Time Series Analysis on Financial Data from the NYSE

Sanchit Singhal, Gaurav Lalwani

Introduction

The modeling of stocks traded on the NYSE into binary classes of gainers, shares that the model expects to appreciate, and losers, shares that the model predicts will lose value, over a time period is a challenge that can be extremely rewarding. The motivation behind this work is to help investors identify financial opportunities and, given a certain risk threshold (probability of price movement), be able to invest in the stock market with more confidence based both on fundamentals of a company as well as the share price history itself. Other desired benefits, apart from predicting price fluctuations for individual securities include the fact that the model should be able to aggregate data for all stocks, given it calculates the probability for each stock anyway, and provide insight into what the overall market should look like across various indexes and sectors (i.e. SP 500, the tech sector, etc), thus further aiding investors to park their money more intelligently through index/mutual funds if a more diversified portfolio is desired.

There is a plethora of existing solutions for the purpose of predicting share prices. Since at least the 1950s, companies have been building models to predict the future value of the stock market. Historically, these analysis could be categorized into fundamental analysis, which looks at the underlying company and its performance, and technical analysis, which focuses on patterns and trends of the past prices. The existing methodology mostly refrains from combining the two approaches. Further, the literature we found on the topic seems to solely focus on the prediction of share prices in the short term - intraday trading or at most daily charts. The team believes there exists a gap in the literature of modelling a longer time frame.

The project aims to build models that help classify stocks into two classes - gainers and losers. Although this itself is nothing new, the team will aim to predict these fluctuations every year, rather than daily; thereby enabling a new approach to solve an existing challenge. The team also plans to augment historical price trends with fundamentals of the company to achieve a more holistic view of the security as opposed to selecting a certain type of analysis. Given the project parameters, the team believes this model will be useful for investors who want to invest for the longer term, rather than benefit from day-trading. Most current technologies use automation software to buy/sell at a faster pace, but in contrast, our project will aim to identify investment opportunities that will appreciate in the longer term.

Related Works

Regression Techniques in Financial Models

- Machine learning in prediction of stock market indicators based on historical data and data from Twitter sentiment analysis. (Porshnev, Redkin, and Shevchenko 2013)
- Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques (Patel et al. 2015)
- Predicting stock market index using fusion of machine learning techniques (Patel et al. 2015)
- Evolving Least Squares Support Vector Machines for Stock Market Trend Mining (Yu et al. 2009)

All the above mentioned literary articles had SVM as the common regression technique to predict stock prices. The articles didnt focus or display information regarding other methods such as Decision tree and KNN - even though it was just vaguely mentioned in the second reference. We propose to use all the regression techniques and display accuracy metrics to select the best possible technique. None of the articles worked on enhancing techniques such as dimension reduction or subset applications to reduce the variance and thus keeping minimal bias. We propose to work on dimension reduction techniques. Also the articles focused more on prediction of stock market prices and none of the papers classified the data into groups to predict the possibility of the market opening up or down for a given year. We propose to work on classification of data so as to predict the open price to be low or high from a given threshold.

Ensemble Learning in Financial Modeling

- Chinese Stock Index Futures Price Fluctuation Analysis and Prediction Based on Complementary Ensemble Empirical Mode Decomposition (Chen and Pan 2016)
- Comparison of individual, ensemble and integrated ensemble machine learning methods to predict Chinas SME credit risk in supply chain finance (Zhu et al. 2017)
- Stock market prediction with multiple classifiers (Qian and Rasheed 2007)

Ensemble learning is popular technique to gain prediction accuracy. These works discuss model architecture that have been applied and some strengths and weaknesses that were found. These analysis have used a combination of artificial neural networks, decision trees, and k-nearest neighbors and through appropriate collaboration achieved 65 percent correct results. From the first reference, it is also important to note that we must consider the type of market when selecting the modelling tools - markets in an emergent state, such as China, will use different analysis than mature ones such as the NYSE. That being said, some learnings from these works can be transferred over to our project: random subspace boosting seemed to provide a substantial increase in accuracy. Our work is different from these works because we plan on building various models, creating ensembles from them over various combinations, and then optimizing our model to select optimal parameters. None of these papers discuss building the ensembles and the effect of different inputs on the results - our analysis aims to cover these subtopics as well.

Deep Learning Techniques for Financial Models

- A deep learning framework for financial time series using stacked autoencoders and long-short term memory (Bao, Yue, and Rao 2017)
- Stock prediction using deep learning (Singh and Srivastava 2017)
- A new hybrid constructive neural network method for impacting and its application on tungsten price prediction (Muzhou et al. 2017)
- A review of unsupervised feature learning and deep learning for time-series modeling (Längkvist, Karlsson, and Loutfi 2014)

Apart from the techniques mentioned in the above articles we might also be using back propagation methods to view the result outcome. Also apart from the mentioned convolution methods we aim to change the pooling layer size to observe the accuracy for different sizes so as to select the best possible method. Although the team is unsure whether we will utilize deep learning in the models, it was important to note some related work that has already been done in the space.

Time Series Analysis for Financial Data

- An Effective Time Series Analysis for Stock Trend Prediction Using ARIMA Model for Nifty Midcap-50 (Devi, Sundar, and Alli 2013)
- Financial time series forecasting using support vector machines (Kim 2003)
- Twitter mood predicts the stock market (Bollen, Mao, and Zeng 2011)

Most previous works focus on using artificial neural networks (ANNS) for the challenging task of financial time-series predictions. However, the third reference points out that these models can have limitations in that learning patterns is difficult when there is a lot of noise and complex dimensionality. Our work is different from these works as we plan to apply SVM to find the maximum margin hyperplane. We might also consider building an ANN and comparing to the SVM - just to see how similar they are. We are also considering building an ensemble which may or may not include both of these models.

References

Bao, W.; Yue, J.; and Rao, Y. 2017. A deep learning framework for financial time series using stacked autoencoders and long-short term memory. *PloS one* 12(7):e0180944.

Bollen, J.; Mao, H.; and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of computational science* 2(1):1–8.

Chen, R., and Pan, B. 2016. Chinese stock index futures price fluctuation analysis and prediction based on complementary ensemble empirical mode decomposition. *Mathematical Problems in Engineering* 2016.

Devi, B. U.; Sundar, D.; and Alli, P. 2013. An effective time series analysis for stock trend prediction using arima model for nifty midcap-50. *International Journal of Data Mining & Knowledge Management Process* 3(1):65.

Kim, K.-j. 2003. Financial time series forecasting using support vector machines. *Neurocomputing* 55(1-2):307–319.

Längkvist, M.; Karlsson, L.; and Loutfi, A. 2014. A review of unsupervised feature learning and deep learning for time-series modeling. *Pattern Recognition Letters* 42:11–24.

Muzhou, H.; Taohua, L.; Yunlei, Y.; Hao, Z.; Hongjuan, L.; Xiugui, Y.; and Xinge, L. 2017. A new hybrid constructive neural network method for impacting and its application on tungsten price prediction. *Applied Intelligence* 47(1):28–43.

Patel, J.; Shah, S.; Thakkar, P.; and Kotecha, K. 2015. Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications* 42(1):259–268.

Porshnev, A.; Redkin, I.; and Shevchenko, A. 2013. Machine learning in prediction of stock market indicators based on historical data and data from twitter sentiment analysis. In *Data Mining Workshops (ICDMW), 2013 IEEE 13th International Conference on*, 440–444. IEEE.

Qian, B., and Rasheed, K. 2007. Stock market prediction with multiple classifiers. *Applied Intelligence* 26(1):25–33.

Singh, R., and Srivastava, S. 2017. Stock prediction using deep learning. *Multimedia Tools and Applications* 76(18):18569–18584.

Yu, L.; Chen, H.; Wang, S.; and Lai, K. K. 2009. Evolving least squares support vector machines for stock market trend mining. *IEEE Transactions on evolutionary computation* 13(1):87–102.

Zhu, Y.; Xie, C.; Wang, G.-J.; and Yan, X.-G. 2017. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict chinas sme credit risk in supply chain finance. *Neural Computing and Applications* 28(1):41–50.