

Predicting Stock Price in Python Using TensorFlow and Keras

Orlunwo Placida Orochi, and Ledesi Kabari

Computer Science Department, Ignatius Ajuru University of Education

Abstract: One of the most important practices in the financial world is stock trading. The act of attempting to forecast the future value of a stock or other financial instrument listed on a stock exchange is known as stock market prediction. This paper discusses how Machine Learning can be used to predict a stock's price. When it comes to stock forecasts, most stockbrokers use technical and fundamental analysis, as well as time series analysis. Python is the programming language used to forecast the stock market. In this paper, we propose a Machine Learning (ML) method that will be trained using publicly accessible stock data to obtain intelligence, and then use that intelligence to make an accurate prediction. In this context, this research builds a neural network in TensorFlow and Keras that predicts stock market, which is basically a Python scraper that extracts finance data from the Yahoo Finance platform; more precisely, a Recurrent Neural Network with LSTM cells was constructed, which is the current state-of-the-art in time series forecasting.

Keywords: Stock, Artificial Neural Network, RNN, LSTM, Machine Learning, Prediction, Tensorflow, Keras, Artificial Intelligence.

I. INTRODUCTION

The financial market is a complex, composite mechanism that enables people via virtual broker sponsored platforms to buy and sell currencies, stocks and equities and derivatives. The stock market allows investors, through either exchange or over-the-counter trading, to own shares in the public company. This market gives investors the opportunity to make a living by invests in small initial sums of capital, a low risk compared with the risk of opening a new company or a high level of wage proficiency. Many factors that create uncertainty and high volatility in the market are affecting stock markets. However, macro or microeconomics like interest rates, exchange rates and monetary policy can easily influence inventory price, which makes prediction into a difficult task. Researchers and speculators based their research on stock market prediction over decades based on great profit in stock market investment. The prediction of the stock price movement was predicted by the traditional statistical methods such as logistic regression, the exponential average, ARIMA, and GARCH [9].

Data mining is the extraction from vast sets of data of valuable and trivial patterns or information. Data driven discovery (DDD), information extraction, pattern analyses, business intelligence are alternate names for data mining. However, goggle search engine simple search or question firing is not data extraction in relation database. Some fields

of data mining are available, such as machine learning (ML), cognitive learning (CL), statistics and algorithms. Files, databases and others are made up of large volumes of data, and so an existing tool for data analysis and explanation must be developed and interesting information extracted to promote decision-making [4]. The identification of concepts or class definitions, associations or similarities, prediction, clustering, trend analysis, outlining, and an analysis of similarity are some of the functionalities of data mining [7].

Machine learning is the technique that can be defined as the data developed through the mining of knowledge. Machines must not be explicitly programmed rather than trained to make decisions driven by data. The data is supplied to the generic algorithms instead of the code of any given problem, and logic is based on those data. As such it can be said very easily that a machine has really been taught, since it improves its exposure depending on its past inferences. The most accurate prediction is based on previous examples, and the artificial Intelligence (AI) strategy is the ideal procedure to produce a program for this. Any AI system (managed or unchecked) is productive enough to create rules for projects that take a superior choice with existing ones. The option here is whether inventory increases or decreases (Stock Analysis).

The aim of this study is to construct in TensorFlow and Keras a neural network which predicts stock market prices. More specifically, as current state of the art in time series forecasting, we are developing a Recurring Neural Network with LSTM cells. The key reason behind this work is for stock market investors to estimate the actions and/or trajectory of stock market prices so that they can participate in an ongoing profit trend in the future.

A neural network is a bio-inspired device type that contains many single processing elements known as neurons. By a joint mechanism consisting of a collection of assigned weight, the neurons are bound to each other. Artificial Neural Networks (ANNs) are an information processing method motivated first by mathematical generalizations in the human neuron. Any signals from other neurons or from the outside are obtained from any neuron. The figure above comprises three neuron layers, with a single input layer. Every neuron uses an active feature when the total input exceeds a certain threshold. Their feedback loops are used to identify Recurring Neural Networks (RNNs). These learning algorithms are primarily used to predict future results, such as stock market forecasts and revenue forecasts by using time series data.

II. RELATED WORK

During the literature review, we gathered knowledge on the latest approaches used in the field of stock market prediction.

As a combination of various methods for selecting stocks, [10] suggested a solution. They used the database of the Taiwan Economic Journal (TEJ). Data from 2000 to 2007 were used in their research. In their work they used a Multi-Layer Perceptron (MLP) based, sliding window process and combined it to produce background multi-layer neural networks as their prediction model. They have used main component (PCA) analysis in their work for the reduction of dimensionality, genetic algorithms (GA) and graded and regressed (CART) tree to pick key characteristics. They were not only dependent on technical indexes. Rather, both fundamental and macroeconomic indicators were used in their study. The authors also documented a comparison of methods for feature selection. The validation component was achieved by the combination of the model results with the study of statistics.

[1] suggested a solution for short-term stock price trend prediction using a neuro-fuzzy system composed of a controller called Adaptive Neuro Fuzzy Inference System (ANFIS). The assessment portion of this work is a notable strength. They compared their proposed framework not only to commonly used data models, but also to investment strategies. While the flaw in their proposed solution that we discovered is that it lacks an optimization component, which could limit the model's success. This study is heuristic for our device design since our proposed approach often concentrates on short-term stock price trend prediction. Meanwhile, their research encouraged us to equate investor strategies to research techniques by comparing them to common trading strategies.

[11] used ensemble methods of four well-known machine learning models to forecast short-term stock prices. There are five sets of data in this study's dataset. These datasets were collected using three open-source APIs and the TTR R kit. The neural network regression ensemble (NNRE), a Random Forest with unpruned regression trees as base learners (RFR), AdaBoost with unpruned regression trees as base learners (BRT), and a support vector regression ensemble were the machine learning models they used (SVRE). A detailed examination of ensemble approaches for predicting short-term stock prices. The authors chose eight technical metrics in this study based on their context expertise and then carefully evaluated five datasets. The key contribution of this paper is that it established an R-based framework for investors that does not require users to enter their own data but instead calls an API to access data from an online source. From a study standpoint, they only looked at market predictions for 1 to 10 days ahead of time, but not for more than two trading weeks or for less than one day. The model could not be generalized to other stock markets or need more revalidation

to see whether it suffered from overfitting issues since they only looked at 20 U.S.-based stocks.

Wavelet Neural Network (WNN) was used [5] to forecast stock price patterns. As an optimization, the author used Rough Set (RS) for attribute reduction. The stock price trend function dimensions were reduced using Rough Collection. It was also used to evaluate the Wavelet Neural Network's structure. This study's dataset contains five well-known stock market indices: the SSE Composite Index (China), the CSI 300 Index (China), All Ordinaries Index (Australian), (4) the Nikkei 225 Index (Japan), and the Dow Jones Index (USA). The model was evaluated using various stock market indexes, and the results were persuasive in terms of generality. The computational complexity is minimized by using Rough Set to optimize the function dimension before processing. However, in the discussion section, the author only emphasized parameter modification and did not mention the model's flaws. Meanwhile, we discovered that since the evaluations were done on indices, the same model might not work as well if applied to a specific stock.

[12] used RNN and LSTM to predict the price of Bitcoin, with the function engineering component optimized using the Boruta algorithm, and it works similarly to the random forest classifier. They used Bayesian optimization to choose LSTM parameters in addition to feature selection. The Bitcoin data was collected between August 19th, 2013, and July 19th, 2016. To boost the efficiency of deep learning approaches, several optimization methods were used. Overfitting is the main issue with their work. The issue of forecasting Bitcoin price trends is close to that of predicting stock market prices. This work is threatened by hidden features and noise embedded in the price details. The research issue was approached as a time series problem by the authors. The feature engineering and optimization section of this paper is the most interesting; we could use the methods they used in our data pre-processing.

In 2019, [8] proposed a scheme of two separate stock investment approaches. The advantages of their proposed solution are self-evident. To begin with, it is a robust framework that involves data pre-processing as well as two separate algorithms for recommending the best investment segments. Second, the system includes a forecasting portion that preserves the time series' characteristics. Finally, their input features are a mixture of fundamental and technological indices that strive to bridge the gap between the financial and technical realms. Their job, however, has a flaw in the assessment section. They selected 25 well-known stocks to test the proposed system instead of a massive dataset. There is a good chance that the well-known stocks have some secret characteristics in common.

In [2] applied the financial market forecast to the long-term memories (LSTM). The data base used is Thomson Reuters' S&P 500 index elements. From December 1989 to September 2015 all month-end S&P 500 lists were collected,

then the lists were unified in one binary matrix to eradicate survivor bias. The authors have used the RMSprop as a mini batch rprop optimizer. The key strength of this paper is that the writers have used the new profound learning technologies to anticipate. They relied on the LSTM technology, lack of financial history. Although the DNN norm and logistic regression algorithms were exceeded by the LSTM, the author did not mention efforts to train an LSTM with long-standing dependencies.

Related works in the investment domain have shown a greater interest in behaviour analysis, such as how herding habits affect stock performance or how the proportion of inside directors who own the firm's common stock affects stock performance. To identify these habits, a pre-processing method using standard technical indices and investment experience is frequently required. Rather than conducting function selections, a rigorous statistical analysis is also conducted based on a special dataset to assert new features. Some information, such as the percentage of an index's fluctuation, has been shown to influence stock performance. We assume that extracting new features from data and integrating them with existing common technological indices would support existing and well-tested prediction models significantly.

III. EXISTING SYSTEM

[3] suggested that a vibrant application should be developed to evaluate and anticipate stock market prices as the basic tool to increase investment interest rates in stock markets. In this report, explained the development and implementation of a stock market prediction application using the machine learning algorithm (Fig 1). Fundamental analysis, technical analysis and the application of machine learning were three different approaches to the solution of the problem proposed.

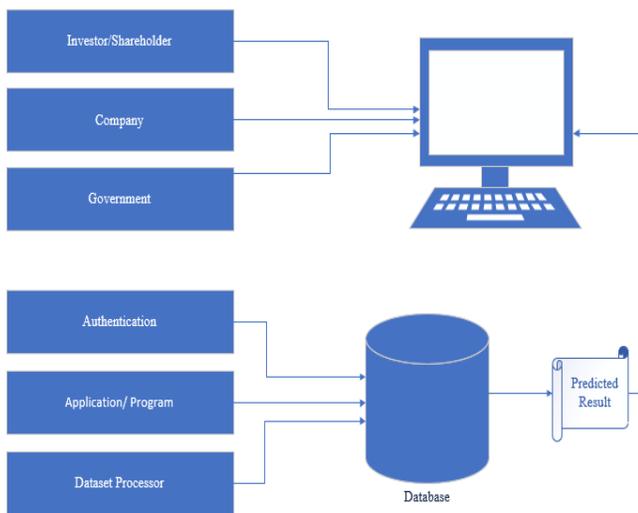


Fig 1: Existing Model Architecture. Source: Samarth (2019)

In the technical analysis methodology, the study demonstrates a common defect to show that limited information is

produced. Based on the study findings, Quant develops and simulates algorithmic trading programs was applied. In recent decades, several machine learning techniques have been applied in order to study by capturing and repetitive patterns the highly theoretical and speculative nature of the stock market. Different companies use different kinds of analytical instruments for forecasting, and the main objective is to predict exactly what stocks would generate the highest profit level. According to a report by Value Walk, machine-learning hedge funds are already substantially above the performance of general hedge funds and conventional quantum funds. ML and AI systems are extremely helpful tools for people to navigate the investment decision-making process and risk evaluation. The effects of human emotions on trade decisions are often the biggest impediment to output. Algorithms and computers make better decisions and carry out trade more quickly than anyone can, free of emotional influence.

IV. METHODOLOGY

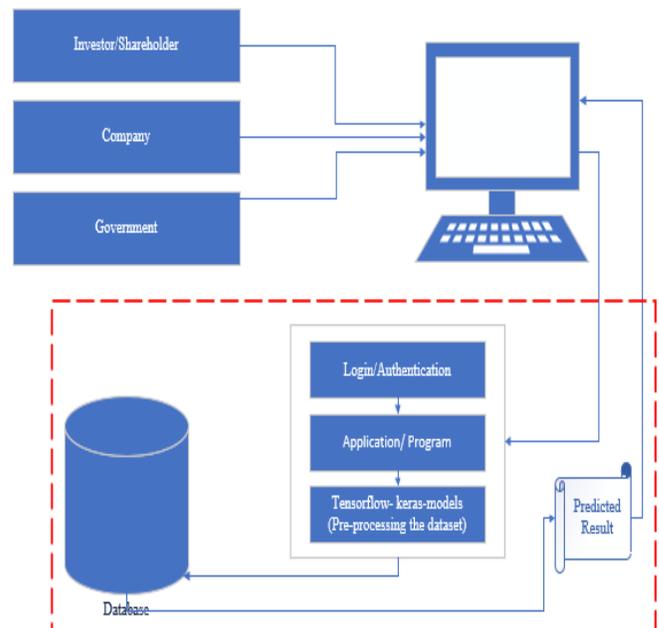


Fig 2: Proposed Architecture

There were several studies on predicting stock price trends for a daily period, mainly for the development of models incorporating various data sources such as news articles, twitter data, google data and Wikipedia data. The effect on stock price movements has been shown by all these external factors in combination with stock prices and technical stock indicators.

The study made use of the yahoo finance module, which is basically a Python scraper that extracts finance data from the Yahoo Finance site. In Fig 2, TensorFlow and Keras, a neural network that predicts stock market prices, are used in this project to forecast stock market prices and building a Recurring Neural Network with LSTM cells as the latest state of the art in time series forecasting.

The main aim of this research is to help stock market investors predict the behaviour and/or direction of stock market prices in order to participate in a long-term profit trend.

V. RESULT AND DISCUSSION

Output:

```
Future price after 15 days is 3171.07$
huber_loss loss: 5.5411544237087697e-05
Mean Absolute Error: 20.882439373811078
Accuracy score: 0.5182978723404256
Total buy profit: 5189.444863319397
Total sell profit: 2290.730220630765
Total profit: 7480.175083950162
Profit per trade: 6.36610645442567
```

Fig 3: Prediction

Using data scraped from Yahoo Finance, which was raw and complex, to train the algorithm. Build an RNN with a dense output layer and one neuron that needs a sequence of features of sequence_length (in this case, 50 or 100) consecutive time steps (days in this dataset) and outputs a single value that indicates the price of the next time phase. Then we used the refined dataset to train the LSTM model, which produced realistic results. TensorBoard was also used to visualize model output during the training phase. The model was trained for 500 epochs (as mentioned previously) after running the code. Fig. 3:

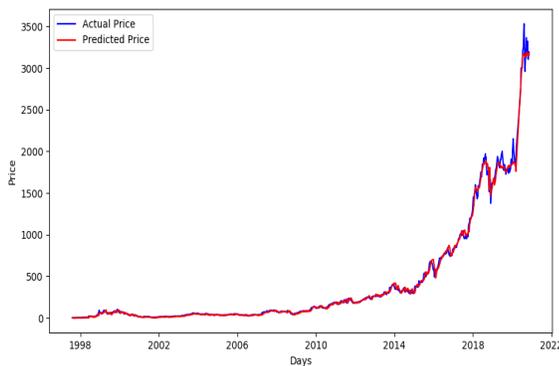


Fig 4: Time Series Prediction Curve showing Yahoo Finance

Chart Explanation and Model Metrics

Fig. 5.4 is a graphical representation of model prediction with the red line indicating prediction curve in Fig. 3. Whilst Fig.

3 shows an almost flat but gradual growth of stock from the year 1998 - 2010. However, 2010 has a significant growth, of which improved through 2018 to 2022. As such our model at the point of testing and training using Yahoo_Fin, assumed that within fifteen (15) days that the price of AMZN amongst other stock will be \$3171.07, using available metrics in python pandas programming language.

Model Metrics

- i. Mean absolute error: we get about 20 error, which means, in average, the model predictions are far by over 20\$ to the true prices, this will vary from ticker to another, as prices get larger, the error will increase as well. Therefore, you can only use this metric to compare the models when the ticker is stable (e.g AMZN).
- ii. Buy/Sell profit: This is the profit we get if we opened trades on all the testing samples, we calculated these on get_final_df() function.
- iii. Total profit: This is simply the sum of buy and sell profits.
- iv. Profit per trade: The total profit divided by the total number of testing samples.
- v. Accuracy score: This is the score of how accurate our predictions are, this calculation is based on the positive profits we get from the all the trades from the testing samples.

VI. CONCLUSION

In recent years, it has been noted that most people are investing in the stock market in order to make fast money. At the same time, an investor stands a good risk of losing all his or her money. As a result, for the consumer to understand future market trends, an effective predictive model is needed. There are several predictive models that tell whether the economy is going up or down, but they do not always produce reliable results. An attempt has been made to develop an effective stock market predictive model that forecasts the next day's trend using Tensorflow and Keras, the model predictions are far by over \$20 to the true prices, this will vary from ticker to another, as prices get larger, the error will increase as well. Therefore, you can only use this metric to compare the models when the ticker is stable.

REFERENCES

- [1] Atsalakis GS, Valavanis KP. (2018). Forecasting stock market short-term trends using a neuro-fuzzy based methodology. *Expert Syst Appl.* 2009;36(7):10696–707. *Computer Systems*, vol. 79, pp. 960–972.
- [2] Fischer T, Krauss C. (2018). Deep learning with long short-term memory networks for financial market predictions. *Eur J Oper Res.* 270(2):654–69.
- [3] Gareja Pradip, Chitrak Bari, J. Shiva Nandhini (2018). Stock market prediction using machine learning. *International Journal of Advanced Research and Development.*
- [4] Khedr, Ayman E., and Nagwa Yaseen. (2017). Predicting stock market behaviour using data mining technique and news sentiment analysis." *International Journal of Intelligent Systems and Applications* 9.7.

- [5] Lei L. (2017). Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. *Appl Soft Comput J.* 2018;62:923–32.
- [6] Lin Z. (2018). Modelling and forecasting the stock market volatility of SSE composite index using garch models, *Future Generation*
- [7] Maini, Sahaj Singh, and K. Govinda. (2017). Stock market prediction using data mining techniques." *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE.
- [8] Nekoeiqachkanloo H, Ghojogh B, Pasand AS, Crowley M. (2019). Artificial counselor system for stock investment. *ArXiv Preprint arXiv :1903.00955*.
- [9] Sun J., Xiao K., Liu C., Zhou W., and Xiong H. (2019). Exploiting intra-day patterns for market shock prediction: a machine learning approach," *Expert Systems With Applications*, vol. 127, pp. 272–281.
- [10] Tsai CF, Hsiao YC. (2010). Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches. *Decis Support Syst.*
- [11] Weng B, Lu L, Wang X, Megahed FM, Martinez W. (2018). Predicting short-term stock prices using ensemble methods and online data sources. *Expert Syst Appl.*
- [12] McNally S, Roche J, Caton S. (2018). Predicting the price of bitcoin using machine learning. In: *Proceedings—26th Euromicro international conference on parallel, distributed, and network-based processing, PDPpp.339–43.*