ISSN 2278-3091

Volume 11, No.3, May - June 2022

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse111132022.pdf

https://doi.org/10.30534/ijatcse/2022/111132022

Comparative study of Sentiment Analysis on Amazon Product Reviews using Recurrent Neural Network (RNN)

Iqbal Ahmed

Dept. of Computer Science and Engineering, University of Chittagong, Bangladesh, iqbal.ahmed@cu.ac.bd

Received Date : April 20, 2022 Accepted Date : May 22, 2022 Published Date : June 06, 2022

ABSTRACT

The problem of sentiment analysis on Amazon products is addressed in this research. In reality, because opinions are at the center of practically all human activity, sentiment analysis tools are used in almost every economic and social arena. They are also major influencers of our actions. The recurrent neural network (RNN) model is used to classify the product reviews of Amazon in this paper. Furthermore, using this family of models, which is particularly well-suited to the processing of sequential data, we were able to construct comprehensible text from an initial sequence on a character-by-character basis. As a result, we used three Amazon review datasets to estimate the authors' attitudes. As a result, we achieve results of 85% accuracy, and which are comparable to the greatest state-of-the-art models in this area.

Key words : Amazon Product Review, Recurrent Neural Network, Sentiment Analysis, Word embedding.

1. INTRODUCTION

Sentiment analysis is a branch of psychology that studies people's thoughts, feelings, attitudes, and emotions. It is, in fact, one of the most active fields of natural language processing study. It is extensively researched in the field of data mining. This research has also been expanded outside science to include management and social sciences. The rise in popularity of sentiment analysis corresponds to the rise in popularity of social media platforms such as chat rooms, blogs, microblogs, and Twitter [1]. For the first time in human history, there is a massive amount of information in digital form that can be analyzed. Because opinions are at the heart of practically all human activity, sentiment analysis technologies are used in almost every business and social industry. As a result, they have a significant impact on our actions. How do others see and value the world affect our beliefs and perceptions of reality, as well as the decisions we make? As a result, when a decision needs to be taken, other people's perspectives are frequently sought. This applies not only to individuals, but also to businesses. Internet users provide their thoughts on things on Amazon. These reviews, on the other hand, differ from one product to the next and are used to improve and warn companies that have received unfavorable feedback on their products [1]. The goal of this paper is to classify reviews from three Amazon review datasets using a recurrent neural network (RNN) model in order to anticipate the writers' attitudes. The remainder of the paper is laid out as follows. Section two defines sentiment analysis, section three discusses the state of the art, section four explains the methodology, and section five summarizes the findings and trials conducted on Amazon datasets. Finally, the concluding section brings the paper to a close.

2. DEFINING SENTIMENT ANALYSIS

Sentiment analysis, often known as opinion mining, is a computer science topic. It's part of automatic natural language processing, and it's used to categorize feelings expressed in texts. Opinion mining, sentiment mining, subjectivity analysis, effect analysis, emotion analysis, and other similar terms are used to refer to similar tasks [2].

Furthermore, rule-based classifiers and machine learning classifiers are the two most common methodologies for sentiment analysis. The former are sentiment analysis rules obtained from the study of a language's linguistics. Statistical machine learning methods are used to learn sentiment signals automatically [1, 2]. As a result, there are two terms that need to be defined: sentiment and opinion. This is due to a misunderstanding between these two terms in the literature and in the active field of research. A feeling, according to the Oxford Dictionary [3], is a point of view or viewpoint that is held or conveyed as an emotion. The term "opinion" refers to a belief or judgment about something that is developed based on the beliefs or opinions of a group or the majority of people, rather than on facts or information. As a result, the phrase feeling refers to a person's emotional response to something, whereas the term opinion refers to a person's point of view. Sentiment analysis, often known as opinion mining, is the study of people's opinions, feelings, evaluations, attitudes, and emotions, according to Bing Liu [4]. After that, the first step is to describe the differences between supervised and unsupervised learning. The former is characterized by the presence of two sets of data: a training set and a test set. Because the system is trained using a training subset of models that have already been processed, the method is called supervised. The latter (unsupervised), which only recommends one data set, needs the system to rearrange the information autonomously so that the most related data is grouped together [5]. Furthermore, there are three types of analytical approaches: lexicon-based, corpus-based, and hybrid (mixed) [5].

In terms of the Lexicon-based method, it determines the polarity of a text by comparing two sets of words, one of which represents a positive emotion and the other a negative emotion. The program then counts the amount of positive and negative terms in the text, with the sum providing an overall assessment of the text's mood. If the number of positive words outnumbers the number of negative terms, the text is deemed positive; if the numbers are equal, the text is considered negative or neutral [5].

The automatic sentiment analysis based on corpora necessitates the production of two manually annotated corpora in the corpus-based approach. The learning corpus is the first, and it is also utilized to teach an autonomous system. The first employs the test corpus, which is trained to check the performance of the automatic system, while the second uses notes made by human annotators so that the system may perform an analysis on its own. In an ideal situation, the outcomes of the automatic system's analysis should exactly match those of the learning corpus. As a result, it is critical that the learning corpus be representative of the test corpus in order to maximize the autonomous system's performance. In [5] presents an example of algorithm like a neural network.

Furthermore, the Hybrid strategy combines the benefits of the preceding two methods in three different ways. The initial step is to build the corpus using linguistic tools, then classify the texts using a supervised learning algorithm. The second option is to utilize machine learning to create the body of knowledge that the lexicon-based method requires. The third method entails combining the two preceding procedures and combining their outcomes [6]. In addition, the next third section will provide related works on sentiment analysis.

3. RELATED WORK ON SENTIMENT ANALYSIS

Despite the fact that sentiment analysis is a growing field in the natural language processing world, research on the non-English language is still scarce [7]. Indeed, most research investigations have focused on document polarity classification to avoid the high cost of sentence annotation, as well as using a machine learning-based approach to avoid the high cost of constructing a good-coverage vocabulary of opinion [8]. As a result, the lack of resources in the form of annotated corpora and opinion lexicons is slowing the growth of the field of Arabic sentiment analysis. The related work in the field of sentiment analysis will be reviewed in this section. Elhawary and Elfeky [7] were among the first researchers to be interested in creating an opinion lexicon for the Arabic language in this context. In fact, their work was crucial in the invention of a technology that allows stakeholders' opinions in the business field to be classified. As a result, more than 600 positive words, 900 negative words, and 100 neutral words were chosen as the beginning point. The accuracy of the evaluation tests was good, but the recall was low.

Furthermore, El-Halees [8] developed a three-step combination approach for classifying Arabic writings according to their polarity. After finishing the translation, he built an opinion lexicon using the English Lexicon SentiStrength and used web dictionaries to supplement the lexicon with synonyms for the words that previously existed. However, the lexicon's size was not specified, and the classification tool's accuracy for the lexicon-based phase is 48.7%. Abdul-Mageed and Diabont [9] have published the "Sifaat" lexicon, a manually produced lexicon with 3325 adjectives divided into three categories: positive, negative, and neutral. The examination revealed a 6% improvement in the categorization of subjectivity and a 40% improvement in the classification of polarity. The authors also advocated expanding the lexicon by translating three English opinion dictionaries: SentiWordNet, Youtube, and the General Inquirer. This strategy, however, was limited to adjectives that had not been appraised. Abdulla et al. [10] also proposed creating a lexicon by translating 300 SentiStrength words and then adding synonyms and emojis. The lexicon's final form has 3479 entries, 1262 of which are good and 2217 of which are negative. In terms of accuracy, the experiment on the collected corpus yielded a score of 59.6 percent. In the case of Alhazmi et al. [11], they proposed to create an Arabic version of SentiWordNet, an opinion lexicon derived from the WordNet database, by going through two stages: updating the base of WordNet Arabic 2.0 by mapping the WordNet 3.0. for English, and also the base obtained from the SentiWordNet 3.0 for English. In fact, the lexicon coverage assessment reports that 5% of the words in the annotated corpus are

In this study, we used an RNN model to predict author sentiments; in sections 4 and 5, we will show how our results outperform the best deep learning models such as LSTM, CNN, and GRU.

4. METHODOLOGY

Things began to alter in 2005, however. Indeed, with the introduction of 'deep learning,' which accounts for the majority of expert research, the outlook on the field of artificial intelligence has radically changed, especially because it intervenes in various fields, such as natural language processing. Deep learning (DL: Deep Learning in English) is a subset of machine learning, which is a subset of artificial intelligence in which machines may learn from their experiences and gain abilities without requiring human intervention [12]. As a result, enormous volumes of data are learned using artificial neural networks, which are algorithms inspired by the human brain. As a result, deep learning allows models with numerous processing layers to learn different degrees of abstraction for data representations.

4.1 Deep learning vs Machine learning

Machine learning (ML) is an artificial intelligence (AI) subfield that focuses on developing systems that learn or improve performance depending on data. These systems are designed to teach a secure set of algorithms for massive volumes of data that will aid in the classification of future data. Supervised and unsupervised machine learning are the two types of machine learning. The former is the most popular type of machine learning, in which the algorithm "learns" by comparing its actual output to "learned" outputs in order to detect faults and adjust the model accordingly. The latter is also known as observational learning. This method's learning algorithm finds common points among its input data on its own. In a machine learning process, on the other hand, the algorithm must be educated on how to create an accurate prediction using more data (for example, by manually extracting key characteristics) [8]. In fact, thanks to the structure of the artificial neural network, the algorithm in Deep Learning (DL) may learn to generate a correct prediction by its own data processing. Manual processes and feature extraction, for example, are ignored, while the modeling process is carried out automatically. Another significant distinction is that deep learning algorithms adapt in response to new data [13]. To succeed in a deep learning application, a big volume of data and a large dimension are required to train the model, which can be accomplished by using one or more GPUs (graphics processors) to handle the data quickly. This means that if you don't require these features, machine learning is preferable than deep learning.

Indeed, one of the most difficult jobs for classical machine learning models is feature extraction, which has been mechanized by deep learning models to enable them to attain a particularly high accuracy rate for computer vision tasks [13]. As a result, when it comes to unstructured data, deep learning's capacity to handle a high number of features makes it extremely strong. Deep learning techniques, on the other hand, may be overkill for less complicated tasks because they require enormous amounts of data to be effective. The success of deep learning algorithms, on the other hand, is dependent on the availability of more training data. Google, Facebook, and Amazon, for example, have already begun to use it to analyze their massive volumes of data [14, 15]. In practice, neural networks, also known as ANNs, are used in all deep learning methods. Information processing models that imitate the operation of a biological nervous system are known as artificial neural networks (ANNs). They work in a similar fashion to how the brain manipulates information on a functional level. All neural networks, in fact, are made up of linked neurons arranged in layers. The features of the neuron, as well as an explanation of how it works, will be discussed in the following section. In reality, the artificial neurons that make up neural networks are inspired by the genuine neuron in human brain [16].

The activation function, on the other hand, is a critical component of the neural network. By calculating the weighted sum of the inputs and applying the bias, this function determines whether the neuron is active or not. The input value undergoes a nonlinear modification. Furthermore, while nonlinearity is critical in neural networks, without the activation function, a neural network devolves into a linear model [17]. There are many different types of these functions, including the sigmoid function, ReLu function, and Softmax function. Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN) are two forms of neural networks that can be used [17, 18].

4.2 Steps of our approach

Deep Learning has been shown to be effective in a variety of complicated challenges involving the use of artificial neural networks to learn and extract meaningful patterns and information from data [18]. As a result, we see a lot of contributions that try to adapt this approach to solve the challenge of sentiment analysis. Also, other types of neural networks will be uses as comparison. Now we are going to present the tasks necessary to carry out this work in figure 1.

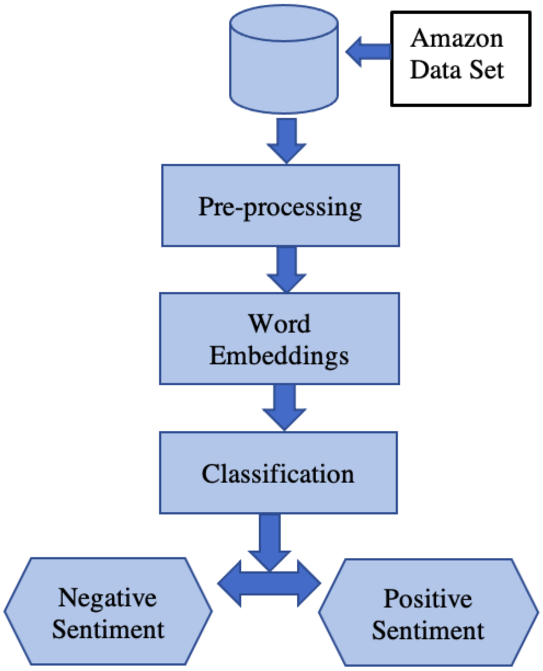

Figure 1: Architecture of our method

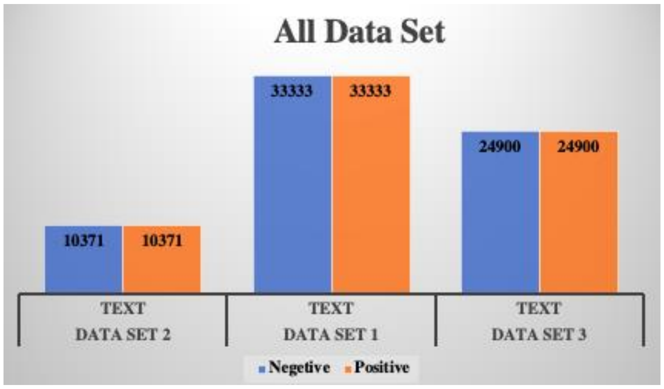

Figure 2: 03 Data set of Amazon Review

4.2.1 Pre-processing

The first step is to ensure that the dataset is of high quality. Figure 2 represents the total number of 3 data set from amazon product review. Before using a dataset for model training, there are several crucial procedures that must be completed. The data used must be representative, clean, precise, full, and correctly labeled in order for the models to be properly trained and produce the intended outcomes. For example, if we wish to forecast the emotions expressed in social media comments, the corpus must contain documents of the same type. As a result, preparing these data is a critical step [5].

4.2.1 Word embedding

The mapping of words to number vectors real in a reduced dimensional space is known as word embeddings [34]. Word embedding vectors describe words and their contexts; as a result, embeddings for words with similar meanings (synonyms) or semantic relationships will be more comparable. Word embeddings should also show how words are related to one another. For example, the inscriptions for "man" and "woman" should be "King" and "queen," respectively. We might start with pre-trained incorporations because learning word embeddings requires time and computing power.

5. EXPERIMENTS AND RESULTS ON DATASETS

The following experiments must be completed in order to get at our system. The corpus in the first Amazon dataset is separated into 4,149 test comments (representing 20% of the dataset) to test our model and determine its precision, and 16,593 training comments, with a precision of 75% for the training and 70% for the test. The results are shown in Figure 3.

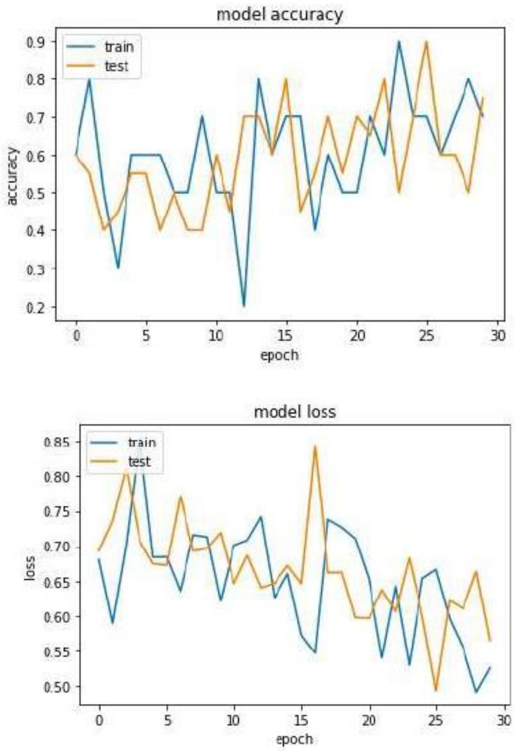

Figure 3: Precision and Error of the model for the first test

In the second trial, we used a dataset of 66666 items to train our binary sentiment classification model, with 33333 positives and 33333 negatives, as shown in Figure 4. For both training and testing, the best accuracy is attained with 25 epochs. With 25 epochs for test and 27 for training, the minimum value is likewise attained for the loss.

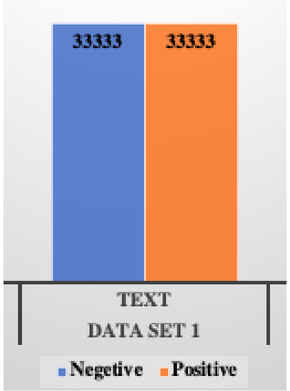

Figure 4: Second trail data set

In addition, the corpus is separated into 6667 test comments (representing 20% of the dataset) for testing and determining the accuracy of our model, and 59999 training comments. Then, for training, we got 90% precision and for testing, we got 85% precision. Figure 5 depicts the results as graphs. With 27 epochs, the best value is obtained for the loss.

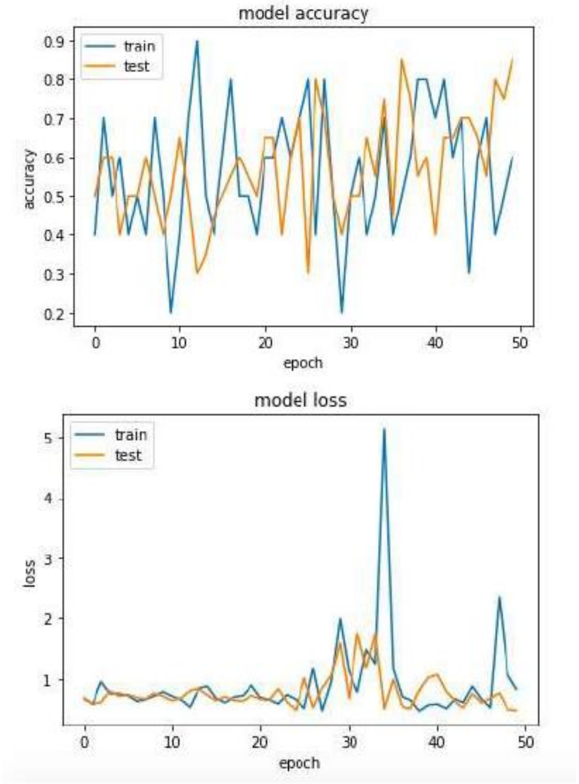

Figure 5: Precision and Error of the model for the second test

In the third trial, we used a dataset 3 of 49,870 items, with over 24,900 positives and 24,900 negatives, to train our binary sentiment classification model, as shown in figure 6.

Furthermore, the corpus is divided into 9973 test comments (representing 20% of the dataset) for testing and determining the accuracy of our model, followed by 39890 training comments. For the test, we received 90 percent for training and 70 percent for accuracy. With 17 epochs for training and 13 epochs for testing, the best results are obtained for the loss. The results are presented in the graphs below in Figure 7.

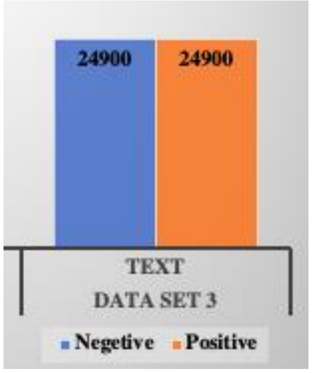

Figure 6: Third trail data set

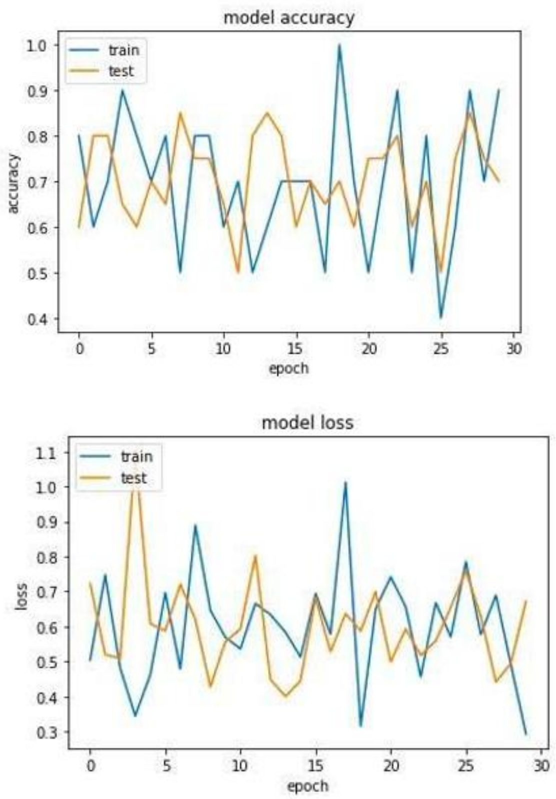

Figure 7: Precision and Error of the model for the second test

Our model was applied to the three data sets mentioned above and the evaluation was made in terms of the "accuracy" precision metrics, as summarized in the Table 1.

Table 1: Results on three amazon datasets

Amazon data sets	Training	Test	Accuracy
20742	16593	4149	70%
66666	59999	6667	85%
49870	39890	9973	70%

We chose distinct datasets based on their size to examine how a DL model performs in two separate scenarios with large and little amounts of data. We used the 30-epoch model to train the datasets for the final tests. The efficiency of our concept and the viability of its usage in this field are based on the findings we have.

We compare the results obtained with RNN to those obtained with three of the most well-known deep learning models. In fact, our results outperform those of CNN. We don't get encouraging results with LSTM and GRU. In the next Table 2 and Figure 8 represents the evidence.

Table 2: Comparison between 04 deep learning model in terms of Accuracy

DL Model	LSTM	GRU	RNN	CNN
Data 1	76	80	85	83
Data 2	68	65	71	70
Data 3	68	64	70	68

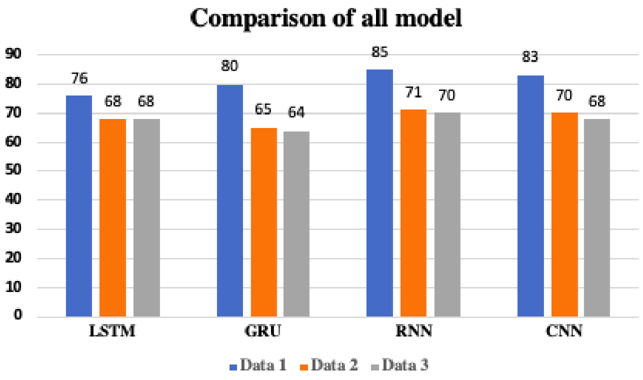

Figure 8: Comparison between 04 deep learning model in terms of Accuracy

Also, we compared our results on the 03 dataset amazon reviews with Bidirectional LSTM amazon review dataset containing 400000 reviews and found that we had the best accuracy with 96%, but we couldn't match their recall. Figure 9 represents the results.

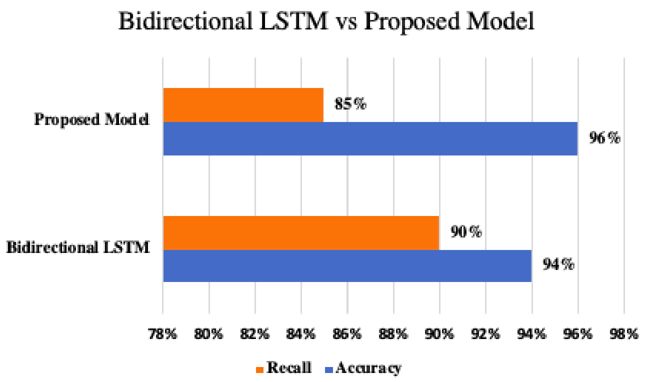

Figure 9: Comparison between bidirectional LSTM amazon review dataset containing 400000 reviews with our model

6. CONCLUSION

We contributed to the sentiment analysis challenge in this work by giving the tools and datasets we used, as well as the actions we took to get the best results for our model. In fact, our sentiment classification model enabled us correctly to identify more than 85% of the 7480 test items, resulting in a total of 5984 items correctly classified by our system. Furthermore, the trials showed that as the amount of the training data grows, so does the accuracy rate, implying that the proportion of the test dataset is just as crucial now as it was in the beginning. As a result, we can conclude that our model works well when applied to the sentiment analysis problem, where the amount of learning data is critical. Our algorithm was able to achieve an accuracy of 85 percent. To anticipate sentiments in the future, we will use multilingual corpora. We will also forecast people's feelings during the virus vaccination time based on Covid-19 reviews from various web forum.

REFERENCES

1. A. Abbasi, H. Chen, and A. Salem. **Sentiment Analysis in multiple languages: Feature selection for opinion classification in Web forum,** *ACM Transactions on Information Systems*, Vol. 26, Issue. 3, Article. 12, pp. 1-34, June 2008.
2. B. Liu. ***Sentiment Analysis and Subjectivity*,** Handbook of Natural Language Processing, 2^{nd} ed. (editors: N. Indurkhya and F. J. Damerau), 2010.
3. ***Definition of Sentiment Analysis***, https://www.lexico.com/definition/sentiment, Retrieved: October 2021.
4. J. K. Waters. ***The Everything Guide to Social Media: All you need to know about participating in today's most popular online communities,*** Adams Media Corporation, ISBN: 978-1-4405-0631-4, pp. 304, November 2010.
5. C. Van Hee. **Sentiment Analysis applied to political tweets: a corpus study,** Master Thesis dissertation, Faculty of Applied Linguistics, University College Ghent (HoGent), Belgium, 2013.
6. D. Poirier, F. Fessant, C. Bothorel, E. G. de Neef, and M. Boulle. **Statistical and Linguistic Approach for the Classification of Opinion texts about Films,** *Journal of New Information Technologies*, RNTI-E-17, pp. 147-169, 2009.
7. M. Elhawary and M. Elfeky. **Mining Arabic Business Reviews.** *In Proceedings of InternationalConference on Data Mining Workshops (ICDMW)*, pages 1108–1113. IEEE, 2010.
8. A. El-Halees. **Arabic Opinion Mining Using Combined Classification Approach.** *In Proceedings of the International Arab Conference on Information Technology (ACIT)*, 2011.
9. M. Abdul-Mageed and M. Diab. **Toward building a large-scale Arabic sentiment lexicon.** *In Proceedings of the 6th International Global WordNet Conference*, Matsue, Japan, 2012.
10. N. A. Abdulla, N. A. Ahmed, Mohammed A. Shehab and Mahmoud Al-Ayyoub. **Arabic Sentiment Analysis: Lexicon-based and Corpus-based.** *2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT)*, 2013.
11. S. Alhazmi, W. Black and J. McNaught. **Arabic SentiWordNet in Relation to SentiWordNet 3.0,** *International Journal of Computational Linguistics*, et 4(1), pp. 1-11, 2013.
12. J. Patterson, and A. Gibson. ***Deep Learning A Practitioner's Approach***, O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472, Mike LoukidesTimMcGovern, 532 p, pp. 28, 2017.
13. **What-is-deep-learning**, Saagie Blog https://www.saagie.com/en/blog/what-is-deep-learning/, Retrieved: October 2021.
14. **Google AI** , https://ai.google, Retrieved: November 2021.
15. **Tools for Advancing the World's AI**, https://ai.facebook.com/tools/#frameworks-and-tools, Retrieved: November 2021.
16. I. Vasilev, D. Slater, G. Spacagna, P. Roelants, and V. Zocca. ***Python Deep Learning,*** 2nd Edition, Packt Publishing, 2019.
17. I. Goodfellow, Y. Bengio, and A. Courville. **Deep Learning,** *MIT Press*, pp. 800, ISBN: 0262035618, 2018.
18. N. Buduma and N. Locascio. ***Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithm,*** O'Reilly Media, Inc., ISBN: 978-1-4919-2561-4, pp. 298, June 2017.