Salim Demiray, et. al. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 11, Issue 1, (Series-III) January 2021, pp. 59-62

RESEARCH ARTICLE OPEN ACCESS

Stock Closing Prediction with Machine Learning Algorithms

Salim Demiray[1], Ahmet Gürhanlı[2]
[1] *(Department of Computer Engineering, Istanbul Aydın University, Turkey*
[2] *(Department of Computer Engineering, Istanbul Aydın University, Turkey*

ABSTRACT
Buying stocks, which is one of the important investment tools today, is important to predict the future values of the stock in order to make a profit. Since the investor has limited capital to be used to buy and sell stocks, she must predict the best time to buy and sell stocks to avoid wasting this capital. Fundamental analysis or technical analysis methods are used for these decisions. Buying stocks, which is one of the important investment tools today, is important to predict the future values of the stock in order to make a profit. Since the investor has limited capital to be used to buy and sell stocks, she must predict the best time to buy and sell stocks to avoid wasting this capital. Fundamental analysis or technical analysis methods are used for these decisions. With the developing software and hardware technology, machine learning algorithm have started to be used in addition to the statistical methods used in technical analysis. In this study, Polynomial Regression, Arima, XGBooster, LSTM (Long Short-Term Memory) and Prophet algorithms were performed on Microsoft stock closing values between 2014-2019.

Keywords **-** Machine Learning, Forecasting, Time Series, Stock Closing Forecasting

---

Date of Submission: 10-01-2021 Date of Acceptance: 25-01-2021

I. INTRODUCTION

The stock market is one of the investment instruments made by the exchange of stocks. Valuable papers purchased by investors to partner with companies are called stocks [8]. Stock investment can be short, medium or long term. There are various indices that show the prices of stocks in the stock exchange by dividing them into certain groups. While reducing risks with these indices, various methods and models are used to further reduce risks [12]. The analysis of the stock sector, the financial structure of the firm, its future financial expectations, the economic situation of the country and the analysis made to decide on the trading of stocks are called fundamental analysis. Technical analysis is the prediction of future prices by looking at past prices and the amount of movement of stocks. In recent years, with the development of software and hardware technology, machine learning methods have been used in addition to basic statistical analysis. Especially, Machine learning techniques have become more popular, since stock market data cannot be predicted linearly [14].

The Machine learning is software modeling that makes predictions such as classification, clustering or slope determination using mathematical and statistical operations and historical data. Machine learning is used to find which class the data belongs to (classification), to find the closest data set in terms of its features (clustering), to determine the direction of movement of the data (to determine the slope).Time series may have different frequency values such as year, month, day [1].In time series analysis, it is the estimation of the values of the future periods according to the previous values[2].The closing value estimation of stocks is a time series analysis and a regression analysis.

In this study, the closing prices of Microsoft stocks between 2014 and 2019 were analyzed using time series machine learning algorithms [3]. The used Polynomial Regression, Arima, XGBooster, LSTM (Long Short Term Memory) and Prophet algorithms were evaluated by making 5, 10 and 20 day predictions.

II. ALGORITHMS

2.1 LSTM (LONG SHORT TERM MEMORY)

Traditional ANNs (Artificial Neural Networks) do not have the ability to behave according to the previous situation like time series. LSTM (Long Short Term Memory) method has been developed to overcome this deficiency [4]. When they receive temporary training in ANN training, they forget their previous gains and value them according to the last situation. LSTM is a structure that feeds itself like a chain.

2.2 POLYNOMIAL REGRESSION ALGORITHM

It is a method that examines variables according to coefficients in a different parabolic

structure when the relationship between data cannot be fully expressed with the linear model [5]. Linear Regression is expressed with a constant and a certain factor such as $y = b_0 + b_1x_1$, while Polynomial Regression is represented by a parabolic expression with the formula $y = b_0 + b_1x_1 + b_2x_1^2 + b_3x_1^3 + \dots + b_nx_1^n$.Thanks to these parabolic parameters, the prediction curve and the true value curve are better matched.

2.3 ARIMA ALGORITHM

Autoregressive Integrated Moving Average (ARIMA) is a popular method of time series analysis that works after stabilizing by the difference or product of series. ARIMA (p, d, q) parameters are ARIMA (p, d, q) and d is the degree of stabilization by taking difference, while p and q are Autoregressive and Moving Average values [11]. Autoregressive is the dependence of the next value.

2.4 XGBOOSTER ALGORITHM

XGBoost is a speed and performance optimized decision tree algorithm published by Tianqi Chen and Carlos Guestrin [6]. Algorithm can work fast with predictive power, empty data management, and prevent over learning. Gradient Boosted, one of the basic parts of the algorithm, tries to turn weak leaves into stronger leaves step by step, trying to give better results [6]. Decision trees are a graphical representation of data that helps find solutions and group them according to conditions. It is called a tree because the grouping starts from a main root and divides deeply into sub-roots and leaves. All h

2.5 FACEBOOK PROPHET ALGORITHM

Prophet Algorithm is a framework developed by Facebook to automatically predict time series data by taking into account special period times such as weekends [7]. Prophet takes the data in two variables, time and values, and is robust against missing and floating data. Prophet also supports trends, seasonality and holidays in time series prediction [8].

2.6 ERROR MEASUREMENT METHODS

Mean Absolute Error, Mean Square Error and Root Mean Square Error measurement methods were used to measure the changes between the measured value and the actual value for the evaluation. Mean Absolute Error (MAE) is the value found by adding the difference between the estimated value and the actual value. Mean Squared Error (MSE) is the value obtained by adding the squared difference between the real value and the estimated value difference. Root Mean Square Error (RMSE) is the value found by the square root of the sum of the squares of the difference in the estimated value [9].

PURPOSE

This study is based on the daily closing values of Microsoft Company stocks between 2014 and 2019 taken from Yahoo Finance website. It has been tried to predict the closing values in future dates with machine learning methods. In this study, using several different software libraries via the Python programming language, with 5 different methods were investigated estimated results. In this study, it is aimed to find the results closest to the real values of stock closing prices.

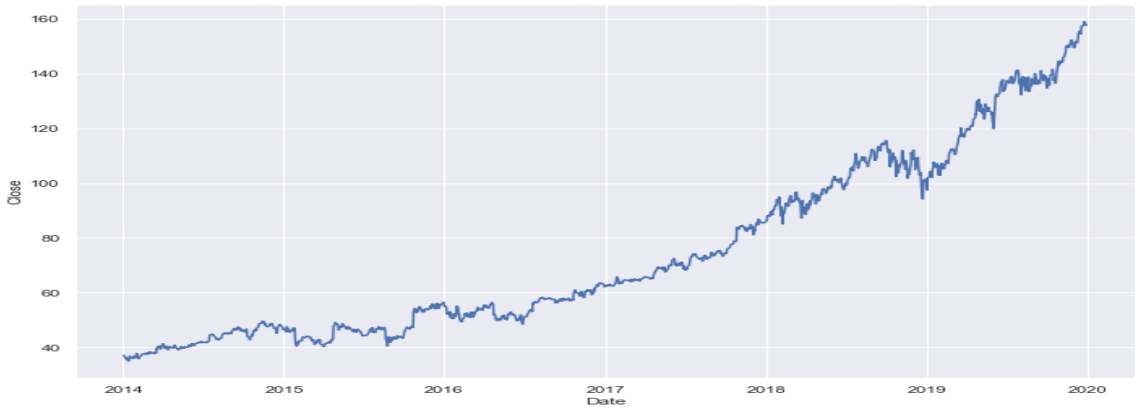
TARGET

Investors prefer various investment tools for their assets to increase in value. Stock market has an important place among these investment instruments. However, stock prices in the stock market can be very variable due to political, economic or industry-related reasons. This variability may turn into a large amount of profit or loss for investors. For this reason, predicting the future values of the stock becomes very important for the investor. In studies conducted with machine learning methods, estimation models for stock closing values have been proposed.

III. METHOD

A study was conducted to estimate the closing values from the data for Microsoft stocks received between 2014 and 2019 from the Yahoo Finance website. In order to measure the performance of the methods, 5, 10 and 20 day estimates were made over daily closing values, without any grouping or averaging. The laptop on which this study was conducted has 16 GB RAM, Intel Core (TM) i7-8550 CPU 1.80GHZ Processor hardware. As the data used for training is time dependent, data other than the last 20 days were used for the 20-day forecast; Tested with closing values for the last 20 days. During the work with Python, it was tried to find the minimum errors by using different learning algorithms and structures[15-10].

The Microsoft stock closing prices analyzed have a rising trend as in the graph-1.Error results of 5, 10 and 20 days for Polynomial Regression, Arima, XGBoost, LSTM (Long Short Term Memory) and Prophet methods used in this study are shown in table-1.When the results are examined, as seen in graphic-2, the Polynomial has the lowest error value.



Graph-1 Stock Price Closed Value

	MSE			RMSE			MAE		
Algorithm	**5**	**10**	**20**	**5**	**10**	**20**	**5**	**10**	**20**
Arima	**0,9393**	**5,4608**	**19,8731**	**0,9692**	**2,3368**	**4,4579**	**0,7496**	**2,0969**	**3,6710**
LSTM	29,8243	10,0034	8,3038	5,4612	3,1628	2,8816	5,3963	3,0374	2,5442
Polynominal	**0,8848**	**0,8238**	**4,1739**	**0,9406**	**0,9076**	**2,0430**	**0,8399**	**0,8114**	**1,6820**
Prophet	118,2301	128,0429	120,4416	10,8734	11,3156	10,9746	10,8429	11,1409	10,3985
XG Booster	**15,4017**	**30,2529**	**49,4574**	**3,9245**	**5,5003**	**7,0326**	**3,5235**	**4,8624**	**5,7572**

MSE: Mean Squared Error
RMSE:Root Mean Square Error
MAE:Mean Absolute Error

Table-1 : Predict Error Table

IV. RESULT

In the study, the best result in stock closing values is obtained with Polynomial Regression. There were slightly more error values that other algorithms can accept. As the number of prediction days increased, the error values increased. As seen in the first 5-day predictions in Table-2, while Polynomial Regression and Arima give very similar results; XG Booster, Prophet and LSTM gave more errors. It is thought that the statistical methods give better results, since this result is a trend of the stock and the closing values in 2014 when the data started were 37.16, reaching the value of 157.69 in the last period of the test. The closing values only examined in this study can be improved by adding the lowest and highest values in subsequent studies. While the model examined here is the stock values of Microsoft Company, different models can be used for the closing prices of other companies.

	Forecast Order				
Algorithm	**1**	**2**	**3**	**4**	**5**
Actual	157,3800	158,6700	158,9600	157,5900	157,7000
Arima	**157,3907**	**157,4299**	**157,2795**	**157,2068**	**157,2665**
LSTM	152,0293	152,2824	152,2975	152,8315	153,5880
Polynominal	**156,9416**	**157,2915**	**157,6440**	**157,9993**	**158,3573**
Prophet	147,7056	147,5504	147,4661	147,4661	145,8971
XG Booster	**155,9004**	**156,3047**	**156,0492**	**153,0982**	**151,3301**

Table-2 Last 5 Days Forecast Closing Values

REFERENCES

[1]. Rounaghi, M. M., & Zadeh, F. N. (2016). Investigation of market efficiency and financial stability between S&P 500 and London stock exchange: monthly and yearly forecasting of time series stock returns using ARMA model. Physica A: Statistical Mechanics and its Applications, 456, 10-21.
[2]. Akgül, I. (2003). Geleneksel Zaman Serisi Yöntemleri. Baskı, Der Yayınları, İstanbul.
[3]. https://finance.yahoo.com/quote/MSFT/history?p=MSFT Access Date:14.11.2020
[4]. http://colah.github.io/posts/2015-08-Understanding-LSTMs/ Access Date:14.11.2020
[5]. http://farukciftler.com/?p=1072 Access Date:14.11.2020
[6]. https://xgboost.readthedocs.io/en/latest/tutorials/model.html Access Date:14.11.2020
[7]. https://facebook.github.io/prophet/docs/trend_changepoints.html#automatic-changepoint-detection-in-prophet Access Date:14.11.2020
[8]. Taylor, S. J., & Letham, B. Forecasting at scale. Peer J Preprints 5: e3190v2 (2017).
[9]. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate research, 30(1), 79-82.
[10]. http://colah.github.io/posts/2015-08-Understanding-LSTMs/ Access Date:14.11.2020
[11]. Tseng, F. M., Yu, H. C., & Tzeng, G. H. (2002). Combining neural network model with seasonal time series ARIMA model. Technological forecasting and social change, 69(1), 71-87.
[12]. Filiz, E., KARABOGA, H. A., & AKOGUL, S. (2017). BIST-50 ENDEKSİ DEĞİŞİM DEĞERLERİNİN SINIFLANDIRILMASINDA MAKİNE ÖĞRENMESİ YÖNTEMLERİ VE YAPAY SİNİR AĞLARI KULLANIMI. Journal of the Cukurova University Institute of Social Sciences, 26(1)
[13]. Egeli, B. (2003). Stock market prediction using artificial neural networks. Decision Support Systems, 22, 171-185.

Salim Demiray, et. al. "Stock Closing Prediction with Machine Learning Algorithms." *International Journal of Engineering Research and Applications (IJERA)*, vol.11 (1), 2021, pp 59-62.