The Prudent Trader: Stock market price predictions using technical analysis and price history.

Arghydeep Giri, Ishaan Paranjape

June 14, 2018

Abstract

Stock market price prediction is one of the most challenging tasks when machine learning is concerned. We built a complete predictor that uses technical analysis indicators and predicts stock prices. Technical analysis takes into account the stock price (and price history) and the price volume. We predict high and close values of the next day. We tried various models and do a comparison of which one is gives us the best prediction score.

1 Introduction

Technical analysis is an analysis methodology for forecasting the direction of prices through the study of past market data, primarily price and volume. The data that we are using is the technical analysis indicators that are calculated with the help of historical stock prices obtained from yahoo finance. The main idea that we are trying to exploit is that prices move in trends and history tends to repeat itself. We try to calculate technical analysis indicators and try to make predictions of the same for the next day to conclude prices and trends for stocks. However some terminology of the indicators we plan to use needs to be defined before moving forward -

1.1 Trend indicators

'Trend' indicates the general direction of the stock market or the price of a security. It is shown with the help of trendlines that connect a series of highs and lows. The most common is the 'moving average' [1] . it helps to smooth out the price by filtering the noise from random price fluctuations. Some of the trend indicators that we are going to use are -

1.1.1 Simple Moving Average

It is an arithmetic moving average that is calculated by adding the closing price of the security for a number of time periods and then dividing this total by the

number of time periods [1]. The weights in a simple moving average are either all equal or not there at all.

1.1.2 Exponential Moving average

It is also a moving average similar to the simple moving average but greater weights are given to the latest data [1].

1.1.3 Moving Average Convergence Divergence

It is a trend following momentum indicator which shows the relationship between two moving averages of the stock prices [1]. This is calculated by subtracting the 26 day exponential moving average from the 12 day EMA. This indicator can tell us a lot about when to buy or sell particular stock. Crossovers are points where MACD falls below a signal line indicating that it is time to sell. When a security price diverges from the MACD means it is the end of a trend. When MACD rises dramatically, i.e. shorter moving average pulls away from the long term moving average indicates that the security is overbought and will soon reach normal levels.

1.2 Momentum Indicators

Momentum is the speed with which the price changes. It is calculated by continually taking the price differences over a fixed price interval. The two momentum indicators we are using are -

1.2.1 Commodity Channel Index - CCI

This tells us the current price level relative to the average price level of a stock. Most of the CCI values are in the range of -100 to 100. A value closer to 100 indicates that the stock is overbought or oversold depending on the uptrend and the downtrend. This indicator alone is not reliable for predicting the trend hence we use relative strength index.

1.2.2 Relative Strength Index - RSI

It is a momentum oscillator which tells us about the speed and change in price movements. RSI oscillates between zero and 100. RSI is considered to be overbought when it is above 70 and oversold when it is below 30. The RSI compares the previous sessions' average gains with respect to the losses. As the number of sessions increases, the RSI value becomes more accurate.

1.3 Volatility Indicators

Degree of fluctuations in the stock market. Shows us the amount of volatility with respect to the moving average.

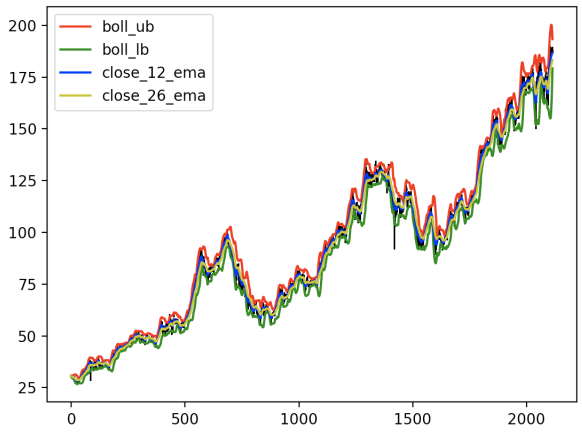

Figure 1: Bollinger bands shown in red and green

1.3.1 Bollinger Bands

Creation of two bands away from the simple moving average based on the standard deviation. Wider bands indicate more volatility. This is shown in fig. 1

2 Prediction Using Technical Analysis Indicators

2.1 Preprocessing

2.1.1 Data [4]

The data that we use is the historical yahoo finance data obtained for various companies from 2010 until the day of the presentation. It is obtained with the $fixyahoofinance$ API. This was used since the official yahoo API has been discontinued. It returns Pandas Dataframe/Panel in the same format as pandas datareader's $getdatayahoo()$. From this the technical analysis indicators are calculated.

2.1.2 Normalization

From the data the redundant and unnecessary data is thrown out and the data is normalized. The data normalization is done using minmax scalar from sklearn. This feature scales and translates each feature individually such that it is in the given range on the training set. Then the N/A values are replaced with the average of the neighbour values. Now the technical indicator values are used as features for learning.

2.1.3 Trend Strength from Technical Analysis indicators

1) Methodology

Now we try to predict the trend strength from these indicators. It is important to note that this is a regression problem. Classification will not give us an accurate trend strength. We use a supervised approach in this case. From the observations it is important to note that values near zero signify a sidewise movement. Many regression models were tried and linear regression was chosen due to its best results. We experimented with Lasso regression, Support vector regressor and ridge regression. From these features, we train the model and try to predict the CCI and RSI values for the next day taking an arbitrary days at a time (upto the day before presenting results).

2) Results

As an example, we take the Apple stock price and do technical analysis on it to obtain the following results -

CCI (20 days)	55
CCI (14 days)	69.97
RSI (12 days)	72.80
RSI (6 days)	69.03
Predicted trend	-0.03

We show the actual stock price for the next day that occured (Source: Google) in Figure 2. From the results we can infer from our previous description that a CCI and RSI value closer to +100 means that the stock is overbought upto that point hence unjustifiably driving the price up and higher the RSI, the more the value has been sustained. Another way of putting it is that it is a sign that the oscillator has reached it's upper bound and is about to fall. This results in the prediction of the downward trend as shown in the value of -0.03.

We obtain a training accuracy of 96.3% and a validtion accuracy of 91.5%. Here the prediction score is an average of many scores. Validation set is used to fine tune the hyperparameters for different predictors. For test data, we take one day's data as the data point to predict the next day's values.

3 Prediction using Price History

3.1 Data

Once again we take the historical data but this time we directly take the open, close, high and low values. This data is then divided into training set, validation set and test set. Each set is normalized using min max scaler before being utilized.

Figure 2: Fall in Apple stock for period of 1 day

3.2 Rolling Regression

In this model, each day's high and low values are predicted from the previous day's Open, high, low and close values. We use the validation set to choose the best regressor as well as for hyper parameter tuning. The regressors to choose from are linear, ridge, lasso and bayesian ridge regressor. These predictions that are calculated give us information about the buying and selling zones for the next day.

3.3 Recurrent Neural Networks

Similar implementation to rolling regression, sequences of length 20 are taken. 19 of those data points are used to predict and train and predict the value of the 20th data point. All 4 values of the data points are predicted and square loss is calculated which is then minimized using AdamOptimizer. It is observed that train square loss reduced from 0.3439 to 0.000058 and validation square loss reduced from 0.5279 to 0.000069.

3.4 Observations and Results

1. Linear regression had a prediction score of 0.98 out of 1 on the validation data.

2. RNN results were more accurate to the rolling regression results. It has an average prediction score of 0.989 out of 1 for all 4 targets (open, close, high and low) using LSTM cells.

3. Basic RNN cells produce results of 0.967 which is quite poor on average.

4. It is also observed that LSTM and GRU cells take almost twice the training time compared to basic RNN cells but have better results.

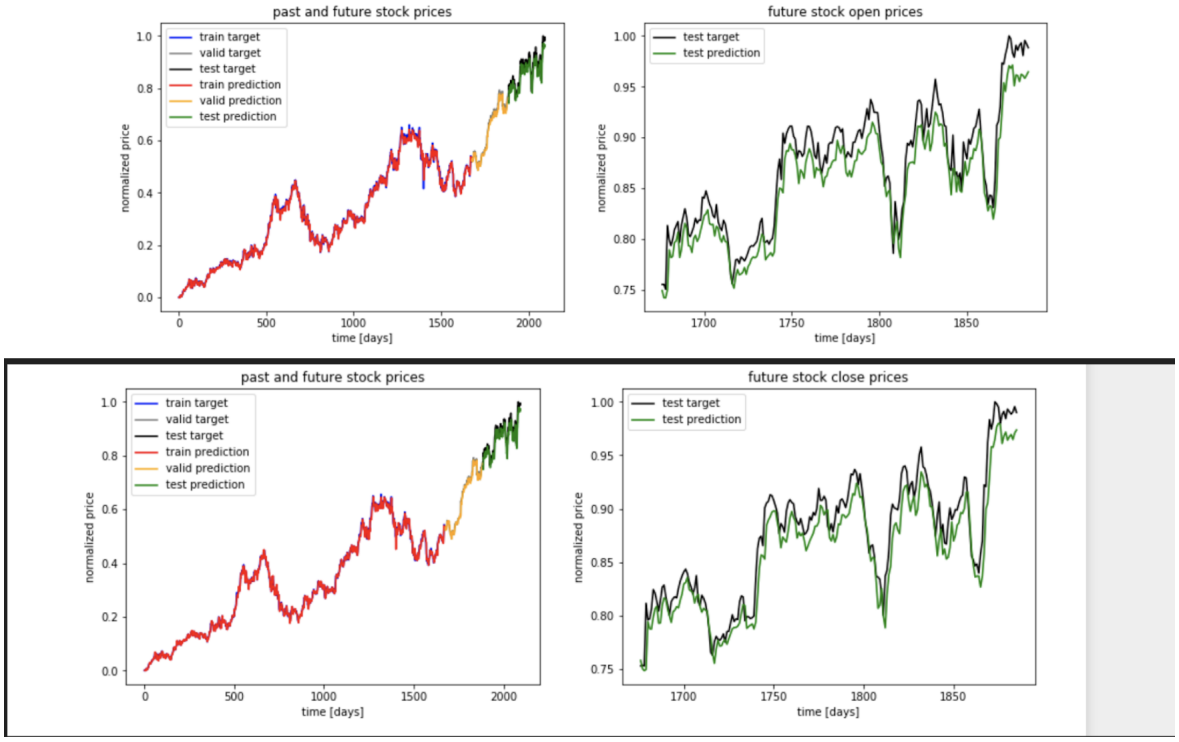

Figure 3: Graph of predicted v. actual stock values

The graphs of predictions with respect to the actual values are shown in Figure 3.

3.5 More Results

We tried our algorithm on some more stocks and got the following results as shown in Table 2.

Stock	**Type of Value**	**Predicted**	**Actual**
Apple	High	192.91	192.88
Apple	Low	191.13	190.70
Tesla	High	348.62	347.20
Tesla	Low	337.53	339.80

4 Suggestions on using this algorithm

We offer some suggestions on using this algorithm. It is important to know the trend and the trend strength so that you know the movement before hand.

Knowing the high and low values to understand the buying and selling regions before the

References

[1] https://www.investopedia.com/terms/m/movingaverage.asp
[2] https://www.investopedia.com/investing/momentum-and-relative-strength-index/
[3] $http://stockcharts.com/school/doku.php?id = chart_school : technical_indicators : commodity_channel_index_cci$
[4] https://finance.yahoo.com/quote/DATA/history?p=DATA
[5] Sezer et. al, "An Artificial Neural Network-based Stock Trading System Using Technical Analysis and Big Data Framework", ACM SE '17 Proceedings of the SouthEast Conference, 223-226
[6] Milosevic, "Equity forecast: Predicting long term stock price movement using machine learning",
[7] Sharma, Bhuriya, et. al, "Survey of stock market prediction using machine learning approach", 2017 International conference of Electronics, Communication and Aerospace Technology (ICECA)