ABSTRACT

There is a large body of research carried out suggesting the predictability of Stock markets. Initially, tests of predictability of stock market returns were motivated by market efficiency, where it is assumed that predictability was inconsistent with constant stock market returns, efficient markets paradigm (Rey, David 2004). For long it was thought that stock markets are not predictable, at least in an economically significant manner (Fama, 1970). Lo and Maculay (1988) in their research paper claim that stock prices do not follow random walks and suggested considerable evidence towards predictability of stock prices. Basu (1977), Fama & French (1992), Lakonishok, Schleifer & Vishney (1997) in their various studies have carried out many cross sectional analysis across the globe and tried to establish the predictability of the stock prices. Ferson & Harvey (1991) showed that predictability in stock returns are not necessarily due to market inefficiency or over-reaction from irrational investors but rather due to predictability in some aggregate variables that are part of the information set.

As a noun “predictability,” refers to the ability to foresee and talk about something that is going to happen. In this context, it refers to the ability to predict the movement of markets from its past behavior. People have been trying to predict stock market for ages as it captivated the attention of
academicians and investors alike. As stock market is a real-time, complex, dynamic, and vibrant system, it poses a challenge to the academia to unravel its dynamics and functions. On the other hand, investors are lured by the prospects of making fortune with the ability to correctly predict the price and the trend of the stock market. The knowledge of market dynamics is of immense value to individual investors, institutional investors, financial institutions, and others whose fortunes are largely dependent on the movement of the stock market. It is also useful for regulating bodies to formulate suitable policies for efficient functioning of the system. No wonder stock market research, unlike many other research areas, always remains as a beehive of activities. Quintessentially, stock market analysts focus on developing models to successfully forecast index values/stock prices, aiming at high profits using well-defined trading strategies. Statistics has been applied for predicting the behavior of the stock market for more than half a century and a hit rate of 54% was considered to be the satisfying result for stock prediction. Mathematical, stochastic, and econometric models have been used to model the share price behavior. However, high volatility in stock prices makes it difficult to predict the stock market movements and it was thought that it would be impossible to predict stock market as it followed random walk. The evolution of computing power, database technology, and new data mining tools threw open hitherto unavailable avenues and tools to the stock market researchers.
Statistical tools have been applied for predicting the behavior of the stock market for more than half a century and a hit ratio of 54% was considered as a satisfying result for stock prediction. Later, analysts started employing more sophisticated mathematical, stochastic and econometric tools to model the share price behavior. But due to high volatility in the prices, the success of prediction (hit rate) remained elusive. The advent of database, data warehouse technologies and machine learning algorithms threw open hitherto unavailable possibilities and tools for stock market research. The current research aims at investigating the predictability of the stock market indices using various Data mining tools. Popular machine learning algorithms such as Artificial Neural Network (ANN), k-Nearest Neighbor algorithm (k-NN), Support Vector machines (SVM), Decision tree and its ensemble variant called Random Forest are employed in this study to investigate predictability of global and Indian stock market indices.

This study attempts to comprehend the phenomenon of predictability among the global markets using data mining tools. Prediction is two dimensional, while one is predicting future values that are numeric and continuous in nature, the latter aims at predicting future trends that are categorical or discrete in nature. The former is often referred to as forecasting, while the latter is known as classification.

The daily stock prices of select global indices and indices of BRIC nations for the period from 2006 to 2013 are considered for this study. The intraday movement of the index along with select technical indicators forms
the predictor set for the predictive models. With these input variables, the four data mining models are put to use to predict the next day’s close price of the index. The predictability of the indices are measured in terms of the hit ratio and forecasting errors obtained from the predictive models. It is found that Emerging markets are less predictable than the developed markets. The study also confirms the emergence of SVM as the most accurate and reliable forecasting and classifying model. Term wise analysis of the predictability on Indian indices shows that markets are better predictable in short term than in long terms. The impressive growth in hit ratios of the classifiers and drastic reduction in forecasting errors as one move from long term to short term prediction is observed in the results. The study also highlights the effect of various global cues on Indian indices. These global cues when used as a predictor set for predictive models, in lieu of the set of technical indicators, show better predictive performance of the models. Application of association rule mining help to identify patterns among the global cues that would result in bullish and bearish trend in Indian indices. These association rules would be of immense help in predicting the trend of Indian markets. This study identifies certain common association rules that are specific to Indian stock indices on the basis of frequently occurring patterns among global cues. A notable finding from these rules is that it is more difficult to predict the bear trend than the bull trend. Thus this study throws more light on the market dynamics and behavior of Indian indices.