2019 International Conference on Computing, Power and Communication Technologies (GUCON)
Galgotias University, Greater Noida, UP, India. Sep 27-28, 2019

Analyzing Varied Approaches for Forecast of Stock Prices by Combining News Mining and Time Series Analysis

1st Pranav Shankar
Dept. of Computer Science
UIET, Panjab University
Chandigarh, India
pranavshankar.work@gmail.com

2nd Manika Arvind Arora
Dept. of Computer Science
Albert Ludwigs Universität
Freiburg, Germany
manika23111993@gmail.com

3rd Rajat Kaushal
Dept. of Computer Science
UIET, Panjab University
Chandigarh, India
rajatkaushal140198@gmail.com

4th Iknoor Singh
Dept. of Information Technology
UIET, Panjab University
Chandigarh, India
iknoor.ai@gmail.com

***Abstract*—Forecasting of Stock Prices has become a trending topic of research relating to varied fields such as machine learning, economy and other fields. Most of the approaches to perform such a task make use of the time series data or news mining. Albeit several approaches have been proposed to combine the two but the choice among those remains a problem. Data mined from the News reports is analyzed with text mining techniques to obtain the mining results which are used for improving the accuracy of analysis algorithms. To perform the aforementioned task, there exist several approaches which may be linear, supervised or even unsupervised. This paper makes use of the financial data to predict the movement of the Dow Jones Industrial Average (DJIA) using varied methods and further delineates the comparison amongst the same. The text from the newspapers is represented using state-of-the-art Word Embedding to help resolve the information loss due to loss of context. Using these embedding, a neural network is trained to forecast markets. The predictions obtained are then evaluated and compared to standard Long Short Term Memory Networks (LSTM) based approach in tandem with popular Convolutional models and other neural network architectures. The results indicate that the neural network based approach out-performs the traditional methods and holds vast potential for further improvements.**

Keywords*— Bi-directional Long Short Term Memory Networks, Forecast, Dow Jones Industrial Average, Financial Times, Gated Recurrent Unit (GRU), Long Short Term Memory Networks (LSTM), News Mining, Stacked Long Short Term Memory Networks, Time Series Analysis, Trend Analysis.

I. Introduction

Time series have become ubiquitous as they are being used to measure data related to various phenomena (e.g., spread of a virus, temperature, sales, etc.). Forecast by time series data is highly beneficial (and necessary) for making optimized decisions, yet it remains to be a very challenging problem; using the historical values of the time series does not suffice. Using text mining and natural language processing (NLP), it is now possible to analyze both the historical market returns and the textual information such as news articles, disclosures, corporate announcements etc., for making even more informed predictions [1, 2]. This makes the task of forecasting markets using textual information a multidisciplinary problem. Approaches which make use of combination of time series and news mining analysis for forecasting have become popular paved; some being Long Short Term Memory Networks, Gated Recurrent Unit Network, Temporal Convolutional Network and many more. All these methods vary in their fundamental working due to which selection of the most suited network is required. Experiments are performed to predict the inter-week stock prices by taking the financial news for time separated intervals as input data and Dow Jones Index of next corresponding week as Target Label. There are multiple variables present in the dataset – date, last, high, close, open, low, turnover and several other fields. By the use of neural network based sequence modeling, a means for handling the drawbacks of the bag-of-words approach by traversing over each document word-by-word is provided and, since the word order is preserved, the contextual information is retained in the numerical representations produced. In this paper, we use varied architectures for modeling the textual data as well as for predicting the movement of market returns of Dow Jones Industrial Average (DJIA). The approach employs the word2vec algorithm provided by [14] that generates word embedding from a corpus, along with pre-trained word embedding of 300-dimensional vectors for a vocabulary of 3 million words. These embedding are then fed to a neural network model, one document (or batches of documents, depending on the network design) at a time to classify the movement of the stock returns.

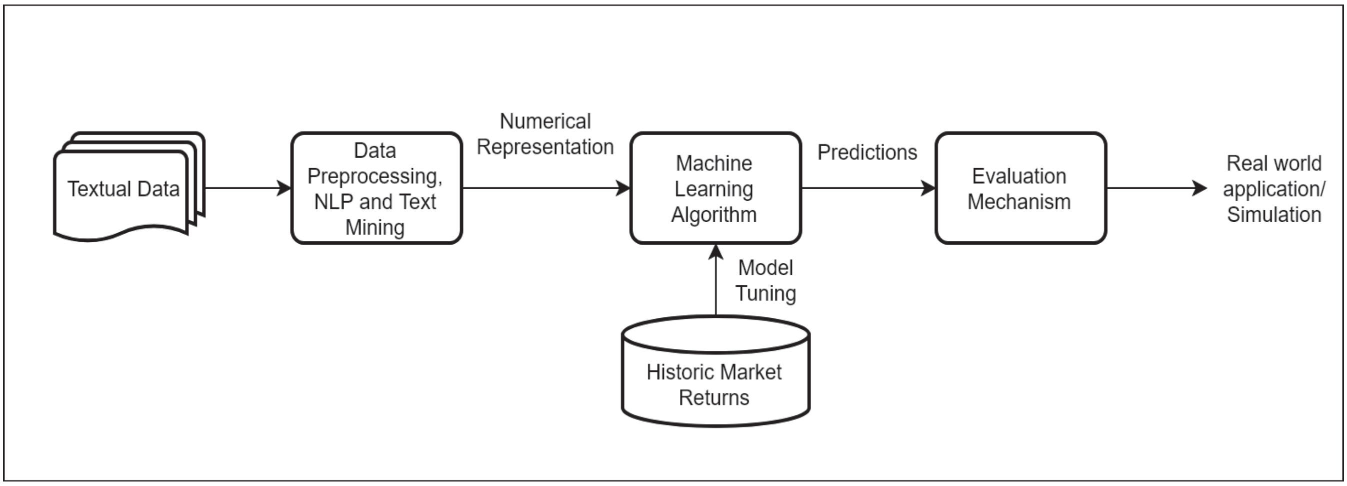

Fig. 1: Simplified pipeline of systems based on the application of text processing in Financial Domain. This pipeline is based on the schematic provided in [3]

This paper is organized into five major sections, the next section is dedicated to discussion on related works followed by the demonstration of our proposed work. Discussion on various experiments conducted for validation of the proposed architecture are grouped in the next to next section. The last and second last sections are dedicated to conclusions and future research directions respectively.

II. RELATED WORK

Time Series forecasting is one of the most widely used applications of machine learning. During the last decade there have been a number of artificial intelligence and machine teaching techniques used for predicting the stock market. For example, Kyoung-jae [4] used support vector machines for financial time series forecasting. Jae Won Lee [13] presented a way to apply reinforcement learning which is suitable for modelling and learning different types of interactions in reality. Furthermore, Jigar et al. [5] presented a comparative study of four prediction models, Support Vector Machine (SVM), Artificial Neural Network (ANN), naive-Bayes and random forest used for predicting stock price index for Indian stock markets. Several hybrid machine learning approaches are also purposed for forecasting. For instance, Rohit et al. [6] purposed a hybrid machine learning system based Support Vector Machines (SVM) and on Genetic Algorithm (GA) for stock market prediction. The study in. [7] combined ANN and the rule-based technique to forecast the direction of change of the S&P 500 stock index futures on a daily basis. Some of the other methods that have been used to forecast the stock market include evolutionary algorithms [9,10], Bayesian belief networks [8], fuzzy sets [12] and classifier systems [11]. Further, over the years, several reviews and surveys have been done by researchers from both financial, and computer science and statistical circles, in order to organize and identify the research gaps in the use of text mining for

Tackling economics and financial behavioral and modeling problems. The most recent reviews are provided by [15, 16, 17], the financial circles have cataloged works based on the source and type of textual information, the methodology of analysis, and the target of study. The work of [16] mainly focuses on the use of large textual-sample corporate disclosures as the source of data, along with the application of natural language processing, statistical and computational linguistic tools for analyzing the text.

The most well-known and easily available source of material is news. Several works, in the past, have made use of financial news for developing predictive models. One of the earliest known works, [18], use a collection of general news from Wall Street Journal, Financial Times and Reuters, to predict Dow Jones Industrial Average (DJIA), NIkkei 225, FT100, Hang Seng and Singapore Straits; using a bag-of-words model together with a predefined dictionary for a text corpus over the period of December 6, 1997 to 6 March, 1998. Another work utilizing news is [1], where US

financial news, a set of 6,602 samples, is used predict intra-day stock prices for the period starting from January 1 to December 2002, using a bag-of-word with TF-IDF transformation approach for representing text and SVM for prediction. Apart from using the entirety of the news, only specific parts of the news can also be used as input. This is demonstrated by [2] in their work, where they use a set of 12,830 headlines from leading electronic newspapers in Taiwan for predicting stock prices from June to November 2005. The advancement in computational resources has enabled the use of neural networks for solving various tasks including text and natural language processing problems. The work of [19], is an illustration of successful application of deep learning models to textual input. One of the uses of neural networks in regards to textual data, is to generate a representation of words and sentences while grasping the context and semantic related information. This is accomplished inherently through the cyclic structure of recurrent neural networks (RNN) as shown in [20]. Another work that uses textual input in parallel with a network model is [21], where word tuples are extracted from a set of 400,000 Bloomberg news headlines and processed for making stock price predictions using a deep learning model. This work, however, doesn't process text as a whole, therefore is not able to demonstrate the advantages of deep learning.

III. PROPOSED WORK

Stock Price prediction can be performed in different manners by changing the target such as inter-day, inter-week, inter-month and other variants but the methodology remains same.

This paper makes use of the variants of some of the most promising architectures to select the most optimal approach.

We have performed experiments using Long Short Term Memory Networks (linear as well as Convolutional), Gated Recurrent Unit Network, Temporal Convolutional Network, and Bidirectional Recurrent Neural Networks. Experiments are performed to predict the inter-week stock prices by taking the financial news for time separated intervals as input data and Dow Jones Index of next corresponding week as Target Label. Refer to Fig 2. For the combined architectural layout which delineates the alternate variations in the experiments being performed and presented in proposed work in the next section.

A. Data Collection and Preprocessing

The text corpus used in this work is obtained from Financial Times Historical Archive [22] which provides a comprehensive and unbiased research tool consisting of newspapers from 1888 to 2016. For this work the original archive which contained newspapers from 1888 to 2010 is used. For the purpose of this work, a version comprising of text files is used. This corpus consists of text files dated from January 1, 1950 to December 31, 2009, a total of 18,792 text files. These text files consist of only the front pages of Financial Times, since they cover the most relevant articles and also help in reducing the data.

The Table 1 consists of some basic statistics for words and sentences in the document collection. There are on average 1,695 words per document and 108 sentences per document in the corpus under study.

TABLE I: STATISTICAL CHARACTERISTICS OF TEXTUAL DATA. MEAN VALUES ARE CALCULATED PER DOCUMENT.

Timeline	From 01/01/1950 to 31/12/2009
Documents	18,792
Words	2,072,545
Mean Words	1,866,852
Sentences	1,694.886
Mean Sentences	108.373

The second part of the data set is formed by the economic time series, Dow Jones Industrial Average (DJIA), obtained from Bloomberg and contains the daily closing values from our period of study. The Table 2 shows the descriptive statistics for the time series. One of the most important statistics is the class ratio, i.e., positive to negative class ratio. These classes represent the direction of daily, weekly and monthly returns. The time series, however, only provides the closing values, therefore, percentage returns were computed before further analysis. For this, the closing values are used for calculating the daily, weekly and monthly percentage returns by calculating the rate of change between the closing value at time t_i belongs to t_1, t_2, t_3.. and the closing value at a time shift of 1,7 and 30 i.e., t_{i+1}, t_{i+7} and t_{i+30} respectively, as shown in the equation below where C_t and C_{t-1} are closing values.

$$R_t = \frac{C_t - C_{t-1}}{C_{t-1}} * 100$$

The obtained returns are then changed into class labels by comparing with zero, i.e., if the percentage return for any given day is less than 0, then the label is "negative", otherwise "positive". Three different labels are generated for daily, weekly and monthly returns each.

TABLE II: STATISTICAL CHARACTERISTICS OF ECONOMIC TIMES, THE DOW JONES INDUSTRIAL AVERAGE (DJIA). THE RATIO VALUES ARE POSITIVE VS. NEGATIVE MOVEMENT OF RETURNS. THE MEAN AND STANDARD DEVIATION ARE CALCULATED FROM PERCENTAGE RETURNS.

Timeline	Observations	Class Ratio	Mean % Return	Standard Deviation
Daily	18,792	0.5191	0.0277	0.7523
Weekly	18,792	0.5559	0.1492	2.0889
Monthly	18,792	0.5988	0.6607	4.2876

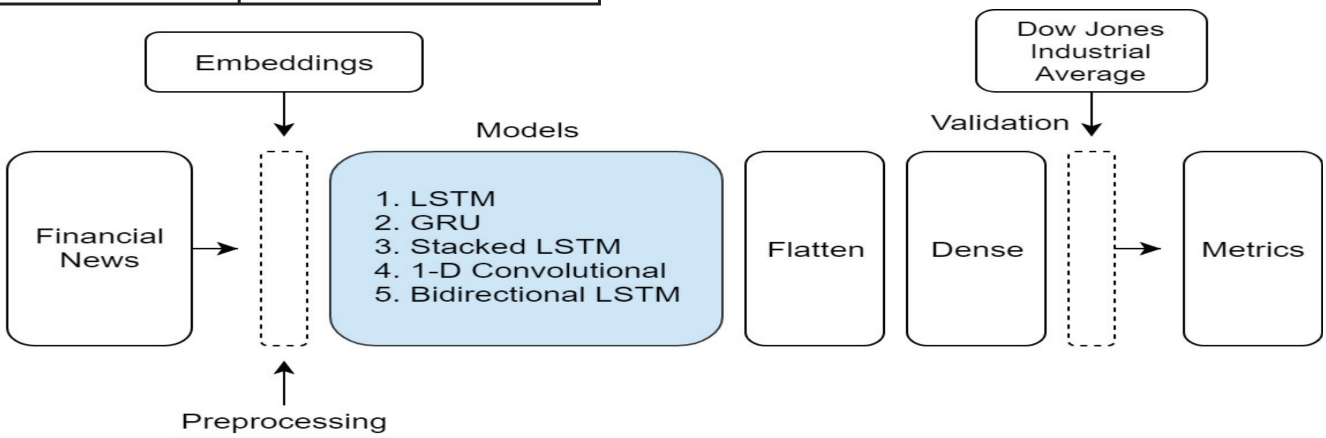

Fig. 2: Architecture Layout

The Financial Times textual data requires several preparation steps, before it can be used for analysis. For any text mining task, one must first determine the lowest level of breakdown of text for analysis, i.e., whether the analysis is to be done on sentence, word or character level. In this case, the analysis is done on word level, and therefore, the text must be broken down into individual words, also known as Tokenization, with each word being a token. The general order of execution of these pre-processing steps is, removal of symbols and numbers, Tokenization. It is important to identify and list these word , as they provide the most information about the subject matter of the text. This implies creation of a task specific vocabulary from the given text corpus, mostly based on the occurrence of a word [23]. We create a vocabulary for the given set of documents using the following steps:

• Create a list of unique words from the text corpus while counting the number of occurrences of each word.

• Using the vocabulary of 3,000,000 words provided in the pre trained word vectors obtained from Google [11], intersect the two lists to obtain a list of common words.

• From this list, remove words with number of occurrence less than 30.

The choice of minimum number of occurrence is 30 because the word2vec algorithm requires several repetitions of a word to generate a meaningful vector representation for it. Meanwhile it is also important to ensure that the vocabulary size is large enough to generate the desired number of vectors, here 300 for each word. By conducting a manual search over 150, 100, 50, 40, 30, the minimum occurrence of 30 is chosen as it provides a relatively well balanced vocabulary of 25,232 words. Extraction of features done by associating with each word in the vocabulary a distributed word feature vector such that the feature vector represents different aspects of the word, i.e., the association of each word with points in vector space [24]. Therefore, the number of features is much lower than the size of vocabulary. These word vectors can then be used to represent a corpus for further analysis.

Experiments are designed for making use of documents having the front page of Financial Times and using those as input, in the form of a sequence of one-hot encoded vectors x_i belongs to x_1, x_2, x_3.. of arbitrary length. A one-hot encoded vector is sparse vector with only one non-zero element, here, this element represents whether the word appears in the document or not. Further, tokenized documents along with embedding vectors are used as the main input.

B. Experiment-1

In this experiment we make use of Long Short Term Memory network architecture to forecast stock prices by combining News Mining and Time Series Analysis. In the domain of sequence prediction problems, LSTMs are being used widely and have proven to be extremely efficacious. Reasoning behind such an efficient working can be derived from the fundamental structure of LSTM to store past information which is important, but forget the information which is not. LSTM comprises of three gates namely input, forget and output gate. Input gate is responsible for the information addition to cell states. The forget gate is responsible for the removal of the information that has become irrelevant for the model whereas output gate selects the information which is to be shown in form of output.

TABLE III. OBSERVED PARAMETER VALUES FOR LSTM ARCHITECTURE.

Parameters	Results
Training loss	0.030
Training accuracy	0.998
Validation loss	0.269
Validation Accuracy	0.575
Testing accuracy	0.548

Observed parameters present a very stable result even though the training accuracy is higher than validation accuracy but still the model cannot be called over fitted due to the variability in the nature and processing of the data of the Financial Times.

C. Experiment-2

In this experiment we make use of Gated Recurrent Unit network architecture for forecasting stock prices by combining News Mining and Time Series Analysis. GRU (Gated Recurrent Unit) was introduced to solve the vanishing gradient problem which comes inherently in a standard recurrent neural network. GRU is often considered to be a variation of the LSTM networks because both have been designed in a similar manner and, in some scenarios, produces equally efficient results. GRU makes use of reset gate and update gate which act as two vectors, deciding what information needs to be passed to the output. The differentiating aspect about them is that they are capable of training which allows keeping information from long ago, without letting it get washed away with time or removal of information which is irrelevant to the process of prediction.

TABLE IV: OBSERVED PARAMETER VALUES FOR GRU ARCHITECTURE.

Parameters	Results
Training loss	0.020
Training accuracy	0.808
Validation loss	0.303
Validation Accuracy	0.576
Testing accuracy	0.536

D. Experiment-3

This experiment comprises of the usage of the Stacked Long Short Term Memory Network to obtain an ensemble architecture capable of using combined technique of news mining and time series analysis. Albeit the same can be obtained using the singular long short term memory networks but addition of layer in form of another stacked linear Long Short Term Memory network increases the level of abstraction of features from data. Also, chunking of observations over passage of time or representing the acquisition problem of features at different time scales.

TABLE V: OBSERVED PARAMETER VALUES FOR STACKED LSTM ARCHITECTURE.

Parameters	Results
Training loss	0.049
Training accuracy	0.959
Validation loss	0.449
Validation Accuracy	0.500
Testing accuracy	0.544

E. *Experiment-4*

Experiment comprises of the use of 1-D Convolutional Network for prediction of the prices of stock using Time Series and News Mining. Albeit Convolutional Networks do not appear to be the inherited choice for such a problem but a single layer network with can also serve the purpose. As depicted in Figure 2, Network is followed by one flatten and then dense layer as well before the validation module. The experiment provides high training accuracy but much lower testing accuracy depicting lower relevance and poorer performance for the model.

TABLE VI: OBSERVED PARAMETER VALUES FOR CONV 1-D ARCHITECTURE

Parameters	Results
Training loss	0.153
Training accuracy	0.959
Validation loss	0.816
Validation Accuracy	0.559
Testing accuracy	0.525

F. *Experiment-5*

In this experiment we make use of Bidirectional Recurrent Neural network architecture to forecast stock prices by combining News Mining and Time Series Analysis. In Bidirectional LSTM (BLSTM) the preceding input sequence and succeeding input sequence are used for exploiting all input data for achieving finest training process performance.

TABLE VII: OBSERVED PARAMETER VALUES FOR BIDIRECTIONAL LSTM ARCHITECTURE.

Parameters	Results
Training loss	0.019
Training accuracy	0.909
Validation loss	0.312
Validation Accuracy	0.556
Testing Accuracy	0.557

As described above in Experiment one, the variability of data causes the training accuracy to be much higher than validation which is noticeable in all the experiments.

The evaluations performed in this work are divided into two parts. The first part aims at forecasting the movement of the DJIA returns i.e., classification for daily or weekly or monthly movements, here depicted for weekly. The second part focuses on predicting the return value, i.e. regression.

Directional accuracy [25], is the most popular metric of evaluation in financial research that employs text mining. As the name suggests, the predicted market directions are compared against the actual directions. The literature in the financial domain, uses this metric to indicate the performance of the respective systems and models [3, 25, 26]. However, it is important to know that market movements as target labels results in an imbalanced data set, i.e., the market moves in one direction (generally, positive) more number of times as compared to the other direction (negative). This is depicted by the statistics in table 2. Since there is an imbalance in the classes, simply reporting accuracy is not enough. Using a meaningful measure of generalizability is a key requirement for evaluating the performance that a classification algorithm has achieved on a given dataset [27]. In order to indicate the performance of the neural network model, balanced accuracy [27] is used as the main metric of evaluation for predicting the DJIA return movements. Balanced accuracy is calculated by taking the average of true positive rate (TPR) and true negative rate (TNR), in a binary classification problem with classes positive and negative. Besides balanced accuracy, two more metrics are calculated and presented in this work. These are precision and recall. Precision measures the correctness achieved in predicting the positive labels, recall indicates how many observations of positive class are labeled correctly. The equations below show the calculation of these three metrics.

$$\text{Balanced Accuracy} = \frac{TPR + TNR}{2}$$

$$\text{Precision} = \frac{TP}{TP + FP}$$

$$\text{Recall} = \frac{TP}{TP + FN}$$

IV. Discussion

The main aim of this paper is to delineate the application of state-of-the-art natural language processing techniques in tandem with deep learning models for forecasting market movements. The work contributes by showing that neural networks can be used with Financial Times as text source for predicting market movements and thus opening up new research avenues in the financial analysis domain.

Figure 3 depicts the comparison of variation of validation accuracies of models during the training period in which Bidirectional LSTM tends to outperform all other models.

Figure 4 gives the comparison of performance metrics of different model architectures used for forecasting Stock Prices. From this figure, we can infer that stacked LSTM gives high precision value whereas LSTM and bidirectional LSTM gives high recall values. In case of F1 Score, Stacked LSTM gives the best results. From our proposed models, Bidirectional LSTM gives the highest accuracy. Thus both Stacked LSTM and Bidirectional LSTM have delineated promising results. Further, for the complete statistical comparison, refer to Table 8. The standard deviation, Root Mean Square Error (RMSE) and Mean Squared Error (MSE) are lowest for Conv 1-D. The Mean Absolute Error (MAE) is lowest for LSTM.

Comparisons

TABLE 8: STATISTICAL COMPARISON OF RESULTS OF AFORE DESCRIBED EXPERIMENTS.

Exp No.	Model Architecture	RMSE	MSE	MAE	Precision	Recall	F1 score	Accuracy
1	LSTM	0.577	0.333	**0.461**	0.301	**0.438**	0.357	0.548
2	GRU	0.564	0.318	0.476	0.354	0.430	0.389	0.536
3	Bidirectional LSTM	0.580	0.336	0.471	0.364	0.433	0.396	**0.557**
4	Stacked LSTM	0.614	0.377	0.493	**0.551**	0.425	**0.480**	0.504
5	Conv 1-D	**0.544**	**0.296**	0.483	0.362	0.418	0.388	0.525

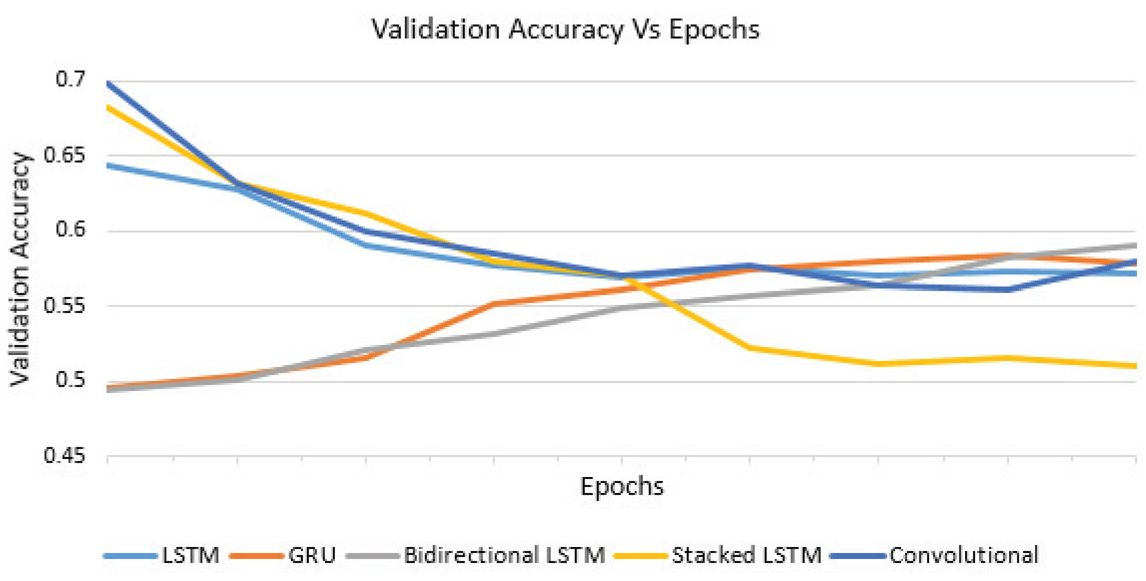

Fig. 3: Comparison of Validation accuracies of different model architectures used for forecasting Stock Prices.

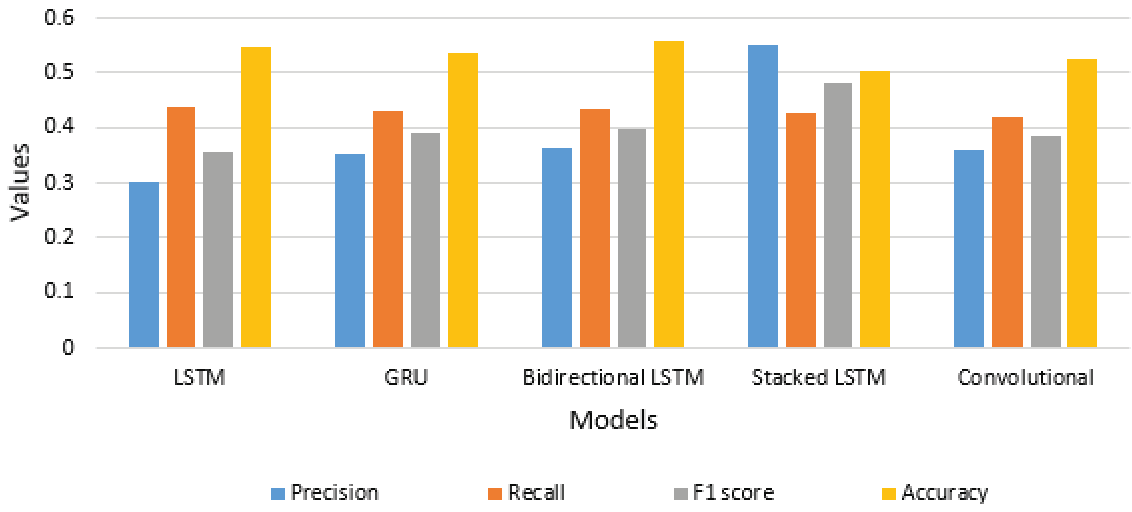

Fig. 4: Comparison of Performance Metrics of different model architectures used for forecasting Stock Prices

V. CONCLUSION

Financial news in any form, newspapers, online articles, disclosures, etc., largely influences investors and decision makers. While language and textual information can be easily processed by humans, it is not so for most automated systems. In this work, we have attempted to study the means of processing large quantity of textual data in regards to its application in the field of financial analysis. The work in this paper further bolsters the work done in the domain of market research by undertaking the task of predicting stock market movements using large quantity of unstructured textual data. Performing predictions using various architecture and further comparing results depicted that neural networks certainly withhold the capability to perform textual data classification and regression. Both Stacked LSTM and Bidirectional LSTM gave promising results. Even though Balanced Accuracy of later was higher but the former had higher values of other performance metrics.

With the constant improvements being made in the computational resources and deep learning methodologies, neural networks will not only help enhance the current market predicting systems by incorporating relevant news and textual information but also the randomness often observed in the markets.

VI. FUTURE SCOPE

The domain of market research is vast and due to impact of different kinds of information sources and implications of human behavior on the markets, any attempts at predicting the market movements have proven to be challenge. The results obtained in this work are indicative of the complexity of this problem. Firstly, improvement in the selection of text documents can be done. In this work we have taken the entirety of the front page of Financial Times as the input. However, the front pages of a newspapers comprise several articles, and it's possible that some of these might not hold any valuable information in context to markets, and therefore, can be dropped from the analysis. The articles that discuss markets can be then be chosen and used as the input.

Secondly, there is a possibility of using more sources of information that influence the market. For example, besides the news articles, corporate disposals and financial announcements can also be used as part of the textual input.

Thirdly, the neural network architecture itself has very high potential for improvements. The architecture used here is simple one-layer network and tuned over a relatively small parameter space. There is a large aspect of deep learning models that can explored in tandem with the financial text corpus for forecasting.

Lastly, the performance of the neural networks here is indicative that, they can be used for predicting other economic variables as well such as GDP, unemployment rates, inflation rates, etc.

ACKNOWLEDGMENT

We acknowledge the support and guidance of the faculty of Computer Science Department, University Institute of technology, Panjab University for providing the necessary guidance and motivation to conduct the research.

REFERENCES

[1] M.-A. Mittermayer, “Forecasting intraday stock price trends with text mining techniques,” in system sciences, 2004. Proceedings of the 37th annual hawaii international conference on, pp. 10–pp, IEEE, 2004.

[2] C.-J. Huang, J.-J. Liao, D.-X. Yang, T.-Y. Chang, and Y.-C. Luo, “Realization of a news dissemination agent based on weighted association rules and text mining techniques,” Expert Systems with Applications, vol. 37, no. 9, pp. 6409–6413, 2010.

[3] M. Kraus and S. Feuerriegel, “Decision support from financial disclosures with deep neural networks and transfer learning,” Decision Support Systems, vol. 104, pp. 38–48, 2017.

[4] Kim, Kyoung-jae. "Financial time series forecasting using support vector machines." Neurocomputing 55.1-2 (2003): 307-319.

[5] Patel, Jigar, et al. "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques." Expert Systems with Applications 42.1 (2015): 259-268.

[6] Choudhry, Rohit, and Kumkum Garg. "A hybrid machine learning system for stock market forecasting." World Academy of Science, Engineering and Technology 39.3 (2008): 315-318.

[7] Tsai, C. F., and S. P. Wang. "Stock price forecasting by hybrid machine learning techniques." Proceedings of the International MultiConference of Engineers and Computer Scientists. Vol. 1. No. 755. 2009.

[8] R. K. Wolfe, Turning point identification and Bayesian forecasting of a volatile time series, Computers and Industrial Engineering, 1988, pp 378–386.

[9] M. A. Kanoudan, Genetic programming prediction of stock prices. Computational Economics, 16, 2000, pp 207–236.

[10] K. J. Kim. Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index. Expert Systems with Applications, 19(2), 2000, pp 125–132.

[11] S. Schulenburg and P. Ross, Explorations in LCS models of stock trading, Advances in Learning Classifier Systems, 2001, pages 151–180.

[12] O. Castillo and P. Melin, Simulation and forecasting complex financial time series using neural networks and fuzzy logic, Proceedings of IEEE Conference on Systems, Man, and Cybernetics, 2001, pages 2664–2669.

[13] Lee, Jae Won. "Stock price prediction using reinforcement learning." ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No. 01TH8570). Vol. 1. IEEE, 2001.

[14] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.

[15] I. E. Fisher, M. R. Garnsey, and M. E. Hughes, "Natural language processing in accounting, auditing and finance: A synthesis of the literature with a roadmap for future research," Intelligent Systems in Accounting, Finance and Management, vol. 23, no. 3, pp. 157–214, 2016.

[16] C. Kearney and S. Liu, "Textual sentiment in finance: A survey of methods and models," International Review of Financial Analysis, vol. 33, pp. 171–185, 2014.

[17] F. Li, "Survey of the literature," Journal of accounting literature, vol. 29, pp. 143–165, 2010.

[18] B. Wüthrich, D. Permunetilleke, S. Leung, W. Lam, V. Cho, and J. Zhang, "Daily prediction of major stock indices from textual www data," Hkie transactions, vol. 5, no. 3, pp. 151–156, 1998.

[19] Y. Goldberg, "A primer on neural network models for natural language processing," Journal of Artificial Intelligence Research, vol. 57, pp. 345–420, 2016.

[20] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," nature, vol. 323, no. 6088, p. 533, 1986.

[21] X. Ding, Y. Zhang, T. Liu, and J. Duan, "Deep learning for event-driven stock prediction." in Ijcai, pp. 2327–2333, 2015.

[22] "Financial times." https://www.ft.com.

[23] T. Mikolov, W.-t. Yih, and G. Zweig, "Linguistic regularities in continuous space word representations," in Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 746–751, 2013.

[24] Y. Bengio, R. Ducharme, P. Vincent, and C. Jauvin, "A neural probabilistic language model," Journal of machine learning research, vol. 3, no. Feb, pp. 1137–1155, 2003.

[25] R. P. Schumaker and H. Chen, "Textual analysis of stock market prediction using breaking financial news: The azfin text system," ACM Transactions on Information Systems (TOIS), vol. 27, no. 2, p. 12, 2009.

[26] R. P. Schumaker, Y. Zhang, C.-N. Huang, and H. Chen, "Evaluating sentiment in financial news articles," Decision Support Systems, vol. 53, no. 3, pp. 458–464, 2012.

[27] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in Pattern recognition (ICPR), 2010 20th international conference on, pp. 3121–3124, IEEE, 2010.