# Sentiment Analysis of Multilingual Tweets Based on Natural Language Processing (NLP)

Abhijit Bera, OmDayal Group of Institutions, India

Mrinal Kanti Ghose, GLA University, Mathura, India

Dibyendu Kumar Pal, Asansol Engineering College (AEC), India

## ABSTRACT

Multilingual sentiment analysis plays an important role in a country like India with many languages as the style of expression varies in different languages. The Indian people speak in total 22 different languages, and with the help of Google Indic keyboard, people can express their sentiments (i.e., reviews) about anything in the social media in their native language from individual smart phones. It has been found that machine learning approach has overcome the limitations of other approaches. In this paper, a detailed study has been carried out based on natural language processing (NLP) using simple neural network (SNN), convolutional neural network(CNN), and long short-term memory (LSTM) neural network followed by another amalgamated model adding a CNN layer on top of the LSTM without worrying about versatility of multilingualism. Around 4,000 samples of reviews in English, Hindi, and in Bengali languages are considered to generate outputs for the above models and analyzed. The experimental results on these realistic reviews are found to be effective for further research work.

## KEYWORDS

Machine Learning, Multilingual Tweets, Natural Language Processing, Neural Networks, Sentiment Analysis

## 1. INTRODUCTION

Sentiment Analysis is a Natural Language Processing and Information Extraction task that aims to evaluate writers' feelings expressed in positive or negative comments, questions and requests, by analyzing a large number of documents. Generally speaking, sentiment analysis aims to determine the attitude of a speaker or a writer with respect to some topic or the overall tonality of a document.

A large variety of machine learning models that perform NLP applications in different ways are available in the literature. Recently, machine learning especially deep learning approaches have obtained very high accuracy across many other different NLP tasks. These models can often be trained with a single end-to-end model and do not require traditional feature-specific classification. Though Neural networks have been extensively used in diverse field of research such as Process modeling and control (Sana Bouzaida & Anis Sakly, 2018), Medical Diagnosis (Majhi, 2018), Targeted Marketing (Milton et.al, 2019), Intelligent searching (Olga Tikhomirova, 2020) and various other applications,

these have also attracted many researchers to obtain superior results on various language-related tasks as compared to traditional machine learning models like SVM or logistic regression or Multinomial Naïve Bayes models (Sarkar & Bhowmick, 2017). (Hasan et. al. 2018) has presented a hybrid approach of sentiment analyzer by applying supervised machine-learning algorithms such as Naïve Bayes and support vector machines (SVM).

Indian native languages are being scarce resourced language, proper datasets and sentiment lexicons are not developed enough. Thus sentiment analysis of multilingual tweets or blogs in Indian scenario is very difficult with traditional lexical analysis of supervised models of sentiment polarity analysis. India has a population of 130 cores speaking in about 22 different languages. As per 2011 census in urban India, it has been reported that already 295 million out of 455 million people are using the internet and from the rural areas of India, about 186 million out of 918 million are internet users. India is the second largest online market after China. Accordingly, the reviews of Indian people in their respective language in social media like Face book, Twitter, YouTube, Amazon, Flip kart etc. bear a very important consequence over international promote. Analyzing the polarity of these multilingual reviews in machine learning method has been proven easier than traditional lexical analysis models as it is independent of the grammar of different languages.

Much of the research work on sentiment analysis has been carried out in the English language, but work in native or regional languages has yet not been explored to a great extent. In this paper, an attempt has been made to study the features of different Neural Network algorithms and to analyze their outcomes in achieving best accuracy in minimum time. The datasets from CoRR (Hassan et al., 2016) and also from Twitter API v1.1 for three popular Indian languages, English Bengali and Hindi have been combined. A Simple Neural Network (SNN) model has been designed and tested on the English, Hindi and Bengali datasets of the reviews of Amazon, IMDB Movie, Restaurants and Cricket matches. The same dataset has been further tested for a second model created by Convolutional Neural Network(CNN) and the same dataset has been tested for a third model created by Long Short Term Memory (LSTM)Neural Network and subsequently for the fourth model of LSTM followed by CNN layers. A comparative study of the outputs of the various models used is presented. And the result shows that LSTM followed by CNN achieves better accuracy performance and minimum time complexity in twitter Sentiment Classification than some of traditional method such as the SVM and Naive Bayes methods.

This paper provides a description of related work on multilingual text analysis and details the methodology and comparison of SNN, CNN and LSTM. A later part of the paper explains background discussion about application of Convolutional Neural Network in NLP and also Recurrent Neural Network with help of Long Short term Memory model. The methodology used is depicted by algorithms and the results from different models with around 4000 samples of tweet texts in English, Hindi and in Bengali languages and different size of training batches are furnished.

## 2. RELATED WORK

Till now various machine-learning techniques are used for document level sentiment analysis detection (Pang et al., 2002; Zhao et al., 2008; Dave et al., 2003) while more linguistically motivated approaches are used for more fine grained analysis (Riloff & Wiebe,2003; Wiegand et al., 2010; Narayanan et al., 2009). Researchers have also studied sentiment analysis of review documents in specific genres: blogs (Ku et al., 2006), discussion boards or forums (Kim et al., 2006), user reviews (Pang & Lee, 2004) and expert reviews (Zhu & Zhang, 2010).

(Dattu, 2015) proposed how the sentiment analysis can be used to interpret the reasons of the sentiment change in public opinion, mining and summarizing products reviews, to solve the polarity shift problem by performing dual sentiment analysis by using various techniques such as LDA approach, DSA model, Naïve Bayes (NB) classifier, Support Vector Machine (SVM) algorithm and so on (Sarkar et al., 2015) used a multinomial Naïve Bayes based model has been used for sentiment

analysis of Tweets. (Kumar Akshi and Jaiswal Arunima, 2020) presented systematic literature review presented a Systematic literature review to understand the feasibility, scope and relevance of using Soft computing techniques for sentiment analysis on Twitter.

(Shankar et al., 2016) presented a survey on Sentiment Analysis and Opinion Mining in different languages like Kannada, Hindi and Malayalam to conclude that in case of Indian regional languages though the resources like reviews and opinions are huge, study on these have not been done to a great extent as it is done for English language. in Hindi and Bengali on the dataset released for SAIL TWEET CONTEST 2015. (Chen et al., 2016) investigated three Sentiment Analysis Techniques viz; lexicon-based approach, a machine learning-based approach, and a hybrid approach for Twitter use by governments with their citizens and concluded that while each technique is developed based upon different rationales, the results are statistically robust and comparable.

(Kaur et al., 2017) provides a survey of recent work done in the Hindi language along with a new proposed approach for sentiment analysis of Hinglish (Hindi + English) text.(Gazi et al.,2019) has attempted to provide a detailed summary of Sentiment Analysis of Indian languages with a special focus on code mixed Indian Languages. (Sharma et.al, 2020) have very recently applied sentiment analysis of twitter data to gain the opinion polarity of the voters concerning general elections held in India. R is used for the acquisition, pre-processing, analyzing the tweets and the results obtained are found to be in full compliance with the actual election results held in May, 2019 in India.

(Rejanimol et al., 2017) used a combination of Lexicon Based and Rule-Based Approach for the Sentiment Analysis of Tweets based on movie reviews. They created a dataset is created using twitter posts of movie reviews. He tweets are collected in a file and the sentences are extracted based on punctuation marks. The sense and context of each sentence were determined by using grammatical rules and the tokenize sentences based on word boundaries to match each tokens against tokens in the positive and negative word corpuses. Matching algorithm was developed to find the polarity of words. Final polarity of sentence is drawn based on the count. (Gupta et al., 2017) studied the hybrid machine learning models for sentiment analysis of tweets by considering two machine learning algorithms such as K-Nearest Neighbours (KNN) and Support Vector Machines (SVM) in a hybrid manner. The study concluded that the hybrid model gives better result in terms of classification accuracy and F-measure as compared to using these approaches in isolation.

(Rani et al., 2018) presented a systematic review of sentiment analysis Indian languages emphasizing the taxonomy of Indian Languages in Sentiment Analysis based on techniques, domains, sentiment levels and classes to assist researchers in finding the available resources such as annotated datasets, pre-processing linguistic and lexical resources and choosing the suitable sentiment analysis technique in Indian languages.( Sharma et al., 2018) successfully applied the data mining and machine learning techniques for sentiment analysis. In addition to resource creation, the task of sentiment classification in Hindi Language has also been taken up. The merits and demerits of each of the above approaches across the different genres for the sentiment classification task have also been studied. The problems and the issues with the user-generated content (reviews and news) in Hindi language are also highlighted.

(Soman et al., 2018) had systematically highlighted the challenges associated with Sentiment Analysis of Tweets in English Language versus Indian Regional Languages viz; Tamil, Malayalam, Telugu, Hindi and Bengali and conceptualized in the form of a framework through systematic review.

(Patra et al., 2015) have used machine learning algorithm called Multinomial Naïve Bayes trained using n-gram and Senti Wordnet Features (Trivedi et al., 2011) with an accuracy of 45% using the data sets collected from SAIL MIKE2015.

(Chowdhury & Chowdhury, 2014) have attempted to Analyze Sentiment from Bangla Tweets using Support Vector Machines (SVM) and Maximum Entropy (MaxEnt) on a combination of various features set and compared the performance of these two machine learning algorithms. As a consequence, they obtained a high accuracy of 93% using SVM against a Feature Unigram and Emoticons (Emotion Symbols). However, the performance of the proposed work is constrained by

some dependency issues, such as presence of specific features like emotion symbols and unsuitability for analyzing complex sentences.

(Hassan et.al., 2016) have applied Deep Recurrent Model more specifically Long Short Term Memory on not just standardized Bengali, but English (Bengali words mixed with English words) and Romanized Bengali using to achieve an accuracy of 78%. (Hannah et.al., 2019) based on experiments with three well-known datasets have proven that by employing consecutive convolutional layers on relatively longer texts are more effective in comparison to other state-of-the-art deep learning models.

The sentiment analysis of Bengali tweets using Deep Convolution Neural Network where a sentence is represented as a $n \times k$ matrix with $n$ number of words in the Tweet and the word vector dimension $k$, have been implemented in (Sarkar & Bhowmick,2017; Zhang & Wallace, 2015). The results are found to be better than previous model of Deep Belief Network.

(Socher et al., 2013) introduced Sentiment Treebank comprising of fine-grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences using a recursive neural tensor network (RNTN) for predicting fine-grained sentiment for all nodes in the parse trees of sentences. Sentiment label assigned at the root of the parse tree of a sentence is considered as the label of the sentence. It has been proven that the RNTN performs better than state-of-the art Sentiment Classification Approaches when it is tested on a test set of positive and negative sentences.

## 3. BACKGROUND DISCUSSION

Deep learning as a part of machine learning has received more attention by the researchers for the last few years, especially for Automatic Speech recognition, Face recognition, Human Language processing etc.( Pang et al., 2002) . Two prime Neural Networks that are used to implement deep learning features are Simple Neural Networks, Convolutional Neural Networks and Recurrent Neural Networks.

CNN being a Feed Forward Neural Network on the other hand has proven to be the best in case of image processing as of now. It contains 4 layers, *viz;* Convolution Layer, ReLU Layer, Pooling and Fully Connected Layer, each layer having its own functionality. It performs feature extractions and finds out hidden patterns. CNN performs Text Classification using word embeddings. The convolution process is applied in 2D matrix where each row represents a word. (Kim, 2014) applied filters to slide over full rows of the matrix covering full word. After that, a pooling technique may be applied on the generated feature maps with an activation function. Pooling will result a feature vector for individual feature map. Then all the feature vectors will be concatenated into one big feature vector. At the end, the big feature vector will be fed into a fully-connected layer for regularization and classification.

While the Recurrent Neural Network (RNN) takes two inputs, the present data and the previous state of output for which the next output depends on, it is preferred to work well on sequential data. Recurrent Neural Networks are traditionally difficult to train. The Long Short-Term Memory (LSTM) Neural Network is perhaps the most successful RNN because it overcomes the problems of training a Recurrent Network and in turn has been used on a wide range of applications (Gers et al., 2000; Hochreiter & Schmidhuber, 1997). LSTMs help preserve the error that can be back propagated through time and layers.

## 4. PROPOSED ALGORITHMS AND IMPLEMENTATIONS

In my present work, a simple but efficient for Sentiment Analysis of Multilingual Tweets based on Natural Language Processing (NLP) using different classifier of the neural network model such as Simple Neural Network (SNN), Convolutional Neural Network (CNN), and Long Short Term Memory (LSTM) Neural Network to identify the positive comments and negative comments of and also increased the classification accuracy of the dataset. The positive comments and negative comments of the dataset are selected for sentiment analysis of multilingual tweets by employing different neural

**Figure 1. Flow Diagram of the Implementation Process**
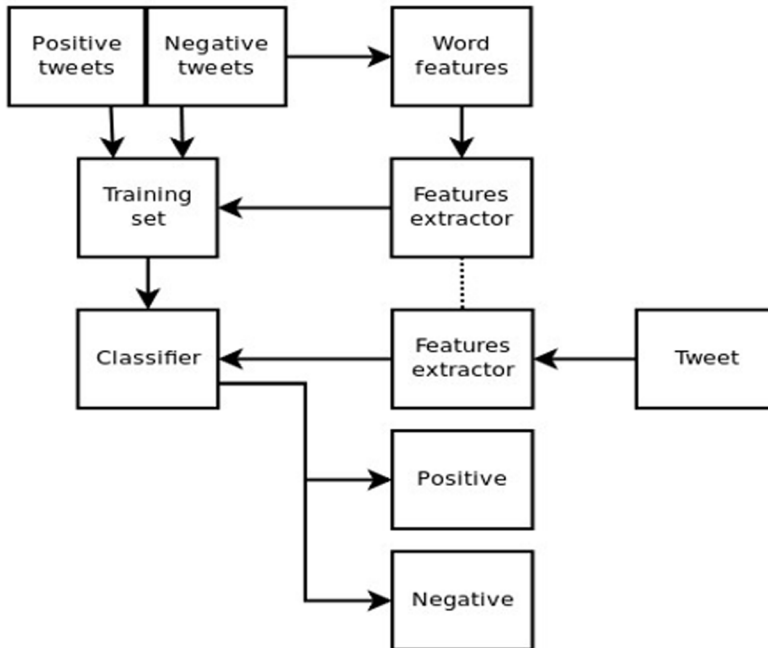


network models. Simple Neural Network (SNN) ,Convolutional Neural Network(CNN), and Long Short Term Memory (LSTM) Neural Network models algorithms for selection and classification of the positive comments and negative comments has been proposed.

A neural network is most important tool for classification. It has various classification methods. In neural networks perform back propagation by activating the neurons in the hidden layer. Neural network model have two phase are (1) training (2) testing. A neural network model involves updating the model weights to create mapping of inputs to outputs using a training dataset. In training phase, of the positive and negative comments of whole dataset are trained and assigned weights. In testing is done on the weighted dictionaries of the dataset. In testing phase we have using the different classifier based on neural network models increase the classification accuracy compares to others classifier .

The following common approach for conducting the experiments and result analysis has been adapted. Figure 1 depicts the flow diagram of the implementation process.

**Step-1:** Preparation of Dataset

The dataset for the present study is a collection of English, Hindi and Bengali Languages Tweets downloaded by the different Twitter API v1.1. The languages filtering of the received posts are supported by different Twitter API. The language parameter in the Twitter Search URL was set to 'eng' to extracted only English, 'ben' to extracted only Bengali tweets and to search Hindi tweets it is set to 'hin'. Also some datasets from amazon review and IMDB movie review (Hassan et al., 2017) are collected and combined with twitter data. Total collected dataset is split into two groups, *viz;* training dataset comprising of 75% of these dataset and test dataset 25% rest.

**4.2 Step-2:** Preprocessing of the Texts:

**2.1 Tokenization: Tokenization** is a process of extracting and splitting text into smaller spices of text called tokens (terms / words) from a corpus. The tokens make it easier to filter out unnecessary tokens such as document into paragraphs or sentences into words and the reviews into words. It has been used to obtain the tokens and their index in the corpus.

**2.2 Normalization**: Identifying punctuation marks, elongated words of both Bengali and English words.

**2.3 Labeling**: The positive tweets are leveled as 1's and negative tweets as 0's and then combine it into single set.

**Step-3:** Build up three similar Neural Network Models: Simple Neural Network (SNN), Convolutional Neural Network (CNN) and Long Short-Term Memory Neural Net work (LSTM).

**Step-4:** Train the Model upon the training dataset and label it.

**Step-5:** The different Neural Network algorithms are compiling by using python compiler

**Step-6:** The data sets are used to obtain a final model fit to the training data set and test dataset and the accuracy of Sentiment Analysis.

**Step-7:** End

The following four models using Neural Networks have been used to carry out the sentiment analysis.

**Model-1:** Using Simple Neural Networks Model Technique (Dattu, 2015).

**Model-2:** Using Convolutional Neural Networks Model Technique (Kim, 2014).

**Model -3:** Using Long Short Term Memory (LSTM) Neural Network (Narayanan et al., 2009)

**Model-4:** Long Short Term Memory (LSTM) Neural Networks followed by another Amalgamated model adding a CNN layer

## 5. EXPERIMENTAL RESULTS

The experiments are carried out by testing the above models with datasets of different sample sizes of English, Hindi and Bengali Languages Tweets downloaded by the different Twitter API v1.1 and some datasets from Amazon review and IMDB movie review given in (Hassan et al., 2017). The preprocessing tasks of tokenization, normalization and labeling of tweets are performed. While the process of tokenization is carried out using the Python's Keras packages, the models for SNN, CNN and LSTM have been built up for testing with training data set by using python compiler. The following output after testing the model with 128 batch size of training data has been found. Table 1 shows the accuracy score and the execution time (in seconds) from creating the model up to validation of test data.

Following the same procedures, a separate experiment has been carried out for Model 4: Long Short Term Memory (LSTM) Neural Networks followed by another Amalgamated model adding a CNN layer and the results are summarized in Table 2 below. It has been observed that the good training accuracy is obtained from the said model irrespective of the amount of sample data sets and Training batch sizes

The progress level of testing and training accuracy and losses for the above Model 4 are shown in the graphs in Figure 2. The x-axis represents no. of epochs and the y-axis represents the amount of accuracy score per epoch.

In addition, the following results are also obtained for the Model 4.

Accuracy: 84.12%
Testing Accuracy: 0.8414
Training Accuracy: 0.4048
Training loss: 0.4972

**Table 1. The variation of accuracy in different models**

| Dataset | Accuracy Obtained | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Model-1 Simple Neural Network (SNN) | | | Model-2 Convolutional Neural Network (CNN) | | | Model-3 Long Short Term Memory Neural Network (LSTM) | | |
| | Test Accuracy | Test Score | Execution Time | Test Accuracy | Test Score | Execution Time | Test Accuracy | Test Score | Execution Time |
| English (1250 samples) | 70.9 | 58.9 | 2.91 | 79.6 | 51.14 | 5.09 | 81.9 | 38.8 | 19.52 |
| Hindi (600 samples) | 75.9 | 66.9 | 1.41 | 74.01 | 63.86 | 3.39 | 74.01 | 58.2 | 11.11 |
| Bengali-1 (700 samples) | 73.5 | 68.6 | 1.65 | 73.05 | 62.9 | 4.51 | 73.1 | 63.4 | 25.08 |
| Bengali-2 (1450 samples) | 72.8 | 67.5 | 1.71 | 73.1 | 59.2 | 5.57 | 72.1 | 59.4 | 28.97 |

**Table 2. Training accuracy in Model-4**

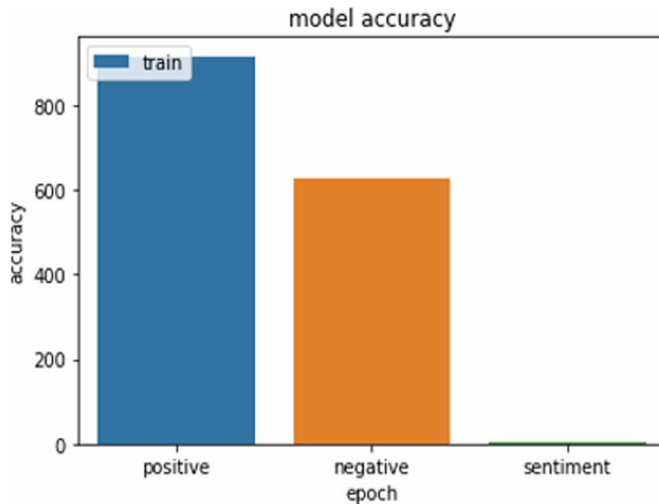| Dataset | Model-4 Long Short Term Memory (LSTM) Neural Network and Convolutional Neural Network( CNN) | | | | | | |
|---|---|---|---|---|---|---|---|
| | Accuracy | Execution Time | Batch size | Testing On (samples) | Validate on (samples) | Training Accuracy | Testing Accuracy |
| English (1250 samples) | 84.1 | 7.18 | 128 | 992 | 258 | 0.4048 | 0.8414 |
| | 78.6 | 7.06 | 256 | 992 | 258 | 0.4747 | 0.7864 |
| Hindi (600 samples) | 73.9 | 5.45 | 128 | 472 | 128 | 0.5716 | 0.7397 |
| | 73.3 | 5.25 | 256 | 472 | 128 | 0.6053 | 0.7331 |
| Bengali-1 (700 samples) | 73.5 | 7.96 | 128 | 539 | 161 | 0.6092 | 0.7353 |
| | 73.1 | 5.65 | 256 | 539 | 161 | 0.6820 | 0.7310 |
| Bengali-2 (1450 samples) | 72.2 | 10.95 | 128 | 1158 | 292 | 0.5903 | 0.7220 |
| | 72.1 | 8.65 | 256 | 1158 | 292 | 0.5898 | 0.7211 |

**Figure 2. Graph representing the classification accuracy and loss in the dataset**
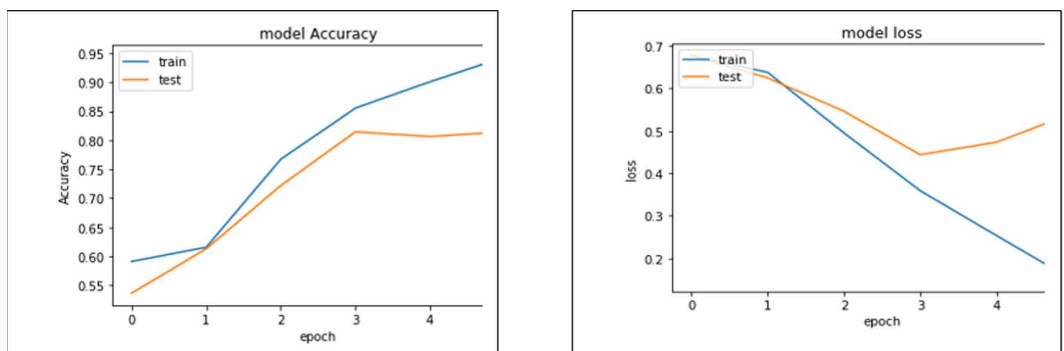
Validation loss: 0.5306
The execution time is 7.18

The accuracy of classification of the number of Positive and Negative Sentences in the data set is shown as bar graph in Figure 3 below.

**Figure 3. Bar Graph representing the accuracy of classification the number of Positive and Negative sentences**



## 6. CONCLUSION

As evident from the result analyses above, the enhancement in the loss of validation is observed for Model-1 Simple Neural Networks (SNN), Model-2 Convolutional Neural Networks (CNN) and Model-3 Long Short Term Memory (LSTM) Neural Network. Amongst all the models, the best accuracy has been achieved through Model-3(LSTM). Though the overall accuracy obtained for Model-4 Long Short Term Memory Neural Network and Convolutional Neural Networks (CNN+LSTM) is the best. The Accuracy of Classification the number of positive and negative sentences in the Sentiment Analysis dataset is 84.1% and with an average time complexity. It is anticipated that better models be designed in the future that gives better accuracy and yet requires minimal execution time to handle the perception capability of humans efficiently.

## REFERENCES

Ahmad, Jimmy, & Nikita. (2019). *Review on Sentiment Analysis of Indian Languages with a Special Focus on Code Mixed Indian Languages*. .10.1109/ICACTM.2019.8776796

Akshi & Arunima. (2020). Systematic literature review of Sentiment Analysis on Twitter using Soft Computing techniques. *Concurrency and Computation: Practice & Experience*. doi:10.1002/cpe.5197

Ankita, Jyotika, & Neha. (2017). Sentiment Analysis of Tweets using Machine Learning Approach. *International Journal of Computer Science and Mobile Computing, 6*(4), 444-458.

Ankita, S., & Udayan, G. (2020). Sentimental Analysis of Twitter Data with respect to General Elections in India. *Procedia Computer Science*, *173*, 325–334.

Balamurali, Aditya, & Bhattachary. (2011). Harnessing WordNet Senses for Supervised Sentiment Classification. *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1081–1091.

Bouzaida & Sakly. (2018). Adaptive Neuro-Fuzzy Sliding Mode Controller. *International Journal of System Dynamics Applications, 7*(2), 34-54.

Bullinaria, J. A. (2013). *Recurrent neural networks, Neural Computation: Lecture*. https://www.cs.bham.ac.uk/~jxb/INC/l12.pdf

Chowdhury, S., & Chowdhury, W. (2014). Performing sentiment analysis in bangla microblog posts. *International Conference on Informatics, Electronics & Vision (ICIEV), 1*, 1–6. doi:10.1109/ICIEV.2014.6850712

Dattu, B. S. (2015). A Survey on Sentiment Analysis on Twitter Data Using Different Techniques. *International Journal of Computer Science and Information Technologies*, *6*(6), 5358–5362.

Dave, K., Lawrence, S., & Pennock, D. M. (2003). Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. *Proceedings of the 12th international conference on World Wide Web*, 519–528. doi:10.1145/775152.775226

Dyuti, S., Prerna, P., Mihika, S., & Kiran, B. (2015). A Proposed Solution for Sentiment Analysis on Tweets to Extract Emotions from Ambiguous Statements. *International Journal of Engineering Research & Technology (Ahmedabad)*, *4*(11), 127–131.

Gers, F. A., Schmidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, *12*(10), 2451–2471. doi:10.1162/089976600300015015 PMID:11032042

Goldberg, A. B., & Zhu, X. (2006). Seeing stars when there aren't many stars: graph-based semi-supervised learning for sentiment categorization. In *Proceedings of TextGraphs: the First Workshop on Graph Based Methods for Natural Language Processing* (pp. 45–52). Association for Computational Linguistics. doi:10.3115/1654758.1654769

Harpreet, K., Veenu, M., & Nidhi, K. (2017). Dictionary based Sentiment Analysis of Hinglish Text. *International Journal of Advanced Research in Computer Science*, *8*(5), 816–822.

Hasan, A., Moin, S., Karim, A., & Shamshirband, S. (2018). Machine Learning-Based Sentiment Analysis for Twitter Accounts. *Journal of Mathematical Computation*, *23*(11), 1–15. doi:10.3390/mca23010011

Hassan, A., Mohammed, N., & Azad, A. K. A. (2016). Sentiment analysis on Bangla and Romanized Bangla text (BRBT) using deep recurrent models. *International Workshop on Computational Intelligence (IWCI)*, 51-56. arXiv:1610.00369v2 [cs.CL]

Herrera, Carvajal-Prieto, Uriona-Maldonado, & Ojeda. (2019). Modeling the Customer Value Generation in the Industry's Supply Chain. *International Journal of System Dynamics Applications, 8*(4), 1-13.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, *9*(8), 735–780. doi:10.1162/neco.1997.9.8.1735 PMID:9377276

Isis, T., & Mohammed-Amine, A. (2017). Natural Language Processing and Fuzzy Tools for Business Processes in a Geolocation context. *Advances in Artificial Intelligence*, 1 – 11. 10.1155/2017/9462457

Jacob, Swaminathan, Anandan, & Kalaivani. (2018). A comparative review of the challenges encountered in sentiment analysis of Indian regional language tweets vs English language tweets. *International Journal of Engineering & Technology, 7*(2.21), 319-322.

Joachims, T. (1999). *Making large scale SVM learning practical*. Advances in Kernel Methods.

Kamal, S. (2017). Sentiment Polarity Detection in Bengali Tweets Using Deep Convolutional Neural Networks. *Journal of Intelligent Systems*, *28*(3), 377–386. https://doi.org/10.1515/jisys-2017-0418

Kamal, S., & Mandira, B. (2017). Sentiment polarity detection in bengali tweets using multinomial Naïve Bayes and support vector machines. *2017 IEEE Calcutta Conference (CALCON),* 31-36.

Kamal & Saikat. (2015). A Sentiment Analysis System for Indian Language Tweets. *Proceedings of the Third International Conference on Mining Intelligence and Knowledge Exploration, 9468*, 694–702. 10.1007/978-3-319-26832-3_66

Kim, H., & Jeong, Y.-S. (2019). Sentiment Classification Using Convolutional Neural Networks. *The Journal of Applied Science*, *9*(11), 1–14. doi:10.3390/app9112347

Kim, J., Chern, G., Feng, D., Shaw, E., & Hovy, E. (2006). Mining and assessing discussions on the web through speech act analysis. *Proceedings of the ISWC'06 Workshop on Web Content Mining with Human Language Technologies*. http://citeseerx.ist.psu.edu/viewdoc/

Kim, Y. (2014). Convolutional neural networks for sentence classification. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 1746–1751. https://www.aclweb.org/anthology/D14-1181.pdf

Ku, L. W., Liang, Y. T., & Chen, H. H. (2006). Opinion Extraction, Summarization and Tracking in News and Blog Corpora. *Proceedings of AAAI Spring Symposium on Computational Approaches to Analyzing Weblogs*. https://www.aaai.org/Papers/Symposia/Spring/2006/SS-06-03/SS06-03-020.pdf

Kumar. (2018). An Efficient Feed Foreword Network Model with Sine Cosine Algorithm for Breast Cancer Classification. *International Journal of System Dynamics Applications, 7*(2), 1-14.

Miao, Q., Li, Q., & Zeng, D. (2010). Fine grained opinion mining by integrating multiple review sources. *Journal of the American Society for Information Science and Technology*, *61*(11), 2288–2299.

Michelle, C. H., & Franks, P. C. (2016). Exploring Government Uses of Social Media through Twitter Sentiment Analysis. *Journal of Digital Information Management*, *14*(5), 290–301. doi:10.6025/jdim/2016/14/5/290-301

Mullen, T., & Collier, N. (2004). Sentiment Analysis using Support Vector Machines with Diverse Information Sources. *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing*, 4, 412-418.

Narayanan, R., Liu, B., & Choudhary, A. (2009). Sentiment analysis of conditional sentences. *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing*, 1, 180-189.

Pang, B., & Lee, L. (2004). A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, 271 – 279. DOI:  doi:10.3115/1218955.1218990

Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*, *2*(1-2), 1–135. doi:10.1561/1500000011

Pang, B., Lee, L., & Vaithyanathan, S. (2002). Thumbs up? Sentiment Classification using Machine Learning Techniques. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2002)*, 79-86.

Patra, B. G., Das, D., Das, A., & Prasath, R. (2015). Shared Task on Sentiment Analysis in Indian Languages (SAIL) Tweets - An Overview. In *Proceedings of the International Conference on Mining Intelligence and Knowledge Exploration*, (pp. 650-655). Springer International Publishing.

Prabowo, R., & Thelwall, M. (2009). Sentiment analysis: A combined approach. *Journal of Informetrics*, *3*(2), 143–157.

Prasath, R., Vuppala, A., & Kathirvalavakumar, T. (Eds.). (2015). Lecture Notes in Computer Science: Vol. 9468. *Mining Intelligence and Knowledge Exploration. MIKE 2015*. Springer. http://doi-org-443.webvpn.fjmu.edu.cn/10.1007/978-3-319-26832-3_61

Razia Sulthana, A., Jaithunbi, A. K., & Sai Ramesh, L. (2018). Sentiment analysis in twitter data using data analytic techniques for Predictive Modeling. In *National Conference on Mathematical Techniques and its Applications (NCMTA 18)*. IOP Publishing. DOI: doi:10.1088/1742-6596/1000/1/012130

Rejanimol, P. R., Sharika, M., & Arun Kumar, N. (2017). Sentimental Analysis of Movie Reviews Based on Twitter Data. *International Journal of Pure and Applied Mathematics*, *114*(11), 137–146.

Riloff, E., & Wiebe, J. (2003). Learning extraction patterns for subjective expressions. In *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing (EMNLP-03)*. Association for Computational Linguistics. DOI: doi:10.3115/1119355.1119369

Shankar, , & Shilpa, , Sridhar, & Suma. (2016). A Survey on Sentimental Analysis in Different Indian Dialects. *International Journal of Advanced Research in Computer and Communication Engineering*, *5*(4), 1072–1076. doi:10.17148/IJARCCE.2016.54262

Sheetal, , Bharti, & Goel. (2018). Sentiment analysis of Indian language. *International Research Journal of Engineering and Technology*, *05*(05), 4251–4253.

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., & Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment Treebank. *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 1642*, 1631–1642. https://www.aclweb.org/anthology/D13-1170

Sujata, R., & Parteek, K. (2018). A journey of Indian languages over sentiment analysis: A systematic review. *Artificial Intelligence Review*, *52*(14), 1–48. doi:10.1007/s10462-018-9670

Tikhomirova. (2020). Entrepreneurial innovative network and the design of socio-economic neural system. *International Journal of System Dynamics Applications, 9*(2), 80-102.

Trivedi, A., Srivastava, A. I., Singh, K., & Gupta, S. K. (2011). Literature Survey on Design and Implementation of Processing Model for Polarity Identification on Textual Data of English. *International Journal of Computer Science Issues, 8*(6), 309-311.

Vilares, Alonso, & Rodr'ıguez. (2017). Supervised Sentiment Analysis in Multilingual Environments. *Information Processing & Management, 53*(3), 595-607. 10.1016/j.ipm.2017.01.004

Wiegand, M., Balahur, A., Roth, B., Klakow, D., & Montoyo, A. (2010). A survey on the role of negation in sentiment analysis. *NeSp-NLP '10: Proceedings of the Workshop on Negation and Speculation in Natural Language Processing*, 60 – 68.

Zhang, Y., & Wallace, B. C. (2015). A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. *Proceedings of the 8th International Joint Conference on Natural Language Processing*, 253–263. arXiv:1510.03820v4 [cs.CL]

Zhao, J., Liu, K., & Wang, G. (2008). Adding redundant features for CRFs based sentence sentiment classification. *Proceedings of 2008 Conference on Empirical Methods in Natural Language Processing*, 117–126. DOI: doi:10.3115/1613715.1613733

Zhu, F., & Zhang, X. (2010). Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics. *Journal of Marketing*, *74*(2), 133–148. https://doi.org/10.1509/jm.74.2

*Abhijit Bera, MCA, M.TECH, working as an Assistant Professor, Department of Computer Science & Engineering (CSE), OmDayal Group of Institutions, Uluberia, Howrah, West Bengal, has experience of academics for last Six years. Research interests are Soft computing, Data Mining, Network & Information Security, Image processing.*

*Mrinal Kanti Ghose, Ph.D, Former Sr. Scientist, ISRO and Retd. Dean (Academics) & Professor in CSE Sikkim Manipal Institute of Technology, Majitar 737136, East Sikkim presently working as Distinguished Professor in CEA Dept, GLA University, Mathua, UP, India. Also served various institutes of repute like RE College (now NIT), Silchar, Assam University, Sikkim University and IIT Kharagpur in academic capacities. Produced 20 PhDs in Computer Engineering and allied fields, published more than 300 research papers in peer reviewed international journals, chaired and organized many conferences and act as reviewer of many reputed journals. Research interests are Geo-Informatics, DIP, Soft computing, Optimization, Simulation and Modeling, Data Mining, Network & Information Security, Reliability and Quality Assurance Analysis. Scopus Author ID: 13310067900; Web Science Researcher ID: F-9733-2018; ORCID ID: 0000-0002-9240-0444.*

*Dibyendu Kumar Pal is presently the Head of the Department -Computer Application, Asansol Engineering College, serving the college since July 2006. He was awarded Ph. D. from Jadavpur University in 2012, with working place at Indian Statistical Institute (I.S.I.) Kolkata. He has more than 20 publications in reputed national & international journals and one book. He is teaching for more than 15 years in different subjects Intelligent Systems, DBMS, Digital Electronics, formal language & automata theory. He completed M.Sc. (Applied Mathematics) Calcutta University, MCA from Vidyasagar University, Ph. D. (Jadavpur University). His research area mainly includes soft computing & flow visualization using image processing technique.*