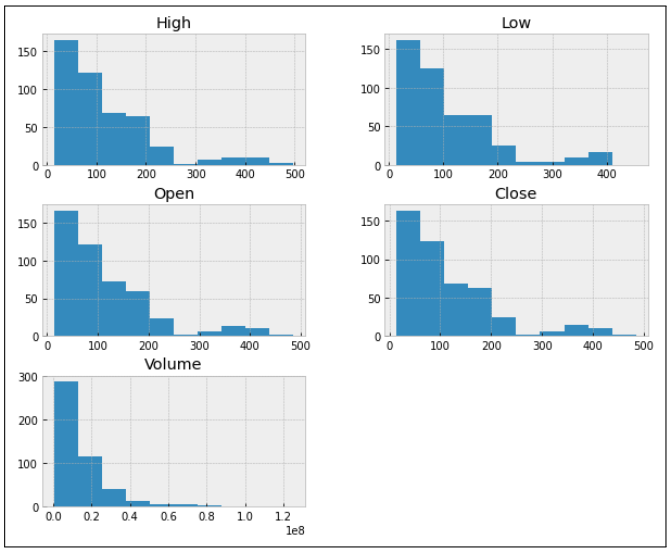

Fig. 9. Skew Distribution

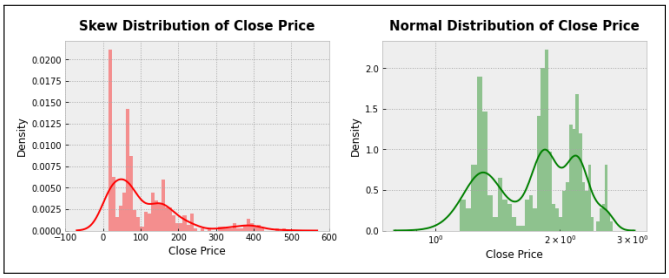

Fig. 10. Skew vs Normal Distribution

The next step as showed on Fig. 11, we checked the value of moving average for the next 20 and 50 days ahead. We could see that the moving average value for the next 20 days is dominantly higher than the moving average for the next 50 days.

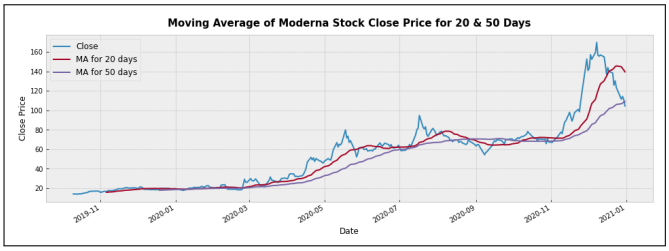

Fig. 11. Moving Average for the next 20 & 50 days

E. Autoregressive Integrated Moving Average (ARIMA)

In this step, we will be using ARIMA (Auto-Regressive Integrated Moving Average) as the algorithm to do the prediction. First of all, we build a scatter plot to see if both of the variable (Date and Close Price), has a good correlation or not. From the visualization result as shown on Fig. 12, it is seen that both of the variable has a good potential connection towards each other, so that we know that those variable is safe to use.

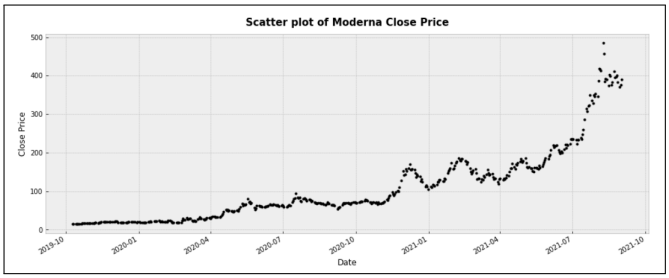

Fig. 12. Scatter Plot Close Price vs Date

Stationer value that's coming from the Rolling Mean and Standard Deviation will be used to see if the result of time series is rather stationer or not.

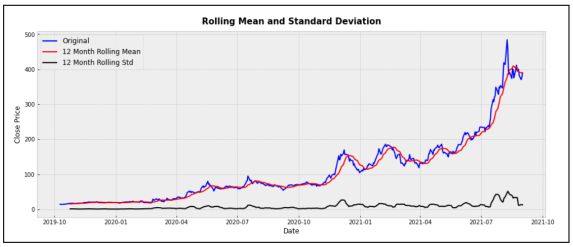

Fig. 13. Stationer Value

From the plot as shown on Fig. 13, we could also found some other value, like dickey fuller test statistics test, p-value, the amount of lags, observation score, and the critical score that's starting from 1%, 5%, and 10%.

The next step is making a decomposition visualization from the variable Close Price, using a multiplicative model, where a model is filled with one or more predictor variable to be seen the interaction effect's result from those predictive. After that, we could make an autocorrelation plot from Close Price to find the correlation towards the lag value mentioned before. From the autocorrelation lag plot, we will then make a visualization of the lagging from the Close Price, to search for the moving average from the mean with the standard deviation value, so we could see the relation between the value and date, as shown on Fig. 14.

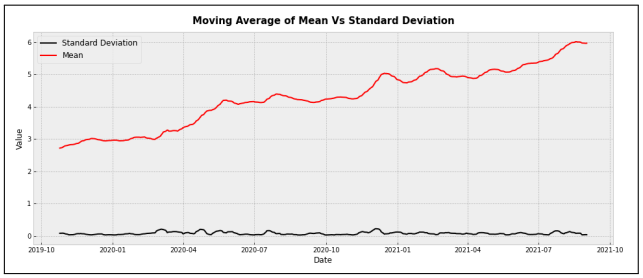

Fig. 14. Moving Average from Mean vs Standard Deviation

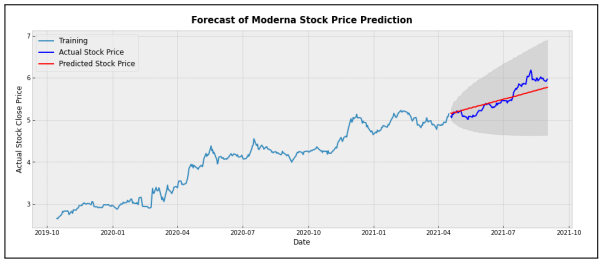

Fig. 15. ARIMA's Prediction Model

We will then make a plot from Training and Testing dataset to build an ARIMA model to search for the optimal order. From the model, we could find a result that has been diagnosed so it could be used to build the final model in predicting Moderna's stock price. After plotting it, don't forget to also check the error performance of the model, i.e. MSE (Mean Squared Error), MAE (Mean Absolute Error), RMSE (Root Mean Squared Error, and MAPE (Mean Absolute Percentage Error). For the example of the close price vs date plot, is shown on Fig. 15.

F. Linear Regression

In this step, we will do a prediction using Linear Regression Algorithm, the researcher wanted to predict on Close Price for the next 60 days, starting from 1st September, 2021. We wanted to find the correlation between Close Price and the other variable to find out whether the correlation between certain variable is quite good to be connected to Close Price variable. After that, we could start to build the Linear Regression model. Then, the result from splitting the data to see the accuracy score from Training and Testing data. Both of them showed the same score, (0,99 or 99% in percentage), which is a good value. At the end, we concluded that on the end of the result, the Moderna Vaccine Stock Close Price has an increased value on 1st September until it reaches the peak around 10th October, 2021. The graph of the forecast mentioned could be seen on Fig. 16.

Fig. 16. Forecast of Close Price for 60 days ahead for Close Price

G. Extreme Gradient Boosting Regressor

In this step, we will be using Extreme Gradient Boosting Regressor Algorithm to make a prediction, or could be called XGBRegressor, researcher uses grid search to find the best hyper-parameter, then do the fitting for training data with x and y as the chosen variables. After the grid search succeeded, we will build the model using the XGBRegressor. Upon building the model, we obtained the accuracy scores for training data (0.999), and testing data (0.997). From that accuracy result, a prediction could be made based on the x variable from testing data. After that, we checked the error rate of the quantitative data prediction from the model, starting from Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R2 Score. R2 Score will be used to evaluate the performance of a model, e.g.Linear Regression model, where a score will be produced to see how well does the model performing. After that, the actual score vs predicted score will be used to check the linearity value. Residual normality and mean score were also be checked to find out their relationship between the Close Price variable with Density variable. Linearity has gotten to know the relation between dependent and independent variable that has a straight red line, represented in Fig. 17.

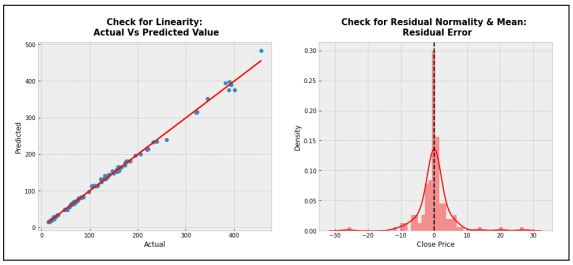

Fig. 17. Linearity value and Residual Normality Mean

The next step is making a prediction for 60 days ahead using XGBRegressor Algorithm, starting with finding the last date variable. After obtaining the last date, we could use it as a benchmark for the final date of the prediction. From that date, we could make a time series plot and adding the value obtained into a dataframe. Then, that dataframe will be added to the function using the join method, and we will get a prediction as a result. The prediction is shown on Fig.18 below.

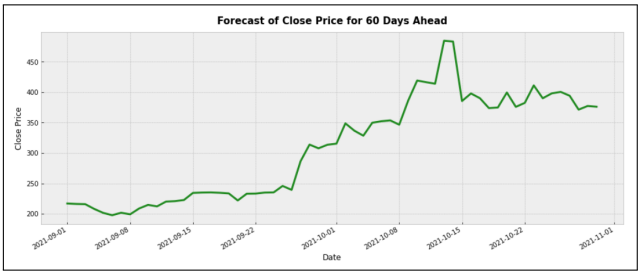

Fig. 18. Forecast of Close Price for 60 days ahead

From the visualization above that shows (*forecast*) of Close Price for 60 days ahead, we could see that the peak of the *Close Price* is around 15th October 2021.

V. Conclusion

From result and discussion section, we can conclude that:

1) The visualization result from the three Algorithm used, Linear Regression has prediction that has similiar result with XGB Regressor algorithm in predicting the Stock's Close Price for 60 days ahead. While ARIMA showed Moderna's Stock price prediction that only based on certain period.
2) From Linear Regression, ARIMA, and XGBRegreesor Algorithm used in this study, we could obtain the value of RMSE (Root Means Square Error) which represents the fidelity of the prediction made. ARIMA model RMSE value is 0.186, Linear Regression is 0.229, and XGBRegressor has a value of 0.439. Usually, RMSE that has value between 0.2 and 0.5 will shows that the model can relatively predict the data accurately. This means that XGBRegressor and Linear Regression's model has an accurate result while predicting the Close Price Stock's data.
3) In the end, all of those three Algorithm could be ran perfectly and generated some nice prediction as the result. Linear Regression dan XGBRegressor Algorithm gave a similiar prediction for 60 days ahead, while ARIMA made prediction based on the and checked if the predictions made were just right and suitable with the existing stock's Close Price.

References

[1] S. T. A. Shah, A. Iftikhar, M. I. Khan, M. Mansoor, A. F. Mirza, and M. Bi-lal, "Predicting covid-19 infections prevalence using linear regression tool," Journal of Experimental Biology and Agricultural Sciences, vol. 8, 2020.

[2] N. N. P. Pinata, I. M. Sukarsa, and N. K. D. Rusjayanthi, "Prediksi kecelakaan lalu lintas di bali dengan xgboost pada python," Jurnal Ilmiah Merpati, vol. 8, no. 3, 2020

[3] W. Li, Y. Yin, X. Quan, and H. Zhang, "Gene expression value prediction based on xgboost algorithm," Frontiers in genetics, p. 1077, 2019.

[4] V. Dear and A. S. Mardiani, "Implementasi model arima untuk algoritma prediksi nilai frekuensi kritis lapisan f ionosfer (fof2) harian," Pros. Semin. Nas. Fis. dan Pendidik. Fis. 2019 [sumber Elektron. Optim. peran Fis. dan Pendidik. Fis. melalui Artif. Intell. dalam menghadapi revolusi Ind, vol. 4, pp. 13–17, 2019.

[5] A. Sadeq, "Analisis prediksi indeks harga saham gabungan dengan metode arima (studi pada ihsg di bursa efek jakarta)," Ph.D. dissertation, program Pascasarjana Universitas Diponegoro, 2008.

[6] D. Hatidja, "Penerapan model arima untuk memprediksi harga saham pt. telkom tbk." Jurnal Ilmiah Sains, vol. 11, no. 1, pp. 116–123, 2011.

[7] A. Ahmad, W. Ahmad, K. Chaiyasarn, K. A. Ostrowski, F. Aslam, P. Zajdel, and P. Joyklad, "Prediction of geopolymer concrete compressive strength using novel machine learning algorithms," Polymers, vol. 13, 10 2021.

[8] P. Bamane and M. Patil, "A comparative study of different lstm neural networks in predicting air pollutant concentrations," Indian Journal of Science and Technology, vol. 13, no. 35, pp. 3664–3674, 2020.