VIT
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

Sentimental Analysis for Stock Prediction

Team 7

In guidance of
Prof. Saravanakumar K

Done by
Harsh Vivek Londhekar - 19BCE0496
Gokul R Nair - 19BCE2245
Hemaksh Chaturvedi - 19BCE2222
Devang Gupta 19BCE0959

ABSTRACT

In this paper we have tackled the forefront of sentimental analysis used in the stock markets. Sentimental analysis refers to the contextual understanding of the desired target, this can be done by using machine learning approach which will classify and analyse the target's sentiments, behaviour, opinions and emotions, etc, about a specific topic (here, the stock market) which they have conveyed via their speech, text, etc. This understanding of text is done with the help of NLP. The stock market is a place where people take part in a continuous exchange (buying and selling) of shares. It has always been a place to earn money and during the pandemic it has experienced a great boost in the number of people interacting with the market. In this survey paper, we have analysed and given our understanding of the topic; which includes the study on a large collection of research papers, articles and existing systems on the topic. Sentimental analysis can be used in stocks in hopes to determine an investor's opinion on a certain group/set of assets or stocks. By understanding these opinions, the analyst may be able to predict the future price action. If a person is able to understand the opinions and sentiments of the general public, he/she would be able to make a prediction of high accuracy. This literature survey is done to study the sentiment analysis of stock market in-depth and to familiarize with other works done on the subject

KEYWORDS

Sentimental analysis, stock, stock market, prediction, investor sentiment, opinions , machine learning, NLP, shares

INTRODUCTION

Sentiment Analysis has various employments. Most prominently, with the ascent of online media destinations like Facebook and Twitter, the expanded fame of websites, and the ascent in audit, rating, and proposal locales, organizations are turning out to be progressively keen on sentiment analysis. With buyers ready to impart their insights across the web so effectively, sentiment analysis has turned into significant money for organizations and organizations attempting to develop their advanced notorieties, distinguish new freedoms, and effectively market their items. In any case, with such a lot of data out there, it may very well be difficult for organizations to focus on the most important pieces of shoppers' remarks. That is the reason sentiment analysis is so helpful. Associations are utilizing the force of sentiment analysis to sift through this significant data to all the more likely to comprehend shoppers' discussion and take more successful, better-designated activity. **[Daudert, T. (2021). Exploiting textual and relationship information for fine-grained financial sentiment analysis. Knowledge-Based Systems, 107389.]**

Sentiment analysis is an approach to recognize the feelings communicated by individuals as text furthermore, to decide if the feelings are good, unbiased or negative. Verifiable feelings show up in a general as opposed to a source like a paper. Subsequently distinguishing feelings in text that don't unmistakably communicate feelings or don't contain such words are hard to measure. If we can further develop opinion investigation by recording settings, this methodology will break down the understanding of text and cycle it. Opinion virus implies people interfacing with one another and catching their feelings to shape an assessment on one another. Hence by utilizing AI models and different systems we can change over ordinary text also, infer the opinion communicated by the text to examine whether the articulation is positive, unbiased or negative.

The main objective of sentiment analysis is to infer the sentiments of people through text based material like social media or other sources and determine if these sentiments are positive, negative or neutral towards a topic.

The stock market is an important sector of the economy and plays an important role in industry development. Social media nowadays is the perfect platform to reflect people's opinion about a company. The task of predicting share prices of a company requires strenuous efforts because of the sensitive financial market, external chaos, political influence, natural calamities and much more. Not only the customers, sellers and buyers will be able to predict the share prices of a company but the company/organisation itself will be benefited a lot with the help of sentiment analysis. The company can figure out how the customers and people feel about them and if the company has negative feedback, they can improve and definitely change their tactics. This will help them stay in the lead when compared to other companies in the market. Companies can incorporate different methods to maintain a positive opinion from the people and this will indirectly increase the stock prices of the company, eventually increasing the valuation of the company.

The problem chosen and discussed in the paper i.e. Sentiment Analysis to predict stock market and share prices of a company is useful for both the people and the company itself. Since the prediction of the stock market is a strenuous task and the stock market is a very volatile market, the process of sentiment analysis makes this extremely unpredictable task an effortless one.

Sentiment analysis can help big companies develop a strategy and improve the public relation. This will change the public opinion into a positive one and indirectly increase the profits of the company because the stock prices will increase due to a positive review of the company. With the help of sentiment analysis a company can develop a customer centric product based on the opinion they gathered from SA.

Not only can this sentiment analysis also help in political analysis. People's opinion regarding political topics like social media like twitter can greatly help to form an opinion regarding the particular political topic.

Sentiment analysis can likewise be utilized for data mining, or assembling serious insight about your rivals. For instance, a brand could without much of a stretch track online media notices or notices of rivals in different spots across the web, and examine how shoppers feel about the contenders and their items. This is an amazing method to acquire a strategic advantage in the present exceptionally cutthroat commercial center.

In this day and age, by far most online media apparatuses have some sort of opinion capacities. Organizations additionally utilize mechanized sentiment analysis dependent on word records, with each word being given a pre-characterized opinion esteem. This has various occupations. For instance, an eatery may participate in web-based media observing to measure how individuals feel about their menu, sort out whether individuals partook in their food, and discover what sentiments individuals related with their general involvement with the café. What's more, fortunately the precision paces of sentiment analysis for online media checking continue to improve. Sentiment analysis organizations accomplish accurate results of more than 75% with its mechanized opinion investigation abilities.

These were the uses and real time application of sentiment analysis. By the above mentioned points people along with the company will be benefited.

Now moving on with the history of sentiment analysis, this method of predicting the nature of a subject has been used for a long time, and has proven to be a very beneficial method. Sentiment analysis was done first in the 1950s. There are several types of sentiment analysis methods which concern with attitude, emotions and opinions. Various sorts of feeling examination utilize various procedures and methods to recognize the opinions contained in a specific text.

Sentiment analysis is broadly classified in two types : Subjectivity/objectivity identification and Feature based identification. Subjectivity/objectivity identification classifies text into subjective or objective type. Feature based identification method determines different sentiments and analyzes them in relation to different aspects of a subject.

The sub topic present in the survey paper is prediction of stock market and share prices of a particular company. Prediction and investing in shares of a company by seeing the fall or rise in the stock price has been practised since a long time and has been a major reason for economic growth. In recent times sentiment analysis has shifted its topic from product reviews to processing texts from social media. Stock prediction, political analysis, and marketing have increased the scope and utilization of sentiment analysis. Tracing back to the 1950s sentiment analysis was first used to process text from documents. From then it started to be used to get product reviews. And evolving so on, sentiment analysis is used for data mining also known as opinion mining, share market prediction, processing text from social media platforms like twitter and facebook and so on. At present sentiment analysis is a crucial component in economic growth of any company and this has benefited both the company and the customer base.

The different approaches which were used in the papers were wavelet analysis, different machine learning models like random forest, gradient boosting genetic algorithm **[Ghosh, I., Chaudhuri, T. D., Alfaro-Cortés, E., Martínez, M. G., & Rubio, N. G. (2021). Estimating the relative effects of raw material prices, sectoral outlook and market sentiment on stock prices. Resources Policy, 73, 102158],** Fine tuned textual representation**[Daudert, T. (2021). Exploiting textual and relationship information for fine-grained financial sentiment analysis. Knowledge-Based Systems, 107389.]**, deep learning models like LSTM (Long Short Term Memory)**[Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Computer Science, 7, e476.]**, adaptive sentiment-aware deep deterministic policy gradients approach**[Koratamaddi, P., Wadhwani, K., Gupta, M., & Sanjeevi, S. G. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. Engineering Science and Technology, an International Journal, 24(4), 848-859.],** convolution neural networks (CNN)**[Deshmukh, R. (2021). Stock Prediction by using NLP and Deep Learning Approach. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(1S), 202-211.]**, recurrent neural networks**[Kilimci, Z. H., & Duvar, R. (2020). An Efficient Word Embedding and Deep Learning Based Model to Forecast the Direction of Stock Exchange Market Using Twitter and Financial News Sites: A Case of Istanbul Stock Exchange (BIST 100). IEEE Access, 8, 188186-188198.]**, deep learning algorithms and Word2Vec, GloVe, and FastText. In other papers sentiment analysis was done using BERT techniques**[Nemes, L., & Kiss, A. (2021). Prediction of stock values changes using sentiment analysis of stock news headlines. Journal of Information and Telecommunication, 1-20.]**, graph based semi supervised learning and autoregressive conditional heteroscedasticity (GARCH) models**[Paramanik, R. N., & Singhal, V. (2020). Sentiment Analysis of Indian Stock Market Volatility. Procedia Computer Science, 176, 330-338.]** were used.

This survey paper is going to compare and discuss different methods of sentiment analysis in order to predict the stock market and share prices of a company. Firstly this survey paper tells about the importance and uses of sentiment analysis, what sentiment analysis can do, how it can help people and the company. Then we will be seeing the architecture of sentiment analysis for stock prediction. Different evaluation methods will be discussed. And later on future implementations to improve sentiment analysis and a conclusion regarding the same will be talked about.

The flow of the survey paper is as follows:

· We discuss the architecture of sentiment analysis for stock market prediction.

· Next we will discuss the evaluation methods which we encountered across different papers.

· Then we will compare the different methods used for sentiment analysis.

· We will discuss the conclusion and other things that can be used in future in order to improve the current implementation.

NEW TERMS

1. Sentiment Analysis: Sentiment analysis is also known as opinion mining and is used to determine whether the particular subject is positive, negative or neutral. Sentiment analysis is usually performed on textual data.

2. Investor Sentiment: Investor sentiment or market sentiment is the overall attitude in the financial world. In other words investor sentiment is the approximate value of the stock market at a given point of time.

3. NLP: Natural Language Processing commonly known as NLP is a sub branch of Artificial Intelligence which concerns manipulation of natural language like speech or text using software. In other terms NLP helps computers understand and process human language and analyze large amounts of natural language data.

4. Stock: Stock means equity of an organisation which is divided in several parts. Collectively stock means all shares of an organisation into which the ownership is divided.

5. Stock Market: Stock market means a place where sellers and buyers collectively sell and buy stocks of various companies and organisations in order to represent ownership.

6. Machine Learning: ML is a branch of Computer Science which concerns the study of various algorithms in order to improve the experience of computers and machines with the help of large samples of data.

ARCHITECTURE

The architecture which we are proposing is made using CoreML and its fundamentals. Here we are using a corpus which is provided with necessary tags. The corpus is provided to the coreML compiler which parses it and feeds it with necessary information.

This is how we are planning to create our Model. The core ML would parse each and every content in the table and feed it in the Model. Then after the model can be integrated to any frontend device.

CoreML Calculation

Tags	Point
Positive	+1
Negative	-1
Neutral	0

For eg: **"Great bid to go with"** in this statement Great is an positive statement remaining are statement compositions.

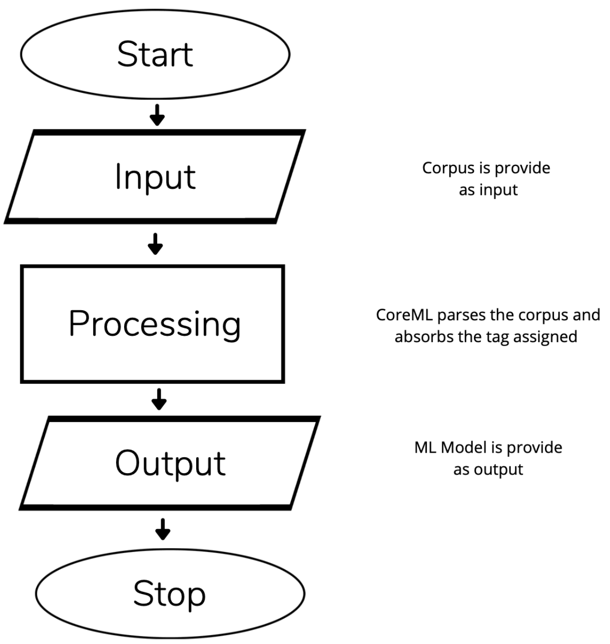
Thus, Total Score = +1, thus this statement is declared as **Positive**.

"Great bid but can make loss too" here in this statement Great is positive and loss is negative.

Thus, total score = +1 -1 = 0, thus this statement is a **Neutral** statement.

This is how the basic architecture detects the sentence.

Flow Chart



YOUR CONTRIBUTION

In our expedition of understanding the ‘sentimental analysis in stock prediction’ we have surveyed a number of papers. In the paper, ‘**The Impact of COVID-19 on the Chinese Stock Market: Sentimental or Substantial?**’ by the authors Sun Y., Wu M., Zeng. X., and Peng Z., they have investigated the impact on the Chinese Stock Market caused by COVID-19 by doing an event study and examining the effects of individual investor sentiment on their returns. This paper serves as a good basis to understand the sentimental analysis involved in stocks because during the pandemic not only was there a huge rollercoaster of stock prices but also a significant increase in the number of traders / investors in the market.

The COVID-19 pandemic, which began in early 2020, has caused financial market turmoil. Circuit breakers struck the American stock market twice in a week, and the situation in other countries was not much better. The majority of academics have noticed drops throughout the epidemic, but the causes are unknown. It is suspected that regions with a higher number of confirmed cases would suffer more significant losses. Naturally, that sector's profitability would be harmed, and its stock returns would suffer as a result, but this hypothesis was proven wrong. Their research demonstrates that this is not the case. The stock returns of Hubei businesses are identical to those of the market. Pharmaceutical stocks' high returns did not endure as long. This oddity supports the theory that stock market volatility during the COVID-19 outbreak was not solely due to economic loss.

In this paper it was found that, Individual investor emotion is positively connected with stock market returns during the outbreak, according to the findings. Stocks with high PB, PE, CMV, net asset, and institutional shareholder ratios, as well as extended listed years, are more susceptible to the epidemic. In consideration of the above stated reasons, they have used a Panel Regression Model. This model is a two-dimensional construct in which the same participants are observed multiple times across different time periods. Panel data is a blend of cross-sectional and time-series data in general. One observation of many objects and accompanying variables at a single point in time (i.e., an observation is taken once) is described as cross-sectional data. In their approach they have also made use of the Fama-French model. This model attempts to explain stock returns using three factors: (1) market risk, (2) small-cap outperformance against large-cap outperformance, and (3) high book-to-market value outperformance versus low book-to-market value outperformance. The advantages of the proposed work are that they have used panel regression implying that they have tested their data over a period of time to give a proper solution which is helpful for performing sentiment analysis.

In their approach the authors used two particular datasets, (1) was a stock-related financial data from the CSMAR database and another (2) was a sentiment data used in this work is GubaSenti, on which they applied three main equations to help find their result. (a) $R_{i,t} = \alpha + \gamma MKT_t + \delta SMB_t + \eta HML_t + \varepsilon_t$ This equation gives the ordinary least squares (OLS) regression where, $R_{i,t}$ represents the return of index i on date t in the estimation window, and MKT_i, SMB_t and HML_t are the three factors of the Fama–French model.
(b) $AR_t = R_t - [\hat{\alpha} + \hat{\gamma} MKT_t + \hat{\delta} SMB_t + \hat{\eta} HML_t]$, this equation gives abnormal returns where, R_t represents the actual return on date t in the event window.
(c) $CAR = \sum_{t=1}^{n} AR_t$, this equation gives individual abnormal returns to create a “cumulative abnormal return (CAR)”.

After experimenting with their approach, they came up with two results (1) The cumulative abnormal return in the event window is positive, indicating that the outbreak has a strong short-term positive impact on the stock price. The second result, regarding pharmaceutical stocks, reveals that the t-value is significantly positive, showing that the epidemic has a strong beneficial impact on pharmaceutical stock prices. And (2) that the findings show that mood can have a big impact on the overall market return during an epidemic. It also backs up the idea that the reverse effect is strong, implying that stock returns fell during the post-event period

The next paper we analyzed was '**Detecting a Risk Signal in Stock Investment Through Opinion Mining and Graph-Based Semi-Supervised Learning'** by the authors Yoon, B., Jeong, Y., & Kim, S. they have turned their head towards the problem as to how one can avoid credit events that might cause a national and global economic crisis which will ultimately lead to socio - economic losses. In this paper, they have shown how the majority of global economic crises are generated by a cascade of tiny events that have a potentially large impact. We can prevent a seriously damaging national or global crisis by recognizing and catching these tiny occurrences before they happen if we can notice and catch them before they happen. There has also been a surge in a number of hazy phenomena, such as cryptocurrency, a new financial service. The widespread practice of taxing imports to protect a country's native industries from outside competition leads to trade wars between countries and, in the long run, may hamper economic progress. All these factors can contribute to a credit event which is defined as an incident that seriously affects the bankruptcy risk of a company. Therefore, it is necessary to pre-determine the factors affecting the credit score so we can work towards monitoring and prevention of the same.

The study's main goal is to use opinion mining and graph-based semi-supervised learning to create an algorithm to aid in stock investment decision-making. Because of the massive increase in data in recent years, not only analysts and professionals, but also individual investors, may now acquire superior-quality financial and non-financial data about companies. This data can be an important source for detecting market moves. As a result, the goal of this work was to use opinion mining and machine learning to construct an algorithm to aid in stock investment decision-making using both objective and subjective information.

The three sections of the algorithm established in this study were as follows: (1) data gathering and filtering, (2) credit risk assessment and early warning signal identification, and (3) credit event prediction First, data was gathered from a variety of stock-related databases, ranging from news and financial statements to social networking sites and online communities. Author analysis and a rule-based method were used to filter bogus material, such as rumors and fake news. Second, sentiment analysis and opinion mining recognized a risk signal, which is an indicator or trigger of credit events such as bankruptcy and delisting. The risk signal was defined by three grades ('dangerous', 'warning', and `caution') in stock investment to provide insights for monitoring and responding to credit events in advance. Third, the likelihood of credit events happening was predicted using logistic regression, which included a binary dependent variable (occurring or not occurring) and independent factors based on signal detection findings.

The authors of this paper have proposed a novel algorithm to recognize risk signals and anticipate the future occurrence of credit events to aid in stock investing decision-making. Sentiment analysis based on opinion data, word2vec, and graph-based semi-supervised learning may be used to detect the danger signal by evaluating the sentiment value of data, including news and views. A logistic regression model comprising indicators based on the sentiment value of views then predicts the likelihood of credit occurrences. To create a logistic regression model for predicting future events, data from companies in the same industry, such as the sentiment value of views collected from surveys, were used. To do so they have performed sentimental analysis based on opinionated data which includes logistic regression model, linguistic rule- based model and a graph-based semi-supervised learning.

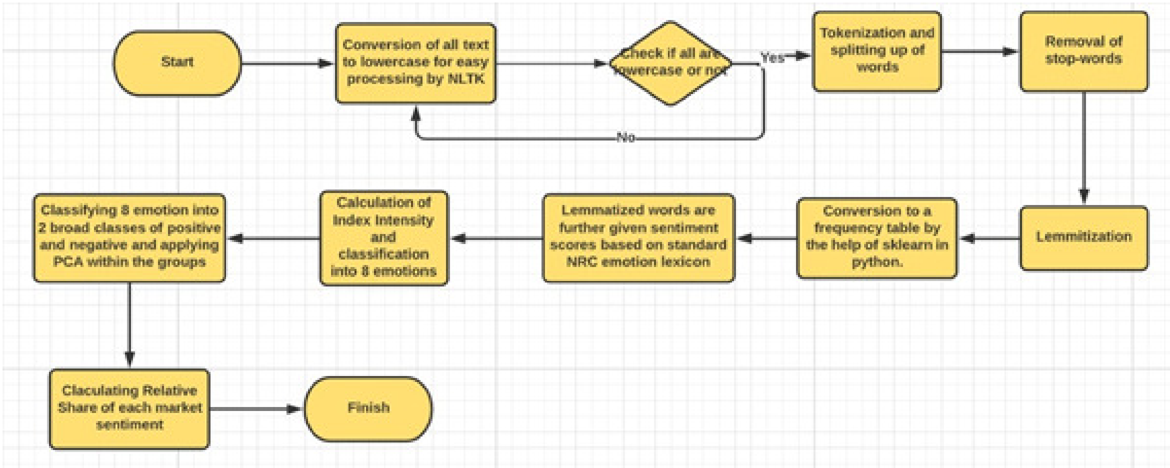
They have also utilized four main equations (a) , this equation provides us with the sentimental value of the keyword. (b) [$P_{i,j}$ = *(Sentimental value of core keyword i)], which gives the proximity index between keyword i and core keyword j, (c) [= (j=1,2,…m)], is an equation for the Euclidean distance between keyword i and core keyword j and (d) I= ,in which 'I' stands for Indicator for monitoring.

This paper suggests behaviour- and language-based approaches in sequence, which means that different characteristics of language in fake or genuine opinions are inspected by sentiment analysis after examining the current trend of opinion occurrences. The behaviour-based approach aims to identify the distribution of opinions by investors, while the language-based approach can pinpoint the pattern of opinions. To estimate the sentiment value of words, the document is preferentially rated, which is conducted in two ways: stock price and review score. The rating differs depending on whether the company is listed on a securities market or not. After deciding the sentiment of each article, this sentiment value is disseminated to all words included in the article through Naïve-Bayes classification, which is based on the co-occurrence of words in the article. Using the logistic regression equation, data of the target firm are put into the prediction model based on logistic regression. The probability of credit event occurrence is finally estimated. To validate the results derived from this prediction model, the authors developed a confusion matrix by comparing the actual number of incidences with the predicted number, which is higher than the cut-off probability and represented as 0 or 1 (binary).

In the paper, '**Sentiment Analysis of Indian Stock Market Volatility**', by the authors Paramanik, R. N., & Singhal, V., have presented the issue of scarcity of literature in the context of India's stock market volatility using investor's sentiment analysis. This paper attempts to shape the volatility of the Indian Stock market using investor sentiment analysis. The recent literature in behavioral finance has challenged the notion of a rational investor in the market since the emergence of noise traders in the market due to their cognitive errors and emotional exuberance. The transitory influence of noise traders were first believed to be eradicated but the argument of traditional financial theory is challenged by many researchers. Normally, sentiment is understood as the overall attitude of an investor's behavior but the influence of such market volatility has been fixed in literature by proxies of market analysis. There is scarcity of literature in the context of Indian stock market's volatility using sentiment analysis.

Three different generalized autoregressive conditional heteroscedasticity (GARCH) models are used to analyze impact of market sentiments. Emotional and sentiment indices are constructed by the help of NLP techniques. Initially the data is gathered and the headlines and summaries are gathered from the resources and NLTK (Natural Language Toolkit) is used to filter the data. In the first step we convert all the texts to lower case so that NLTK can process it easily. Next tokenization is done so that the words in the string. Next, Stop-words are removed from the string by the help of NLTK so that common words can be removed that do not generate any significance. In the further step, Lemmatization is done on various similar split words to get one common word as this helps us get better frequency insights. The pre-processed data is then converted to a frequency table using sklearn feature extractor of python. This table contains date wise frequency of each lemmatized word. These lemmatized words are further given sentiment scores based on standard NRC emotion lexicon (Emolex) consisting of 14,181 words with eight basic emotions (i.e., anger, fear, anticipation, trust, surprise, sadness, joy, and disgust). Now the mentioned 8 emotions are further classified into two broad categories which are positive and negative sentiments. Anger, fear, sadness and disgust are negative sentiments and others are positive sentiments. Principal component analysis (PCA) is employed to respective groups of emotions were derived factor loading is assigned as suitable weights for each emotion to construct scores for the two sentiments. At last, the relative share of each type of sentiment is calculated and relative dominance of two types of market sentiments is measured.

The approach used in this paper is known as NLTK with volatility modelling. The authors use the algorithm of Natural Language Toolkit (NLTK) and Principal Component Analysis (PCA) in order to complete the proposed work. Sklearn feature extractor is also used in order to complete the proposed work. The architecture of their work can be seen below



This paper attempts to augment the existing GJR GARCH model with previously proposed variables, share of positive market sentiment POS) and share of negative market sentiment (NEG) to assess more rigorously how these two contradictory sentiments shape the dynamics of conditional volatility using the augmented conditional volatility equation as shown in the equation section. When we analyse the finding it is very much clearly evident that the noise traders play a dominant role. An advantage of finding a real stream of data which is more useful in daily life. This stream of data is found to be more dynamic in nature than traditional monthly and quarterly indicators. Also, the given model does not consider positive and negative errors in the mean equation and instead of that the model has generated separate market sentiments on conditional volatility of the Indian financial market. This approach is better and more appealing in today's financial market.

In the paper, **Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation** the problem of stock portfolio allocation and to get maximum profit out of it was addressed.According to many financiers and investors, portfolio allocation is the most difficult and challenging task. Due to the stock market's complexity many people are unable to understand and analyze the stock and its future performance. Because the stock market is continuously changing, predicting the stock prices and its future return is not an easy task.

Authors of this paper **Koratamaddi, P., Wadhwani, K., Gupta, M., & Sanjeevi, S. G. (2021)** chose this problem as people are confused and find themselves in difficult situations when it comes to maintaining their stock portfolio.They have to maintain their stocks in their portfolio such that they not only get maximum returns but also with minimum risk involved. To help people from this situation the authors propose an algorithm that will analyze the news and tweets to help find a stock that can give maximum return with minimum risk. Stock prediction in previous years was very naive and was purely based on mathematical models that used quadratic programming, stochastic calculus, numerical analysis and other techniques. With the rise of supervised machine learning algorithms such as artificial neural networks became more prominent in predicting the values.The ability of neural networks to learn complex nonlinear functions is why they outperform other machine learning approaches in predicting market returns. Data set used in this paper was google news and twitter tweets. This dataset is useful because twitter is a social media platform where people can share their genuine opinion about the product, using that sentiment whether it is bad or good the company's stock price will be predicted. Google news is another platform where authentic news is shared and using that the predicted value will also be accurate.Few metrics were considered to evaluate the performance of the algorithms like, Sharpe ratio, Annualized returns and std error, Final portfolio Value and etc.

Following Equations were used to evaluate the stock values:

Equation 1: $$NS_{(c,d)} = \frac{\sum_{i=1}^{N} PS(a_{(i,c,d)})}{N}$$

Description: Sentiment analysis was performed to obtain a google news sentiment score.

Equation 2: $$TS_{(c,d)} = \frac{\sum_{i=1}^{N} w(t_{(i,c,d)})PS(t_{(i,c,d)})}{N}$$

Description: Sentiment analysis was performed to obtain a twitter sentiment score

Equation 3: $$CS_{(c,d)} = \frac{NS_{(c,d)} + TS_{(c,d)}}{2}$$

Description: The confidence score is computed using the financial news sentiment score (NS) and twitter sentiment score (TS)

In the paper, **Stock Prediction by using NLP and Deep Learning Approach** by **Deshmukh, R.** problem to properly recognize which shares to promote with a purpose to get more profits was addressed and solved.People have a tendency to analyze existing strategies and so planned new strategies for inventory prediction. They predict the share price based on random assumptions. People are not able to keep track of the stock prices whether it is going up or down. Financial analysts invest in stocks usually, but they are not aware about the inventory marketplace conduct. They usually go through the problem of trading as they do not properly recognize which shares to shop or which shares to promote with a purpose to get greater profits. In today's world, all the information pertaining to the inventory market is available. Maintaining and analyzing all these stocks data manually is a very challenging task.This manual process should be automated and should be beneficial for people to use it to get profit out of it. This is where Data mining techniques help. According to the authors' numerical time series offers close results when analysed , wise traders use system learning techniques in predicting the inventory market conduct. This will allow financial analysts to foresee the conduct of the inventory that they may be interested in and consequently act accordingly. In this paper ANN algorithm was used for the share market prediction. On the basis of the studies ANN had claimed that it can achieve a high percentage of accuracy while predicting the values in the stock market.

In **An Efficient Word Embedding and Deep Learning Based Model to Forecast the Direction of Stock Exchange Market Using Twitter and Financial News Sites: A Case of Istanbul Stock Exchange (BIST 100)** by **Kilimci, Z. H., & Duvar, R.** In this study, unlike the recent research on forecasting the stock market direction, they focus on financial sentiment analysis using the Turkish data sets collected from both a social media platform and websites including technical analysis and news to analyze the stock market direction by evaluating high volume stocks.In some previous researches the combination of technical and fundamental analysts approaches applied to market trend forecasting through the use of conventional machine learning techniques applied to time series prediction and sentiment analysis on the same data but the results were not as expected. In this paper long short-term memory networks, recurrent neural networks, convolutional neural networks as deep learning algorithms and Word2Vec, GloVe, and FastText as word embedding models are evaluated. To demonstrate the effectiveness of the proposed model, four different sources of Turkish news are collected. The news articles about stocks from Public Disclosure Platform (KAP), text-based technical analysis of each stock from Bigpara, user comments from both Twitter and Mynet Finans platforms are gathered.

The following equations are used for the research:

Equation 1: $$J_\theta = \frac{1}{T}\sum_{t=1}^{T}\sum_{-n\leq j\leq n} log\, p\left(w_{t+j}\middle| w_t\right)$$

Description: Given a center word and sequence of training words w1, w2, w3, . . ., w_t skip-gram model maximizes the average log probability of n surrounding words of the center word w_t, n denotes the size of training context.

Equation 2: $$J_\theta = \sum_{i,j=1}^{V} f\left(X_{ij}\right)\left(w_i^T w_j + b_i + b_j - logX_{ij}\right)^2$$

Description: Glove method first contracts a word co-occurrence matrix X. Each element of Xij shows the number of times word i appears in the context word j. The Glove model utilizes (3) to calculate cost.

Equation 3: $$s(w,c) = \sum_{g\in G_w} z_g^T v_c$$

Description: FastText uses the skip-gram model with negative sampling proposed for Word2Vec with a modified skip-gram loss function. Let Gw ⊃ {1, . . . , G} be the set of n-grams appearing in a word w, the score of the word is calculated by the sum of the vector representations of its n-grams:

In ***Prediction of stock values changes using sentiment analysis of stock news headlines*** paper we realised a few things. We did an in depth study which led us here.

Predictions and speculations regarding stock market values, particularly the valuations of global firms, are a fascinating and appealing topic. In this post, we look at how stock value varies and how to anticipate stock value changes using freshly scraped economic news about companies. We're concentrating on economic news headlines.To analyse the sentiment of the headlines, we use a variety of technologies. We use BERT as a baseline and compare the sentiment results to stock fluctuations over the same time period using three other tools: VADER, TextBlob, and a Recurrent Neural Network.In contrast to the other two instruments, the BERT and RNN were far more accurate; these tools were able to assess emotional values without neutral parts. We can determine the moment of the change in stock values by comparing these results to the movement of stock market values over the same time periods using sentiment analysis of economic news headlines.In contrast to the other two instruments, the BERT and RNN were far more accurate; these tools were able to assess emotional values without neutral parts. We can determine the moment of the change in stock values by comparing these results to the movement of stock market values over the same time periods using sentiment analysis of economic news headlines.

In the paper **A Method of Using News Sentiment for Stock Investment Strategy** , we went through the idea proposed by them and at the end we were clea with few points

In this study, we investigate whether news sentiment quantified by a sentiment analyzer utilizing deep learning can be utilized for investment strategies. Concretely, we aggregated sentiment scores by circadianly, hebdomadally(week), and monthly frequency and calculated the performance of the investment at that magnitude. The results show that the performance of the investment strategy is high at the quotidian frequency but not at all at the hebdomadal or monthly frequency. These results show that the impact of news sentiment on the stock market is very short, more than daily and less than hebdomadally. Several antecedent studies have substantiated that the impact of sentiment on the overall market is short in the Japanese market, but the results are homogeneous for individual stocks. This betokens that news is expeditiously woven into both the market as a whole and individual stocks.

In the paper **Augmented Textual Features-Based Stock Market Prediction** we were pretty clear with the logic and how we need to achieve it. The third paper helped us to make our prospect more clear and wider. The paper helped to understand following things:

Due to its dynamics, non-linearity, and complexity, the stock market is inherently difficult to predict. One of the attractive goals is to predict the direction of the stock market movement by using public opinion analysis. However, there is intense debate about the usefulness of this method and the strength of the causal relationship between stock market trends and sentiment. The researchers' opinions range from rejecting this relationship to confirming that there is a clear causal relationship between sentiment and stock market transactions. However, many advanced computing methods have adopted emotion-based features, but have not yet reached maturity and performance. In this article, they have used improved sentiment analysis methods to conduct empirical research on the predictability of stock market trends and make a constructive contribution to this debate. To be precise, they experimented with stock price history, sentiment polarity, subjectivity, Ngrams, text-based custom features, and feature lag for more detailed analysis. Five research questions have been investigated to answer questions related to the use of sentiment analysis to predict stock market trends.

In the paper - **Estimating the relative effects of raw material prices, sectoral outlook and market sentiment on stock prices** - the main objective was to determine the relation between movement of raw material prices and share price of the company.

We went through this paper and understood the relation between the raw material prices with the market sentiment. The problem addressed is also to ascertain the relative strength of the above mentioned factors depending on the time period. Important internal factor which affects the companies performance and stock prices is raw materials, the main aim is to understand the relation between stock prices and prices of the raw material of a particular company.

The particular problem is selected because the task of predicting share prices of a company requires strenuous efforts because of the sensitive financial market, external chaos, political influence, natural calamities and much more. The selected research paper contributes to relevant literature by combining or uniting wavelet analysis and machine learning to determine the relation between the movement of raw materials and share prices of a company, which affects the end consumer. The scope of the problem is limited to the following sectors : Oil and Gas, Metal, FMCG and Healthcare. The suggested methods and frameworks mentioned in the paper can be used to determine the required relation for other sectors as well.

As per the microeconomics rules, the prices of the raw material affects the end consumer, and this depends on the demand in the market and position of the company amongst their competitors. Increase in raw material prices also affects the sector’s outlook. Stock prices of the company reflects the market view of the current working of the organisations and furthermore their future performances. Stock prices of a company get affected in two ways: through sectoral outlook and and decrease in profitability. Moreover, the influence of raw materials over different time horizons has never been surveyed properly, this research paper helps to deal with this problem and contributes to literature by combining raw materials and share prices of a company.

The particular paper resorts to wavelet analysis and machine learning models to predict the relation between. Wavelet coherence and correlation analysis have been done to determine relations between raw material, sector outlook and market sentiment over a set of Indian companies for a short, medium and long period of time. Certain machine learning algorithms like Random forest, gradient boosting and genetic algorithms have also been used to determine the rank of the three factors mentioned above over different time periods. Below diagram shows the equation used in the following research paper :

Equation 1: CWT of a time series

$$W_x(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-\tau}{s}\right)dt$$

Description: Continuous wavelet transform based coherence analysis

Equation 2: MODWT estimator of wavelet correlation

$$\rho_{xy}(\lambda_j) = Corr\left(w_{ijt}, \widetilde{w}_{ijt}\right) = \frac{Cov\left(w_{ijt}, \widetilde{w}_{ijt}\right)}{\sqrt{Var(w_{ijt})Var\left(\widetilde{w}_{ijt}\right)}}$$

Description: The MODWT estimator of wavelet correlation has been used in this study which basically considers the covariance of two series ($x(t)$, $y(t)$) and wavelet variance of individual series.

In the paper - **Exploiting textual and relationship information for fine-grained financial sentiment analysis -** the main objective was to identify expressions (positive neutral negative) towards the subject by way of expressing sentiments in text. Novel approach to capture implicit sentiments and the contagion process.

We got to know about the problem of stock prediction. In a financial firm, sentiment could be read across platforms like company, analyst reports, news articles and blogs. The aim is to capture implicit sentiments and the contagion process. To apply the solution of sentiment analysis across multiple domains and text types, such as product reviews. To demonstrate the impact of implicit sentiment as well as importance of different relationship or sentiment prediction on company and analyst reports news articles and blogs.

This problem is chosen in order to showcase that textual context can be modelled on lithography, to study and gain further insight into sentiment analysis and improve it. Future scope of this paper includes exploring dynamic vertices like GraphSAGE. This will reduce re-calculation of vertex representation for the entire graph. Different classifiers can be used to further optimise the performance.

We also got some more insight into sentiment analysis, and found about sentiment contagion. . Implicit sentiment appears in a general rather than a source like a newspaper. Thereby detecting sentiments in text that do not clearly express emotions or do not contain such words are difficult to process. If we can improve sentiment analysis by historical context, this approach will analyse the interpretation of text and process it. The role of sentiment contagion is not considered. Sentiment contagion means individuals interacting with each other and capturing their sentiments to form an opinion about each other. Therefore by using machine learning models and other frameworks we can convert normal text and derive the sentiment expressed by the text to analyse whether the expression is positive, neutral or negative.

Proposed approach as per given in the paper includes text and graph FFNN (feed forward neural network) or a Fine Tuned Textual Representation FFNN (feed forward neural network).

Below diagram shows the equation used in the following research paper :

Equation 1: Equation to calculate mean squared error (MSE) :

$$MSE(P,T) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(T_i - P_i)^2}$$

Description: mean square error measures the average of the squares of the error.
Equation 2: Equation to calculate Cosine Similarity(CS)

$$CS(T,P) = \frac{\sum_{i=1}^{n} T_i \times P_i}{\sqrt{\sum_{i=1}^{n} T_i^2} \times \sqrt{\sum_{i=1}^{n} P_i^2}}.$$

Description : cosine similarity is the cosine of the angle between to N-dimensional vectors in an N-dimensional space.

In the paper – **Harvesting social media sentiment analysis to enhance stock market prediction using deep learning** – the main problem addressed was to identify how movements in a company's stock prices correlate with expressed opinions of the public regarding that company.

This problem was selected because the current stock market is affected by social mood and historical prices, and also by people's sentiment which play a major role in moving the prices of stock prices of different companies. The use of news articles, social media like Twitter, Facebook explains how a company performs in the share market. This will help the common user to predict the stock market and invest wisely and get a good return in the long term. We can get stock accuracy so that the user can buy or sell stock of a particular company. There are future opportunities for research in this area.

And after thorough reading of this paper we came to know that Opinions of others play a very important role in an individual person's decision making. People who use social media sites like Facebook or Twitter can communicate with other people and tell their opinion on a specific topic like a news, company, sports and many more. This data can be usefully extracted to predict the future of a particular object or a topic, this method can also be used to predict a company's stock prices and is known as sentiment analysis. The analysis of sentiment and emotions, is the study of opinions, thoughts, experiences, feelings and behaviours in the text form. The entire market is dependent on sentiments of people and the society and sentiment analysis can be used to predict the share market.

And the main objective of this paper was to make a stock price prediction tool which considers public sentiment and also other parameters. Data will be gathered from social networking sites like Twitter, Facebook, Google plus etc. Social networking sites perfectly reflect People's opinion on a particular company or a particular news. It is found from a survey that financial news has an impact on stock prices of a particular company.

So after all the observation and ideation we have tried something new by gaining knowledge from all the papers. While going through the papers we noticed that the main issue was that all the methods were quite hard to implement and not only that the given model was tough to implement. When it comes to any NLP model at last it's all about how we can implement them on various devices. And the above proposed models were really tough to implement. With this search we did our research about how to make this procedure easy to use as well as easy to implement.

During our research we went through various ML model production tools, various coding styles , etc. At the end we came across this wonderful tool called CoreML. This is a proprietary tool provided by Apple to develop all your NLP oriented development along with a few extra features.

The basic function of CoreML is to design a ML model for you according to the need. You need to provide the input accordingly and the model will be generated. The best part is using those models is the easiest one. Integrating them on projects is really easy.

With our project we created an NLP model which identifies the sentence emotion and accordingly predicts the stock price. We have used twitter API to fetch tweets which belong to various companies and using the latest tweets we determine the emotion of people towards a company. This helps us to understand and predict the stock price fluctuations. Proposed method in the paper was using the Deep Learning Model and LSTM(Long short term memory) a reliable predictive model for stock movement. LSTM is a form of RNN and is likely to learn long-term dependencies. LSTM allows RNN to keep track of their input data over a long time. Below diagram shows the equation used in the following research paper :

Precision	Recall	F-score	Accuracy
∑Ŧp ∑Ŧp +∑Fp	∑Ŧp ∑Tp + ∑Fn	2∑ Ŧp/(2∑ Ŧp + ∑FP + ∑FN)	∑ Ŧp + ∑Tn ∑Ŧp + ∑Tn + ∑Fp + ∑Fn

Description: the above formula returns a value between -1 and +1, this value is also called as the sentiment of a particular company and depending on the sentiment value whether it is negative or positive the analysis is made about a particular company or a topic.

Equation 2: Various LSTM formulas were used :

$$\mathrm{FT} = \sigma(\mathrm{Wf} * [\mathrm{ht}-1, \mathrm{xt}] + \mathrm{Bf})$$

$$\Delta\mathrm{Ct} = \tanh(\mathrm{WC} \cdot [\mathrm{ht}-1,\ \mathrm{xt}] + \mathrm{bC})$$

$$\mathrm{it} = \sigma(\mathrm{Wi} \cdot [\mathrm{ht}-1,\ \mathrm{xt}] + \mathrm{bi})$$

$$\mathrm{Ct} = \mathrm{ft} * \mathrm{Ct} - 1 + \mathrm{it} * \Delta\mathrm{Ct}$$

$$\ddot{\mathrm{O}}\mathrm{T} = \Sigma(\mathrm{Wo} \cdot [\mathrm{ht}-1,\ \mathrm{xt}] + \mathrm{Bo})$$

EVALUATION METHOD

Most common evaluation metrics:

Accuracy

Whenever the accuracy metric is used, we aim to learn the closeness of a measured value to a known value. It's therefore typically used in instances where the output variable is categorical or discrete — Namely a classification task.

Precision

In instances where we are concerned with how exact the model's predictions are we would use Precision. The precision metric would inform us of the number of labels that are actually labeled as positive in correspondence to the instances that the classifier labeled as positive.

Recall

Recall measures how well the model can recall the positive class (i.e. the number of positive labels that the model identified as positive).

F1 Score

Precision and Recall are complementary metrics that have an inverse relationship. If both are of interest to us then we'd use the F1 score to combine precision and recall into a single metric.

Mean Reciprocal Rank (MRR)

The Mean Reciprocal Rank (MRR) evaluates the responses retrieved, in correspondence to a query, given their probability of correctness. This evaluation metric is typically used in informational retrieval tasks quite often.

Mean Average Precision (MAP)

Similar to MRR, the Mean Average Precision (MAP) calculates the mean precision across each retrieved result. It's also used heavily in information retrieval tasks for ranked retrieval results.

Root Mean Squared Error (RMSE)

When the predicted outcome is a real value then we use the RMSE. This is typically used in conjunction with MAPE — which we will cover next — in the case of regression problems, from tasks such as temperature prediction to stock market price prediction.

Mean Absolute Percentage Error (MAPE)

The MAPE is the average absolute percentage error for each data point when the predicted outcome is continuous. Therefore, we use it to test evaluate the performance of a regression model.

Area Under the Curve (AUC)

The AUC helps us quantify our model's ability to separate the classes by capturing the count of positive predictions which are correct against the count of positive predictions that are incorrect at different thresholds.

Datasets used

- BSE(Bombay stock exchange) and NSE(National Stock Exchange) websites - all the information related to the stock prices of the companies have been taken from bse and nse websites, which are the best resources for such data as they update it regularly and are the most reliable sources for stock prices.
- IMDB dataset & Yelp dataset - Yelp Data set contains information about eight metropolitan areas in the USA and Canada. IMDB data set contains data over 25,000 reviews labelled according to the sentiment (Positive or negative).

- News from Money control, Google News and Economic Times - News sites like money control, Economic Times, IFL are trusted sources and they have enough information about stock related stuff.. Twitter is the one of largest social media to house tweets related to stock market and share prices of numerous companies.

- Twitter and other social media platforms - Here customers share their genuine opinion about the product and using those as the input can be really helpful in detecting the stock and portfolio future prices.

- NASDAQ Stock Price and TOPIX500 Stock Price

- Stock-related financial data are from the CSMAR database

- Sentiment data used in this work is GubaSenti

The most common methodologies that are used in our problem domain are:

- Mean Squared error (avg - 0.0693)
- Cosine Similarity (avg - 0.79945)
- Sentiment Value (positive sentiment value 0.01355 , negative sentiment value -0.0063)
- Sharpe Ratio
- Annualized return comparisons
- Annualized standard errors
- Final portfolio value vs predicted value

Approach	Sentiment-Aware ADDPG	Adaptive DDPG	DDPG	Mean Variance	Min Variance
Sharpe Ratio	**2.07**	1.49	0.93	1.25	0.99
Annualized Return (%)	**22.05**	18.84	14.7	15.86	11.48
Annualized Std. Error	**0.096**	0.116	0.147	0.127	0.116
Final Portfolio Value (USD)	**25,051**	21881	18156	19632	16333

- F-criterion and accuracy
- Q values of Equal Weighted and Market value Weighted
- Dickey-fuller test to check stationarity.

COMPARISON OF BASE PAPERS

Paper Titles:

1. Prediction of stock values changes using sentiment analysis of stock news headlines

2. A Method of Using News Sentiment for Stock Investment Strategy

3. Augmented Textual Features-Based Stock Market Prediction

Parameters	**Comparison**

Paper Title	**1**	**2**	**3**
Datasets	IMDB Review Dataset	TOPIX500 Stock Price	NASDQ Stock Price
Performance	Average	Average	Best
Scope	Movie Review prediction	TOPIX500 Stock price prediction	NASDQ stock price prediction
Application	Used to predict movie reviews	Used to predict stock prices of Japan	Used to predict stock prices of US
Algorithm	RNN	LSTM	DNN, Deep CNN, LSTM
Model Type	Supervised	Supervised	Supervised
Measures	Rating Analysis	Graph Analysis	Graph Analysis
Evaluation	NLTK, Text Blob	Market Value Comparison	Tweet Mining, Granger Causality test
Output Type	Ratings	Stock Price	Stock Price

So in the above mentioned table we have compared three research papers on various parameters.

1. Datasets: Every paper uses datasets which are required to achieve their evaluation domain
2. Performance: Performance of 1 & 2 paper is less as compared to 3 since the third paper is using more complex algorithms with more structured queries .
3. Scope: All papers have their specific scope, the first one provides the prediction of ratings for movies, second paper helps in stock prediction of TOPIX500 stocks. The third paper helps to predict NASDQ listed companies stock prices.
4. Application: Every paper has their own application, depending upon their needs.
5. Model Type: In this all of them are supervised models. In the start the model learn and absorbs the respective domain through supervised learning.
6. Measures: Measurements help us to compare respective models. The first one uses ratings of previous movies for measurement. Similarly in the second and third stock prices are taken for measurement.
7. Evaluation: Every model has their own evaluation model, which shows their precision.
8. Output: The output produced by each model depends on their domain and prediction.

Paper Titles:

1. **The Impact of COVID-19 on the Chinese Stock Market: Sentimental or Substantial?**
2. **Detecting a Risk Signal in Stock Investment Through Opinion Mining and Graph-Based Semi-Supervised Learning**
3. **Sentiment Analysis of Indian Stock Market Volatility**

Parameters	**Comparison**		
Paper Title	**1**	**2**	**3**
Datasets	Stock related financial data from CSMAR database and Sentiment data used is GubaSenti	Hyundai Merchant Marine	web sources for Indian financial market and business:Reuters India Livemint business news, The Hindu Business Line and Moneycontrol.com.
Performance	Average	Average	Better
Scope	Chinese market Sentiment	Hyundai merchant	Indian Stock volatility
Approach	Panel Regression Model	Sentiment analysis based on opinion data	NLTK approach with Volatility Modelling
Algorithm	Panel Regression Model and Fama-French Model	a. Logistic regression b.Linguistic rule based c.Graph based semi-supervised	algorithm of Natural Language Toolkit(NLTK) and Principal Component Analysis(PCA)
Model Type	Supervised	Supervised	Unsupervised
Future Work	comparative analysis of different sectors on Indian stock to have a better understanding of the dynamics in the marketing sector	process of screening false information should be expanded in many ways	the function of (food) industry influences in epidemics.

So in the above mentioned table we have compared three research papers on various parameters.

1. Datasets: Every paper uses datasets which are required to achieve their evaluation domain
2. Performance: All the research papers use complex algorithms for stock price prediction and the accuracy of all the paper's models are good.
3. Scope: All papers have their specific scope.
4. Approach: Every paper has its own method to approach the problem which is mentioned here.
5. Algorithm: The algorithm involved in the three papers is compared in this row. Each has its own unique algorithm.
6. Model Type: The papers follow a variety of model types like supervised neural networks etc.
7. Future work: in this column it is shown what according to the author of those papers thinks is future work in this paper or field.

Paper Titles:

1. **Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation**
2. **Stock Prediction by using NLP and Deep Learning Approach**
3. **An Efficient Word Embedding and Deep Learning Based Model to Forecast the Direction of Stock Exchange Market Using Twitter and Financial News Sites: A Case of Istanbul Stock Exchange (BIST 100)**

Parameters	**Comparison**		
Paper Title	**1**	**2**	**3**
Datasets	Google News, Twitter	Twitter Tweets, top 5 performing companies' data in IT sector	Twitter, KAP, Mynet Finans, Bigpara
Performance	Good	Average	Good
Scope	Stock Prediction using sentiments on Google news and twitter	Stock prediction using twitter tweets and company data	Stock Price Prediction for Istanbul stock Exchange
Application	Used to arrange the stocks and predict the portfolio value	To recognize which shares to promote to get more profit	Used to predict stock prices of US
Algorithm	DDPG	CNN	CNN, RNN, Word

			embedding models
Model Type	deep reinforcement learning	Neural Network	Supervised
Measures	Sentiment Analysis	Data Analysis	Graph Analysis
Evaluation	Sharpe Ratio Annualized Return Annualized Std. Error Final Portfolio Value	Market Value Comparison	F-criterion and accuracy
Output Type	Predicted value of stock price	Stock Price value	Stock price values for Istanbul stock exchange

So in the above mentioned table we have compared three research papers on various parameters.

8. Datasets: Every paper uses datasets which are required to achieve their evaluation domain
9. Performance: All the research papers use complex algorithms for stock price prediction and the accuracy of all the paper's models are good.
10. Scope: All papers have their specific scope. All three papers predict the prices of the stock values but on different datasets with different metrics for evaluations.
11. Application: Every paper has their own application, depending upon their needs.
12. Model Type: The papers follow a variety of model types like supervised neural networks etc.
13. Measures: Measurements help us to compare respective models.
14. Evaluation: Every model has their own evaluation model, which shows their precision.
15. Output: The output produced by each model depends on their domain and prediction.

CONCLUSION AND FUTURE DIRECTIONS

With this all papers we found that every paper has their own way of prediction. With our strategy we plan to make it more easy and convenient for developers to use it. Many of the papers just focused on the headlines of the news to get the prediction. What we have planned is to provide the whole article as an input thus we can make our prediction more stronger and reliable.

With this we have planned some new features for our project. We are planning to make the set more unique by providing an additional feature of predicting the occurrence of the same kind of scenario with companies which fall in the same categories. For eg: we know how the market fell in 1992, during the Harshad Mehta Scam, at that time banking sector stocks saw a huge crash So if we are keeping a track on the falls in the market on a daily basis there are chances that we can predict the market even before days to the fall. For this we need to take care of three things

1. Company Sector
2. Fall percent
3. Time and date

Using these three things we can track the market and predict the future of stocks. The main reason to bring this feature is because we saw in the research papers there was huge uncertainty when we predicted the stocks only on a daily basis. Thus having a track on the market on a daily basis with a precise note can help us to make this prediction more powerful.

REFERENCE SECTION

1) Rao, T., & Srivastava, S. (2012). Analyzing stock market movements using twitter sentiment analysis
2) Chen, R., & Lazer, M. (2013). Sentiment analysis of twitter feeds for the prediction of stock market movement. *stanford edu Retrieved January*, *25*, 2013.
3) Patel, A. P., Patel, A. V., Butani, S. G., & Sawant, P. B. (2017). Literature Survey on Sentiment Analysis of Twitter Data using Machine Learning Approaches. *IJIRST-International journal for Innovative Reasearch in Scinece & Technology, 3(10).*
4) Varghese, R., & Jayasree, M. (2013). A survey on sentiment analysis and opinion mining. *International journal of Research in engineering and technology*, *2*(11), 312-317.
5) Giachanou, A., & Crestani, F. (2016). Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys (CSUR)*, *49*(2), 1-41
6) https://www.dailyfx.com/education/understanding-the-stock-market/stock-market-sentiment-analysis.html
7) Sun, Y., Wu, M., Zeng, X., & Peng, Z. (2021). The impact of COVID-19 on the Chinese stock market: Sentimental or substantial? *Finance Research Letters*, *38*, 101838.
8) Yoon, B., Jeong, Y., & Kim, S. (2020). Detecting a Risk Signal in Stock Investment Through Opinion Mining and Graph-Based Semi-Supervised Learning. *IEEE Access*, *8*, 161943-161957.
9) Paramanik, R. N., & Singhal, V. (2020). Sentiment Analysis of Indian Stock Market Volatility. *Procedia Computer Science*, *176*, 330-338.
10) Al-Awadhi, A. M., Alsaifi, K., Al-Awadhi, A., & Alhammadi, S. (2020). Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns. Journal of behavioral and experimental finance, 27, 100326.
11) Ashraf, B. N. (2020). Stock markets' reaction to COVID-19: Cases or fatalities?. *Research in International Business and Finance*, *54*, 101249.
12) Shan, L., & Gong, S. X. (2012). Investor sentiment and stock returns: Wenchuan Earthquake. *Finance Research Letters*, *9*(1), 36-47
13) Baig, A. S., Butt, H. A., Haroon, O., & Rizvi, S. A. R. (2021). Deaths, panic, lockdowns and US equity markets: The case of COVID-19 pandemic. *Finance research letters*, *38*, 101701.
14) Ichev, R., & Marinč, M. (2018). Stock prices and geographic proximity of information: Evidence from the Ebola outbreak. International Review of Financial Analysis, 56, 153-166.
15) https://www.gtarsc.com/
16) https://ifind.bnu.edu.cn/
17) Chapelle, O., Schölkopf, B., & Zien, A. (2006). A discussion of semi-supervised learning and transduction. In *Semi-supervised learning* (pp. 473-478). MIT Press.
18) Hussain, A., & Cambria, E. (2018). Semi-supervised learning for big social data analysis. Neurocomputing, 275, 1662-1673.
19) Yu, D., Chen, N., Jiang, F., Fu, B., & Qin, A. (2017). Constrained NMF-based semi-supervised learning for social media spammer detection. Knowledge-Based Systems, 125, 64-73.

20) Rout, J. K., Dalmia, A., Choo, K. K. R., Bakshi, S., & Jena, S. K. (2017). Revisiting semi-supervised learning for online deceptive review detection. IEEE access, 5, 1319-1327.
21) Lašek, I., & Vojtáš, P. (2011). Semantic information filtering-beyond collaborative filtering. In 4th International Semantic Search Workshop. Retrieved from: http://km. aifb. kit. edu/ws/semsearch11/11. pdf.
22) Moseki, K. K., & Madhava Rao, K. S. (2017). Analysing stock market data—Market sentiment approach and its measures. Cogent Economics & Finance, 5(1), 1367147.
23) Rao, K. M., & Ramachandran, A. (2014). Exchange rate market sentiment analysis of major global currencies. *Open Journal of Statistics*, *2014*.
24) Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. Journal of computational science, 2(1), 1-8.
25) Atkins, A., Niranjan, M., & Gerding, E. (2018). Financial news predicts stock market volatility better than close price. The Journal of Finance and Data Science, 4(2), 120-137.
26) Kuzmina, J. (2010). Emotion's component of expectations in financial decision making. Baltic Journal of Management.
27) Billah, M., Waheed, S., & Hanifa, A. (2016, December). Stock market prediction using an improved training algorithm of neural network. In 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) (pp. 1-4). IEEE.
28) Das, S., Behera, R. K., & Rath, S. K. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. *Procedia computer science*, *132*, 956-964.
29) Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.
30) Khedr, A. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. *International Journal of Intelligent Systems and Applications*, *9*(7), 22.
31) Billah, M., Waheed, S., & Hanifa, A. (2016, December). Stock market prediction using an improved training algorithm of neural network. In 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) (pp. 1-4). IEEE.
32) Laney, D. (2001). 3D Data management: Controlling data volume, velocity and variety. Meta Group. *Lakshen, Guma Abdulkhader*, 1-4.
33) Suwabe, T. (2015). Data Revolution and Equity Investment. *Securities Analysts Journal*, *53*(4), 6-17.
34) Galbraith, J. W., & Tkacz, G. (2018). Nowcasting with payments system data. *International Journal of Forecasting*, *34*(2), 366-376.
35) Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. *The Journal of Finance*, *67*(1), 1-43.
36) Katayama, D., & Tsuda, K. (2020). A Method of Using News Sentiment for Stock Investment Strategy. *Procedia Computer Science*, *176*, 1971-1980.
37) Yoo, P. D., Kim, M. H., & Jan, T. (2005, November). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06) (Vol. 2, pp. 835-841). IEEE.
38) Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, *89*, 14-46.
39) Bollen, J., & Mao, H. (2011). Twitter mood as a stock market predictor. *Computer*, *44*(10), 91-94.

40) Zembik, M. (2014). Social media as a source of knowledge for customers and enterprises. *Online Journal of Applied Knowledge Management*, *2*(2), 132-148.
41) Bollen, J., Mao, H., & Pepe, A. (2011, July). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In *Proceedings of the international AAAI conference on web and social media* (Vol. 5, No. 1).
42) Koratamaddi, P., Wadhwani, K., Gupta, M., & Sanjeevi, S. G. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. Engineering Science and Technology, an International Journal, 24(4), 848-859.
43) Koyano, S., & Ikeda, K. (2017, November). Online portfolio selection based on the posts of winners and losers in stock microblogs. In 2017 IEEE Symposium Series on Computational Intelligence (SSCI) (pp. 1-4). IEEE.
44) Abe, M., & Nakayama, H. (2018, June). Deep learning for forecasting stock returns in the cross-section. In PacificAsia conference on knowledge discovery and data mining (pp. 273-284). Springer, Cham.
45) Bao, W., & Liu, X. Y. (2019). Multi-agent deep reinforcement learning for liquidation strategy analysis. arXiv preprint arXiv:1906.11046.
46) Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Intelligent asset allocation via market sentiment views. ieee ComputatioNal iNtelligeNCe magaziNe, 13(4), 25-34.
47) Yu, P., Lee, J. S., Kulyatin, I., Shi, Z., & Dasgupta, S. (2019). Model-based deep reinforcement learning for dynamic portfolio optimization. arXiv preprint arXiv:1901.08740.
48) Deshmukh, R. (2021). Stock Prediction by using NLP and Deep Learning Approach. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(1S), 202-211
49) Batra, R., & Daudpota, S. M. (2018, March). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-5). IEEE.
50) Wang, Y., & Wang, Y. (2016, March). Using social media mining technology to assist in price prediction of stock market. In 2016 IEEE International conference on big data analysis (ICBDA) (pp. 1-4). IEEE.
51) Zhang, Z., Shen, Y., Zhang, G., Song, Y., & Zhu, Y. (2017, November). Short-term prediction for opening price of stock market based on self-adapting variant PSO-Elman neural network. In 2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS) (pp. 225-228). IEEE
52) Firdaus, M., Pratiwi, S. E., Kowanda, D., & Kowanda, A. (2018, October). Literature review on artificial neural networks techniques application for stock market prediction and as decision support tools. In 2018 Third International Conference on Informatics and Computing (ICIC) (pp. 1-4). IEEE
53) Sharma, N., & Juneja, A. (2017, April). Combining of random forest estimates using LSboost for stock market index prediction. In 2017 2nd International conference for convergence in technology (I2CT) (pp. 1199-1202). IEEE
54) Kilimci, Z. H., & Duvar, R. (2020). An Efficient Word Embedding and Deep Learning Based Model to Forecast the Direction of Stock Exchange Market Using Twitter and Financial News Sites: A Case of Istanbul Stock Exchange (BIST 100). IEEE Access, 8, 188186-188198
55) Tekin, S., & Çanakoğlu, E. (2019, April). Analysis of price models in istanbul stock exchange. In 2019 27th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE.
56) PERVAN, N., & KELEŞ, H. Y. (2017). Sentiment analysis using a random forest classifier on turkish web comments. Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering, 59(2), 69-79.
57) Kilimci, Z. H. (2020). Financial sentiment analysis with Deep Ensemble Models (DEMs) for stock market prediction. Journal of the Faculty of Engineering and Architecture of Gazi University, 35(2), 635-650.
58) Othan, D., Kilimci, Z. H., & Uysal, M. (2019, December). Financial sentiment analysis for predicting direction of stocks using bidirectional encoder representations from transformers (BERT) and deep learning models. In Proc. Int. Conf. Innov. Intell. Technol. (Vol. 2019, pp. 30-35).

59) Tekin, S., & Çanakoğlu, E. (2018, May). Prediction of stock returns in Istanbul stock exchange using machine learning methods. In 2018 26th Signal Processing and Communications Applications Conference (SIU) (pp. 1-4). IEEE
60) Ghosh, I., Chaudhuri, T. D., Alfaro-Cortés, E., Martínez, M. G., & Rubio, N. G. (2021). Estimating the relative effects of raw material prices, sectoral outlook and market sentiment on stock prices. Resources Policy, 73, 102158
61) Sen, J., & Chaudhuri, T. D. (2018). Understanding the sectors of Indian economy for portfolio choice. International Journal of Business Forecasting and Marketing Intelligence, 4(2), 178-222.
62) JHUNJHUNWALA, A., CHAUDHURI, T. D., & BHAMRAH, G. K. (2019). Value creation by Indian companies: A comparative study over two time periods. Turkish Economic Review, 6(1), 44-61.
63) Jhunjhunwala, A., & Chaudhuri, T. D. (2021). Innovation, growth and value creation: a study of Indian companies. International Journal of Business Innovation and Research, 25(3), 328-352.
64) Jareño, F., de la O González, M., Tolentino, M., & Sierra, K. (2020). Bitcoin and gold price returns: a quantile regression and NARDL analysis. Resources Policy, 67, 101666.
65) Fama, E. F., & French, K. R. (2021). The cross-section of expected stock returns (pp. 349-391). University of Chicago Press.
66) Daudert, T. (2021). Exploiting textual and relationship information for fine-grained financial sentiment analysis. Knowledge-Based Systems, 107389.
67) Daudert, T. (2021). A multi-source entity-level sentiment corpus for the financial domain: the FinLin corpus. Language Resources and Evaluation, 1-24.
68) Sinha, N. R. (2016). Underreaction to news in the US stock market. Quarterly Journal of Finance, 6(02), 1650005
69) Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Representation learning on graphs: Methods and applications. arXiv preprint arXiv:1709.05584.
70) Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 701-710).
71) Gori, M., Monfardini, G., & Scarselli, F. (2005, July). A new model for learning in graph domains. In Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005. (Vol. 2, pp. 729-734). IEEE.
72) Mehta, P., Pandya, S., & Kotecha, K. (2021). Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Computer Science, 7, e476.
73) Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies, practices and applications. International Journal of Scientific and Technology Research, 9(2), 601-609.
74) Milosevic, N. (2016). Equity forecast: Predicting long term stock price movement using machine learning. arXiv preprint arXiv:1603.00751.
75) Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2020). An innovative neural network approach for stock market prediction. The Journal of Supercomputing, 76(3), 2098-2118.
76) Das, S., Behera, R. K., & Rath, S. K. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. Procedia computer science, 132, 956-964.
77) Sun, A., Lachanski, M., & Fabozzi, F. J. (2016). Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. International Review of Financial Analysis, 48, 272-281.