Chapter 27
Time Series for Forecasting Stock Market Prices Based on Sentiment Analysis of Social Media

Babu Aravind Sivamani
SSN College of Engineering, Chennai, India

Dakshinamoorthy Karthikeyan
SSN College of Engineering, Chennai, India

Chamundeswari Arumugam
SSN College of Engineering, Chennai, India

Pavan Kalyan
SSN College of Engineering, Chennai, India

ABSTRACT

This paper attempts to find a relation between the public perception of a company and its stock value price. Since social media is a very powerful tool used by a lot of people to voice their opinions on the performance of a company, it is a good source of information about the public sentiment. Previous studies have shown that the overall public sentiment collected from sites like Twitter do have a relation to the market price of a company over a period of time. The goal is to build on their research to improve the accuracy of predictions and determine if the public perception surrounding a company is a driving factor of its stock growth.

DOI: 10.4018/978-1-6684-6303-1.ch027

1. INTRODUCTION

Previous studies(Mankar et al., 2018) on the effect of social media on the stock market have shown that the aggregate public mood towards a company over a short time span has a relation to the closing price of that company at the end of the time span. Studies have been able to utilize data collected from any one reputed social media site (Ex: Twitter, Stocktwits, Weibo, etc.) to produce a model that predicts stock market prices with ~70% accuracy(Acosta et al., 2017). This paper attempts to more accurately gauge the public sentiment of a company from social media websites such as Twitter by implementing time series analysis at minute intervals to find correlations that will likely produce a better stock estimate.

Stock price of a company is determined by a large number of independent traders all over the world. Previous studies have not taken into account the reasons why an individual trader makes the decision to buy or sell. As social media has been shown to offer an insight into the mindset of people, it was realized that the posts online may be an indication of how the market at large is inclined towards a company. The main objective of this paper is to find whether the public sentiment surrounding a company is able to determine the growth of its stock price. Here in this paper, the company Apple (NASDAQ: AAPL) was selected because it is prominent in the public spotlight and hence ideally suited for an analysis of this kind.

First the selected social media platform is queried for posts in the time period containing any of the keywords in our search term. The search term must be carefully selected to ensure that the number of off-topic posts is limited, while not missing out on any messages with important content. Then any irrelevant posts which passed through the search query are found and filtered out. Data pre-processing procedures such as the removal of non-English characters, stop words, hashtags and user mentions is carried out. Sentiment analysis is performed on the pre-processed text data and each post is classified as positive, negative or neutral corresponding to whether the market for Apple is bullish, bearish or not having any effect. Finally, the aggregate sentiment values from all collected websites will be fed into the model which would use a machine learning algorithm to produce a correlation between the media posts and the stock market price which can then be used to predict the closing market value, given the opening price and overall public sentiment.

The organization of this paper proceeds as follows. Section 2 discusses the literature survey, while Section 3 elaborates on the proposed methodology. Section 4 details with the result and discussion, and Section 5 details the conclusion and future work.

2. LITERATURE SURVEY

Venkata et al.(2016) used Word2vec and N-gram representation of text to train a classifier model to predict the stock market movements and picked Word2vec representation due to its high accuracy in large datasets. Rakhi et al.(2018) collected the sentiment data, and the stock price data to predict stock market price using a support-vector machine (SVM) classifier and observed that if the data size increases the accuracy obtained will also increase. Scott et al.(2017) used smart user classification to filter the tweets by computing scoring weights based on number of likes, number of followers count and how often the user is correct. Further, they used Tf-Idf vectorizer for textual representation and linear regression classifier for the sentiment prediction. Zhaoxia et al.(n.d.) used the sentiments of the news data to predict the stock market price using neural networks.