

THE APPLICATION OF ARTIFICIAL INTELINGENCE IN DJIA STOCKS TO IMPROVE THE INVESTMENT PROFITABILITY USING PHYTON

Widhiyo Sudiyono, ST, MAB
University of Muhammadiyah Malang

E-mail: widhiyo@umm.ac.id

Abstract:

Technological developments, competitive economics climate and demanding competition have led the investment industry to experienced rapid and continuous development in the last few decades.

Some of the rapid and continuous key developments are transformation in financial microstructures, development of investment strategies, the progression in computing capacity and the new trend of the investment performance of pioneers in algorithmic traders surpassing that of the human, discretionary investors (Jansen, Stefan 2017)

These four key factors have driven the investment company and hedge fund to develop algorithmic trading methods even further to achieve a more stable and reliable profit over time.

Therefore, to manifest aforementioned concerns, this research will conduct the process of building hybrid machine learning in Dow Jones Industrial Average stocks by using Long Short Term Memory (LSTM) Method to improve the investment profitability using phyton programming language.

The Result of this research shows that the prediction made by the software has acceptable rate of errors. The several measurements of errors used are namely, Median Absolute Error, Mean Absolute Percentage Error and Median Absolute Percentage Error.

Keywords: *Artificial Intelligence, Stock Market, Data Science , LSTM, Phyton*

1. Introduction

In this twenty-first century, the well-being of every rising economy, nation or society depends primarily on its consumer economies and stock prices, with the pivot being the financial market (Nassirtoussi et al. 2014; Göçken et al. 2016).

Technological developments, competitive economics climate and demanding competition have led the investment industry to experienced rapid and continuous development in the last few decades.

Some of the rapid and continuous key developments are

1. Transformation in financial microstructures such as the proliferation of electronic trading and the convergence of markets through asset classes and geographies.

2. Development of investment strategies in terms of risk factors exposure as opposed to groups of assets.

3. The progression in computing capacity, data-generation and management, and analytical approaches.

4. The new trend of the investment performance of pioneers in algorithmic traders surpassing that of the human, discretionary investors (Jansen, Stefan 2017)

These four key factors have driven the investment company and hedge fund to develop algorithmic trading method even further to achieve more stable and reliable profit over time. Therefore, this research will discuss more about the efficacy of algorithmic trading method to predict the movement of the stock market.

Problem Formulation

Based on the background above, the problem formulations in this research are:

1. How to implement machine learning method to make investment recommendation in the stock market using Python programming language.
2. How effective is the machine learning method in recommending investment decision?
3. How to optimize portfolio based in the machine learning recommendation
4. How to execute trades in the real stock market.. [Times New Roman, 12, normal].

Research Objectives

The objectives of this study are as follows:

1. To build a machine learning to help stock investors and trader making investment decision using Python programming language.
2. To examine the effectiveness of machine learning to predict the stock market movement

Research Benefits

This research is expected to provide benefits for researchers and potential investors

For potential stock trader, the machine learning built of this study can be used as a basis for decision making for investing.

Additionally, for further research, this research recommends to develop modified machine learning method to refine the machine learning built in this research.

Limitation of Problems

Limitation problems in this research are:

1. The stock price data used is the stock price data listed on the US stock market. The data will be retrieved from yahoo finance using yahoo finance API.
2. The duration of observation in this final project is 10 years, starting in January 2010 to December 2020

2. Research Method

This research uses a quantitative approach in the form of numbers or qualitative data created in the form of numbers. This study uses data in the form of a daily adjusted closing stock price of all companies incorporated in Dow Jones Industrial Average Stock data, from January 2010 to December 2020.

The data will be processed using a statistical formula that will find the results to be analyzed in order to obtain a conclusion. The research method explains the design of activities, scope or objects, main materials and tools, places, data collection techniques, operational definitions of research variables, and analysis techniques.

Data Population And Sample

The data population in this study are all shares of the companies included in the Dow Jones Industrial Average(DJIA) in the US stock market for the period 1 January 2010 to December 2020, as many as 30 company shares. Because the list of Dow Jones Industrial Average shares from time to time always change, all shares that have been included in the DJIA list in the past 10 years will be combined.

Source And Type Of Data

The stock price data used is the stock price data listed on the US stock market. The data and library of python programming language will be retrieved from :

1. Math library that support Pythons for processing mathematical functions
2. Numpy Library that support Pythons for processing scientific calculation
3. Pandas library that support Pythons with flexible, fast and expressive data structures
4. Mathplotlib library that support Pythons for data visualization and formatting
5. Sklearn library that support Pythons for measuring model and performance and errors
6. Keras library that support Pythons for building Deep Learning and Neural Network
7. Yahoo Finance that provide stock market data that can be accessed directly from Python programming language

The Method Of Analyzing And Processing Data

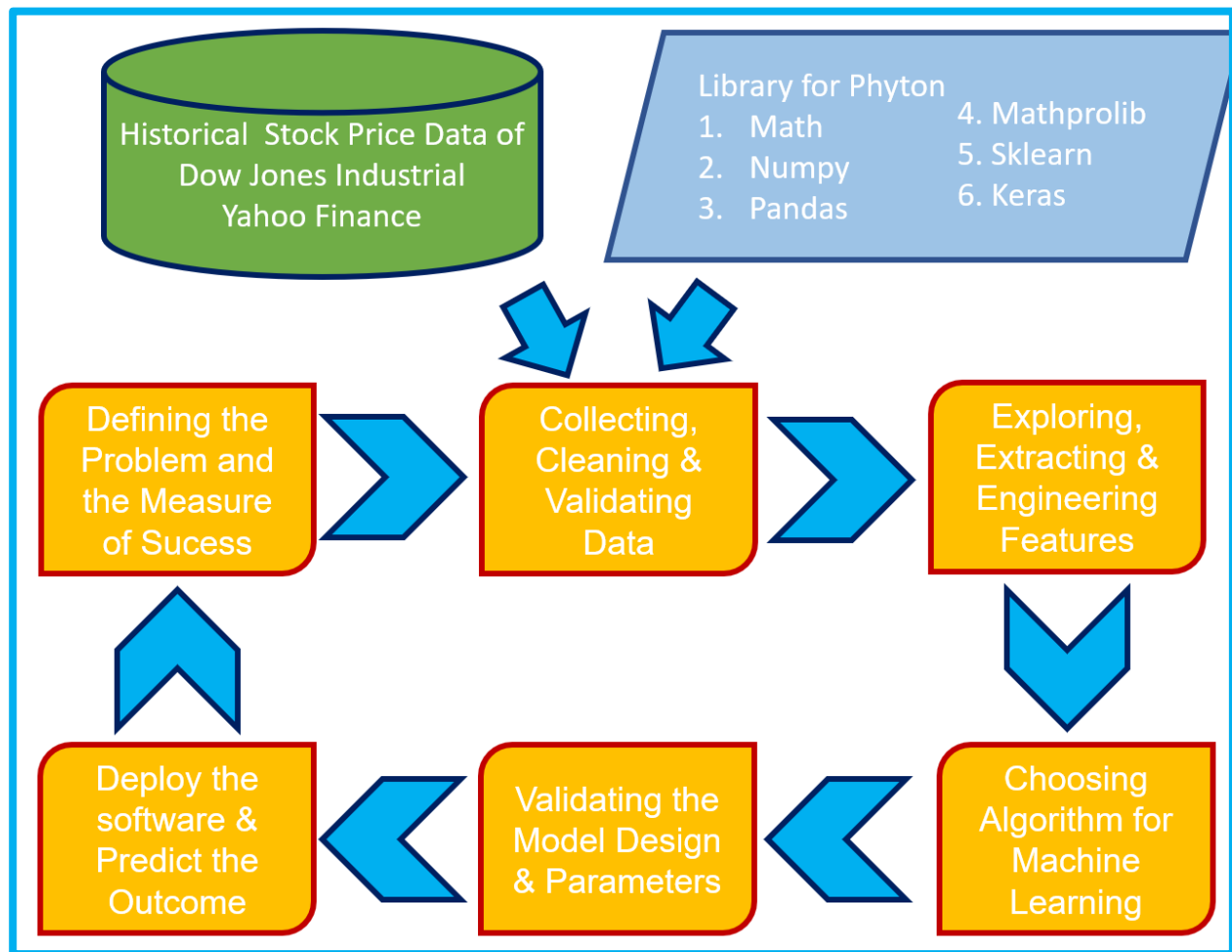


Figure 1 The Method of Analyzing and Processing Data

The above processes are conducted iteratively throughout the sequence, and the effort required at different stages will vary according to the project, but this process should generally include the following steps:

1. Frame the problem, identify a target metric, and define success
2. Source, clean, and validate the data
3. Understand your data and generate informative features
4. Choose machine learning algorithms suitable for your data
5. Train, test, and tune your models
6. Use your model to solve the original problem

The Machine Learning will be built using Python programming language.

Long Short Term Memory (LSTM)

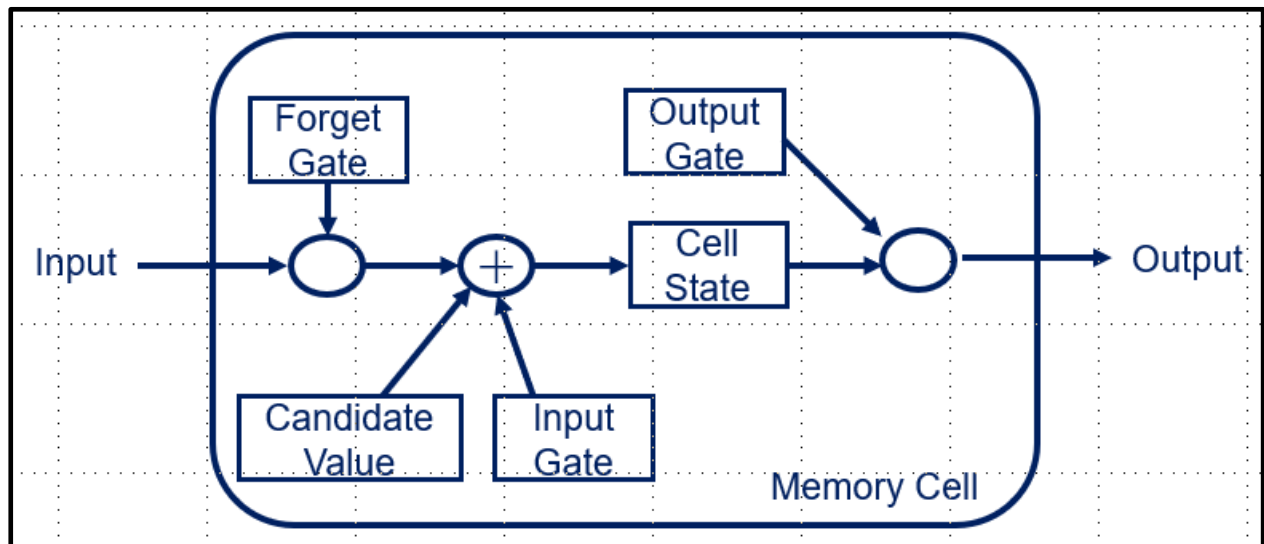


Figure 2 LSTM Structure

The structure of Long Short Term Memory above is identical with that of (Graves, 2014), (Olah, 2015) and (Chollet, 2016). Furthermore, several essential introductions about LSTM method with clear stage-by-stage illustration can be found in (Karpathy, 2015) and (Britz, 2015).

LSTM method is included under the class of recurrent neural network (RNN). LSTM network is built to enhance the capability to learn long-term dependencies and surmounts the innate problems previously faced by RNN, such as exploding gradient (Sak et al, 2014).

LSTM consist of an input layer, an output layer and one or more hidden layers in the middle. The hidden layer is called memory cell. The structure of memory cell is shown in the Figure 2 above. There are three essential gates in the memory cell, and each gate has its own role and different purpose, namely

- The forget gate decides which information to eliminate from the memory cell
- The input gate selects which information to add to the memory cell
- The output gate decides which information to send as output

3. Results and Discussion

The research which uses LSTM Method, an artificial recurrent neural network (RNN) architecture in the deep learning, made calculation and prediction of the stock market. This research divides the calculation into 8 steps to ensure the training process to be more accurate.

At first, the software will show the data from 1 January 2010 to the latest data available in the market to ensure the data processed is correctly retrieved. The initial display of the software is shown in the figure 3 below.

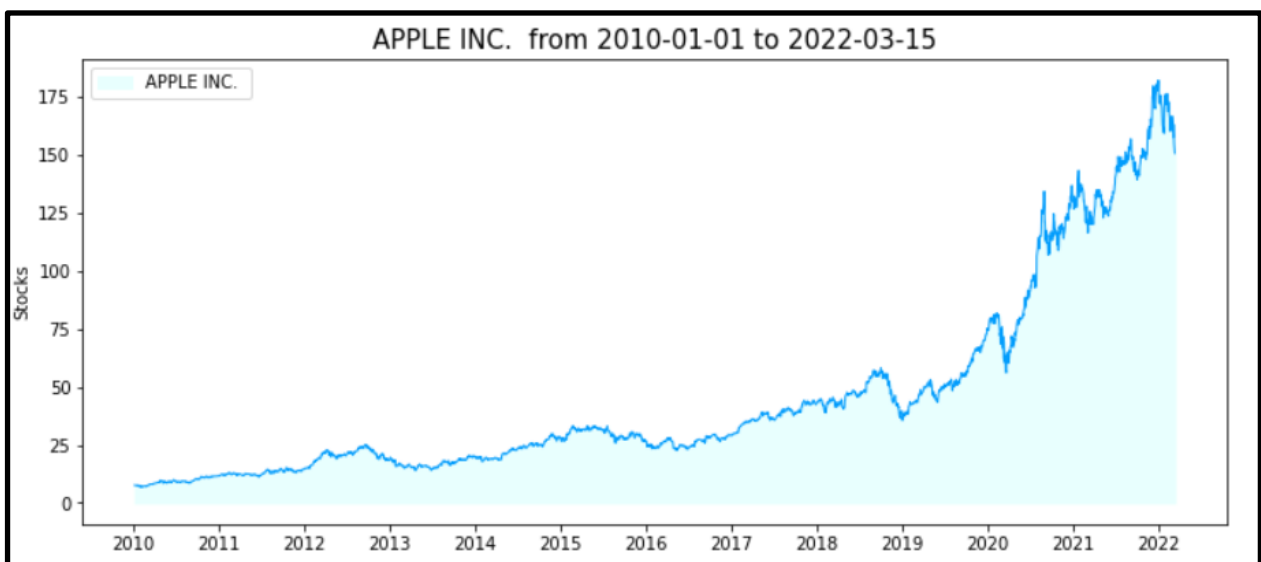


Figure 3 The Initial Display of the Software.

After that, the software will calculate the several absolute errors namely, Median Absolute Error (MAE), Mean Absolute Percentage Error and Median Absolute Percentage Errors. The purpose of this process is to ensure that the error generated by the system is not too significant. These three absolute errors are illustrated in the figure 4 below.

```
Median Absolute Error (MAE) = 9.05
Mean Absolute Percentage Error (MAPE) = 6.92 %
Median Absolute Percentage Error (MDAPE) = 7.16 %
This software is developed by Widhiyo Sudiyono
```

Figure 4 the calculation of absolute errors

Finally, the software shows the data used for training and also the comparison between test prediction and the real stock price. Furthermore, the software also shows its prediction whether the price will increase or decrease in percent as illustrated the figure 5 below.

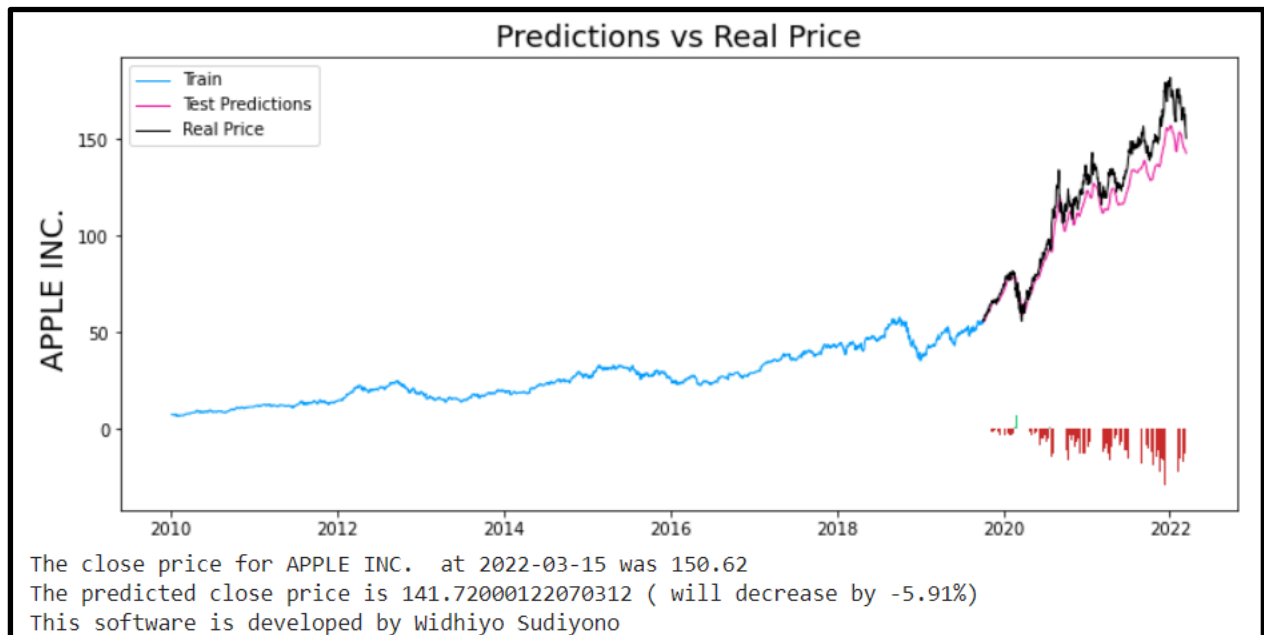


Figure 5 Prediction Result

As shown above, the software depicts the historical data in the light blue line and black line.. The light blue line are the historical data used to train the LSTM network. The black line and the pink line are shown to compare the real price with the prediction made by the software. It can be seen that the prediction has acceptable accuracy and the several absolute errors calculation made reflect that the prediction made is acceptable.

4. Conclusion

Long Short Term Memory(LSTM) is an improved version of Recurrent Neural Network(RNN). Unlike the RNN which unable to learn to the linked information because the old memory saved will be overwritten or replaced with new memory, LSTM can maintain and manage the memory of its each input through employing memory cells and gate units.

The software developed has shown significant successful rate on providing prediction of stock market price which included in the Dow Jones Industrial Average Index.

Further research should focus on modified LSTM layers, such as introducing additional dropout and adjusting the number of epochs, and using different oscillation datasets to test how satisfactory the prediction results are, or attempt to use sentiment analysis combined with the LSTM method to see the impact of the uncertainty in the value of bitcoin, based on the results of the study. sting the number of epochs, and employing different instability datasets to test how good the prediction results are, or try to use sentiment analysis combined with the LSTM method to see the impact of the uncertainty in the stock market, based on the findings of this study.

Reference

- "Standard & Poor's 500 Index – S&P 500". Investopedia. Archived from the original on June 14, 2012. Retrieved September 15, 2019.
- Bishop, C. M. (2006), Pattern Recognition and Machine Learning, Springer, ISBN 978-0-387-31073-
- Britz, D., 2015. Recurrent neural network tutorial, part 4 - Implementing a GRU/LSTM RNN with Python and Theano. URL <http://www.wildml.com/2015/10/recurrent-neural-network-tutorial-part-4-implementing-a-grulstm-rnn-with-python-and-theano/27>
- Chollet, F., 2016. Keras. URL <https://github.com/fchollet/keras>
- Deporre, James (September 7, 2018). "Ignore the Misleading Dow Jones Industrial Average". TheStreet.com. Archived from the original on August 12, 2019. Retrieved August 12, 2019.
- Ethem Alpaydin (2020). Introduction to Machine Learning (Fourth ed.). MIT. pp. xix, 1–3, 13–18. ISBN 978-0262043793.
- Floyd, David (June 25, 2019). "Discover What Makes the Dow Jones Industrial Average Stupid". Investopedia. Archived from the original on August 12, 2019. Retrieved August 12, 2019.
- Friedman, Jerome H. (1998). "Data Mining and Statistics: What's the connection?". Computing Science and Statistics. 29 (1): 3–9.
- Graves, A., 2014. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850.
- Jansen, Stefan(December 2018) Machine Learning for Algorithmic Trading
- Judge, Ben (May 26, 2015). "26 May 1896: Charles Dow launches the Dow Jones Industrial Average". MoneyWeek. Archived from the original on October 6, 2019. Retrieved October 6, 2019.
- Karpathy, A., 2015. The unreasonable effectiveness of recurrent neural networks.URL <http://karpathy.github.io/2015/05/21/rnn-effectiveness/>
- Lo, Andrew W., The Journal of Portfolio Management 30th Anniversary Issue 2004, 30 (5) 15-29; DOI: <https://doi.org/10.3905/jpm.2004.442611>
- Mitchell, Tom (1997). Machine Learning. New York: McGraw Hill. ISBN 0-07-042807-7. OCLC 36417892.
- Sak, H., Senior, A., Beaufays, F., 2014. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. arXiv preprint arXiv:1402.1128.