Visualizing and Forecasting Stocks Using Machine Learning

Harshal Pujari, Akshata Ubale, Shubham Patil, Atharva Shrivastav, Prof. Vrushali Kondhalkar

Students, Dept. of Computer Engineering, Jayawantrao Sawant College of Engineering, Pune, Maharashtra, India

---***---

Abstract - *An Accurately predicting stock market returns is a veritably grueling task due to its unpredictable, arbitrary, and rapid-changing nature. With the preface of Machine Learning techniques, programmed style prediction of stock market returns has been proved to be more effective. Although there are various models in Machine Learning to prop prediction, this work substantially focuses on the use of the Regression model and the LSTM model for the prediction of stock market returns.*

*Key Words***:   LSTM, Prediction, Stock, Regression, Dataset, etc.**

1. INTRODUCTION

The stock market is characterized as dynamic, arbitrary, and unsystematic in nature. There are various factors that affect stock prices like political conditions, the global economy, etc., making stock prediction a grueling task. Thus, using Machine Learning techniques to predict stock values by observing trends, could prove highly effective. [1]

In Machine Learning, the dataset is the most significant part. Even a little change in data can immortalize massive changes. So, data should be as refined and concrete as possible. For this work, the dataset is attained from Yahoo Finance. Yahoo Finance is a part of the Yahoo network which allows accessing datasets of stocks of several companies.

The regression model and LSTM model are considered for this work. Regression serves the purpose of keeping errors as low as possible and LSTM grants memory for the data and results to be used for the long run. The graph of the actual and the predicted value of stocks is plotted using both regression and LSTM techniques. The remaining paper consists of the following: Section 2 discusses the related work. Section 3 puts forward the two models used and the methods used in them in detail. Section 4 discusses the results produced with different plots for both models in detail. Section 5 has the conclusion and the last section contains the references.

2. RELATED WORK

From the literature survey, it is clear that the machine learning techniques is applied for stock market vaticination across the world. Compared to contemporary vaticination techniques, these techniques are much more accurate.

The model developed by Kim and Ha in [2] is a blend of artificial neural networks (ANN) and genetic algorithms (GAs). They discretized the features for predicting the stock price index. They incorporated data from technical indicators and the daily Korea stock price index (KOSPI). The data accommodated 2928 trading days, stretching from January 1989 to December 1998. They applied optimization of feature discretization, which is a technique akin to dimensionality reduction. They introduced genetic algorithms (GA) to enhance the Artificial Neural networks (ANN). Limitation of their work is that they focused only on two factors in optimization. They believed that the genetic Algorithm has a substantial prospective for feature discretization optimization.

Qiu and Song in [3] also introduced a solution that was based on an optimized artificial neural network (ANN) model. In this work, the authors have utilized genetic algorithms and an artificial neural network-based model and named it as GA-ANN model.

For data mining applications, Piramuth in [4] organized an in-depth evaluation of various feature selection methods. The datasets which were credit approval data, tam, and kiang data, were used. It compared how various feature selection methods optimized decision tree performance. The featured selection methods like probabilistic distance measures: the Bhattacharyya measure, the Mahalanobis distance measure, the Matusita measure, the divergence measure, and the Patrick-Fisher measure; were compared. The advantage of this paper is that the author analyzed both feature selection methods, i.e., probabilistic distance-based and several inter-class feature selection methods. Another strength is that the evaluation was performed, based on different datasets. However, only decision tree was used in this work as evaluation algorithm. So, it is difficult to conclude if the feature selection method still works the same way on the larger and complex dataset or model.

Hasan and Nat in [5] forecasted the stock market for stock prices of four distinct Airlines. They used the Hidden Markov Model (HMM) for prediction. The states of the model were reduced down to four states: The opening price, the closing price, the highest price, and the lowest price. The strength of this paper is that no expert knowledge is needed in this approach to design a prediction system. On the other hand, the dataset used for the training and testing purposes of this model is really small. A maximum of 2 years is selected as the data range of the training and testing dataset.

Lei in [6] applied Wavelet Neural Network to forecast patterns in stock price. For optimization, the author also used Rough Set (RS) as an attribute reduction technique. A rough set was used to reduce the dimension for stock price pattern features. For determining the structure of the Wavelength Neural Network, a rough set is used. The dataset used for this work is made up of the five popular stock market indices, specifically: (1) SSE Composite Index (China), (2) Nikkei 225 Index (Japan), (3) Dow Jones Index (USA), (4) CSI 300 Index (China), and (5) All Ordinaries Index (Australian). The computational complexity can be diminished by using Rough Set for optimizing feature dimensions before processing. On the other hand, the author only focused on the parameter adjustment and the discussion part and did not mention the flaw of the model itself. In this model, appraisals were performed on the indices, but this model may not have an identical performance if used on a distinct stock.

Lee in [7] used the support vector machine (SVM) and hybrid feature selection method for estimating the stock market trends the dataset used in this work is from dataset of the NASDAQ Index in the Taiwan Economic Journal Database (TEJD) in 2008. The role of the wrapper was played by Supported sequential forward search (SSFS) whereas the feature selection part was done using a hybrid method. The strong point of this work is that they built an elaborate procedure of parameter adjustment with performance under various parameter values. The downside of this work is that the performance of SVM was compared to the back-propagation neural network (BPNN) only. They did not compare the performance of SVM to the other machine learning algorithms.

Ni el at. in [8] used SVM for predicting stock price patterns. Here fractal feature selection was used for optimization. The author used a dataset from Shanghai Stock Exchange Composite Index (SSECI), along with 19 technical indicators as features. They also used a K-cross validation which is a grid search method, for searching the best parameter combination. The weak point of this work is that they only considered the technical indicators. The macro and micro factors in the financial domain were ignored. While testing hyper-parameter combinations, the author also mentioned a method called k cross-validation.

3. METHODOLOGY

There are innumerable factors that affect stock values. They may not seem statistical at first, making stock trend prediction a complex problem. But by proper knowledge and application of Machine Learning, we can find patterns and trends in data and train our machine learning model to make appropriate predictions

The dataset being employed for analysis is picked from Yahoo Finance. The data reflects the stock prices at certain time intervals for each day of the year. It comprises various sections videlicet date, symbol, open, close, low, high, and volume. The data was considered from only one company for simulation and analysis purposes. The complete data was available in CSV format. It was first to read and then, with the help of Pandas library in Python, it was transformed into a data frame. From this, the data for one particular company was uprooted by segregating data on the basis of the symbol field. After this, with the help of the sklearn library in Python, the data was normalized and then divided into two sets: the training dataset and the testing dataset.

Although machine learning is similar and has numerous models, this paper focuses on two of the most important amongst them and made the prognostications using these.

A. Regression Based Model

The regression model is a supervised machine learning model. Regression analysis is a statistical method. It helps to build a relationship model between a dependent i.e., target variable, and one or more independent i.e., predictor variables. Regression analysis explains the variation of the value of the target variable in tune with a predictor variable when other predictor variables are constant.

$$Y = mX + c$$

Here,

Y   =   dependent   variables   (Output),
X   =   Independent   variables   (Input),

m and c are the linear coefficients

B. Long Short-Term Memory (LTSM)

LSTM is a special version of RNN which solves the short-term memory problem. Long Short-Term Memory (LSTM) networks are a type of recurrent neural network. They are proficient in learning order dependence in sequence prediction problems. A regular LSTM unit has four components: a Cell, an Input Gate, an Output Gate, and a Forget Gate. The values are held by the cell for a random time interval. The inflow and outflow of information are maintained by three gates. LSTM is usually preferred for classifying, processing, and prognosticating the given time series of unknown duration.

The main reason behind using the LSTM model for stock trend vaticination is that the prediction of trends depends on a large dataset and thus, they are dependent on long - term history of the market. Test accuracy of the LTSM algorithm is around 72% which is a high number.

LSTM also solves the problem of Vanishing Gradient which occurs because of the processing of huge data. This makes LSTM a better choice for this work. In this, the gradient with respect to the weight matrix may become genuinely minute. This make deteriorates the learning rate of the model. The Remembering Cell in LSTM remembers the

value for long-term propagation and Gates regulates the flow of information.

4. CONCLUSIONS

This paper is an attempt to determine the future stock trends of various companies with greater accuracy and reliability using machine learning techniques.

Both the techniques have shown an enhancement in the accuracy of prognostications, thereby yielding positive results with the LSTM model proving to be more effective. The results are quite promising and have led to the conclusion that it is possible to prognosticate the stock market with further accuracy and efficiency using machine learning techniques.

In the future, the accuracy of the stock market prediction system can be further bettered by exercising a much bigger dataset than the one being utilized currently. Furthermore, other emerging models of Machine Learning could also be studied to check for the accuracy rate resulting from them. Sentiment analysis through Machine Learning on how news affects the stock prices of a company is also a very promising area. Other deep learning-based models can also be used for prediction purposes.

REFERENCES

[1] Masoud, Najeb MH. (2017) "The impact of stock market performance upon economic growth." International Journal of Economics and Financial Issues 3 (4): 788–798.

[2] Kim K, Han I. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Syst Appl. 2000; 19:125–32.

[3] Qiu M, Song Y. Predicting the direction of stock market index movement using an optimized artificial neural network model. PLoS ONE. 2016;11(5): e0155133.

[4] Piramuthu S. Evaluating feature selection methods for learning in data mining applications. Eur J Oper Res. 2004;156(2):483–94.

[5] Hassan MR, Nath B. Stock market forecasting using Hidden Markov Model: a new approach. In: Proceedings—5th international conference on intelligent systems design and applications 2005, ISDA'05. 2005. pp. 192–6.

[6] Lei L. Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. Appl Soft Comput J. 2018; 62:923–32.

[7] Lee MC. Using support vector machine with a hybrid feature selection method to the stock trend prediction. Expert Syst Appl. 2009;36(8):10896–904.

[8] Ni LP, Ni ZW, Gao YZ. Stock trend prediction based on fractal feature selection and support vector machine. Expert Syst Appl. 2011;38(5):5569–76.