

TEXT MINING FOR MARKET PREDICTION: A SYSTEMATIC REVIEW

Arman Khadjeh Nassirtoussi , Saeed Aghabozorgi , Teh Ying Wah , David Chek Ling Ngo b

Introduction: The text mining also referred to as text data mining roughly equivalent to text analysis, is to process unstructured information, extract meaningful numeric indices from the text and thus make the information contained in the text accessible to data mining. Understanding market movements primarily facilitates one with the ability to predict future movements. Most of the time it's extremely difficult to predict the nature of markets but, being able to predict the economy is equal to being able to generate wealth by avoiding financial losses.

Goal: Fundamental data available in unstructured text are the most challenging aspect of analysis which includes social media, news, blog and forums and therefore is the main idea of this work to extend and see relevant fundamental economic concepts with cutting edge pieces of work and comparing the contrast of the current and the identified works on text mining.

Descriptions: When predicting the future prices of the stock market there are several theories available. The strong form includes historical, public and private information, such as share price. From the tenets of EMH, it is believed that the market reacts instantaneously to any given news and that it is impossible to consistently outperform the market. A consistent comparison with AMH which figures out the dilemma of market efficiently with regard to its degree and applicability to different markets is one major research going and has been analyzed with overall results suggesting a better description of the behavior of stock return than the EMH. Like textual representation there are also a variety of machine learning algorithms available.

Despite being multiple research going on, no comprehensive comparative analysis are available. The major systems that have been developed in the current market analysis are the overview of this paper. The process of having the text mining approaches for stock market predictions are followed by multiple steps including inputting dataset which involves having famous market analysis websites and new websites. In recent time's social media websites like FB, twitter are utilized for the market predictions. Other sources of input data comes from numeric value in financial markets which will be in purpose for machine learning algorithm.

Earlier research was mainly focused on stock market prediction with value companies considering the timeframe from point of news release to a market impact observation. The data is preprocessed into a machine learning algorithm. In text mining specifically the pre-processing phase is of significant impact on the overall outcomes including the feature selection. Several machine learning algorithm are mainly based on classification and regression analysis is used in order to better comprehend the reviewed systems. While sampling the training and testing volume a special kind of interest here is to know if a linear sampling have been followed as in essence the samples are on the time series. It is possible to pass technical data like average, relative strength or signals along with the text features into the classification algorithm as additional independent variables but, it is always better to encode the variables as per the algorithm needed. In general researches have used evaluation mechanisms and experimental data that widely vary and makes an objective comparison in terms of concrete levels of effectiveness unreachable. There are several literature review made in this paper regarding the present effectiveness which will be discussed later.

Future Work: I will be focusing on how other related sited papers have used the techniques and do a comparative studies on which will be the best way to predict the market analysis using text mining.