Natural Language Processing and Bi-Directional LSTM for Sentiment Analysis

Mohamed Muftah Abubaera
Computer Department
Elmergib University

Salman Mohammed Jiddah
Computer Engineering Department
Cyprus International University

ABSTRACT
In e-commerce, one of the most critical and important aspects of the business model is customer reviews. Customer reviews reflect the satisfaction of customers with respect to the products and services offered. E-commerce is driven by significant amounts of data which poses a huge challenge of collection and evaluation to have an insight before decision-making and business strategy implementations. The field of natural language processing and machine learning techniques have provided significant leaps in helping the analysis of big data and business analytics. Also, Recurrent Neural Networks (RNN) evolved in so many powerful algorithms and one of those is the Bi-LSTM variation of RNNs. Bi-LSTM has been identified in the literature as a suitable machine learning classification algorithm for natural language processing due to its sequential learning process. This study is an implementation of the lemmatization natural language processing technique coupled with the Bi-LSTM machine learning classification technique for customer review sentiment analysis. The application of these two techniques has reported a significant performance accuracy in sentiment analysis of customer review data. The results in this study are reported as 96.06%, 91%, and 90% for accuracy, precision, and recall respectively.

Keywords
Amazon customer reviews, Bi-LSTM, deep learning, natural language processing.

1. INTRODUCTION
Language is a naturally occurring phenomenon for humans, humans use verbal communications methods to relay both meaning and intentions using language. Language to humans is also a method of emotional expression through the messages being communicated. The expression of emotions using language as a tool has been identified to affect and impact both listener and speaker of the emotions [17]. The expression of language is primarily done using words, and the transmission and decoding of emotions using language is a conscious cognitive process carried out by both listeners and speakers, otherwise, the emotional messaging in languages will be inadequately understood, unfortunately, there is no specific empirical proof on the human cognitive process of constructing and deconstructing language messaging, and hence the existing challenges of language and emotion processing research [15].

Emotional communication has been presumed by researchers to be a process requiring emotional intelligence in communication, sentiment analysis is the process of using computational models based on machine learning techniques for the decoding of intended emotions in messages [10]. Computational models used for sentiment analysis aid in gaining insight into messages in the same manner humans cognitively would. Sentiment analysis has gained a lot of research attention and has become one of the most established dimensions of machine learning due to the high increase in data generation [21].

2. STUDY BACKGROUND
Amazon is arguably one of the largest e-commerce platforms in the world, it is a place where people can buy and sell products. Amazon also has product review options where customers can leave reviews of the products they have purchased on the platform [5]. According to [18], customer product review data are both critical and important in the purchase decisions of customers, such data can also be used by the business organization to enable them to carry out good business development decisions based on the sentiments communicated by customers on a product or service. A study carried out by [6] reported that around 91% of customers using e-commerce platforms engage in purchase decisions after reading reviews about products and or services. Product reviews play significant roles to both customers and business organizations in purchase decisions and business strategy decisions respectively.

2.1 Recurrent Neural Networks (RNN)
Recurrent Neural Networks (RNNs) are a category of artificial neural networks that have been popularly used in the field of handwriting and speech recognition [19]. RNNs are models which have similar node connections in comparison to the human network of connected neurons in the brain. The connected nodes in RNNs use a sequence in a directed graph where each node in the network has its input from the backward nodes and the outputs from forward nodes in the network. The nodes in RNNs use weighted edges for connection which are modified at the final output where the strength of the signals is passed through. A peculiar feature of RNNs is their backpropagation characteristics which are absent in feedforward neural networks [1].

Conventional RNNs have however been reported to have challenges of exploding gradients and vanishing which also implies challenges and difficulty in training, gradients sometimes shrink and become so small which then stops the learning, or in some cases, the gradients become enlarged which then makes the weights exceed the maximum limits [12]. These challenges have been mitigated by the addition of gating mechanisms in RNNs, which have been seen in Long Short-Term Memory (LSTM) models [9] and Gated Recurrent Unit [4].

2.2 Long Short-Term Memory (LSTM
[9] developed LSTM as a unit, LSTM has seen several changes and modifications from its original state, this includes the addition of the forget gate. LSTMS are generally regarded as complex models but have been very successful in applications of sentiment analysis, machine translation and handwriting

recognition analysis. LSTMs are particularly very efficient in long-term dependent characterized problems which require consistent sequential training such as sentiment analysis and handwriting recognition analysis [16]. LSTM uses repeating modules that have four distinct interacting layers which are referred to as gates in the LSTM architecture illustrated in Figure 1, the gates are as follows [14]:

- Forget gate: this is the memory gate of the LSTM architecture, this gate is used for long-term memory storage of information which is considered not useful for training.
- Learn gate: this is the input module that is combined with the short-term memory of the LSTM model to apply recently learned features to the current learning phase of the LSTM model.
- Remember gate: this is the gate used for information gain with regards to all information not parsed into the forget gate module. The remember gate combines the short-term memory and current training events to update the forget gate.
- Use gate: this gate combines both short-term and long-term memories to predict outputs after which it uses new information to update the short-term memory

2.3 Bi-Directional Long Short-Term Memory (Bi-LSTM)

Bi-LSTM is a modified LSTM that is capable of recognizing contextual features from contextual features learned from both previous and future states [14]. Bi-LSTM uses hidden and forward layers to calculate and generate output sequences, the use of Bi-LSTM implies an input is run in two ways; one in a backward direction and the other in a forward direction. The bi-directional characteristics of the Bi-LSTM give it an edge over LSTM which is unidirectional because it enables it to store in memory any information from both the backward direction and forward direction to use in the gate of the model.

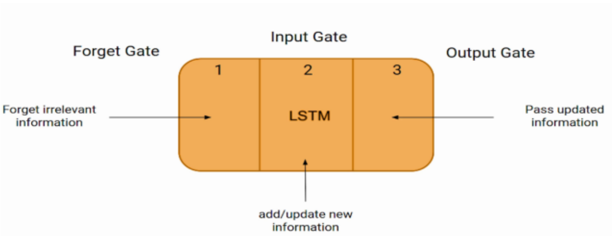

Fig 1: Simplified LSTM architecture

3. STUDY OBJECTIVES

The most important tools of sentiment analysis have seen an increasing technological advancements and maturity in recent times; data mining and machine learning [8]. Customer review data are more accessible to people now more than they have ever been, people have access to review data about products and services with regard to pricing, value, quality, and other such variables. This research is aimed at carrying out sentiment analysis using mined Amazon customer review data with the following subsets of objectives:

- Sentiment dataset analysis
- Using natural language processing techniques for data preprocessing
- Machine learning classification of customer sentiments

This research contributes to sentiment analysis literature by showing the efficiency of applying natural language processing techniques can impact already powerful machine learning and deep learning models in the classification of customer reviews and sentiment data.

4. RELATED WORKS

There have been numerous studies and experiments carried out in product review and sentiment analysis in recent years, the use of machine learning has contributed immensely in natural language processing and sentiment analysis [3]. Earlier research on sentiment analysis of customer product reviews has been reported to be more inlined into finding relationships between the sentiment data syntactic factors and the semantic data [23]. Recent researchers have however been more inclined into seeking to explore the relationships between the service or products and the given customer feedback, a study reported by [13] carried out movie reviews using IMDB review dataset while applying a CNN classification methodology, their methodology used a CNN 1 dimensional text data classification which is then reordered to suit the CNN layers and then retrained. The reported method can detect patterns in the review data text which carries out adjustments of the CNN kernel size to suit the output size. Some peculiar words and phrasing have been associated with sentiments with particular kinds of reviews, such words have also been reported to be utilized in machine learning sentiment analysis studies, according to [5] words like *"dislike", "really nice", and "i dont like"* can be tracked using machine language processing. The use of specific words and word processing from text data for sentiment analysis is reported in the work of [22], they carried out the categorization of subjects to enable pattern trending in their machine learning algorithm training.

A study by [8] used a combination of BagofWords text processing techniques and the application of Multilayer Perceptrons (MLP), they reported a performance accuracy of 92% performance on the Amazon sentiment analysis dataset. The same dataset was experimented on by [7] using a methodology of Linear Support Vector Machine (LSVM) and the application of BagofWords text processing techniques, their own study reported a performance accuracy of 91.72%. Another study carried out on the Amazon sentiment analysis dataset is the work of [20], in their study they carried out their machine learning training on a Gated Recurrent Unit (GRU) of a RNN algorithm, they reported a performance accuracy of 81.82% without a reported application of NLP technique in their experiments.

5. METHODOLOGY

This study carries out a customer review sentiment analysis on the Amazon Review Polarity Dataset (ARPD). The methodology of this research is designed to apply NLP techniques which literature has shown the increased performance of machine learning training of text data. The chosen NLP techniques of this study are lemmatization and stemming, this processed data is then trained using a Bi-directional LSTM RNN machine learning architecture. Figure 2 shows an experimental flowchart of the research methodology.

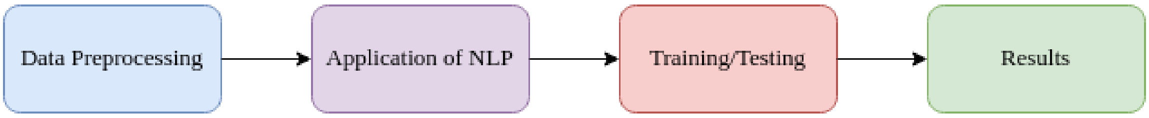

Fig 2: Research methodology flowchart

5.1 Bi-Directional LSTM

This study uses an RNN architecture machine learning algorithm in the form of a modified LSTM known and referred to as a Bi-LSTM. This form of LSTM is chosen because of its bidirectional learning advantage as previously explained in the LSTM literature review section. The architecture of the implemented Bi-LSTM in this study is designed as a Bi-LSTM architecture padded with a 1-dimension convolution layer and a dense layer that is flattened at the output as shown in Figure 3.

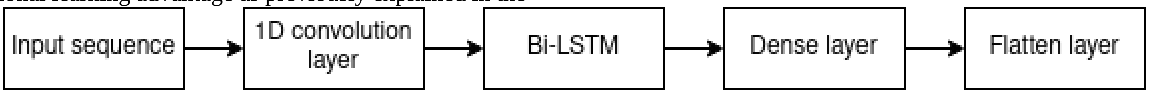

Fig 3: Proposed Bi-LSTM architecture of the study

5.2 Dataset

The chosen dataset for our study is the Amazon Review Polarity Dataset (ARPD), ARPD is a text-based amazon product customer review dataset collected by the works of [23], and has been annually updated since its creation. The dataset contains a collection of about 34 million customer product reviews for over 2 million products. The labeling of the dataset is carried out using the Stanford Network Analysis Project (SNAP), where a rating of 5 points is used to label a sentiment as either negative or positive, product reviews with a point of 3 and above are labeled positive and reviews with points below 3 are labeled negative. Due to the computing cost associated with high-intensity data processing, this study is limited to the use of only a segment of this dataset; 2 million product reviews. The segment of the data was used for training and testing using the 80:20 training and testing data ratio [11].

5.3 Data preprocessing

Like most datasets, the ARPD dataset also has some noisy and redundant data which require preprocessing to enable optimal training and testing of machine learning models. Some of the customer review data contain repeating words, use of non-English language, use of emoticons in reviews for emotional expression, and even null data. Data preprocessing was applied to rid the dataset of such noise, the data preprocessing applied for data noise reduction are as follows: URL removal from the product review texts, case conversion of all text to lowercase, removal of repeating words, and the elimination of all null data from the data rows.

5.4 Lemmatization

Lemmatization is an NLP algorithm that is applied in text data normalization. Lemmatization helps in the contextualization of text data by normalizing text data into a common root, this enables computing algorithms to target words that have a common root as a group [2].

5.5 Performance Evaluation

The methodology of our study employs a standard performance evaluation of using the following criteria: accuracy, recall, f-score, and precision. These criteria are used to evaluate the classification of the sentiments by our proposed model based on the weighted parameters of true positive (tp), false positive (fp), true negative (tn), and false negative (fn). The weighted parameters are measured based on a classification matrix as illustrated in Figure 4.

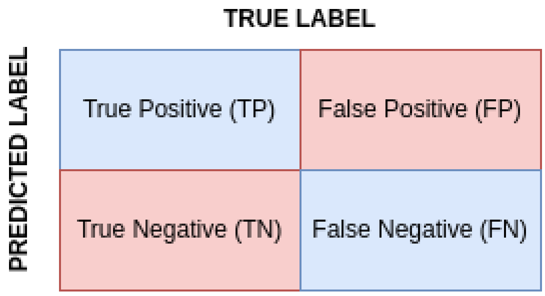

Fig 4: Illustration of a classification confusion matrix

The confusion matrix classification weights are used to calculate the evaluation metrics for accuracy, precision, recall, and f1-score as shown in equations 1, 2, 3, and 4 respectively.

$$\text{Accuracy} = (TN + TP) / (FP + TN + TP + FN) \tag{1}$$

$$\text{Recall} = TP / (TP + FN) \tag{2}$$

$$\text{Precision} = TP / (FP + TP) \tag{3}$$

$$\text{F1-score} = TP / (TP + 0.5\,(FP + FN)) \tag{4}$$

6. TRAINING AND TESTING

Training and testing of our model were carried out in two different sets of experiments, both sets of experiments used the same training and testing ratio proportion of 80:20, and training was set up using 10 epochs with efficient stoppage enabled for stopping when there is no training accuracy improvement during training. The difference between the two sets of experiments was one set applied stemming and lemmatization after the standard data preprocessing and the other was trained without the application of stemming and lemmatization. Figure 5 shows the training and validation accuracy charts for both experiments.

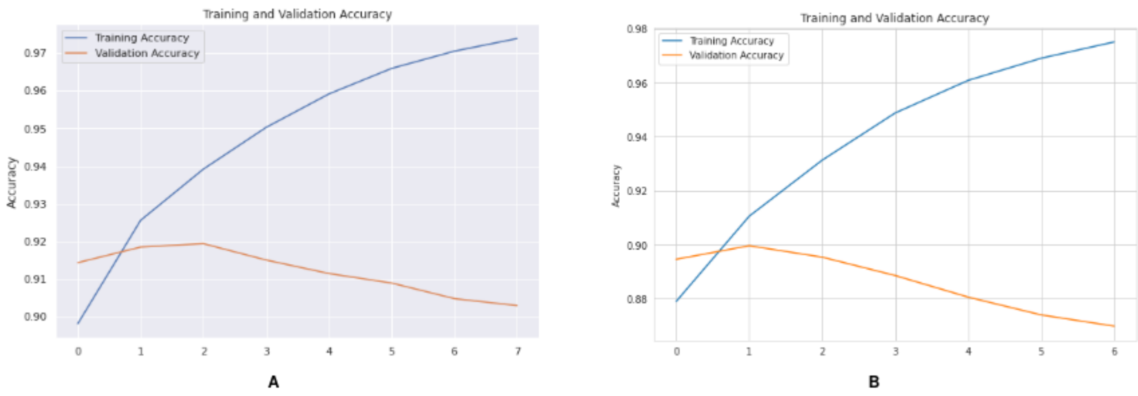

Fig 5: Training and validation graphs: (A) NLP, (B) without NLP

7. RESULTS

The performance evaluation of our experiments is ported comprehensively for both sets of experiments; with stemming and lemmatization and without. Performance accuracy of the experiments without stemming and lemmatization was recorded at 93.6% while accuracy with stemming and lemmatization was reported at 96.06%. Table 1 shows a report of all evaluation metrics measured for both experiments. Performance evaluation showed a significant improvement in performance accuracy by up to 3% when stemming and lemmatization NLP are applied to the dataset processing.

Table 1. Performance evaluation results

Evaluation criteria	**Bi-LSTM**	**Bi-LSTM + Lemmatization**
Accuracy	93.6%	96.06%
Precision	87%	91%
Recall	87%	90%
F1-score	87%	90%

8. RESULTS COMPARISON

Our proposed study was also compared with related words which carried out sentiment analysis for the Amazon product review dataset to compare the performance of our methods with other research methods. The study of [8] used MLP and BagOfWords, the study of [7] used BagofWords and LSVM, and the study of [20] utilized GRU and RNN used for comparison with our study. The comparison shows the use of Bi-LSTM with and without stemming and lemmatization yields more performance accuracy compared to all the previous studies Table2 shows the summary of the studies and their performance accuracy results.

Table 2. Comparison of studies carried out on the Amazon sentiment analysis dataset

Author(s)	**Methodology**	**Accuracy (%)**
Hague et al. (2018)	LSVM and BagofWords	91.72%
Hawlader et al. (2021)	MLP and BagofWords	92%
Shresth & Nazos (2019)	RNN and GRU	81.82%
This study	Bi-LSTM + Stemming/lemmatization	96.06%

9. CONCLUSION

Sentiment analysis is a critical component of natural language processing and machine learning. Sentiment analysis has a wide range of applicability in real-life scenarios with one of the main applications in e-commerce. This study carries out a sentiment analysis on e-commerce customer review data using lemmatization natural language processing technique and Bi-LSTM classification algorithm. The results of the experiments carried out in this study have shown the application of the lemmatization NLP technique significantly improves the classification accuracy of Bi-LSTM for sentiment analysis in sentiment analysis carried out on the ARPD dataset of e-commerce customer reviews. The application of lemmatization and Bi-LSTM for sentiment analysis on the ARPD dataset in this study has shown significant performance improvement compared with other state-of-the-art proposed methods on the same dataset. We recommend the use of this technique in future studies over different robust set of datasets to test its performance on different sentiment analyses for efficiency and reliability.

10. ACKNOWLEDGMENTS

We would like to thank our universities for their contribution and support of our research.

11. REFERENCES

[1] Banerjee, I., Ling, Y., Chen, M. C., Hasan, S. A., Langlotz, C. P., Moradzadeh, N., ... & Lungren, M. P. (2019). Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification. Artificial intelligence in medicine, 97, 79-88.

[2] Boban, I., Doko, A., & Gotovac, S. (2020). Sentence retrieval using stemming and lemmatization with different length of the queries. Advances in Science, Technology and Engineering Systems, 5(3), 349-354.

[3] Chowdhary, K., & Chowdhary, K. R. (2020). Natural language processing. Fundamentals of artificial intelligence, 603-649.

[4] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

[5] Dang, N. C., Moreno-García, M. N., & De la Prieta, F. (2020). Sentiment analysis based on deep learning: A comparative study. *Electronics*, 9(3), 483.

[6] Du, C., Sun, H., Wang, J., Qi, Q., Liao, J., Xu, T., & Liu, M. (2019, November). Capsule network with interactive attention for aspect-level sentiment classification. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP) (pp. 5489-5498).

[7] Haque, T. U., Saber, N. N., & Shah, F. M. (2018, May). Sentiment analysis on large scale Amazon product reviews. In 2018 IEEE international conference on innovative research and development (ICIRD) (pp. 1-6). IEEE.

[8] Hawlader, M., Ghosh, A., Raad, Z. K., Chowdhury, W. A., Shehan, M. S. H., & Ashraf, F. B. (2021, September). Amazon Product Reviews: Sentiment Analysis Using Supervised Learning Algorithms. In 2021 International Conference on Electronics, Communications and Information Technology (ICECIT) (pp. 1-6). IEEE.

[9] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[10] Jiddah, S. M., Abushakra, M., & Yurtkan, K. (2021). Fusion of geometric and texture features for side-view face recognition using svm. Istatistik Journal of The Turkish Statistical Association, 13(3), 108-119.

[11] Katić, T., & Milićević, N. (2018, September). Comparing sentiment analysis and document representation methods of amazon reviews. In 2018 IEEE 16th International Symposium on Intelligent Systems and Informatics (SISY) (pp. 000283-000286). IEEE.

[12] Kaur, M., & Mohta, A. (2019, November). A review of deep learning with recurrent neural network. In 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 460-465). IEEE.

[13] Kumar, H. M., Harish, B. S., & Darshan, H. K. (2019). Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method. International Journal of Interactive Multimedia & Artificial Intelligence, 5(5).

[14] Murthy, A. R., & Kumar, K. A. (2021, March). A review of different approaches for detecting emotion from text. In IOP Conference Series: Materials Science and Engineering (Vol. 1110, No. 1, p. 012009). IOP Publishing.

[15] Otter, D. W., Medina, J. R., & Kalita, J. K. (2020). A survey of the usages of deep learning for natural language processing. IEEE transactions on neural networks and learning systems, 32(2), 604-624.

[16] Pal, S., Ghosh, S., & Nag, A. (2018). Sentiment analysis in the light of LSTM recurrent neural networks. International Journal of Synthetic Emotions (IJSE), 9(1), 33-39.

[17] Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N., & Huang, X. (2020). Pre-trained models for natural language processing: A survey. Science China Technological Sciences, 63(10), 1872-1897.

[18] Ruder, S., Peters, M. E., Swayamdipta, S., & Wolf, T. (2019, June). Transfer learning in natural language processing. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: Tutorials (pp. 15-18).

[19] Sherstinsky, A. (2020). Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 404, 132306.

[20] Shrestha, N., & Nasoz, F. (2019). Deep learning sentiment analysis of amazon. com reviews and ratings. arXiv preprint arXiv:1904.04096.

[21] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... & Rush, A. M. (2020, October). Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations (pp. 38-45).

[22] Yang, W., Lu, W., & Zheng, V. W. (2019). A simple regularization-based algorithm for learning cross-domain word embeddings. arXiv preprint arXiv:1902.00184.

[23] Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8(4), e1253.