International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 06 | June 2020 www.irjet.net p-ISSN: 2395-0072

STOCK MARKET PREDICTION USING SENTIMENT ANALYSIS & LSTM

Karan Soni[1], AbdulRehman Shariff[2], Muskan Shah[3], Allen Lopes[4]

[1]*Karan Soni (BE IT) - Kandivali(west), Mumbai-67*
[2]*AbdulRehman Shariff (BE IT) – Andheri(west), Mumbai-58*
[3]*Muskan Shah (BE IT) – Malad(west), Mumbai-95*
[4]*Professor Allen Lopes, Dept. of Information Technology Engineering, Universal College of Engineering, Vasai, India.*

Abstract - *Recently, there has been a rapidly growing interest in deep learning research and their applications to real-world problems. In this paper, we aim at evaluating and comparing LSTM deep learning architectures for short- and long-term prediction of financial time series. This problem is often considered as one of the most challenging real-world applications for timeseries prediction. Unlike traditional recurrent neural networks, LSTM supports time steps of arbitrary sizes and without the vanishing gradient problem. We consider both bidirectional and stacked LSTM predictive models in our experiments and also benchmark them with shallow neural networks and simple forms of LSTM networks. The evaluations are conducted using a publicly available dataset for stock market closing prices. Overall, the ultimate goal of this project is to forecast how the market will behave in the future via sentiment analysis on a set of tweets over the past few days, as well as to examine if the theory of contrarian investing is applicable. The final results seem to be promising as we found correlation between sentiment of tweets and stock prices.*

Key Words: datasets, api, lstm, sentiments, python libraries.

1. INTRODUCTION

Neural networks are widely used for estimation and approximation of unknown complicated functions and systems that may depend on a large number of input variables. Artificial neural networks are commonly modelled as layers of neurons, exchanging information with each other. Connections between neurons have some numeric weights that are adjusted during the learning process. Neural networks obtain intelligent behaviour by learning from the provided data or from interactions with environment. Due to the fact that they can learn complex non-linear mappings, neural networks are widely used for solving advanced problems, such as pattern recognition and classification. The process of learning the data by the neural network is called training, and while there are several approaches to neural network training, for this study the supervised learning (training) methods are employed. In supervised training pairs of input-output data must be provided to the network, and the network will try to learn the mapping implied by that data. The robustness of the neural network often depends on the training algorithm that was used to learn the data. There are two training methods that will be compared in this paper – backpropagation and resilient backpropagation over the last decade neural networks have been proven to be one of the most powerful tools in modelling and forecasting. Recently neural networks have expanded their forecasting applications to many areas, such as: urban traffic state predictions, disease predictions, earthquake magnitude forecasting, river flow forecasting, air quality and pollution forecasting, solar power forecasting, weather forecasting. Forecasting with neural networks is also widely used in analysing complex financial systems and market based relations: credit risk evaluations, gas prices and production level predictions, forecasting of vehicle sales, forecasting demand on consumable parts in production predicting tourism demand, forecasting airline data forecasting stock index prices and currency exchange rates Since neural networks proved to be reliable for chaotic time series forecasting, financial firms worldwide are employing neural networks to solve difficult prediction problems; and it is anticipated that neural networks will eventually outperform even the best traders and investors Farmers and plant caretakers in nursery could be benefited a lot with an early disease detection application that can be used anywhere, in order to prevent their plants and let the human know what has to be done beforehand.

2. LITERATURE SURVEY

All ANNs are being used in various fields of application including business forecasting, credit scoring, bond rating, business failure prediction, medicine, pattern recognition, image processing, speech processing, computer vision and control systems. In the context of financial forecasting, Kuan and Liu [2] show that a properly designed ANN has lower out-of-sample mean squared prediction error relative to the random walk model. Basic and Wood [3] discuss the profitability of trading signals generated from the out-of sample short-term predictions for daily returns of S&P 500, DAX, TOPIX and FTSE stock market indices evaluated over the period 1965–99. The out of sample prediction performance of neural networks is compared against a benchmark linear autoregressive model. They found that the buy and sell signals derived from neural network predictions are significantly different from unconditional one-day mean return and are likely to

provide significant net profits for reasonable decision rules and transaction cost assumptions. They report that ANNs outperform the linear models from financial forecasting literature in terms of its predictive power. It is reported that the neural network models yield statistically lower forecast errors for the year-over year International Journal of Computer Applications (0975 – 8887) Volume 99– No.9, August 2014 5 growth rate of real GDP relative to linear and univariate models. Huang et al. [4] report a comparative study of application of Support Vector Machines (SVM) and Backpropagation Neural Networks (BPNN) for an analysis of corporate credit ratings. They report that the performances of SVM and BNN in this problem were comparable and both these models achieved about 80 per cent prediction accuracy. It is reported that ANNs perform better than the statistical discriminant analysis both for training and hold-out samples. All the recent literatures [5, 6] have also echoed the same about neural networks.

3. PROPOSED SYSTEM

1) Sentimental Analysis

Sentiment Analysis is a term that you must have heard if you have been in the Tech field long enough. It is the process of predicting whether a piece of information (i.e. text, most commonly) indicates a positive, negative or neutral sentiment on the topic.

Section A: Preparing the Test Set

- Step A.1: Getting the authentication credentials
- Step A.2: Authenticating our Python script
- Step A.3: Creating the function to build the Test set

Section B: Preparing the Training Set

Section C: Pre-processing Tweets in The Data Sets

Section D: Naive Bayes Classifier

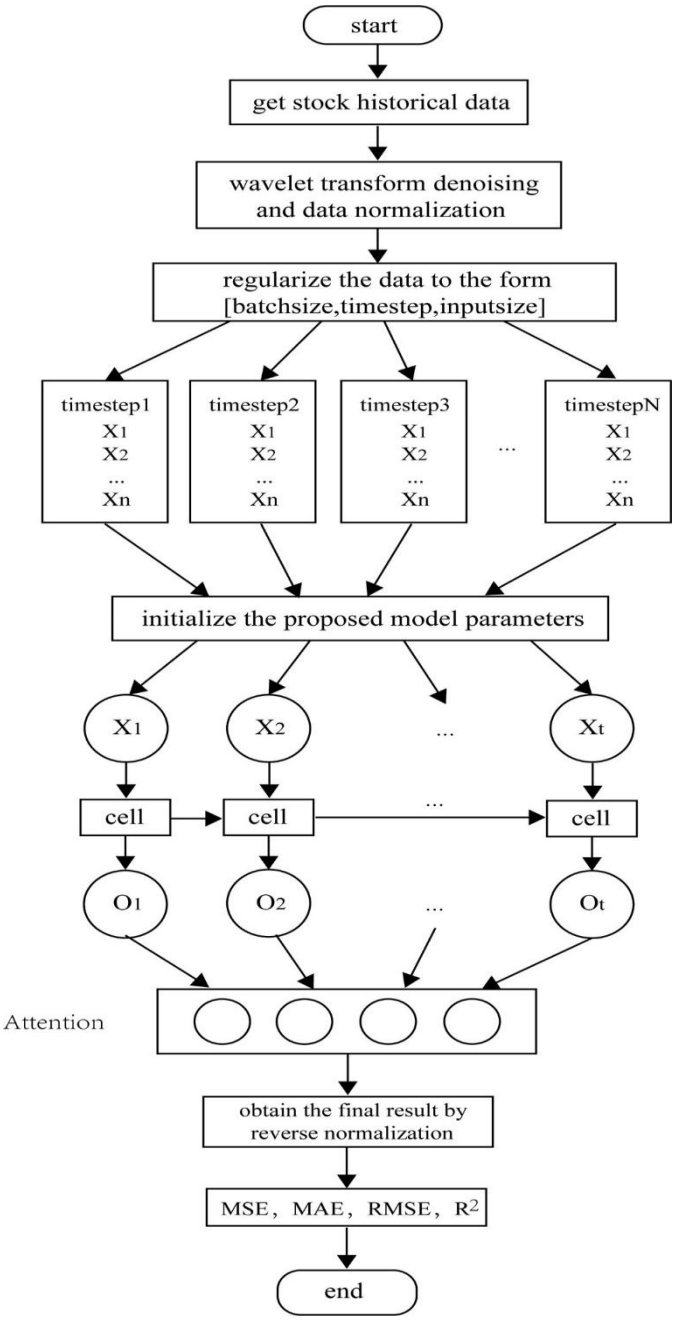
- Step D.1: Building the vocabulary
- Step D.2: Matching tweets against our vocabulary
- Step D.3: Building our feature vector
- Step D.4: Training the classifier

When all the above process is completed the output i.e. The sentimental analysis will be generated for the selected company. It will display the Percentage (%) of Positive (+), Negative (-) and Neutral Comments passed in tweeter to determine what opinion does people have for the stock prices of the company.

2) Long shot-term memory neural networks (LSTM)

LSTM uses one of the most common forms of RNN. This time recurrent neural network is meant to avoid long-term dependence problems and is suitable for processing and predicting time series. Proposed by Sepp Hochreiter and Jurgen Schmidhuber in 1997.

Proposed prediction model



To establish a stock index price forecasting model has three stages: 1) data collection and preprocessing, 2)

model establishment and training, 3) and evaluation and results as shown in Fig.

Data Source

We obtain the data from the datasets of different company's which contains previous years stock data prices.

Data Preprocessing

We implemented the proposed stock forecasting method in Python using TensorFlow. We used zero-mean normalization to the data and divided it into training and test datasets.

Evaluation

We processed stock index datasets in the LSTM model: the LSTM model with the wavelet transform. We run the datasets with the trained data in order to generate the prediction. In this process the most important aspect is to get the highest accuracy possible of the result.

4. PROPOSED SYSTEM

1) Sentimental Analysis

The result will display the percentage (%) of positive (+), negative (-) and neutral tweets for the company. Below the percentage it displays the positive, negative and neutral comments of Tweeter.

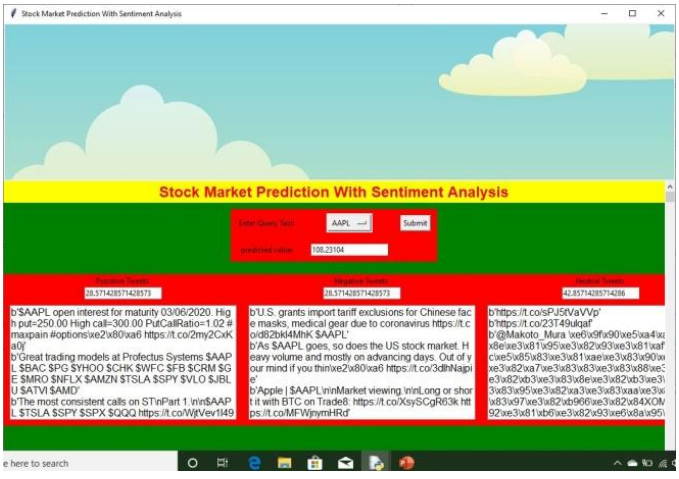


Fig. 1

2) LSTM

The Predicted stock prices will be displayed along with a graph showing the prediction for the next one month.

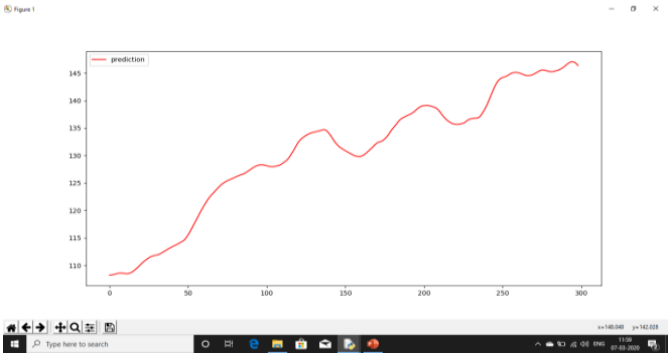


Fig. 2.1

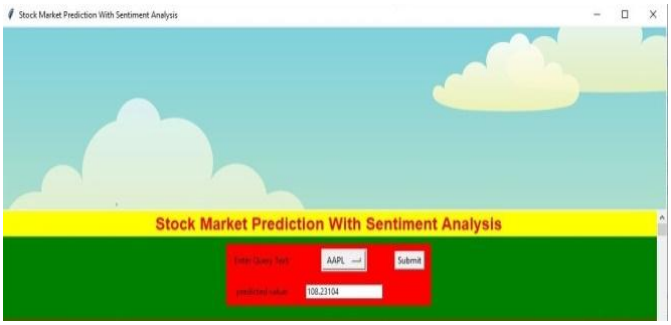


Fig. 2.2

5. CONCLUSION

This paper establishes a forecasting framework to predict the prices of stocks. We processed stock data through a wavelet transform and used LSTM neural network to predict the stock price, with excellent results. Our proposed model has a better fitting degree and improved accuracy of the prediction results. Therefore, the model has broad application prospects and is highly competitive with existing models.

6. FUTURE SCOPE

Our future work has several directions. Our work has found that LSTM has more predictive outcomes for price prediction than other methods. However, simply considering the impact of historical data on price trends is too singular and may not be able to fully and accurately forecast the price on a given day. Therefore, we can add data predictions related to stock-related news and basic information, so as to enhance the stability and accuracy of the model in the case of a major event.

REFERENCES

[1] W. Antweiler and M. Frank. Do US stock markets typically overreact to corporate news stories? Working Paper, (1998):1–22, 2006.

[2] K. A. A. Devitt, "Sentiment polarity identification in financial news: A cohesion-based approach, Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL-07), pp. 984-991, June 2007.".

[3] "Prediction with LSTM," [Online]. Available: https://lilianweng.github.io/lillog/2017/07/08/predict-stock-prices-using-RNN-part-1.html. K. Elissa, "Title of paper if known," unpublished.

[4] "Deep learning for complete beginners: convolutional neural networks with keras," [Online]. Available: STOCK PRICE PREDICTION USING DEEP LEARNING 53 https://cambridgespark.com/content/tutorials/convolutional-neural-networks-withkeras/index.html.

[5] W. L. Shaosheng Cao, "Improving Word Embeddings with Convolutional".

[6] "Recurrent neural Network - Wikipedia," [Online]. Available: https://en.wikipedia.org/wiki/Recurrent_neural_network.

[7] S. Hochreiter, "The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 06, no. 02, 1998.

[8] J. S. Sepp Hochreiter, "Long Short Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, 1997.

[9] "Understanding LSTM networks," [Online]. Available: http://colah.github.io/posts/2015-08-Understanding-LSTMs/.