


# Data-Driven Trend Forecasting in Stock Market Using Machine Learning Techniques

Puneet Misra, University of Lucknow, Lucknow, India

 <https://orcid.org/0000-0003-2297-9072>

Siddharth Chaurasia, University of Lucknow, Lucknow, India

## ABSTRACT

Stock market movements are affected by numerous factors making it one of the most challenging problems for forecasting. This article attempts to predict the direction of movement of stock and stock indices. The study uses three classifiers - Artificial Neural Network, Random Forest and Support Vector Machine with four different representation of inputs. First representation uses raw data (open, high, low, close and volume), The second uses ten features in the form of technical indicators generated by use of technical analysis. The third and fourth portrayal presents two different ways of converting the indicator data into discrete trend data. Experimental results suggest that for raw data support vector machine provides the best results. For other representations, there is no clear winner regarding models applied, but portrayal of data by the proposed approach gave best overall results for all the models and financial series. Consistency of the results highlight the importance of feature generation and right representation of dataset to machine learning techniques.

## KEYWORDS

Artificial Neural Network, Classification, Feature Representation, Machine Learning, Prediction, Random Forest, Stock Market, Support Vector Machine

## INTRODUCTION

Financial domain presents one of the most complex fields which can be influenced by numerous factors making it susceptible to unexpected changes. Financial time series, for example, daily prices of security, index or currency, presents an example of high dimensional, non-stationary and noisy data. Due to its practical importance, the analysis of financial market movements has been widely studied in the fields of finance, engineering and mathematics in the last decades (Yoo, Kim, & Jan, 2007).

Adding on to challenges inherent in data, economists have also highlighted the unpredictability of financial markets. Efficient Market Hypothesis (Fama, 1965) and Random Walk Theory (Godfrey, Granger, & Morgenstern, 1964) both signify that the market movements are random and unpredictable thus ridiculing utility of technical analysis. (Lendasse et al. 2001) references to statement by Campbell which says, “Recent econometric advances and empirical evidence seem to suggest that financial asset returns are predictable to some degree.” hence, stressing on the utility of technological advancements for prediction of market.

As successful prediction in this field can result into financial gains by guiding investment decisions, the market prediction has been the forerunner in the adoption of technology. Consequently, forecasting which was considered as forte of statisticians in the last decade has seen substantial

DOI: 10.4018/JITR.2020010109

Copyright © 2020, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.

implementation of AI-based techniques like machine learning, evolutionary computation and fuzzy logic.

Complex, chaotic and noisy nature of financial data presents the challenges that require non-parametric methods which do not use a statistical assumption about its nature. Machine learning approach achieves this as no apriori knowledge about data is required (Misra & Siddharth, 2017). A lot of efforts has been made to predict the price or movement direction of the security. Still, accurate forecast of the stock price, even its movements, is not easy to achieve.

For short-term market predictions, technical analysis is widely employed as it provides a framework for taking informed investment decisions by applying a supply and demand methodology to market prices. Fundamental principles of the study of technical analysis are governed by the changes in the supply and demand of traded securities affect their current market prices (Scott, Carr, & Cremonie, 2016). Raw time series data in the form of Open, High, Low, Close and Volume (OHLCV) is utilized to compute technical indicators (TI). Mathematical nature of technical analysis enables its smooth blending with data-driven techniques which bring in insights that may not be obvious in raw data.

Attribute or feature generation and selection represent a critical aspect in the construction of the ML models. A good set of attributes derived from the financial time series will ease the process of classification (Gerlein, McGinnity, Belatreche, & Coleman, 2016). Another vital aspect after selection of attributes is to present the selected input in the format that can provide the inherent information in an interpretable form such that maximum information gain can be accomplished. This paper provides empirical evidence on the second aspect which is depiction of features in the right format as per the prediction goal.

Predictors generated by the application of technical analysis can be considered as proxies for the true but unobservable latent factors. Hence for this study, we hypothesize that the use of TIs should assist in improved prediction. The improvement can be in the form of accuracies or stability of the predicted outcomes. Moreover, trend information generated from TIs should be more appropriate for trend prediction. Hence, after selection of features, the format in which the selected inputs are offered to machine learning system may also play a crucial role in the effectiveness of the outcomes.

According to a recent study (Gerlein et al., 2016), the average accuracy of machine learning models for the task of trend prediction is around 48% to 54%. Though in this study the models used were simpler ones like C4.5, K\*, Naïve Bayes (NB), JRip, OneR. Other studies too have reported accuracies in the range of 50s and 60s like (Kim, 2003), (Qian & Rasheed, 2007), (Lin, Guo, & Hu, 2013). Even after decades of research, inherent challenges in financial forecasting keeps research community alive to learn the intricacies of market forecasting and improve upon the prediction accuracy.

Literature shows that forecasting the direction is enough to execute profitable trading strategies (Ballings, Van Den Poel, Hespeels, & Gryp, 2015). Hence, the emphasis of this study is to predict the next day direction of movement instead of absolute price. Different techniques have been utilized in directional predictions. Above paper lists the researches focusing on directional forecast which indicates ANN, SVM to be preferred methods. Recently, ensembles like RF have also been tried with reasonable accuracy. Thus, this paper uses said models for next day directional forecast for three financial securities.

This study contributes to existing literature in three ways: First, it proposes an approach for directional forecast which is independent of interpretations from the finance domain. The proposed approach still provides results that are at par with an approach which utilizes financial knowledge to interpret TI values generated by mathematical formulations. Second, it extensively compares four different approaches to data presentation on three data series for next day directional forecasting. Third, extensive experimentation is carried out to evaluate the performance of machine learning models in conjunction with technical analysis for each representation of data thus highlighting the utility of ML techniques in this field of forecasting.

Rest of paper is divided into four sections. Section 2 describes the experimental setup. It provides details of the generation of four datasets used in this paper along with testing strategy of

18 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/article/data-driven-trend-forecasting-in-stock-market-using-machine-learning-techniques/240726](http://www.igi-global.com/article/data-driven-trend-forecasting-in-stock-market-using-machine-learning-techniques/240726)

## Related Content

---

### Borderline Fields of Information Architecture: Information Overload, the Literacies, and Personal Information Management

Tibor Koltay (2016). *Handbook of Research on Information Architecture and Management in Modern Organizations* (pp. 111-129).

[www.irma-international.org/chapter/borderline-fields-of-information-architecture/135764](http://www.irma-international.org/chapter/borderline-fields-of-information-architecture/135764)

### Enterprise Information Portals: Efficacy in the Information Intensive Small to Medium Sized Business

Wita Wojtkowski and Marshall Major (2004). *Annals of Cases on Information Technology: Volume 6* (pp. 90-103).

[www.irma-international.org/article/enterprise-information-portals/44571](http://www.irma-international.org/article/enterprise-information-portals/44571)

### The Value of Government Mandated Location-Based Services in Emergencies in Australia

Anas Aloudat, Katina Michael, Roba Abbas and Mutaz Al-Debei (2011). *Journal of Information Technology Research* (pp. 41-68).

[www.irma-international.org/article/value-government-mandated-location-based/68961](http://www.irma-international.org/article/value-government-mandated-location-based/68961)

### The E-Pabelan National ICT4PR Pilot Project: Experiences and Challenges of Implementation in an Indonesian Context

Alex Robinson (2008). *Information Communication Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 2052-2068).

[www.irma-international.org/chapter/pabelan-national-ict4pr-pilot-project/22799](http://www.irma-international.org/chapter/pabelan-national-ict4pr-pilot-project/22799)

### The Diffusion Of Customer-Related IT Among Large And Mid-Sized Companies

Shana L. Dardan, Ram L. Kumar and Antonis C. Stylianou (2007). *Information Resources Management Journal* (pp. 12-24).

[www.irma-international.org/article/diffusion-customer-related-among-large/1324](http://www.irma-international.org/article/diffusion-customer-related-among-large/1324)