<u>Research Questionnaire</u>

<u>Note:</u>

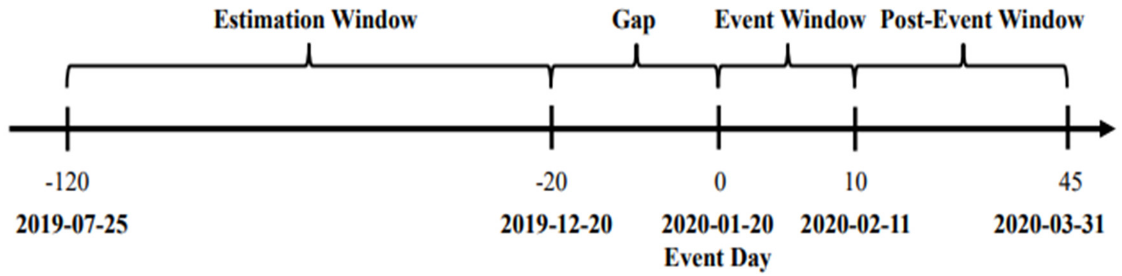
- **Fill up *Table 2* for each paper. (COPY TABLE 2 AND PASTE AT THE END OF THIS FILE FOR NEXT PAPER)**
- ***Green* – Write few or more lines of required in your own words.**
- ***Red* – write down the list of what is required and description about each in the list**

TABLE 1	
Reg. No. & Name	19BCE2222 - Hemaksh Chaturvedi
Team No.	**7**
Paper Title	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
***Citation** (APA style)*	1. 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. 2. 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. 3. Paramanik, R. N., & Singhal, V. (2020). Sentiment Analysis of Indian Stock Market Volatility. *Procedia Computer Science*, *176*, 330-338.

1.

TABLE 2	
Problem answered in this paper. **(1-2 lines)**	In this paper the authors have investigated the impact on the Chinese Stock Market caused by COVID-19 by doing an event study and examine the effect of individual investor sentiment on their returns.
Detailed description about the problem **(5-8 lines)**	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 region 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,
Why that problem is chosen in this paper? Scope of the problem and solution (*Refer Introduction*) **(5-8 lines)**	The above 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.
History of the problem. (*Refer Introduction*) **(8-10 lines)**	The following assumptions are tested in this article to see how sentiment affected stock market volatility during the outbreak. When important events have an impact on stock returns via sentiment, two requirements must be met. To begin with, the occurrence triggers intense negative emotions such as panic and anxiety. Previous research has suggested that public health threats like SARS and Ebola can influence market sentiment. Furthermore, the occurrence results in lower-than-normal yields on associated stocks.
List of the related/similar problems (*Refer Related work*) – Describe each with proposed solutions 1. Death and contagious infectious diseases: Impact of the COVID-19 virus on stock market returns 2. Stock markets' reaction to COVID-19: Cases or fatalities? 3. Investor sentiment and stock returns: Wenchuan Earthquake 4. Deaths, panic, lockdowns and US equity markets: The case of COVID-19 pandemic 5.	
Related problem 1 – Describe **(3-4 lines)**	This study looks into whether communicable infectious diseases have an impact on stock market performance. The data show that daily increases in total confirmed cases and total deaths induced by

	COVID-19 have a significant negative impact on stock returns across the board.
Paper in APA style	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.
Related problem 2 – Describe **(3-4 lines)**	After examining the stock markets' response to the COVID-19 pandemic they found that Stock markets reacted negatively to the increase of COVID-19 confirmed cases, according to our findings. That is, as the number of confirmed instances increased, stock market returns decreased. We also discovered that, in comparison to Journal Pre-proof the increase in the number of confirmed cases, financial markets reacted more proactively to the increase in the number of verified cases.
Paper in APA style	Ashraf, B. N. (2020). Stock markets' reaction to COVID-19: Cases or fatalities?. *Research in International Business and Finance*, *54*, 101249.
Related problem 3 – Describe **(3-4 lines)**	We find that stock returns for companies based near the epicenter are much lower than for companies located further away in the 12 months after the earthquake. Further research shows that this pattern of stock returns did not exist prior to the earthquake or for a long time afterward, and that it cannot be explained by actual economic losses or a shift in systemic risk.
Paper in APA style	Shan, L., & Gong, S. X. (2012). Investor sentiment and stock returns: Wenchuan Earthquake. *Finance Research Letters*, *9*(1), 36-47.
Related problem 4 – Describe **(3-4 lines)**	According to the findings, an increase in verified coronavirus cases and deaths is linked to a large increase in market illiquidity and volatility. Declining sentiment, as well as the imposition of limits and lockdowns, contribute to the degradation of market liquidity and stability. Negative sentiments from Coronavirus related news deteriorate stock market liquidity and stability.
Paper in APA style	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.
Related problem 5 – Describe **(3-4 lines)**	According to behavioural finance research, investor sentiment influences investment decisions and, as a result, stock pricing. This study looks at how stock prices in the United States were affected by the geographic proximity of information released by the 2014–2016 Ebola outbreak events, as well as extensive media coverage. Result suggests that the information about Ebola outbreak events is more relevant for companies that are geographically closer to both the birthplace of the Ebola outbreak events and the financial markets.
Paper in APA style	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.
What is the proposed solution in this paper for the problem chosen? (*Refer Proposed work*) **(5-8 lines)**	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.

Architecture of the proposed solution. (*Refer proposed work*) **Diagram**	 Time frames used for Panel Regression model
Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)	Panel Regression Model
List of existing algorithms used by the authors to complete the proposed work. **(1-2 lines for each algorithm)**	1. Panel regression model Panel data 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. 2. Fama–French model. The Fama-French 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.
List of datasets used. (*Refer experimental evaluation/result discussion*) **(3-4 lines)**	1. Stock-related financial data are from the CSMAR database 2. Sentiment data used in this work is GubaSenti
References/links to each of the dataset used in this paper (in APA style)	1. https://www.gtarsc.com/ 2. https://ifind.bnu.edu.cn/
Why the above dataset(s) used? (*Refer experimental evaluation/result discussion*)	1. This covers the period from July 25th, 2019 to March 31st, 2020 Companies with an A-share market capitalization are used as examples. In the annual report, samples having negative net assets are excluded from the panel data.

(3-4 lines)	2. captures the individual investor sentiment by text analytics on opinions from Guba – the biggest online financial social platform in China for individual investors to share and exchange their opinions and experiences on stocks.
List of equations that are very well applied in this problem domain	Equation 1: $R_{i,t} = \alpha + \gamma MKT_t + \delta SMB_t + \eta HML_t + \varepsilon_t$ Description: Ordinary least squares (OLS) regression is shown above 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 Equation 2: $AR_t = R_t - [\hat{\alpha} + \hat{\gamma} MKT_t + \hat{\delta} SMB_t + \hat{\eta} HML_t]$ Description: Abnormal returns is shown above where R_t represents the actual return on date t in the event window Equation 3: $CAR = \sum_{t=1}^{n} AR_t$ Description: individual abnormal returns to create a "cumulative abnormal return (CAR)"

<table>
<tr>
<td>List of method(s)/metrics used to evaluate the proposed approach. (Refer experimental evaluation/result discussion)
(5-8 lines)</td>
<td>
Table 1.. Summary Statistics.
<table>
<tr><th>Variable</th><th>Observations</th><th>Mean</th><th>SD</th><th>Min</th><th>Max</th></tr>
<tr><td colspan="6">Panel A: Estimation Window</td></tr>
<tr><td>Market return</td><td>100</td><td>0.001</td><td>0.008</td><td>-0.019</td><td>0.022</td></tr>
<tr><td>Stock return</td><td>324,729</td><td>0.001</td><td>0.022</td><td>-0.102</td><td>0.103</td></tr>
<tr><td>Sentiment</td><td>324,729</td><td>0.486</td><td>2.636</td><td>-19.880</td><td>21.463</td></tr>
<tr><td colspan="6">Panel B: Event Window</td></tr>
<tr><td>Market return</td><td>10</td><td>-0.005</td><td>0.030</td><td>-0.080</td><td>0.020</td></tr>
<tr><td>Stock return</td><td>31,475</td><td>-0.005</td><td>0.046</td><td>-0.103</td><td>0.104</td></tr>
<tr><td>Sentiment</td><td>31,475</td><td>0.431</td><td>2.728</td><td>-10.912</td><td>12.024</td></tr>
<tr><td colspan="6">Panel C: Post-event Window</td></tr>
<tr><td>Market return</td><td>36</td><td>-0.001</td><td>0.018</td><td>-0.039</td><td>0.034</td></tr>
<tr><td>Stock return</td><td>106,164</td><td>0.000</td><td>0.034</td><td>-0.109</td><td>0.103</td></tr>
<tr><td>Sentiment</td><td>106,164</td><td>0.627</td><td>2.550</td><td>-10.533</td><td>11.527</td></tr>
</table>
Notes: Table 1 reports summary statistics of the comprehensive A-share market daily return, sample stock daily return and investor sentiment. In panel A, the sample period is from July 25, 2019 to December 19, 2019. In panel B, the sample period is from January 20, 2020 to February 10, 2020. In panel C, the sample period is from February 11, 2020 to March 31, 2020. The market return and stock return data are derived from CSMAR database. The sentiment data is GubaSenti established by International Institute of Big Data in Finance, BNU(http://ifind.bnu.edu.cn/).
</td>
</tr>
<tr>
<td>List of supporting tools/concepts
(3-4 lines)</td>
<td>1. Panel regression model
2. Fama–French model</td>
</tr>
<tr>
<td>What are the similar approaches with which the proposed approach is compared?
(Refer experimental evaluation/result discussion)
Explain each of these approach
(3-4 lines)</td>
<td>Approach/method 1: Event Study
Event study is applied in this work to identify abnormal returns in the stock market from the outbreak of COVID-19.

Approach/method 2: Panel regression model
Panel regression can better capture the time-varying relationship between dependent and independent variables due to its ability to extract changes from panel data and minimize estimation bias</td>
</tr>
</table>

How the results of proposed approach are compared with other similar approaches? (*Refer experimental evaluation/result discussion*)	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. 2. 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.
Advantages/merits of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	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 sentimental analysis.
Disadvantages/limitations of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	This paper is mainly focused on only 7 main industries so this cannot be used as a basis for other industries.
Future work as stated by authors (*Refer conclusion / result discussion / experimental evaluation*)	Due 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.
Your one page write-up about this paper	
The COVID-19 pandemic, which began in early 2020, has caused financial market turmoil. The stock market in the United States witnessed circuit breakers twice in a single week, and the situation in other countries was not much better .Based on an event analysis and panel regression, this research assesses the impact of COVID-19 on China's stock market. This study adds to the body of knowledge by investigating the unanticipated impact of a feared disease's emergence on Chinese financial markets. Data also demonstrate that pandemics can induce widespread negative sentiment, resulting in investor concern and market volatility. Stock return volatility during the epidemic is influenced by sentiment and is not just due to economic losses. Stocks in various businesses and with various financial features are affected in different ways. During the	

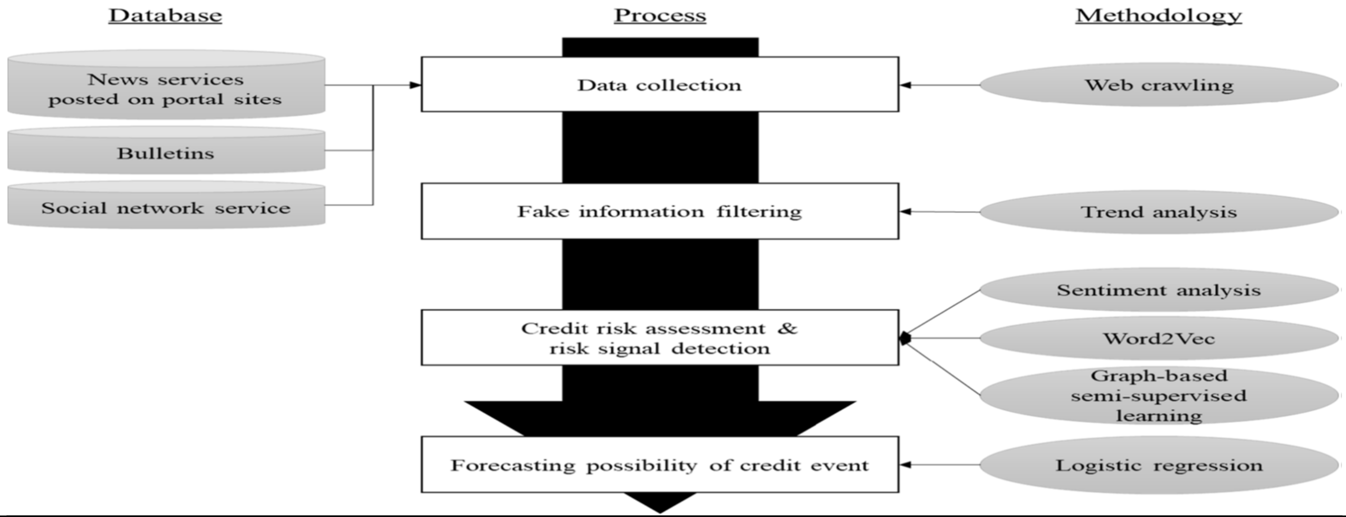
middle and late phases of the epidemic, equities with high risk factors, such as high P/E and P/B ratios, high CMV, low institutional shareholding ratio, and low net assets, should be avoided.
<u>Your findings: (possible alternate for the solution proposed)</u> • Stocks with high PB, PE and CMV, low institutional shareholding ratio and low net assets are found to be more sensitive to the turbulence. • Only 7 industries related to Pharmacy, Digitalization, and Agriculture are boosted during the merging window of event and post-event

2.

TABLE 2	
Problem answered in this paper. (1-2 lines)	How to avoid credit events that might cause a national and global economic crisis ultimately leading to socioeconomic losses?
Detailed description about the problem (5-8 lines)	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 recognising 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 practise 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.
Why that problem is chosen in this paper? Scope of the problem and solution (*Refer Introduction*) (5-8 lines)	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 rumours and fake news. Second, sentiment analysis and opinion mining recognised 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.
History of the problem. (*Refer Introduction*) **(8-10 lines)**	Since the significance of monitoring and detecting signals has been emphasised, attempts have been made to identify risk signals or bankruptcy by taking into account a variety of elements that may influence the likelihood of a firm failing. These studies have discovered that a variety of elements (internal and external factors, financial and noneconomic factors, corporate culture, and management and investor attitudes) interact to impact a firm's propensity for failure. In addition, as natural language processing (NLP) has advanced, there have been attempts to employ text data for financial analysis, resulting in a new discipline known as natural language-based finance. Natural language based financial forecasting These studies, however, have certain drawbacks. To begin with, the majority of research have concentrated on predicting stock price gains or declines, but attempts to discover an early indication of credit events have been rare. Second, many studies have focused on numerical data, such as stock prices and financial statements, whereas there has been little study on social data. Recent research have used news headlines to detect business-related events, but the results have not led to stock investing decisions. Furthermore, several elements impacting business bankruptcy, such as consumer behaviour, were still not taken into account in most research. Finally, in the situation of firms that aren't listed on the exchange, Finally, there is no information accessible for firms that are not listed on stock exchanges since they are not required to disclose their information to the public. As a result, there is a scarcity of study on these private businesses.
List of the related/similar problems (*Refer Related work*) – Describe each with proposed solutions 1. A Discussion of Semi-Supervised Learning and Transduction 2.	
Related problem 1 – Describe **(3-4 lines)**	It involves three researchers, who will be referred to as A, B, and C just for simplicity, without implying any one-to-one correspondence to real people. Talking about "What is the Difference Between Semi-Supervised and Transductive Learning? "
Paper in APA style	Chapelle, O., Schölkopf, B., & Zien, A. (2006). A discussion of semi-supervised learning and transduction. In *Semi-supervised learning* (pp. 473-478). MIT Press.
Related problem 2 – Describe **(3-4 lines)**	In this paper they have investigated the possibilities of a novel semi-supervised learning approach that combines the use of random projection scaling as part of a vector space model with the use of support vector machines to do reasoning on a knowledge base. The latter is created by combining a commonsense graph with a linguistic resource for the lexical representation of affect.

Paper in APA style	Hussain, A., & Cambria, E. (2018). Semi-supervised learning for big social data analysis. Neurocomputing, 275, 1662-1673.
Related problem 3 – Describe **(3-4 lines)**	We describe a novel semi-supervised social media spammer detection system that makes extensive use of message content, user behaviour, and social relationship data.
Paper in APA style	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.
Related problem 4 – Describe **(3-4 lines)**	In this paper we see that before proving its usefulness with a data set of hotel reviews, we describe how semi-supervised learning techniques can be utilised to detect spam reviews.
Paper in APA style	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.
Related problem 5 – Describe **(3-4 lines)**	This paper proposes a procedure that makes use of web-based semantic information. In order to optimise the process of extracting information from unstructured data sources, our system uses structured information crawled from the semantic web. We also make recommendations for how to incorporate user interaction into the process.
Paper in APA style	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.*
What is the proposed solution in this paper for the problem chosen? (*Refer Proposed work*) **(5-8 lines)**	The study proposes a novel algorithm to recognise 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 comprised of 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.

Architecture of the proposed solution. (*Refer proposed work*) **Diagram**	
Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)	Sentiment analysis based on opinion data
List of existing algorithms used by the authors to complete the proposed work. **(1-2 lines for each algorithm)**	1. Logistic regression model Logistic regression measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function 2. Linguistic rule-based model linguistic rule is a rule describing a linguistic practice. rule. concept, conception, construct - an abstract or general idea inferred or derived from specific instances. 3. graph-based semi-supervised learning An important class of SSL methods is to naturally represent data as graphs such that the label information of unlabelled samples can be inferred from the graphs, which corresponds to graph-based semi-supervised learning (GSSL) methods
List of datasets used. (*Refer experimental evaluation/result discussion*)	1. Data related to Hyundai Merchant Marine Objective Data Subjective Data

(3-4 lines)	
References/links to each of the dataset used in this paper (in APA style)	1.No links/refrences were mentioned in the paper.
Why the above dataset(s) used? (*Refer experimental evaluation/result discussion*) **(3-4 lines)**	Data related to Hyundai Merchant Marine were collected from diverse databases. To obtain objective data, news and numeric data, such as the stock prices and operating statuses of firms, were collected. In addition, subjective data, including those from SNSs and articles posted on portal sites, to reflect the opinions of general users, were collected through web scraping. Hyundai Merchant Marine has a large amount of data arising from a major crisis in the shipping industry (81,425 articles including both objective and subjective data).
List of equations that are very well applied in this problem domain	Equation 1: i = $\sum_{j=1}^{m}(\mathrm{Pi,j})$ Description: Sentimental value of keyword Equation 2: $\mathrm{P_{i,j}} = \frac{1}{r^2}$*(Sentimental value of core keyword i) Description: Proximity index between keyword i and core keyword j Equation 3: $r_j = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ (j=1,2,…m) Description: Euclidean distance between keyword i and core keyword j Equation 4: I= $\sum_{\mathrm{i}=1}^{n} W_i S_i$ Description: I stands for Indicator for monitoring
List of method(s)/metrics used to evaluate the proposed approach. (*Refer experimental evaluation/result discussion*) **(5-8 lines)**	(1) data collection and filtering, (2) credit risk assessment and early signal detection, and (3) prediction of credit events.
What are the similar approaches with which	Approach/method 1: DATA COLLECTION AND INFORMATION FILTERING This paper utilizes all information related to stock investment obtained from news and opinions posted

the proposed approach is compared? (Refer experimental evaluation/result discussion) **Explain each of these approach** **(3-4 lines)**	on websites and SNSs. opinions are also collected from finance websites, online communities, and SNSs. After collecting raw data from web services, these data need to be refined to clarify documents and derive more accurate results. Approach/method 2: CREDIT RISK ASSESSMENT AND RISK SIGNAL DETECTION This paper attempts to propagate the sentiment value of core keywords to relevant words after allocating sentiment value for each document with naïve-Bayes classification, word2vec, and graph-based semi-supervised learning. After filtering fake information, all textual data are preprocessed by natural language processing to assess credit risk. Data preprocessing includes the following processes: (1) splitting sentence, (2) tokenizing, and (3) part-of-speech (POS) tag- ging and parsing. This preprocessing aims to rate documents using the sentiment values of sentences and words. Approach/method 3: FORECASTING OCCURRENCE POSSIBILITY OF CREDIT EVENT After identifying the sign of a credit event, the actual possibility of credit event occurrence is predicted by logistic regression. The prediction model based on logistic regression is composed of sub-indices for assessing risk at the prior step. The regression equation for credit event occurrence is estimated through logistic regression.
How the results of proposed approach are compared with other similar approaches? (*Refer experimental evaluation/result discussion*)	This paper suggests behavior- 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 behavior-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, we 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).
Advantages/merits of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	Our article handles false information from the perspective of data processing since the stock market is impacted by investor information, and there is a lot of it regardless of honesty. As a result, based on a vast quantity of opinion data, we presented an algorithm for recognising risk signs early. By presenting a method for when investors decide whether or not to trade stocks, our work increased the availability of social data in the finance market. Semi-supervised graph-based learning aids in the situational learning and classification of words or texts. Other data is smoothed and consistent with the labelled data using graph-based semi-supervised learning.
Disadvantages/limitations of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	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.
Future work as stated by authors (Refer conclusion / result discussion / experimental evaluation)	To begin, the process of screening false information should be expanded in many ways. Purifying a significant volume of raw data that has collected over time will take more time and effort. Because the goal of this work was to build an algorithm for identifying early signals and aiding decision-making, little thought was given to developing an investing strategy. While the proposed algorithm can give initial risk or opportunity for stock investment, the comprehensive approach is not provided. For each industry, the sentiment dictionary must be improved using data from a variety of databases. Because each industry has its own set of features, there is some uncertainty in terms of terminology. Although the sentiment dictionary used in this work was based on a database for a specific industry, it should be expanded and enhanced in the near future by using another database for generality.
Your one page write-up about this paper	
The goal of this research is to use opinion mining and graph-based semi-supervised learning to create an algorithm to aid in stock investing decision-making. This study focuses on the following fundamental procedures to achieve this goal: Using sentiment analysis, word2vec, and graph-based semi-supervised learning, we can (1) filter false information, (2) assess credit risk and discover	

risk signals, and (3) forecast future occurrences of credit events. First, financial data was gathered, including news, messages from social networking sites, and financial accounts. Fake material, such as rumours and fake news, was filtered out of these data using author analysis and a rule-based approach. Second, credit risk was determined using sentiment analysis and opinion mining for both social data and news in the form of a sentiment score and document trend for each stock. The degree of evaluated risk was then used to detect a signal for a credit event. As a result, based on the risk signal, the likelihood of credit events such as delisting and bankruptcy in the near future was forecasted using logistic regression. This study used a real-world case study to demonstrate the applicability of the suggested method. The findings of this study can assist investors in monitoring a vast quantity of previously gathered data and detecting concealed danger signs ahead of time.

In conclusion, this research proposes a new method to aid stock investing decision-making by recognising early signals and forecasting the likelihood of credit events using opinion mining and logistic regression models. When investors decide to purchase or sell stocks, news and official reports produced by securities analysts or stock experts have long been significant and plentiful sources. When making stock investment decisions, however, with the development of IT devices and the expansion of SNS use, information or individual intentions are fiercely exchanged through online communities and private SNSs.

Some unscrupulous investors, in instance, might purposefully write and distribute false information in order to manipulate the stock price. Individual investors may be harmed as a result of such a capitulation bottom. As a result, filtering out false information and anticipating true signals is important. As a result, based on a vast quantity of opinion data, we presented an algorithm for recognising risk signs early. The suggested method is focused on the stock market in this study, but it may be extended to other contexts where human behaviour is heavily affected, such as social media commerce. Although prior research has concentrated on predicting the rise and fall of stock prices, our findings assist stock investing decisions. The suggested algorithm may be used by both individuals and businesses, allowing the government to deal with financial problems on a national scale.

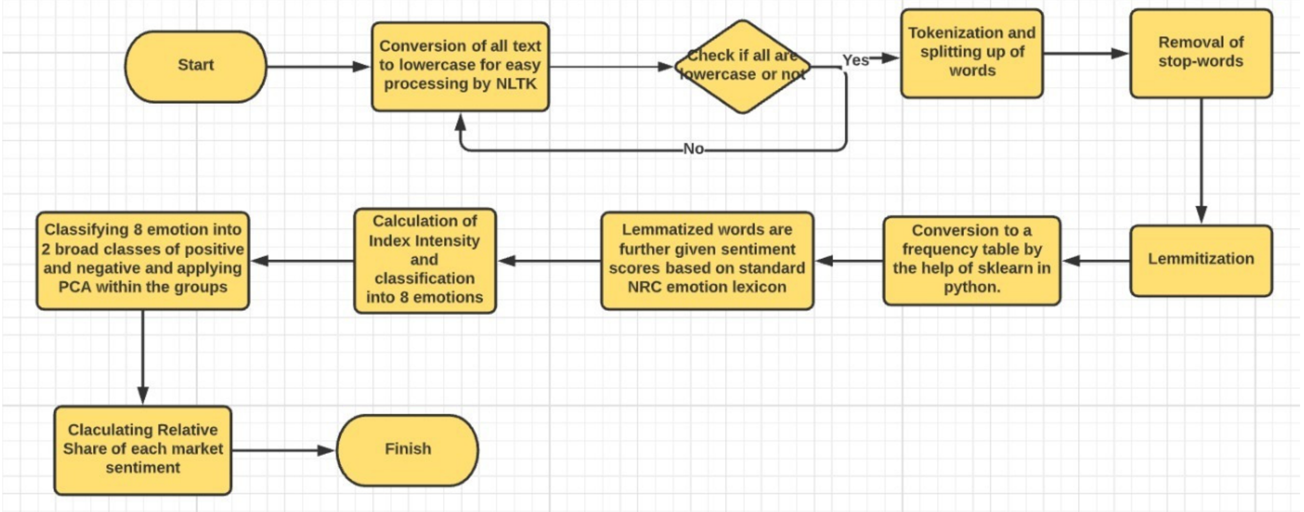
Your findings: (possible alternate for the solution proposed)

- Using a panel regressing model
- Using a semi-supervised learning model

3.

TABLE 2	
Problem answered in this paper. (1-2 lines)	There is a 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 Indian Stock market using investor sentiment analysis.
Detailed description about the problem (5-8 lines)	The recent literature in the behavioural 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 behaviour 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.
Why that problem is chosen in this paper? Scope of the problem and solution (*Refer Introduction*) (5-8 lines)	Economic agents make decisions based on their expectations about the economy's future state since the economic system is an expectation feedback system. These decisions have an impact on the actual manifestation of economic variables, resulting in new expectations. By understanding the importance of sentiment we can make better decisions by understanding the expectations through sentiment
History of the problem. (*Refer Introduction*) (8-10 lines)	The traditional financial modeling of market volatility using time series analysis is experiencing a regime change in the domain and dominance of behavioral finance. It is believed that mean and variance is believed to be the sole factory that plays a decisive role in shaping the decision that the investors tend to take. But now the impact of noise traders has been found to impact the stock market. It was believed that the transitory influence of noise traders has been eradicated by rational investors but now this is being challenged by many researchers. A link is being found between financial trading and prevailing market sentiment.
List of the related/similar problems (*Refer Related work*) – Describe each with proposed solutions	
Related problem 1 – Describe (3-4 lines)	Market feelings, according to this article, lie at the heart of every financial data analysis. There is a clear gap between examining financial data in terms of volatility vs analysing financial data in terms of market sentiments. The former is an established and widely used strategy, whereas the latter is a proposed strategy.
Paper in APA style	Moseki, K. K., & Madhava Rao, K. S. (2017). Analysing stock market data—Market sentiment approach and its measures. *Cogent Economics & Finance*, *5*(1), 1367147.

Related problem 2 – Describe **(3-4 lines)**	The study examines market attitudes in currency rates, a topic of great interest to both individual traders and institutional investors. To reflect the uncertainties in market attitudes, a multinomial probability model is created.
Paper in APA style	Rao, K. M., & Ramachandran, A. (2014). Exchange rate market sentiment analysis of major global currencies. *Open Journal of Statistics*, *2014*.
Related problem 3 – Describe **(3-4 lines)**	The goal of this study is to see if assessments of collective emotional states collected from large-scale Twitter feeds are related to the Dow Jones Industrial Average (DJIA) value over time. We use two mood tracking tools to examine the text content of daily Twitter feeds: OpinionFinder, which measures positive vs. negative mood, and Google-Profile of Mood States, which measures positive vs. negative mood (GPOMS)
Paper in APA style	Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, *2*(1), 1-8.
Related problem 4 – Describe **(3-4 lines)**	In this paper it is shown that information derived from news sources is better at forecasting the direction of underlying asset volatility movement, or its second order statistics, than its price movement direction.
Paper in APA style	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.
Related problem 5 – Describe **(3-4 lines)**	The article demonstrates how emotions influence investor assumptions and can be translated into price movements. The emotional investor was defined as a type of agent who solely relies on his intuition and whose presence has an impact on market values. As a result, there is no doubt that an acceptable rational strategy necessitates the adoption of a new type of market agent, and the theoretical considerations offered in this study may aid this process.
Paper in APA style	Kuzmina, J. (2010). Emotion's component of expectations in financial decision making. *Baltic Journal of Management*.
What is the proposed solution in this paper for the problem chosen? (*Refer Proposed work*) **(5-8 lines)**	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 covert 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

<table>
<tr><td></td><td>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 where 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.</td></tr>
<tr><td>Architecture of the proposed solution. (Refer proposed work) Diagram</td><td>
</td></tr>
<tr><td>Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)</td><td>NLTK approach with Volatility Modelling</td></tr>
<tr><td>List of existing algorithms used by the authors to complete the proposed work. (1-2 lines for each algorithm)</td><td>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</td></tr>
<tr><td>List of datasets used. (Refer</td><td>The datasets that are used in the above study are the prominent web sources for Indian financial</td></tr>
</table>

experimental evaluation/result discussion) **(3-4 lines)**	market and business. The web sources that are used in the given study include data from Reuters India (Business and Economic section), Livemint business news, The Hindu Business Line and Moneycontrol.com.
List of equations that are very well applied in this problem domain	Equation 1 $$return = \ln(\frac{P_t}{P_{t-1}}) * 100$$ Description: - This formula calculates the return of Sensex. P_t is sensex index on day t. Equation 2: - $$IE_{e,t} = \sum_{\forall w} (f_{w,t})(I_{w,e})$$ Description: - Intensity index for each emotion is calculated by the above formula. IEe,t refers to calculated index of emotion e on day t, fw,t refers to frequency of occurrence of word w on day t and Iw,e refers to intensity of emotion e associated with word w. Equation 3: - $$POS = \frac{S_p}{S_p + S_n}$$ Description: - This formula is used to calculate the relative share of each market sentiment. POS refers to share of positive sentiment on a particular day, S_p refers to score of positive sentiment for that day and S_n refers to score of negative sentiment for that day. Equation 4: - NEG=1−POS Description: - Negative market sentiment is calculated by the above formula. Equation 5: - $$Y_t = \mu + \alpha * Y_{t-1} + \beta * \varepsilon_{t-1} + \varepsilon_t$$

	Description: - This is the mean equation of a standard GARCH model. $YYtt$ is return of Sensex and follows an ARMA (1,1) process and ε_t is error term in model. μ, α, β are parameters in mean equation and t suffix represents time. Equation 6: - $$h_t = \rho + c*\varepsilon^2_{(t-1)} + \delta*h_{(t-1)} + \gamma*\varepsilon^2_{(t-1)}*D_{(t-1)}$$ Description: - This is the conditional volatility equation and this equation follows a deterministic path. h_t denotes the deterministic path and depends on its own past as well as past error variance ε^2_{t-1}. a dummy variable (D_{t-1}) is introduced to assess the asymmetric impact of error. Coefficient γ signifies asymmetric impact of error on conditional volatility. Equation 7: - $$h_t = c*\varepsilon^2_{(t-1)} + \delta*h_{(t-1)} + \gamma*\varepsilon^2_{(t-1)}*D_{(t-1)} + \varphi_1*POS + \varphi_2*NEG$$ Description: - This is the augmented conditional volatility equation. Here the constant term of above equation is purposefully dropped since POS and NEG are linearly related and inclusion of the constant will lead to omission of either of these terms during estimation.
List of method(s)/metrics used to evaluate the proposed approach. (*Refer experimental evaluation/result discussion*) **(5-8 lines)**	Three different generalized auto regressive conditional heteroscedasticity (GARCH) models are used to analyze impact of market sentiments. Standard GARCH model along with two variants of it, namely, EGARCH or exponential GARCH model and GJR-GARCH model have been employed. GARCH models have gained popularity for conditional volatility but simple GARCH models fails to show the asymmetric impact of volatility. Distinct impact of positive and negative shocks led to conceptualization of leverage effect and this in turn causes extension to GARCH model. Recent trend to capture market sentiment through news articles, online information, etc. has been gathering attention of researchers. 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. In the augmented conditional volatility equation, (c+δ) is less then unity which confirms stability of the model and since (c+δ) is close to unity we can infer there exists volatility persistence in market. Coefficient φ_1 shows relative influence of positive sentiment on conditional volatility and it is found to bear a negative sign whereas parameter φ_2 measures relative importance of negative market

	sentiment on market volatility and found to be positive.
What are the similar approaches with which the proposed approach is compared? (Refer experimental evaluation/result discussion) **Explain each of these approach** **(3-4 lines)**	Approach/method 1: - A standard GARCH model is used for conditional volatility. It can successfully estimate the conditional variance in the model but it fails to depict the asymmetric impact of volatility. A standard GARCH(1,1) model consists of two equations, the mean equation and conditional volatility equation Approach/Method 2: - A GJR-GARCH model follows the deterministic path as the conditional volatility. Also a dummy variable is used in order to assess the asymmetric impact. Though traditional assumption of this GJR GARCH model to consider positive error ($\varepsilon_{t-1} > 0$) as a proxy for good news and its counterpart as bad news in market has been proven empirically very successful in literature but recent trend to capture market sentiment through news articles, online information, etc. has been gathering attention of researchers.
How the results of proposed approach are compared with other similar approaches? (*Refer experimental evaluation/result discussion*)	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.
Advantages/merits of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	When we analyse the finding it is very much clearly evident that the noise traders play a dominant role. An advantage of finding is use of 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 mean equation and instead of that the model has generated separate market sentiments on conditional volatility of Indian financial market. This approach is better and more appealing in today's financial market.
Disadvantages/limitations of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	The first disadvantage of the proposed work is that only 8 emotions are used in which they are divided into 2 classes only. The number of emotions should be increased another broader class known as neutral can be added in order to analyse the sentiment. Future scope of this study lies in a comparative analysis of different sectors of Indian stock market like energy, telecommunication or metal.
Future work as stated by authors (*Refer conclusion / result discussion / experimental evaluation*)	As per the authors, the future scope of this study lies in a comparative analysis of different sectors on Indian stock market like the different areas of stock companies like energy, telecommunication and metal. There can be assessed and the respective impact on the overall stock market can be analysed to have a better understanding of the dynamics in the marketing sector.

Your one page write-up about this paper
Traditional empirical models use macroeconomic fundamentals or financial indicators to analyze the impact of sentiment on financial market fluctuations. This article uses the latest market sentiment analysis method based on the text from economic and financial market-related news articles. Two different market sentiments, positive and negative sentiments, are constructed using different sentiments identified by standard natural language processing methods. In addition, Document aims to propose an enhanced version of the asymmetric GARCH conditional volatility model for the Indian Stock Exchange Sensex from April 19, 2007, to January 10, 2020, which contains the aforementioned market sentiment. The empirical results show that negative market sentiment has a dominant influence on positive sentiment, and it also provides evidence of noisy transactions in the financially immature Indian stock market.
Your findings: (possible alternate for the solution proposed) •