Stock Price Prediction Using LSTM

UNITED INTERNATIONAL UNIVERSITY

Project Report

Machine Learning

By

MD Al-Amin,
ID : 011161268

Contents

1 Abstract

Predicting stock market is one of the most difficult tasks in the field of computation. There are many factors involved in the prediction – physical factors vs. physiological, rational and irrational behavior, investor sentiment, market rumors,etc. All these aspects combine to make stock prices volatile and very difficult to predict with a high degree of accuracy. We investigate data analysis as a game changer in this domain.As per efficient market theory when all information related to a company and stock market events are instantly available to all stakeholders/market investors, then the effects of those events already embed themselves in the stock price. So, it is said that only the historical spot price carries the impact of all other market events and can be employed to predict its future movement. Hence, considering the past stock price as the final manifestation of all impacting factors we employ Machine Learning (ML) techniques on historical stock price data to infer future trend. ML techniques have the potential to unearth patterns and insights we didn't see before, and these can be used to make unerringly accurate predictions. We propose a framework using LSTM (Long ShortTerm Memory) model to predict the future stock price.Along with the advance of opinion mining techniques, public mood has been found to be a key element for stock market prediction. However, how market participants' behavior is affected by public mood has been rarely discussed. Consequently, there has been little progress in leveraging public mood for the asset allocation problem, which is preferred in a trusted and interpretable way. In order to address the issue of incorporating public mood analyzed from social media, we propose to formalize public mood into market views, because market views can be integrated into the modern portfolio theory.

2 Introduction

STOCK time series forecast is one of the main challenges for machine learning technology because the time series analysis is required [1]. Two methods are usually used to predict financial time series: machine learning models and statistical methods [2]. Statistical methods can be used to predict a financial time series. The common methods are autoregressive conditional heteroscedastic (ARCH) methods [3], and autoregressive moving average (ARMA) [4] or an autoregressive integrated moving average (ARIMA) methods. However, traditional statistical methods generally assume that the stock time series pertains to a linear process, and model the generation process for a latent time series to forecast future stock prices [5]. However, a stock time series is generally a dynamic nonlinear process [6]. Many machine learning models can capture nonlinear characters in data without prior knowledge [7]. These models are always used to model a financial time series. The most commonly used models for stock forecasts are artificial neural networks (ANN), Long-Short Term Memory (LSTM), and hybrid and ensemble methods. Artificial neural networks have found many applications in business because they can deal with data that is non-linear, non-parametric, discontinuous or chaotic for a stock time series [8].

Sales and macroeconomic factors are some of the driving forces behind stock movements but there are many others. For example, the subjective views of market participants also have important effects. Along with the growing popularity of social media in the past decades, people tend to rapidly express and exchange their thoughts and opinions. As a result, the importance of their views has dramatically risen [6]. Currently, stock movements are considered to be essentially affected by new information and the beliefs of investors.

The paper that we have presented modeled and predicted the stock returns of NIFTY 50 using LSTM. We collected 5 years of historical data of NIFTY 50 and used it for the training and validation purposes for the model. The next section of the paper will be methodology where we will explain about each process in detail. After that, we will have pictorial representations of the analysis that we have used and we will also reason about the results achieved.

3 Motivation

Machine learning has many applications, one of which is to forecast time series. One of the most interesting (or perhaps most profitable) time series to predict are, arguably, stock prices.A good stock market predictor can helps us many ways.It can saves/utilize our money.Their is a big risk to invest our money in business.And a stock market predictor can helps us when and where we have to invest our money. As a example the bullish market turned bearish during November 2010, with the exchange losing 1,800 points between December 2010 and January 2011. Millions of investors have been rendered bankrupt as a result of the market crash. The crash is believed to be caused artificially to benefit a handful of players at the expense of the big players. so in this situation it could be play a big role.

4 Methodology

Various types of neural networks can be developed by the combination of different factors like network topology, training method etc. For this experiment, we have considered Recurrent Neural Network and Long Short-Term Memory.
This section we will discuss the methodology of our system. Our system consists of several stages which are as follows:-
Stage 1: Raw Data:
In this stage, the historical stock data is collected from https://www.quandl.com/data/NSE and this historical data is used for the prediction of future stock prices.
Stage 2: Data Preprocessing:
The pre-processing stage involves
a) Data discretization: Part of data reduction but with particular importance, especially for numerical data.
b) Data transformation: Normalization.
c) Data cleaning: Fill in missing values
. d) Data integration: Integration of data files.
After the dataset is transformed into a clean dataset, the dataset is divided into training and testing sets so as to evaluate. Here, the training values are taken as the more recent values. Testing data is kept as 5-10 percent of the total dataset.
Stage 3: Feature Extraction:
In this layer, only the features which are to be fed to the neural network are chosen. We will choose the feature from Date, open, high, low, close, and volume.
Stage 4: Training Neural Network:
In this stage, the data is fed to the neural network and trained for prediction assigning random biases and weights. Our LSTM model is composed of a sequential input layer followed by 2 LSTM layers and dense layer with ReLU activation and then finally a dense output layer with linear activation function.
Stage 5: Output Generation:
In this layer, the output value generated by the output layer of the RNN is compared with the target value. The error or the difference between the target and the obtained output value is minimized by using back propagation algorithm which adjusts the weights and the biases of the network.

Here is the some sample data below:

Date	High	Low	Open	Close	Volume	Adj Close
2011-01-03	47.180000	46.405716	46.520000	47.081429	111284600.0	40.868607
2011-01-04	47.500000	46.878571	47.491428	47.327145	77270200.0	41.081905
2011-01-05	47.762856	47.071430	47.078571	47.714287	63879900.0	41.417946
2011-01-06	47.892857	47.557144	47.817142	47.675713	75107200.0	41.384472
2011-01-07	48.049999	47.414288	47.712856	48.017143	77982800.0	41.680836

Table 1: Sample Data

5 Analysis

For analyzing the efficiency of the system we are used the Root Mean Square Error(RMSE). The error or the difference between the target and the obtained output value is minimized by using RMSE value. RMSE is the square root of the mean/average of the square of all of the error. The use of RMSE is highly common and it makes an excellent general purpose error metric for numerical predictions. Compared to the similar Mean Absolute Error, RMSE amplifies and severely punishes large errors.

$$RMSE = \sqrt{\frac{1}{N}\sum(X_i - Y_i)^2}$$

6 Discussion

First of all it is little bit tough to find a suitable paper for my project topic with the proper data set the have used. After a lot of searching in the google scholar i found a paper that is ”Predicting Stock Prices Using LSTM ” with the data set they have used.After Finding the paper with the data set the main challenges are come.I have to build something that that i have know idea.I have search a lot about ”LSTM” in the google and YouTube .I have read some blog and saw a lots of video about this topic that's help me a lot.After understand the model i started to build the model.I also found difficulty to build the ”Sigmoid” function.After all i can build the model but their i face another problem that is the accuracy.The model i build that give us the error seven point something but the paper which i am trying to replicate they have error below zero.And that is a huge different .But i doesn't get the mistake i did.After searching my mistake a lot of time i found that was a very silly mistake.That was the error function.I am use the error function which one i saw in the tutorial but the in the paper they uses a different error function.After fix that the result are so close but not the same.Last of all i have learned a lot for finish the project.

7 Results

Data Set	Paper Result	Our Result
Training Data	0.00983	0.29536
Test Data	0.00859	0.53948

Table 2: This Table show the comparison of result

8 Graphs

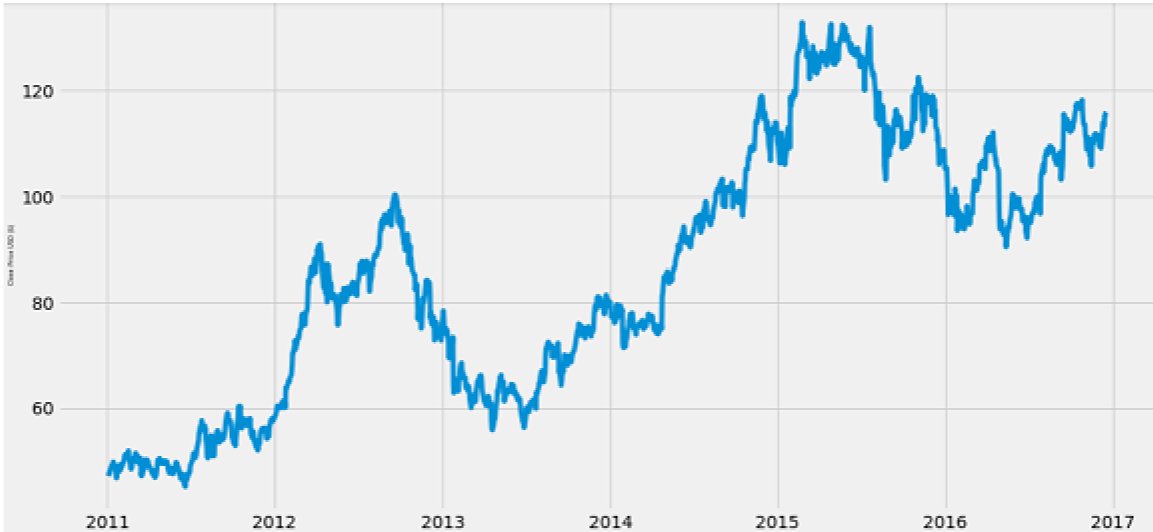

Figure 1: Training Data graph

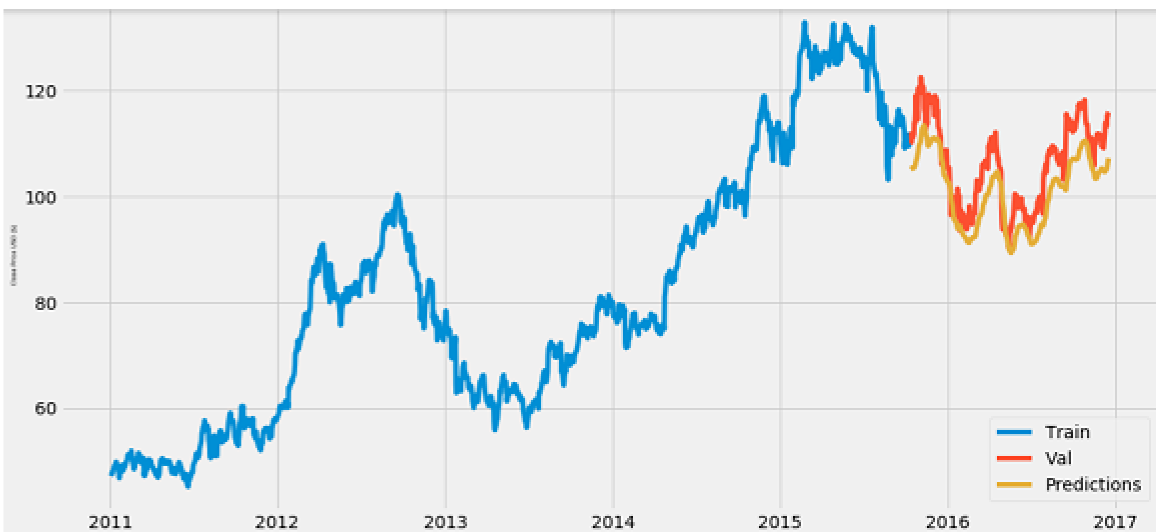

Figure 2: Testing Data graph

9 Conclusion

The popularity of stock market trading is growing rapidly, which is encouraging researchers to find out new methods for the prediction using new techniques. The forecasting technique is not only helping the researchers but it also helps investors and any person dealing with the stock market. In order to help predict the stock indices, a forecasting model with good accuracy is required.

10 References

[1] F. E. Tay and L. Cao, “Application of support vector machines in financial time series forecasting,” Omega, vol. 29, no. 4, pp. 309–317, 2001.
[2] J.-Z. Wang, J.-J. Wang, Z.-G. Zhang, and S.-P. Guo, “Forecasting stock indices with back propagation neural network,” Expert Systems with Applications, vol. 38, no. 11, pp. 14 346–14 355, 2011.
[3] R. F. Engle, “Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation,” Econometrica: Journal of the Econometric Society, pp. 987–1007, 1982.
[4] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time Series Analysis: Forecasting and Control. John Wiley Sons, 2015.
[5] D. A. Kumar and S. Murugan, “Performance analysis of indian stock market index using neural network time series model,” in Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on. IEEE, 2013, pp. 72–78
[6] Y.-W. Si and J. Yin, “Obst-based segmentation approach to financial time series,” Engineering Applications of Artificial Intelligence, vol. 26, no. 10, pp. 2581–2596, 2013.
[7] G. S. Atsalakis and K. P. Valavanis, “Surveying stock market forecasting techniques–part ii: Soft computing methods,” Expert Systems with Applications, vol. 36, no. 3, pp. 5932–5941, 2009.
[8] F. Liu and J. Wang, “Fluctuation prediction of stock market index by legendre neural network with random time strength function,” Neurocomputing, vol. 83, pp. 12–21, 2012.