NLP and Stock Market: Role of NLP Techniques and Applications in Stock Market Analysis

Harjit Singh

Assistant Professor
Neighbourhood Campus Dehla Seehan,
Punjabi University, Punjab, India

Abstract: The financial markets are unpredictable, complex and volatile due to which it is very challenging to predict the market. Quantitative analysis on historical data is mostly used for stock market prediction using conventional methods. But these methods have their own limitations because they cannot analyse sentiment and news driven events which largely affect the stock prices. Natural language processing is a very powerful tool which can be used to analyze textual information such as financial reports, social media posts, news articles etc. to get insights into the stock market prices. This paper discusses NLP applications for predicting the stock market prices with its benefits and challenges.

Index Terms – **Natural Language Processing, Sentiment Analysis, Predicting Financial Markets. Market Transactions, Stock Market and Computers.**

I. INTRODUCTION

Many factors affect the stock market prices such as corporate earnings, market sentiments, political events, economic indicators etc. Traditionally historical data is being used to analyse the market movements based on quantitative models but it does not include qualitative aspects that impact the stock prices such as public sentiments and news events. By using Natural Language Processing techniques we can analyse huge amounts of unstructured data related to the public sentiments and market behaviour that cannot be captured by quantitative models. NLP processing of social media posts, financial reports and news articles allows us to understand their impact on stock market movements.

The objective of this paper is to discuss how NLP techniques can be used to predict stock market price movements including its effectiveness and challenges and also the future aspects of this field.

II. STOCK MARKET PREDICTION

Stock market prediction means the procedure used to forecast the future movements in stock price using various data sources and information. Conventional methods include the following:

Fundamental analysis: Using fundamental analysis we evaluate the financial health of a company. Company earnings, revenue, dept and other data is used to predict the value of the stock and future performance.

Technical analysis: Using technical analysis we valuate statistical indicators and chart patterns based on data volume and historical prices to predict the future movements in price.

Both these methods use quantitative data rather than qualitative factors such as new events and public sentiments which can greatly affect the stock prices.

III. STOCK MARKET PREDICTION AND NLP

NLP is a part of artificial intelligence which is based on the human computer interaction using the natural language. This field makes the computers capable to understand the human language and interpret it. There is huge amount of textual information available in the financial market, that can be analysed using Natural Language Processing techniques easily:

3.1 Social Media Posts

Social media platform such as facebook, twitter etc. generate huge amount of real time sentimental data. The NLP techniques can analyse this social media information to determine sentiments and trends that greatly effect market movements. The collective sentiments posted on these platforms and speculation of users generally precede price movements.

3.2 Financial Newspapers

Financial newspapers publish daily reports on company earnings, acquisitions, economic events, and regulatory changes etc. which greatly impact stock prices. Using NLP algorithms we can apply topic modelling techniques and sentiment analysis to extract useful information from these data sources to predict the trends.

3.3 Earning Reports

The company publishes their earning reports such as income statements, balance sheets etc that contain valuable information about the company. NLP techniques can be used to extract important information from these reports such as debt levels, profit margins, revenue growth etc. From financial filings the sentiment and tone of the language used by company executives can also be processed to predict the price movements.

3.4 Financial Analysts

The financial analysts give their views on stocks to buy, sell or hold them along with other details. These sentiments and recommendations of financial analysts can be processed using NLP techniques to predict the market behaviour.

IV. NLP TECHNIQUES FOR STOCK MARKET PREDICTION

Many NLP techniques can be used for the prediction in stock market price movements:

4.1 Topic Modeling

Large data sets of text are processed to find the thematic structures that remain hidden in it. Latent Dirichlet Allocation is a popular topic modeling algorithm that can extract the key topics from social media and financial news. These topics can further we processed to correlate them with stock movements. Topic Modeling is helpful to identify any changes in market conditions that impact stock prices. Any political event that is related to the economy greatly effect the market reaction.
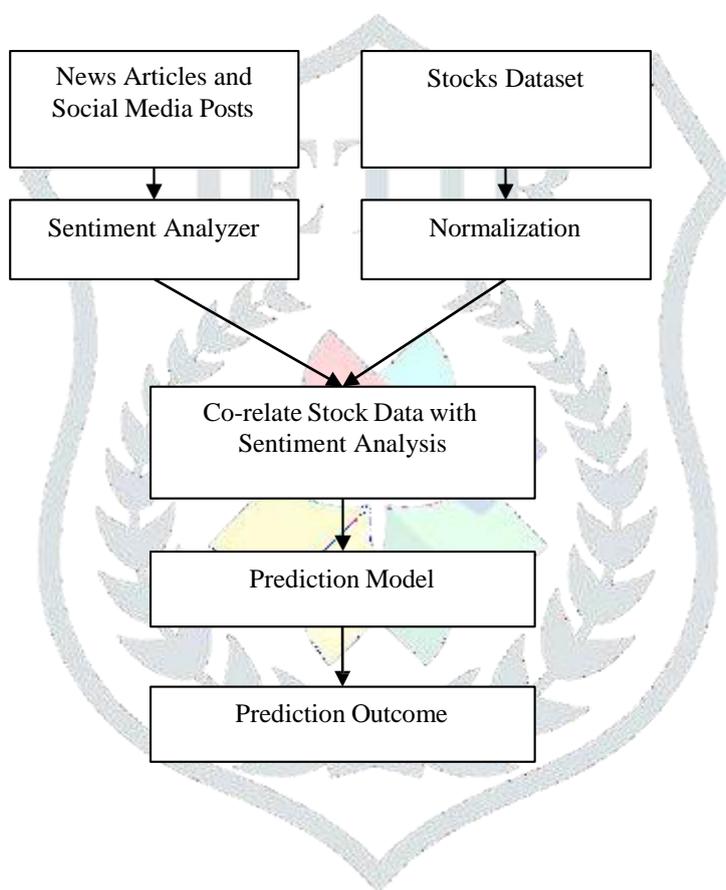

Figure 1. Sentiment Analysis for Prediction

4.2 Sentiment Analysis

This technique uses the emotional tone in the textual information to categories it as positive, negative or neutral, based on the sentiment. Using this analysis, financial news, social media contents, earning reports etc. can be processed to predict future movement of prices. NLP technology is helpful to detect the positive or negative sentiment in the text such as earning report which if positive, signals the upward price movement. The news articles can be processed to detect emotions such as surprise or fear to get further estimates about the market behaviour.

4.3 Named Entity Recognition

Named entity recognition helps us to identify named entities from the text such as name of the company, stock symbol etc. Named entry recognition can be used to extract information about specific company or some stock from the earning reports, social media discussions or news articles. The analysis of named entities and the appearance frequency in the text may help to predict the stock price.

4.4 Machine Learning Technique

Machine learning can be used along with NLP models to develop advanced algorithms for stock prediction. Historical data can be used to train these models which may also include sentiments, named entities and topics. To handle complex patterns in financial and textual data, transformer based models and LSTM networks of deep learning techniques can be used.
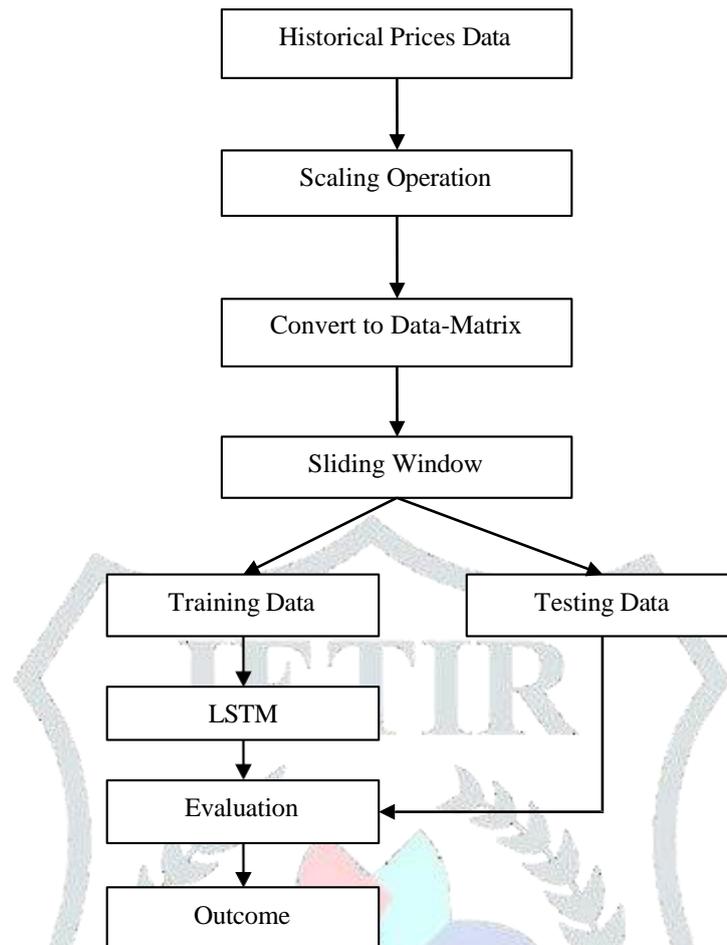

Figure 2. Machine Learning and LSTM

V. CHALLENGES

All though there are many advantages of using NLP techniques for stock market prediction but at the same time it suffers from several challenges in using it:

5.1 Language Ambiguity

Natural languages or human languages are much ambiguous because a single word may have multiple meanings as per the context in which it is used. NLP techniques must be made so capable that they can distinguish the word meaning depending upon the context in which it is used.

5.2 Data Quality

The raw data captured from earning reports, financial news and social media post maybe of low quality containing lot of noise and biases. This data need to be preprocessed to remove stop words and perform stemming of words. Then only the NLP models will work well with this type of data.

5.3 Generalization

It has been seen that many machine learning models have the problem of overfitting. In this problem the model works well on training data but does not work well when it is operated on new data. Powerful validation techniques can be used to overcome this problem so that the model can generalize in every condition.

5.4 Processing in Real Time

Stock market prices goes on continues changes. The movement of stock prices fluctuates in real time, so NLP models must be capable to analyse huge data in real time to provide better predictions. To make it possible high end infrastructure is needed having efficient algorithms to handle live data.

VI. CONCLUSION

It is the beginning stage where NLP is being used for stock market prediction, so it needs much improvement. In future, data from many sources including numerical and textual data and audio and images from various data sources will include. More sophisticated models will be used to process that are more accurately using advanced deep learning techniques. The computational power is being improved at a large extent, it will definitely help to extend the ability of models to analyse large real time data and provide more accurate and timely predictions.

NLP provides better approach for market prediction based on qualitative data and can better predict the markets than quantitative models used traditionally. But at the same time there are several challenges regarding the ambiguity of language data quality and processing of data in real time to accurately predict the market behaviour. The growth of NLP will also assist the analysts, investors and traders to predict the future price changes in financial markets.

REFERENCES

[1] Yu, L., & Liu, B. (2006). Sentiment analysis for financial news and stock prediction. Proceedings of the AAAI Conference on Artificial Intelligence.

[2] Das, S. R., & Chen, M. Y. (2007). Yahoo! for Amazon: Sentiment extraction from small talk on the web. Proceedings of the 8th ACM Conference on Electronic Commerce, 106-112.

[3] - Zhang, Y., & Skiena, S. (2010). Improving Financial Prediction by Combining News Articles and Historical Prices. Proceedings of the 4th International Conference on Weblogs and Social Media (ICWSM).

[4] Liu, B., & Zhang, L. (2007). A survey of opinion mining and sentiment analysis. Proceedings of the 5th International Conference on Computational Linguistics and Intelligent Text Processing.

[5] Antweiler, W., & Frank, M. Z. (2004). Do news articles improve stock prediction? Journal of Financial and Quantitative Analysis, 39(3), 527-546.

[6] Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of Computational Science, 2(1), 1-8.

[7] Siriwardane, M., & Mathur, S. (2009). Using textual data for predicting stock returns. Proceedings of the 15th International Conference on Data Mining (ICDM), 1010-1014.

[8] Schumaker, R. P., & Chen, H. (2009). Textual analysis of stock market prediction using financial news articles. Proceedings of the International Conference on Information Systems (ICIS), 175-186.