SENTIMENT BASED STOCK TRADE PREDICTION

[1]**Sarthak Gupta,** [2]**Vipul Garg,** [3]**Tushar Agrawal**

[1]Computer Engineering, [2]Computer Engineering, [3] Computer Engineering
[1] Poornima Institute of Engineering and Technology,
Jaipur, Rajasthan, India

Abstract: We live in a digital age where social media usage has increased to previously unheard-of levels in recent years. One such well-known social networking and microblogging website is Twitter, which enables hundreds of millions of users to communicate informally and in real time about events that merit wide attention as well as to express opinions on such occurrences. The results of this study suggest that emotion on Twitter and stock market movement are related. We are particularly curious about whether and to what extent sentiment data gathered from Twitter might be used to predict future changes in stock prices. Market forecasting is a well-known and important topic in financial research, academic study, and the financial sector. Time series analysis is the most common and fundamental method used to carry out this task. The tweets mentioning a certain company over the previous three months are collected together.

The results of this experiment show a statistically significant correlation between daily stock price fluctuations and changes in tweet polarity, which was determined via sentiment analysis of tweets.

Keywords: Sentiment Analysis, Twitter data, Share market prediction, Data analysis.

I. INTRODUCTION

In the past 40 years, data has mostly been used to record and report business operations and scientific transactions. In the following 40 years, data will also be used to influence business decisions and hasten scientific advancements. The Internet has evolved in recent years into a public arena where everyone can voice their opinions. People can submit their opinions on a variety of online social media platforms, including forums, wikis, blogs, social networking sites, and Twitter. These articles provide in-depth, practical knowledge about a variety of topics or events that can be applied successfully in a variety of contexts.

One of the major challenges in data mining and knowledge discovery is using this internet data and extracting valuable information from it.

These unstructured data provide crucial information that can be used to develop new services for organisations, people, or governments. Every investor hopes to be able to predict future stock market action. Fundamental analysis and technical analysis are the two basic stock market forecasting techniques. Despite all the research that has been done, this problem is still difficult. Information is enormous and difficult to understand because it comes from so many different news sources. Numerous studies have been conducted recently in a variety of fields that have addressed this problem and suggested other solutions.

Recent advancements in the field of machine learning have refocused the attention of many academics on this subject. Artificial intelligence has advanced quickly during the past three decades, moving from multilayers to deep neural networks. Long-term memory (LSTM) and neural convolution (CNN) are similar to DNNs. Undoubtedly, one of the most well-liked industries where people perceive potential for effective DNN application is the finance sector. Chen et al. employed LSTM and fundamental Stock statistics, including Open, Close, High, Low (OCHL) stock-market prediction prices, while Bollenet al. used Twitter to predict market movement.

enhancement of conventional methods. Nelson, al.With few exceptions, LSTM experiments and stock technical indicators from OCHL outperform conventional methods. However, the results demonstrated novel techniques capable of achieving acceptable precision.

II. Literature Survey

Platforms for social networking have developed into quick and inexpensive means of communication that let consumers readily and quickly access political information. Social media users express their opinions and thoughts about various parties and candidates during elections.

Initial research generally yielded favourable results on Twitter's capacity to predict election results. According to some research, just the volume of referrals to a candidate or party could predict election results [Gordon, 2013].

On the internet, people can express themselves in a variety of media. Examples of these media include blogs, wikis, forums, and social networks where users can post content, express their opinions, and receive feedback from other users. Together, they provide a wealth of information on a variety of topics, including politics, health, product reviews, and travel, as well as a variety of life domains. Opinion data will likely play a large role in online textual data as seen by the rising popularity of personal publishing services of all kinds.

i) Social Media Analysis
Due to the abundance of real-time unstructured data available, social media data has becoming more widely used in the domains of information retrieval (IR) and text mining. Every blog post, comment, and tweet they make could be an expression of their feelings [Jain and Kumar, 2015a]. These unstructured data collections include important data that could be leveraged to create new services for organizations, enterprises, and people. By using this unstructured data, a new field called opinion mining and feeling analysis was created.

ii) Sentiment Analysis
Existing sentiment analytical work has been categorised as a text classification challenge by Jain and Kumar (2016) using a range of techniques, including document sentiment analysis, sentencing analysis, aspect-based feeling analysis, comparative sentiment analysis, and lexical feeling acquisition.

III. System Design

This section provides a quick overview of the process used to predict stock prices using historical data and user sentiment on social media.

The ARIMA and LSTM algorithm predicts stock market deviation by gathering data, processing it, and then displays the result of the prediction in table and graph style. It is divided into three main parts. These are the Module of Current Market Analysis, Module of Result and Prediction, and Module of Historic Data. The historical data module is further divided into the data gathering and preprocessing, the sentiment analysis, and the data training modules for the current market analysis.
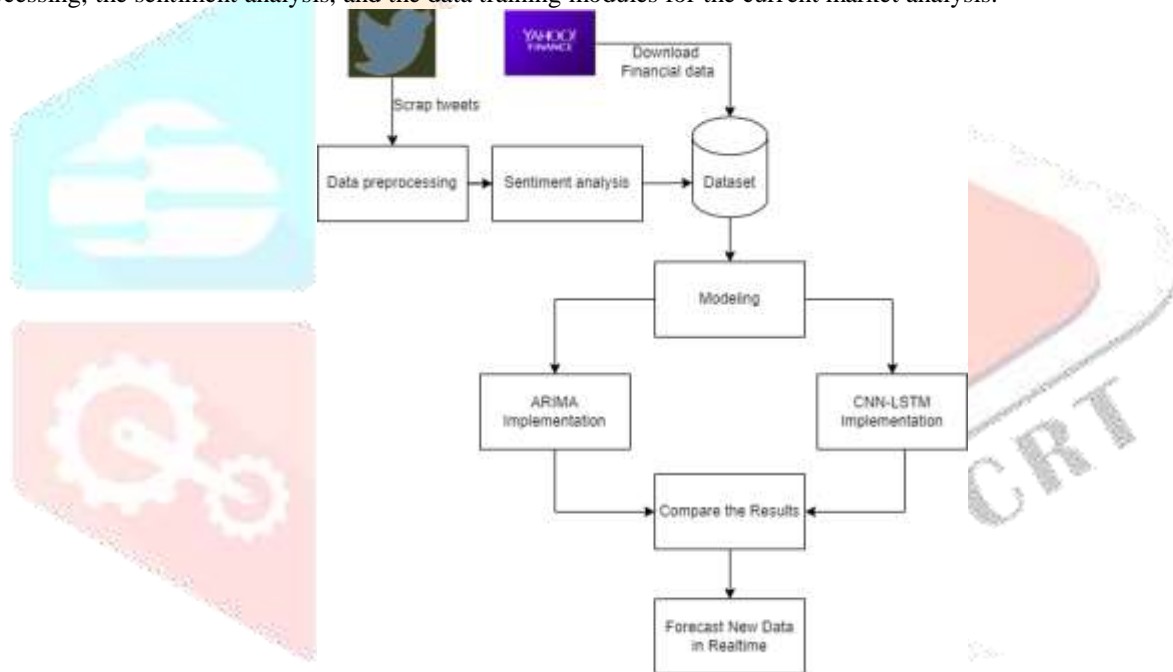
Data Gathering
The "get symbol" feature of the system is used by the module to gather real-time data. The company's opening and closing valuations are used to gather data [16]. It then shows a date-to-value graph on the x and y axes in accordance to show the results of the home screen.

Module for Pre-processing Data:
The preprocessing module eliminates any erroneous values and only accepts the essential parameters. This method's predictive variables include the date, volume, and opening and closing stock prices. The preprocessing programme removes the leftover parameters and provides a clear set of input data for the system.
structure because it is present in the form of lists with strings.Let's use the feature's literal_eval() function.

Data Training and Forecasting Module:
Two methods, ARIMA and Long Short-Term Memory (LSTM), are used to train the algorithm.

ARIMA Algorithm:
This module considers sentiment analysis data as well as historical data. Sentiment analysis is most frequently used when a dramatic shift in market interest occurs. It often occurs when a company changes its policy, when new software is released or upgraded, when there has been terrible publicity or an adverse media effect brought on by influencers, etc. This module is the most crucial since it is used to predict potential stock levels that the company may encounter in the future (based on user requirements). The prediction module looks into the sentiment analysis module to get data on the company's present status with investors and analyses the pattern revealed by the past data that has been provided.

It collects information from corporate Twitter users and categorises it into positive or negative classifications. If a class contains more tweets than the predicted number, the forecast is changed. The direction change in the exhibited graph serves as a clue. The current settings for the ARIMA algorithm, often known as ARIMA, are 2 for auto regression, 2 for integration, and 2 for moving parts.

Moving Average Process:
The current value of the moving averaging process is a linear mixture of the current disturbance with one or more past disturbances, with the current disturbance being the most recent. The shifting average's order shows how many prior periods have been factored into the present figure.

Long Short-Term Memory (LSTM)
This study suggests an attention-based, long-term memory model for predicting the trend of stock prices. The input layer, the hidden layer, the attention layer, and the output layer are the four layers that make up the model. The data is cleaned up in the input layer to make sure it complies with model input specifications. The LSTM unit links the line model network to the hidden layer. The vector provided a numerical value to the attention layer. The output layer receives the computations' results. The issue of model training is dealt with using the gradient descent technique.

Twitter Sentiment Analysis
A feeling is frequently used to describe an emotional point of view or perspective. When users from various user groups are examined and their attitudes towards the entire relevant extreme or enthusiastic response to a report, message, or occasion are determined, sentiment analysis is used to capture these intense feelings.
Instead of understanding the reviewer's profound views, analysis of the sensitivity enables the consumer to realise the circumstances of what the review is about, whether it be a specific product or ongoing concerns.
Twitter is one of the many well-known social media sites where users may share their opinions about products, services, and brands. Determining whether someone has good, negative, or neutral views of you can be done very well by looking into their twitter sentiments. People may want to share their knowledge on the subject if they discover important or fascinating information. The focus could be on a product, service, social message, or anything else.

Technical Analysis utilizing indicators
Technical analysis is the art of analysing historical stock data, such as price changes, volume, and market patterns. In order to help investors anticipate what's more likely to happen to prices over a shorter length of time and to enable them to make informed decisions, future stock movement predictions are made using graphs and charts.
In order to make the most profits or suffer the fewest losses, one must choose the best period of time to enter or invest in a certain asset and exit. Such decisions are dependent on how prices function, the signals they produce, and the market's key turning points. Technical indicators that help investors find these movements and patterns are typically technical analyses.

There are four primary categories of technical indicators:
- Trend
- Volatility
- Momentum
- Volume.

Trend: These technical indicators help determine the direction and strength of the stock. The pricing plot is often levelled and shown as a single line. Due to this process, the indicator lacks the abrupt price change known as trends. One of the primary negatives is that when the market is unstable, these indicators lose money. The moving average convergence divergence (MACD) indicator is one of the important trend indicators.

Volatility Indicators: The volatility indicators used by the analysts help them choose exact entry and exit points for each transaction. They are employed to calculate movement rate while disregarding direction. The Bollinger Bands are a key indicator.

Momentum: Momentum indicators use price comparisons over a wide range of time periods to help spot price movements. It is calculated by comparing the most recent closing price to the previous closing price.

Volume is the total number of shares traded during a specific period of time. A period of time might range from 1 to 1 Year. Volume has a significant impact on the price and direction of a particular stock.
Technical indicators with a focus on volume use volume-based computations.

IV. Result:
When we examined the outcomes of the two models, we discovered that:
The tweets had a favourable effect on accuracy.
A greater performance was anticipated if the model trained, once more with more data and with more parameter tuning. Our model has a fast-learning curve, achieving the Arima's accuracy in a few epochs and without any optimisation work. LSTM and CNN would considerably improve with more data, taking this into account.

V. Conclusion:

The stock market is becoming increasingly popular, which has prompted researchers to create fresh, innovative, and predictive methodology. The projection method is useful for investors and anyone with a stock market in addition to academics.

A precise projection model is required to help forecast stock indices. Many gauge advances are currently being employed to help financial professionals, examiners, and anybody else interested in investing money in the securities exchange by giving them superior knowledge about the future conditions of the securities exchange.

VI. Future Scope

The future scope provides options for scholars to advance in their discipline, including the following research areas:

1. To increase accuracy, sentiment analysis should be taught in more than just two classes.
2. To broaden sentiment research, in addition to data from Twitter, consider data from blogs, online polls, and Facebook.
3. The automatic messaging system for investors during volatile market swings.

VII. REFERENCES

[1]. Vaibhav Chavan1 , Dr. Sanjeev S. Sannakki2 , Kuldeep P. Sambrekar3 Share Market Prediction using Twitter Sentiment Analysis 1,2,3KLS's Gogte Institute of Technology, Belagavi, India

[2]. 1Hardik Lad, 2Jairaj Saraf, 3Siddhi Pawar, 4Dr. Vidya Chitre Stock prediction and Analysis using Twitter Sentiments: A Survey 1,2,3UG Student, 4Assistant Professor, Vidyalankar Institute of Technology, Mumbai, India,

[3] P. Chakraborty, U. S. Pria, M. R. A. H. Rony and M. A. Majumdar, "Predicting stock movement using sentiment analysis of Twitter feed," 2017 6th International Conference on Informatics, Electronics and Vision & 2017 7th International Symposium in Computational Medical and Health Technology (ICIEV-ISCMHT), 2017, pp. 1-6, doi: 10.1109/ICIEV.2017.8338584.

[4] Sarkar, Abhijoy & Sahoo, Abhaya & Sah, Sourav & Pradhan, Chittaranjan. (2020). LSTMSA: A Novel Approach for Stock Market Prediction Using LSTM and Sentiment Analysis. 1-6. 10.1109/ICCSEA49143.2020.9132928.

[5] Aman, S. and Szpakowicz, S. (2008). Using Roget's Thesaurus for fine-grained emotion recognition. In Proceedings of the Third International Joint Conference on Natural Language Processing, Hyderabad, India, pp.296- 302.