Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 89 (2016) 441 – 449

Procedia Computer Science

Twelfth International Multi-Conference on Information Processing-2016 (IMCIP-2016)

Prediction Models for Indian Stock Market

Aparna Nayak, M. M. Manohara Pai* and Radhika M. Pai

Manipal Institute of Technology, Manipal University, Manipal 576 104, India

Abstract

Stock market price data is generated in huge volume and it changes every second. Stock market is a complex and challenging system where people will either gain money or lose their entire life savings. In this work, an attempt is made for prediction of stock market trend. Two models are built one for daily prediction and the other one is for monthly prediction. Supervised machine learning algorithms are used to build the models. As part of the daily prediction model, historical prices are combined with sentiments. Up to 70% of accuracy is observed using supervised machine learning algorithms on daily prediction model. Monthly prediction model tries to evaluate whether there is any similarity between any two months trend. Evaluation proves that trend of one month is least correlated with the trend of another month.

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016

Keywords: Boosted Decision Tree; Logistic Regression; Sentiment Analysis; Stock market; Support Vector Machine.

1. Introduction

Stock price prediction is very important as it is used by most of the business people as well as common people. People will either gain money or lose their entire life savings in stock market activity. It is a chaos system. Building accurate model is difficult as variation in price depends on multiple factors such as news, social media data, fundamentals, production of the company, government bonds, historical price and country's economics[1]. Prediction model which considers only one factor might not be accurate. Hence incorporating multiple factors news, social media data and historical price might increase the accuracy of the model.

There are two common methods to predict the stock market prices[2]. One among that is chartist or technical theories and the second one is fundamental or intrinsic value analysis. Proposed method is built on the principle of technical theories. Basic assumption of this theory is history tends to repeat itself. Prediction model can be applied on the historical data to get future trend. As researchers have discussed in S. J. Grossmara and R. J. Shiller[3] and L. Andrew and M. A. Craig[4], as and when new information comes in the market stock market value varies. Technical analysis and semi strong form of efficient market hypothesis are followed, to build prediction model in the proposed work. The goal of this research work is to build a model which predicts stock trend movement (trend will be up or down) using historical data and social media data. Two models are built as part of research work. Both models use supervised machine learning algorithm. First model is daily prediction model, considers both sentiment and historical data.

*Corresponding author. Tel.: +91-9945202361.
E-mail address: mmm.pai@manipal.edu

Peer-review under responsibility of organizing committee of the Organizing Committee of IMCIP-2016
doi:10.1016/j.procs.2016.06.096

This model predicts the future trend for the next day. Sentiment of the company has been computed by using twitter data and news of the company. Outcome of sentiment analysis is considered along with open price, close price of stock with extracted statistical parameters to build model. Second model is monthly prediction model, considers only historical data and predicts the trend for next one month. Proposed work 3 investigates whether the outcome of model is inline with the actual trend movement.

The rest of this paper is organized is as follows. Section 2 introduces some previous research work on sentiment analysis for stock market prediction and stock trend movement using historical price. Section 3 describes proposed method. Section 4 shows the dataset used and evaluates the results of the experiments. Finally, Section 5 concludes the contribution of this research work.

2. Related Work

Many research groups are exploring stock market trend prediction using social media analytics. Architecture for building the model has been referred from P. Paakkonen, D. Pakkala[5]. Many use case architecture have been discussed in the same paper. Multiple methods are there to detect the polarity of each tweet/news. Initially moods of a user on the specific company was considered to analyze the stock price as shown in X. Zhhang *et al.*[6] and J. Bollen *et al.*[7]. Now polarity of each item in news/tweet has been found to get the sentiment. To find the polarity of each news/tweet item one can use either dictionary based approach or semi supervised algorithm. In case of dictionary based algorithm, polarity to each word is assigned by comparing each word of news with dictionary word. In case of semi supervised algorithm as discussed by K. Mizumoto *et al.*[8], initial level of dictionary is built manually then new words are categorized as either positive or negative based on occurrence of new words along with words in the built dictionary. Dictionary based approach has been used in the proposed method, as semi supervised learning might not cover all possible combination of words.

X. Zhang *et al.*[6] and J. Bollen *et al.*[7] have analyzed that mood of individual affects stock market price. In[6] they have also mentioned that twitter sentiment might effect stock market trend only for few company. W. Antweilwer and M. Z. Frank[9] have discussed that information which is available about any company is not noise. One can get useful information such as prediction of future value from it. R. Ahuja *et al.*[10] have analyzed twitters on stock market by collecting 3 months BSE data. N. Lin *et al.*[11] have shown that news effects future market trend of stock market. They have considered two market places America and China. M. Hagenau *et al.*[12] have considered German Adhoc messages as input and for feature selection, Chi square method has been used. SVM algorithm has been used for which 65% of accuracy is obtained.

J. Gong and S. Son[13] have implemented stock prediction model using logistic regression considering feature index variables. They have mentioned that daily stock trading prediction with logistic regression out performs other methods such as RBF – ANN prediction model.

3. Proposed Method

The model predicts the price movement on t_n by considering all the available historical data i.e. from $t_{n-1}, t_{n-2}, \ldots t_1$, where t_n stands for transaction data of prediction. All the available data is trained by supervised machine learning algorithm. Sentiment from social media data and news are extracted. Extracted sentiments later will be integrated with historic price to build prediction model. Conflicting opinions has been reported by researchers about effect of sentiment on stock market. Few research[14] reported sentiment extracted from social media has no effect on stock price movement whereas in[7], they have reported the sentiment has either strong or weak effect on stock price movement.

Two different models have been built to predict stock market trend. First model predicts the stock market trend for the next day (Daily prediction model) by considering all available data on daily basis as input. Second model predicts the stock market trend for the next month(Monthly prediction model) by considering available data on monthly basis.

First contribution of the proposed work is that few features has been deduced from the historical data available by using statistics. One of the statistical parameter considered is relationship between trend of a day and volume of stock traded on the same day[15]. Volume traded feature in historical data will reflect both bought and sold stocks on a daily