



# THE ANALYTICAL STUDY ON INFLUENCING STOCK MARKET USING TWITTER DATA

Dr.C K Gomathy  
Assistant Professor,  
Department of CSE

SCSVMV Deemed to be University, TN, India

Swecha Thumu  
UG Student

Department of CSE

SCSVMV Deemed to be University, TN, India

T Vandana  
UG Student

Department of CSE

SCSVMV Deemed to be University, TN, India

**Abstract**— This paper work system finds out the relation between sentiment extracted from twitter and particular stock price. Sentiment Analysis is a machine learning method that can determine whether a specific statement is positive or negative. This technique is most widely used in analyzing customer reviews or feedback. The data required in sentiment analysis must be specific and should have a considerable volume. There are several sources of Amazon product reviews on the internet, but some of them are fraudulent. Thus, it is crucial to find relevant and legitimate datasets to produce accurate results. This proposals main objective is to generate a sentiment analysis for tweets provided on a particular stock item and compare it with market prices.

**Keywords:** Stock market prices, sentiment Analysis, and Stock influencing system.

## I. INTRODUCTION

Twitter, have you ever thought one tweet could bring the world's richest person to the 4th position? It has happened to Elon Musk and lost 150 Million dollars for just one tweet mentioning his bitcoin purchase. Nowadays, social media has a prominent role in our lives, where most social figures use Twitter as a platform for their freedom of speech. However, this could also impact political, financial and relations between countries as well. This Paper have two important objectives:

- To extract any relation between people sentiment on twitter and stock price.
- Predicting stock price using machine learning.

## II. LITERATURE SURVEY

Social network analysis is the study of people's interactions and communications on different topics and nowadays it has

received more attention. Millions of people give their opinion of different topics on a daily basis on social Medias like Facebook and Twitter. It has many applications in different areas of research from social science to business. Twitter nowadays is one of the popular social media which according to the statistics currently has over 300 million accounts. Twitter is the rich source to learn about people's opinion and sentimental analysis. For each tweet it is important to determine the sentiment of the tweet whether it is positive, negative, or neutral. Another challenge with twitter is only 140 characters is the limitation of each tweet which cause people to use phrases and words which are not in language processing. Recently twitter has extended the text limitations to 280 characters per each tweet. Social networks is a rich platform to learn about people's opinion and sentiment regarding different topics as they can communicate and share their opinion actively on social medias including Facebook and Twitter. There are different opinion oriented information gathering systems which aim to extract people's opinion regarding different topics. The sentiment-aware systems these days have many applications from business to social sciences. Since social networks, especially Twitter, contains small texts and people may use different words and abbreviations which are difficult to extract their sentiment by current Natural Language processing systems easily, therefore some researchers have used deep learning and machine learning techniques to extract and mine the polarity of the text. Some of the top abbreviations are FB for Facebook, B4 for before, OMG for oh my god and so on. Therefore sentimental analysis for short texts like Twitter's posts is challenging. This technical paper reports the implementation of the Twitter sentiment analysis, by utilizing the APIs provided by Twitter itself. There are great works and tools focusing on text mining on social networks. In this project the wealth of available libraries has been used. The approach to extract sentiment from tweets is as follows. Start with downloading and caching the sentiment dictionary. Download twitter testing data sets, input it in to the program. Clean the tweets

by removing the stop words. Tokenize each word in the dataset and feed in to the program.

. For each word, compare it with positive sentiments and negative sentiments word in the dictionary. Then increment positive count or negative count. Finally, based on the positive count and negative count, we can get result percentage about sentiment to decide the polarity. Researchers have done different sentimental analysis on Twitter for different purposes for example the work designed by Wang, et.al is a real-time twitter sentimental analysis of the presidential elections. The latest work also proposes a similar hybrid neural network architecture, integrating a convolutional neural network with a bidirectional long short-term memory to predict the stock market index While the researchers frequently proposed different neural network solution architectures, it brought further discussions about the topic if the high cost of training such models is worth the result or not. There are three key contributions of our work a new dataset extracted and cleansed a comprehensive feature engineering, and a customized long short-term memory (LSTM) based deep learning model. We have built the dataset by ourselves from the data source as an open-sourced data API called Tushare.

### III. PROBLEM STATEMENT

In general tweets have major impact on different sectors. For example, Elon musk, CEO of Tesla made a tweet about purchasing of bitcoins and this made the surge in price of bitcoin and took a dip on his own stock Tesla. Taking the above example as a reference, we are trying to prove it through coding whether the twitter data is influencing the stock prices or not. Using Natural language processing and machine learning we are trying to prove the above fact

### IV. PROPOSED SYSTEM

We are trying to prove it through coding whether the twitter data is impacting the stock prices. Using Natural language processing and machine learning we are trying to prove above known accepted fact. Proposed feature engineering along with a fine-tuned system instead of just an LSTM model only. We observe from the previous works and find the gaps and proposed a solution architecture with a comprehensive feature engineering procedure before training the prediction model. With the success of feature extension method collaborating with recursive feature elimination algorithms, it opens doors for many other machine learning algorithms to achieve high accuracy scores for short-term price trend prediction. It proved the effectiveness of our proposed feature extension as feature engineering. We further introduced our customized LSTM model and further improved the prediction scores in all the evaluation metrics. The

proposed solution outperformed the machine learning and deep learning-based models in similar previous works.

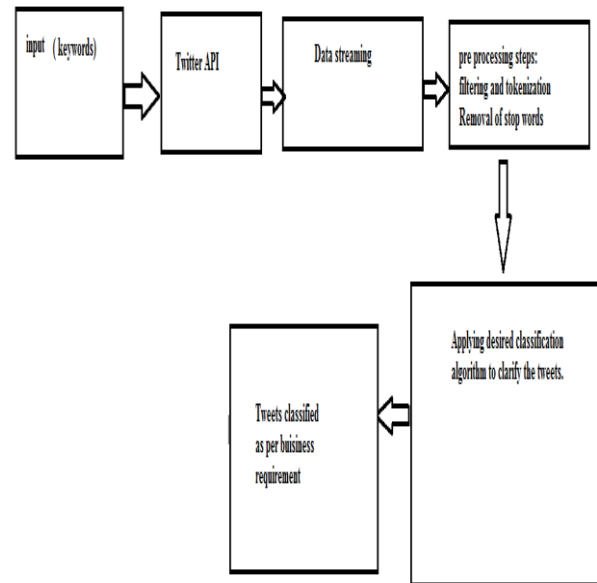


Fig 1: Business Requirement function system

### V. METHODOLOGY

The methodology we have used is a hybrid approach to build the system, which is a combination of all the statistical, natural language processing and machine learning-based calculation techniques. Below mentioned are the steps we follow in the project:

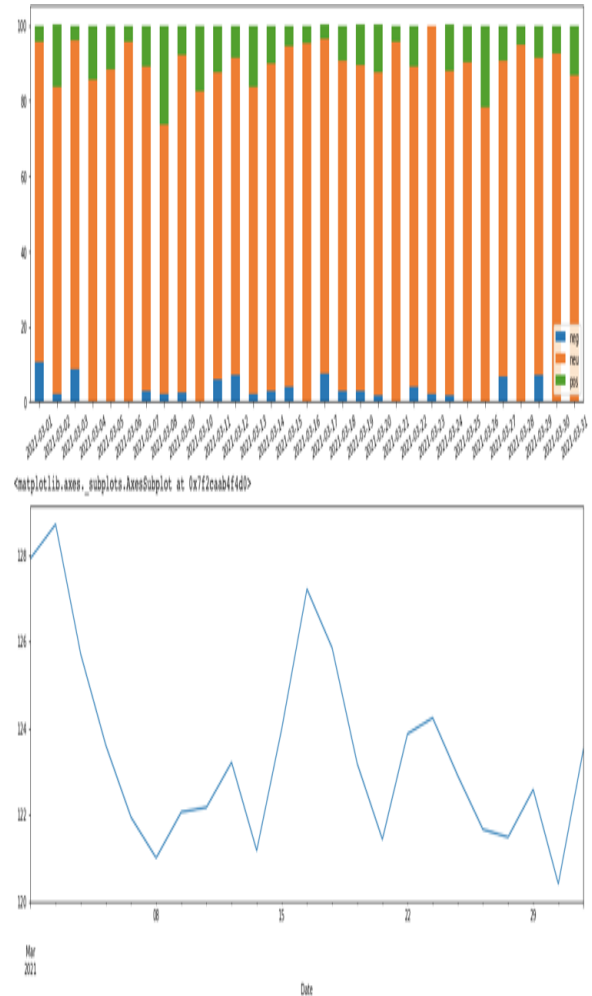
- Collected the data from using Twitter web scraper named scrape and stock data using Yahoo, data reader function by pandas.
- Data cleaning is followed next NLP (Natural Language Processing) techniques were uses such as punctuation removal, special characters removal., translation of text in tweets to English language.
- Calculating sentiment analysis of users using Vader Sentiment Analyzer.
- Visualising the Sentiment Analysis data.
- Extract stock prices using yahoo Api, standardisedata,
- Applying LSTM machine learning algorithms to the data.
- Evaluation is done along with the accuracy check.
- The libraries used for our project include NumPy, pandas, Seaborn, Spicy, Matplotlib, Sklearn, NLTKand snsrape.

## VI. IMPLEMENTATION PROCEDURES ON TWITTER DATA

Collection of data from Twitter was a challenging part due to Twitter API (Application Programming Interfaces) limitations like no of tweets could be extracted can only be extracted for the past week (Maximum of seven days). So, Python package named sncscrape had been used to extract data from Twitter for desired period. This package will scrape the data, which is visible over web. After collection of tweets data, the dataset was cleaned, tokenised, lowered, translated using Google Translator API. Sentiment calculation had been done on these tweets using Vader Sentiment Analyzer and grouped based on dates. This sentiment data has been visualised using seaborn and Matplotlib. Prediction of stock price was the next step using LSTM technique. The Implementation is the process/ML algorithm life cycle steps that we will follow in the project. Some steps are given below:

- Collect the Twitter data for the last 30 days on a particular stock tick Ex: AAPL, TSLA, GME etc.
- Preprocess, clean and tokenize data to make it suitable for sentiment analysis.
- Get open and closing stock prices for a mentioned tick, process the data and create a data frame.
- Create a visualization chart that overlaps the sentiment bar graph and stock price line chart.
- Then apply a machine learning model and predict close market prices.
- After applying models, we will evaluate models and check the accuracy. If the accuracy is not good, we re-evaluate it and can try another model

## VI. RESULTS:



**Fig.1. Tweets sentiment and stock price**

Model: "sequential\_1"

| Layer (type)    | Output Shape   | Param # |
|-----------------|----------------|---------|
| lstm_2 (LSTM)   | (None, 60, 50) | 10400   |
| lstm_3 (LSTM)   | (None, 50)     | 20200   |
| dense_1 (Dense) | (None, 1)      | 51      |

Total params: 30,651  
 Trainable params: 30,651  
 Non-trainable params: 0

**Fig 2: LSTM Model specifications**

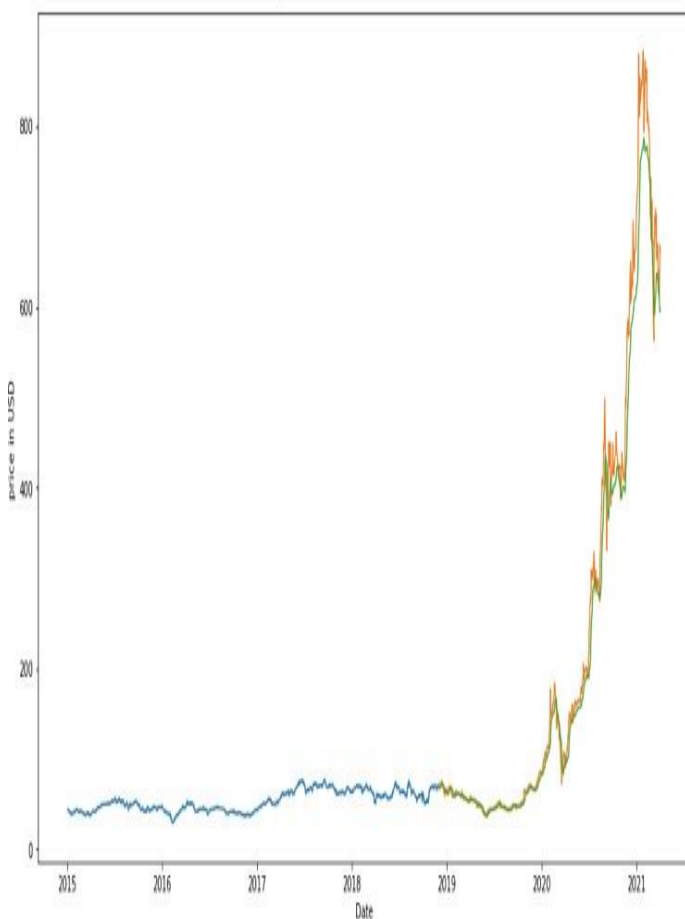


Fig 3: Predicted vs Original values

## VII. CONCLUSION

In this paper, predicting whether the twitter data is influencing the stock market prices. We have observed that it is not influencing the stock data but the prediction is given almost correctly. Couldn't any find any relation between sentiment generated by tweets and stock price. These results might vary depending on the selected stock, no of tweets collected on a given day etc. But our data is not providing any relation. Model to predict closing price based on past 60 days of data has been done and working as expected. Results are satisfying but more investigation in needed in model configuration part.

## VIII. REFERENCES

1. Zexin Hu, "A Survey of Forex and stock price prediction using machine learning", 2010
2. Jingyi Shen, 2012 "Short term stock market trend prediction using comprehensive deep learning system.
3. C.K.Gomathy and Dr.S.Rajalakshmi., "An Efficient Business Integration and Quality Service for Service Oriented

Architecture", World Academy of informatics and management sciences(WAIMS), Volume 1, Issue II (WAIMS/0019/0084), March 2012, P.No: 74-79, ISSN : 2278-1315.

4. C.K.Gomathy and Dr.S.Rajalakshmi, "A proficient Business incorporation and Quality for SOA. ", International journal on information Science and computing, Volume 6, Issue II, July-2012, ISSN : 0973-9092.

5. Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijrcst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-10-13, April '2017

6. Dr.C K Gomathy, Article: A Semantic Quality of Web Service Information Retrieval Techniques Using Bin Rank, International Journal of Scientific Research in Computer Science Engineering and Information Technology ( IJSRCSEIT ) Volume 3 | Issue 1 | ISSN : 2456-3307, P.No:1563-1578, February-2018

7. Dr.C K Gomathy, Article: A Web Based Platform Comparison by an Exploratory Experiment Searching For Emergent Platform Properties, IAETSD Journal For Advanced Research In Applied Sciences, Volume 5, Issue 3, P.No-213-220, ISSN NO: 2394-8442, Mar/2018

8. Dr.C K Gomathy, Article: A Study on the Effect of Digital Literacy and information Management, IAETSD Journal For Advanced Research In Applied Sciences, Volume 7 Issue 3, P.No-51-57, ISSN NO: 2279-543X, Mar/2018

9. Dr.C.K.Gomathy, A.V.Sripadh Kaustthub, K.Banuprakash, Article: An Effect of Big Data Analytics on Enhancing Automated Aviation , International Journal Of Contemporary Research In Computer Science And Technology (Ijrcst) E-Issn: 2395-5325 Volume 4, Issue 3, P.No-1-7. March -2018

10. Dr.C K Gomathy, Article: A Semantic Quality of Web Service Information Retrieval Techniques Using Bin Rank A Cloud Monitoring Framework Perform in Web Services, International Journal of Scientific Research in Computer Science Engineering and Information Technology IJSRCSEIT | Volume 3 | Issue 5 | ISSN : 2456-3307, May-2018

11. Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and Information Technology ( IJSRCSEIT ) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018

12. Dr.C K Gomathy, Article: A Scheme of ADHOC Communication using Mobile Device Networks, International



Journal of Emerging technologies and Innovative Research (JETIR ) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:320-326, Nov-2018

13.Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research ( JETIR ) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

14.Dr.C.K.Gomathy,P.Sarvani Divya jyothsna,M.Srimayi, Article: A study on the Mobile Application Advancements in Anti-Ragging, SSRG International Journal of Computer Science And Engineering(SSRG-IJCSE)-Volume 6 issue 3, March 2019.

15. C.K.Gomathy and Dr.S.Rajalakshmi.(2014),"A Software Quality Metric Performance of Professional Management in Service Oriented Architecture", Proceedings of ICCTET'14, organized by Akshaya College of Engineering, Coimbatore. Archived in IEEE Xplore Digital Library, July 2014,ISBN:978-1-4799-7986-8.

16. C.K. Gomathy, V. Geetha, Venkata Nagendra, Venkata Sai Ram, Article: The Secured Proficient Smart Electronic Voting System, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-4, April 2020

17. V.Geetha, C.K.Gomathy, P.Manojkumar, N.S.L.S.V.Manohar, Article: The Sentimental Analysis for E-Commerce Application, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-5, June 2020

18.C K Gomathy,V Geetha,P.S.lavakumar,K.V.Rahul,The Efficient Automatic Water Control Level Management Using Ultrasonic Sensor, International Journal of Computer Applications (0975 – 8887) Volume 176 – No. 39, July 2020

19.VGeetha, C.K. Gomathy, Maddu Pavan Manikanta Kiran, Gandikota Rajesh, The Implementation of Fleet Monitoring System using Mobile based Software, International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-9 Issue-4, April 2020

20.C K Gomathy, Student Location Tracking Inside College Infrastructure, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 07 Issue: 06 | June 2020.