International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 10, pp: 155-158 www.ijaem.net ISSN: 2395-5252

Prediction of Stock Prices (A Machine Learning Approach)

Kirtika Gupta, Prof Dr KC Tripathi and Prof Dr M.L.Sharma
Information Technology Department
Maharaja Agrasen Institute of Technology New Delhi , India

Date of Submission: 20-11-2020 Date of Acceptance: 06-12-2020

ABSTRACT: We employ both random forests and LSTM networks and In this paper, we propose a novel way to minimize the risk of investment in stock market by predicting the returns of a stock using a class of powerful machine learning algorithms known as ensemble learning. Some of the technical indicators such as Relative Strength Index (RSI), stochastic oscillator etc are used as inputs to train our model as training methodologies to analyze their effectiveness in forecasting out-of-sample directional movements of constituent stocks of the S&P 500 from January 1993 till December 2018 for intraday trading. We introduce a multi-feature setting consisting not only of the returns with respect to the closing prices, but also with respect to the opening prices and intraday returns. As trading strategy, we use Krauss et al. (2017) and Fischer & Krauss (2018) as benchmark and, on each trading day, buy the 10 stocks with the highest probability and sell short the 10 stocks with the lowest probability to outperform the market in terms of intraday returns – all with equal monetary weight. Our empirical results show that the multi-feature setting provides a daily return, prior to transaction costs, of 0.64% using LSTM networks, and 0.54% using random forests. Hence we outperform the single-feature setting in Fischer & Krauss (2018) and Krauss et al. (2017) consisting only of the daily returns with respect to the closing prices, having corresponding daily returns of 0 .41% and of 0 .39% with respect to LSTM and random forests, respectively.

I. INTRODUCTION

In the last decade, machine learning methods have exhibited distinguished development in financial time series prediction. Huck **(2009)** and Huck **(2010)** construct statistical arbitrage strategies using Elman neural networks and a multi-criteria-decision method. Takeuchi & Lee **(2013)** evolve a momentum trading strategy. Moritz & Zimmermann **(2014)** apply random forests to construct a trading decision. Tran et al **(2018)**, and Sezer & Ozbayoglu **(2018)** use neural networks for predicting time series data. Borovykh et al. **(2018)** and Xue et al. **(2018)** employ convolutional neural networks, and Siami-Namini & Namin **(2018)** use long short-term memory networks (LSTM).

In my work, I use the results in Krauss et al. **(2017)** and Fischer and Krauss **(2018)** as benchmark.I introduce a multi feature setting consisting not only of the returns with respect to the closing prices, but also with respect to the opening prices and intraday returns to predict for each stock, at the beginning of each day, the probability to outperform the market in terms of intraday returns. As a data set I use all stocks of the S&P 500 from the period of January **2005** until December **2018.** I employ both random forests on the one hand and LSTM networks (more precisely CuDNNLSTM) on the other hand as training methodology and apply the same trading strategy as in Krauss et al. **(2017)** and Fischer & Krauss **(2018)**. My empirical results show that the multi-feature setting provides a daily return, prior to transaction costs, of 0.64% for the LSTM network, and 0.54% for the random forest, hence outperforming the single-feature setting in Fischer & Krauss **(2018)** and Krauss et al. **(2017),** having corresponding daily returns of 0.41% and of 0.39%, respectively.

II. RELATED WORKS

The use of prediction algorithms to determine future trends in stock market prices contradict a basic rule in finance known as the Efficient Market Hypothesis (Fama and Malkiel (1970)). It states that current stock prices fully reflect all the relevant information. It implies that if someone were to gain an advantage by analyzing historical stock data, then the entire market will become aware of this advantage and as a result, the price of the share will be corrected. This is a highly controversial and often disputed theory. Although it is generally accepted, there are many researchers

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 10, pp: 155-158 www.ijaem.net ISSN: 2395-5252

who have rejected this theory by using algorithms that can model more complex dynamics of the financial system.
Several algorithms have been used in stock prediction such as SVM, Neural Network, Linear Discriminant Analysis, Linear Regression, KNN and Naive Bayesian Classifier. Literature survey revealed that SVM has been used most of the time in stock prediction research.

Multiple algorithms were chosen to train the prediction system. These algorithms are Logistic Regression, Quadratic Discriminant Analysis, and SVM. These algorithms were applied to next day model which predicted the outcome of the stock price on the next day and long term model, which predicted the outcome of the stock price for the next n days. The next day prediction model produced accuracy results ranging from 44.52% to 58.2%. Dai and Zhang (2013) have justified their results by stating that US stock market is semi-strong efficient, meaning that neither fundamental nor technical analysis can be used to achieve superior gain. However, the long-term prediction model produced better results which peaked when the time window was 44. SVM reported the highest accuracy of 79.3%. In Xinjie (2014), the authors have used 3 stocks (AAPL, MSFT, AMZN) that have time span available from 2010-01-04 to 2014-12-10. Various technical indicators such as RSI, On balance Volume, Williams %R etc are used as features. Out of 84 features, an extremely randomized tree algorithm was implemented as described in Geurts and Louppe (2011), for the selection of the most relevant features. These features were then fed to an rbf Kernelized SVM for training. Devi, Bhaskaran and Kumar (2015) has proposed a model which uses hybrid cuckoo search with support vector machine.The literature survey helps us conclude that Ensemble learning algorithms have remained unexploited in the problem of stock market prediction. We will be using an ensemble learning method known as Random Forest to build our predictive model. Random forest is a multitude of decision 2 May 3, 2018 Applied Mathematical Finance main trees whose output is the mode of the outputs from the individual trees

III. METHODOLOGY

Our methodology is composed of five steps. In the first step, we divide our raw data into study periods, where each study period is divided into a training part (for in-sample-trading), and a trading part (for out-of-sample predictions). In the second step, we introduce our features, whereas in the third step we set up our targets. In the forth step, we define the setup of our two machine learning methods we employ, namely random forest and CuDNNLSTM.
Finally, in the fifth step, we establish a trading strategy for the trading part.

Dataset creation with non-overlapping testing period.

We follow the procedure of Krauss and divide the dataset consisting of 29 years starting from January 1990 till December 2018, using a 4-year window and 1-year stride, where each study period is divided into a training period of approximately 756 days (≈ 3 years) and a trading period of approximately 252 days (≈ 1 year). As a consequence, we obtain 26 study periods with non-overlapping trading part.

Features selection

Technical Indicators are important parameters that are calculated from time series stock data that aim to forecast financial market direction. They are tools which are widely used by investors to check for bearish or bullish signals. The technical indicators which we have used are listed below:

Relative Strength Index
The formula for calculating RSI is:

RSI = 100 − (100/ (1 + RS))
RS = Average Gain Over past 14 days Average Loss Over past 14 days

RSI is a popular momentum indicator which determines whether the stock is overbought or oversold. A stock is said to be overbought when the demand unjustifiably pushes the price upwards. This condition is generally interpreted as a sign that the stock is overvalued and the price is likely to go down. A stock is said to be oversold when the price goes down sharply to a level below its true value. This is a result caused due to panic selling. RSI ranges from 0 to 100 and generally, when RSI is above 70, it may indicate that the stock is overbought and when RSI is below 30, it may indicate the stock is oversold.

Stochastic Oscillator
The formula for calculating Stochastic Oscillator is:

$$\%K = 100 * ((C - L14) / (H14 - L14))$$
where, C = Current Closing Price
L14 = Lowest Low over the past 14 days
H14 = Highest High over the past 14 days

International Journal of Advances in Engineering and Management (IJAEM)
Volume 2, Issue 10, pp: 155-158 www.ijaem.net ISSN: 2395-5252

Stochastic Oscillator follows the speed or the momentum of the price. As a rule, momentum changes before the price changes. It measures the level of the closing price relative to low-high range over a period of time.

Moving Average Convergence Divergence
MACD = EMA12(C) − EMA26(C)
SignalLine = EMA9(MACD) (9)

where,
MACD = Moving Average Convergence Divergence
C = Closing Price series
EMAn = n day Exponential Moving Average
EMA stands for Exponential Moving Average.
When the MACD goes below the SingalLine, it indicates a sell signal. When it goes above the SignalLine, it indicates a buy signal.

Target selection
Model training specification
Model specification for Random forest
• Number of decision trees in the forest = 1000
• Maximum depth of each tree4 = 10
• For every split, we select m := ⌊ √p⌋ features randomly
Prediction and trading methodology

IV. RESULTS

The empirical results5 show that our multi-feature setting consisting not only of the returns with respect to the closing prices, but also with respect to the opening prices and intraday returns, outperforms the single feature setting of Krauss et al. (2017) and Fischer & Krauss (2018), both with respect to random forests and LSTM. We refer to ”IntraDay” for our setting and ”NextDay” for the setting in Krauss et al. (2017) and Fischer & Krauss (2018) in Tables 1–3 and Figures 1–3. Indeed, our setting involving LSTM obtains, prior to transaction costs, a daily return of 0.64%, compared to the setting in Fischer & Krauss (2018) obtaining a 0.41% daily return. Also for random forests, our setting obtains a higher daily return of 0.54%, compared to 0.39% when using the setting as in Krauss et al. (2017). The share of positive returns is at 69.67% and 65.85% for LSTM and random forests. In addition, our setting obtains a higher sharpe ratio and lower standard deviation (i.e. typical annualized risk-return metrics) in comparison with the one in Krauss et al.

To show the importance of using three features instead of having a single feature, we additionally analyze in Tables 2 & 3 the performance in the case of intraday-trading, but where only intraday returns i (s) t,m as a single feature is used. The experimental results show massive improvement in all metrics when using the three features.

In Figures 1–3, we have divided the time period from January 1993 until December 2018 into three time-periods, analog to Fischer & Krauss (2018) and similar to Krauss et al. (2017). Roughly speaking, the first time-period corresponds to a strong performance caused by, among others, the dot-com-bubble, followed by the time-period of moderation with the bursting of the dot-com bubble and the financial crisis of 2008, ending with the time-period of deterioration; probably since by that time on, machine learning algorithms are broadly available and hence diminishes the opportunity of creating statistical arbitrage having a technological advantage. We refer to Krauss et al. (2017) and Fischer & Krauss (2018) for a detailed discussion of these sub-periods. We see in Figures 1–3 that in each of these sub-periods, our setting outperforms the one in Krauss et al. (2017) and Fischer & Kraus.

REFERENCES

[1]. Avellaneda, M., & Lee, J.-H. (2010). Statistical arbitrage in the US equities market. Quantitative Finance, 10, 761–782. Borovykh, A., Bohte, S., & Oosterlee, C. W. (2018). Dilated convolutional neural networks for time series forecasting. Journal of Computational Finance, Forthcoming.
[2]. Braun, S. (2018). LSTM benchmarks for deep learning frameworks. preprint, arXiv:1806.01818. Breiman, L. (2001). Random forests. Machine learning, 45, 5–32.
[3]. Chetlur, S., Woolley, C., Vandermersch, P., Cohen, J., Tran, J., Catanzaro, B., & Shelhamer, E. (2014). cuDNN: Efficient primitives for deep learning. preprint, arXiv:1410.0759.
[4]. Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270, 654–669.
[5]. Ho, T. K. (1995). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (pp. 278–282). IEEE volume 1.
[6]. Huck, N. (2009). Pairs selection and outranking: An application to the S&P 100 index. European Journal of Operational Research, 196, 819–825. Huck, N. (2010). Pairs trading and outranking: The multi-

step-ahead forecasting case. European Journal of Operational Research, 207, 1702–1716.

[7]. Krauss, C., Do, X. A., & Huck, N. (2017). Deep neural networks, gradient-boosted trees, random forests: Statistical arbitrage on the S&P 500. European Journal of Operational Research, 259, 689–702.

[8]. Moritz, B., & Zimmermann, T. (2014). Deep conditional portfolio sorts: The relation between past and future stock returns. In LMU Munich and Harvard University Working paper .

[9]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: machine learning in python. Journal of machine learning research, 12, 2825–2830.

[10]. Schmidhuber, J., & Hochreiter, S. (1997). Long short-term memory. Neural Comput, 9, 1735–1780.

[11]. Sezer, O. B., & Ozbayoglu, A. M. (2018). Algorithmic financial trading with deep convolutional neural networks: Time series to image conversion approach. Applied Soft Computing, 70, 525–538.

[12]. Siami-Namini, S., & Namin, A. S. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. preprint, arXiv:1803.06386.

[13]. Takeuchi, L., & Lee, Y.-Y. A. (2013). Applying deep learning to enhance momentum trading strategies in stocks. In Technical Report. Stanford University.

[14]. Tran, D. T., Iosifidis, A., Kanniainen, J., & Gabbouj, M. (2018). Temporal attention-augmented bilinear network for financial time-series data analysis. IEEE transactions on neural networks and learning systems, 30, 1407–1418.

[15]. Xue, J., Zhou, S., Liu, Q., Liu, X., & Yin, J. (2018). Financial time series prediction using ℓ2,1RF-ELM. Neurocomputing, 277, 176–186. 8

[16]. Widom, J. (1995). Research problems in data warehousing. In Proceedings of the fourth international conference on information and knowledge management, CIKM '95 (pp. 25- 30). New York, NY, USA: ACM. 10.1145/221270.221319.

[17]. R. Gencay, "Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules," Journal of International Economics, vol. 47,no.!, pp. 91-107,1999.

[18]. A. Timmermann and C. W Granger, "Efficient market hypothesis and forecasting," International Journal of Forecasting, vol. 20,no.!, pp. 15- 27,2004.

[19]. D. Bao and Z. Yang, "Intelligent stock trading system by turning point confirming and probabilistic reasoning," Expert Systems with Applications, vol.34,no. 1,pp. 620-627,2008.

[20]. Haoming Li, Zhijun Yang and Tianlun Li (2014). Algorithmic Trading Strategy Based On Massive Data Mining. Stanford University.

[21]. Yuqing Dai, Yuning Zhang (2013). Machine Learning in Stock Price Trend Forecasting. Stanford University.

[22]. Xinjie (2014). Stock Trend Prediction With Technical Indicators using SVM. Stanford University.

[23]. Pierre Geurts, Gilles Louppe . Learning to rank with extremely randomized tree. JMLR: Workshop and Conference Proceedings 14 (2011) 4961

[24]. Felipe Giacomel, Renata Galante, Adriano Pareira. An Algorithmic Trading Agent based on a Neural Network Ensemble: a Case of Study in North American and Brazilian Stock Markets. 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology