

Khalil, Faisal; Pipa, Gordon

**Article — Published Version**

## Is Deep-Learning and Natural Language Processing Transcending the Financial Forecasting? Investigation Through Lens of News Analytic Process

Computational Economics

**Provided in Cooperation with:**

Springer Nature

*Suggested Citation:* Khalil, Faisal; Pipa, Gordon (2021) : Is Deep-Learning and Natural Language Processing Transcending the Financial Forecasting? Investigation Through Lens of News Analytic Process, Computational Economics, ISSN 1572-9974, Springer US, New York, NY, Vol. 60, Iss. 1, pp. 147-171,  
<https://doi.org/10.1007/s10614-021-10145-2>

This Version is available at:

<https://hdl.handle.net/10419/286994>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



# Is Deep-Learning and Natural Language Processing Transcending the Financial Forecasting? Investigation Through Lens of News Analytic Process

Faisal Khalil<sup>1</sup> · Gordon Pipa<sup>1</sup>

Accepted: 22 June 2021 / Published online: 22 July 2021  
© The Author(s) 2021

## Abstract

This study tries to unravel the stock market prediction puzzle using the textual analytic with the help of natural language processing (NLP) techniques and Deep-learning recurrent model called long short term memory (LSTM). Instead of using count-based traditional sentiment index methods, the study uses its own sum and relevance based sentiment index mechanism. Hourly price data has been used in this research as daily data is too late and minutes data is too early for getting the exclusive effect of sentiments. Normally, hourly data is extremely costly and difficult to manage and analyze. Hourly data has been rarely used in similar kinds of researches. To built sentiment index, text analytic information has been parsed and analyzed, textual information that is relevant to selected stocks has been collected, aggregated, categorized, and refined with NLP and eventually converted scientifically into hourly sentiment index. News analytic sources include mainstream media, print media, social media, news feeds, blogs, investors' advisory portals, experts' opinions, brokers updates, web-based information, company' internal news and public announcements regarding policies and reforms. The results of the study indicate that sentiments significantly influence the direction of stocks, on average after 3–4 h. Top ten companies from High-tech, financial, medical, automobile sectors are selected, and six LSTM models, three for using text-analytic and other without analytic are used. Every model includes 1, 3, and 6 h steps back. For all sectors, a 6-hour steps based model

---

✉ Faisal Khalil  
[fkhalil@uni-osnabrueck.de](mailto:fkhalil@uni-osnabrueck.de)

Extended author information available on the last page of the article

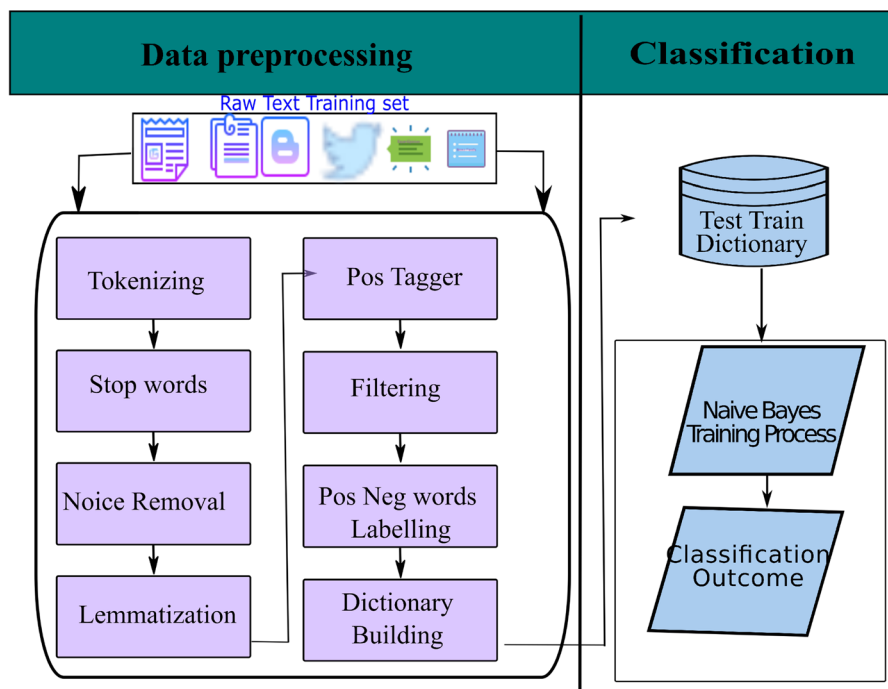
outperforms the other models due to LSTM specialty of keeping long term memory. Collective accuracy of textual analytic models is way higher relative to non-textual analytic models.

**Keywords** LSTM · Natural language processing · News analytic · Sentiment analysis · Stock prediction

## 1 Introduction

Accurate forecasting of returns is crucial for individual investors, investment banks and corporate investment managers. It is also equally important for investors to foresee the returns accurately and design the investment or trading strategies keeping in view all relevant aspects of forecasting. For many years stock market forecasting studies have been emphasizing the volatility models. Few studies inculcate the role of technological forecasting i.e. artificial intelligence. The efficient market hypothesis (EMH) is proposed by Fama (1998) is under criticism, because the proposed model is in contrast to the behavioral finance concept (Kahneman & Tversky, 1979; Kahneman, 2003; Shefrin, 2008). It has been much debated and considered as a limitation of EMH that this model is not considering the role of investor's sentiments and their behavioral aspect. Technological advancements and inventions of the new artificial intelligence-based model are reshaping the method of forecasting (Wang et al., 2018; Kuo & Huang, 2018; Makridakis et al., 2018). Normally, the artificial intelligence-based model takes previous stock prices and other variables into account but news analytic consideration is less researched. In this specific context news analytic is in the early stages and needs advancements for better forecasting and efficiency of intelligent trading systems. In the last decade, soft computing methods and techniques have grown rapidly that entice researchers to explore more sophisticated techniques for the stock market and time-series predictions. Time series financial modeling has a long history and time-series data is characterized by hidden relationships, high uncertainty, and unstructured in nature. To estimate the behavior of financial time series there are two types of models are available; linear model and non-linear models. Whereas, linear models are affected by techniques like Box Jenkins and Kalman filters, piece-wise regression, and Brown's exponential smoothing. All these theories are turning data into the linear functions. However, recent evidence shows that financial markets behave in a non-linear fashion. In addition to these problems, there are other factors that intact with financial markets like, general economic conditions, political events, news, and investor's psychology that makes a stock market prediction so difficult (Cheng & Chan, 2018; Huang et al., 2007). To address these issues, artificial intelligence has been evolved as a very good technique due to its learning, generalization, and non-linear behavior to overcome these problems and to give better forecasting (Makridakis et al., 2018; Li & Ma, 2010). In this connection, the most relevant techniques are; Recurrent Neural networks, Neural

Networks, fuzzy logic, and genetic algorithm (Hiransha et al., 2018; Ergen et al., 2017; Nelson et al., 2017; AlFalahi et al., 2014). Artificial Neural Networks model are pretty good with flexibility and adaptability to learn from changes and previous trend in a given set of input and predicts the trends based on network training. There is a fair deal of evidence that exists in the literature that models that based on artificial neural networks outperform the traditional time series model, for example, see Adebisi (2012), AlFalahi et al. (2014), Trippi and DeSieno (1992), Correa et al. (2009) and Hansson (2017). There are many soft-computing techniques available under the umbrella of artificial intelligence but finding appropriate techniques is very important to get accurate forecasting results. Study of Li et al. (2018) and Atsalakis and Valavanis (2009) can be referred here each has surveyed more than 100 articles by researchers who have used fuzzy logic, genetic algorithms, and neural networks and recurrent neural network as modeling techniques in their studies. It is evident from these articles that mostly researcher have used feed-forward neural networks (FFNN), currently, some studies use Recurrent Neural Networks(RNN) multilayer perceptron (MLP) to forecast the stock markets (Arora et al., 2019; Pawar et al., 2019). This survey study also testifies the magnitude of the importance of non-conventional tools for stock market prediction. For the stock market prediction process we cannot rely upon past stock prices and some other variable but we need to embed the impact of market news to achieve maximum accuracy. In the prediction process, it can be very tedious for managers to focus on every news that just pops up and align their investment strategies. A human being can miss much information and even information can be out of his reach as well. So, here natural language processing (NLP) techniques come into play. So, there is an urgent need to automate the news analysis process based on NLP technique so that the investment manager and the corporations can be benefited as well as AI-based predictive models can be supplied with more relevant information instead of just past prices. Natural language processing is a subfield of AI where Algos and deep learning model tries to make computers understand language intuitively near to the human level (Nadkarni et al., 2011). A human being has evolved from thousands of year training to understanding the emotion and feeling of language elicits but computers are struggling with the help of deep learning and AI-based models. In this study, we have used the NLP model (see Fig. 1) with the help naive Bayes classifier to process the raw information that is parsed out of many sources. These sources include mainstream media, print media, social media news feeds, blogs, investors' advisory portals, expert's opinions, brokers updates, web-based information, company' internal news and public announcements regarding policies and reforms. Detail of the news analytic and sentiment analysis can be seen in Sect. 3.1.2. Many studies propose soft computing techniques for better and most of the researches have focused on the comparison of traditional time series stock prediction models and artificial neural embedded network models. This study contributes to the existing body of knowledge in the following ways: Normally, studies use news information and stock price data for indices. Apart from other motivations to choose indices for the prediction process, one benefit is that data collection and aggregation is relatively easier



**Fig. 1** Natural language processing model

because of its ready availability. However, collecting news information for each company individually and make meaningful sentiments for that stock is challenging. However, this study focuses on individual-level stock and news information that makes this study bit challenging because not only news from all possible sources need to accumulate but also company internal news is also taken care. For example, the company changes the top echelon due to any reason or decides to change the level of dividends, any commentary on ‘hashtags’ is not covered by prominent media sources but still, they impact upon the prediction. Secondly, this study is emphasizing NLP techniques and the way how to raw news text can be used for sentiments building processes. So, NLP based models are simply efficient in extracting emotion, feelings, and sentiments out of a raw text. Thirdly, this study not using simple neural networks for predictions process but Long Short Term Memory (LSTM) model based upon the newly developed and highly proven performance in different fields. LSTM models are specifically designed to remember the long-term dependencies. A point that makes it different is mostly, LSTM model is supplied with past stock prices as an input to predict the future price of the stock, however, this study has used sentiments, extracted with help of NLP techniques, to predict the stock price and it is evident from results that model with sentiments has significantly increased the accuracy of the model. This study will be generically beneficial to all institutional and individual investors, all kinds of traders, portfolio managers, and

specifically for short-term and long-term investors who invest in the equity market, future marks, derivative and foreign exchange market.

The rest of the paper is divided into the following sections; Literature review, Methodology section that discusses data collection processes, sentiment index development process, NLP techniques, and implementation of the study model. Then comes results and their interpretations and finally the conclusion of the study.

## 2 Literature Review

By exploring existing literature on the application of neural networks and machine learning in the area of business and finance we found that neural network literature is rooted back to 1988. Research articles have started publishing in 1990 exclusively in the area of finance (Zhang et al., 1998). The neural network are gaining popularity in organizations that are investing money in neural network and data mining solutions for the resolution of their problems (Smith & Gupta, 2000). In the year 1988 researchers were more focused on the application of artificial intelligence in the area of production and operational management. In the area of finance more articles are published in the field of financial firm's bankruptcy prediction. Whereas the focus of stock market prediction was restrained to comparison of traditional time series models with ANNS. Despite the fact that a substantial effort has been made for time series prediction via kernel methods (Chang & Liu, 2008), ensemble methods (Qin et al., 2017), and Gaussian processes (Frigola & Rasmussen, 2013), the drawback is that most of these approaches employ a predefined linear form and may not be able to capture the true underlying non-linear relationship appropriately. Recurrent neural networks (RNNs) (Rumelhart et al., 1986; Werbos, 1990; Elman, 1991), a type of deep neural network specially designed for sequence modeling, have received a great amount of attention due to their flexibility in capturing non-linear relationships. In particular, RNNs have shown their success in NARX time series forecasting in recent years (Diaconescu, 2008; Gao & Er, 2005). Traditional RNNs, however, suffer from the problem of vanishing gradients (Bahdanau et al., 2017) and thus have difficulty capturing long-term dependencies. Recently, long short term memory units (LSTM) (Hochreiter & Schmidhuber, 1997) and the gated recurrent unit (GRU) (Cho et al., 2014) have overcome this limitation and achieved great success in various applications, e.g., The reason suggested by researchers is that the neural network has the capability to outperform the time series models because these models can efficiently predict without the requirement of data being following any distribution and linearity. In addition to the comparison of ANNs and traditional model, in literature evidence exists where models based upon artificial intelligence are compared with each other e.g. Tan et al. (2011) have compared three models; ANN, decision tree and hybrid model with the conclusion that ANN has the highest accuracy in stock price prediction.

In early stages of development for financial forecasting using ANNS, researcher and professionals emphasizes on a comparison of traditional time series models and ANNs to measure the better accuracy in forecasting process for an instant see

Swanson and White (1997), Yoon et al. (1993), Kaastra and Boyd (1996), Lawrence (1997) and Kryzanowski et al. (1993). ANNs gives 72 percent accuracy of predicting stock market returns and also able to accurately predict the positive and negative returns by training and validating the neural networks (Kryzanowski et al., 1993).

As for as methods of artificial neural networks are concerned researches have used different ways to mimic the neural networks of the human brain.

Many efforts have been made to solve the issue of linearity, for example, Kernam method has been used by Chang and Liu (2008), Bouchachia and Bouchachia (2008) and Frigola and Rasmussen (2013), with the help of traditional non-machine learning-based model that are unable to capture underlying non-linear relationships. Stock market data is always stochastic and noisy in nature so, LSTM is more suitable. Normally, statistical and metamathematical models are used for financial prediction and these model are handcrafted and aligned with respect to observation and thus compromise accuracy (Tsantekidis et al., 2017). Fischer and Krauss (2018) suggested that LSTM performs well as compared to Random forecast, Deep neural network, and logistic classifier. Recurrent Neural Network(RNN) model gained popularity due to the flexibility of use and coping up the problem of linearity in time series (Rumelhart et al., 1986; Werbos, 1990; Elman, 1991).

Artificial intelligence based expert system is also catering to the needs of auditing, banking sector, credit risk management but along with it matchless benefits there is the dark side of these expert systems of being costly. Omoteso (2012) have studied the cost and benefit analysis of an intelligent system that can predict the future direction and softwares development in this area. It is concluded that in small and medium organization it may be not suitable to apply such system to achieve the marginal benefit by incurring heavy cost. Oreski et al. (2012) apply neural networks to reduce the data dimensionality by coping redundant data and removing irrelevant factors to enhance the predictive ability of genetic algorithm. Similarly, López Iturriaga and Sanz (2015) designed the artificial neural network-based model that have predicted the financial distress of US bank 3 years before the bankruptcy occurs.

## 2.1 NLP

Recently, natural language processing (NLP) has grown up as powerful techniques for many fields due to its capability to capture sentiments and feeling into the text in more nuanced way. Many applications have started adopting the NLP techniques to give their users better experience (Xing et al., 2018). Though it relatively easy to get the external news with help of many sources but it difficult to access and parse the data through financial statement of company. So, developing information content from companies financial statements is tedious and difficult. Here information means voluntary information disclosed by firm that is not obligatory by law to disclose to stakeholders (Xing et al., 2018). With help of databases this paper includes all sort of internal information whether it reaches to external media or not as well as external news and information.

Textual information extraction and news articles processing rooted back to 1934 (Bühler, 1934; Chomsky, 1956). Previous two decade people have been giving much focus upon bag of word approach to seek sentiments of text with help of stop words and frequencies. Serious drawback of these model is that they are unable to capture the context of sentence. For example, company A is gaining advantage over company B or company B is gaining advantage over A are two completed opposite sentiments but belongs to same bag- of-word. Recent advances like, word to vector representation, word embedding and LSTM have addressed these problems very well. Sentiment analysis is very important phenomena for stock market and financial forecasting (Poria et al., 2016). With increasing use of web.2.0 Standards (Cooke & Buckley, 2008) users have easy access and ways to sharing the information across platform like Facebook, twitter, etc thus market sentiments become importance for financial market. Businesses dealing in financial products and services reshaping their approach to make their application more informed and sophisticated to gain competitive edge over rivals. New NLP techniques are promising them for their required edge. Existing sentiment technique can be broadly categorized into three domains; namely, hybrid, knowledge-based and statistical approach (Poria et al., 2017). Knowledge-based sentiment analysis is based upon list of words and its frequencies- a relatively old approach that categorizes text into different categories and then further compares the frequencies with the lexicon. Second is statistical method, this approach is not only focusing on list of word but also use statistical model to classify the text with help of probabilities. Third category is mixture of these two Lenat (Lenat et al., 1990; Liu & Singh, 2004; Fellbaum, 1998). This study has used hybrid approach with help of modern available NLP techniques that supports programming languages environments as well.

### 3 Methodology

#### 3.1 Data Pre-processing

This section describes a summary of approaches and methods that have been used to process the data from raw text to machine-readable data. Data preprocessing has been divided into four major sections, namely; hourly stock returns, News Analytics preprocessing, Naive Bayes Classifier and sentiment index development. All three sections give the snapshot of preprocessing of data. Let's briefly describe one by one.

##### 3.1.1 Hourly Stock Data

Hourly stock returns are calculated with help of opening and closing price of all 10 companies. Hourly stock data is obtained from Thomson Reuters data portal. Simple formula for calculating the stock return is as follows:



$$R_{ij} = \frac{closing_{ij} - opening_{ij}}{opening_{ij}} \quad (1)$$

whereas  $R_{ij}$  is  $j_{th}$  stock at  $i_{th}$  hours,  $closing_{ij}$  is closing price of  $j_{th}$  stock at  $i_{th}$  hours,  $opening_{ij}$  is opening price of  $j_{th}$  stock at  $i_{th}$  hours.

### 3.1.2 News Analytic Processing

There are many sources through which information flows into the stock exchanges related to a specific stock. News and information sources that have been used in this paper are: mainstream media, print media, social media news feeds, blogs, investors' advisory portals, experts opinions, brokers updates, web-based information, company' internal news and public announcements regarding policies and reforms. We have collected the news stories from a very well known and reliable database, named; Thomson Returns. Using Thomson Reuters's API we were able to collect new stories if these stories would be related to any of ten stocks which, we have chosen for analysis. The reason for choosing individual stock instead of the stock exchange is; stock exchanges absorb and react to collective level information and thus, specifically event-level information is hard to be separated. Every news story has its timestamp according to GMT and precise at the millisecond level. The time frame for news collection is 10 years, so, collecting every news resulted to have very large text corpus. The timestamp for news is strictly matched with the stock exchange's opening and closing time. Although, we have thrown a lot of use full collected news information that lies outside of the stock exchange opening and closing time window. However, it was necessary to gauge the impact of news analytics on the stock price movement.

### 3.1.3 Naive Bayes Classifier

After the raw text regarding news is extracted from sources, the text is refined in the way that it can be used in the Navie Bayes Classification model. Originally text was in 'HTML' form with a lot of unnecessary information, but with help of parser and some lines of coding, 'HTML' based- text is refined and filtered into 'lxml' form. 'XML' form of text is accurately and quickly readable by machines. Naive Bayes Classification model has been used to calculate the sentiments out of news text. The Fig. 1 shows how information filters though raw sources to sentiment score. The left column of the diagram shows that Raw text, which includes 'HTML' meta-information in it. The first step is to split the complete sentences into a list of unique words, the process is called tokenizing. Next comes, creating a filter of stop words, these stop words are mostly related to pronouns. At the next stage, the text is filtered from hyperlinks and unnecessary information. In the next step, lemmatization is applied to address spelling mistakes. The list of all words is labeled with a part of speech. Then, data is a little bit more refined to see any redundancies. As a next step, with the help of the already available NLTK database, each word has been assigned with negative or positive labels. In the next two steps data is prepared for test and train

dataset - ready to feed to the 'Naive Bayes' Model for training. After the training process is completed each sentence is tested to get sentiments scores out of it. The outcome of the NLP model is utilized in building the sentiment index and LSTM data at the later stages.

### 3.1.4 Sentiment Index

Following variables are taken into account while building the sentiments index: 'sentiment time window', 'score value', 'class of sentiment score', 'relevance' of score towards the underpinning stock; time window means how many times news/information, related to the selected stock appeared during 1 h time period. The logic behind keeping the time window to 1 h is that stock exchanges need a bit of time to absorb the information related to an individual stock. secondly, Minute level analysis is too early and day level analysis is too late. Next factor is 'score value'. Score value is the outcome of a trained NLP model, the process is given in the Fig. 1. Sentiment score values are classified into three categories based on their scores; positive, negative, and neutral. All the negative score are carrying the negative signs and neutral sentiment are equal to zero. The scores for all three type of classes are ranges from 0 to 1. Thus 'score value' are summed up during the 1 h time window, if the sum of the score is negative and greater than  $-0.10$ , it is labeled as negative score, if the sum is between  $-0.10$  and  $0.10$  it is considered as neutral score and, from  $0.10$  to  $0.90$ , the 'score value' is positive. In the Next step, the sentiment score outcome is finally multiplied by variable 'relevance' to weight the sentiment with respect to its relevance score. 'Relevance score' is percentage number, calculated; the number of times news story mentioned the name of a stock divided by the total count of words in the news story. The mathematical expression of the sentiment index is as under:

$$S_{ij} = \sum_{i \in I} \left( e_i \times R_i \times C_i \right) \quad (2)$$

whereas  $I$  = time windows for every  $i_{th}$  and  $j_{th}$  stock.  $e = \max(pos_i, neg_i, neut_i) \exists$ ,

$$C_i = \begin{cases} +1 & \text{argmax}(pos_i, neg_i, neut_i) = 1 \\ 0 & \text{argmax}(pos_i, neg_i, neut_i) = 3. R_i = \text{Relevance.} \\ -1 & \text{argmax}(pos_i, neg_i, neut_i) = 2 \end{cases}$$

$$P(w_j|x_i) = \frac{P(x_i|w_j).P(w_j)}{P(x_i)} \quad (3)$$

whereas  $w_i$  is a particular class (e.g. Negative or positive) and  $x_i$  is an given features,  $P(x_i|w_j)$  is called the posterior or in other word probability of feature  $x_i$  belongs to class  $w_j$ ,  $P(w_j)$  probability of class itself with respect to total sample also called the prior and finally,  $P(x_i)$  is called the marginal probability or evidence. based upon above-stated Bayes theorem, conditional class probabilities of the equation can be calculated as follows:

$$P(\mathbf{x}_i|\omega_j) = P(x_1|\omega_j) \cdot P(x_2|\omega_j) \cdot \dots \cdot P(x_d|\omega_j)$$

Posterior probabilities can be calculated with following expression:

$$P(\mathbf{x}_i|\omega_j) = \prod_{k=1}^d P(\mathbf{x}_k|\omega_j)$$

So, probability of class can be calculated with this expression:

$$P(\omega_j) = \frac{N_{\omega_j}}{N_c}$$

### 3.2 Model Equation

Artificial intelligence-based models have proved their importance and efficiency in almost all spheres of life and the field of economics and finance can not be excluded. Our model can be used practically in a variety of ways. For example, online trading expert systems are forced to integrate advanced ways for the prediction process. The current model could be specifically very relevant for the trading system to reshape the prediction process and reduces the effort of organizing and search the relevant market info through millions of text records with either human-based effort or the traditional text filtering approaches. The model already uses sophisticated NLP techniques to include the sentimental-based market information into the model. For example, building the information-related index is very crucial. Keep this point in view we have built a customized sentiment index that collects the market information at one minute level and sums it up for a 1-h window. On one hand, it enables LSMT model to capture high-level precision and on other hand, its overcome the limitation to rely upon daily-based market information. There are many traditional models which try to achieve precise forecasting using economic data i.e. simple regression, Moving Averages, and autoregressive-based models (see. ARMA, ARIMA, ARCH GARCH), simple regression, and a bunch of other time series forecasting models. The universal problem for all these models is the limitation to handle the assumption of linear distribution, handling long past lags, and very strict criteria of data structure. These limitations come with a lot of compromises in terms of efficiency and accuracy. Artificial neural network-based model and most specifically, LSTM is very good at handling long-term dependencies i.e. you can keep tracing the past data without losing the information it carries. Moreover, With help of different activation functions and specific approaches model works flexibly without setting many assumptions Let's elaborate how this model works.

The current model is based upon original scientific publications made by Hochreiter and Schmidhuber (1997). The research is regarded highly by the research community because of its ability to work on long-term dependencies and the ability to remember important information in previous steps. The cases where the dependency of information does not matter much, simple neural network models work fine, but this is not an ideal situation in the practical business world. Stock market prediction,

natural language processing, sentimental analysis, and language translation are the example where information of model is highly dependent and context is very important thus recurrent neural network model are good alternatives of simple neural networks. Here is a short description of how the model of this study is fitted.

Hidden state function can be written in the following way:

$$h_t = f(h_{t-n}, X_t) \quad (4)$$

So, the hidden state of LSTM model has been written with the help of the following equation.

$$h_t = \tanh(W_h h_{t-n} + W_x X_t) \quad (5)$$

Weight matrix is first multiplied with current input. Previous time steps hidden states are one by one multiplied with weight matrix for hidden state. Finally, tanh has been applied on result after adding both, current input and previous time steps hidden states. Now output layer of LSTM model is as under:

$$O_t = \sigma(W_o \cdot [h_{t-1}, X]_t + b_o) \quad (6)$$

whereas  $W$  is weight matrix for output layer and  $h_t$  we have calculated in Eq. 5.

Equations 5 and 6 simply shows how hidden and output layers of the LSTM model are formulated but this formulation is not much different from simple neural network models. The true secret of LSTM model lies in its unique way of developing cell and memory state with help of gating mechanism.

### 3.2.1 Signalling and Gates

Gates are basically fully connected feed-forward networks that receive information, applies functions, usually sigmoid activation functions, and do point-wise operations and then return outputs. Thus, we have applied here sigmoid activation function that spits outputs between the range of 0 and 1. So, all the outputs values closer to 0 are considered unimportant and cell deletes them, on the other hand, all information that is close to 1 is important for the prediction process and therefore updated in cell state. In this section, we will describe how signals and gates for LSTM work. Not all information in cell state is important to know for the prediction process and overflow of unnecessary information means disinformation. Primarily there are three gates of LSTM, namely: forget gate, input gates, and the output gate.

**Forget Gates** Forget gate receives information from current input and earlier hidden layer input, it applies the sigmoid function on this number and multiplies it with previous cell state. This decides that whether we want information in previous cell state with respect to new information and  $t - 1$  information in state  $C_{t-1}$ . The mathematical equation of forget gate is as under:

$$f_t = \sigma(W_f \cdot [h_{t-1}, X]_t + b_f) \quad (7)$$

**Input Gate** This is the second part of the signalling process. In the first part, we have decided that the previous cell state is importation to keep or not. Now it is time to store new essential information on cell state, that will be later judged again by forgetting gate with respect to its importance for the model learning process. Input gate is a multiplication of  $t - 1$  hidden state and  $t$  input by input weight matrix, that will be later merged into the new candidate. The activation function of the input gate is sigmoid. Mathematical equation of  $i_t$  is as under:

$$i_t = \sigma(W_i \cdot [h_{t-1}, X]_t + b_i) \quad (8)$$

**New Candidate** Similar to input gate new candidate is multiplication of hidden state's current input with weighted matrix of new candidate denoted with symbol  $\tilde{C}_t$  with combination of  $i_t$  new candidate will decide with how much information model wants to write on new cell state. Mathematical equation of  $\tilde{C}_t$  is as under:

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, X]_t + b_c) \quad (9)$$

Now cell state is updated with help of input gate and new candidate the equation is as follows:

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (10)$$

Output layer is the multiplication of the weight matrix of the output layer by previously hidden state and current input.

$$O_t = \sigma(W_o \cdot [h_{t-1}, X]_t + b_o) \quad (11)$$

Finally output  $h_t$  is product of output layer and hidden state and mathematical expression of  $h_t$  is as under:

$$h_t = o_t \times \tanh(C_t) \quad (12)$$

### 3.2.2 Model Optimization

As a model optimization function Stochastic Gradient Descent (SGD) has been used in this study. As our model is not supposed to be linear so slop of non-linear error between two point can be calculated with help of derivative as under:

$$\frac{f(x) = \Delta f(x)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x} \quad (13)$$

Cost of the model is always an outcome of the specific function. In our model cost is the difference between the actual price of the entity—predicted price of the entity and based on Mean Square Errors. There are two major parameters that need to be tuned to reach the global minimum level of error.

$$\frac{\delta f}{\delta \beta} \quad (14)$$

As there in our function of cost two parameters are involved namely,  $\alpha$  and  $\beta$ . Because there are two parameters we need partial derivation  $\delta$ .

$$\frac{\delta f}{\delta \alpha} \quad (15)$$

In the direction of the slop we can calculate all possible partial derivatives and map them on a vector and can be called gradient vector. Mathematical expression is as under:

$$f : R^n \rightarrow R : \nabla f = \begin{bmatrix} \frac{\delta f}{\delta \theta_1} \\ \frac{\delta f}{\delta \theta_2} \\ \vdots \\ \frac{\delta f}{\delta \theta_n} \end{bmatrix} \quad (16)$$

$\theta$  is the point toward slop to achieve the global minima and  $\delta f$  changes in function due to change in slop. So, in this way, we can make a vector of all possible partial derivatives to go down to hill.

So, gradient descent update rule is as under:

$$\theta_{\text{new}} = \theta_{\text{old}} - \eta \nabla_{\theta} f \quad (17)$$

whereas  $\theta_{\text{new}}$  is updated parameter  $\theta_{\text{old}}$  is old parameter '-' sign means we want to go downhill  $\eta$  is step size that model should take on slop line to go down hill,  $\nabla_{\theta}$  is gradient with respect to parameters.

### 3.2.3 RMS Prop

To really speed up the model learning and error reduction, RSMprop algorithm has been used in the model. The idea behind this algo is to divide the gradient decent into two parts, a gradient that moves in vertical and gradients that moves in a horizontal direction. Vertical movement is called oscillation that is not much beneficial of error reduction. Thus, this algorithm focus on horizontal movement to achieve the global minima.

$$s_{dw} = \beta s_{dw} + (1 - \beta)(s_{dw})^2 \quad (18)$$

$$W = W - \alpha \frac{s_{dw}}{\sqrt{s_{dw}} + \epsilon}$$

$$s_{db} = \beta s_{db} + (1 - \beta)(s_{db})^2$$

$$b = b - \alpha \frac{s_{db}}{\sqrt{s_{db}} + \epsilon}$$

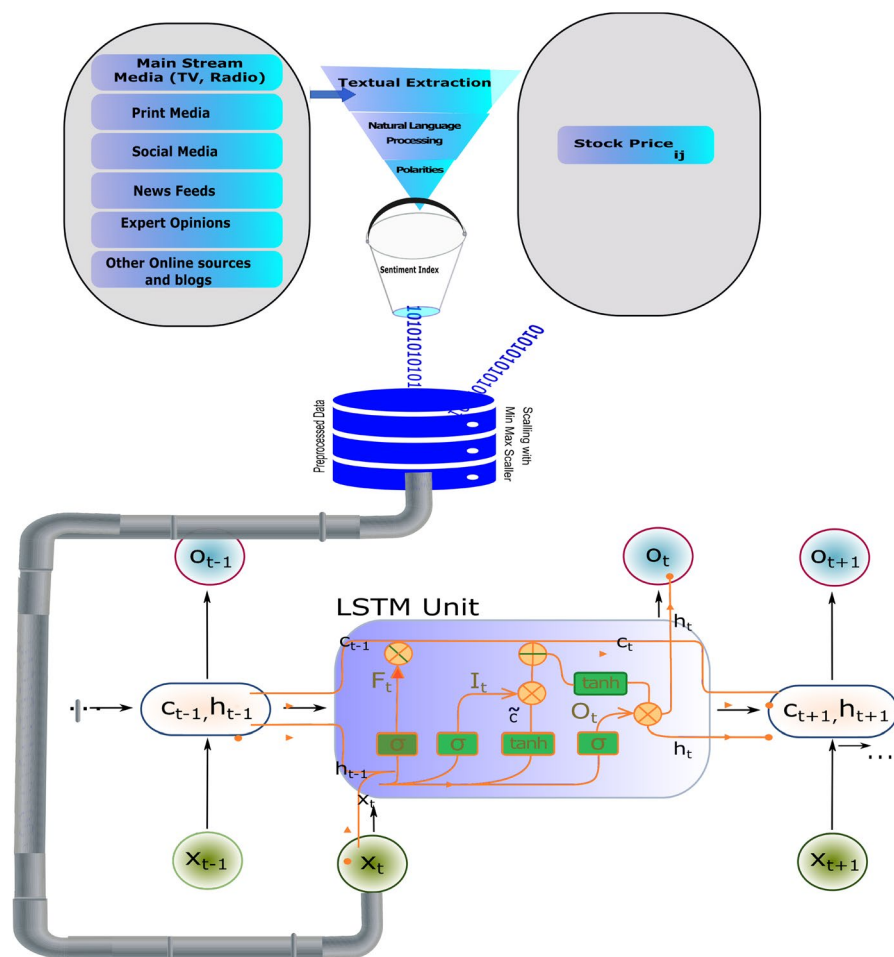


Fig. 2 Study model

whereas  $s_{dw}$  is gradient in horizontal direction and  $s_{db}$  is gradient in vertical direction.  $\alpha$  is learning rate and  $\beta$  is simply parameter for moving average that separate for  $s_{dw}$  and  $s_{db}$ . Whereas,  $(s_{dw})^2$  square of past gradient.  $\epsilon$  is very small value to avoid dividing by zero. Moving average is effective in this also because it gives higher weight to current value of gradient and less weight to square of past gradient.

Overall schematic of the study model is as follows.

Now, we will start next section where we have described our results of model (Fig. 2).

**Table 1** Training loss for all models

Name/error	With sentiments			Without sentiments		
	MSE	MAE	RMS	MSE	MAE	RMS
Apple shift1	2389.420	48.087	48.882	2590.420	49.087	50.899
Apple shift3	0.546	0.602	0.739	0.606	0.662	0.778
Apple shift6	0.449	0.485	0.670	0.700	0.660	0.880
Bank of America shift1	3.066	1.230	1.751	2048.201	45.225	45.257
Bank of America shift3	0.004	0.043	0.060	8344.511	90.952	91.348
Bank of America shift6	0.010	0.071	0.101	8391.930	91.208	91.607
Cisco shift1	57.922	7.370	7.611	121.695	10.872	11.032
Cisco shift3	6.259	2.416	2.502	112.721	10.374	10.617
Cisco shift6	0.027	0.122	0.163	112.424	10.360	10.603
Ford shift1	6.314	2.385	2.513	45.308	6.676	6.731
Ford shift3	0.002	0.029	0.041	137.890	11.625	11.743
Ford shift6	0.005	0.048	0.070	200.572	13.990	14.162
General Electric shift1	54.135	7.189	7.358	197.547	13.958	14.055
General Electric shift3	0.356	0.573	0.597	140.152	11.659	11.839
General Electric shift6	0.010	0.068	0.100	140.001	11.653	11.832
Intel shift1	127.881	11.022	11.308	202.233	13.988	14.221
Intel shift3	47.959	6.733	6.925	57.391	7.208	7.576
Intel shift6	0.037	0.141	0.193	53.460	6.927	7.312
JP Mogen shift1	511.131	21.853	22.608	1637.782	40.050	40.470
JP Morgen shift3	421.752	19.834	20.537	1359.767	36.464	36.875
JP Morgen shift6	79.236	8.529	8.901	874.040	29.119	29.564
Microsoft shift1	508.827	21.728	22.557	144.701	10.353	12.029
Microsoft shift3	387.778	18.945	19.692	99.047	8.231	9.952
Microsoft shift6	28.659	5.090	5.353	29.534	4.361	5.435
Pfizer shift1	138.511	11.664	11.769	17.021	3.797	4.126
Pfizer shift3	32.070	5.604	5.663	3.812	1.680	1.952
Pfizer shift6	0.032	0.143	0.179	170.684	12.307	13.065
Well Frago shift1	361.245	18.773	19.006	874.213	29.413	29.567
Well Frago shift3	214.780	14.464	14.655	568.722	23.650	23.848
Well Frago shift6	0.852	0.867	0.923	326.240	17.755	18.062
Sum of errors	5379.275	236.108	243.427	29,003.325	624.264	637.437

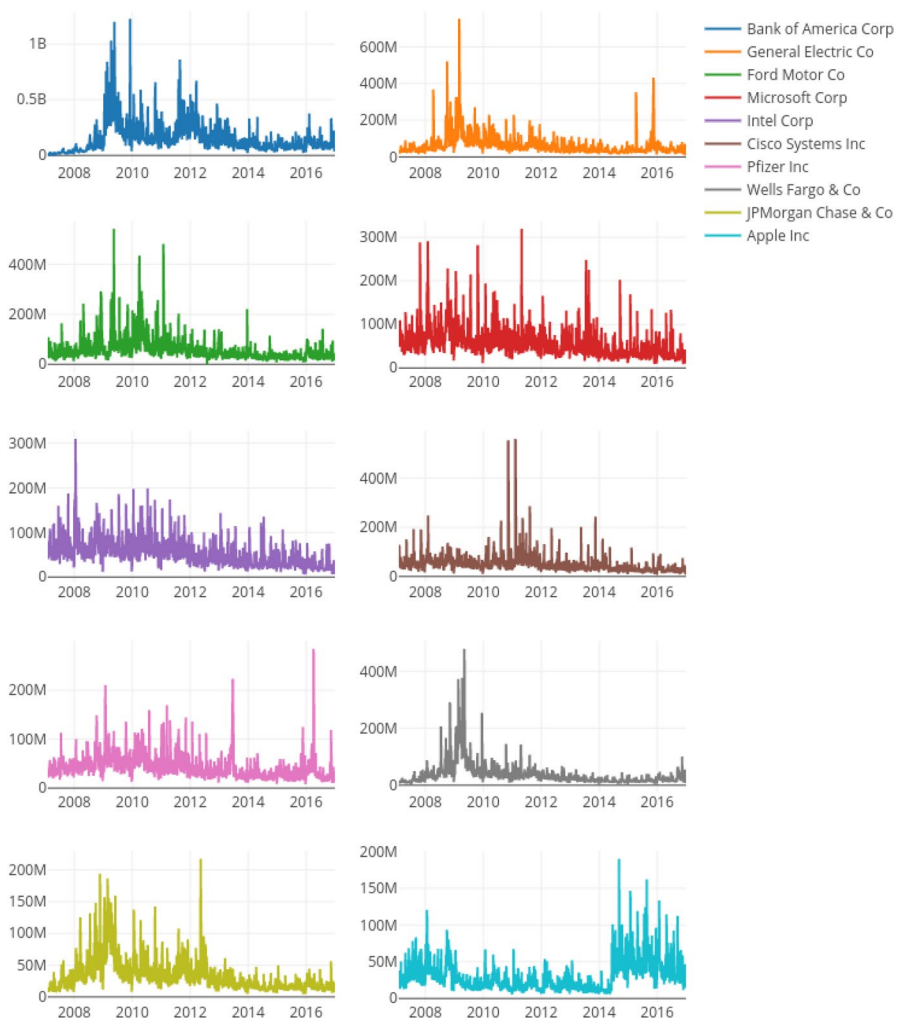
## 4 Results and Interpretations

This section shows the result of model and gives short detail and analysis of the results. Volume wise top ten companies from four major sectors has been selected for analysis purpose. Prediction accuracy results are given in the Table 1. Some figures and tables are omitted from the result section on account of brevity.

The Fig. 3 shows the top ten companies with the highest trade volume during the period of 2008–2016. These top ten companies are the sample that is under the



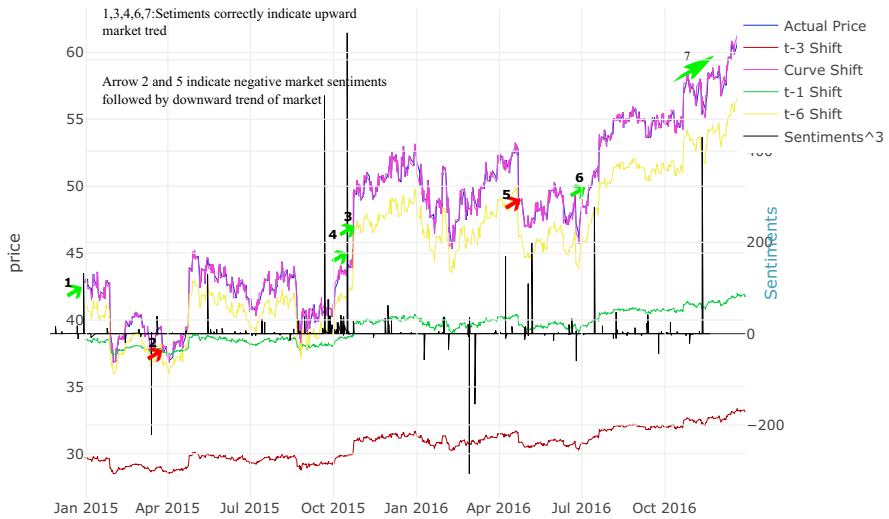
## Top Ten Companies by Trade Volume



**Fig. 3** Top ten companies from different sectors

study in this paper. These ten companies roughly are big names in the financial, IT, medical, electronics and auto-mobile sectors. The reason for selecting diversified companies is to show the reflection of big sectors onto study model. Due to data collection issue, the latest year of study is 2016 but year of the study does not matter in study because purpose of the study to investigate prediction accuracies with machine learning based models and importance of textual analytic (Figs. 4, 5, 6, 7, 8, 9, 10).

There are some interesting results to show regarding the above-shown figures. Six types of different models have been applied to each of the ten companies. The first

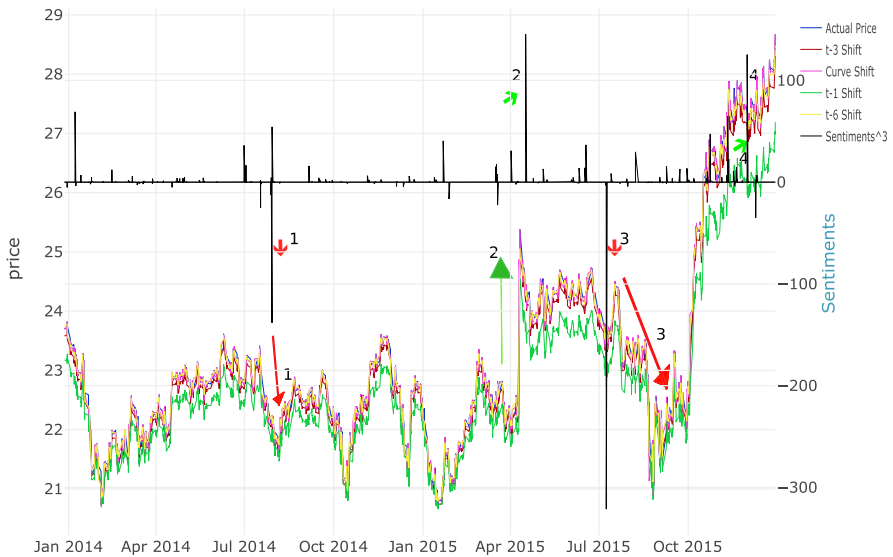


**Fig. 4** Microsoft price predictions with sentiments

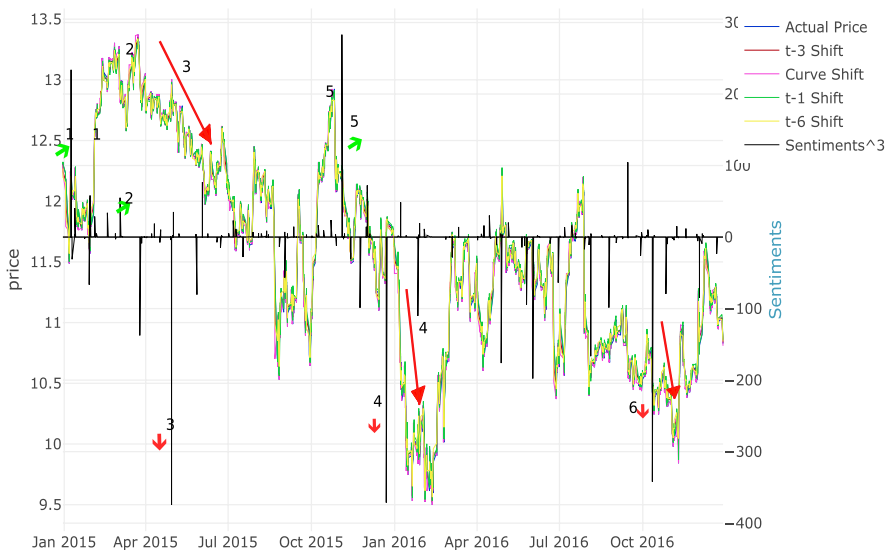


**Fig. 5** Intel price predictions with sentiments

three models are related to the stock market prediction with embedding the company related sentiments and the other three are related to the forecasting without sentiments. There are six types of models for each group i.e with sentiment and without sentiments. The legends in the figures for both types of groups i.e with sentiments and without sentiments is the same. Let's describe the six types of models. The actual price is a simple plot of Actual price for a given period. 't - 3' is a model

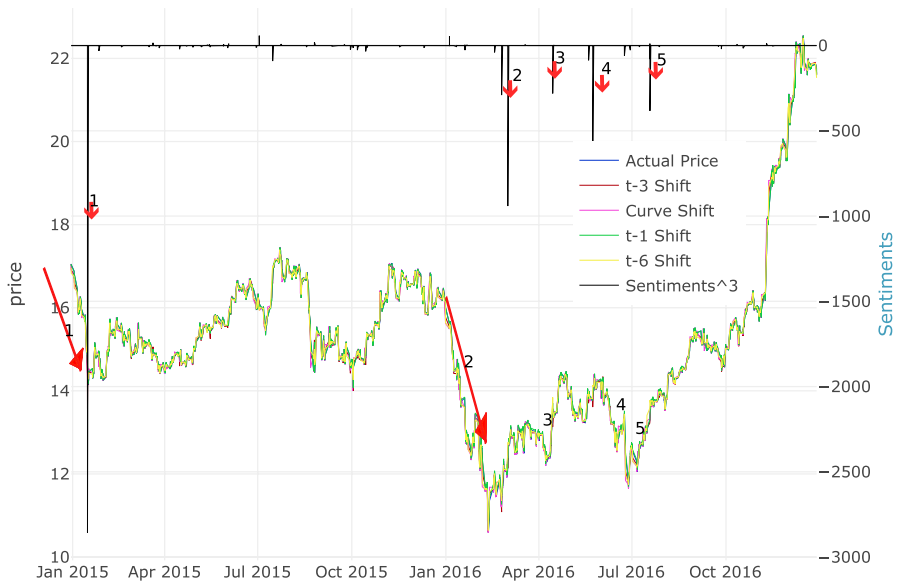


**Fig. 6** GE price predictions with sentiments

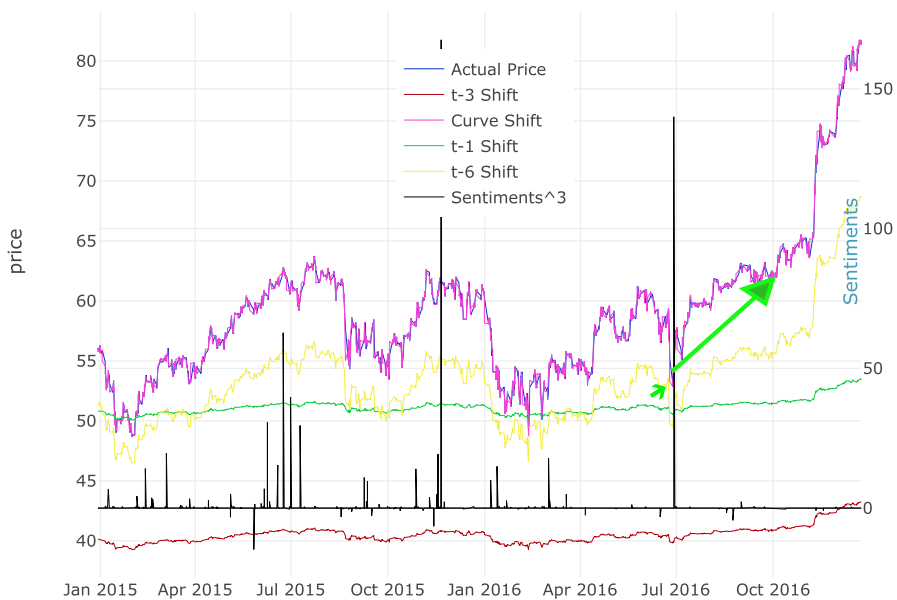


**Fig. 7** Ford stock predictions with sentiments

that is based upon 3 h time windows. That essentially means that the model knows the actual price of 1 h in the future as a label for  $t - 3$  time price. ' $t - 3$ '. The curve shift is a simple 1-h curve shift without sliding windows. For the sliding window, we mean that model gets a price and certain  $t$  and gets the output as a window of 1, 3, and 6 h future prices as the label. So in the curve shift is just the next day as and as

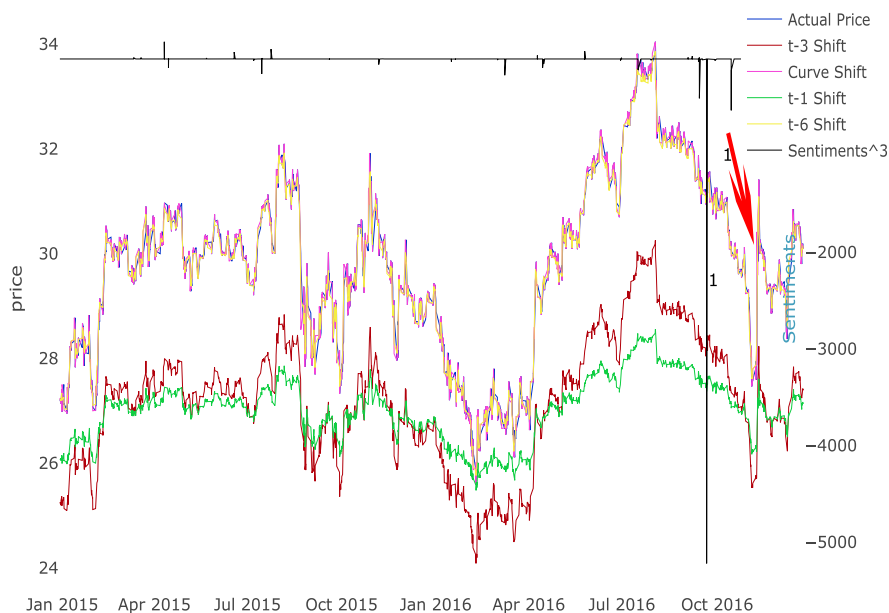


**Fig. 8** Bank of America price prediction without sentiments



**Fig. 9** JP Morgan price prediction with sentiments

the label for current  $t$  price. Similarly,  $t - 1$  and  $t - 6$  are sliding windows of 1 and six-time steps respectively. sentiment curve legend only appears in the graphs where the sentiment index has been used as an input in all input models.



**Fig. 10** pfe price prediction with sentiments

The model ( $t - 1$ ) is looking back to 1 h past data along with sentiments and try to predict the price of the next hour and so on. The second model is looking 3 h back and the third model is looking 6 h back. The purpose behind selecting these different models is to get the idea to what extent the model needs past information to be able to render better results. It can be observed from figures that model is consistently giving very good results when we have given it 6 h of information of the company, as compared to three and one time steps respectively. In most of the cases, the one-time step is relatively the least accurate model and the reason is obvious that the model is getting less information. The sentiment line is plotted on the secondary axis of the figures, as scales of both axes are very different, so to avoid convolution and getting a better overview, we have used the secondary axis. All sentiment scores are exponential with  $x_i^3$ . Referred to method section for detailed formulation and algorithm for sentiment scores building. Selected companies are quite large and famous around the world thus the frequency of the company's information is high. So there are many cases of small sentiments that don't influence the market much. So, in the exponentiation process most strong sentiments get prominent and gives better visualization for analysis purpose. The sentiment line is giving a very insightful and meaningful indication for the next market direction. Sentiments on average are getting 2–3 h advance the company-specific information and that information reflects stock direction very effectively.

The results of all six models of study are given in the Table 1. For comparison purposes, three different criteria of model accuracies are given. This is a very comprehensive table that shows a complete training process and achieved the prediction

accuracy of the study model. To get a generic overview, the sum for two panels, namely, with sentiments and without sentiment is given at the end of the table. It is obvious from the results that the error sum using all criteria is greater in a case where models don't use company-specific textual analytic.

## 5 Conclusion

In the recent past, the basic way of operating businesses and corporations, penetrating the new market and reaching to the customers and providing the financial services is exponentially influenced by the new wave of data sciences and artificial intelligence. The research study is motivated by the same phenomenon and empirically investigates the forecasting the stock prices with out-of-box cutting edge soft-computing techniques. The forecasting process is inherited with three unique parts: text analytic, hourly Sentiment index building process, and LSTM AI-based model. First, company-specific text information has been collected, aggregated, classified, and cleansed from thousands of different Thomson Reuters's based information channels that include, mainstream media, print media, social media, blogs, investors advisory services, discussion forums, brokers commentaries. Useful information was lurking in a pile of unwanted information, using natural language processing techniques information is cleaned and useful features have been extracted to be fed to Naive Bayes Classifier to get its sentiments. The second part is building an hourly sentiment index. Though much information from raw text has been collected at the end only three important features related to the sentiment have been preserved for the index building process, namely class of sentiment, the direction of sentiment, and relevance of the sentiments to a specific company. Sentiment information can come anytime round the clock but stock exchange only works for a specific time range. So first, the time of sentiment is matched with the operational time of stock exchange, then based on the self-discovered equation sentiment index has been built. Most of the research studies in similar direction use daily bases stock market values along with other variables but uniquely this study uses an hourly based model for the forecasting process. The reason for the hourly-based model is that getting the accurate influence of the information because 1 day is too late and minute-interval is too early, thus, the direction of stock may not be aligned with sentiments. Third, and final part of the study is the usage of the LSTM neural network model that works in a very special way when it comes to time series or long term dependency of the information.

The results of the study show that sentiments are playing a very important role in the prediction process. Exponentiated sentiments are concisely followed by the big companies traded at US major stock exchanges. That makes our new way of measuring the sentiments robust. Top ten companies from High-tech, financial, medical, automobile sectors are selected, and six LSTM models are applied, three for using text-analytic and other three without analytic being used. Every model includes 1, 3, and 6 h steps back. For all sectors, a 6-h steps based model outperforms the others

due to LSTM specialty of keeping long term memory, collective accuracy of having textual analytic models is way higher relative to non-textual analytic models.

**Limitation of study** Limitation of the study includes a waste of a lot of useful information due to matching the time of news and information strictly with an opening and closing time of stock exchange. However, top companies are operating worldwide and universal time varies across the globe. Almost every hour round the clock information regarding these companies is coming in, but during time matching processing very useful information of almost 18 h has been thrown out. The next challenge in this connection is to come up with a sophisticated mechanism to cope with the issue.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

#### Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Adebiyi, A. A. (2012). A model for stock price prediction using the soft computing approach.
- AlFalahi, K., Atif, Y., & Abraham, A. (2014). Trading and fuzzy logic. *Int J Intell Syst*, 29(2), 1–23.
- Arora, N., et al. (2019). Financial analysis: Stock market prediction using deep learning algorithms. In *Proceedings of International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM)*, Amity University Rajasthan, Jaipur-India.
- Atsalakis, G. S., & Valavanis, K. P. (2009). Surveying stock market forecasting techniques—Part II: Soft computing methods. *Expert Systems with Applications*, 36(3 Part 2), 5932–5941. <https://doi.org/10.1016/j.eswa.2008.07.006>.
- Bahdanau, D., Cho, K., & Bengio, Y. (2017). Learning to compute word embeddings on the fly. *Iclr*. <https://doi.org/10.2507/26th.daaam.proceedings.070>. arXiv: 1409.0473v7
- Bouchachia, A., & Bouchachia, S. (2008). Ensemble learning for time series prediction. In *First Int. Work. Nonlinear Dyn. Synchronization*.
- Bühler, K. (1934). *Sprachtheorie: Die Darstellungsfunktion der Sprache* [Linguistics theory: Representation function of language]. Jena Fischer.
- Chang, P. C., & Liu, C. H. (2008). A TSK type fuzzy rule based system for stock price prediction. *Expert Systems with Applications*, 34(1), 135–144.
- Cheng, C. H., Chan, C. P., & Yang, J. H. (2018). A seasonal time-series model based on gene expression programming for predicting financial distress. *Computational Intelligence and Neuroscience*, 2018, 1067350.
- Cho, K. R., Huang, C. H., & Padmanabhan, P. (2014). Foreign ownership mode, executive compensation structure, and corporate governance: Has the literature missed an important link? Evidence from

- Taiwanese firms. *International Business Review*, 23(2), 371–380. <https://doi.org/10.1016/j.ibusrev.2013.06.005>.
- Chomsky, N. (1956). Three models for the description of language. *IEEE Transactions on Information Theory*, 2(3), 113–124. <https://doi.org/10.1109/TIT.1956.1056813>.
- Cooke, M., & Buckley, N. (2008). Web 2.0, social networks and the future of market research. *International Journal of Market Research*, 50(2), 267–292. <https://doi.org/10.1177/147078530805000208>.
- Correa, M., Bielza, C., & Pamies-Teixeira, J. (2009). Comparison of Bayesian networks and artificial neural networks for quality detection in a machining process. *Expert Systems with Applications*, 36(3), 7270–7279.
- Diaconescu, E. (2008). The use of NARX neural networks to predict chaotic time series. *Computer (Long Beach Calif)*.
- Elman, J. L. (1991). Distributed representations. Simple recurrent networks and grammatical structure. *Machine Learning*. <https://doi.org/10.1023/A:1022699029236>. arXiv:1206.2944
- Ergen, T., Kozat, S. S., & Member, S. (2017). Based on LSTM neural networks. *IEEE Transactions on Neural Networks and Learning Systems Efficiency*, 1–12.
- Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics*, 49(3), 283–306.
- Fellbaum, C. (1998). A semantic network of English: The mother of all WordNets. In *EuroWordNet A Multiling. database with Lex. Semant. networks*, Dordrecht, pp. 137–148. [https://doi.org/10.1007/978-94-017-1491-4\\_6](https://doi.org/10.1007/978-94-017-1491-4_6)
- Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669. <https://doi.org/10.1016/j.ejor.2017.11.054>.
- Frigola, R., & Rasmussen, C. E. (2013). Integrated pre-processing for Bayesian nonlinear system identification with Gaussian processes. In *Proceedings of IEEE Conference on Decision and Control*. <https://doi.org/10.1109/CDC.2013.6760734>. arXiv:1303.2912
- Gao, Y., & Er, M. J. (2005). NARMAX time series model prediction: Feedforward and recurrent fuzzy neural network approaches. *Fuzzy Sets Systems*. <https://doi.org/10.1016/j.fss.2004.09.015>.
- Hansson, M. (2017). *On stock return prediction with LSTM networks*. <https://lup.lub.lu.se/student-papers/search/publication/8911069>
- Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018). NSE stock market prediction using deep-learning models. *Procedia Computer Science*, 132(Iccids), 1351–1362. <https://doi.org/10.1016/j.procs.2018.05.050>.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*. <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Huang, W., Lai, K. K., Nakamori, Y., Wang, S., & Yu, L. (2007). Neural networks in finance and economics forecasting. *International Journal of Information Technology & Decision Making*, 6(1), 113–140. <https://doi.org/10.1142/S021962200700237X>.
- Kaastra, I., & Boyd, M. (1996). Designing a neural network for forecasting financial and economic time series. *Neurocomputing*, 10(3), 215–236. [https://doi.org/10.1016/0925-2312\(95\)00039-9](https://doi.org/10.1016/0925-2312(95)00039-9).
- Kahneman, D. (2003). Maps of bounded rationality: Economist psychology for behavioral. *American Economic Review*, 93(5), 1449–1475. <https://doi.org/10.1257/000282803322655392>.
- Kahneman, D., & Tversky, A. (1979). Prospect theory—An analysis of decision under risk.pdf. *Econometrica*, 47, 263–292. <https://doi.org/10.2307/1914185>.
- Kryzanowski, L., Galler, M., & Wright, D. W. (1993). Using artificial neural networks to pick stocks. *Financial Analysts Journal*, 49, 21–27. <https://doi.org/10.2469/faj.v49.n4.21>.
- Kuo, P. H., & Huang, C. J. (2018). A green energy application in energy management systems by an artificial intelligence-based solar radiation forecasting model. *Energies*, 11(4), 819.
- Lawrence, R. (1997). Using neural networks to forecast stock market prices. *Methods*, pp. 1–21. <http://people.ok.ubc.ca/rlawrenc/research/Papers/nn.pdf>
- Lenat, D., Guha, R., & Pittman, K. (1990). Cyc: Toward programs with common sense. *dlacmorg*. <https://dl.acm.org/citation.cfm?id=79176>
- Li, Y., & Ma, W. (2010). Applications of artificial neural networks in financial economics: A survey. In *2010 International symposium on computational intelligence and design*, pp. 211–214. <https://doi.org/10.1109/ISCID.2010.70>
- Li, Y., Jiang, W., Yang, L., & Wu, T. (2018). On neural networks and learning systems for business computing. *Neurocomputing*, 275, 1150–1159. <https://doi.org/10.1016/J.NEUCOM.2017.09.054>.



- Liu, H., & Singh, P. (2004). ConceptNet—A practical commonsense reasoning tool-kit. *BT technology Journal*, 22(4), 211–226. <https://doi.org/10.1023/B:BTTJ.0000047600.45421.6d>.
- López Iturriaga, F. J., & Sanz, I. P. (2015). Bankruptcy visualization and prediction using neural networks: A study of U.S. commercial banks. *Expert Systems with Applications*, 42(6), 2857–2869. <https://doi.org/10.1016/j.eswa.2014.11.025>.
- Makridakis, S., et al. (2018). Forecasting the impact of artificial intelligence, part 3 of 4: The potential effects of AI on businesses, manufacturing, and commerce. *Foresight: The International Journal of Applied Forecasting*, 49, 18–27.
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2018). Statistical and Machine Learning forecasting methods: Concerns and ways forward. *PLoS ONE*, 13(3), e0194889.
- Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: An introduction. *Journal of the American Medical Informatics Association*, 18(5), 544–551.
- Nelson, D. M., Pereira, A. C., & de Oliveira, R. A. (2017). Stock market's price movement prediction with LSTM neural networks. In *2017 International joint conference on neural networks (IJCNN)* (pp. 1419–1426). IEEE.
- Omoteso, K. (2012). The application of artificial intelligence in auditing: Looking back to the future. *Expert Systems with Applications*, 39(9), 8490–8495. <https://doi.org/10.1016/j.eswa.2012.01.098>.
- Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications*, 39(16), 12605–12617. <https://doi.org/10.1016/j.eswa.2012.05.023>.
- Pawar, K., Jalem, R. S., & Tiwari, V. (2019). Stock market price prediction using LSTM RNN. In *Emerging trends in expert applications and security* (pp. 493–503). Springer.
- Poria, S., Cambria, E., & Gelbukh, A. (2016). Aspect extraction for opinion mining with a deep convolutional neural network. *Knowledge-Based Systems*. <https://doi.org/10.1016/j.knosys.2016.06.009>.
- Poria, S., Chaturvedi, I., Cambria, E., & Hussain, A. (2017). Convolutional MKL based multimodal emotion recognition and sentiment analysis. In *Proceedings of IEEE international conference on data mining, ICDM*. <https://doi.org/10.1109/ICDM.2016.178>
- Qin, Y., Song, D., Cheng, H., Cheng, W., Jiang, G., & Cottrell, G. W. (2017). A dual-stage attention-based recurrent neural network for time series prediction, pp 2627–2633. <https://doi.org/10.24963/ijcai.2017/366>. [arXiv:1704.02971](https://arxiv.org/abs/1704.02971)
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*. <https://doi.org/10.1038/323533a0>. [arXiv: 1011.1669v3](https://arxiv.org/abs/1011.1669v3)
- Shefrin, H. (2008). A behavioral approach to asset pricing. *America (NY)*, 71, 046123.
- Smith, K. A., & Gupta, J. N. D. (2000). Neural networks in business: Techniques and applications for the operations researcher. *Computers & Operations Research*, 27(11–12), 1023–1044.
- Swanson, N. R., & White, H. (1997). A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 79(4), 540–550. <https://doi.org/10.1162/003465397557123>.
- Tan, S., Wang, Y., & Wu, G. (2011). Adapting centroid classifier for document categorization. *Expert Systems with Applications*. <https://doi.org/10.1016/j.eswa.2011.02.114>.
- Trippi, R., & DeSieno, D. (1992). Trading equity index futures with a neural network. *Journal of Portfolio Management*, 19, 27–27.
- Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., & Iosifidis, A. (2017) Using deep learning to detect price change indications in financial markets. In *25th Eur Signal Process Conf EUSIPCO 2017*, 2017-January, pp. 2511–2515. <https://doi.org/10.23919/EUSIPCO.2017.8081663>
- Wang, M., Zhao, L., Du, R., Wang, C., Chen, L., Tian, L., & Stanley, H. E. (2018). A novel hybrid method of forecasting crude oil prices using complex network science and artificial intelligence algorithms. *Applied Energy*, 220, 480–495.
- Werbos, P. J. (1990). Backpropagation through time: What it does and how to do it. *Proceedings of the IEEE*, 10(1109/5), 58337.
- Xing, F. Z., Cambria, E., & Welsch, R. E. (2018). Natural language based financial forecasting: A survey. *Artificial Intelligence Review*, 50(1), 49–73. <https://doi.org/10.1007/s10462-017-9588-9>.
- Yoon, Y., Swales, G. S., & Margavio, T. M. (1993). A comparison of discriminant analysis versus artificial neural networks. *Journal of the Operational Research Society*, 44(1), 51–60. <https://doi.org/10.2307/2584434>.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*, 14, 35–62. [https://doi.org/10.1016/s0169-2070\(97\)00044-7](https://doi.org/10.1016/s0169-2070(97)00044-7).

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

Faisal Khalil<sup>1</sup>  · Gordon Pipa<sup>1</sup>

Gordon Pipa  
gpipa@uni-osnabrueck.de

<sup>1</sup> Institute of Cognitive Science, 49090 Osnabrück, Germany