

Research Questionnaire

Note:

- 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	19BCE0496 - Harsh Vivek Londhekar
Team No.	7
Paper Title	<ol style="list-style-type: none">1. Estimating the relative effects of raw material prices, sectoral outlook and market sentiment on stock prices2. Exploiting textual and relationship information for fine-grained financial sentiment analysis3. Harvesting social media sentiment analysis to enhance stock market prediction using deep learning
Citation (APA style)	<ol style="list-style-type: none">1. 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, 102158.2. Daudert, T. (2021). Exploiting textual and relationship information for fine-grained financial sentiment analysis. Knowledge-Based Systems, 107389.3. 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.

TABLE 2

Problem answered in this paper. (1-2 lines)	To determine the relation between movement of raw material prices and share price of the company.
Detailed description about the problem (5-8 lines)	To determine the effect of raw material prices in different time periods, outlook of different sectors in the market and market sentiment on the share prices of the company. 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.
Why that problem is chosen in this paper? Scope of the problem and solution (Refer Introduction) (5-8 lines)	The particular problem is selected because the task of predicting share prices of a company requires strenuous efforts because of the sensitive financial market, external chaos, political influence, natural calamities and much more. The selected research paper contributes to relevant literature by combining or uniting wavelet analysis and machine learning to determine the relation between the movement of raw materials and share prices of a company, which affects the end consumer. The scope of the problem is limited to the following sectors : Oil and Gas, Metal, FMCG and Healthcare. The suggested methods and frameworks mentioned in the paper can be used to determine the required relation for other sectors as well.
History of the problem. (Refer Introduction) (8-10 lines)	As per the micro economics rules, the prices of the raw material affects the end consumer, and this depends on the demand in the market and position of the company amongst their competitions. Increase in raw material prices also affects the sector's outlook. Stock prices of the company reflects the market view of the current working of the organisations and furthermore their future performances. Stock prices of a company get affected in two ways: through sectoral outlook and decrease in profitability. Moreover influence of raw materials over different time horizons has never been surveyed properly, this research paper helps to deal with this problem and contributes to literature by combining raw materials and share prices of a company.
List of the related/similar problems (Refer Related work) – Describe each with proposed solutions	

Related problem 1 – Describe (3-4 lines)	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 characteristic of stock market over different time periods.
Paper in APA style	Sen, J., & Chaudhuri, T. D. (2018). Understanding the sectors of Indian economy for portfolio choice. <i>International Journal of Business Forecasting and Marketing Intelligence</i> , 4(2), 178-222.
Related problem 2 – Describe (3-4 lines)	To derive economic profit of Indian companies over two different time periods and two assess the performance change in the company and sector to which they belong. And also to compare how Indian companies have worked over these two periods and see if any structural changes are there or not.
Paper in APA style	JHUNJHUNWALA, A., CHAUDHURI, T. D., & BHAMRAH, G. K. (2019). Value creation by Indian companies: A comparative study over two time periods. <i>Turkish Economic Review</i> , 6(1), 44-61.
Related problem 3 – Describe (3-4 lines)	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 also assess the strategies formulation by these companies.
Paper in APA style	Jhunjhunwala, A., & Chaudhuri, T. D. (2021). Innovation, growth and value creation: a study of Indian companies. <i>International Journal of Business Innovation and Research</i> , 25(3), 328-352.
Related problem 4 – Describe (3-4 lines)	To analyse the change in value of Bitcoin returns, change in Gold price returns, US Stock Market returns, interest rates, crude oil and rapid change in American Stock Market. To make the outcomes more robust , two different time periods have been taken into consideration.
Paper in APA style	Jareño, F., de la O González, M., Tolentino, M., & Sierra, K. (2020). Bitcoin and gold price returns: a quantile regression and NARDL analysis. <i>Resources Policy</i> , 67, 101666.

Related problem 5 – Describe (3-4 lines)	To combine the cross sectional variation in stock returns related to the market B, size, leverage, book-to-market equity, and earnings-price ratios. Moreover, when the tests allow for variation in B that is unrelated to size, the relation between market B and average return is flat, even when B is the only explanatory variable.																																																															
Paper in APA style	Fama, E. F., & French, K. R. (2021). <i>The cross-section of expected stock returns</i> (pp. 349-391). University of Chicago Press.																																																															
What is the proposed solution in this paper for the problem chosen? (Refer Proposed work) (5-8 lines)	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 short, medium and long period of time. Certain machine learning algorithms like Random forest, gradient boosting and genetic algorithm have also been used to determine the rank of the three factors mentioned above over different time periods.																																																															
Architecture of the proposed solution. (Refer proposed work) Diagram	<p>The figure consists of three separate line graphs, each plotting Wavelet Correlation on the y-axis against Wavelet Scale on the x-axis (with values 1, 2, 4, 8, 16, 32). Each graph contains two lines: an upper line labeled 'U' and a lower line labeled 'L'. In all three cases, the correlation generally increases as the wavelet scale increases, with the 'U' line consistently showing higher correlation than the 'L' line across most scales.</p> <table border="1"> <caption>Reliance Stock Price and Crude Oil Price</caption> <thead> <tr> <th>Wavelet Scale</th> <th>U (Wavelet Correlation)</th> <th>L (Wavelet Correlation)</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.05</td><td>-0.10</td></tr> <tr><td>2</td><td>0.08</td><td>-0.12</td></tr> <tr><td>4</td><td>0.08</td><td>-0.15</td></tr> <tr><td>8</td><td>0.08</td><td>-0.20</td></tr> <tr><td>16</td><td>0.25</td><td>-0.25</td></tr> <tr><td>32</td><td>0.35</td><td>-0.30</td></tr> </tbody> </table> <table border="1"> <caption>Reliance Stock Price and Oil & Gas Index</caption> <thead> <tr> <th>Wavelet Scale</th> <th>U (Wavelet Correlation)</th> <th>L (Wavelet Correlation)</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.05</td><td>-0.05</td></tr> <tr><td>2</td><td>0.15</td><td>-0.10</td></tr> <tr><td>4</td><td>0.45</td><td>0.20</td></tr> <tr><td>8</td><td>0.65</td><td>0.45</td></tr> <tr><td>16</td><td>0.75</td><td>0.55</td></tr> <tr><td>32</td><td>0.78</td><td>0.50</td></tr> </tbody> </table> <table border="1"> <caption>Reliance Stock Price and SENSEX</caption> <thead> <tr> <th>Wavelet Scale</th> <th>U (Wavelet Correlation)</th> <th>L (Wavelet Correlation)</th> </tr> </thead> <tbody> <tr><td>1</td><td>0.45</td><td>0.40</td></tr> <tr><td>2</td><td>0.55</td><td>0.45</td></tr> <tr><td>4</td><td>0.55</td><td>0.40</td></tr> <tr><td>8</td><td>0.45</td><td>0.35</td></tr> <tr><td>16</td><td>0.75</td><td>0.65</td></tr> <tr><td>32</td><td>0.20</td><td>0.45</td></tr> </tbody> </table>	Wavelet Scale	U (Wavelet Correlation)	L (Wavelet Correlation)	1	0.05	-0.10	2	0.08	-0.12	4	0.08	-0.15	8	0.08	-0.20	16	0.25	-0.25	32	0.35	-0.30	Wavelet Scale	U (Wavelet Correlation)	L (Wavelet Correlation)	1	0.05	-0.05	2	0.15	-0.10	4	0.45	0.20	8	0.65	0.45	16	0.75	0.55	32	0.78	0.50	Wavelet Scale	U (Wavelet Correlation)	L (Wavelet Correlation)	1	0.45	0.40	2	0.55	0.45	4	0.55	0.40	8	0.45	0.35	16	0.75	0.65	32	0.20	0.45
Wavelet Scale	U (Wavelet Correlation)	L (Wavelet Correlation)																																																														
1	0.05	-0.10																																																														
2	0.08	-0.12																																																														
4	0.08	-0.15																																																														
8	0.08	-0.20																																																														
16	0.25	-0.25																																																														
32	0.35	-0.30																																																														
Wavelet Scale	U (Wavelet Correlation)	L (Wavelet Correlation)																																																														
1	0.05	-0.05																																																														
2	0.15	-0.10																																																														
4	0.45	0.20																																																														
8	0.65	0.45																																																														
16	0.75	0.55																																																														
32	0.78	0.50																																																														
Wavelet Scale	U (Wavelet Correlation)	L (Wavelet Correlation)																																																														
1	0.45	0.40																																																														
2	0.55	0.45																																																														
4	0.55	0.40																																																														
8	0.45	0.35																																																														
16	0.75	0.65																																																														
32	0.20	0.45																																																														
Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)	Discrete Wavelet Transformation (MODWT) Complete Wavelet Transformation (CWT)																																																															

<p>List of existing algorithms used by the authors to complete the proposed work. (1-2 lines for each algorithm)</p>	<p>Random Forest : It is a machine learning algorithm used to solve regression and classification problems.</p> <p>Gradient Boosting : It is based on Greedy algorithm and can overfit a dataset quickly. It penalise various parts of algorithm and improve the performance by reducing overfitting.</p> <p>Genetic Algorithm : It is a method for solving constrained and unconstrained optimisation problems which are based on natural selection process.</p>
<p>List of datasets used. (<i>Refer experimental evaluation/result discussion</i>) (3-4 lines)</p>	<p>Data set related to the stock prices have been taken from the following paper : https://journals.sagepub.com/doi/10.1177/2321022217724978 And BSE(Bombay stock exchange) and NSE(National Stock Exchange) websites.</p>
<p>References/links to each of the dataset used in this paper (in APA style)</p>	<ol style="list-style-type: none"> 1. https://journals.sagepub.com/doi/10.1177/2321022217724978 2. https://www.bseindia.com 3. https://www.nseindia.com
<p>Why the above dataset(s) used? <i>(Refer experimental evaluation/ result discussion)</i> (3-4 lines)</p>	<p>Since the eight companies mentioned are Indian companies, all the information related to the stock prices of these companies have been taken from bse and nse websites, which are the best resources for such data as they update it regularly and are the most reliant sources for stock prices.</p>

	<p>Equation 1: CWT of a time series</p> $W_x(\tau, s) = \frac{1}{\sqrt{ s }} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t - \tau}{s} \right) dt$ <p>Description: Continuous wavelet transform based coherence analysis</p> <p>Equation 2: MODWT estimator of wavelet correlation</p> $\rho_{xy}(\lambda_j) = \text{Corr}\left(w_{ijt}, \tilde{w}_{ijt}\right) = \frac{\text{Cov}\left(w_{ijt}, \tilde{w}_{ijt}\right)}{\sqrt{\text{Var}(w_{ijt}) \text{Var}(\tilde{w}_{ijt})}}$ <p>Description: The MODWT estimator of wavelet correlation has been used in this study which basically considers the covariance of two series ($x(t)$, $y(t)$) and wavelet variance of individual series.</p>
<p>List of method(s)/metrics used to evaluate the proposed approach. <i>(Refer experimental evaluation/result discussion)</i> (5-8 lines)</p>	<p>Boruta Algorithm Random Forest Gradient Boosting Genetic Algorithm Monto Carlo Methods for deriving significance Y axis - Period: used on a scale of weeks (2 weeks to 512 weeks) X axis - Timeline: scale 1 unit = 500</p>
<p>List of supporting tools/concepts (3-4 lines)</p>	tsfeatures' package of R, Jarque-Bera Test , ADF Test , ZA Test, ARCH statistics, SpectralEntropy, Hurst Exponent

<p>What are the similar approaches with which the proposed approach is compared? (Refer experimental evaluation/result discussion)</p> <p>Explain each of these approach (3-4 lines)</p>	<p>Approach/method 1: FTT transform (Fast Fourier Transform). In FTT is an important measurement method in the science of audio and acoustic measurements. It converts the signal into individuals spectral components and thereby provide frequency information about the signal.</p>
<p>How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)</p>	<p>There are many packages available for wavelet analysis like Mathematica, Maple, Matlab, R, and PyWavelets. Which is not the case with FTT.</p> <p>The wavelet transform improves on the Fourier Transform in that it can analyse a signal by time and frequency simultaneously, thereby easily recovering localised signal information.</p> <p>The results which were obtained by wavelet analysis were as follows:</p> <ol style="list-style-type: none"> 1. In the short run company stock prices are not affected by raw material. 2. In terms of petroleum prices, overall sentiment is affected by prices of crude oil. 3. Company stock prices are more affected by society view point. 4. In the medium and long run time, sentiment and society viewpoint are highly correlated.
<p>Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)</p>	<ol style="list-style-type: none"> 1. One of the main advantage of wavelet is that they offer a simultaneous localisation in time and frequency domain 2. The use of larger DWT basis functions or wavelets filters produces blurring and ringing noise near edge regions in images or video frames 3. The second main advantage of wavelet is that, using fast wavelet transform it is computationally very fast. 4. Provide a way for analysing web forms in both frequency and duration. 5. Representation of functions that have discontinuities and sharp pics. 6. Accurately deconstructing and reconstructing finite, non-periodic and non-stationary signals. 7. Allow signals to be stored more efficiently than by Fourier transform.

<p>Disadvantages/limitations of proposed solution in your view. <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<ol style="list-style-type: none"> 1. A poor directional selectivity for diagonal features, because the wavelet filters are separable and real. 2. Longer compression time. 3. And lack of shift invariance, in which means that small shifts in the input signal can cause major variations in the distribution of energy between DWT coefficients at different scales. 4. The cost of computing DWT as compared to DCT maybe higher.
<p>Future work as stated by authors <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<p>Presently the scope of this research paper is limited to market sectors namely : Oil and Gas, Metal, FMCG and Healthcare. By the proposed frameworks the study can include different sectors of the market too in the future.</p>
<p>Your one page write-up about this paper</p>	

Recognisable proof of key determinants answerable at driving stock costs across the world is of central reasonable significance. The undertaking is amazingly strenuous attributable to affect-ability of monetary business sectors to macroeconomic shocks, outside tumult, political shakiness and regular disasters. In this work, exertion has been made to fundamentally assess the impact of crude material costs, sectoral viewpoint, and market feeling on stock costs at a granular level in the Indian setting. The examination resorts to wavelet investigation and machine learning models to gauge time fluctuating reliance and logical capacities of particular builds. Wavelet analysis and machine learning models have been used to compare the relationship between movement of raw materials and stock prices of a particular company. This observation has been made over different time periods that is short, medium and long time periods. Other machine learning algorithms like random forest, gradient boosting and genetic algorithms are used for ranking three features which are raw material prices, sectoral outlook and market sentiment.

The Indian companies which were analysed are: Reliance, Aurobindo, Arvind stock price and cotton price,Nahar spinning and cotton, Hindalco stock price, Divi's laboratory and crude oil , National aluminium and aluminium.

The scope of this research is limited to market sectors like oil and gas, metal, FM CG, healthcare. In future using the same framework we can extend the market sector to other domains also.

The results which were obtained by wavelet analysis were as follows:

1. In the short run company stock prices are not affected by raw material.
2. In terms of petroleum prices, overall sentiment is affected by prices of crude oil.
3. Company stock prices are more affected by society view point.
4. In the medium and long run time, sentiment and society viewpoint are highly correlated.

Your findings: (possible alternate for the solution proposed)

1. Possible solution could be by using FTT ie Fourier transform but it is not better than wavelet transform which is used in the paper.

TABLE 2	
Problem answered in this paper. (1-2 lines)	Identification of expressions (positive neutral negative) towards the subject by way of expressing sentiments in text. Novel approach to capture implicit sentiments and the contagion process.
Detailed description about the problem (5-8 lines)	<p>In a financial firm, sentiment could be read across platforms like company, analyst reports, news articles and blogs. The aim is to capture implicit sentiments and the contagion process. To apply the solution of sentiment analysis across multiple domains and text types, such as product reviews.</p> <p>To demonstrate the impact of implicit sentiment as well as importance of different relationship or sentiment prediction on company and analyst reports news articles and blogs.</p>

<p>Why that problem is chosen in this paper? Scope of the problem and solution (<i>Refer Introduction</i>) (5-8 lines)</p>	<p>This problem is chosen in order to showcase that textual context can be modelled lithograph, to study and gain further insight into sentiment analysis and improve it. Future scope of this paper includes exploring of dynamic vertex like GraphSAGE. This will reduce re-calculation of vertex representation for the entire graphs. Different classifiers can be used to further optimise the performance.</p>
<p>History of the problem. (<i>Refer Introduction</i>) (8-10 lines)</p>	<p>Sentiment analysis is a way to identify the sentiments expressed by people in the form of text and to determine whether the sentiments are positive neutral or negative. Implicit sentiment appear in a general rather than a source like a newspaper. Thereby detecting sentiments in text that do not clearly express emotions or do not contain such words are difficult to process. Does we can improve sentiment analysis by historical context, this approach will analyse the interpretation of text and process it.</p> <p>The role of sentiment contagion is not considered. Sentiment contagion means individuals interacting with each other and capturing their sentiments to form an opinion about each other. Therefore by using machine learning models and other frameworks we can convert normal text and derive the sentiment expressed by the text to analyse whether the expression is positive, neutral or negative.</p>
<p>List of the related/similar problems (<i>Refer Related work</i>) – Describe each with proposed solutions</p>	
<p>Related problem 1 – Describe (3-4 lines)</p>	<p>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 potential application in behavioural science.</p>
<p>Paper in APA style</p>	<p>Daudert, T. (2021). A multi-source entity-level sentiment corpus for the financial domain: the FinLin corpus. <i>Language Resources and Evaluation</i>, 1-24.</p>
<p>Related problem 2 – Describe (3-4 lines)</p>	<p>To measure qualitative information that predicts returns of a company using news articles and other sentimental data. To form a relation between under reaction of the market and content of news articles.</p>

Paper in APA style	Sinha, N. R. (2016). Underreaction to news in the US stock market. <i>Quarterly Journal of Finance</i> , 6(02), 1650005.
Related problem 3 – Describe (3-4 lines)	To find a way to represent and encode, graph structure so that it can be easily used by machine learning models. Currently machine learning models rely on user defined heuristics to extract the different features and encode structural information.
Paper in APA style	Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Representation learning on graphs: Methods and applications. <i>arXiv preprint arXiv:1709.05584</i> .
Related problem 4 – Describe (3-4 lines)	To find a approach for learning latent representations of vertices in a network. To use language modelling and deep learning in order to change sequence of words into graphs. The above-mentioned latent representations encode social relations in a continuous vector space, which is easily exploited by statistical models.
Paper in APA style	Perozzi, B., Al-Rfou, R., & Skiena, S. (2014, August). Deepwalk: Online learning of social representations. In <i>Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining</i> (pp. 701-710).
Related problem 5 – Describe (3-4 lines)	Graphs process set of flat vectors, this method loses important topological information and the results which are heavily dependent on pre-processing. The objective is to directly process graphs of different kinds like directed, and directed, labelled and cycling graphs.
Paper in APA style	Gori, M., Monfardini, G., & Scarselli, F. (2005, July). A new model for learning in graph domains. In <i>Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005</i> . (Vol. 2, pp. 729-734). IEEE.
What is the proposed solution in this paper for the problem chosen? (Refer Proposed work) (5-8 lines)	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).

<p>Architecture of the proposed solution. (Refer proposed work)</p> <p>Diagram</p>	<p>Fig. 3. Overview of the graph structure of the record presented in Fig. 1. The vertices are represented as circles and the edges as arrows. V^T of the type stocktwit and news articles appear in orange and dark green, respectively. V^W and V^E are presented in yellow and cyan, respectively, and V^Z are colored in blue, brown, purple, and red depending on the type of temporal class, as specified in Section 3.4.</p>
<p>Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)</p>	<p>Fine Tuned Textual Representation FFNN (feed forward neural network)</p>

<p>List of existing algorithms used by the authors to complete the proposed work.</p> <p>(1-2 lines for each algorithm)</p>	<p>1. 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.</p>
<p>List of datasets used. (Refer experimental evaluation/result discussion)</p> <p>(3-4 lines)</p>	<p>1. IMDB dataset 2. Yelp dataset</p>
<p>References/links to each of the dataset used in this paper (in APA style)</p>	<p>1. IMDB dataset : https://www.imdb.com/interfaces/ 2. Yelp dataset : https://www.kaggle.com/yelp-dataset/yelp-dataset</p>
<p>Why the above dataset(s) used?</p> <p>(Refer experimental evaluation/result discussion)</p> <p>(3-4 lines)</p>	<p>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).</p>

<p>List of equations that are very well applied in this problem domain</p>	<p>Equation 1: Equation to calculate mean squared error (MSE) :</p> $MSE(P, T) = \frac{1}{n} \sqrt{\sum_{i=1}^n (T_i - P_i)^2}$ <p>Description: mean square error measures the average of the squares of the error.</p> <p>Equation 2: Equation to calculate Cosine Similarity(CS)</p> $CS(T, P) = \frac{\sum_{i=1}^n T_i \times P_i}{\sqrt{\sum_{i=1}^n T_i^2} \times \sqrt{\sum_{i=1}^n P_i^2}}.$ <p>Description : cosine similarity is the cosine of the angle between to N-dimensional vectors in an N-dimensional space.</p>
<p>List of method(s)/metrics used to evaluate the proposed approach. <i>(Refer experimental evaluation/ result discussion)</i> (5-8 lines)</p>	<ol style="list-style-type: none"> 1. Mean squared error (MSE) 2. Cosine Similarity (CS)
<p>List of supporting tools/concepts (3-4 lines)</p>	<p>SOTA Single Text sentiment analysis Contextual sentiment analysis Deep Learning Implicit sentiments</p>

<p>What are the similar approaches with which the proposed approach is compared? <i>(Refer experimental evaluation/result discussion)</i></p> <p>Explain each of these approach (3-4 lines)</p>	<p>Approach/method 1: Fine-tuned BERT (Fine-Tuned BERT). This is designed by pre training deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context.</p> <p>Approach/method 2: FinBERT (FinBERT) . FinBERT is a pre-trained NLP model to analyse sentiment of financial text. This is modified version of Bert language model in the finance domain by training it further more than the regular Bert.</p> <p>Approach/method 3: aFFNN</p>
<p>How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)</p>	<p>The improvement from baseline was seen from 0.68 - 15.1% and 59.46–234.15% for the CS and MSE, respectively.</p> <p>A comparison of fine tuned BERT and fine tuned FFNN , showed that FFNN outperforms fine tuned BERT. Fine tuned FFNN had the best performance as compared to the other two models.</p>
<p>Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)</p>	<ol style="list-style-type: none"> Problems in FFNN are represented by attribute-value pairs. These learning methods are quite robust to noise in the training data. The training examples may contain errors, which do not affect the final output. It is used where the fast evaluation of the learned target function required.
<p>Disadvantages/limitations of proposed solution in your view. <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<ol style="list-style-type: none"> It is highly dependent on hardware. Lack of assurance of proper network structure. The difficulty to show the problem to the network.
<p>Future work as stated by authors <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<p>Future scope of this paper includes exploring of dynamic vertex like GraphSAGE. This will reduce re-calculation of vertex representation for the entire graphs.</p> <p>Different classifiers can be used to further optimise the performance.</p>
<p>Your one page write-up about this paper</p>	

The selected paper's objective is Identification of expressions (positive neutral negative) towards the subject by way of expressing sentiments in text. Novel approach to capture implicit sentiments and the contagion process.

In a financial firm, sentiment could be read across platforms like company, analyst reports, news articles and blogs. The aim is to capture implicit sentiments and the contagion process. To apply the solution of sentiment analysis across multiple domains and text types, such as product reviews.

This paper also aims to demonstrate the impact of implicit sentiment as well as importance of different relationship or sentiment prediction on company and analyst reports news articles and blogs. The following paper gave a novel idea to approach sentiment analysis as a multi text problem. Textual context can be modelled with a graph and provided a method for this. The results showed that contextualised fine- grained sentiment analysis outperforms several approaches that employ SOTA single-text sentiment analysis. The paper showed an ablation study to gain further insight into the contribution of individual parts of the graph toward improvement of the sentiment analysis.

This problem is chosen in order to showcase that textual context can be modelled lithograph, to study and gain further insight into sentiment analysis and improve it.

This paper saw the challenge of fine grained sentiment analysis of financial texts. The proposed solution saw an improvement of 15% and 234% relative to the baselines. Future scope of this paper includes exploring of dynamic vertex like GraphSAGE. This will reduce re-calculation of vertex representation for the entire graphs.

Different classifiers can be used to further optimise the performance.

Your findings: (possible alternate for the solution proposed)

Possible alternatives for FFNN can be words Bayes or Markov algorithm .

TABLE 2	
Problem answered in this paper. (1-2 lines)	To identify how movements in a company's stock prices correlate with expressed opinions of the public regarding that company.
Detailed description about the problem (5-8 lines)	To make a stock price prediction tool which considers public sentiment and also other parameters. Data will be gathered from social networking sites like Twitter, Facebook, Google plus etc. Social networking sites perfectly reflects People's opinion on a particular company or a particular news. It is found from a survey that financial news have impact on stock prices of a particular company.

<p>Why that problem is chosen in this paper? Scope of the problem and solution (Refer Introduction) (5-8 lines)</p>	<p>The current stock market is affected by social mood and historical prices, and also by people sentiment which play a major role in moving the prices of stock prices of different companies. The use of news articles, social media like Twitter, Facebook explains how a company performance 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 stocks accuracy so that the user can buy or sell stock of a particular company. There are future opportunities for research in this area.</p>
<p>History of the problem. (Refer Introduction) (8-10 lines)</p>	<p>Opinions of others play a very important role in an individual person's decision making. People who use social media sites like Facebook or Twitter can communicate with other people and tell their opinion on a specific topic like a news, company, sports and many more. This data can be usefully extracted to predict the future of a particular object or a topic, this method can also be used to predict a company's stock prices and is known as sentiment analysis. The analysis of sentiment and emotions, is the study of opinions, thoughts, experiences, feelings and behaviours in the text form. The entire market is dependent on sentiments of people and the society and the sentiment analysis can be used to predict the share market.</p>
<p>List of the related/similar problems (Refer Related work) – Describe each with proposed solutions</p>	
<p>Related problem 1 – Describe (3-4 lines)</p>	<p>To match on opinion mining or feeling assessment which is an area of them data mining and machine learning. Dr active is to utilise and examine different ML and lexicon investigation methodologies. And two analyse a study and check estimation of the present composition.</p>
<p>Paper in APA style</p>	<p>Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies, practices and applications. <i>International Journal of Scientific and Technology Research</i>, 9(2), 601-609.</p>
<p>Related problem 2 – Describe (3-4 lines)</p>	<p>Calculating value of some company and the price of their stock and evaluating the long-term return is not very easy because people have to take into consideration many factors. No objective is to predict stock value of a company over a long period of time using machine learning models. To implement a ML model which can evaluate equities future price over the long time.</p>

Paper in APA style	Milosevic, N. (2016). Equity forecast: Predicting long term stock price movement using machine learning. <i>arXiv preprint arXiv:1603.00751</i> .
Related problem 3 – Describe (3-4 lines)	To develop an innovative neural network approach to achieve better stock market predictions. To propose the deep long short term memory neural network LSTM with embedded layer and long short term memory neural network to predict the stock market.
Paper in APA style	Pang, X., Zhou, Y., Wang, P., Lin, W., & Chang, V. (2020). An innovative neural network approach for stock market prediction. <i>The Journal of Supercomputing</i> , 76(3), 2098-2118.
Related problem 4 – Describe (3-4 lines)	To make financial decisions such as stock market prediction, to predict the prices of companies stock using Twitter data.
Paper in APA style	Das, S., Behera, R. K., & Rath, S. K. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. <i>Procedia computer science</i> , 132, 956-964.
Related problem 5 – Describe (3-4 lines)	To find the potential use of textual information from user generated blogs and to predict the stock market. Also to correlate movements of stock prices and social media content and thereby in for the relation between social media and stock price of a company.
Paper in APA style	Sun, A., Lachanski, M., & Fabozzi, F. J. (2016). Trade the tweet: Social media text mining and sparse matrix factorization for stock market prediction. <i>International Review of Financial Analysis</i> , 48, 272-281.
What is the proposed solution in this paper for the problem chosen? (Refer Proposed work) (5-8 lines)	Using Deep Learning Model and LSTM(Long short term memory) a reliable predictive model for stock movement is build. LSTM is a form of RNN and are likely to learn long-term dependencies. LSTM allow RNN to keep track of their input data over a long time.

Architecture of the proposed solution. (Refer proposed work)
Diagram

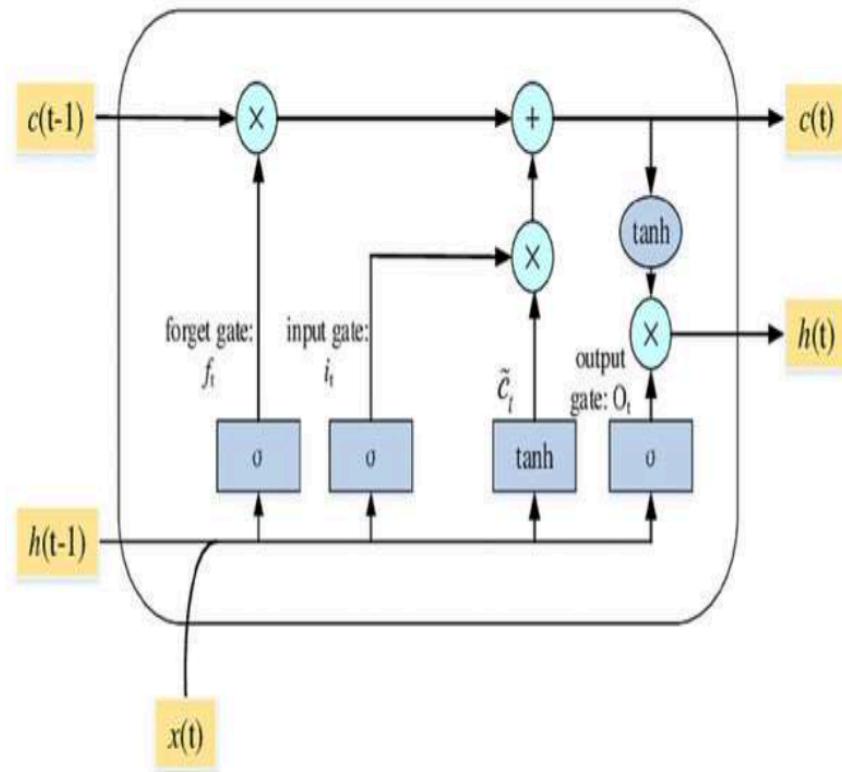


Figure 4: The LSTM model with its three layers.

DOI: [10.7717/peerj-cs.476/fig-4](https://doi.org/10.7717/peerj-cs.476/fig-4)

Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)	Deep Learning and LSTM based model
List of existing algorithms used by the authors to complete the proposed work. (1-2 lines for each algorithm)	RNN Algorithm : RNN stands for recurrent neural network which is a class of artificial neural networks where connections between nodes form a directed graph along a temporal sequence. LSTM algorithm : Long short-term memory is an artificial recurrent neural network architecture used in the field of deep learning. It can process not only single data points, but also entire sequences of data.
List of datasets used. (Refer experimental evaluation/result discussion) (3-4 lines)	1. News from Money control 2. IFL 3. Economic Times 4. Twitter 5. NSE Stock Data
References/links to each of the dataset used in this paper (in APA style)	1. http://moneycontrol.com 2. https://www.iflhousingfinance.com 3. https://economictimes.indiatimes.com 4. https://twitter.com/search?q=stock 5. https://www1.nseindia.com/products/content/equities/equities/eq_security.htm
Why the above dataset(s) used? <i>(Refer experimental evaluation/result discussion)</i> (3-4 lines)	News sites like Money Control, Economic Times, IFL are trusted sources and they have enough information about stock related stuff. Twitter is one of the largest social media to house tweets related to stock market and share prices of numerous companies. NSE stock data is also the most reliable when it comes to stocks of Indian companies.

Equation 1: Sentiment Formulas

Table 1:
Predominant sentiment formulas.

Precision	Recall	F-score	Accuracy
$\frac{\sum \text{Tp}}{\sum \text{Tp} + \sum \text{Fn}}$	$\frac{\sum \text{Tp}}{\sum \text{Tp} + \sum \text{Fn}}$	$\frac{2\sum \text{Tp}}{2\sum \text{Tp} + \sum \text{FP} + \sum \text{FN}}$	$\frac{\sum \text{Tp} + \sum \text{Tn}}{\sum \text{Tp} + \sum \text{Tn} + \sum \text{Fn} + \sum \text{Fn}}$

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

Equation 2: Various LSTM formulas were used :

$$FT = \sigma(W_f * [ht - 1, xt] + B_f)$$

$$\Delta Ct = \tanh(W_C \cdot [ht - 1, xt] + b_C)$$

$$it = \sigma(W_i \cdot [ht - 1, xt] + bi)$$

$$Ct = ft * Ct - 1 + it * \Delta Ct$$

$$Ot = \Sigma(W_o \cdot [ht - 1, xt] + Bo)$$

List of equations that are very well applied in this problem domain

<p>List of method(s)/metrics used to evaluate the proposed approach. <i>(Refer experimental evaluation/result discussion)</i> (5-8 lines)</p>	<p>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.</p>
<p>List of supporting tools/concepts (3-4 lines)</p>	<p>Social Media Sites like Twitter, Facebook were used. RNN algorithm was used for LSTM method. Deep Learning Techniques were also helpful.</p>
<p>What are the similar approaches with which the proposed approach is compared? <i>(Refer experimental evaluation/result discussion)</i> Explain each of these approach (3-4 lines)</p>	<p>Approach/method 1: Naive Bayes Technique Approach/method 2: Linear Regression Approach/method 3: Maximum entropy Approach/method 4: Decision Tree Approach/method 5: Linear SVC Regression</p>
<p>How the results of proposed approach are compared with other similar approaches? <i>(Refer experimental evaluation/result discussion)</i></p>	<p>The average sentiment estimates the regular sentiment of any topic over a given period. The experiments done considered opinions, primary sentiment, precision and recall. Stock prediction using different social media platforms resulted in more accurate and reliable than previous predictions that were made.</p>
<p>Advantages/merits of proposed solution in your view. <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<ol style="list-style-type: none"> 1. LSTM provide with various range of parameters like learning rates, and input and output biases. 2. There is no need of fine adjustments. 3. Complexity is reduced to O(1) with LSTM.

<p>Disadvantages/limitations of proposed solution in your view. <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<ol style="list-style-type: none"> 1. LSTMs get influenced by various irregular weight introduction and consequently act very like that of a feed-forward neural net. They incline toward little weight introduction all things being equal. 2. LSTMs are inclined to overfitting and it is hard to apply the dropout calculation to check this issue. Dropout is a regularisation technique where input and intermittent associations with LSTM units are probabilistically avoided from initiation and weight refreshes while preparing an organisation.
<p>Future work as stated by authors <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<p>The use of news articles, social media like Twitter, Facebook explains how a company performance 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 stocks 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.</p>
<p>Your one page write-up about this paper</p> <p>Predicting the stock market is a very strenuous task and it depends on number of factors. Researchers have found that news articles social media and sentiments of people are interrelated with the stock market. For extracting the sentiment data and social media information regarding a particular company different ML classifiers and deep learning techniques can be used. These techniques can't predict whether the stock value of a particular company can increase or decrease depending on its sentiment value. The experiments carried out by the following paper used sentiment analysis along with historical stock price, data of different news websites like money control, economic times etc. and social media data like Twitter. The results of the experiment is conducted were able to successfully predict the nature of stock market and also tell whether a company's stock value will increase or decrease in the near future. This will help people to invest and earn in the long run.</p> <p>The following research paper provided the following:</p> <p>It suggested combination of economic news and social network analysis to forecast SM developments.</p> <p>Implemented a collection of features from final data set to improve the performance of predictions.</p> <p>This paper recommended impact on SM from social networking sites like Twitter, Facebook, Google plus and financial forecast.</p>	

Your findings: (possible alternate for the solution proposed)

1. Naive Bayes Algorithm : It is based on bases theorem which share a common rule, ie every pair of feature classified is independent of each other.
2. Maximum Entropy : The principle of maximum entropy states that the probability distribution which best represents the current state of knowledge about a system is the one with largest entropy, in the context of precisely stated prior data

Research Questionnaire

Note:

- 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	19BCE0959 Devang Gupta
Team No.	7
Paper Title	Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation Stock Prediction by using NLP and Deep Learning Approach 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)
Citation (APA style)	Koratamaddi, P., Wadhwani, K., Gupta, M., & Sanjeevi, S. G. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. <i>Engineering Science and Technology, an International Journal</i> , 24(4), 848-859. Deshmukh, R. (2021). Stock Prediction by using NLP and Deep Learning Approach. <i>Turkish Journal of Computer and Mathematics Education (TURCOMAT)</i> , 12(1S), 202-211. 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). <i>IEEE Access</i> , 8, 188186-188198.

TABLE 2

Problem answered in this paper. (1-2 lines)	To solve stock portfolio allocation and maintain it to get maximum return with minimum risk involved.
Detailed description about the problem (5-8 lines)	One of the most difficult and fascinating challenges in modern finance is portfolio allocation. This is due to the stock market's complexity, which is characterized by a web of interconnected return effects that necessitates a significant amount of computational work to decipher and predict return regularities. Furthermore, because the stock market is continuously changing, anticipating stock price fluctuations is not an easy process.
Why that problem is chosen in this paper? Scope of the problem and solution (Refer Introduction) (5-8 lines)	People are confused and find themselves in difficult situation when it comes to maintain their stock portfolio. They have to maintain the stocks in their portfolio such that then not only get maximum returns but also with minimum risk involved. To help people from this situation this paper proposes an algorithm that will analyze the news and tweets to help find a stock that can give maximum return with minimum risk involved.
History of the problem. (Refer Introduction) (8-10 lines)	Some of the earlier efforts on portfolio allocation were based on mathematical models that used quadratic programming, stochastic calculus, numerical analysis, and other techniques. Statistical learning approaches were initially utilized for simple numerical analysis problems – for example, the Newton-Raphson algorithm was used to solve logistic regression. However, when supervised machine learning methods such as artificial neural networks became more prominent in the 1990s, multiple deep learning algorithms for diverse stock market applications were proposed. The ability of neural networks to learn complex nonlinear functions is why they outperform other machine learning approaches in predicting market returns.
List of the related/similar problems (Refer Related work) – Describe each with proposed solutions	
Related problem 1 – Describe (3-4 lines)	To solve the financial problem of portfolio selection.
Paper in APA style	Koyano, S., & Ikeda, K. (2017, November). Online portfolio selection based on the posts of winners and losers in stock microblogs. In <i>2017 IEEE Symposium Series on Computational Intelligence (SSCI)</i> (pp. 1-4). IEEE.
Related problem 2 – Describe (3-4 lines)	To predict one-month-ahead stock returns in the cross-section in the Japanese stock market and investigates the performance of the method
Paper in APA style	Abe, M., & Nakayama, H. (2018, June). Deep learning for forecasting stock returns in the cross-section. In <i>Pacific-Asia conference on knowledge discovery and data mining</i> (pp. 273-284). Springer, Cham.

Related problem 3 – Describe (3-4 lines)	To find an appropriate modeling system that can incorporate the complexities of the stock market and generate practical trading strategies.
Paper in APA style	Bao, W., & Liu, X. Y. (2019). Multi-agent deep reinforcement learning for liquidation strategy analysis. <i>arXiv preprint arXiv:1906.11046</i> .
Related problem 4 – Describe (3-4 lines)	To investigate the role of market sentiment in an asset allocation problem.
Paper in APA style	Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Intelligent asset allocation via market sentiment views. <i>ieee Computational Intelligence magazine</i> , 13(4), 25-34.
Related problem 5 – Describe (3-4 lines)	Dynamic portfolio optimization is the process of sequentially allocating wealth to a collection of assets in some consecutive trading periods, based on investors' return-risk profile. Automating this process with machine learning remains a challenging problem.
Paper in APA style	Yu, P., Lee, J. S., Kulyatin, I., Shi, Z., & Dasgupta, S. (2019). Model-based deep reinforcement learning for dynamic portfolio optimization. <i>arXiv preprint arXiv:1901.08740</i> .
What is the proposed solution in this paper for the problem chosen? (Refer Proposed work) (5-8 lines)	In this they proposed an adaptive, sentiment-aware deep deterministic policy gradients approach to solve portfolio allocation that not only learns from historical stock price trends, but also from market sentiment – which is an influential environment input that captures the overall mood of investors. They've consolidated an extensive dataset of Google News and Twitter tweets that reflect the sentiment of the 30 Dow Jones companies. It also provided the methodology and mathematical definitions used to calculate market sentiment, and enable the adaptive DDPG algorithm to leverage it sufficiently.

<p>Architecture of the proposed solution. <i>(Refer proposed work)</i></p> <p>Diagram</p>	<p>The diagram illustrates the architecture of a sentiment-aware deep reinforcement learning approach. It features a central loop involving a State, Actor, Critic, Q-Value, and Environment. The Actor receives the State and outputs Actions. The Environment provides Observations and Rewards. The Critic receives Observations and Q-Value as input and outputs Critic Loss. The Actor receives Q-Value and Actor Loss as input and outputs an Update. The Q-Value receives Actor Loss and Rewards as input and outputs Prediction Error Inference. The Environment also provides Closing Prices of Companies, Market Sentiment Confidence Score (CS = average[NS, TS]), Financial News Sentiment Score (NS), and Twitter Sentiment Score (TS). A graph titled "The U.S. Stock Market" shows a line chart of closing prices.</p>
<p>Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)</p>	<p>Sentiment-aware deep reinforcement learning approach</p>

List of existing algorithms used by the authors to complete the proposed work. (1-2 lines for each algorithm)	<p>(DDPG) Deep Deterministic Policy Gradients algorithm Deep Deterministic Policy Gradients (DDPG) is a deep reinforcement learning algorithm that concurrently learns a Q-function (learnt by the critic network) and a policy (learnt by the actor network). The algorithm is a model-free, off-policy actor-critic algorithm using deep function approximators that can learn policies in high-dimensional, continuous action spaces</p> <p>Adaptive Deep Deterministic Policy Gradients (Adaptive DDPG) As an improvement to the DDPG algorithm that allowed for learning the bullish and bearish aspects of the market actively</p>
List of datasets used. <i>(Refer experimental evaluation/result discussion)</i> (3-4 lines)	Google News Twitter
References/links to each of the dataset used in this paper (in APA style)	1. https://www.kaggle.com/getting-started/8223 2. https://www.kaggle.com/kazanova/sentiment140 ...
Why the above dataset(s) used? <i>(Refer experimental evaluation/result discussion)</i> (3-4 lines)	Twitter and Google News are used as the data set in this paper as, Twitter tweets was found useful to predict the stock prices because twitter is a platform where they share their opinions and in a smaller number of words which can be easily analyzed. Google news is a reliable news platform and using its news and twitter tweets the predicted price of the stock was expected to be more precise.
List of equations that are very well applied in this problem domain	Equation 1: $NS_{(c,d)} = \frac{\sum_{i=1}^N PS(a_{(i,c,d)})}{N}$ <p>Description: Sentiment analysis was performed to obtain a google news sentiment score.</p>

	<p>Equation 2:</p> $TS_{(c,d)} = \frac{\sum_{i=1}^N w(t_{(i,c,d)}) PS(t_{(i,c,d)})}{N}$ <p>Description: Sentiment analysis was performed to obtain a twitter sentiment score</p> <p>Equation 3:</p> $CS_{(c,d)} = \frac{NS_{(c,d)} + TS_{(c,d)}}{2}$ <p>Description: The confidence score is computed using the financial news sentiment score (NS) and twitter sentiment score (TS)</p>
List of method(s)/metrics used to evaluate the proposed approach. (Refer experimental evaluation/result discussion) (5-8 lines)	Sharpe Ratio Annualized Return Annualized Std. Error Final Portfolio Value
List of supporting tools/concepts (3-4 lines)	VADER (Valence Aware Dictionary and sentiment Reasoner). VADER is a lexicon and rule-based sentiment analysis tool implemented in Natural Language Toolkit.
What are the similar approaches with which the proposed approach is compared? (Refer experimental evaluation/result discussion)	<p>Approach/method 1: Supervised machine learning tools such as artificial neural networks. Neural networks approach was more successful as compared to other machine learning approaches in predicting stock returns is attributed to their ability to learn complex nonlinear functions.</p> <p>Approach/method 2: Deep learning approaches, LSTMs are very powerful in sequence prediction problems because they're able to store past information. This is important in our case because the previous price of a stock is crucial in predicting its future price.</p>

Explain each of these approach (3-4 lines)	
How the results of proposed approach are compared with other similar approaches? <i>(Refer experimental evaluation/result discussion)</i>	It is noteworthy that the sentiment aware approach has significant improvements across all the considered metrics compared to the baselines. With an initial investment of 10,000 dollars, the final portfolio value of our approach reaches 25,051 dollars which is much higher than 21,881 dollars by Adaptive DDPG and 18,156 dollars by DDPG. This approach achieves an annualized rate of return of 22.05% as compared to 18.85% and 14.7% returns by Adaptive and DDPG respectively. The risk taken by this model is also lesser compared to the other baselines as shown by obtaining the least annualized standard deviation error of 0.096. With a 2.07 Sharpe ratio value, our sentiment-aware approach is shown to be more robust and effective in balancing return and risk compared to 1.49, 0.93 for the adaptive DDPG and DDPG respectively.
Advantages/merits of proposed solution in your view. <i>(Refer conclusion / result discussion / experimental evaluation)</i>	Deep Deterministic Policy Gradient (DDPG) is a model-free off-policy algorithm for learning continuous actions. It combines ideas from DPG (Deterministic Policy Gradient) and DQN (Deep Q-Network). It uses Experience Replay and slow-learning target networks from DQN, and it is based on DPG, which can operate over continuous action spaces.
Disadvantages/limitations of proposed solution in your view. <i>(Refer conclusion / result discussion / experimental evaluation)</i>	In environments with continuous state and action spaces, Deep Deterministic Policy Gradient (DDPG) algorithms can solve very complex problems, yet can also fail in environments that seem trivial, but the reason for such failures is still poorly understood.
Future work as stated by authors <i>(Refer conclusion / result discussion / experimental evaluation)</i>	Future work could focus on acquiring more tweets per day, expanding acquisition to get insights from different sources such as stock market-specific news websites (CNBC, Business Standard, etc.), and processing photos, as most tweets and news online are now provided as image snippets. For this case, multi-agent reinforcement learning algorithms can be investigated. We're also excited to address the presence of several exogenous restraints on retail and institutional traders, such as transaction fees, trading restrictions, cash holding restrictions, and liquidity shortages. Furthermore, due to the extensive use of metaphors, sarcasm, domain specific terminology, and other indirect linguistic references in common language, especially in material that expresses an opinion, natural language processing on financial data is a non-trivial effort. Being able to grasp such language might aid in more accurately predicting market sentiment.

Your one page write-up about this paper

The stock market is still one of the most challenging financial systems to model. As a result, stock portfolio allocation is a difficult problem to solve, as an optimal investing strategy for a curated group of companies must be developed that efficiently maximizes return while minimizing risk. When used to automate portfolio allocation by training an intelligent agent on historical stock prices, deep reinforcement learning algorithms have demonstrated promising outcomes. Modern investors, on the other hand, are using digital platforms like social media and online news websites to better understand and analyses their portfolios. Market sentiment refers to the overall attitude that investors have about a certain stock or financial market. Existing techniques ignore market mood, which has been proved to influence investor decisions empirically. They suggested a revolutionary deep reinforcement learning strategy for efficiently training an intelligent automated trader that not only uses historical stock price data but also detects market sentiment for a Dow Jones stock portfolio. They showed that their method is more robust than existing baselines in terms of standardized metrics like the Sharpe ratio and annualized investment return.

Your findings: (possible alternate for the solution proposed)

- ARIMA is a very popular statistical method for time series forecasting. ARIMA models take into account the past values to predict the future values.

TABLE 2

Problem answered in this paper. (1-2 lines)	To properly recognize which shares to promote with a purpose to get more profits.
Detailed description about the problem (5-8 lines)	People have a tendency to analyze existing strategies and so planned new strategies for inventory prediction. They predict the share price based on random assumption. People are not able to keep track of the stock prices whether it is going up or down. Financial analysts investing in stocks usually, but they are not aware about the inventory market place conduct. They usually go through the problem of trading as they do not properly recognize which shares to shop or which shares to promote with a purpose to get greater profits.
Why that problem is chosen in this paper? Scope of the problem and solution (Refer Introduction)	In today's world, all the information pertaining to inventory market is available. Analyzing all these records in my opinion or manually is pretty much difficult. As such, automation of the method is required. This is where Data mining techniques help. Understanding that analysis of numerical time series offers close results, wise traders use system learning techniques in predicting the inventory market conduct. This will allow financial

(5-8 lines)	analysts to foresee the conduct of the inventory that they may be interested by and consequently act accordingly.
History of the problem. <i>(Refer Introduction)</i> (8-10 lines)	They had used one system to find out the opening value of stocks in the financial market. However, their developed system was self-learner so that they were able to predict the opening value of the market. They had given the stocks data to their developed system to find out the forecasted value. Last, they developed another network system and compared both the system with each other to predict the starting day value of the stock. Used ANN algorithm for the share market prediction. On the basis of their studies of ANN they had claimed that they had achieved the high percentage of accuracy while predicting the values in the stock market.
List of the related/similar problems <i>(Refer Related work)</i> – Describe each with proposed solutions	
Related problem 1 – Describe (3-4 lines)	They employed a sentiment analysis technique for stock tweets that were related to a different type of Apple product; for this, she had extracted stock-related tweets from various social networking sources for an eight-year period. These individuals had elected to use stocks-related data from the Yahoo Finance source for that time period in addition to share data. To determine the polarity of those tweets, they applied the SVM approach.
Paper in APA style	Batra, R., & Daudpota, S. M. (2018, March). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. In <i>2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)</i> (pp. 1-5). IEEE.
Related problem 2 – Describe (3-4 lines)	The focus of this article was on the market's share price swings. They employed additional important information in addition to mining approaches for stock predictions. Their findings revealed that they calculated the polarity of the stocks in order to better anticipate the stock's price.
Paper in APA style	Wang, Y., & Wang, Y. (2016, March). Using social media mining technology to assist in price prediction of stock market. In <i>2016 IEEE International conference on big data analysis (ICBDA)</i> (pp. 1-4). IEEE.
Related problem 3 – Describe (3-4 lines)	They had 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.
Paper in APA style	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 <i>2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)</i> (pp. 225-228). IEEE.
Related problem 4 – Describe (3-4 lines)	They employed the ANN algorithm to predict the stock market. They stated that their research of ANN had resulted in a high percentage of accuracy in predicting stock market prices. They researched several approaches for this, and after doing so, they discovered accurate and proper findings.
Paper in APA style	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 <i>2018 Third International Conference on Informatics and Computing (ICIC)</i> (pp. 1-4). IEEE.

Related problem 5 – Describe (3-4 lines)	They were primarily interested in predicting share values using historical data. They used decade data from the NSE and BSE, two well-known stock exchanges. They used the SVM method to create a model for this.
Paper in APA style	Sharma, N., & Juneja, A. (2017, April). Combining of random forest estimates using LSboost for stock market index prediction. In <i>2017 2nd International conference for convergence in technology (I2CT)</i> (pp. 1199-1202). IEEE.
What is the proposed solution in this paper for the problem chosen? <i>(Refer Proposed work)</i> (5-8 lines)	In this paper they tend to area unit planned System that is working with Improved level of recommendation. System is developed with Natural Language Processing (NLP) technique of computer science and Convolutional Neural Network (CNN) of Deep Learning. Natural Language Processing technology is used facilitate system to search out companies with excellent news in terms of live performance in market. That helped to facilitate to create selection of best performer in market. NLP is used to classify news in positive and negative sets and to provide performance graph of selected organization. Supported to that we got to know the best performing company. Natural Language Processing provides to system NLP (Natural Language Processing) that worked on our twits for detection merchandise and unhealthy of its impact.
Architecture of the proposed solution. <i>(Refer proposed work)</i> Diagram	

Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)	Prediction of stock prices using CNN and VADAR Sentiment analysis algorithm.
List of existing algorithms used by the authors to complete the proposed work. (1-2 lines for each algorithm)	<p>CNN Algorithm - A Convolutional Neural Network (ConvNet/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.</p> <p>VADER Algorithm - VADER (Valence Aware Dictionary for Sentiment Reasoning) is a model used for text sentiment analysis that is sensitive to both polarity (positive/negative) and intensity (strength) of emotion. VADER sentimental analysis relies on a dictionary that maps lexical features to emotion intensities known as sentiment scores.</p>

List of datasets used. <i>(Refer experimental evaluation/result discussion)</i> (3-4 lines)	Twitter Tweets top 5 performing companies' data in IT sector
References/links to each of the dataset used in this paper (in APA style)	1. https://www.kaggle.com/kazanova/sentiment140 2. ...
Why the above dataset(s) used? <i>(Refer experimental evaluation/result discussion)</i> (3-4 lines)	Twitter tweets was found more useful to predict the stock prices as twitter is a platform where they share their opinions and in a smaller number of words which can be easily analyzed.
List of equations that are very well applied in this problem domain	Equation 1: Description: Equation 2: Description: Equation 3: Description:
List of method(s)/metrics used to evaluate the proposed approach. <i>(Refer experimental evaluation/result discussion)</i> (5-8 lines)	The real data of the stock prices were used to evaluate the CNN model for prediction and forecasting. Comparing both the data they measured the prediction accuracy. The model first analyzed the previous data then compared the future data which was then compared with the stock price data. And using this the model achieved a polarity and based on this polarity they measured the top 10 well performing companies in the IT sector.
List of supporting tools/concepts (3-4 lines)	VADER sentiment analysis
What are the similar approaches with which	Approach/method 1: ANN Algorithm to predict the stock prices.

<p>the proposed approach is compared? (Refer experimental evaluation/result discussion)</p> <p>Explain each of these approach (3-4 lines)</p>	<p>They used ANN algorithm for the share market prediction. On the basis of their studies of ANN they had claimed that they had achieved the high percentage of accuracy while predicting the values in the stock market. For this they had studied different methods and after studying that they had find out the accurate and proper results.</p>
<p>How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)</p>	<p>CNN has the characteristic of paying attention to the most obvious features in the line of sight, so it is widely used in feature engineering. With comparison to other approach i.e., ANN algorithm CNN performs much better.</p>
<p>Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)</p>	<p>The proposed CNN achieved relatively higher prediction accuracy of 84.6%, while the ANN, SVM, and KNN algorithms obtained prediction accuracies of 73.5%, 67.9%, and 65.9% using 11 HRV features, respectively.</p>
<p>Disadvantages/limitations of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)</p>	<p>CNN do not encode the position and orientation of object. Lack of ability to be spatially invariant to the input data. Lots of training data is required. If CNN-LSTM was used it can provide a reliable stock price forecasting with the highest prediction accuracy.</p>
<p>Future work as stated by authors (Refer conclusion / result discussion / experimental evaluation)</p>	<p>In future they will attempt to execute more calculations and all the newer methods planning to give live proposal to securities exchange financial specialists. Additionally, their emphasis will be on entire securities exchange for forecasting.</p>
<p>Your one-page write-up about this paper</p>	

Early stock prediction research relied entirely on random walks, machine learning, numerical prediction, and support vector machines, but with the introduction of behavioral finance, people's market literacy was taken into account when predicting stock movement. They have proposed a model that will use sentiment analysis on twits associated with special sectors such as Information Technology, Banking, Pharmaceuticals, Automobiles, and Infrastructure that are extracted from twitter in order to exploit the benefits of sentiment analysis on enterprise associated inventory. These twits were taken from Twitter and used to calculate polarity. To make it more efficient, they applied NLP technology to analyses the sentiment of Stock Tweets. They put the theory into action by gathering sentiment and stock price data and creating a CNN model for prediction and forecasting, as well as measuring prediction accuracy. The results showed that we had attained polarity, and we measured the top 10 performing companies in each category using this polarity.

Your findings: (possible alternate for the solution proposed)

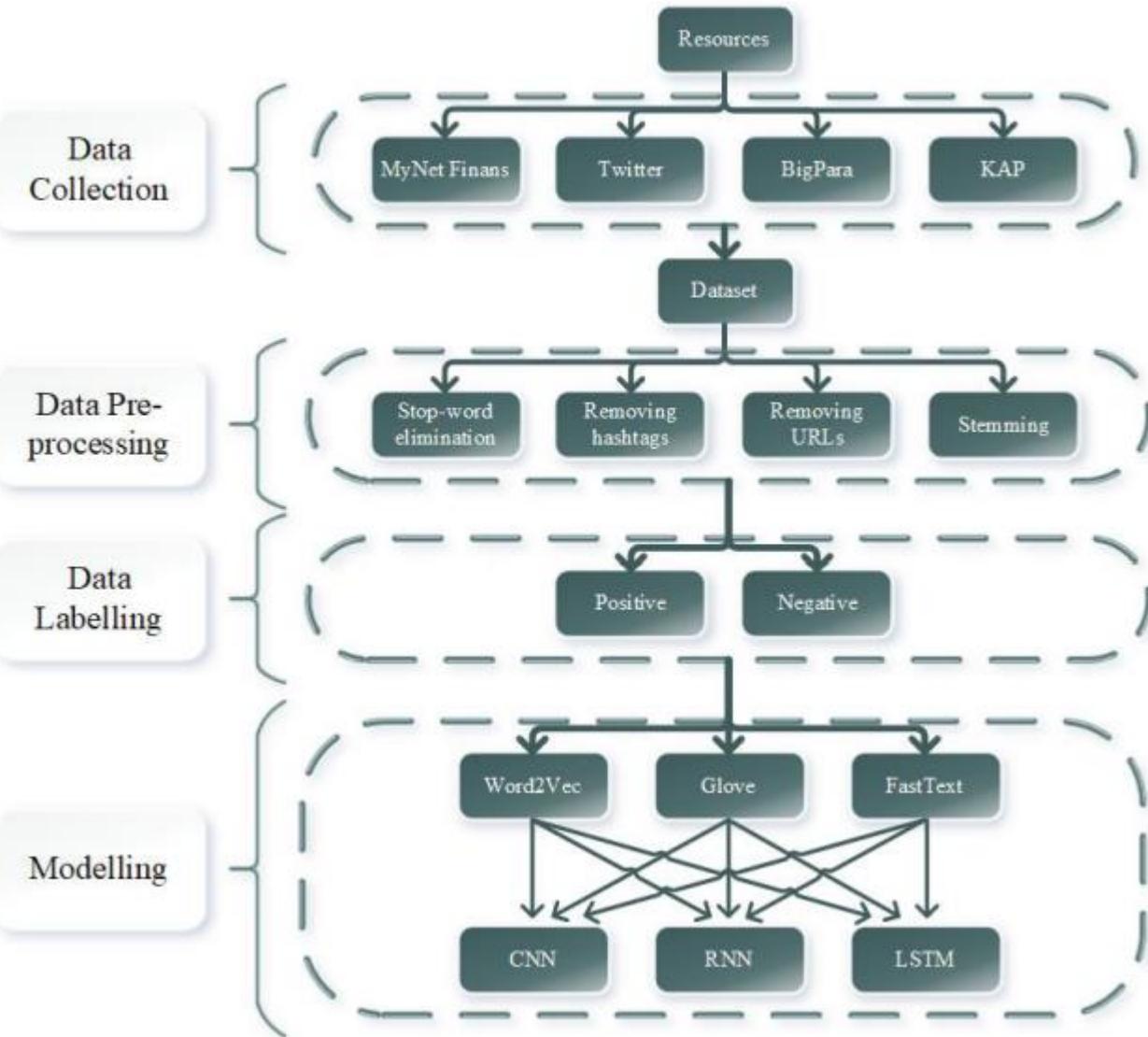
- Convolution Neural Network (CNN) and Long–Short-Term Memory Neural Network (LSTM). This new method is aptly named stock sequence array convolutional LSTM (SACLSTM).

TABLE 2

Problem answered in this paper. (1-2 lines)	Prediction of stock price in Istanbul Stock Exchange
Detailed description about the problem (5-8 lines)	In this study, unlike the recent researches on forecasting the stock market direction, they focus on financial sentiment analysis using the Turkish data sets collected from both a social media platform and websites including technical analysis and news to analyze the stock market direction by evaluating high volume stocks.
Why that problem is chosen in this paper? Scope of the problem and solution (Refer Introduction) (5-8 lines)	To forecast the movement directions of stocks, exchange rates, and stock markets are significant and an active research area for investors, analysts, and researchers. So, in this paper word embedding and deep learning-based direction prediction of Istanbul Stock Exchange (BIST 100) is proposed by analyzing nine banking stocks with high volume in BIST 100.
History of the problem. (Refer Introduction) (8-10 lines)	In some previous researches the combination of technical and fundamental analysts approaches applied to market trend forecasting through the use of conventional machine learning techniques applied to time series prediction and sentiment analysis on the same data but the results were not as expected.

List of the related/similar problems (Refer Related work) – Describe each with proposed solutions	
Related problem 1 – Describe (3-4 lines)	Equity investments are one of the most important asset classes. Equity investments have high return yield however also high risk due to the variability of share prices. Therefore, precise share price modeling is essential.
Paper in APA style	Tekin, S., & Çanakoğlu, E. (2019, April). Analysis of price models in istanbul stock exchange. In <i>2019 27th Signal Processing and Communications Applications Conference (SIU)</i> (pp. 1-4). IEEE.
Related problem 2 – Describe (3-4 lines)	Sentiment Analysis on turkish web comments using random forest classifier on Turkish Web comments.
Paper in APA style	PERVAN, N., & KELEŞ, H. Y. (2017). Sentiment analysis using a random forest classifier on turkish web comments. <i>Communications Faculty of Sciences University of Ankara Series A2-A3 Physical Sciences and Engineering</i> , 59(2), 69-79.
Related problem 3 – Describe (3-4 lines)	To estimate the direction of Borsa Istanbul 100 Index by using financial sentiment analysis and to enrich the datasets with various techniques from a semantic perspective and improve the classification performance of system by blending ensemble learning approach with deep learning algorithms.
Paper in APA style	Kilimci, Z. H. (2020). Financial sentiment analysis with Deep Ensemble Models (DEMs) for stock market prediction. <i>Journal of the Faculty of Engineering and Architecture of Gazi University</i> , 35(2), 635-650.
Related problem 4 – Describe (3-4 lines)	To forecast the direction of stocks is a significant for investors, analysts, and researchers. In this study, we propose to predict the direction of stocks in Turkish stock market (BIST100) by employing Turkish texts such as social media platforms. For this purpose, different deep learning methodologies
Paper in APA style	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 <i>Proc. Int. Conf. Innov. Intell. Technol.</i> (Vol. 2019, pp. 30-35).
Related problem 5 – Describe (3-4 lines)	The stock market parameter forecasting is an important research subject both for financial professionals and the machine learning experts due to the challenges and opportunities it possess. Despite the difficulties in financial data, interest in this research area is growing rapidly.
Paper in APA style	Tekin, S., & Çanakoğlu, E. (2018, May). Prediction of stock returns in Istanbul stock exchange using machine learning methods. In <i>2018 26th Signal Processing and Communications Applications Conference (SIU)</i> (pp. 1-4). IEEE.
What is the proposed solution in this paper for the problem chosen? (Refer Proposed work) (5-8 lines)	In this paper long short-term memory networks, recurrent neural networks, convolutional neural networks as deep learning algorithms and Word2Vec, GloVe, and FastText as word embedding models are evaluated. To demonstrate the effectiveness of proposed model, four different sources of Turkish news are collected. The news articles about stocks from Public Disclosure Platform (KAP), text-based technical analysis of each stock from Bigpara, user comments from both Twitter and Mynet Finans platforms are gathered.

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)	Word embedding techniques and deep learning approaches
List of existing algorithms used by the authors to complete the proposed work. <i>(1-2 lines for each algorithm)</i>	Word embedding models namely, FastText, GloVe, Word2Vec, deep learning techniques such as CNN, LSTM, RNN as well as combinations of these methods are approaches that are evaluated in this study. For the deep learning models and word embedding algorithms, Keras library is utilized.
List of datasets used. <i>(Refer experimental evaluation/result discussion)</i> <i>(3-4 lines)</i>	Twitter as a social media platform KAP as public disclosure platform Mynet Finans as news platform Bigpara as a technical analysis platform
References/links to each of the dataset used in this paper (in APA style)	1. https://www.kaggle.com/kazanova/sentiment140 2. ...
Why the above dataset(s) used? <i>(Refer experimental evaluation/result discussion)</i> <i>(3-4 lines)</i>	Though English news articles have been employed for forecasting of market direction previously, to the best of our knowledge, Turkish news articles and user comments from social media and different platforms have not been utilized with the combination of deep learning techniques and word embedding methods to predict the direction of Turkish stocks and market
List of equations that are very well applied in this problem domain	<p>Equation 1: $J_{\theta} = \frac{1}{T} \sum_{t=1}^T \sum_{-n \leq j \leq n} \log p(w_{t+j} w_t)$</p> <p>Description: Given a center word and sequence of training words $w_1, w_2, w_3, \dots, w_t$ skip-gram model maximizes the average log probability of n surrounding words of the center word w_t, n denotes the size of training context.</p> <p>Equation 2: $J_{\theta} = \sum_{i,j=1}^V f(X_{ij}) \left(w_i^T w_j + b_i + b_j - \log X_{ij} \right)^2$</p>

	<p>Description: Glove method first contracts a word co-occurrence matrix X. Each element of X_{ij} shows the number of times word i appears in the context word j. The Glove model utilizes (3) to calculate cost.</p> $\text{Equation 3: } s(w, c) = \sum_{g \in G_w} z_g^T v_c$ <p>Description: FastText uses the skip-gram model with negative sampling proposed for Word2Vec with a modified skip-gram loss function. Let $G_w \supseteq \{1, \dots, G\}$ be the set of n-grams appearing in a word w, the score of the word is calculated by the sum of the vector representations of its n-grams:</p>
List of method(s)/metrics used to evaluate the proposed approach. (Refer experimental evaluation/result discussion) (5-8 lines)	<p>The success of proposed model is appraised with evaluation metrics namely, F-criterion, and accuracy. F-criterion is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled.</p>
List of supporting tools/concepts (3-4 lines)	<p>Word2Vec is accepted as a pioneer word embedding method that starts a new trend in natural language processing. Word2Vec tries to express words in a vector space and it is a prediction-based and unsupervised model.</p> <p>Global Vectors (GloVe) is another popular word embedding algorithm.</p> <p>FastText is an artificial neural network library developed for text classification. Converts text or words into continuous vectors that can be used in any language, such as a speech-related task.</p> <p>Bidirectional Encoder Representations from Transformers (BERT) is a machine learning framework which is designed for natural language processing. In this every output element is connected to every input element and it is designed to pre-train deep bidirectional representations from unlabeled text by jointly conditioning on both left and right context in all layers</p>
What are the similar approaches with which the proposed approach is compared? (Refer experimental evaluation/result discussion) Explain each of these approach	<p>Approach/method 1: Evolutionary Neural Network - In this work, the prediction of status of stocks in BIST 100 are proposed by using ESA. The classification performance obtained by ESA is evaluated to be higher than those obtained with the chi-square feature selection and logistic regression classifier.</p> <p>Approach/method 2: LSTM, multi-layer perceptron (MLP), and random forest classifier (RFC) are employed. an approach is proposed to compute the asset-level market sentiment from social media data stream, and integrate it to the state-of-the-art asset allocation method using market views.</p>

(3-4 lines)	
How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)	The combination of RNN and LSTM models with word embedding models exhibit remarkable experiment results for all data sets. When data sets are evaluated, it is observed that either Word2Vec or FastText word embedding models present more successful results. Mynet Finans and Twitter data sets contain user comments which have much shorter text content compared to the Bigpara and KAP data sets.
Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)	The main Advantage is that since the model uses RNN, LSTM, Machine Learning and Deep Learning models the prediction of stock prices will be more accurate. And also, in the model it can predict the future 30 days Stock Prices and it can show it in a graph. Also, the main feature is that the model can show an output of the Individual Predicted Close prices of the Predicted 30 days.
Disadvantages/limitations of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)	LSTM require 4 linear layer (MLP layer) per cell to run at and for each sequence time-step. Linear layers require large amounts of memory bandwidth to be computed, in fact they cannot use many computes unit often because the system has not enough memory bandwidth to feed the computational units.
Future work as stated by authors <i>(Refer conclusion / result discussion / experimental evaluation)</i>	In future work authors may try to increase the data set and also tune the parameters to predict more accurate value of stock prices.
Your one-page write-up about this paper	
For investors, analysts, and researchers, forecasting the movement directions of stocks, exchange rates, and stock markets is a key and active research topic. By evaluating nine banking stocks with large volume on the Istanbul Stock Exchange (BIST 100), this paper proposes word embedding and deep learning-based direction prediction. Although English news articles have previously been used to forecast market direction, to our knowledge, Turkish news articles and user comments from social media and other platforms have not been combined with deep learning techniques and word embedding methods to forecast the direction of Turkish stocks and markets. t. For this objective, long short-term memory networks, recurrent neural networks, convolutional neural networks as deep learning algorithms and Word2Vec, GloVe, and FastText as word embedding models are evaluated. 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.	

Your findings: (possible alternate for the solution proposed)

- Instead of word embedding approach text mining approach offers good prediction and great profit.

Research Questionnaire

Note:

- 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

1.

TABLE 1	
Reg. No. & Name	19BCE2245 - Gokul R Nair
Team No.	7
Paper Title	1. Prediction of stock values changes using sentiment analysis of stock news headlines 2. A Method of Using News Sentiment for Stock Investment Strategy 3. Augmented Textual Features-Based Stock Market Prediction
Citation (APA style)	1. Nemes, L., & Kiss, A. (2021). Prediction of stock values changes using sentiment analysis of stock news headlines. <i>Journal of Information and Telecommunication</i> , 1-20. 2. Katayama, D., & Tsuda, K. (2020). A Method of Using News Sentiment for Stock Investment Strategy. <i>Procedia Computer Science</i> , 176, 1971-1980. 3. Bouktif, S., Fiaz, A., & Awad, M. (2020). Augmented textual features-based stock market prediction. <i>IEEE Access</i> , 8, 40269-40282.

TABLE 2

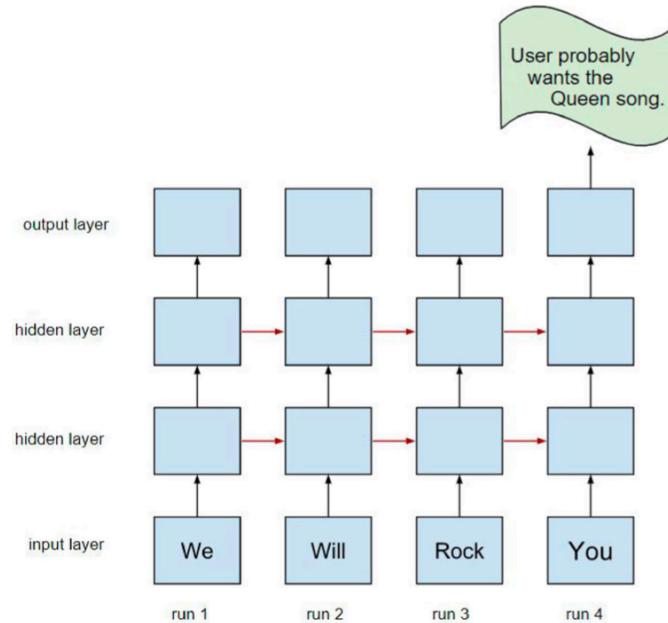
Problem answered in this paper. (1-2 lines)	How to analyse stock value changes and predictions of the stock values using fresh scraped economic news about the companies
Detailed description about the problem (5-8 lines)	Stock market prediction is highly dynamic, in-order to predict its change and analyse it we need some high computation technology. The following work focuses on analysing the stock market using news headlines.
Why that problem is chosen in this paper? Scope of the problem and solution (<i>Refer Introduction</i>) (5-8 lines)	The stock market is one of the most important economic participants. Many people try to interpret and define the different stock market movements in many ways. In this article, we use different tools to the sentiment analysis, especially focussing on the economic news, but in terms of economic news, focussing only on the headlines of economic news.
History of the problem. (<i>Refer Introduction</i>) (8-10 lines)	The traditional method which we use are analysing the graphs and making a prediction on it. The graphs analytics are still acceptable, but with the dynamics of Stock market changing every second we need to find some new methods.

List of the related/similar problems (*Refer Related work*) – Describe each with proposed solutions

Related problem 1 – Describe (3-4 lines)	Stock market prediction using an improved training algorithm of neural network
Paper in APA style	Billah, M., Waheed, S., & Hanifa, A. (2016, December). Stock market prediction using an improved training algorithm of neural network. In 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) (pp. 1-4). IEEE.
Related problem 2 – Describe (3-4 lines)	Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction
Paper in APA style	Das, S., Behera, R. K., & Rath, S. K. (2018). Real-time sentiment analysis of twitter streaming data for stock prediction. <i>Procedia computer science</i> , 132, 956-964.
Related problem 3 – Describe (3-4 lines)	BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

Paper in APA style	Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. <i>arXiv preprint arXiv:1810.04805</i> .
Related problem 4 – Describe (3-4 lines)	Predicting Stock Market Behavior using Data Mining Technique and News Sentiment Analysis
Paper in APA style	Khedr, A. E., & Yaseen, N. (2017). Predicting stock market behavior using data mining technique and news sentiment analysis. <i>International Journal of Intelligent Systems and Applications</i> , 9(7), 22.
Related problem 5 – Describe (3-4 lines)	A survey on opinion mining and sentiment analysis: Tasks, approaches and applications
Paper in APA style	Billah, M., Waheed, S., & Hanifa, A. (2016, December). Stock market prediction using an improved training algorithm of neural network. In 2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE) (pp. 1-4). IEEE.
What is the proposed solution in this paper for the problem chosen? (Refer Proposed work) (5-8 lines)	Sentiment analysis with BERT is one of the most powerful tool that we can use, but we can also create a Recurrent Neural Network (RNN) as well or use the NLTK tool with VADER Lexicon with SentimentIntensity Analyzer.

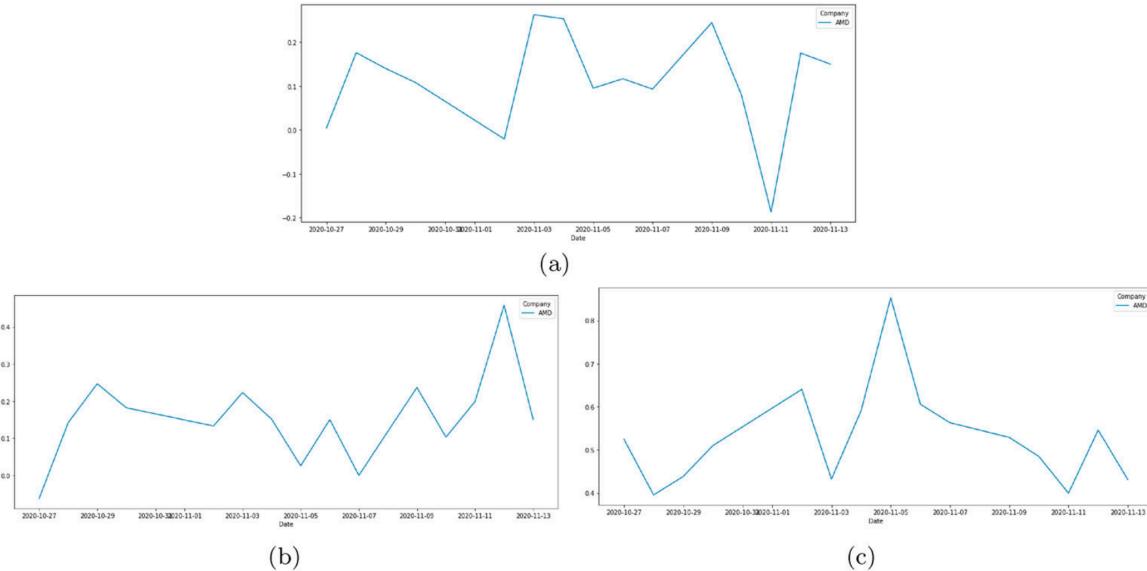
Architecture of the proposed solution. (*Refer proposed work*)
Diagram



Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)

Bidirectional encoder representations from transformers (BERT)

	<p>Recurrent neural network (RNN)</p> <p>Recurrent Neural Network is a neural network that is intentionally run multiple times, where parts of each run feed into the next run. Specifically, hidden layers from the previous run provide part of the input to the same hidden layer in the next run.</p> <p>NLTK -- VADER lexicon</p> <p>NLTK stands for Natural Language Toolkit. This toolkit is one of the most powerful NLP libraries which contains packages to make machines understand human language and reply to it with an appropriate response.</p> <p>TextBlob</p> <p>TextBlob is a powerful NLP library for Python, which is built upon NLTK and provides an easy to use interface to the NLTK library. This tool can be used to perform a variety of NLP tasks ranging from parts-of-speech tagging to sentiment analysis, and language translation to text classification, but we focus on the sentiment analysis.</p>
<p>List of datasets used. (Refer experimental evaluation/result discussion) (3-4 lines)</p>	<p>IMDB review dataset</p> <p>A dataset comprising of 50k plus movie reviews with 250000 highly polar reviews.</p>
<p>References/links to each of the dataset used in this paper (in APA style)</p>	<ol style="list-style-type: none"> 1. https://www.kaggle.com/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews 2. https://www.imdb.com/interfaces/

<p>Why the above dataset(s) used? <i>(Refer experimental evaluation/result discussion)</i> (3-4 lines)</p>	<p>IMDB dataset having 50K movie reviews for natural language processing or Text analytics. This is a dataset for binary sentiment classification containing substantially more data than previous benchmark datasets. We provide a set of 25,000 highly polar movie reviews for training and 25,000 for testing. So, predict the number of positive and negative reviews using either classification or deep learning algorithms.</p>
<p>List of equations that are very well applied in this problem domain</p>	<p>Equation 1: RNN Algo Description: The state of the art algorithm for sequential data and are used by Apple's Siri and Google's voice search.</p>
<p>List of method(s)/metrics used to evaluate the proposed approach. <i>(Refer experimental evaluation/result discussion)</i> (5-8 lines)</p>	<p>RNN, Text Blob, NLTK</p>  <p>Figure 12. Sentiment analysis of different models by daily separation. (a) TextBlob. (b) NLTK -Vader Lexicon and (c) RNN.</p>

List of supporting tools/concepts (3-4 lines)	<p>NLTK -- VADER lexicon</p> <p>TextBlob</p>
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)	<p>Approach 1: Graph Analysation The traditional method of graph analysation is compared with sentiment analysis based graphs to find the similarity.</p> <p>Approach 2: Statistical Analysation Past Stock prices of AMD was compared with the sentiment based statistics to match the precision.</p>
How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)	<p>Time based fluctuations of stock price. A normal for loop based code which compares data added in both the arrays(array 1 with historical data and array 2 with sentimental analysis data) to find acceptance rate.</p>
Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)	<p>The advantage of this method is we can predict the future prices without the help of any statistics of previous price chart. This will help us to free billions of data storage which comprises of time based stock prices.</p>
Disadvantages/limitations of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)	<p>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 $10^{-9} \times 10^9$ chances of happening.</p>

Future work as stated by authors

(Refer conclusion / result discussion / experimental evaluation)

Future work could include further expansion of the analyses, possible additions of new features. In addition, the inclusion of other tools to compare stock market predictions with different sentiment analysis tools. That can be built into an easy-to-use format by developing a platform incorporating various future changes of tensorflow into the current model.

Your one page write-up about this paper

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

Your findings: (possible alternate for the solution proposed)

- One of the alternate is using LSTM rather than RNN, because the specificity of LSTM would helps us to target the core problem
- Using pre built ML model synthesiser(eg: CoreML) which will help to create ML model based on the data provided and precision

2.

TABLE 2	
Problem answered in this paper. (1-2 lines)	This study evaluates the sentiment of Japanese news and attempts to apply it to investment strategies in individual stocks.
Detailed description about the problem (5-8 lines)	When the effectiveness of investment was lost in weekly and monthly frequencies Japanese news made an attempt to make individual stocks stable. This reveals that the validity period of sentiment in individual stocks in the investment was made at daily, weekly, and monthly frequencies, the effectiveness of the investment was found in daily, but stocks is as short as daily. its effectiveness was lost in weekly and monthly frequencies. This reveals that the validity period of sentiment in individual stocks is as short as daily.
Why that problem is chosen in this paper? Scope of the problem and solution (Refer Introduction) (5-8 lines)	The reason was people were loosing interest and faith in Japanese Stocks. To make the flow of money constant in the market they made this attempt of boosting stock market.
History of the problem. (Refer Introduction) (8-10 lines)	The problem started with deduction in flow of money in stock market. Not only this Tokyo Stock Exchange was observing a huge descent in share price.
List of the related/similar problems (Refer Related work) – Describe each with proposed solutions	
Related problem 1 – Describe (3-4 lines)	3D Data management: Controlling data volume, velocity and variety. Meta Group

Paper in APA style	Laney, D. (2001). 3D Data management: Controlling data volume, velocity and variety. Meta Group. <i>Lakshen, Guma Abdulkhader</i> , 1-4.
Related problem 2 – Describe (3-4 lines)	Data Revolution and Equity Investment
Paper in APA style	Suwabe, T. (2015). Data Revolution and Equity Investment. <i>Securities Analysts Journal</i> , 53(4), 6-17.
Related problem 3 – Describe (3-4 lines)	Nowcasting with payments system data
Paper in APA style	Galbraith, J. W., & Tkacz, G. (2018). Nowcasting with payments system data. <i>International Journal of Forecasting</i> , 34(2), 366-376.
Related problem 4 – Describe (3-4 lines)	The Power of Voice: Managerial Affective States and Future Firm Performance
Paper in APA style	Mayew, W. J., & Venkatachalam, M. (2012). The power of voice: Managerial affective states and future firm performance. <i>The Journal of Finance</i> , 67(1), 1-43.
Related problem 5 – Describe (3-4 lines)	A Method of Using News Sentiment for Stock Investment Strategy
Paper in APA style	Katayama, D., & Tsuda, K. (2020). A Method of Using News Sentiment for Stock Investment Strategy. <i>Procedia Computer Science</i> , 176, 1971-1980.

<p>What is the proposed solution in this paper for the problem chosen? (Refer Proposed work) (5-8 lines)</p>	<p>The term "big data" has been attracting attention as a new keyword since 2010. According to Laney ,data with three V's as its characteristics is often defined as a set of data so huge and complex that its difficult to handled by general data management and processing software. In the asset management industry as well, the use of big data is progressing, and there is a study of Nowcasting GDP with electronic payment data, an analysis of the voice of a company's representative at an earnings conference to use it for forecasting, and a study of the relationship between policy change and the expression of the central bank governor. In fact, a large amount of information arrives at the fund managers who manage the assets entrusted to them. In addition to</p>																		
<p>Architecture of the proposed solution. (Refer proposed work) Diagram</p>	<p>Table 2. Summary statistics of sentiment score.</p> <table border="1" data-bbox="1072 731 1691 1155"> <thead> <tr> <th></th> <th style="text-align: right;">Sentiment model</th> </tr> </thead> <tbody> <tr> <td>Total</td> <td style="text-align: right;">367,215</td> </tr> <tr> <td>Mean</td> <td style="text-align: right;">0.566</td> </tr> <tr> <td>Standard deviation</td> <td style="text-align: right;">0.180</td> </tr> <tr> <td>Minimum</td> <td style="text-align: right;">0.000</td> </tr> <tr> <td>25%</td> <td style="text-align: right;">0.451</td> </tr> <tr> <td>Median</td> <td style="text-align: right;">0.579</td> </tr> <tr> <td>75%</td> <td style="text-align: right;">0.691</td> </tr> <tr> <td>Maximum</td> <td style="text-align: right;">1.000</td> </tr> </tbody> </table>		Sentiment model	Total	367,215	Mean	0.566	Standard deviation	0.180	Minimum	0.000	25%	0.451	Median	0.579	75%	0.691	Maximum	1.000
	Sentiment model																		
Total	367,215																		
Mean	0.566																		
Standard deviation	0.180																		
Minimum	0.000																		
25%	0.451																		
Median	0.579																		
75%	0.691																		
Maximum	1.000																		
<p>Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)</p>	<p>Sentimental Analyser Model</p>																		

<p>List of existing algorithms used by the authors to complete the proposed work. (1-2 lines for each algorithm)</p>	<p>LSTM Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications.</p>
<p>List of datasets used. (<i>Refer experimental evaluation/result discussion</i>) (3-4 lines)</p>	<p>TOPIX500 Stock Prices Consists of data from January 1983 to the end of December 2016.</p>
<p>References/links to each of the dataset used in this paper (in APA style)</p>	<p>1. https://in.investing.com/indices/topix-500</p>
<p>Why the above dataset(s) used? <i>(Refer experimental evaluation/result discussion)</i> (3-4 lines)</p>	<p>The top 20 stocks in TOPIX are approximately 25% of the total weighting, and the top 70 stocks make up half the market cap of TOPIX. This makes the reason of selecting TOPIX since they form a huge share of Japanese Stock Market. Thus a fluctuation in them can be considered as a fluctuation in whole market.</p>
<p>List of equations that are very well applied in this problem domain</p>	<p>Description: $NAV_{i,T}$ is the market capitalisation of the stock i at time T.</p> $\mathbf{w}_{eq,i,T} = \frac{\mathbf{1}}{n}$ $\mathbf{w}_{cap,i,T} = \frac{NAV_{i,T}}{\sum_{i=1}^n NAV_{i,T}}$

		Table 3. Performance summary statistics of daily rebalance strategy.							
		Equal-Weighted				Market-Value-Weighted			
		Q1(positive)	Q2(negative)	ALL	Q1-Q2	Q1(positive)	Q2(negative)	ALL	Q1-Q2
Return		13.8%	4.0%	8.8%	--	7.8%	2.9%	5.4%	--
Risk		21.1%	21.4%	20.5%	--	23.2%	23.1%	22.2%	--
Return/Risk		0.66	0.19	0.43	--	0.34	0.12	0.24	--
Alpha		5.0%	-4.8%	--	9.8%	2.4%	-2.5%	--	5.0%
T.E.		5.7%	5.3%	--	11.0%	7.5%	7.4%	--	14.4%
IR		0.89	-0.89	--	0.89	0.32	-0.34	--	0.34
p-value		0.00%	0.00%	--	0.00%	5.86%	4.55%	--	4.43%
Turnover		21667%	20968%	18170%	--	19743%	19198%	13812%	--
Names		23	24	46	--	23	24	46	--

(a) Equal-Weighted

(b) Market-Value-Weighted

Fig. 1. Cumulative excess return of daily rebalance strategy (a) equal-weighed; (b) market-value-weighted.

List of method(s)/metrics used to evaluate the proposed approach. *(Refer experimental evaluation/result discussion)* **(5-8 lines)**	
List of supporting tools/concepts **(3-4 lines)**	**CoreML** A ML model generator to generate a ML model which selects algorithm based on the input and precession.
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: Performance Graph Observation** In this approach we study previous stats to analyse the market prediction. This method was the old tradition which we used.

<p>How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)</p>	<p>Statistical Comparison</p> <p>Statistics of stock prices are compared with the results obtained through this method. The comparison is done directly using precision rate.</p>
<p>Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)</p>	<p>This method seems to be really helpful when market seems to crash or descent. The descent rate of market can be highly reduced using this method.</p>
<p>Disadvantages/limitations of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)</p>	<p>The disadvantage is making news headlines in order to maintain the share market is against of human faith.</p> <p>Not only this, by this way any government body can easily manipulate future results.</p>
<p>Future work as stated by authors (Refer conclusion / result discussion / experimental evaluation)</p>	<p>Future works include making this method more versatile by adding social media app based reviews, comments made on a particular brand. By this way we can achieve more precision rate.</p>
<p>Your one page write-up about this paper</p>	

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

Your findings: (possible alternate for the solution proposed)

- This method seems to be life saviour when worst case arise.
- There are certain flaws which might make wrong prediction, but the precision rate is high and cause of irregularity is highly low so we can trust this method.

3.

TABLE 2

Problem answered in this paper. (1-2 lines)	Due to its dynamics, non-linearity and complexity nature, stock market is inherently difficult to predict.
Detailed description about the problem (5-8 lines)	Stock market prediction has become highly unpredictable due to various kind of dynamics which are changing with high rate. With this high rate of change, prediction had become really hard. The reason is with every second a new data gets added and to manage prediction of such huge dataset is tough. Not only that, with the traditional methods predictions are made with repetition of certain events at every uniform duration.
Why that problem is chosen in this paper? Scope of the problem and solution (<i>Refer Introduction</i>) (5-8 lines)	Stock market movement prediction has massive benefits in academia and industry. In particular, accurate prediction helps investors make decisions and gain profit in the stock exchange. However, this prediction task is challenging due to the financial data nature that comprises noise, non-stationary, high degree of uncertainty, and chaotic characteristics. More- over, the complex interaction of political and economic factors makes market prediction more difficult.
History of the problem. (<i>Refer Introduction</i>) (8-10 lines)	The need of this matter arose when traditional methods couldn't manage to predict precise price. This was the reason why we were supposed to find a new method.
List of the related/similar problems (<i>Refer Related work</i>) – Describe each with proposed solutions	
Related problem 1 – Describe (3-4 lines)	Machine Learning Techniques and Use of Event Information for Stock Market Prediction: A Survey and Evaluation
Paper in APA style	Yoo, P. D., Kim, M. H., & Jan, T. (2005, November). Machine learning techniques and use of event information for stock market prediction: A survey and evaluation. In International Conference on Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce (CIMCA-IAWTIC'06) (Vol. 2, pp. 835-841). IEEE.

Related problem 2 – Describe (3-4 lines)	A survey on opinion mining and sentiment analysis: Tasks, approaches and applications
Paper in APA style	Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. <i>Knowledge-based systems</i> , 89, 14-46.
Related problem 3 – Describe (3-4 lines)	Twitter Mood as a Stock Market Predictor
Paper in APA style	Bollen, J., & Mao, H. (2011). Twitter mood as a stock market predictor. <i>Computer</i> , 44(10), 91-94.
Related problem 4 – Describe (3-4 lines)	Social media as a source of knowledge for customers and enterprises
Paper in APA style	Zembik, M. (2014). Social media as a source of knowledge for customers and enterprises. <i>Online Journal of Applied Knowledge Management</i> , 2(2), 132-148.
Related problem 5 – Describe (3-4 lines)	Modeling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena
Paper in APA style	Bollen, J., Mao, H., & Pepe, A. (2011, July). Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In <i>Proceedings of the international AAAI conference on web and social media</i> (Vol. 5, No. 1).
What is the proposed solution in this paper for the problem chosen? (Refer Proposed work) (5-8 lines)	A comparison of our approach with other sentiment-based stock market prediction approaches including Deep learning, establishes that our proposed model is performing adequately and predicting stock movements with a higher accuracy of 60%.

<p>Architecture of the proposed solution. (Refer proposed work)</p> <p>Diagram</p>	<pre> graph LR I[Step I: Data Collection] --> II[Step II: Data Preprocessing & Features Extraction LDA & Granger Causality] II --> III[Step III: Fitting & Evaluation] III --> ACC{ACC > T} ACC -- YES --> V[Step V: Models Stacking & Regularization] ACC -- NO --> IV[Step IV: Feature Selection/Extraction] IV --> II </pre>
<p>Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)</p>	<p>Dickey-fuller test to check stationarity.</p> <p>Machine learning techniques.</p> <p>Granger causality test for the four stocks.</p>

<p>List of existing algorithms used by the authors to complete the proposed work. (1-2 lines for each algorithm)</p>	<p>DNN A DNN is a collection of neurons organized in a sequence of multiple layers, where neurons receive as input the neuron activations from the previous layer, and perform a simple computation</p> <p>Deep CNN A Convolutional Neural Network (ConvNet/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</p> <p>LSTM Long short-term memory (LSTM) is an artificial recurrent neural network (RNN) architecture used in the field of deep learning. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications.</p>
<p>List of datasets used. (Refer experimental evaluation/result discussion) (3-4 lines)</p>	<p>NASDAQ Stock Price The Nasdaq Stock Market is an American stock exchange based in New York City. It is ranked second on the list of stock exchanges by market capitalization of shares traded, behind the New York Stock</p>
<p>References/links to each of the dataset used in this paper (in APA style)</p>	<p>https://www.nasdaq.com</p>
<p>Why the above dataset(s) used? (Refer experimental evaluation/result discussion) (3-4 lines)</p>	<p>NASDAQ is the second largest stock market in the world in terms of its capitalisation. Being such huge share holders their data sets will be numerous, such that we can achieve high precision.</p>

	<p>where p_t is the price of the stock price at time t. In other words, y_i indicates whether the stock price is ‘1 = up’ or ‘0 = down’. We will apply different machine learning algorithms to predict y_i.</p> $y_t = \begin{cases} 1, & \text{if } p_t > p_{t-1} \\ 0, & \text{otherwise} \end{cases}$
<p>List of method(s)/metrics used to evaluate the proposed approach. <i>(Refer experimental evaluation/result discussion)</i> (5-8 lines)</p>	<p>Dickey-fuller test to check stationarity.</p> <p>Machine learning techniques.</p> <p>Granger causality test for the four stocks.</p> <p>Tweet mining</p>
<p>List of supporting tools/concepts (3-4 lines)</p>	<p>Twitter Developer APIs In order to access Twitter APIs, you must first apply for a developer account</p> <p>NLP Kit Natural Language Processing, or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software.</p>

<p>What are the similar approaches with which the proposed approach is compared? (Refer experimental evaluation/result discussion)</p> <p>Explain each of these approach (3-4 lines)</p>	<p>Approach/method 1: Statistical Comparison Comparing statistics of traditional and sentimental datas. By this we can find the precision rate.</p> <p>Approach/method 2: Graphical Comparison Comparing graphical data to find difference between both methods.</p>
<p>How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)</p>	<p>Statistical Comparison Statistics of stock prices are compared with the results obtained through this method. The comparison is done directly using precision rate.</p>
<p>Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)</p>	<p>The best advantage is we can predict stock price more accurately and more precise. Using this method will help us in avoiding high mathematical computation.</p>
<p>Disadvantages/limitations of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)</p>	<p>The demerit is still there are chances that data might get false in some exceptional case when false reviews are done.</p>
<p>Future work as stated by authors (Refer conclusion / result discussion / experimental evaluation)</p>	<p>For the sake of maturing sentiment analysis for stock market predictions, more researches are needed towards (1) improving the representation of sentiment as a set textual features and (2) leveraging the abilities of machine learning algorithms. Accordingly, given the dynamic and complex nature of stock time series data, our future research involves mainly the investigation of variations of deep learning techniques for both sentiment features engineering and prediction modelling of the stock movement.</p>

Your one page write-up about this paper

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

Your findings: (possible alternate for the solution proposed)

- The work shows computation is really easy with this method rather than the traditional method.
- This work targets on one of the largest stock exchange market, thus the precision rate will be really high.

Research Questionnaire

Note:

- 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	<ol style="list-style-type: none">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 Learning3. Sentiment Analysis of Indian Stock Market Volatility
Citation (APA style)	<ol style="list-style-type: none">1. Sun, Y., Wu, M., Zeng, X., & Peng, Z. (2021). The impact of COVID-19 on the Chinese stock market: Sentimental or substantial? <i>Finance Research Letters</i>, 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. <i>IEEE Access</i>, 8, 161943-161957.3. Paramanik, R. N., & Singhal, V. (2020). Sentiment Analysis of Indian Stock Market Volatility. <i>Procedia Computer Science</i>, 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	<ol style="list-style-type: none"> 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. <i>Journal of behavioral and experimental finance</i> , 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?. <i>Research in International Business and Finance</i> , 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. <i>Finance Research Letters</i> , 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. <i>Finance research letters</i> , 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. <i>International Review of Financial Analysis</i> , 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	<p>Time frames used for Panel Regression model</p>
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)	<ol style="list-style-type: none"> 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)	<ol style="list-style-type: none"> 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)	<ol style="list-style-type: none"> 1. https://www.gtarsc.com/ 2. https://ifind.bnu.edu.cn/
Why the above dataset(s) used? (Refer experimental evaluation/result discussion)	<ol style="list-style-type: none"> 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)	<p>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.</p>
<p>List of equations that are very well applied in this problem domain</p>	<p>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</p> <p>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</p> <p>Equation 3: $CAR = \sum_{t=1}^n AR_t$ Description: individual abnormal returns to create a “cumulative abnormal return (CAR)”</p>

	Table 1.. Summary Statistics.				
	Variable	Observations	Mean	SD	Min
Panel A: Estimation Window					
	Market return	100	0.001	0.008	-0.019
	Stock return	324,729	0.001	0.022	-0.102
	Sentiment	324,729	0.486	2.636	-19.880
Panel B: Event Window					
	Market return	10	-0.005	0.030	-0.080
	Stock return	31,475	-0.005	0.046	-0.103
	Sentiment	31,475	0.431	2.728	-10.912
Panel C: Post-event Window					
	Market return	36	-0.001	0.018	-0.039
	Stock return	106,164	0.000	0.034	-0.109
	Sentiment	106,164	0.627	2.550	-10.533
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/).					
List of method(s)/metrics used to evaluate the proposed approach. (Refer experimental evaluation/result discussion) (5-8 lines)	1. Panel regression model 2. Fama–French model				
List of supporting tools/concepts (3-4 lines)					
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: 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				

<p>How the results of proposed approach are compared with other similar approaches? <i>(Refer experimental evaluation/result discussion)</i></p>	<ol style="list-style-type: none"> 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.
<p>Advantages/merits of proposed solution in your view. <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<p>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.</p>
<p>Disadvantages/limitations of proposed solution in your view. <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<p>This paper is mainly focused on only 7 main industries so this cannot be used as a basis for other industries.</p>
<p>Future work as stated by authors <i>(Refer conclusion / result discussion / experimental evaluation)</i></p>	<p>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.</p>
<p>Your one page write-up about this paper</p>	
<p>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</p>	

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.

Your findings: (possible alternate for the solution proposed)

- 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. <i>(Refer Introduction)</i> (8-10 lines)	<p>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.</p> <p>Natural language based financial forecasting</p> <p>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,</p> <p>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.</p>
List of the related/similar problems (<i>Refer Related work</i>) – 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 <i>Semi-supervised learning</i> (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. <i>Neurocomputing</i> , 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. <i>Knowledge-Based Systems</i> , 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. <i>IEEE access</i> , 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 <i>4th International Semantic Search Workshop</i> . Retrieved from: http://km.aifb.kit.edu/ws/semssearch11/11.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	<p>Database</p> <ul style="list-style-type: none"> News services posted on portal sites Bulletins Social network service <p>Process</p> <ul style="list-style-type: none"> Data collection Fake information filtering Credit risk assessment & risk signal detection Forecasting possibility of credit event <p>Methodology</p> <ul style="list-style-type: none"> Web crawling Trend analysis Sentiment analysis Word2Vec Graph-based semi-supervised learning Logistic regression
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. <i>(1-2 lines for each algorithm)</i>	<ol style="list-style-type: none"> 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. <i>(Refer experimental evaluation/result discussion)</i>	<ol style="list-style-type: none"> 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/refsrences 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	<p>Equation 1: $i = \sum_{j=1}^m (P_i, j)$ Description: Sentimental value of keyword</p> <p>Equation 2: $P_{i,j} = \frac{1}{r^2} * (\text{Sentimental value of core keyword } i)$ Description: Proximity index between keyword i and core keyword j</p> <p>Equation 3: $r_j = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$ ($j=1,2,\dots,m$) Description: Euclidean distance between keyword i and core keyword j</p> <p>Equation 4: $I = \sum_{i=1}^n W_i S_i$ Description: I stands for Indicator for monitoring</p>
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

<p>the proposed approach is compared? (Refer experimental evaluation/result discussion)</p> <p>Explain each of these approach (3-4 lines)</p>	<p>on websites and SNSs. opinions are also collected from finance websites, online communities, and SNSs.</p> <p>After collecting raw data from web services, these data need to be refined to clarify documents and derive more accurate results.</p> <p>Approach/method 2: CREDIT RISK ASSESSMENT AND RISK SIGNAL DETECTION</p> <p>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.</p> <p>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) tagging and parsing. This preprocessing aims to rate documents using the sentiment values of sentences and words.</p> <p>Approach/method 3: FORECASTING OCCURRENCE POSSIBILITY OF CREDIT EVENT</p> <p>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.</p>
<p>How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)</p>	<p>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.</p> <p>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.</p> <p>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.</p>

	<p>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.</p> <p>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).</p>
Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)	<p>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.</p>
Disadvantages/limitations of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)	<p>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.</p>
Future work as stated by authors (Refer conclusion / result discussion / experimental evaluation)	<p>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.</p> <p>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.</p> <p>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.</p>
Your one page write-up about this paper	
<p>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</p>	

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. <i>Cogent Economics & Finance</i> , 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. <i>Open Journal of Statistics</i> , 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. <i>Journal of computational science</i> , 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. <i>The Journal of Finance and Data Science</i> , 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. <i>Baltic Journal of Management</i> .
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 convert all the texts to lower case so that NLTK can process it easily. Next tokenization is done so that the words in the string. Next, Stop-words are removed from the string by the help of NLTK so that common words can be removed that do not generate any significance. In the further step, Lemmatization is done on various similar split words to get one common word as this helps us get better frequency insights. The pre-processed data is then converted to a frequency table using sklearn

	<p>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.</p>
Architecture of the proposed solution. (Refer proposed work) Diagram	<pre> graph LR Start([Start]) --> Lowercase[Conversion of all text to lowercase for easy processing by NLTK] Lowercase --> Check{Check if all are lowercase or not} Check -- No --> Lowercase Check -- Yes --> Tokenize[Tokenization and splitting up of words] Tokenize --> Remove[Removal of stop-words] Remove --> Lemmatize[Lemmatization] Lemmatize --> Convert[Conversion to a frequency table by the help of sklearn in python.] Convert --> Lemmatized[Lemmatized words are further given sentiment scores based on standard NRC emotion lexicon] Lemmatized --> Calculate[Calculation of Index Intensity and classification into 8 emotions] Calculate --> Classify[Classifying 8 emotion into 2 broad classes of positive and negative and applying PCA within the groups] Classify --> Share[Calculating Relative Share of each market sentiment] Share --> Finish([Finish]) </pre>
Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)	NLTK approach with Volatility Modelling
List of existing algorithms used by the authors to complete the proposed work. (1-2 lines for each algorithm)	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
List of datasets used. (Refer	The datasets that are used in the above study are the prominent web sources for Indian financial

<i>experimental evaluation/result discussion) (3-4 lines)</i>	<p>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.</p>
List of equations that are very well applied in this problem domain	<p>Equation 1 $return = \ln\left(\frac{P_t}{P_{t-1}}\right) * 100$ Description: - This formula calculates the return of Sensex. P_t is sensex index on day t.</p> $IE_{e,t} = \sum_{w} (f_{w,t}) (I_{w,e})$ <p>Equation 2: -</p> <p>Description: - Intensity index for each emotion is calculated by the above formula. $IE_{e,t}$ refers to calculated index of emotion e on day t, $f_{w,t}$ refers to frequency of occurrence of word w on day t and $I_{w,e}$ refers to intensity of emotion e associated with word w.</p> $POS = \frac{S_p}{S_p + S_n}$ <p>Equation 3: -</p> <p>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.</p> <p>Equation 4: - $NEG = 1 - POS$</p> <p>Description: - Negative market sentiment is calculated by the above formula.</p> $Y_t = \mu + \alpha * Y_{t-1} + \beta * \varepsilon_{t-1} + \varepsilon_t$ <p>Equation 5: -</p>

	<p>Description: - This is the mean equation of a standard GARCH model. YY_{tt} 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.</p> $h_t = \rho + c * \varepsilon_{(t-1)}^2 + \delta * h_{(t-1)} + \gamma * \varepsilon_{(t-1)}^2 * D_{(t-1)}$ <p>Equation 6: -</p> <p>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_{t-1}^2. 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.</p> $h_t = c * \varepsilon_{(t-1)}^2 + \delta * h_{(t-1)} + \gamma * \varepsilon_{(t-1)}^2 * D_{(t-1)} + \varphi_1 * POS + \varphi_2 * NEG$ <p>Equation 7: -</p> <p>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.</p>
<p>List of method(s)/metrics used to evaluate the proposed approach. (Refer experimental evaluation/result discussion) (5-8 lines)</p>	<p>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.</p> <p>In the augmented conditional volatility equation, $(c+\delta)$ is less than unity which confirms stability of the model and since $(c+\delta)$ 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</p>

	<p>sentiment on market volatility and found to be positive.</p>
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)	<p>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</p> <p>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.</p>
How the results of proposed approach are compared with other similar approaches? (Refer experimental evaluation/result discussion)	<p>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.</p>
Advantages/merits of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)	<p>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.</p>
Disadvantages/limitations of proposed solution in your view. (Refer conclusion / result discussion / experimental evaluation)	<p>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.</p>
Future work as stated by authors (Refer conclusion / result discussion / experimental evaluation)	<p>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.</p>

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)

-