



# VIT<sup>®</sup>

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

**SCHOOL OF COMPUTER SCIENCE ENGINEERING**

**STUDY OF RECENT ADVANCES IN SENTIMENT ANALYSIS FOR  
STOCK PREDICTION**

**Team 7**

**HARSH VIVEK LONDHEKAR (19BCE0496)**

**DEVANG GUPTA (19BCE0959)**

**HEMAKSH CHATURVEDI (19BCE2222)**

**GOKUL R NAIR (19BCE2245)**

**CSE 4022 - NATURAL LANGUAGE PROCESSING**

**FALL SEMESTER 2021-22**

**SUBMITTED TO**

**FACULTY: Dr. SARAVANAKUMAR K**

**SLOT : A1**

# Study of Recent Advances in Sentiment Analysis for Stock Prediction

Dr. Saravanakumar Kandasamy<sup>1,2,3</sup> . Harsh Vivek Londhekar<sup>2,3</sup> . Devang Gupta<sup>2,3</sup> . Hemaksh Chaturvedi<sup>2,3</sup> . Gokul R Nair<sup>2,3</sup>

## Abstract

Predictions and speculations regarding stock market values, particularly the values of global corporations, 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 will be using various kinds of NLP techniques to find the sentiment of a news headline. With use of these techniques we will be generating sets of results. We can determine the timing of the change in stock prices by comparing these results to the movement of stock market values over the same time periods using sentiment analysis of economic news headlines. The correlation matrices also revealed a substantial difference between the three models in terms of the influence of emotional values on the change in the stock market's value.

**Keywords** Deep Learning · Neural Networks, Augmented Textual Features, Sentimental Context, BERT, NLPTK, VADER, Big Data, financial sentiment analysis, stock market prediction, word embedding methods, Price forecasting, sentiment analysis, CNN, SVM, ANN, NSE, BSE, financial market, Data mining, Reinforcement learning Portfolio allocation, Stock prices, Sectoral outlook, Market sentiment, Wavelet coherence and correlation, Machine learning, contextual sentiment analysis, sentiment contagion, implicit sentiments, volatility, investor sentiment, stock return

## 1. Introduction

### 1.1 Problem definition

The rational behavior of investors and market fundamentals are regarded to be the only factors that shape investors' decisions in the stock market, according to fundamental theories of finance such as Markowitz's (1952) mean variance analysis [6]. According to them the whole stock market can be predicted using the Mean Variance theory. Mean Variance theory works on pure mathematics, but when it comes to stock prediction there is one more factor called sentimental context. Every trader has a sentiment analysis before buying a stock. The opening and closing price of every stock depends on the sentiment of the trader towards the stock.

Stock market is a highly volatile place where a stable stay can never be experienced [1]. Thus to bear with the volatility we need to predict the stock. Thus here comes Sentiment Analysis. As discussed above, a trader has hundreds of sentiments before buying a stock. No trader buys a stock without a motive and when we talk about this sentiment it's all derived through news headlines, tweets, paper headlines etc. Thus these headlines and tweets create a huge impact on stock price [2]. People tend to believe what they see, and with news and paper headlines being in favor of a company makes people buy their stocks and vice versa. Thus these headlines get woven with the companies without physical existence.

From these given points we would be defining how sentiment analysis plays a vital role in prediction of stocks rather than the old technique of Mean variance by the end of the paper.

---

✉ Harsh Vivek Londhekar  
harshvivek.londhekar2019@vitstudent.ac.in

✉ Devang Gupta  
devang.gupta2019@vitstudent.ac.in

✉ Hemaksh Chaturvedi  
hemakshamit.chaturvedi2019@vitstudent.ac.in

✉ Gokul R Nair  
gokul.nair2019@vitstudent.ac.in

✉ Dr. Saravanakumar Kandasamy  
ksaravanakumar@vit.ac.in

1. Department of Software Systems

2. School of Computer Science and Engineering

3. VIT, Vellore, India

## 1.2 Impact

With the prospect of what we are believing and trying to point out we believe the stock market prediction can be improved by 60% [3]. Relying on the old theory of Mean Variance and applying the same theory on every stock is not an ideal solution. Every stock has its own state of volatility and non linearity. Considering them all the same and applying the same theory may lead to false results. Thus we need some solution which can be unique for every stock and predict it with higher precision.

Sentiment analysis works on headlines denoted for individual stock or company. With focus in particular stock we can derive results which are highly precise and accurate. Thus adopting the method of sentiment analysis can create a huge impact on stock prediction.

## 1.3 History of Problem

History of the problem started when people considered stocks only as a market of mathematics. It's not all about mathematics in stock prediction. The early fundamental theories of finance believe in the theory of Mean Variance. Fundamental finance believes the stock market can be completely predicted using Mean Variance theory. Applying the same theory on every Stock can provide them an accurate result. But it's not true, with every stock purchase every trader has their own sentimental context behind it. The sentimental context can be due to any reason.

This application of fundamental theory on every stock is the root cause of irregularity in stock prediction. We need a firm process which can get the best precision possible for predicting stock. Not only this, people generally believe in past day prediction, which means predicting stock on the performance given on previous days. This theory is also wrong, we can't and could never judge a stock with just past performance of a couple of days. We need a firm set of data to conclude with this theory.

These are the root causes, why we need a new technique to predict the stock market.

## 1.4 Different approaches taken by researchers

The different approaches which were used in the papers were wavelet analysis, different machine learning models like random forest, gradient boosting genetic algorithm [7], Fine tuned textual representation [8], deep learning models like LSTM (Long Short Term Memory) [2][3][9], adaptive sentiment-aware deep deterministic policy gradients approach [10], convolution neural networks (CNN) [3][11][12], recurrent neural networks [12], deep learning algorithms and Word2Vec, GloVe, and FastText. In other papers sentiment analysis was done using BERT techniques [1], graph based semi supervised learning and autoregressive conditional heteroscedasticity (GARCH) models [6] were used. Recurrent neural networks (RNN) was also seen to be used in some of the papers [1][2][9][12]. Apart from these certain models like NLTK Vader Lexicon, Text Blob [1], Panel regression model, Fama French Model [4], Logistic regression model, Linguistic rule-based model, graph-based semi-supervised learning [5], Principal Component Analysis (PCA), SKlearn feature extractor [6], were seen to be used. And lastly Deterministic Policy Gradient, or DPG and DDPG [10] were also used.

## 1.5 Purpose of this paper

In This survey paper we are going to compare and discuss different methods of sentiment analysis in order to predict the stock market and share prices of a company. We will be exploring the most recent trends in the field of stock prediction and discuss the improvements and advancements in performance and development of models that improve the prediction of stock prices for maximum profit.

The main purpose of this paper is to compare different papers and their algorithms on the problem "stock prediction using sentiment analysis NLP" and then to come up with the best algorithm among all these algorithms. For this we collected the recent research and also used different existing research to compare the methodology in the architecture. Our main research involved analysing the social media data and news data to predict the stock value. Our main objective was to solve stock portfolio allocation and maintain it to get maximum return with minimum risk involved [10] and to analyze stock value changes and predictions of the stock values using fresh scraped economic news about the companies [1].

We analysed Japanese news and the impact of COVID 19 to come up with good investment strategies in various stock exchanges [2][4]. Opinion Mining and graph based Sentiment analysis were the common approaches to predict the stock values [5].

## 1.6 Document layout

The flow of the survey paper is as follows:

- Firstly this survey paper gives an introduction about this paper. Then we discuss the importance and uses of sentiment analysis, what sentiment analysis can do, how it can help people and the company to predict the stock values for maximum profit involving minimum risk..
- Then, we discuss the architecture of sentiment analysis for stock market prediction.
- Then, we will discuss the evaluation methods which we encountered across different papers.
- Then, section 4 details the popular evaluation methods and datasets used in the base papers.
- A comparative analysis of the base papers has been included in section 5.
- Section 6 contains the conclusion and future work.

## 2. Definition of Important of Terms

1. Stock Prices : Stock means equity of an organisation which is divided in several parts. Stock price means the current price of the stock that is trading for in the market.
2. Sectoral Outlook : It means the trend of different industries in the financial sector.
3. Market Sentiment : Market sentiment means the overall trend or the behaviour of investors towards a particular market in the financial market.
4. Wavelet Coherence and correlation : Wavelet coherence refers to the measure of the correlation between two signals. It is used to measure linear interactions.
5. 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.
6. Contextual sentiment analysis : It is related to gathering user data and analyzing it to form an opinion on a particular topic, which will be either positive or negative or neutral.
7. Sentiment contagion : sentiment contagion or emotion contagion refers to the process of observing behavior changes in an individual, which results in a similar behaviour in other individuals.
8. Implicit sentiments : Implicit sentiments refer to the expressions which are related to an idea without explicitly stating them, and they convey factual information which leads to a positive, negative or neutral response towards that idea.
9. Data mining: The practice of analysing large databases in order to generate new information.
10. Portfolio allocation: spreading the investments across various asset classes. Broadly speaking, that means a mix of stocks, bonds, and cash or money.
11. Sentiment analysis: contextual mining of text which identifies and extracts subjective information in source material, and helps a business to understand the social sentiment of their brand, product or service while monitoring online conversations.
12. Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.
13. A Convolutional Neural Network (CNN) is a Deep Learning algorithm which can take in an input image, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other.
14. NSE and BSE are stock exchanges of India. National Stock Exchange and Bombay Stock Exchange.
15. Volatility is the degree of variation of a trading price series over time, usually measured by the standard deviation of logarithmic returns
16. Stock return is the total return for a stock includes both capital gains/losses and dividend income, while the nominal return for a stock only depicts its price change

## 3. Architecture and analysis

The method that every model uses to find sentiments is called the Scoring technique. In this method tags are provided with a score. With the score generated for a sentence the sentiment is predicted. Thus for every sentiment analysis to work on a sentence a scoring system is to be established, which scores a word according to the Language Processing model. The tagging technique which we use is shown below

Tags	Point
Positive	+1
Negative	-1
Neutral	0

For better understanding let's take an example and understand how the whole scoring technique works.

For eg:

1. **“Great bid to go with”** in this statement Great is a positive statement remaining are statement compositions.

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

2. **“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.

3. **“Will lose price”** in this statement lose is a negative word and the remaining are statement compositions

Thus the Total Score = -1, thus the statement is declared as **Negative**

This is how the basic architecture detects the sentence.

This way every sentence is parsed and provided with a score, and at the end if the score is Positive then there are high chances of increase in price of that stock and if negative then high chance of fall in price.

### 3.1 Prediction of stock values changes using sentiment analysis of stock news headlines [1]

#### a) Aim

Stock Prediction is one of the sectors where no one can achieve 100% efficiency. When it comes to prediction of stock it's really tough since dynamics of stock market change in seconds. With such volatility it's really hard to predict it. But what we can do is try our best to get the best possible precision. Prediction of stock is directly dependent on the moves taken by companies and trader nature towards the company. Thus predicting it with company moves is not possible. So we will be predicting it in accordance to the trader/public nature towards the company. And the best source to get the nature of people towards companies is through news headlines.

#### b) Methodology and Results

Stock value predictions based on newly collected economic news regarding firms. With an emphasis on recent fresh news and utilising techniques such as BERT as a baseline, compare the findings with three additional tools: VADER, TextBlob, and a Recurrent Neural Network, and compare the sentiment results to stock fluctuations over the same time period.

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 timing of the change in stock prices by comparing these results to the movement of stock market values over the same time periods using sentiment analysis of economic news headlines. The correlation matrices also revealed a substantial difference between the three models in terms of the influence of emotional values on the change in the stock market's value [18].

c) Advantages

With the proposed conclusion we came to know that there is no neutral emotion in sentiment analysis. This is because every sentiment caused an impact on the market. Thus while classifying the contextual texts we can eliminate the neutral emotion. This will help to improve the precision of prediction. Since now we have only two emotions to take care of. Thus every headline would give a direct result of positive or negative. This study eliminated the existence of a third emotion which directly helped in increasing the precision of the result [13].

d) Disadvantages

Since this analysis was done on a small corpus it is not that evident if this proposed change will help in increasing the precision of result. Testing it on a wide variety of corpora and with a variety of techniques can only conclude what is the change in result derived. Thus considering it in such an early stage is not a good move. We need to have further experiments which can prove the existence of this conclusion. By the time we can only consider it as a prospect and avoid it from getting generalised.

### **3.2 A Method of Using News Sentiment for Stock Investment Strategy**

a) Aim

The aim over here is to predict daily, weekly as well as monthly stock movement Stock exchanges. The study focuses on getting a universal method to predict stocks for desired period of time, which practically is impossible. As discussed the movement of stock price is directly dependent on company moves and nature of traders towards the company. So with such volatile proportionality we can't predict the stock for more than 2-3 days of time. So in order to check if it's possible or not this study was done.

b) Methodology and Results

The methodology is quite simple over here. The system uses a set of deep learning techniques to identify stock prices. Once the stock price list is obtained according to sentiment analysis then the obtained result is plotted on a graph. Now we will be comparing the graphs generated by sentiment analysis and economical graphs obtained by daily market value.

Thus we will be comparing and finding the common points, regular patterns etc to prove sentiment analysis is as good as economical prediction and most of the times better than too.

c) Advantages

This process makes the process really easy to find. With the usual process we need to read each line and give a verdict. But with this every line is parsed and a score is generated along with its sentimental context. This method can be applied on any model and it is really versatile to use. The method can be applied and adopted by any corpora and predicts a result with the same precision [19].

d) Disadvantages

The issue lies over here is the whole method was found to be short adopting. Like the effect of this method on any company will be short lived. This method is best suitable for daily trading, when it comes to long term trading or mid term this method seems to be a false case. The reason is very simple since the volatility of the market is dependent on the company moves and trader nature. Thus for short trading this method seems best case.

Other than that the method we adopted provides three emotional contexts(i.e Positive, Negative, Neutral), but with our study we found every news creates an impact on Stock price, so considering a stock headline as neutral and declaring it to have no impact on the stock price is wrong.

### **3.3 Augmented Textual Features-Based Stock Market Prediction**

a) Aim

There is a high volatility, non-linearity and complexity in stock market prediction. It's inherently difficult to predict the Stock market due to its volatility. With the explosive non linearity predicting a stock seems really tough and unreal. So the aim is to predict the stock price using consumer behaviour and nature. Analysis of customer/trader behaviour towards a

company can help to identify how well a stock can perform. Taking the top 10 companies of NASDAQ will help to generate more precise results, since those companies are always in headlines. Consuming data from such companies is really helpful for experimenting.

#### b) Methodology and Results

They presented a strategy for enriching the representation of feelings that begins with extracting several text-based elements. It then use a variety of feature selection algorithms to dynamically choose the right feature sets for various situations, before stacking separate models to obtain the best of base stock direction classifiers [13].

Different machine learning algorithms and feature selection approaches performed differently for different stocks in the empirical study, which would not be the case if the stock market followed a random pattern. We infer that a company's stock price rises and falls in response to public sentiment or emotions. Only more complex methods of sentiment research are required to forecast stock market direction [23].

#### c) Advantages

The advantage of this method is it's easy to adopt on other corpus too. With its empirical investigation process things get really sorted out and finding a genuine prediction is really easy and precise. Along with that it was observed certain stock prices followed a regular pattern. Thus getting such patterns will help to make prediction easy for the next time, since those stocks have a regular pattern. Such patterns are hard to find if we go on with the random charts [14][15].

#### d) Disadvantages

As mentioned, getting a pattern is really helpful but recurring of the same pattern in future is doubtful. So trusting such a pattern without solid proof may lead to wrong judgement. Thus before adopting a pattern we need to know the recurring of such patterns again is negligible sometimes [16]-[18].

### 3.4 The impact of COVID-19 on the Chinese stock market: Sentimental or substantial?

#### a) Aim

Most researchers have observed plummets during the pandemic, but the reasons remain unclear [24]. The goal of this article is to determine if COVID-19's impact on the Chinese market was driven by stock owners' sentiments or by more substantial factors. The market was seen to have high volatility as a result of the pandemic. According to the efficient market hypothesis, a rational explanation for the volatility would be significant economic loss. If this stays true, the region with the most confirmed cases will incur the most significant losses. Naturally, that sector's profitability would be harmed, and its stock returns would suffer as a result. [4]. But it was observed that the rational reasoning did not stand and no such abnormal change was observed.

This paper goes ahead to show how sentiment contributes to the volatility of the stock market. They put forth 2 main hypotheses to test the contribution of sentiment:-

- The event that leads to strong negative sentiment, such as panic and anxiety. Previous studies argued that public health hazards such as SARS and Ebola can affect market sentiment. In the case of COVID-19, Liu, H. found that the virus outbreak had raised investors' fear of uncertainty [25]. Baig, A. S., Butt, H. A., Haroon, O., & Rizvi, S. A. R. found that the overall sentiment declined during the pandemic [26]
- The event caused lower yields on related stocks than usual. They said that media coverage of pandemics had an impact on the stock prices of companies closer to the origin area and in the pharmaceutical industry.

#### b) Methodology and Results

The authors broke down their methodology into main parts:-

- Event Study
- Panel regression

They have first calculated the abnormal returns of the stock market during the pandemic and conduct a significance test by doing an event study. Then, they went on to explore whether sentiment is explanatory to abnormal returns by regression.

The event study was done so as to identify abnormal returns in the stock market due to the outbreak of COVID-19. To do this they have divided their data on the basis of time to create event windows which can be seen in **Fig 3.4**. They have used the Fama- French Model to derive their expected returns and cumulative abnormal returns.

The result of the event study can be observed in 3 divided segments as follows: -

- After the event day => stock returns and individual sentiment both react negatively
- During event window => the standard deviation increases, leading to stock market yield decreasing and increase in volatility
- Post event window => the return and investor sentiment both rise, higher than normal before the pandemic outbreak
- Reversal effect was also observed



**Fig 3.4: Set up of the Event Study**

Due to its capacity to extract changes from panel data and minimise estimation bias, panel regression is better at capturing the time varying relationship between dependent and independent variables than event studies. As a result, panel regression is employed to control heterogeneity during sentiment effect calculation. [4]. Feasible generalised least squares(FGLS) was also done to ensure robustness of results.

The result of panel regression observed was as follows: -

- results show that the overall market return was significantly affected by sentiment
- Reversal effect was also observed meaning the stock returns depreciated during post-event window

#### c) Advantages

The advantages of the proposed work are that they have used event study and 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.

#### d) Disadvantages

This paper is mainly focused on only 7 main industries so this cannot be used as a basis for other industries.

### 3.5 Detecting a Risk Signal in Stock Investment Through Opinion Mining and Graph-Based Semi-Supervised Learning

#### a) Aim

A global economic crisis arose from the subprime mortgage crisis of 2008, resulting in corporate bankruptcies or delistings from securities markets amid a shrinking national economy [27]. A potentially catastrophic national or global disaster can be avoided if these little instances can be noticed and treated in advance.



The goal of this study is to use opinion mining and graph-based semi-supervised learning to create an algorithm to aid in stock investment decision-making [5] to avoid credit events that might cause a national and global economic crisis ultimately leading to socioeconomic losses.

The sentiment analysis used in this paper analyzes emotions expressed by people as a subfield of opinion mining and has recently been used in many fields of application, including the financial field [28] while, Semi-supervised learning has considerably better efficiency and accuracy than the abovementioned techniques since the number of independent variables is sufficient while the number of dependent variables is insufficient (a general situation) [29]

In order to complete their objective they have divided the processes involved as follows: -

- filtering fake information,
- assessing credit risk and detecting risk signals,
- predicting future occurrences of credit events through sentiment analysis, word2vec, and graph-based semi-supervised learning [5]

#### b) Methodology and Results

In order to evaluate the core processes following methods were used: -

- Filtering of fake information was done with the help of author analysis and a rule-based approach
- Credit risk was assessed with the help of opinion mining and sentiment analysis
- A signal for a credit event was then detected by the degree of assessed risk
- Predicting future occurrences of credit events was done based on the risk signal using logistic regression.

When investors are concerned about trading in a company's stock, a risk signal is characterised as a warning that they should pay attention to the company's status and management issues. This study introduces a novel approach for recognising risk signals and predicting the occurrence of credit events in the future, which can help with stock investing decision-making.

This paper uses naive-Bayes classification, word2vec, and graph-based semi-supervised learning to propagate the sentiment value of core keywords to relevant words after allocating sentiment value for each text. All textual content is preprocessed by natural language processing once it has been filtered for fraudulent information in order to determine credit risk. Logistic regression is used to forecast the chance of a credit event occurring after detecting the sign of a credit event.. The logistic regression prediction model is made up of sub-indices for assessing risk in the first phase.

The regression line made it possible to predict the occurrence of credit events by using events occurring in recent times. The credit event's projected occurrence was then compared to the actual occurrence. If the probability calculated using a regression line based on significant factors is greater than the threshold value, credit occurrences are more likely to occur.

#### c) Advantages

Because the stock market is influenced by investor information, and there is a lot of it regardless of honesty, our article deals with incorrect information from the perspective of data processing. As a result, we provided an algorithm for early detection of risk indications based on a large amount of opinion data. Our work expanded the availability of social data in the finance industry by giving a mechanism for when investors decide whether or not to exchange stocks. Situational learning and classification of words or texts are aided by semi-supervised graph-based learning. Graph-based semi-supervised learning is used to smooth out the other data and make it consistent with the labelled data.

#### d) Disadvantages

Despite the fact that each company is required to provide information about its financial health and significant changes in operations and management, some companies may conceal their unfavourable position. As a result, certain events are not visible or are concealed, making it harder to discover these hidden occurrences, which might result in a significant loss.

### 3.6 Sentiment Analysis of Indian Stock Market Volatility

#### a) Aim

The goal of this research is to demonstrate how standard empirical models (which analyse the impact of sentiments on financial market volatility using financial indicators or macroeconomic fundamentals) may be used to predict future market volatility. Further, the paper aims at proposing an augmented version of asymmetric GARCH model of conditional volatility for Indian stock exchange [6].

Fundamental theories of finance like mean variance analysis of Markowitz (1952) posits the rational behaviour of investors and market fundamental is believed to be the sole factor that plays a decisive role in shaping the decision of such investors [30]. During their study of literatures they found that, unlike rational investors, noise traders impact market return and volatility due to their cognitive errors and emotional exuberance [31]. This paper aims to enhance the statement, 'Sentiment is conceptualized as the overall attitude of an investor towards the market or any specific stock and this is independent of market fundamentals' by Atoniou [32]

#### b) Methodology and Results

The study employs basic macroeconomic or financial factors as sentiment components to examine their impact on market conditional volatility, which is a novel approach to modelling conditional volatility. To quantify different forms of attitudes in the Indian market, they used a novel approach of leveraging news driven sentiment research and augmenting volatility models with such sentiment factors.

Here they have implemented a GARCH model which is a statistical model that may be used to examine a variety of financial data, such as macroeconomic data. This model is commonly used by financial organisations to estimate the volatility of stock, bond, and market indices returns.

They have used NLP to find sentiment over their data following the steps in **Fig 3.6**.

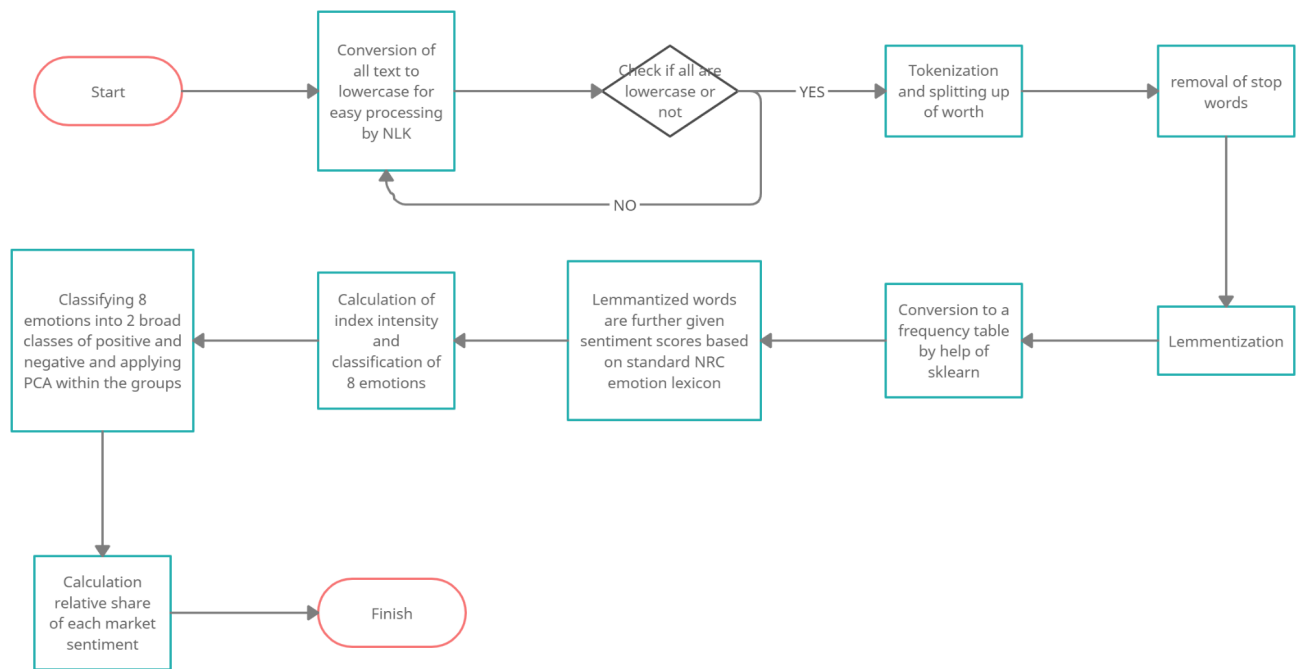
Empirical findings suggest dominant impact of negative market sentiment over positive one and it also provides evidence of noise trading in financially immature Indian stock market.

#### c) Advantages

When examining the findings, it is clear that the noise traders have a significant impact. Finding an actual stream of data has the advantage of being more useful in everyday life. Traditional monthly and quarterly measures are found to be less dynamic than this stream of data. 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.

#### d) Disadvantages

The suggested work's first flaw is that it only uses eight emotions, which are divided into two categories. The number of emotions should be raised, and a new broader class called neutral might be created to help with sentiment analysis. The study's future scope includes a comparison of diverse sectors of the Indian stock market, such as energy, telecommunications, and metal.



**Fig 3.6 Methodology used in paper ‘Sentiment Analysis of Indian Stock Market Volatility’**

### 3.7 Estimating the relative effects of raw material prices, sectoral outlook and market sentiment on stock prices [7]

#### a) Aim

To decide the effect of raw fabric charges in different time periods, outlook of different sectors in the market and market sentiment on the share charges of the organisation/company/corporation. 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.

#### b) Methodology and Result

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 relation 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.

The algorithms used in the particular paper were :

Random Forest : It is a machine studying algorithm used to solve regression and type troubles.

Gradient Boosting : It is based on the Greedy algorithm and can overfit a dataset quickly. It penalises various parts of the algorithm and improves the performance by reducing overfitting.

Genetic Algorithm : It is a technique for solving confined and unconstrained optimization issues that are primarily based on the herbal selection technique.

Formula used :

1. CWT of a time series

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

Description: Continuous wavelet transform primarily based coherence analysis

## 2. MODWT estimator of wavelet correlation

$$\rho_{xy}(\lambda_j) = \text{Corr}(w_{ijt}, \overline{w_{ijt}}) = \frac{\text{Cov}(w_{ijt}, \overline{w_{ijt}})}{\sqrt{\text{Var}(w_{ijt})\text{Var}(\overline{w_{ijt}})}}$$

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

A related problem was found in another paper also. It is stated that different sectors in an economy behave differently depending upon their trend pattern, characteristic of the sector and also some uncertainties that lead to randomness. The objective is to find out for certain sectoral characteristics of the stock market over different time periods [33]. A similar trend of comparing Indian companies was also seen in another paper. The objective was to assess the innovation of Indian companies and to analyse the changes due to these innovations. Three types of innovations are taken into consideration : Product, marketing and process innovation. It additionally assesses the strategies components by way of these businesses [34]. Regarding attention of different time periods, one of the research papers analysed the change in cost of Bitcoin returns, alternate in Gold charge returns, US Stock Market returns, hobby prices, crude oil and fast change in American Stock Market. To make the outcomes more robust , two different time periods have been taken into consideration [35].

### c) Advantages

1. One of the principal advantages of wavelets is they offer a simultaneous localisation in time and frequency domain.
2. The use of large DWT foundation features or wavelets filters produces blurring and ringing noise near part regions in pictures or video frames.
3. The second foremost benefit of wavelets is that, the usage of a fast wavelet remodel, it is computationally very speedy.
4. Provide a way for analysing web forms in both frequency and duration.
5. Representation of features that have discontinuities and sharp photos.
6. Accurately deconstructing and reconstructing finite, non-periodic and non-stationary signals.
7. Allow alerts to be saved extra efficaciously than by means of Fourier remodel.

### d) Disadvantages

1. A terrible directional selectivity for diagonal capabilities, because the wavelet filters are separable and real.
2. Longer compression time.
3. And lack of shift invariance, in because of this that small shifts inside the input signal can purpose.
4. fundamental versions within the distribution of electricity among DWT coefficients at specific scales.
5. The value of computing DWT in comparison to DCT may be higher.

## 3.8 Exploiting textual and relationship information for fine-grained financial sentiment analysis [8]

### a) Aim

Identification of expressions (positive neutral negative) towards the subject by way of expressing sentiments in text. Novel approach to seize implicit sentiments and the contagion procedure.

In a monetary firm, sentiment might be examine throughout structures like company, analyst reports, information articles and blogs. The goal is to seize implicit sentiments and the contagion method. To observe the answer of sentiment evaluation across more than one domain names and text sorts, inclusive of product opinions. And to demonstrate the effect of implicit sentiment in addition to significance of various relationship or sentiment prediction on organization and analyst reviews, news articles and blogs.

### b) Methodology and Result

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).

The paper uses aFFNN for its calculations, FFNN is based on LMBP algorithm, LMBP uses all the sample information to modify the weight and threshold and it can adjust or modify weight and threshold of the model.

The improvement from baseline became visible from zero. Sixty eight - 15.1% and 59. Forty six-234.15% for the CS and MSE, respectively.

A contrast of excellent tuned BERT and quality tuned FFNN, showed that FFNN outperforms high-quality tuned BERT. Fine tuned FFNN had excellent performance as compared to the opposite fashions.

Equation used :

1. 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.

2. Cosine Similarity(CS)

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

Description : Cosine similarity is the cosine of the perspective among N-dimensional vectors in an N-dimensional space.

One of the papers used the following datasets to meet their objectives : IMDB corpus, Sentiment stanford sentiment treebank, Sentiment140, SemEval2017 Task 4, SemEval2017 Task 5, The SSIX Corpora, FiQA 2018 Challenge, [36]. The objective of this paper is to form a novel corpus which contains various reports, Company reports, articles from the newspaper and micro-blogs from StockTwits and to analyse the entire corpus in order to determine the sentiment of a company. To foster on the financial sentiment analysis and capability application in behavioural science [36]. Another paper resorted to content material of news articles to form a relation among reaction of the marketplace and news articles [37]. One of the papers worked with graphs and processed graphs of different sorts like directed, and directed, labelled and biking graphs [38]. On a comparable basis, one paper worked on locating a manner to represent and encode graph structure in order that it may be without problems used by machine mastering fashions. Currently system studying models rely upon person defined heuristics to extract the one-of-a-kind functions and encode structural facts [42].

#### c) Advantages

1. Problems in FFNN are represented by way of attribute-value pairs.
2. These gaining knowledge of strategies are pretty robust to noise within the education facts. The education examples may additionally incorporate errors, which do not affect the very last output.
3. It is used wherein the quick evaluation of the found out goal function is required.

#### d) Disadvantages

1. It is highly dependent on hardware.
2. Lack of assurance of proper network structure.
3. The difficulty to show the problem to the network.

### 3.9 Harvesting social media sentiment analysis to enhance stock market prediction using deep learning [9]

#### a) Aim

To become aware of how actions in an employer's inventory costs correlate with expressed evaluations of the public regarding that employer. And to make a stock charge prediction device which considers public sentiment and also different parameters. Data may be collected from social networking websites like Twitter, Facebook, Google plus and so forth. 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.

#### b) Methodology and Result

Using Deep Learning Model and LSTM(Long quick time period reminiscence) a reliable predictive model for inventory movement is constructed. LSTM is a shape of RNN and is able to analyze lengthy-time period dependencies. LSTM allows RNN to maintain music of their entered facts over a long time. Stock Values of companies were taken from NSE stock data, Sentiment value was used as a metric to compare. Sentiment value lies between -1 and +1 , and depending on this value the different companies stock value can be determined whether it will increase or decrease.

The average sentiment estimates the normal sentiment of any topic over a given length. The experiments done considered opinions, primary sentiment, precision and recall. Stock prediction using different social media platforms resulted in more accuracy and reliability than previous predictions that were made.

Similar Machine studying models and deep studying fashions have been extensively utilized that allows you to expect inventory price of an agency over a long period of time using device getting to know fashions [39]. LSTM is also used to expand an modern neural community approach to acquire higher inventory market predictions. To advise the deep long brief term reminiscence neural community LSTM with embedded layer and long brief time period memory neural network to be expecting the inventory market [40]. Making use of user generated blogs is also a good method to predict the stock prices [41]. This is a potential use of textual information.

#### c) Advantages

1. LSTM provides various parameters like learning rates, and input and output biases.
2. There is no need for fine adjustments.
3. Complexity is reduced to  $O(1)$  with LSTM.

#### d) Disadvantages

1. LSTMs get influenced by various irregular weight introductions and consequently act very like that of a feed-forward neural net. They incline toward little weight introducing all things being equal.
2. LSTMs are inclined to overfitting and it's miles hard to apply the dropout calculation to check this trouble. Dropout is a regularisation technique where enter and intermittent institutions with LSTM gadgets are probabilistically prevented from initiation and weight refreshes while preparing an agency.

### **3.10 Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation [10]**

#### a) Aim

To solve stock portfolio allocation and maintain it to get maximum return with minimum risk involved.

#### b) Methodology and Results

Portfolio allocation is solved using an adaptive, sentiment-aware deep deterministic policy gradient that not only learns from historical stock price movements, but also from market sentiment — an influential environment input that captures investors' overall mood. They've compiled a large collection of Google News and Twitter tweets that represent the mood of the big firms. It also offered the methodology and mathematical definitions for calculating market sentiment, allowing the adaptive DDPG algorithm to take full advantage of it. Different methodologies were employed in other studies to handle the financial challenge of portfolio selection. One of the publications used sentiment analysis to generate a situation of winners and losers in the selection of stock microblogs.[43]. Deep learning is also used in some cases for forecasting the stock returns in different stock markets like the Japanese stock market etc and the performance of methods is investigated [44].Other papers found a suitable modelling system that can handle the complexities of the stock market and generate strategies that can help in practical trading[45]. As a result, the majority of papers in this domain investigated the role of market sentiment in an asset allocation problem, and the main conclusion was to develop Dynamic portfolio optimization.[46].

$$\begin{aligned}\text{News Sentiment Score} \quad \text{NS}_{(c,d)} &= \frac{\sum_{i=1}^N PS(a(i,c,d))}{N} \\ \text{Twitter Sentiment Score} \quad \text{TS}_{(c,d)} &= \frac{\sum_{i=1}^N TS(t_{(i,c,d)})PS(t_{(i,c,d)})}{N} \\ \text{Confidence Score} \quad \text{CS}_{(c,d)} &= \frac{NS(c,d) + TS(c,d)}{2}\end{aligned}$$

When compared to the baselines, the sentiment aware strategy offers considerable gains across all of the studied criteria. With a \$10,000 beginning investment, our approach's eventual portfolio value is \$25,051, which is significantly greater than Adaptive DDPG's 21,881 dollars and DDPG's 18,156 dollars. This strategy generates a 22.05 percent annualised rate of return, compared to 18.85 percent and 14.7 percent for Adaptive and DDPG, respectively. This model likewise takes less risk than the other baselines, as seen by the fact that it has the lowest annualised standard deviation error of 0.096. Our sentiment-aware strategy is demonstrated to be more robust and effective in balancing return and risk with a Sharpe ratio of 2.07, compared to 1.49 and 0.93 for the adaptive DDPG, respectively.

c) Advantages

A model-free off-policy approach for learning continuous actions is Deep Deterministic Policy Gradient (DDPG). It combines DPG (Deterministic Policy Gradient) and DQN (Deep Q-Network) principles. It is based on DPG, which can operate over continuous action spaces, and it uses Experience Replay and slow-learning target networks from DQN.

d) Disadvantages

Deep Deterministic Policy Gradient (DDPG) algorithms may handle exceedingly complicated issues in environments with continuous state and action spaces, but they can also fail in seemingly small and easy environments, for reasons that are currently unknown.

### 3.11 Stock Prediction by using NLP and Deep Learning Approach [11]

a) Aim

Identifying the stocks that will perform best in future and will give maximum profits using a deep learning approach.

b) Methodology and Results

In this paper a system was created using computer science techniques such as Natural Language Processing (NLP) and Deep Learning's Convolutional Neural Network (CNN). Natural Language Processing (NLP) is being used to help a system find organisations that have outstanding news in terms of market performance in real time. This simplified the process of selecting the best performers on the market. NLP is used to categorise news into good and negative sets, as well as to generate a performance graph for specific firms. We learned about the most successful business. Natural Language Processing offers an NLP (Natural Language Processing) technique for detecting merchandise and its effect on our tweets. The other related papers utilized a single system to determine the stock market's opening value. However, their built system was self-learner, allowing them to estimate the market's opening value. They had fed the stock data into their custom-built algorithm, which calculated the anticipated value[48]. Some papers also tried the ANN algorithm approach to predict the stock market prices and also discovered proper findings[49].

CNN is commonly employed in feature engineering because it has the property of focusing on the most evident elements in the line of sight. In comparison to alternative approaches, such as the ANN algorithm, CNN performs significantly better. Using 11 HRV features, the suggested CNN achieved a prediction accuracy of 84.6 percent, whereas the ANN, SVM, and KNN algorithms produced prediction accuracies of 73.5 percent, 67.9%, and 65.9%, respectively. Other research used a sentiment analysis technique to extract stock-related tweets from various social networking sources, while others used a sentiment analysis technique for stock tweets that were related to a different sort of product. They used the SVM technique to determine the polarity of tweets[47].

c) Advantages

A fully linked CNN learns an appropriate kernel, resulting in a filtered image that is less template-based. In comparison to a CNN, a fully linked network with one hidden layer exhibits fewer symptoms of being template-based. CNN is commonly employed in feature engineering because it has the property of focusing on the most evident elements in the line of sight. In comparison to other approaches, such as the ANN algorithm, CNN outperforms other approaches in terms of stock price prediction.

d) Disadvantages

The position and orientation of objects are not encoded by CNN. Inability to be spatially invariant when dealing with incoming data. It is necessary to collect a large amount of training data. When CNN-LSTM is applied, it is capable of providing the most accurate stock price forecasts.

### **3.12 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) [12]**

a) Aim

Prediction of stock price in Istanbul Stock Exchange

b) Methodology and Results

For investors, analysts, and academics, predicting the direction of stocks is crucial. We propose in this study to use Turkish texts such as social media platforms to anticipate the direction of equities in the Turkish stock market (BIST100). Various deep learning methodologies have been used for this goal. Deep learning algorithms, long short-term memory networks, recurrent neural networks, and convolutional neural networks are examined, as are Word2Vec, GloVe, and FastText as word embedding models. Four distinct sources of Turkish news are gathered to demonstrate the efficiency of the suggested model. Public Disclosure Platform (KAP) news articles about stocks, Bigpara text-based technical analysis of each stock, and user comments from both Twitter and Mynet Finans platforms are aggregated[51].

Due to the challenges and opportunities it presents, stock market parameter forecasting is an essential study topic for both financial professionals and machine learning researchers. Despite the difficulty in obtaining financial data, interest in this field of study is fast growing[50]. So, using financial sentiment analysis, estimate the direction of the Borsa Istanbul 100 Index, enrich the datasets using various methodologies from a semantic perspective, and improve the classification performance of the system by combining ensemble learning with deep learning algorithms.[51].

c) Advantages

The key benefit is that stock price predictions will be more accurate because the model employs RNN, LSTM, Machine Learning, and Deep Learning models. In addition, the algorithm can forecast future stock prices for the next 30 days and display them in a graph. The model may also output Individual Predicted Close prices for the Predicted 30 days, which is a key feature.

d) Disadvantages

To run at and for each sequence time-step, LSTM requires four linear layers (MLP layers) per cell. To be computed, linear layers demand a lot of memory bandwidth; in reality, they can't use a lot of compute units since the system doesn't have enough memory bandwidth to feed them.

## **4. EVALUATION METHODS AND DATASETS**

### **4.1 Evaluation methods and metrics used**

F1 Score Precision and recall are measurements which are together unique and feature an inverse relationship. If we're inquisitive about both, the F1 score may be used to combine precision and bear in mind right into a single metric.

Mean Reciprocal Rank (MRR) The Mean Reciprocal Rank (MRR) evaluates the responses retrieved, in correspondence to a query, given their possibility of correctness. This assessment metric is typically utilized in information retrieval duties pretty regularly.



**Mean Average Precision (MAP)** Similar to MRR, the Mean Average Precision (MAP) calculates the suggest precision across each retrieved end result. It's also used heavily in fact retrieval responsibilities for ranked retrieval outcomes.

**Root Mean Squared Error (RMSE)** When the predicted outcome is a real cost then we use the RMSE. This is commonly used at the side of MAPE — which we are able to cover subsequent — within the case of regression troubles, from obligations including temperature prediction to inventory marketplace price prediction.

**Mean Absolute Percentage Error (MAPE)** The MAPE is the average absolute percent errors for every information point whilst the anticipated outcome is continuous. Therefore, we use it to test compare the performance of a regression version.

**Area Under the Curve (AUC)** The AUC enables us quantify our model's capacity to split the classes with the aid of taking pictures the count of nice predictions that are correct in opposition to the rely of high quality predictions that are incorrect at exceptional thresholds.

**Metrics 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

## 4.2 Popular datasets used

- BSE(Bombay stock exchange) and NSE(National Stock Exchange) websites - all of the information associated with the stock charges of the agencies had been taken from bse and nse websites, which can be the fine resources for such statistics as they replace it regularly and are the most reliable resources for inventory fees.
- 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.
- CSMAR database for stock related data

## 5. Evaluating Different Systems for Sentiment Analysis

Title	Scope of work	Existing	Datasets used	Evaluation metrics and	Limitations faced
-------	---------------	----------	---------------	------------------------	-------------------

		<b>algorithms used</b>		<b>corresponding performance gains</b>	
Prediction of stock values changes using sentiment analysis of stock news headlines [1]	Analyze changes in stock market values and forecast stock values using new economic news about companies	RNN, NLTK Vader Lexicon, Text Blob	IMDB review dataset	NLTK Vader lexicon , Text blob, RNN	Disadvantage is the whole system depends on news headlines and user response. Due to this we need user response in-order to habitat with this method. If time arise when we have no news headline based on stock prises which might cause failure of this service. But the above mentioned problem have 10^-9%* chances of happening.
A Method of Using News Sentiment for Stock Investment Strategy [2]	This study assesses the sentiment of Japanese news and attempts to apply it to individual stock investment methods.	LSTM, RNN	TOPIX500 Stock Prices	Equal weighted, Market Value weighted of cumulative excess return of daily rebalance energy	The disadvantage is making news headlines in order to maintain the share market is against human faith.  Not only this, by this way any government body can easily manipulate future results.
Augmented Textual Features-Based Stock Market Prediction [3]	To predict stock market, due to its dynamics, non-linearity and complexity nature	DNN, Deep CNN, LSTM	NASDAQ Stock Price	Dickey-fuller test to check stationarity. Machine learning techniques.  Granger causality test for the four stocks. Tweet mining	The demerit is still there are chances that data might get false in some exceptional case when false reviews are done.
The Impact of COVID-19 on the Chinese Stock Market: Sentimental or Substantial?	The authors of this work employed an event study to look into the effects of COVID-19 on the Chinese stock market, as well as the effect of individual investor emotion on their returns.	Panel regression model, Fama French model	Stock-related financial data are from the CSMAR database, Sentiment data used in this work is	Share market daily return, sample stock daily return	This paper is mainly focussed on 7 industries so this cannot be used as a basis for other industries.

[4]			GubaSenti		
Detecting a Risk Signal in Stock Investment Through Opinion Mining and Graph-Based Semi-Supervised Learning [5]	To avoid credit events that might cause a national and global economic crisis ultimately leading to socioeconomic losses	Logistic regression model  Linguistic rule-based model  graph-based semi-supervised learning	Data related to Hyundai Merchant Marine	(1) data collection and filtering,  (2) credit risk assessment and early signal detection, and  (3) prediction of credit events.	Despite the fact that each firm is expected to report on its financial health and significant changes in operations and management, some companies may hide their unfavourable situation. As a result, certain events are buried or not visible, making it more difficult to detect these hidden occurrences, which could result in a big loss.
Sentiment Analysis of Indian Stock Market Volatility [6]	This paper attempts to shape the volatility of Indian Stock market using investor	Natural Language Toolkit(NLTK) and Principal Component Analysis(PCA), Sklearn feature extractor	The datasets that are used in the above study are the prominent web sources for the Indian financial market and business.	GARCH model	The suggested work's first flaw is that it only uses eight emotions, which are divided into two categories. In order to analyse the feeling, the number of emotions should be raised, and a new broader class called as neutral should be added. The study's future scope includes a comparison of diverse sectors of the Indian stock market, such as energy, telecommunications, and metal.

Estimating the relative effects of raw material prices, sectoral outlook and market sentiment on stock prices [7]	<p>To determine the relation among movement of uncooked cloth expenses and share rate of the organization.</p> <p>The scope of the problem is restrained to the following sectors : Oil and Gas, Metal, FMCG and Healthcare.</p>	Random Forest, Gradient Boosting, Genetic Algorithm	BSE(Bombay stock exchange) and NSE(National Stock Exchange) websites	<p>Boruta Algorithm</p> <p>Random Forest</p> <p>Gradient Boosting</p> <p>Genetic Algorithm</p> <p>Monte Carlo Methods for deriving significance.</p>	<p>A bad directional selectivity for diagonal functions, because the wavelet filters are separable and real.</p> <p>Longer compression time.</p> <p>And loss of shift invariance, in which means that small shifts inside the input signal can reason</p> <p>fundamental variations in the distribution of power between DWT coefficients at distinctive scales.</p> <p>The cost of computing DWT in comparison to DCT can be better.</p>
Exploiting textual and relationship information for fine grained financial sentiment analysis [8]	Identification of expressions (superb impartial terrible) toward the concern by way of expressing sentiments in textual content. Novel approach to capture implicit sentiments and the contagion method.	aFFNN, LMBP algorithm.	IMDB dataset and Yelp dataset	Mean squared error (MSE) , Cosine Similarity (CS)	<p>1. It is highly dependent on hardware.</p> <p>2. Lack of assurance of proper network structure.</p> <p>3. The difficulty to show the problem to the network.</p>
Harvesting social media sentiment analysis to enhance stock market prediction using deep learning [9]	To discover how moves in a organisation's stock charges correlate with expressed reviews of the public concerning that employer.	RNN algorithm, LSTM algorithm.	Money Control, IFL, Economic Times, Twitter, NSE Stock Data	Sentiment value was used as a metric to compare.	LSTMs get influenced by various irregular weight introductions and consequently act very like that of a feed-forward neural net. They incline toward little weight introducing all things being

					<p>equal.</p> <p>LSTMs are inclined to overfitting and it is difficult to apply the dropout calculation to check this difficulty. Dropout is a regularisation approach wherein input and intermittent associations with LSTM gadgets are probabilistically prevented from initiation and weight refreshes at the same time as getting ready an organisation.</p>
Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation [10]	To solve stock portfolio allocation and maintain it to get maximum return with minimum risk involved.	DPG and DDPG	Google News, Twitter	Annual Return, Sharpe Ratio, Annual Std. Error and Final Portfolio Value	Deep Deterministic Policy Gradient (DDPG) algorithms may handle exceedingly complicated issues in environments with continuous state and action spaces, but they can also fail in seemingly small and easy environments, for reasons that are currently unknown.
Stock Prediction by using NLP and Deep Learning Approach [11]	Identifying the stocks that will perform best in future and will give maximum profits using a deep learning approach.	CNN	Twitter Tweets, top 5 performing companies' data in IT sector	Market Value Comparison	The position and orientation of objects are not encoded by CNN. Inability to be spatially invariant when dealing with incoming data. It is necessary to collect a large amount of training data. When CNN-LSTM is applied, it is capable of providing the most accurate stock price forecasts.
An Efficient Word Embedding and Deep Learning Based Model to Forecast the Direction of	Prediction of stock price in Istanbul Stock Exchange	CNN, RNN, Word embedding models	Twitter, KAP, Mynet Finans, Bigpara	F-criterion and accuracy	To run for each sequence time-step, LSTM requires four linear layers (MLP layers) per cell. To be computed, linear layers demand a lot of memory bandwidth; in reality, they can't use a lot of compute units since the system

Stock Exchange Market Using Twitter and Financial News Sites: A Case of Istanbul Stock Exchange (BIST 100) [12]					doesn't have enough memory bandwidth to feed them.
---	--	--	--	--	--

## 6. Conclusion and future work

With this paper our motive was to generalise and represent various recent trends happening in the stock market with respect to sentiment analysis. The papers discussed above give a wide prospect about various techniques which can be used to predict the market. Along with prediction the papers present some alerts which must be taken care of when handling Sentimental analysis in stock prediction.

In the first paper the authors analysed the difference of emotional graph and trading graph correlation. They applied various techniques and made a set of results. In which some techniques were giving high opening and closing differences as well as minor differences in highest and lowest value of a stock price. Along with this they came to a conclusion that Economic news has an impact on the stock market irrespective of its textual context. They even concluded that there is no neutral emotion, since every emotion caused an impact on the market [1].

In the second paper the focus was made to find the prediction duration and its life expectancy. On doing various experiments it was found that the sentimental analysis prediction is really short lived. The impact caused on stock price due to a certain sentimental headline is really short living. None of the sentimental predictions can predict stock for a week or month. Thus for people who are doing day to day trading for them sentimental analysis is the best technique [2].

Not only this if a certain headline is on the front page or the first bullet news then the impact caused on that stock is high as compared to headlines on the next page. It was even found that articles with more texts can cause major impact as compared to articles with less words [2].

In the third paper it is observed that the sentimental prediction provided a regular pattern. This regular pattern can be really helpful for intra day trading. Since as discussed before sentimental prediction world for a day or two. Thus getting such a regular pattern can be helpful in such a short duration. But along with the derivation and usage of regular patterns one should always remember there is a high chance of getting that pattern to mislead a prediction. Because being in such a volatile market where linearity is near to nil we can totally depend on a pattern and trade. Thus such patterns must only be accepted if found true for a large amount of data [3].

According to the fourth study, pandemics can cause widespread negative emotion, resulting in investor worry and market instability. During the epidemic, stock return volatility is determined by sentiment rather than economic losses. Due of the great unpredictability of epidemics, investors may be able to earn a higher return by holding bellwether pharmaceutical stocks in the early stages. Then, as time goes on, investors should gradually cut stockholdings in the pharmaceutical industry while increasing stockholdings in government-favored industries. In addition, stocks with high risk factors, such as high P/E and P/B ratios, high CMV, low institutional shareholding ratio, and low net assets, should be avoided during the middle and late stages of the epidemic [4]. The future scope of this is to the UN's food crisis alarm, the food business has become a new emphasis in the post-event window. As a result, it's worth looking into the function of industry influences in epidemics.

In the fifth paper, they have suggested a new algorithm to support decision-making in stock investment by detecting early signals and predicting the occurrence probability of credit events through opinion mining and logistic regression models[5]. A key goal of this article is to use sentiment analysis to detect danger in the stock market; once a person is aware of risk, he or she can be prepared and adjust their course of action. They used sentiment

analysis, word2vec, and graph-based semi-supervised learning to filter bogus information, assess credit risk and discover risk signals, and forecast future occurrences of credit events, all of which have a lot of potential for future work in real-world challenges.

In the sixth paper, their main focus is to oppose the point made by Markovits which states that the rational behaviour of investors and market fundamentals is believed to be the sole factor that plays a decisive role in shaping the decision of such investors[30]. This research takes a different approach to modelling conditional volatility by employing common macroeconomic or financial variables as sentiment components to determine their impact on market conditional volatility. To quantify different forms of feelings in the Indian market, we used a novel approach of leveraging news-driven sentiment research and augmenting volatility models with such sentiment components. Major findings of the paper suggest the dominant role of negative sentiments in shaping conditional volatility in the Indian stock market. Findings also support evidence of noise traders which signifies immaturity of the Indian financial market. [6].

In the 7th paper, the prevailing scope of the paper is limited to market sectors particularly : Oil and Gas, Metal, FMCG and Healthcare. By the proposed frameworks the observer can consist of distinct sectors of the marketplace too within the future. The paper contributes to applicable literature by way of combining or uniting wavelet evaluation and machine getting to know to determine the relation among the motion of raw substances and percentage costs of an organization, which impacts the cease client. The scope of the trouble is confined to the subsequent sectors : Oil and Gas, Metal, FMCG and Healthcare. The counseled strategies and frameworks referred to in the paper can be used to determine the desired relation for other sectors as well [7].

In the eighth paper, 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.

The particular paper mainly focuses on showcasing the textual context that can be modelled on lithography, to study and gain further insight into sentiment analysis and improve it [8].

In the ninth paper, 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. The method can be made more accurate and more optimised in the future by making some changes or using a different algorithm [9].

Future work in the tenth paper might concentrate on getting more tweets each day, broadening acquisition to include stock market-specific news websites (CNBC, Business Standard, etc. ), and processing photographs, as most tweets and news online are now given as image snippets. Multi-agent reinforcement learning techniques can be researched in this instance. Transaction fees, trading limitations, cash holding restrictions, and liquidity shortages will all be investigated as external restraints on retail and institutional traders. Furthermore, natural language processing on financial data is a non-trivial endeavour due to the extensive use of metaphors, sarcasm, domain specialised terminology, and other indirect linguistic references in common English, especially in material that expresses an opinion. Being able to understand such language could help predict market emotion more accurately[10].

Future work in the eleventh and twelfth paper could concentrate on performing more calculations and implementing all of the current methodologies in order to provide live proposals to stock exchange financial specialists. Furthermore, for forecasting, they will focus on the complete securities market, and the authors will try to expand the data set and fine-tune the parameters in order to predict more accurate stock prices.[11][12]

**Acknowledgments :** This research has been supported by Dr. Saravanakumar Kandasamy from Department of Software Systems, School of Computer Science and Engineering, VIT Vellore, India. We feel grateful to Dr. Saravanakumar Kandasamy who once offered us valuable suggestions during the study period. Our sincere thanks are also given to the anonymous reviewers, from whose comments we have benefited greatly.

## References

- [1] Nemes, L., & Kiss, A. (2021). Prediction of stock values changes using sentiment analysis of stock news headlines. *Journal of Information and Telecommunication*, 1-20.
- [2] Katayama, D., & Tsuda, K. (2020). A Method of Using News Sentiment for Stock Investment Strategy. *Procedia Computer Science*, 176, 1971-1980.
- [3] Bouktif, S., Fiaz, A., & Awad, M. (2020). Augmented textual features-based stock market prediction. *IEEE Access*, 8, 40269-40282.
- [4] 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.
- [5] 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.
- [6] Paramanik, R. N., & Singhal, V. (2020). Sentiment Analysis of Indian Stock Market Volatility. *Procedia Computer Science*, 176, 330-338.
- [7] 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.
- [8] Daudert, T. (2021). Exploiting textual and relationship information for fine-grained financial sentiment analysis. *Knowledge-Based Systems*, 107389.
- [9] 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.
- [10] Koratamaddi, P., Wadhvani, 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.
- [11] Deshmukh, R. (2021). Stock Prediction by using NLP and Deep Learning Approach. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(1S), 202-211.
- [12] 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.
- [13] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowl.-Based Syst.*, vol. 89, pp. 14–46, Nov. 2015.
- [14] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *Proc. 5th Int. AAAI Conf. Weblogs Social Media*, 2011, pp. 450–453.
- [15] T. Sun, J. Wang, P. Zhang, Y. Cao, B. Liu, and D. Wang, "Predicting stock price returns using microblog sentiment for chinese stock market," in *Proc. 3rd Int. Conf. Big Data Comput. Commun. (BIGCOM)*, Aug. 2017, pp. 87–96.
- [16] K. Ichinose and K. Shimada, "Stock market prediction from news on the Web and a new evaluation approach in trading," in *Proc. 5th IIAI Int. Congr. Adv. Appl. Informat. (IIAI-AAI)*, Jul. 2016, pp. 78–81.
- [17] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of Twitter data for predicting stock market movements," in *Proc. Int. Conf. Signal Process., Commun., Power Embedded Syst. (SCOPEs)*, Oct. 2016, pp. 1345–1350.
- [18] D. Lv, S. Yuan, M. Li, and Y. Xiang, "An empirical study of machine learning algorithms for stock daily trading strategy," *Math. Problems Eng.*, vol. 2019, Apr. 2019, Art. no. 7816154.
- [19] Daisuke Katayama and Kazuhiko Tsuda. (2018) "A Method of Measurement of The Impact of Japanese News on Stock Market." 22nd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems KES2018
- [20] S. Bouktif and M. Adel Awad, "Predicting stock market movement: An evolutionary approach," in *Proc. 7th Int. Joint Conf. Knowl. Discovery, Knowl. Eng. Knowl. Manage.*, 2015, pp. 159–167.
- [21] D. Shah, H. Isah, and F. Zulkernine, "Predicting the effects of news sentiments on the stock market," in *Proc. IEEE Int. Conf. Big Data*, Dec. 2018, pp. 4705–4708.
- [22] A. Picasso, S. Merello, Y. Ma, L. Oneto, and E. Cambria, "Technical analysis and sentiment embeddings for market trend prediction," *Expert Syst. Appl.*, vol. 135, pp. 60–70, Nov. 2019.



- [23] P. D. Yoo, M. H. Kim, and T. Jan, "Machine learning techniques and use of event information for stock market prediction: A survey and evaluation," in *Proc. IEEE Int. Conf. Comput. Intell. Modelling, Control Automat., Int. Conf. Intell. Agents, Web Technol. Internet Commerce*, vol. 2, Nov. 2005, pp. 835–841.
- [24] 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.
- [25] Liu, H., Manzoor, A., Wang, C., Zhang, L., & Manzoor, Z. (2020). The COVID-19 outbreak and affected countries stock markets response. *International Journal of Environmental Research and Public Health*, 17(8), 2800.
- [26] 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.
- [27] Löbner, H. (2014). When trust makes it worse—Rating agencies as disembedded service systems in the US financial crisis. *Service Science*, 6(2), 94-105.
- [28] Bakshi, R. K., Kaur, N., Kaur, R., & Kaur, G. (2016, March). Opinion mining and sentiment analysis. In *2016 3rd international conference on computing for sustainable global development (INDIACom)* (pp. 452-455). IEEE.
- [29] Chapelle, O., Schölkopf, B., & Zien, A. (2006). A discussion of semi-supervised learning and transduction. In *Semi-supervised learning* (pp. 473-478). MIT Press.
- [30] Markowitz, H. M. (1968). *Portfolio selection*. Yale university press.
- [31] Schmitt, N., & Westerhoff, F. (2017). Herding behaviour and volatility clustering in financial markets. *Quantitative Finance*, 17(8), 1187-1203.
- [32] Antoniou, C., Doukas, J. A., & Subrahmanyam, A. (2016). Investor sentiment, beta, and the cost of equity capital. *Management Science*, 62(2), 347-367.
- [33] 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.
- [34] 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.
- [35] 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.
- [36] Daudert, T. (2021). A multi-source entity-level sentiment corpus for the financial domain: the FinLin corpus. *Language Resources and Evaluation*, 1-24.
- [37] Sinha, N. R. (2016). Underreaction to news in the US stock market. *Quarterly Journal of Finance*, 6(02), 1650005.
- [38] 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.
- [39] Milosevic, N. (2016). Equity forecast: Predicting long term stock price movement using machine learning. *arXiv preprint arXiv:1603.00751*.
- [40] 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.
- [41] 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.
- [42] Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584*.
- [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 di*
- [45] Bao, W., & Liu, X. Y. (2019). Multi-agent deep reinforcement learning for liquidation strategy analysis. *arXiv preprint arXiv:1906.11046*.
- [46] 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*.

- [47]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.
- [48]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.
- [49]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
- [50]Tekin, S., & Çanakoglu, 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.
- [51]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)
- [52]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.

### Basic Information

Title NLP Paper  
Time 2021-12-08 12:55:57 ( +00:00 UTC )  
Words 12596

### Overall Similarity

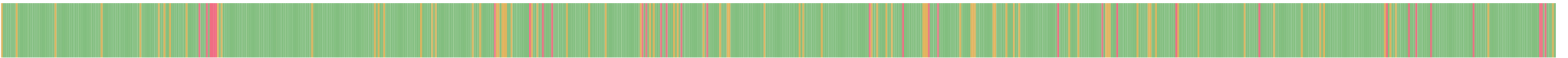
Overall Similarity  
Similarity on academic resources  
Similarity on internet resources



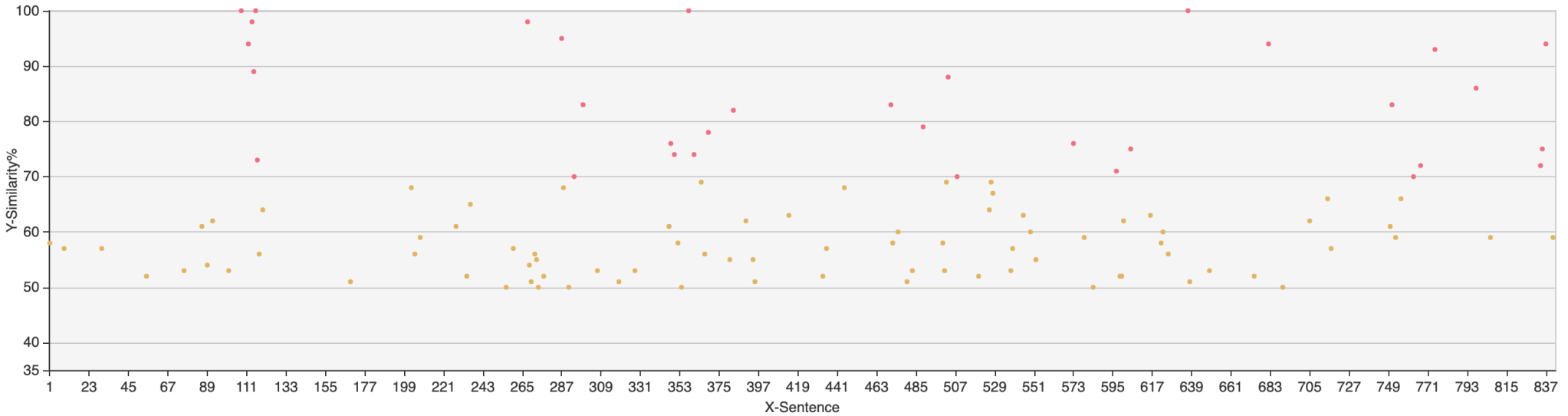
### Similarity Calculation Formula

(Sentence1 similarity + Sentence2 similarity..... / n)  
Sentence similarity scale from 0.0~1.0 Green sentence represent similarity-free

### Sentences similarity distribution



### Unqualified sentences dotted distribution map



### Similar resources on academic

Highly similar resources on academic is empty.

### Similar resources on internet

1. 0.5%  **Source :** Internet **Title :** Sentiment Analysis of Indian Stock Market Volatility - Science ... [🔗](#)

This report is powerd by paperpass.net similarity detecting system  
Copyright © 2021 PaperPass.Net