Research Questionnaire

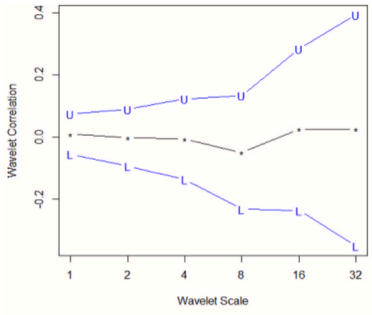
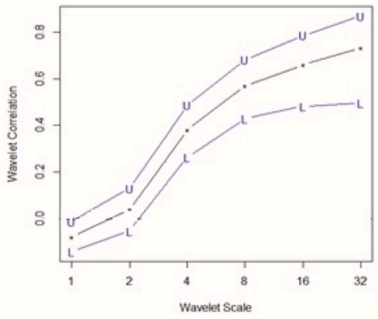
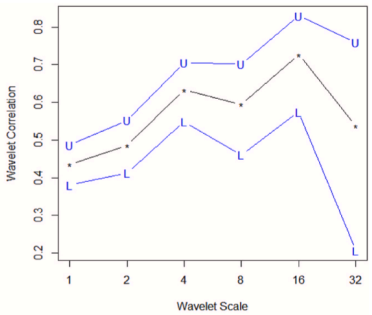
Note:

- **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	1. Estimating the relative effects of raw material prices, sectoral outlook and market sentiment on stock prices 2. Exploiting textual and relationship information for fine-grained financial sentiment analysis 3. Harvesting social media sentiment analysis to enhance stock market prediction using deep learning
Citation *(APA style)*	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 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. *International Journal of Business Forecasting and Marketing Intelligence*, *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. *Turkish Economic Review*, *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. *International Journal of Business Innovation and Research*, *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. *Resources Policy*, *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). *The cross-section of expected stock returns* (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**	 Reliance Stock Price and Crude Oil Price  Reliance Stock Price and Oil & Gas Index  Reliance Stock Price and SENSEX
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)

List of existing algorithms used by the authors to complete the proposed work. **(1-2 lines for each algorithm)**	Random Forest : It is a machine learning algorithm used to solve regression and classification problems. 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. Genetic Algorithm : It is a method for solving constrained and unconstrained optimisation problems which are based on natural selection process.
List of datasets used. (*Refer experimental evaluation/result discussion*) **(3-4 lines)**	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.
References/links to each of the dataset used in this paper (in APA style)	1. https://journals.sagepub.com/doi/10.1177/2321022217724978 2. https://www.bseindia.com 3. https://www.nseindia.com
Why the above dataset(s) used? (*Refer experimental evaluation/ result discussion*) **(3-4 lines)**	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.

List of equations that are very well applied in this problem domain	Equation 1: CWT of a time series $$W_x(\tau, s) = \frac{1}{\sqrt{\lvert s \rvert}} \int_{-\infty}^{+\infty} x(t)\psi^*\left(\frac{t-\tau}{s}\right)dt$$ Description: Continuous wavelet transform based coherence analysis Equation 2: MODWT estimator of wavelet correlation $$\rho_{xy}(\lambda_j) = Corr\left(w_{ijt}, \tilde{w}_{ijt}\right) = \frac{Cov\left(w_{ijt}, \tilde{w}_{ijt}\right)}{\sqrt{Var\left(w_{ijt}\right) Var\left(\tilde{w}_{ijt}\right)}}$$ Description: The MODWT estimator of wavelet correlation has been used in this study which basically considers the covariance of two series ($x(t)$, $y(t)$) and wavelet variance of individual series.
List of method(s)/metrics used to evaluate the proposed approach. (*Refer experimental evaluation/ result discussion*) **(5-8 lines)**	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
List of supporting tools/concepts (3-4 lines)	tsfeatures' package of R, Jarque-Bera Test , ADF Test , ZA Test, ARCH statistics, SpectralEntropy, Hurst Exponent

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: 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.
How the results of proposed approach are compared with other similar approaches? (*Refer experimental evaluation/result discussion*)	There are many packages available for wavelet analysis like Mathematica, Maple, Matlab, R, and PyWavelets. Which is not the case with FTT. 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. 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.
Advantages/merits of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	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 discontinuties 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.

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

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.

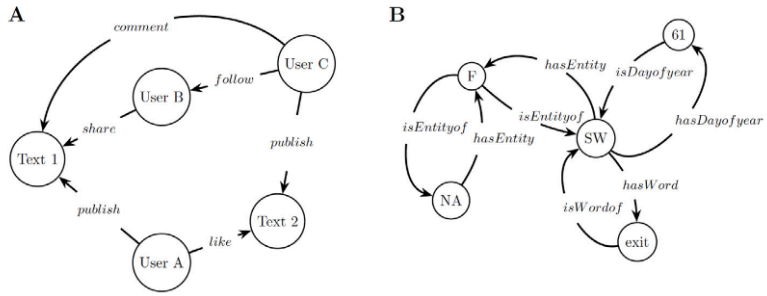
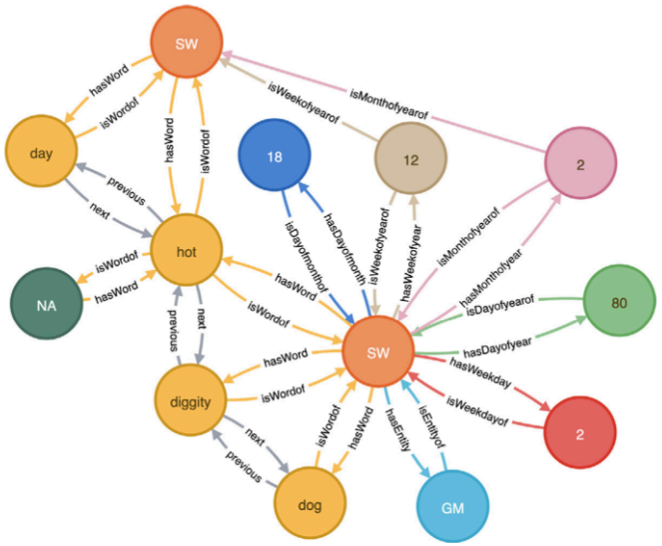
Your findings: (possible alternate for the solution proposed)

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

Why that problem is chosen in this paper? Scope of the problem and solution (*Refer Introduction*) **(5-8 lines)**	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.
History of the problem. (*Refer Introduction*) **(8-10 lines)**	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. 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.
List of the related/similar problems (*Refer Related work*) – Describe each with proposed solutions	
Related problem 1 – Describe **(3-4 lines)**	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.
Paper in APA style	Daudert, T. (2021). A multi-source entity-level sentiment corpus for the financial domain: the FinLin corpus. *Language Resources and Evaluation*, 1-24.
Related problem 2 – Describe **(3-4 lines)**	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.

Paper in APA style	Sinha, N. R. (2016). Underreaction to news in the US stock market. *Quarterly Journal of Finance*, *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. *arXiv preprint arXiv:1709.05584*.
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 *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (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 *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.* (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).

Architecture of the proposed solution. (*Refer proposed work*) **Diagram**	 **Fig. 2.** Overview of a social network graph structure compared to the graph proposed in this work. NA denotes a news article and SW stockswit.  **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^r 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, green, and red depending on the type of temporal class, as specified in Section 3.4.
Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)	Fine Tuned Textual Representation FFNN (feed forward neural network)

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

List of equations that are very well applied in this problem domain	Equation 1: Equation to calculate mean squared error (MSE) : $$MSE(P,T) = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(T_i - P_i)^2}$$ Description: mean square error measures the average of the squares of the error. Equation 2: Equation to calculate Cosine Similarity(CS) $$CS(T,P) = \frac{\sum_{i=1}^{n} T_i \times P_i}{\sqrt{\sum_{i=1}^{n} T_i^2} \times \sqrt{\sum_{i=1}^{n} P_i^2}}.$$ Description : cosine similarity is the cosine of the angle between to N-dimensional vectors in an N-dimensional space.
List of method(s)/metrics used to evaluate the proposed approach. (*Refer experimental evaluation/ result discussion*) **(5-8 lines)**	1. Mean squared error (MSE) 2. Cosine Similarity (CS)
List of supporting tools/concepts (3-4 lines)	SOTA Single Text sentiment analysis Contextual sentiment analysis Deep Learning Implicit sentiments

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: 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. 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. Approach/method 3: aFFNN
How the results of proposed approach are compared with other similar approaches? (*Refer experimental evaluation/result discussion*)	The improvement from baseline was seen from 0.68 - 15.1% and 59.46–234.15% for the CS and MSE, respectively. 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.
Advantages/merits of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	1. Problems in FFNN are represented by attribute-value pairs. 2. 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. 3. It is used where the fast evaluation of the learned target function required.
Disadvantages/limitations of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	1. It is highly dependent on hardware. 2. Lack of assurance of proper network structure. 3. The difficulty to show the problem to the network.
Future work as stated by authors (*Refer conclusion / result discussion / experimental evaluation*)	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 one page write-up about this paper	

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 to 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.

Why that problem is chosen in this paper? Scope of the problem and solution (*Refer Introduction*) **(5-8 lines)**	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.
History of the problem. (*Refer Introduction*) **(8-10 lines)**	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.
List of the related/similar problems (*Refer Related work*) – Describe each with proposed solutions	
Related problem 1 – Describe **(3-4 lines)**	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.
Paper in APA style	Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies, practices and applications. *International Journal of Scientific and Technology Research*, *9*(2), 601-609.
Related problem 2 – Describe **(3-4 lines)**	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.

Paper in APA style	Milosevic, N. (2016). Equity forecast: Predicting long term stock price movement using machine learning. *arXiv preprint arXiv:1603.00751*.
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. *The Journal of Supercomputing*, *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. *Procedia computer science*, *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. *International Review of Financial Analysis*, *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

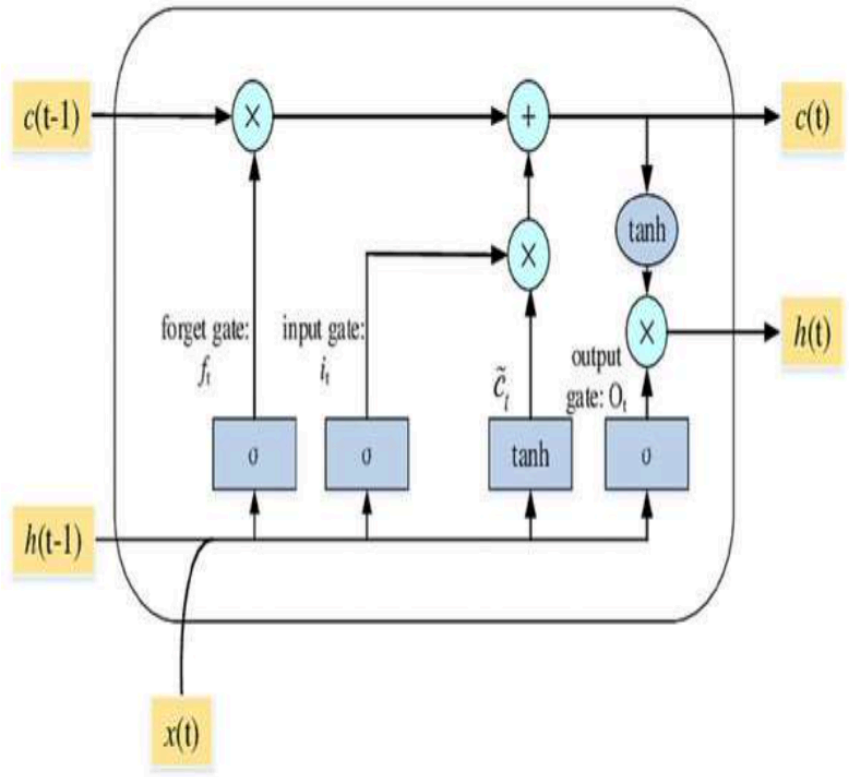


Figure 4: The LSTM model with its three layers.

DOI: 10.7717/peerj-cs.476/fig-4

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 were connections between notes former 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 sequence 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? (*Refer experimental evaluation/ result discussion*) **(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 the one of largest social media to house tweets related to stock market and share prices of numerous companies. NSE stock data is also the most reliant when it comes to stocks of Indian companies.

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

Equation 1: Sentiment Formulas

Table 1:
Predominant sentiment formulas.

Precision	Recall	*F*-score	Accuracy
∑Ŧp ∑Ŧp +∑Fp	∑Ŧp ∑Tp + ∑Fn	2∑ Ŧp/(2∑ Ŧp + ∑FP + ∑FN)	$\frac{\sum Ŧp + \sum Tn}{\sum Ŧp + \sum Tn + \sum Fp + \sum Fn}$

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

Equation 2: Various LSTM formulas were used :

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

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

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

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

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

List of method(s)/metrics used to evaluate the proposed approach. (*Refer experimental evaluation/ result discussion*) **(5-8 lines)**	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 determines whether it will increase or decrease.
List of supporting tools/concepts (3-4 lines)	Social Media Sites like Twitter, Facebook were used. RNN algorithm was used for LSTM method. Deep Learning. Techniques were also helpful.
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: Naive Bayes Technique Approach/method 2: Linear Regression Approach/method 3: Maximum entropy Approach/method 4: Decision Tree Approach/method 5: Linear SVC Regression
How the results of proposed approach are compared with other similar approaches? (*Refer experimental evaluation/result discussion*)	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.
Advantages/merits of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	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.

Disadvantages/limitations of proposed solution in your view. (*Refer conclusion / result discussion / experimental evaluation*)	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.
Future work as stated by authors (*Refer conclusion / result discussion / experimental evaluation*)	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.

Your one page write-up about this paper

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.
The following research paper provided the following:
It suggested combination of economic news and social network analysis to forecast SM developments.
Implemented a collection of features from final data set to improve the performance of predictions.
This paper recommended impact on SM from social networking sites like Twitter, Facebook, Google plus and financial forecast.

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