

# Predictive Analytics with Structured and Unstructured data - A Deep Learning based Approach

Lipika Dey, Hardik Meisheri and Ishan Verma, TCS Research, New Delhi, India

**Abstract**— Predictive analytics over structured time-series data has always been an active area of research. With plentitude of textual information generating across different sources on the web, researchers have started combining relevant structured and unstructured data to improve predictions. In this paper, we have present a generic deep learning framework for predictive analytics utilizing both structured and unstructured data. We also present a case-study to validate the functionality and applicability of the proposed framework where we use LSTM for prediction of structured data movement direction using events extracted from news articles. Experimental results shows that the proposed model outperforms existing baseline.

**Index Terms**— Predictive analytics, deep neural networks, text-driven analytics

## I. INTRODUCTION

**B**USINESS Intelligence (BI) refers to a collection of technologies that help organizations analyze data to derive actionable intelligence. Traditionally, application of BI technologies has been restricted to structured, numerical business data like those reporting sales figures, customer acquisition figures etc. With the advances in unstructured data analytics techniques over the last one and half decade, businesses have also started looking at unstructured data like customer feedback, social media content, organizational communications etc. to gather intelligence around consumer sentiments, competition landscape etc. However, what is still missing is a common framework that can pull and analyze a multitude of heterogeneous data of both structured and unstructured types, to provide a richer set of insights.

The advantages of linking data from a multitude of sources are many. First of all, while business figures can clearly indicate how the business performed, they can hardly explain why it performed so. Often times, the attributable causes can be known from sources extraneous to the enterprise. For example, possible reasons for the dip in sales for a newly launched car may be learnt from social media. Further, the causal events themselves may be extraneous to the organization. For example, the actual reason for a dip in sales may be due to a higher interest rate levied on automobile

loans. Similarly, for certain periods, the reasons for a bad market may be attributed to socio-political disturbances over a region, rather than anything to do with the product or the organization itself.

While Business Analysts do use this kind of information, there exists no formal framework where structured and unstructured information can be simultaneously used for gathering business insights to make informed decisions. One of the key challenges was to learn these models from large volumes of historical data due to the computational complexities of text processing, both from quantitative and qualitative perspectives. Any kind of text processing requires handling of high volumes of data due to the potentially large vocabulary size in any natural language document collection. Further complexities are introduced due to issues like synonyms, lack of fixed structure, spelling errors, ever-changing vocabulary etc.

Deep learning based text-processing methods have provided an alternative way to represent words and text documents as vectors of numbers in a fixed dimensional space. It has been shown that these representations are capable of preserving and detecting semantic similarities without the use of additional linguistic resources. Deep neural networks have been employed to learn predictive models from large volumes of text data. Among these, the particularly interesting variants are those of Long Short-Term Memory (LSTM) networks [27], which exploit the sequential nature of words in a text for text classification. Obviously, LSTMs are not restricted to be used only for text classification, rather can be utilized for any kind of sequence-based prediction tasks.

In this paper, we propose the use of deep neural networks to build predictive models using both structured and unstructured data. The text data is first classified into pre-determined categories using supervised learning. Thereafter, quantification of the text data is done using volumes of category labels aggregated over pre-specified time units. An LSTM based network is thereafter employed to consume a multitude of time-series and predict the future events. We have presented the feasibility and effectiveness of the approach by predicting stock-market movements using both historic stock data and News data. We show that the combined model can predict stock movements more accurately than traditional predictive methods, which uses the time series for

Lipika Dey, Hardik Meisheri and Ishan Verma are with the TCS Research, New Delhi, India; email: (lipika.dey, hardik.meisheri, ishan.verma)@tcs.com

stock values only.

The model can be applied towards predicting any business data like sales, market-share etc. However, availability of large volumes of relevant historical text data is not easily available for most organizations. Typically, the kind of data that would be useful for such tasks would be customer communications, product- or service-related complaints, company policy-related documents etc.

## II. PREDICTIVE ANALYTICS – A BRIEF OVERVIEW OF TECHNIQUES USED BY BUSINESS ANALYSTS

For ages, business analysts have relied on predictive analytics models to identify risks and opportunities. These models learn relationships among many factors that are likely to affect business performance. These models are then employed to predict future outcomes depending on current factors, thus allowing analysts to obtain an assessment of risks or opportunities for a particular situation. Analysts also usually run “what if” simulations, wherein they experiment with various values of current factors for potential assessment of risks associated with different situations. Predictive analytics thus enables guided decision making.

Predictive analytics has a wide-range of applications in business. The predictive models provide a probability score for measurable business data elements. It may provide the probability of a customer churning or buying a product or service in the near future. It may provide an estimate of the number of vehicles or insurance policies likely to be sold in the next three months. It may also provide the likelihood of a customer defaulting on a loan based on his or her personal history or characteristics. In yet another set of applications, predictive analytics may be applied to determine whether a transaction is fraudulent or not, based on certain patterns learnt from the past. Applying predictive analytics for predicting stock data movement is an age-old statistical problem. Statistical models are also available for predicting the actual stock price or stock volume of a commodity.

In this section, we will try to provide a very brief overview of various classes of predictive techniques, which will just help understand the different classes of models. Providing in-depth knowledge about any specific technique is beyond the scope of the paper.

As seen from the examples, the context of applying predictive analytics can be quite diverse. The outcomes expected can also vary from simple binary values like YES / NO or TRUE /FALSE for churn or fraud prediction to predicting actual numerical values for stocks or sales. Consequently, the types of methods applied for the tasks are also quite diverse.

Predictive analytics techniques can be broadly grouped into regression techniques and machine learning techniques.

### A. Regression Techniques

Regression analysis is a predictive modelling technique which explores the dependence relationship of a target variable, like sales trends, churn possibility or possibility of a transaction to be fraudulent etc. on predictor variables.

Regression techniques focus on finding a mathematical equation that can capture the interactions between the different variables in consideration, typically trying to minimize the overall error between the model-predicted values and the real values.

Linear regression models learn to predict the response variable as a linear function of the parameters. These parameters are learnt or adjusted so that a measure of fit like the sum of squared residuals is minimized.

Logistic regression models [2] on the other hand assign probabilities to possible outcomes. A binary outcome variable is transformed to an unbounded continuous variable, and a regular multivariate model is estimated.

Time series models are used for forecasting future behaviour of variables when the past data points exhibit internal structures like autocorrelation, trends or seasonal variations. Stock or sales data are perfect examples of such data. Standard regression techniques cannot model these internal structures. Time series models are capable of decomposing these trends and seasonal components and thereby produce better models. Popular time series models are autoregressive model (AR), moving average (MA), a combination of the two called auto-regressive moving average (ARMA) and auto-regressive integrated moving average (ARIMA) [1].

Classification and regression trees are non-parametric decision tree learning methods that produce classification or regression trees depending on whether the dependent variable is categorical or numeric. These are based on hierarchical optimal discriminant analysis that are generalizations of optimal discriminant analysis. Decision trees [3] are a collection of rules based on variables in the modelling data set, where the rules are deduced to obtain the best split to differentiate observations belonging to different target classes. Decision tree rules are explanatory and are often preferred by analysts. For example, to decide whether a new loan applicant is likely to default or not, a decision tree based model not only predicts the decision but also provides the rule that was applied to come to the decision thereby helping the analyst understand the reasons for it.

### B. Machine Learning Techniques

Machine learning based models are also applied for predictive analytics in applications like medical condition diagnostics, fraud detection etc. However, unlike regression or classification trees, the model here remains a black-box without explicit insight into the underlying relationships among the predictor variables, which can be terribly complex, and it is deemed sufficient to predict the dependent variable only.

Among the various machine learning models, the artificial neural networks [4, 5] that were introduced in the sixties, inspired by the human nervous system have gained a huge surge in popularity in recent times, due to their capability to learn very complex relationships among large numbers of predictor variables. There are a wide variety of neural network architectures that are useful for different types of classification

tasks. While earlier, neural networks were restricted to using three layers of neurons, the input layer, a hidden layer and an output layer, deep neural networks [6,7], with more than one hidden layers have gained popularity. A large number of neurons and their interconnections are capable of modelling highly non-linear relationships between input and output variables. These architectures are also found to extract useful features by themselves from large volumes of training samples without explicit feature engineering.

Below are some of the most commonly used networks that are suitable for prediction tasks:

#### 1) *Multilayer Perceptron*

These networks use more than one hidden layer of neurons and are also known as deep feed-forward neural networks.

#### 2) *Convolutional Neural Networks*

Convolutional Neural Networks (CNNs) [8] are a class of neural networks which performs convolutions between the desired filter and the input data. These Networks are highly efficient in learning hierarchical features from the data by capturing neighbouring relationships among features.

#### 3) *Recurrent Neural Network*

These are types of neural network in which hidden layer neurons have self-connections, thereby making it possible for a neuron to possess memory. Recurrent neural networks [9] are very suitable for text processing tasks since the interpretation of a word in a text is dependent on its context or its neighbouring words. Thus, networks that can model these interrelationships of words by taking their sequence into account are better at text class prediction tasks than their counterparts, which treat a text as a bag of words.

#### 4) *Long-Short Term Memory Network (LSTM)*

These are extensions of recurrent neural networks in which memory cell is incorporated inside each hidden layer neuron. LSTMs are good at modelling long-distance relationships also among variables like words within a text separated by many words in between. LSTMs are applicable for analyzing any kind of sequential data.

### III. USING UNSTRUCTURED DATA FOR PREDICTIVE ANALYTICS – RELATED WORK

As discussed earlier, till very recently much of predictive analysis in business dealt with structured data alone. With the organizations opening up to the idea of using consumer-generated unstructured text like complaints, service logs, social media data etc. several researchers are exploring the effect of these types of text data on business outcomes. One popular area of research has been to study the effects of online reviews on sales of products and services. Effects of online reviews on few areas like fashion and movies have been studied in depth. While most of these works study the impact of the reviews, few provide mechanisms for including the factors in a predictive model.

In [11], authors have provided a detailed survey of various

types of predictive techniques used for predicting the fashion market. The fashion market is affected by several factors like changing weather conditions, trans-continental production facilities, holidays, public events as well as economic situations etc. Along with these, lack of historical data for new kinds of fashion items motivated analysts working in this area to especially look for methods that can combine unstructured data into the predictive process to make better predictions.

One stream of research in this area focuses on the integration of expert judgment and combining it with statistical forecasts, as [13] and [14] have shown that adjustments to the statistical models based on past knowledge of experts led to more accurate forecasts.

However, a second stream of research focuses on the use of machine learning based models that could easily integrate unconventional factors like the ones stated above. Machine learning based models were found to produce much better results than traditional regression techniques for this domain.

Sun et al. [15] proposed the use of extreme learning methods (ELM) for forecasting sales at item levels. In [16], Thomassey and Happiette propose the use of soft computing methods like fuzzy inference systems and neural networks to predict sales. Teucke et al. [17] proposed using a combination of decision trees to determine articles likely to be re-ordered and then support vector machines for the actual forecasts to obtain more accurate results. While several other variations of these models have been proposed by other researchers, as presented in [11], the point to be noted is that most of these works have used only numbers from the retail industry to build their models.

In [10], Yu et al. analyzed large volumes of online movie reviews and showed that both the sentiments expressed in the reviews and quality of reviews have a significant impact on the future sales performance. Sentiments are detected using Probabilistic Latent Semantic analysis (PLSA). Based on the sentiments, they propose an autoregressive sentiment-aware model for sales prediction. This model is further improved by considering the quality factor of a review.

An area where text inputs have been heavily used for prediction is that of News for predicting stock market data. Researchers in a number of studies have analyzed texts from social network services (SNS), blogs and news to analyze correlations between stock prices and public emotion as a reaction to social events and news [28-32]. In [33] Luss and Aspremont show that information extracted from news articles can be used to predict intraday price movements of financial assets using support vector machines. In [34], Verma et. al. presented results to show that stock trends can be better predicted by taking News events into consideration along with actual stock values.

From the above discussion, we find that though machine-learning based models are gaining traction with business analysts, using unstructured data for prediction is yet to pick up. We also see that though they have potential to handle large volumes of data, deep-learning based models have not gained

wide attention of researchers in this field. In the next few sections, we propose a deep-learning based model, which can effectively do this.

IV. DEEP LEARNING MODELS FOR PREDICTIVE ANALYTICS

We choose deep learning architectures for predictive analytics since these models can learn from large volumes of sequential input data without explicit feature engineering. Additionally, the same architectures can easily handle both structured and unstructured sequential data. Before going into the details of the proposed model, we provide a brief introduction of word vectors and document vectors, which are used for predictions.

Semantic vector space models of language represent each word as a real-valued vector. These vectors can be used as features in a variety of text-based applications, such as information retrieval, document classification, question answering, and named entity recognition. Bengio et al. [22] introduced a model that learns word vector representations as part of a simple neural network architecture for language modelling.

Recently, Mikolov et al. [21] introduced a scheme for learning the representations using skip-grams. They also proposed a new evaluation scheme based on word analogies that probe the finer structure of the word vector space by examining not the scalar distance between word vectors, but rather their various dimensions of difference. For example, the analogy “king is to queen as man is to woman” should be encoded in the vector space by the vector equation king – queen = man – woman. Though this model does well on analogy tasks, they do not utilize the statistics of the corpus since they train on separate local context windows instead of on global co-occurrence counts, and hence perform poorly for most text-based applications. In 2014, Pennington et al. [18] present a set of global vectors for words, called GloVe vectors which uses a specific weighted least squares model that trains on global word-word co-occurrence counts and thus makes efficient use of statistics. This model produces a word vector

space with meaningful substructure that are found to be suitable for a large number of downstream text processing tasks.

Vector-based representation for words can be further extended to a paragraph or whole document. Mikolov et. al. [23], proposed a scheme where along with learning word representation if explicit information about the paragraph or the document is given, it can effectively encode the whole document into fixed sized vectors. These representations are useful over word representations while processing long text.

In the next section, we present the complete architecture for our proposed predictive systems.

A. Text Classification for Business Data Prediction

In order to effectively quantify without losing human understandability, unstructured business data needs to be assigned to pre-specified categories. One can treat these categories as event labels pertaining to different event type which are known and have defined explanation. One such set of business events is the PESTEL framework where PESTEL represents Political, Economic, Social, Technological, Environmental and Legal events. The set of PESTLE events [12] had been proposed for analyzing general environments of a business organization. These events represent six broad macro-economic factors that may have an impact on an organization.

To evaluate the impact of PESTEL events (or any set of pre-determined business events) on business outcomes, one can use Granger Causality test. The Granger causality test is a statistical hypothesis test for determining whether a time-series X is useful in forecasting another time-series Y. Time-series X is accepted as impacting Y, if and only if, prediction of future values of Y improves after taking values of X into consideration.

In this case our null hypothesis is that ‘PESTEL factors reported in News do not Granger-cause changes in indices’. We apply Granger causality test using PESTEL event occurrence time-series and stock indices time-series.

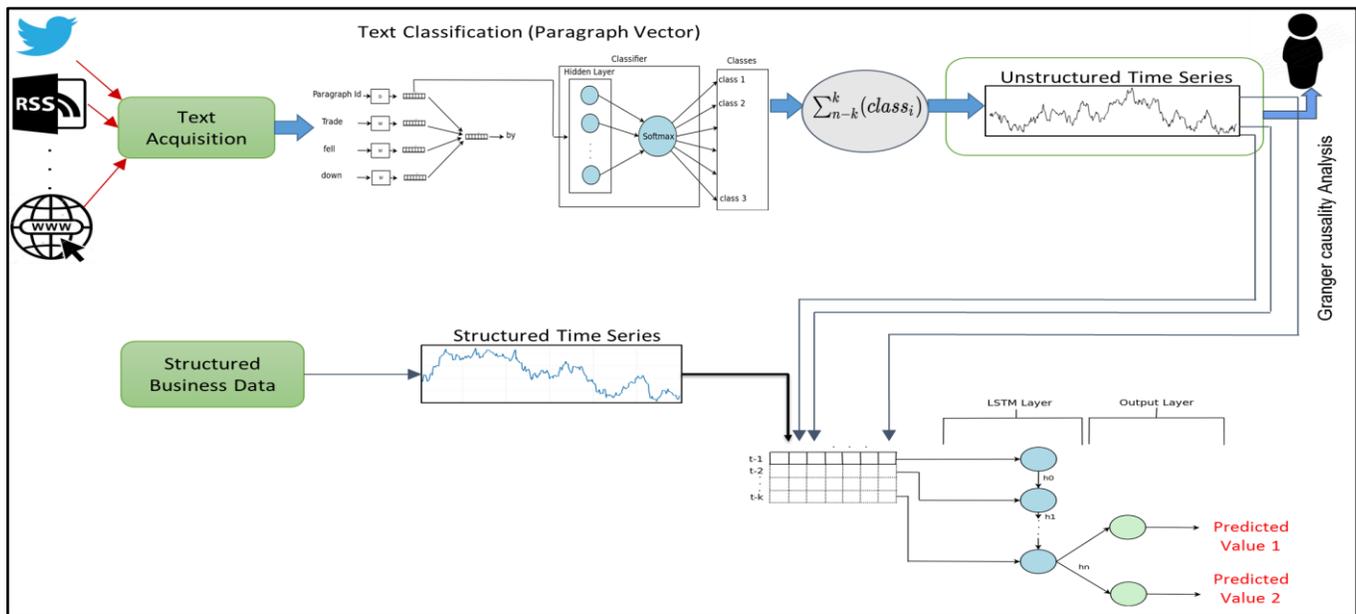


Figure 1 System Architecture

$$y_t = \alpha + \sum_{i=1}^n \beta_i y_{t-i} + \varepsilon_t$$

$$y_t = \alpha + \sum_{i=1}^n \beta_i y_{t-i} + \sum_{i=1}^n \gamma_i x_{t-i} + \varepsilon_t$$

The first one tries to predict future values of Y from past values. The second model takes past values of both X and Y to predict future values of Y, where series X represents quantified PESTEL factors.

## V. SYSTEM ARCHITECTURE

Figure 1 shows the complete generic system pipeline from information acquisition to final prediction for the proposed predictive system. The upper layer is dedicated for acquiring and processing unstructured text. Text documents are acquired from various sources like news websites, consumer forums, twitter as well as other social media sites, using dedicated crawlers, APIs or RSS feeds. This layer supports stream processing. Structured information can be obtained from various business sources such as market capital, market share, stock etc. These can also be streaming in from different sources or read from internal databases.

All text content is subjected to classification. A multi-label classifier that can assign probabilities for each class to each document is designed for the purpose.

This is achieved through a two-step procedure. First, we use paragraph vector model [23] as stated earlier to represent the long text into a vector with fix dimension. Second, we map this fix dimension vector to our predefined classes using a softmax classifier which gives us the probability of each class in that document.

Paragraph vector model is an extension of learning distributed word representations using gradient descent method. While learning, for every window document ID is given to the model to learn the overall representation of that document. Window here is defined as the number of words model uses to predict the next word in the series. Paragraph vector  $D$  is learnt along with the word vectors  $W$ .

Since external world data like tweets, News or social media sentiments stream in continuously, but business data is only aggregated over fixed time-periods like days, weeks or months, an aggregator is introduced over the softmax outputs. The aggregator aggregates the classifier content over fixed time-periods for each class. This aggregation over temporal domain generates a time series from unstructured data. This is then combined with the structured data time-series in the LSTM network to provide the final prediction.

### A. LSTM

LSTMs are a special case of RNNs where they tackle the classical problem of vanishing (or exploding) gradients. Information flow in LSTMs are controlled through three different gates, forget gate, output gate and input gate. Along with these gates, there is memory cell which is present, which helps in remembering long sequences. The whole process of learning can be formulated into information flow over a temporal domain. The function of input gate is to regulate new information flowing, output gate regulates the outward flow of

information for that time step and forget gate regulates the amount of flow that should be discarded. Single cell of LSTM is shown in Figure 2.  $\langle \rangle$  represents a vector,  $p$  denotes the number of hidden units in an LSTM cell. Input gate (denoted by  $i$ ), Output gate (denoted by  $o$ ), forget gate (denoted by  $f$ ) and memory cell (denoted by  $c$ ) govern the LSTM cell.  $t$  indicates the time-step or  $t^{\text{th}}$  time instance. LSTM follows following equations:

$$i_t^p = \sigma(w_{xi}^p x_t + w_{hi}^p h_{t-1}^p + w_{ci}^p c_{t-1}^p + b_i^p)$$

$$f_t^p = \sigma(w_{xf}^p x_t + w_{hf}^p h_{t-1}^p + w_{cf}^p c_{t-1}^p + b_f^p)$$

$$c_t^p = f_t^p c_{t-1}^p + i_t^p \tanh(w_{xc}^p x_t + w_{hc}^p h_{t-1}^p + b_c^p)$$

$$o_t^p = \sigma(w_{xo}^p x_t + w_{ho}^p h_{t-1}^p + w_{co}^p c_{t-1}^p + b_o^p)$$

$$h_t^p = o_t^p \tanh(c_t^p)$$

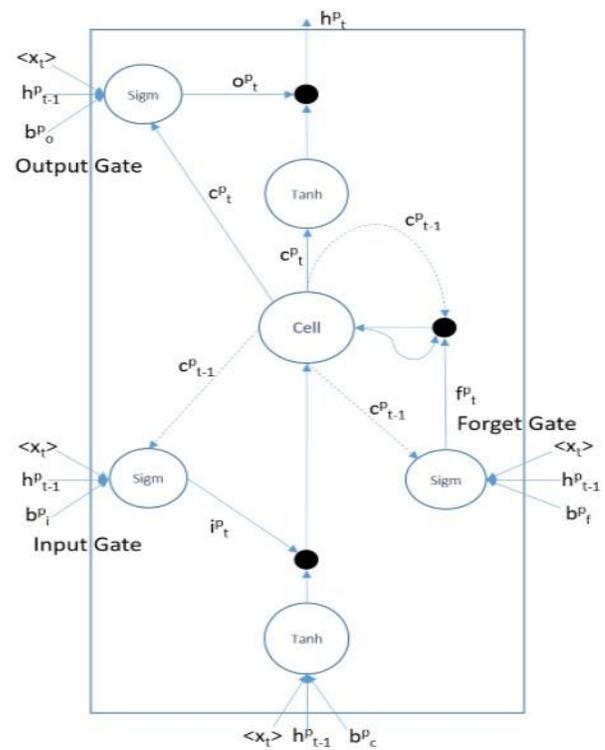


Figure 2 LSTM Cell

Training procedure in LSTM are done through truncated Back Propagation through time (BPTT) [24], which applies backpropagation over time and make changes in the weights after the whole sequence has been passed through LSTM cell.

The non-stationarity present in most of the time series can be modelled using LSTM, where it can store seasonality as well as the overall trend which effectively uses long term and short term dependencies.

## VI. CASE STUDY

Due to confidentiality issues associated to internal structured and unstructured business data, we are presenting results for publically available news and stock market data. There is also huge consensus in the business community on the importance of news articles over stock prices. Research results are also available for the same. Hence, we show results over this domain using the predictive analysis framework presented earlier. The accuracy of trend prediction is shown to better than other machine learning models.

For this task, we have used News articles as the source of text information. News articles are pulled from the RSS feeds provided by a number of agencies and crawling news websites. Stock acquisition is done from the official website of National Stock Exchange of India (<https://www.nseindia.com/>). Next, we have used paragraph vector model with a window size of 8 and fix learning rate of 0.25.

Next, vector representation of each news document is mapped to one the six PESTEL classes as described earlier.

We have used 5 different Indexes from the National Stock Exchange namely, NIFTY 50 (represents about 65% of free floating market capitalization), NIFTY Bank (consists of 12 large Indian banking stocks), NIFTY Auto (15 stocks consisting of Automobile sector including cars & motorcycles, heavy vehicles, auto ancillaries, tires, etc.), NIFTY IT (Companies in this index are those that have more than 50% of their turnover from IT related activities), NIFTY Energy (Index includes companies belonging to Petroleum, Gas and Power sub sectors).

We have investigated the dependency through Granger causality between structured and unstructured time series. In addition to that, we use these PESTEL classes as input to LSTM model for predicting weather stock will rise or fall.

### A. Data Description

Documents are collected as mentioned earlier from different sources. The collection contains over 258144 documents collected for the duration of Jan-2013 to Feb-2017. For training paragraph vector model we have used approximately 1200 labelled document equally distributed for each PESTEL class.

Data aggregated over a day is then split into training and testing set. Data from 1<sup>st</sup> Jan 2013 till 31<sup>st</sup> Aug 2016 is used for training purpose while data from 1<sup>st</sup> Sep 2016 till 31<sup>st</sup> Jan 2017 is used for testing set. Data corresponding to stock is not available for weekend and public holidays due to which total instances for prediction tasks are 912, out of which 820 are used for training and 92 for testing.

### B. Experiments and Results

Table 1 shows the confusion matrix of classification over PESTEL classes using paragraph vector model. After 10 fold cross-validation of 70-30 train test split, we achieved 91.94% accuracy for PESTEL classes.

Table 1 Confusion Matrix for PESTEL Classifier

	Pol.	Eco.	Soc.	Tech.	Env.	Leg.
Political	187	0	0	0	0	13
Economy	0	174	0	26	0	0

Social	3	0	184	10	0	3
Technology	0	0	3	190	0	7
Environment	0	10	0	13	174	3
Legal	0	0	0	3	0	197

Table 2 shows results for Granger causality between different stock indices and the PESTEL classes aggregated over a day. Highlighted cells in the table indicate which PESTEL features have a high impact on indices. In addition to this following observations can be made:

Table 2 Granger causality results

Index	No of Days	Political	Economy	Social	Technology	Environment	Legal
NIFTY 50	1 Day	0.52	0.02	0.54	0.39	0.4	0.6
	2 Day	0.31	0.04	0.84	0.34	0.04	0.38
	Week	0.47	0.16	0.94	0.67	0.06	0.53
NIFTY Auto	1 Day	0.35	0.02	0.92	0.32	0.55	0.47
	2 Day	0.29	0.04	0.85	0.36	0.26	0.4
	Week	0.43	0.17	0.8	0.66	0.34	0.49
NIFTY IT	1 Day	0.23	0.02	0.2	0.08	0.09	0.11
	2 Day	0.15	0.01	0.1	0.04	0.03	0.04
	Week	0.22	0.02	0.09	0.12	0.01	0.09
NIFTY Bank	1 Day	0.62	0.05	0.29	0.56	0.52	0.73
	2 Day	0.35	0.07	0.54	0.57	0.1	0.49
	Week	0.54	0.29	0.74	0.99	0.17	0.71
NIFTY Energy	1 Day	0.76	0.17	0.1	0.93	0.95	0.45
	2 Day	0.98	0.44	0.11	0.98	0.44	0.69
	Week	0.77	0.89	0.13	0.49	0.49	0.52

1) NIFTY 50 and NIFTY Bank have a significant long-term impact on social and Technology features respectively.

2) NIFTY Auto also get impacted from social class but from a short-term perspective.

3) NIFTY Energy is affected by multiple PESTEL classes which include the short-term impact from technology and environment while technology and political also affect over two-day lag.

Above observation shows that there is a significant impact of news events over stock value, and the introduction of these features leads to better prediction of stock indices.

Table 2 Stock Index Prediction with PESTEL Events and previous Index Values

Index	No of Days	SVM		LSTM	
		Accuracy	MCC	Accuracy	MCC
India Volatility Index	1	61.96	0.169	61.96	0.154
	2	56.52	0.071	<b>67.39</b>	<b>0.304</b>
	5	58.24	0.099	<b>64.83</b>	<b>0.250</b>
NIFTY 50	1	53.26	0.014	<b>58.69</b>	<b>0.163</b>
	2	51.09	0.043	<b>52.17</b>	<b>0.098</b>
	5	52.75	0.035	<b>54.95</b>	<b>0.113</b>
NIFTY Bank	1	57.61	0.267	51.08	0.019
	2	51.09	0.01	<b>55.43</b>	<b>0.114</b>
	5	54.95	0.049	<b>56.05</b>	<b>0.070</b>
NIFTY Auto	1	53.26	0.063	50	0.103
	2	52.17	0.051	51.09	0.146
	5	58.24	0.165	51.65	0.105
NIFTY IT	1	55.43	0.097	<b>56.52</b>	<b>0.118</b>
	2	51.09	0.036	<b>54.35</b>	<b>0.071</b>

	5	52.75	0.032	<b>56.04</b>	<b>0.122</b>
<b>NIFTY Energy</b>	1	51.09	0.188	<b>58.69</b>	<b>0.118</b>
	2	54.35	0.093	<b>57.61</b>	<b>0.079</b>
	5	59.35	0.133	58.24	0.083

We have conducted experiments taking variable length sequences with sequence length been, 1 day, 2 days and 1 week (5 days since the weekend is not counted). We have considered state of the art SVM with Radial Basis kernel as the baseline for our experiments. It can be observed that LSTM outperforms the SVM in most of the cases. We have used accuracy and Matthews correlation coefficient (MCC) as evaluation metrics'. Prediction results are shown in Table 3.

It is observed that in addition, looking at larger sequences improves the performance of LSTMs as compared to SVMs. It can be observed from Table 3 that for most of the indices LSTM predictions are outperforming SVMs by a margin of 2-11% in terms of accuracy.

## VII. CONCLUSION

In this paper, we have presented a generic deep learning framework for predictive analytics using both structured and unstructured data. It is notable that the same architecture can easily handle both structured and unstructured sequential data. We have presented a case study which demonstrate the capability of the proposed framework.

We intend to extend this work in future to cover a wider range of factors along with a more fine-grained representation of generic events and also experiment with recent advancements in deep learning based prediction algorithms.

## REFERENCES

- [1] Box, George; Jenkins, Gwilym M.; Reinsel, Gregory C. (1994). *Time Series Analysis: Forecasting and Control* (Third ed.). Prentice-Hall. ISBN 0130607746.
- [2] Cox, DR (1958). "The regression analysis of binary sequences (with discussion)". *J Roy Stat Soc B*. 20: 215–242.
- [3] Rokach, Lior; Maimon, O. (2008). *Data mining with decision trees: theory and applications*. World Scientific Pub Co Inc.
- [4] Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain". *Psychological Review*. 65 (6): 386–408.
- [5] Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain". *Psychological Review*. 65 (6): 386–408.
- [6] Hinton, G. E.; Osindero, S.; Teh, Y. W. (2006). "A Fast Learning Algorithm for Deep Belief Nets" (PDF). *Neural Computation*. 18 (7): 1527–1554.
- [7] Hinton, G. E.; Osindero, S.; Teh, Y. W. (2006). "A Fast Learning Algorithm for Deep Belief Nets" (PDF). *Neural Computation*. 18 (7): 1527–1554.
- [8] LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013.
- [9] J. L. Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.
- [10] Xiaohui Yu, Yang Liu, Xiangji Huang and Aijun An: Mining Online Reviews for Predicting Sales Performance: A Case Study in the Movie Domain, *IEEE Transactions on Knowledge and Data Engineering*, Volume: 24, Issue: 4, April 2012
- [11] Samaneh Beheshti-Kashi, Hamid Reza Karimi, Klaus-Dieter Thoben, Michael Lütjen & Michael Teucke: A survey on retail sales forecasting and prediction in fashion markets, *Systems Science & Control Engineering*, Dec. 2014, Pages 154-161.
- [12] Dare, C. (2006). The UK tour-operating industry: A competitive analysis. *Journal of Vacation Marketing*, 6(4), 357-367.
- [13] Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510–522.
- [14] Franses, P. H., & Legerstee, R. (2013). Do statistical forecasting models for SKU-level data benefit from including past expert knowledge? *International Journal of Forecasting*, 29(1), 80–87.
- [15] Sun, Z.-L., Choi, T. M., Au, K-F., & Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, 46(1), 411–419.
- [16] Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. *International Journal of Production Economics*, 128(2), 470–483.
- [17] Teucke, M., et al. (2014). Forecasting of seasonal apparel products. In H. Kotzab, J. Pannek, & K.-D. Thoben (eds.), *Dynamics in Logistics. Fourth International Conference, LDIC 2014 Bremen, Germany, February 2014 Proceedings*. Springer.
- [18] Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).
- [19] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013a. Efficient Estimation of Word Representations in Vector Space. In *ICLR Workshop Papers*.
- [20] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013b. Distributed representations of words and phrases and their compositionality. In *NIPS*, pages 3111–3119.
- [21] Tomas Mikolov, Wen tau Yih, and Geoffrey Zweig. 2013c. Linguistic regularities in continuous space word representations. In *HLT NAACL*.
- [22] Yoshua Bengio, Rejean Ducharme, Pascal Vincent, and Christian Janvin. 2003. A neural probabilistic language model. *JMLR*, 3:1137–1155.
- [23] Le, Q. V., & Mikolov, T. (2014, June). Distributed Representations of Sentences and Documents. In *ICML* (Vol. 14, pp. 1188-1196).
- [24] Werbos, P. J. (1990). Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, 78(10), 1550-1560
- [25] Dare, C. (2006). The UK tour-operating industry: A competitive analysis. *Journal of Vacation Marketing*, 6(4), 357-367.
- [26] Richardson Jr, J. V. (2006). The library and information economy in Turkmenistan. *IFLA journal*, 32(2), 131-139.
- [27] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997.
- [28] J. R. Nofsinger, "Social Mood and Financial Economics," *J. Behav. Finance*, vol. 6, no. 3, pp. 144–160, Sep. 2005.
- [29] W. S. Chan, "Stock price reaction to news and no-news: drift and reversal after headlines," *J. Financ. Econ.*, vol. 70, no. 2, pp. 223–260, Nov. 2003.
- [30] N. Strauß, R. Vliegthart, and P. Verhoeven, "Lagging behind? Emotions in newspaper articles and stock market prices in the Netherlands," *Public Relat. Rev.*
- [31] G. Ranco, D. Aleksovski, G. Caldarelli, M. Grčar, and I. Mozetič, "The Effects of Twitter Sentiment on Stock Price Returns," *PLOS ONE*, vol. 10, no. 9, p. e0138441, Sep. 2015.

- [32] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011.
- [33] Luss, R., & d'Aspremont, A. (2015). Predicting abnormal returns from news using text classification. *Quantitative Finance*, 15(6), 999-1012.
- [34] Ishan Verma, Lipika Dey, and Hardik Meisheri. 2017. Detecting, quantifying and accessing impact of news events on Indian stock indices. In *Proceedings of the International Conference on Web Intelligence (WI '17)*. ACM, New York, NY, USA, 550-557.