JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Sentiment Polarity Detection using Machine Learning and Deep Learning.

Dr. P. Sumalatha[1]
[1]Asst.Prof., Dept. of Artificial Intelligence and Data Science.
Central University of Andhra Pradesh, Anantapur.
Tammineani Sarath Chowdary[2]
[2]P.G. Student., Dept. of Artificial Intelligence and Data Science.
Central University of Andhra Pradesh, Anantapur.

Abstract:

As e-commerce has grown in recent years, so online shopping has increased with the number of product reviews posted online. The consumer's recommendations or complaints influence significantly customers and their decision to purchase. Sentiment polarity analysis is the interpretationand classification of text-based data. The main goal of our work is to categorize each customer's review into a class that represents its quality (positive or negative). Our sentiment polarity detection consists of the following steps: preprocessing, feature extraction, training, classification, and generalization. First, the reviews were transformed into vector representation using different techniques of Tf-Idf and Tokenizer. Then, we trained with a machine learning model of SVM Linear, RBF, Sigmoid kernel, and a deep learning model LSTM. After that, we evaluated the model's using accuracy, f1-score, precision, and recall. Our LSTM model predicts an accuracy of 86% for Amazon-based customer reviews and an accuracy of 85% for Yelp customer reviews.

Keywords: Sentiment Analysis, Sentiment Polarity, Opinion Mining, Natural Language Processing (NLP), Machine Learning, Text Classification, Emotion Detection, Sentiment Classification, Text Analysis, Lexicon-based Sentiment Analysis, Deep Learning, Supervised Learning.

1. INTRODUCTION

Purchasing something online has become a regular practice for millions of individuals all around the world. The number of people purchasing goods and services online has recently expanded more than ever before. One of the reasons why online shopping has developed so rapidly over the years is the experience that businesses can offer their customers. 2.14 billion Individuals worldwide are anticipated to make online purchases in 2021. The prediction for worldwide e-commerce revenues is

$4.891 trillion at the same time. If these online purchasing statistics are not mind-blowing enough, forecasts indicate that global e-commerce sales will increase to $6.4 trillion by 2024 [1]. Online purchasing is increasingly popular for a variety of reasons, including ease and affordable rates. Salesmen typically try to persuade customers to purchase products in physical locations. You can conduct as you desire while shopping online.

Additionally, online retailers are making every effort to ensure that the online purchasing experienceis comparable to in-person shopping experiences. Online buying provides a huge range of possibilitiesthat are not available when shopping in person. You can look through many websites and select the merchandise based on your needs. Although you have the option of paying with cash on delivery,

online buying does not require you to physically carry currency. Instead, payments can be made with debit or credit cards. Online buying makes up 63% of all shopping occasions [2]. The number of people shopping online, especially for groceries and daily necessities has surged because of the severelockdown measures that are being implemented by nations throughout the world to stop the virus' spread. During the Covid-19 pandemic period, it was not possible to go to the market for shopping, so people were inclined to buy things online. Because of Covid-19, 42% of US consumers purchased goods online in March 2020, nearly double the rate of 22% in 2018 [3].

2. Related Work

For the last few years, some research has been conducted to predict sentiment polarity detection using machine learning and deep learning on customer reviews. Tanjim Ul Haque, Nudrat Nawal Saber, and Faisal Muhammad Shah have shown they used cross-validation methods and found that a 10-fold increase in accuracy was the best for Linear SVM [4]. They use the best classifier for three different types of product reviews. Aiming to improve all extraction methods and preprocessing steps, they chose the most accurate one for their research. It was shown that all datasets had the best outcomes when the common features from TF-IDF and the bag of words were used in the feature selection procedure. Using a support vector machine is a better option because the dataset is huge and it doesn'thave to be overfitted. According to these findings, the highest level of accuracy was 94.02%. In this study, Sanjay Dey, Sarhan Wasif, Dhiman Sikder Tonmoy, Subrina Sultana, Jayjeet Sarkar, and Monisha Dey applied two machine learning models: Support Vector Machine (SVM), NB [5]. This paper represents a comparison between two machine learning approaches for analyzing the sentiment of customer reviews on Amazon products. In this work, their models were trained by almost 2250 features with almost 6000 datasets after the preprocessing procedure. In the meantime, almost 4000 test sets have been passed through the models for statistical measurement. The system provides a precision of 82.85%, recall of 82.88%, an accuracy of 84%, and an f1-score of 82.662% for the SVM classifier. Atiqur Rahman and Md. Sharif Hossen used the SVM method, the dataset contains 2000 movie reviews where 1000 are negative and the remaining is positive [6]. According to this paper SVM model accuracy is 87.33%, precision is 85.90%, recall is 89.33% and f1-score is 87.58%. Arwa S. M. Al Qahtani analyzes the dataset of Amazon reviews and investigates the classification of the sentiment using various classification models [7]. The used classifiers include BERT, NB, BiLSTM, RF and Logistic Regression. With 94% and 98% accuracy in binary and multiclass classification, the BERT model has shown good results which gave a precision of 98.4%, recall of 98.4%, and f1-score of 98.4%.

The author also used RF with Glove. The RF with Glove displays 90% accuracy. In this paper, Akanksha Halde, Aditi Uttekar, and Amit Vishwakarma also used the BERT model along with RF, NB, and SVM [8]. The BERT classifier is the most accurate in estimating the sentiment of a review with a precision of around 90% after numerous model iterations and tests. Naveen Kumar Gondhi, Chaahat, Eishita Sharma, Amal H. Alharbi, Rohit Verma, and Mohd Asif Shah used the LSTM, CNN, SCA, NBmodels for sentiment analysis [9]. After around 10 epochs of training, they calculated their model's training and validation loss and accuracy. It was established that 0.78 was the ultimate threshold for classifying sentiment, based on the ROC curve. They used the f1-score as the best indicator of the model's effectiveness because the dataset was imbalanced. Their research goal was to test the models functionality with a large amount of dataset. Roobaea Alroobaea proposed long short-term memory (LSTM), Gated Recurrent Unit (GRU), and Convolutional Neural Network (CNN), RNN models[10]. The author has used Arabic datasets and tools to analyze sentiment and showed the result of multiple deep-learning models on three different datasets. He used balanced datasets and models have predicted well for large datasets than shorter datasets. He compared it with Palsah's BiLstm method and showed his RNN model accuracy which was better for three datasets. Shivaprasad T K and Jyothi Shetty have reviewed many papers of other researchers [11]. They tried to give a review of different classifier methods on sentiment analysis of different product reviews. They have explained many sophisticated methods that define sentiment analysis. They have shown that for different datasets SVM classifier gives much more accuracy than NB and Max Entropy. Zeenia Singla, Sukhchandan Randhawa, and Sushma Jain have used SVM, NB, and DT classifier models to predict the sentiment of over 400000 reviews in positive or negative [12]. They have used an inbuilt Syuzhet package to conduct sentiment analysis. They evaluated their models with 10-fold cross-validation and got the highest accuracy for the SVM model which was 81.77%.

3. LITERATURE SURVEY ANALYSIS

Although most of the research papers on SA present approaches to SA of English texts, work on otherlanguages is still growing. There are some previous research works that have dealt with SA of non-English texts such as German, Chinese, Arabic, Bengali, and Hindi. (1) determining each sentence's sentiment based on word dependency, and (2) aggregating sentences to predict the document sentiment. Wan (2009) made use of bilingual knowledge including both Chinese resources and English resources for sentiment polarity identification of Chinese product reviews. In this work, Wan (2009) translated a corpus of Chinese documents into English by using Google Translate and Yahoo Babel Fish. In addition, he applied ensemble methods to combine the individual results over Chinese and English datasets. The results for the combination methods improved the performance of individual results.

Ghorbel and Jacot [13] used a corpus with movie reviews in French. They applied a supervised classification combined with SentiWordNet in order to determine the polarity of the reviews.

Rushdi-Saleh et al. presented a corpus of movies reviews in Arabic annotated with polarity and performed several experiments using machine learning techniques. Al-Ayyoub et al. [1] presented Lexicon-based SA of Arabic tweets.

Banea et al. [3] proposed several approaches to cross-lingual subjectivity analysis by directly applyingthe translations of the opinion corpus in English to train an opinion classifier in Romanian and Spanish. This work showed that automatic translation is a viable alternative for the construction of resources and tools for subjectivity analysis in a new target language. A comparative study of NLP techniques applied to SA and topic detection of Spanish tweets has been presented in [2]. Martínez- Cámara et al. [23] applied the supervised approach to SA of Spanish movie reviews using different machine learning algorithms such as SVM, naive Bayes, Bayesian-based logistic regression, K-nearestneighbor (KNN), and C4.5.

Das and Bandyopadhyay [9] presented an approach to opinion polarity classification on news texts forBengali using SVM. Their system identifies the semantic orientation of an opinionated phrase as either positive or negative. They also developed a subjectivity classifier for performing sentence-level subjectivity classification. Chowdhury and Chowdhury [6] used a semi-supervised bootstrapping approach for the development of the training corpus which avoids the need for labor-intensive manual annotation. For classification, they used SVM and maximum entropy with a combination of various sets of features. SA of Bengali tweets using multinomial NB with n-gram and SentiWordNet features have been presented in. An approach presented uses a more sophisticated system that uses an unsupervised approach for expanding a (small) Indian sentiment lexicon, leveraging distributional thesauri, sentence-level co-occurrence statistics, and SVM.

Joshi et al. [15] presented a study on SA for Hindi. They studied three approaches to perform SA in Hindi: the first approach involves training a classifier on this annotated Hindi corpus and using it to classify a new Hindi document, the second approach translates the given document into English and uses a classifier trained on standard English movie reviews to classify the document and third approach uses a lexical resource called Hindi-SentiWordNet and implements a majority-score-based strategy to classify the given document. They proposed a fallback strategy for SA in Hindi which applies any one of the above-stated three strategies based on the availability of in-language training data or a translation system from Hindi to a resource-rich language or SentiWordNet. Sharma et al. presented unsupervised lexicon-based SA of Hindi tweets. Mittal et al. presented SA of Hindi review based on a negation and discourse relation.

4. EXISTING APPROCHES

In the present day, social media is a popular technology that uses microblogging platforms to connect millions of people. People can freely express their thoughts, ideas, and views as short messages called tweets on many micro-blogging platforms in social networks (like Twitter) and business websites or web forums [1]. Researchers gather these unstructured tweets and use a variety ofmethods to extract information from them.

This analysis of tweets or opinions provides predictions or measures in a variety of application domains such as business, government, education, sports, tourism, biomedicine, and telecommunication services [2]. Sentiment analysis or opinion mining is the study of opinions and prediction. Sentiment analysis (SA) is one of the text mining approaches that use natural language processing for binary text classification. Sentiment analysis can be performed in four levels based on the scope of the text. They are document-level, sentence-level, aspect-level, and word-level sentiment analysis [6]. In Document level SA, the overall opinion of the document about the single entity is grouped into positive or negative. In sentence level SA, the opinion expressed in a sentence is classified as either positive or negative. In aspect level SA, opinions about entities are grouped based on specific entity elements. Atword level SA, opinions about entities are grouped based on specific words. In the proposed work, the word-level sentiment analysis model is developed for a restaurant review data set based on machine-learning and deep-learning algorithms to classify sentiments as positive or negative automatically. Thesections in this paper are organized as follows: Section 2 examines the various sentiment analysis models that have been discussed in the literature. Section 3 discusses the various machine learning techniques. Section 4 analyzes the performance of machine learning techniques used in sentimentanalysis and finally discusses the conclusion and future work that needs to be carried out in sentiment analysis.

5. PROPOSED METHOD

After analyzing different research papers described in section 2, it is clear that deep learning-based text classification using NLP is a popular topic nowadays. Identifying positive and negative emotions from social media is a very popular and challenging topic for research. As many people share their emotions through social media, it can be easy to collect huge data forresearch. In this paper, we have proposed some models to classify emotions from tweets of Twitter users. For this purpose, data preprocessing is an essential step after collecting the dataset. We have proposed both baseline models (traditional machine learning models) and deep learning models (Stacked LSTM, Stacked LSTM with 1D convolution, CNN with pre-trained word embedding, and a BERT-based model). For baseline models, tf-idf(term frequency-inverse document frequency) and count vector features have been used for Multinomial Naive Bayes, Support Vector Machine(SVM), and Logistic Regression separately as input.
tf-idf is a weighting scheme that assigns each term in a document a weight based on its term frequency (tf) and inverse document frequency (idf). The terms with higher weight scores areconsidered to be more important.

Count Vectorizer works on Terms of Frequency, i.e. counting the occurrences of tokens and building a sparse matrix of documents x tokens.

For all deep learning models, epoch size has remained the same and it is 3 because of thelarge size of the dataset. But batch size, dropout probability, activation, and optimization have varied from model to model.

The main target of this proposed method is to obtain improved accuracy in identifying emotions. We have trained all the models mentioned in Fig 1 and found the accuracy. At last, the best-performed model has been found among them after comparing the accuracy results. Thedetails of these models have been explained in section 4.

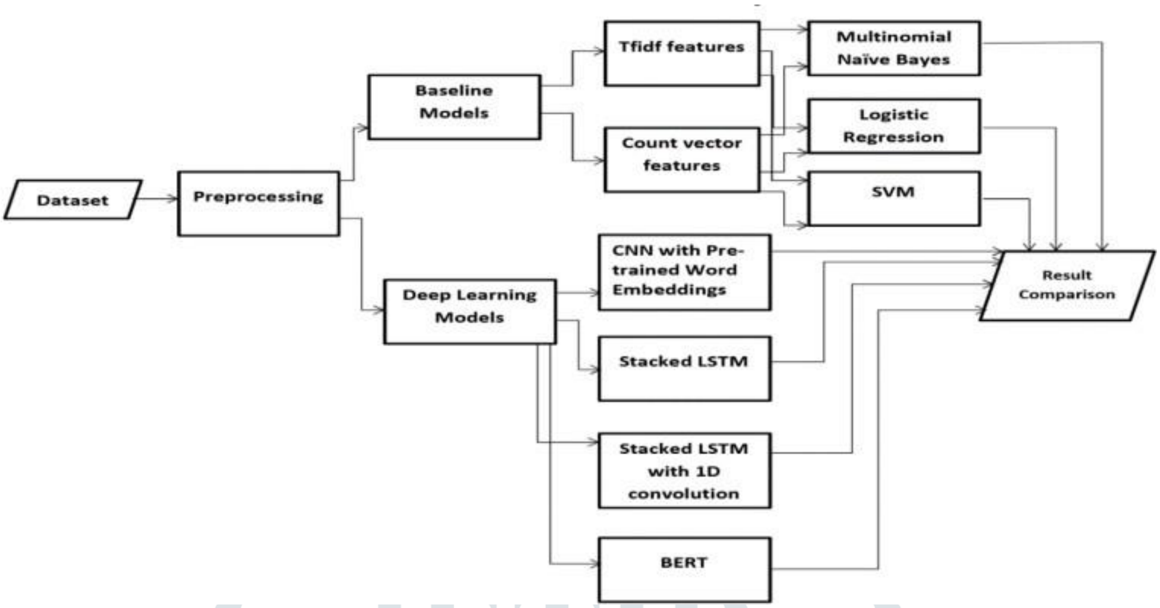

Fig 1: Proposed Methodology

6. RESULT

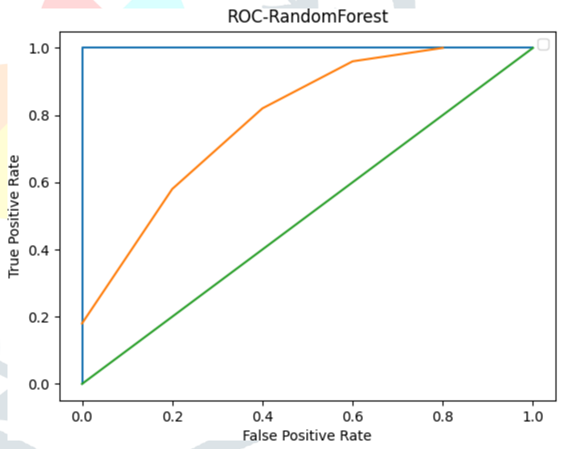

Figure 6f: ROC curve of Random Forest

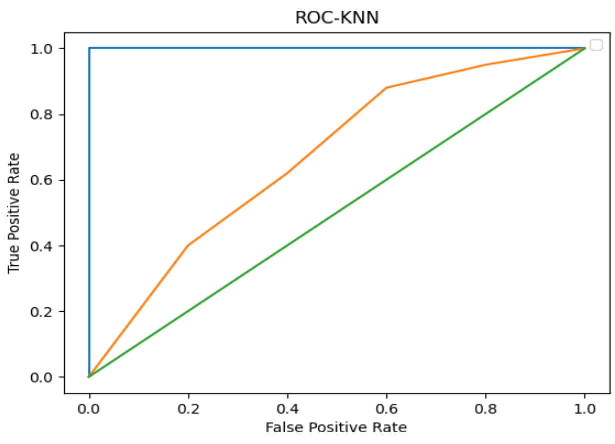

Figure 6e: ROC curve of KNN

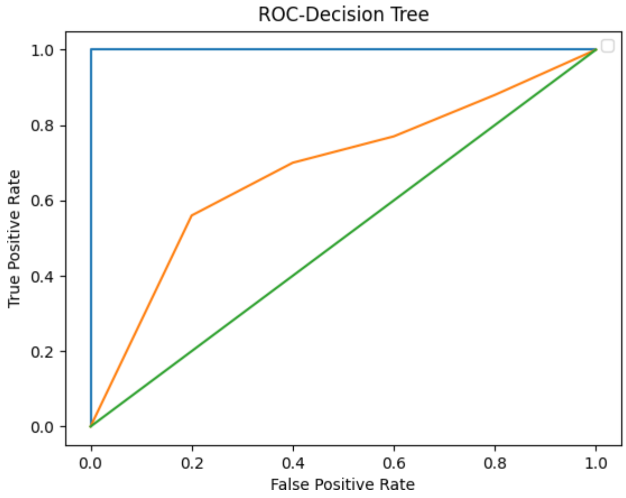

Figure 6d: ROC curve of Decision Trees

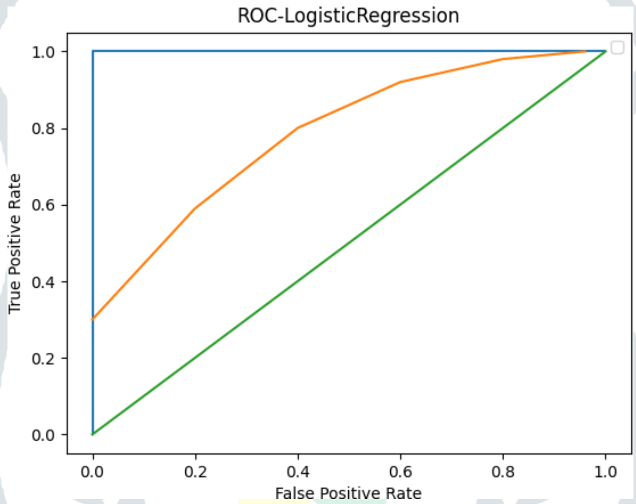

Figure 6c: ROC curve of Logistic Regression

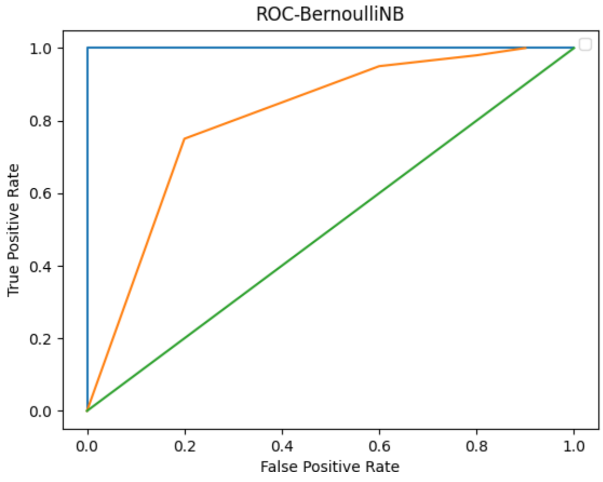

Figure 6b: ROC curve of Bernoulli Naïve Bayes

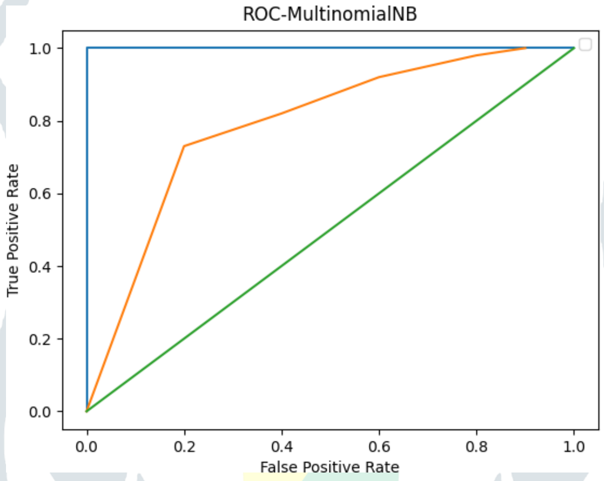

Figure 6a: ROC curve of Multinomial Naïve Bayes

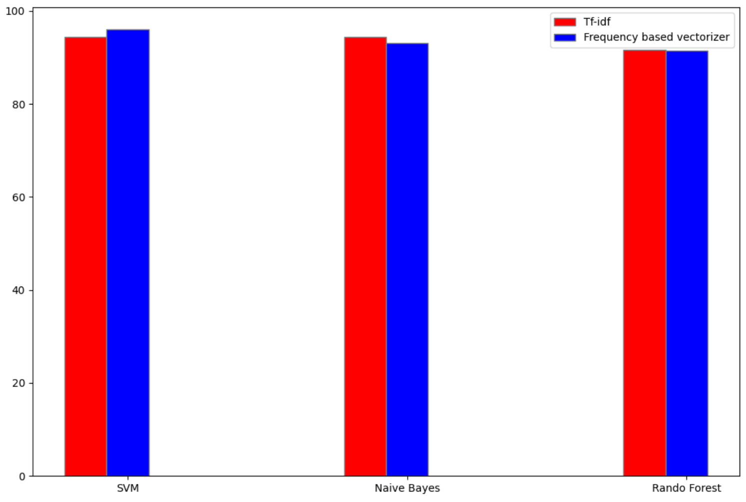

Figure 7: Accuracy of SVM, Naive Bayes, random forest classifiers with vectorizationmethods

Figure 1 demonstrates the accuracy results obtained from the implementation of tf-idf and frequency-based vectorization methods with each classification algorithm. As can be seen from the figure, the combination of SVM and tf-idf outperforms other classifiers with the 96 percent accuracy. The minimum accuracy was obtained from the combination of random forest classifier and tf-idf vectorizer with 91.4 percent accuracy. While considering accuracy result it is observable that accuracy range changes between 91 percent and 96 percent. Random forest classifier does not offer huge difference between two vectorization methods asthere is only 0.2 percent accuracy difference between tf-idf and frequency-based vectorization methods. When analyzing Naive Bayes classifier, the best accuracy we got was
94.4 percent and was obtained from the frequency based vectorizer.

Feature extraction	n-grams	F1-Score	Precision	Recall
Frequency based vectorizer	Unigram	95.51	95.4	95.63
	Bigram	95.45	94.17	96.76
	Trigram	92.82	88.73	97.31
Tf-idf	Unigram	96.79	96.48	97.1
	Bigram	95.9	94.45	97.41
	Trigram	93.35	89.19	97.93

Table 3: Linear SVM Result

Table 3 depicts the F1, recall, and precision results for the SVM classifier using various feature selection and n-gram models. As can be seen from the table the best F1 score we got 96.79 percent was from the combination of tf-idf vectorizer and unigram model. While considering frequency-based vectorizer the highest F1 score which is 95.51 percent was obtained from the unigram model. When comparing the F1 score of unigram and bigram models we do not observe a huge difference between results. Recall values for the SVM classifier ranged from

95.63 percent to 97.93 percent. The highest recall score was 97.93 which was obtained from the combination of tf-idf and trigram model. As described in the table the lowest F1 score 92.82 percent was gained from the trigram model and frequency-based vectorization.

7. CONCLUSION

After evaluating the proposed model, we have seen that the highest accuracy 86% is shown by the LSTM model for the Amazon test dataset. RBF, Linear, and Sigmoid kernel of the SVM model have shown the same accuracy for the Amazon test dataset which is 84%. For the Yelp test dataset, we havegotten the highest accuracy from the LSTM model. For both test datasets, LSTM model performance parameters are higher than SVM models.

We have seen that the performance parameters of our models are improved when we hyper-tune the models more. In the future, we shall increase our dataset size and will improve the model performanceparameters. We will use the BERT model. We will research more on feature extraction of the dataset and will use different lemmatizers and model parameters to improve the system.

In this paper, we have presented an approach to sentiment detection of Bengali tweets using deep convolutional neural networks. We have compared the performance of our proposed CNN-based Bengali sentiment polarity detection model with the DBN-based model. One of the critical problems that we faced for our experimentations was a scarcity of benchmark datasets for SA in Indian languages. Although we have used the datasets released for the SAIL contest 2015, it is insufficient and noisy. In the future, we have planned to rectify errors in the training data and create my data for developing a more accurate sentiment classification model for Bengali tweets.

Another plan is the exploitation of unlabeled data in the training process of DBN- as well as CNN-based models. For improving system performance, some other kinds of deep neural networks such as deep recursive neural networks can also be used.

The model proposed in this paper can be easily extended to other Indian languages such as Hindi and Tamil if training data for the corresponding language are available.

REFERENCES:

1. “Online shopping statistics,” last accessed 18 August,2022. [Online].Available: https://www.oberlo.com/blog/online-shopping-statistics
2. “Online shopping statistics second,” last accessed 18 August 2022. [Online]. Available: https://www.oberlo.com/blog/online-shopping- statistics
3. “Coronavirus effect on online purchasing,” last accessed 18 August 2022.[Online].Available: https://www.nytimes.com/2020/04/05/technology/coronavirus-amazon- workers.html

4. T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," IEEE International Conference on Innovative Research and Development, 2018.

5. S. Dey, S. Wasif, D. S. Tonmoy, S. Sultana, J. Sarkar, and M. Dey, "A comparative study of support vector machine and naive Bayes classifier for sentiment analysis on Amazon product reviews," International Conference on Contemporary Computing and Applications (IC3A), 2020.

6. A. Rahman and M. S. Hossen, "Sentiment analysis on movie review data using machine learning approach," International Conference on Bangla Speech and Language Processing (ICBSLP), 2019.

7. A. S. M. AlQahtani, "Product sentiment analysis for amazon reviews," International Journal of Computer Science and Information Technology, 2021.

8. A. Halde, A. Uttekar, and A. Vishwakarma, "Sentiment analysis on ama- zon product reviews," International Research Journal of Modernization in Engineering Technology and Science, 2022.

9. N. K. Gondhi, Chaahat, E. Sharma, A. H. Alharbi, R. Verma, and M. A. Shah, "Efficient long short-term memory-based sentiment analysis of e-commerce reviews," Hindawi Computational Intelligence and Neuro- science, 2022.

10. R. Alroobaea, "Sentiment analysis on amazon product reviews using the recurrent neural network (rnn)," International Journal of Advanced Computer Science and Applications, 2022.

11. S. T. K and J. Shetty, "Sentiment analysis of product reviews: A review," International Conference on Inventive Communication and Computational Technologies (ICICCT 2017), 2017.

12. Z. Singla, S. Randhawa, and S. Jain, "Sentiment analysis of customer product reviews using machine learning," 2017 International Conference on Intelligent Computing and Control, 2017.

13. M. Al-Ayyoub, S. B. Essa and I. Alsmadi, Lexicon-based sentiment analysis of Arabic tweets, Int. J. Soc. Netw. Min.2 (2015), 101–114.10.1504/IJSNM.2015.072280Search in Google Scholar

14. A. F. Anta, L. N. Chiroque, P. Morere and A. Santos, Sentiment analysis and topic detection of Spanish tweets: a comparative study of NLP techniques, Procesamiento del Lenguaje Natural50 (2013), 45–52.Search in Google Scholar

15. C. Banea, R. Mihalcea, J. Wiebe and S. Hassan, Multilingual subjectivity analysis using machine translation, in: EMNLP, pp. 127–135, ACL, Honolulu, Hawaii, 2008.