Predicting the Direction of Stock Market Prices Using Tree-Based Classifiers

Suryoday Basak[a], Saibal Kar[b], Snehanshu Saha[a], Luckyson Khaidem[a], Sudeepa Roy Dey[a]

[a] *Authors are affiliated with the Department of Computer Science and Engineering, and Center for Applied Mathematical Modeling and Simulation (CAMMS), PESIT South Campus, Bangalore, India*
[b] *Author is affiliated with the Centre for Studies in Social Sciences, Calcutta, India and IZA, Bonn*

Abstract

Predicting returns in the stock market is usually posed as a forecasting problem where prices are predicted. Intrinsic volatility in the stock market across the globe makes the task of prediction challenging. Consequently, forecasting and diffusion modeling undermines a diverse range of problems encountered in predicting trends in the stock market. Minimizing forecasting error would minimize investment risk. In the current work, we pose the problem as a direction-predicting exercise signifying gains and losses. We develop an experimental framework for the classification problem which predicts whether stock prices will increase or decrease with respect to the price prevailing n days earlier. Two algorithms, random forests, and gradient boosted decision trees (using XGBoost) facilitate this connection by using ensembles of decision trees. We test our approach and report the accuracies for a variety of companies as improvement over existing predictions. A novelty of the current work is about the selection of technical indicators and their use as features, with high accuracy for medium to long-run prediction of stock price direction.

Keywords: stock price movement, xgboost, random forests, machine classification

1. Introduction and Motivation

For a long time, it was believed that changes in the prices of stocks is not forecastable. The well known Random Walk hypothesis (Malkiel & Fama, 1970; Malkiel, 2003), and the Efficient Market Hypothesis (Jensen, 1978), which state that a market is efficient with respect to a current information set $I(t)$ if it is impossible to make economic gains in this market, led to this belief. In other words, if it is impossible to outperform the market owing to the randomness in stock prices, then unless a different (often excessive) type of risk is considered, economic profits cannot rise. It is unclear, however, how such risk would be measured (see the work by Timmermann & Granger (2004), where stock prices are treated as a martingale). Therefore, it should be of little doubt that predicting the trends in stock market prices is a challenging task and that there is high returns from improving value at risk forecasts (Halblieb & Pohlmeier, 2012). On the contrary, the Wisdom of Crowd hypothesis, which emerges from the theory of collaborative filtering, states that many individuals, each with limited information, can provide very accurate assessments if their information is elicited in an appropriate fashion. It is, however, not known to be useful for predicting stock market returns; nonetheless, some individual, as well as institutional investors are able to beat the market to make profits (Avery, Chevalier & Zeckhauser, 2016). The inefficiency of prediction gets accentuated due to various uncertainties involved and owing to the presence of multiple variables all of which can potentially influence the market value on a particular day. Over time, a number of explanatory variables have been added to this enormous literature (see a history of EMH in (Sewell, 2011; Beechey, Gruen & Vickery, 2000) etc.): these include country-specific economic conditions, investors' sentiments towards a particular company, political events, etc. Consequently, stock markets are susceptible to quick changes, which often turn into random fluctuations in the stock prices (for calibration of agent-based dynamics, refer to the work by Recchioni et al. (2015)). Notwithstanding, EMH has certain fault lines that should be mentioned before we develop two algorithms to observe the quality of predictability for stock prices.

Indeed, (Malkiel, 2003) offers a masterly discussion of the critique of the EMH, and suggests that, way back in 1973, he in general advised investors to purchase broad-based index funds that bought and held all the stocks in the market and that charged very low expenses. He admits that by the start of the twenty-first century, the intellectual dominance of the efficient market hypothesis had become far less universal. Many