From Forecasting to Classification: Predicting the Direction of Stock Market Price 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 Institute of Labor Economics, Bonn*

Abstract

Predicting trends in stock market prices has been an area of interest for researchers for many years due to its complex and dynamic nature. Intrinsic volatility in the stock market across the globe makes the task of prediction challenging. Forecasting and diffusion modeling, though effective, cannot be the panacea for the diverse range of problems encountered in predicting trends in the stock market, short-term or otherwise. Market risk, strongly correlated with forecasting errors, needs to be minimized to ensure minimal risk in investment. This paper is the outcome of experiments with a completely different approach: instead of defining this problem as a traditional forecasting-style problem, we just try to predict whether prices will increase or decrease. The problem is posed as a classification problem, where the class labels may be ± 1, indicating an increase or a decrease in the price of a stock with respect to n days back. For this purpose, the potential of Random Forests and XGBoosted trees is explored. Random Forests use an ensemble of Decision Trees to improve the accuracy of classification. XGBoost is an engineering solution which aims to speed up the process of growing Gradient Boosted Decision Trees (GBDT). Technical indicators such as Relative Strength Index (RSI), Stochastic Oscillator, etc. are used as features to train the model. The algorithms are shown to outperform the algorithms used in the existing literature.

Keywords: stock price prediction, xgboost, random forests, classification, binary options

1. Introduction and Motivation

For a long time, it was believed that changes in the price of stocks is not forecastable. The well known Random Walk hypothesis (Malkiel & Fama, 1970; Malkiel & Burton, 2003), and the Efficient Market hypothesis (Jensen, 1978), which states 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 a risk would be measured (see (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. On the contrary, the Wisdom of Crown 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 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 may seem to be random fluctuations in the stock prices.

Stock market series are generally dynamic, non-parametric, chaotic and noisy in nature making investments intrinsically risky. Stock market price movement is considered to be a random process with fluctuations, that are more prominent in the short-run. However, some stocks tend to develop linear trends in the long-run. It is needless to mention that advanced knowledge of stock price movement in the future should help to minimize this risk. Traders are more likely to buy a stock whose value is expected to increase in the future and conversely for falling prices. So there is a need for accurately predicting the trends in stock market prices in order to maximize capital gain and minimize loss. Regardless, it had best be admitted that adding value to this complex and deeply researched topic is not easy. To this end, this paper presents the use of techniques of machine learning (ML) to predict stock prices at the level of a firm to get better insights into the accuracy

Contact Information of Authors:
Suryoday Basak: suryodaybasak@pes.edu; Snehanshu Saha: snehanshusaha@pes.edu
Luckyson Khaidem: khaidem90@gmail.com; Saibal Kar: saibal@cssscal.org; Sudeepa Roy Dey: sudeepar@pes.edu