4/18/2023

Report on Applications of Data Science in the Stock Market

BY
SHREYASH SABLE

Abstract

Background:

Stock market process is full of uncertainty; hence stock prices forecast- ing very important in finance and business. For stockbrokers, understanding trends andsupported by prediction software for forecasting is very important for decision making. This paper proposes a data science model for stock prices forecasting in Indonesian exchange based on the statistical computing based on R language and Long Short- Term Memory (LSTM).

Findings:

The first Covid-19 (Coronavirus disease-19) confirmed case in Indonesia is on 2 March 2020. After that, the composite stock price index has plunged 28% since the start of the year and the share prices of cigarette producers and banks in the midst of the corona pandemic reached their lowest value on March 24, 2020. We use the big data from Bank of Central Asia (BCA) and Bank of Mandiri from Indonesia obtained fromYahoo finance. In our experiments, we visualize the data using data science and predictand simulate the important prices called Open, High, Low and Closing (OHLC) with various parameters.

Conclusions:

Based on the experiment, data science is very useful for visualizationdata and our proposed method using Long Short-Term Memory (LSTM) can be used aspredictor in short term data with accuracy 94.57% comes from the short term (1 year) with high epoch in training phase rather than using 3 years training data.

Keywords:

Data science, LSTM, Forecasting, Stock market, Finance, Deep learning

Introduction

Data science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data. Using data science and fore- casting method, we can see get many insight such as the financial health of a company. A forecasting algorithm is an information process that seeks to predict future values basedon past and present data. The forecasting is so important because prediction of future events is a critical input into many types of planning and decision-making processessuch as finance, industrial process control risk management. Time series analysis has significance in financial analytic and forecasting and it canbe utilized in any field. In finance, time series analysis is used for financial forecasting, such as stock prices, assets, and commodities. Stock is the most volatile investment with high risk, but with high return to investors if carefully managed in their portfolio.

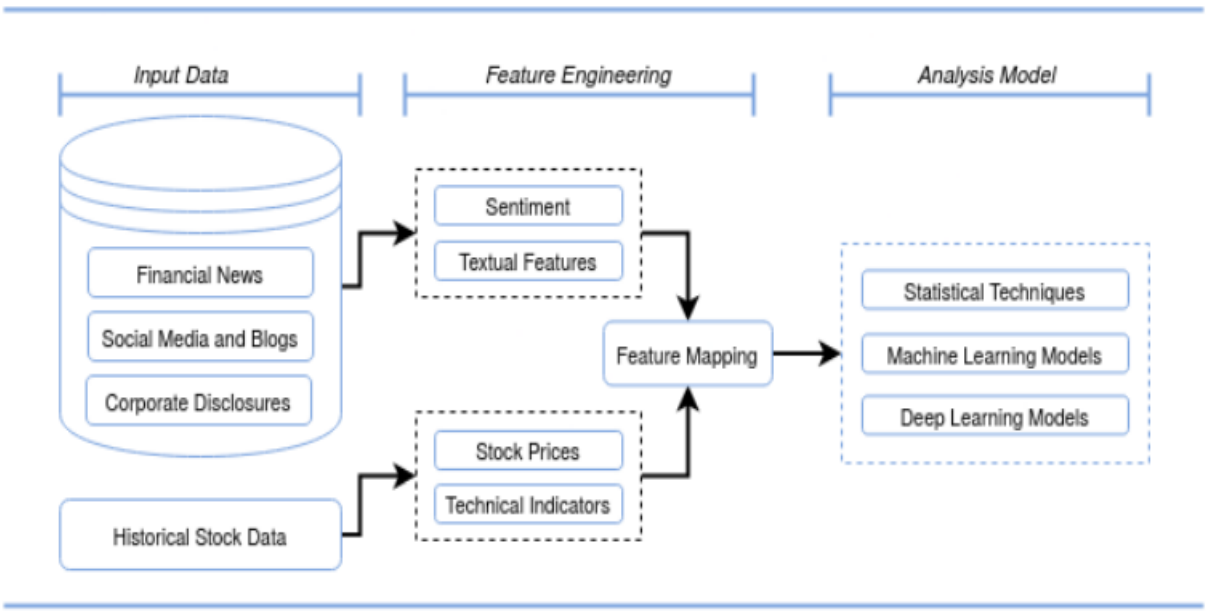

Fig. 1. A standard structure of a stock market analysis workflow

In managing stocks, information on their prices is of utmost importance. Capital marketsare markets for buying and selling equity and debt instruments, it also has activitiesrelated to public offering and trade of stock and issuance stock of public company. Stock exchanges are considered major players in financial sectors of many countries included Indonesia. Stockbrokers, who execute stock trade, use technical, fundamental or time series analysis in trying to predict stock prices, so as to advise client.

The capital market on the Indonesia Stock Exchange (IDX) in 2020 is in an uncertain condition since the outbreak of the corona virus (Covid-19) in Indonesia. Many issuers' shares have dropped, including state companies, aka state-owned enterprises. There are at least 10 state-owned enterprises (BUMN) shares whose prices have dropped considerably since the end of 2019 until now, such as the share price of PT Adhi Karya Tbk (ADHI) which experienced a drastic drop in just the last 60 trading days. As well as PT Semen Indonesia Tbk (SMGR). In the midst of the COVID-19 pandemic and the dynamics of the global financial market during Semester I 2020, the Jakarta Composite Index (JCI) and the majority of global stock index reference indexes experienced a significant decline. As of August 7, 2020, JCI was still closed in the red zone with − 18.34%. The IDX suspended short selling as the Jakarta Composite Index (JCI) was in a free-fall, continuing its losses since the start of 2020. The IDX believed the stock market corrections in Indonesia was mirroring similar losses around the world over fears of the corona- virus pandemic.

Today, artificial intelligence (AI) is a thriving field with many practical applicationsand active research topics. Many researcher on data science and deep learning try to predict stock prices forecasting such as using LSTM. This paper proposes an efficient, simple model and algorithm for big data analysis using R language and LSTM for stock forecasting with improvement and innovation in selecting only short-term datafor training phase and able to gives future prediction value and of course should be very useful for stock prices prediction in Indonesia. The section of paper consists of introduction, literature review, proposed method, result and discussion and conclusion section.

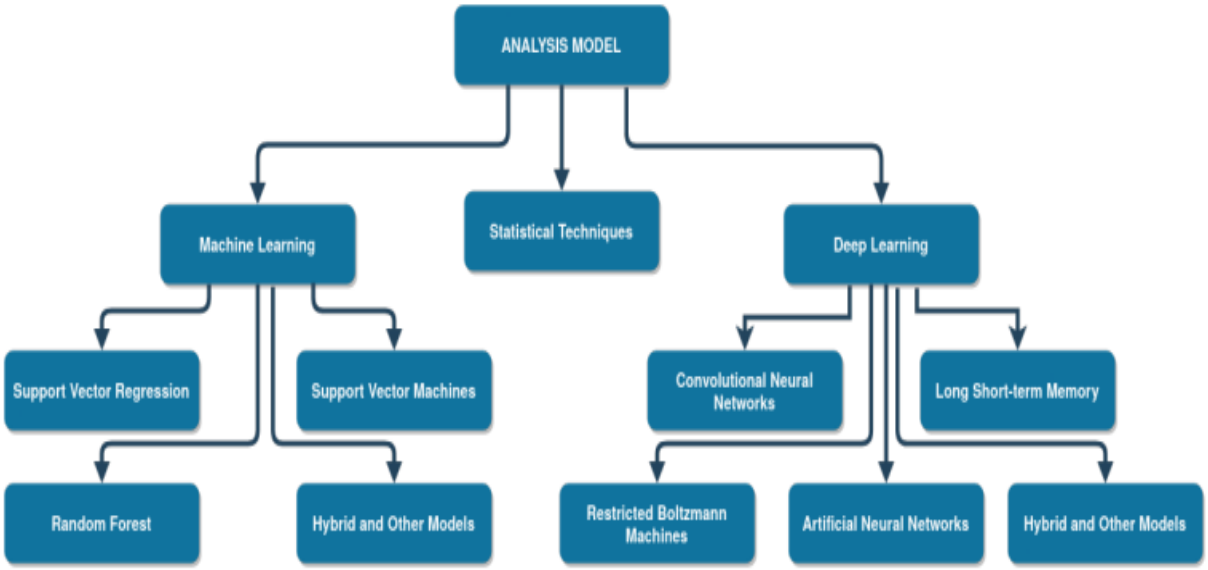

Fig. 3. A taxonomy of stock market analysis models

In contrast to these theories, behavioral economists claim that investors can be emotional and thus their behavior can be explained using psychology based theories. Behavioral finance mainly focuses on understanding the affect of investors' psychology in their trading strategy and on the market. LeBaron shows that there is a delay between the time new information is being introduced and market correcting itself by reflecting the new information and Gidófalvi reports that this lag is approximately 20 minutes. These results support the idea that new information can be used in stock price prediction for a short duration. Recently, a relatively new theory - the Adaptive Market Hypothesis - has been proposed to bridge EMH and behavioral finance - efficiency and inefficiency of the markets - in order to understand investor behavior better. This theory assumes that markets can be predicted by analyzing investor behavior.

Various approaches have been developed to analyze and beat the market. Fundamental and technical analytical techniques are mostly based on human knowledge and reasoning in areas such as locations of reversal patterns, market patterns, and trend forecasting. Although these techniques take historic stock data into consideration, most of the existing studies approach stock prices as stationary data. Also, human factors might cause inaccuracy in the results. The traditional statistical methods, including moving average, exponential smoothing and linear regression, have been used in the prediction of stock prices. However, they haven't been really effective as these models mostly assume linear relationship in the data.

Literature review

Stock prices forecasting

Predicting stock prices is very important for finance practitioners to best allocate their assets and to academics to build better and more accurate asset pricing models. Predicting stock returns gives crucial implications about market efficiency. Prediction of future movement of stock prices has always been a challenging task for the researchers. In fact, investors are highly interested in the research area of stock price prediction. Time series forecasting analyzes past data and projects estimates of future data values. Basically, this method attempts to model a nonlinear function by a recurrence relation derived from past values. A comparative study of LSTM and Deep Neural Network for Stock MarketForecasting has been conducted by. The Efficient Market Hypothesis (EMH) statesthat at any time, the price of a share fully captures all known information about the share. Since all known information is used optimally by market participants, price variations are random, as new information occurs randomly. Thus, share prices perform a "random walk", and it is not possible for an investor to beat the market. Many researchers propose a model for stock price forecasting, such as proposed a model for predictions using neural network to discover nonlinear relationships in input data makes them ideal for modeling nonlinear dynamic systems such as the stock market.

Another method for stock prices prediction is using ARIMA (Autoregressive Inte- grated Moving Average). In an ARIMA model, the future value of a variable is supposed to be a linear combination of past values and past errors. Assumptions of ARIMA model that data should be stationary—by stationary it means that the properties of the seriesdo not depend on the time when it is captured. A white noise series and series withcyclic behavior can also be considered as stationary series. Based on our previous result, LSTM is better compared with ARIMA model.

Dataset of stock prices from Yahoo Finance :-

Yahoo Finance is the largest business and financial news site in the world, with unrivaled access to data, insights, and content. The example of datasheet from Yahoo Finance is shown in Fig. 3. Data science approach focus on how to display data that easily understood by the decision

maker. Data visualization is an important feature in data science approach, as shown in Fig. 4, we can see data between 2018 and 2020, that the best stock prices condition of Bank BCA at the beginning of year 2020.

Artificial intelligence (AI) for stock prices prediction :-

Sequence prediction problems have been around for a long time especially in financial markets. LSTM built from the Recurrent Neural Network (RNN). In the figure shown, a chunk of neural network **A**, looks at some input x_i and outputs a value h_i. A loop allowsinformation to be passed from one step of the network to the next as shown in Fig. 5.

A typical LSTM network is comprised of different memory blocks called cells. Thereare two states that are being transferred to the next cell; the cell state and the hidden state. The memory blocks are responsible for remembering things and manipulations to this memory is done through three major mechanisms, called gates. LSTMs are

Time Period: Aug 26, 2019 - Aug 26, 2020 Show: Historical Prices Frequency: Daily Apply

Currency in IDR Download

Date	Open	High	Low	Close*	Adj Close**	Volume
Aug 26, 2020	31,925.00	31,950.00	31,525.00	31,700.00	31,700.00	4,023,900
Aug 25, 2020	31,600.00	31,900.00	31,525.00	31,825.00	31,825.00	9,842,200
Aug 24, 2020	31,700.00	31,900.00	31,400.00	31,575.00	31,575.00	13,376,500
Aug 19, 2020	31,800.00	32,050.00	31,550.00	31,650.00	31,650.00	11,758,700
Aug 18, 2020	32,025.00	32,125.00	31,500.00	31,800.00	31,800.00	16,312,700
Aug 14, 2020	31,500.00	32,025.00	31,200.00	32,025.00	32,025.00	11,463,200
Aug 13, 2020	31,300.00	31,500.00	30,900.00	31,500.00	31,500.00	9,716,200
Aug 12, 2020	30,875.00	31,300.00	30,800.00	31,300.00	31,300.00	7,122,400

Fig. 4 Example of datasheet from Yahoo Finance

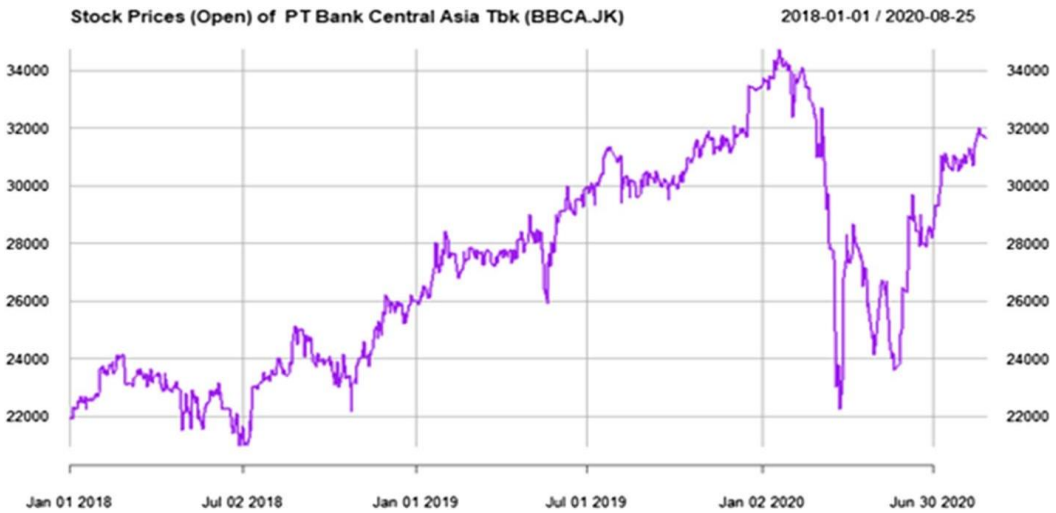

Fig. 5 Example of data visualization using data science approach

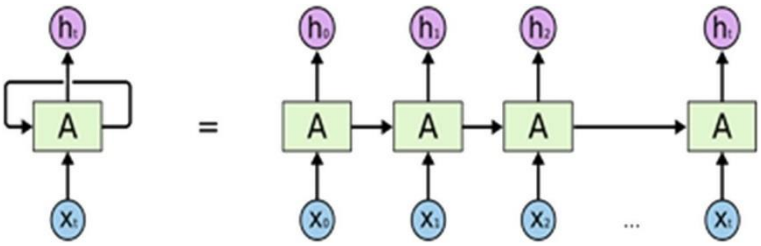

Fig. 6 The basic model of Recurrent neural networks (RNN)

particularly well suited to time-series prediction because they can "learn" and "remem-ber" in long-term memory things like market regimes, whereas short-term memory andgood interaction with look back windows (and even time-irregular data or large steps between significant events) leads to solid performance in short-term trend prediction.

Proposed method :-

The flow of data science approach for data visualization and stock prices predictionbased on big data from Yahoo is shown in Fig. 7.

Considering the complexity of financial time series, combining deep learning with concept of financial market prediction is regarded as one of the most charming topics. Based on that idea, we propose the algorithm for predicting of future values and the RNN model that has LSTM. We use values from the very beginning in the first

sliding window to predict the price *p* in the following window W_{t+1}:

$$W_{t+1} = p_{(t+1)w}, p_{(t+1)w+1}, \ldots, p_{(+2)w-1} \quad (1)$$

Figure 8 shows our model for stock prices prediction.

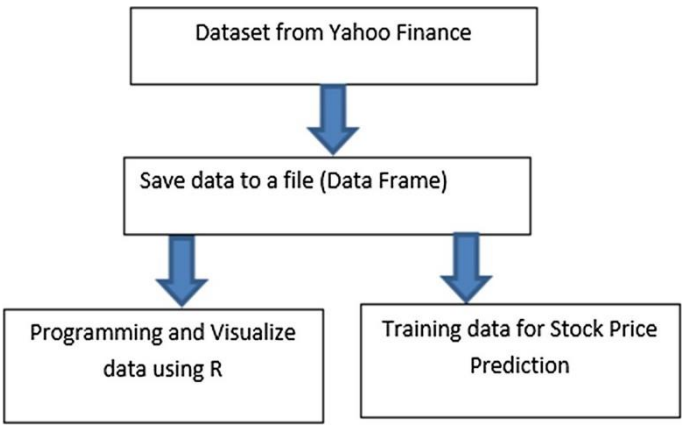

Fig. 7 The model for stock prices prediction using big data from Yahoo Finance

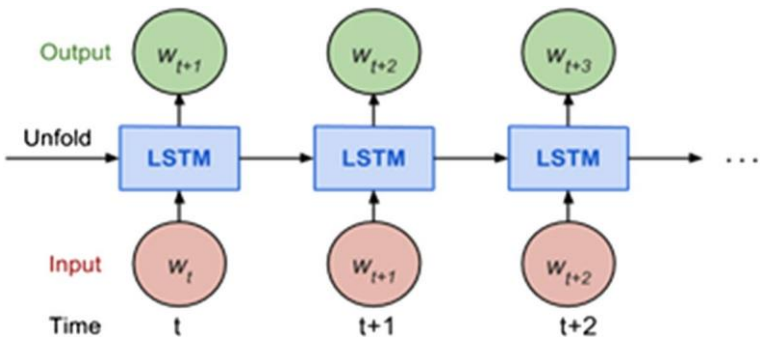

Fig. 8 The Rnn model for stock prices prediction

The efficient algorithm based on Tensorflow and LSTM for prediction of stock prices

Algorithm 1. *Predicting stock prices using LSTM using Stock Prices of PT Bank Central Asia Tbk (BBCA.JK)* declare epoch and variables load data BBCA.JK from yahoo finance begin get and format important columns for processing prepare for time series dataset *80% training set and 20% testing set* train-test split modeling LSTM calculate prediction 1 year and 3years and RMSE plotting predictions displaying prediction of future values end.

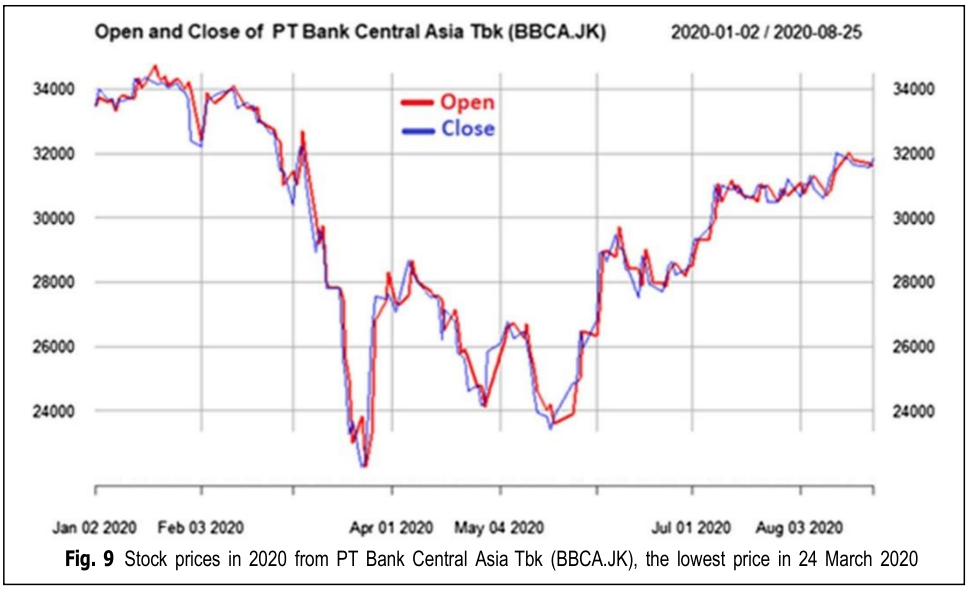

Fig. 9 Stock prices in 2020 from PT Bank Central Asia Tbk (BBCA.JK), the lowest price in 24 March 2020

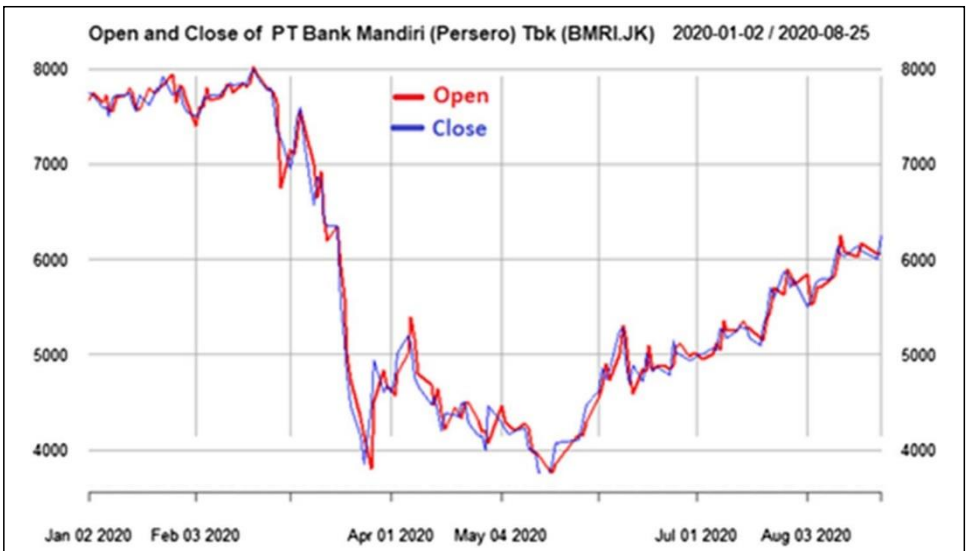

Fig. 10 Stock prices in 2020 from PT Bank Mandiri (Persero) Tbk(BMRI.JK), the lowest price also in 24 March 2020

Application Of Data Science In Stock Market

Data Science is everywhere In Stock markets, analyzing products, etc. Stock investments analysis is a theme that can be deeply explored in programming. If we include R language, which already has a big literature, packages and functions developed/import in this point.

Someone says that “In the short term, a market is a voting machine. But, in the long term, the market is a weighing machine”. — Ben Graham

Nowadays, predicting how the stock market will perform is one of the most difficult things to do. There are so many features involved in the prediction.

(a). physical factors vs. psychological,

(b). rational and irrational behavior, etc.

Working with historical data about the stock prices of a publicly listed company and implementing a mix of machine learning algorithms to predict the future stock price of this company, and starting with simple algorithms like (1)averaging and (2)linear regression. All these aspects link up to make share prices volatile and very difficult to predict with a high degree of accuracy.

We all are in the era of online discount brokerages and super-fast connections for both wireless and wired and Combining this with companies taking away pension plans is a huge thing to make up. One software name as Stock-Forecasting software which helps traders to make decisions to buy a favourite stock and sell it at the right a moment in maximum profit.

The Use of Data Science in Stock Market Analysis

Data science plays with numbers and lets us view financial data and the stock market from a different angle. The best use of data sciences is for forecasting future data results. In data science, data classification is done through testing, applying algorithms, and experimenting.

Before putting them into practice, ensure the concept is understood and that technology is utilized correctly. Fundamental analysis will be incorrect if data science in the stock market is not executed flawlessly. Data science makes it easier to execute trades and helps to generate profits. Data Science displays artificial intelligence-powered data analytics

by Data Science the numbers that could be profitable. The goal of technology is to provide accurate results and facilitate user interaction.

The Role of Data Science

Focus &Target :-

Data science aids in concentrating on and identifying the key elements in the stock market. A column or table format is made using data science to separate the data. The information helps you concentrate on crucial information for your stock market study by illustrating its significance. The column will provide market insights and stock values.

For the stock market, data sciences display dependent and independent variables. Big Data is crucial to the data's ability to forecast the future. For the prediction of future values, technology like artificial intelligence and machine learning models is helpful.

Algorithms :-

The programming of data science has a set algorithm. It is a series of instructions designed to carry out specific tasks and activities. The algorithm aids in the stock market trading and is applied to the timing of stock purchases and sales. It keeps track of the users' stock purchases and notifies them if stock prices or rates change. It entails making predictions and analyzing the current situation, which changes the stock market data.

Since the algorithm doesn't need human intervention, you can buy or sell using less-powerful trading strategies. You needed a data science specialist or data scientist to complete the work.

Training :-

When we say training, we don't necessarily mean you have to walk them through using the technique. It implies that some data or a subset of the data is chosen using data science and machine learning to train the model. Data science is initially trained before being tested.

The full dataset was used for training, and historical data was consulted for better training. Because it aids in forecasting and implementing data in the stock market, it is a crucial activity. Even datasets from the past and future are usable for the data model. The dataset model and stock prices are made clearer as a result.

Testing :-

It is crucial to pursue testing after the training model is finished. After testing, the model performed satisfactorily. The testing model can reveal whether the model is operating satisfactorily or not. A collection of experiment sets serve as the testing data, which aids in comparison for stock market analysis.

Datasets used for training and testing are two sides of the same coin. As a result, implementing a training set before testing is necessary. We want to reduce the error between the forecasts and the actual data as we experiment with our model.

Alternative Data :-

The idea of using data to forecast stock performance is not new. Investors have historically analyzed a company's overall health and investment prospects using financial records, sales data, buyer data, and other data.

On the other hand, data scientists today rely on non-traditional data sources or data sets that are frequently beyond the organization's control. Examples of alternative data include cell phone usage, social media activity, product reviews, credit card transactions, news sources, and satellite technology. Nearly infinite amounts of alternative data are available.

Future of Data Science in the Stock Market

Big Data, or any massive data set, is being utilized to spot patterns and trends and foretell how specific events will turn out. Whether structured or unstructured, data may frequently overwhelm a firm. The amount of data isn't what counts, though. How firms use the data is important.

Gain insights from big data analytics to inform your strategic planning and decision-making. As a result, it is crucial to the stock market. The use of the stock market in data science is a wonderful idea with a bright future. Modern technologies improve everything and aid in financial success.

Result and discussion :-

Based on data science approach, we can have insight Example of Stock Prices of PT Bank Central Asia Tbk (BBCA.JK) and PT Bank Mandiri from Indonesia at Yahoo finance are shown in Figs. 9 and 10. The first Covid-19 confirmed case in Indonesia is on 2 March 2020. After that, the composite stock price index has plunged 28% since the start of the year 2020, the share prices of cigarette producers and banks in the midst of the corona pandemic reached their lowest value on March 24, 2020 easily can be seen from data visualization.

We developed LSTM program using Python and Tensorflow for stock prices predic- tion.

We use 80% for training data and 20% for testing data, and the result shown in Figs. 9,10.

We compare result of the experiment by varying epoch and historical data between 1 and 3 years as shown in Table 3. It shows that the best prediction using 1 year data with the best accuracy 94.59% at 100 epoch. Epoch is one of the best methods to compare various data for forecasting. For analyzing the efficiency of the system we are used the Root Mean Square Error (RMSE). Comparing with other research for stock price forecasting, our method is better (usually neural network method only about 90% accuracy).

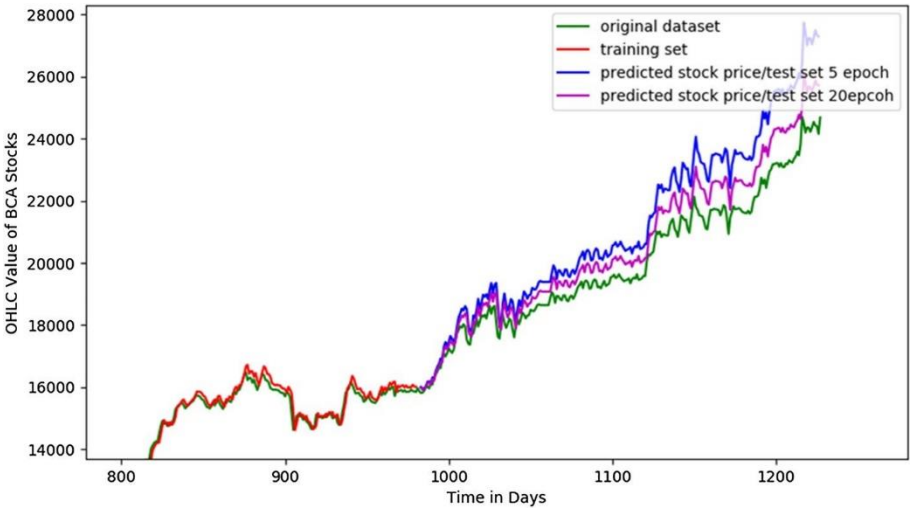

Fig. 11 Detailed result of prediction based on the 80% training set and 20% testing set. It shows thesignificant accuracy from 5 to 20 epochs

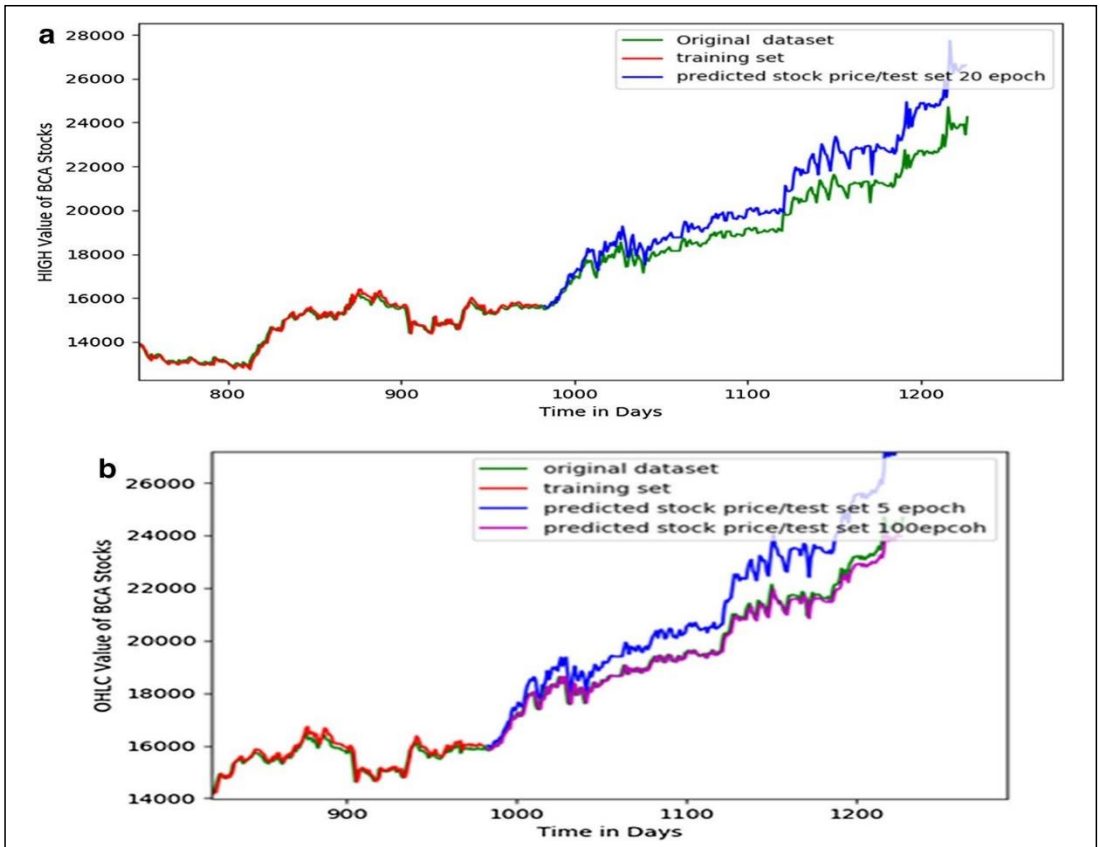

Fig. 12 Result of prediction of high price using 20 epochs (**a**) and OHLC price using 5 and 100 epochs.Prediction using 100 epochs is better than 5 epochs (**b**)

Table 1 Result of experiment with various historical data and epoch (with epoch 100 in 1 year, the best accuracy 94.59% is reached)

Historical data	OHLC value			High value		
	Epochs					
	5	20	50	20	50	100
Test RMSE value						
1 year	537.07	335,33	257.42	565.18	394.76	205.65
	Prediction value (next day)		25,545.20	25,341.33	25,343.35	25,344.05
	Accuracy		93.83%	94.57%	94.58%	*94.59%*
3 years	1023.43	1022.43	1023.45	929.80	193.30	193.31
	Prediction value (next day)		30,619.65	28,690.60	27,094.52	27,092.09
	Accuracy		78.28%	83.54%	88.46%	88.47%

Conclusion

This report develops a model and program for stock prices prediction using data from Yahoo finance. Efficient and accurate prediction systems for stock prices help traders, investors, and analyst by providing supportive information like the future direction of the stock market. We found that for LSTM, it should use short term historical data forthe best accuracy. Historical data using 1 year is the best compared with 3 years and5 years data. Deep learning technology is expanding the options available to data scientists to solve interesting problems with high accuracy. LSTM also superior in short term data until 94.59% as shown in Table 1. Data science approach proved to be usedeasily for decision maker and companies to get better view of stock prices or their financial health condition. For future work, we will improve our method using recent deep learning methods. At the end of 2020, amid the COVID-19 Pandemic, the number of Indonesia Capital Market investors continues to increase rapidly. The number of Indonesia Capital Market investors, according to data recorded in KSEI as of December 29,2020, increased by more than 50% to 3,871,248 from the previous 2,484,354 at the end of 2019. So, we have to optimize to solve the pandemic with empowering the business at all sectors.

References

1. Brockwell PJ, Davis RA. Introduction to time series and forecasting (Springer text in statistics). Cham: Springer Inter-national Publishers; 2016.
2. Introduction to Capital market and problem in Indonesia. http://www.indonesiaforeigninvestmentlaw.com/capit al-market/indonesia-invesment-capital-market-in-indonesia/. Accessed 7 Dec 2020.
3. Indonesia Stock Index [IDX]. https://www.idx.co.id/en-us/products/index/. Accessed 7 Dec 2020.
4. Analysis and Forecast of Indonesia's Jakarta Composite Index (IHSG) and state-owned enterprises (BUMN). https ://money.kompas.com/read/2020/03/25/182353726/10-bumn-yang-sahamnya-rontok-parah-saat-corona-menye rang. Accessed 20 April 2020.
5. Goodfellow I, Bengio Y, Courville A. Deep learning (Adaptive Computation and Machine Learning series). Cam-bridge: MIT Publisher; 2016.
6. Saad EW, Prokhorov DV, Wunsch DC. Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. IEEE Trans Neural Netw. 1998;9:1456–70.
7. Manurung AH, Prabowo H, Budiharto W. Algorithm and modeling of stock prices modeling based on long shortterm memory (LSTM). ICIC Express Lett. 2018;12:12.
8. Shah D, Campbell W, Zulkernine F. A comparative study of LSTM and DNN for stock market forecasting. In: IEEEinternational conference on Big Data. 2018.
9. Lawrence R. Using neural networks to forecast stock market prices. 1997. https://people.ok.ubc.ca/rlawrenc/resea rch/Papers/nn.pdf.
10. Yahoo Finance of Bank BCA. https://finance.yahoo.com/quote/BBCA.JK/. Accessed 28 Dec 2020.
11. Chen K, Zhou Y, Dai F. A LSTM-based method for stock returns prediction: a case study of China stock market. In: IEEE international conference on Big Data. 2015.
12. Stock price prediction tutorial. https://github.com/NourozR/Stock-Price-Prediction-LSTM. Accessed 10 Apr 2020.
13. Cavalcante RC, Brasileiro RC, Souza VLF, Nobrega JP, Oliveira ALI. Computational intelligence and financial markets: asurvey and future directions. Expert SystAppl. 2016;55:194–211.
14. Bao W, Yue J, Rao Y. Deep learning framework for financial time series using stacked autoencoders and long-shortterm memory. PLoS ONE. 2017;12(7):1–24.
15. Introduction about LSTM. https://www.analyticsvidhya.com/blog/2017/12/fundamentals-of-deep-learning-intro duction-to-lstm/. Accsessed 10 Jan 2019.
16. Tutorial on stock prediction using RNN. https://lilianweng.github.io/lil-log/2017/07/08/predict-stock-prices-using -RNN-part-1.html. Accessed 10 Sept 2019.
17. Patel MB, Yalamalle SR. Stock price prediction using artificial neural network. Int J Innov Res Sci Eng Technol. 2015;3(6):13755–62.