International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9, Issue-7, May 2020

Twitter Sentiment Analysis using Deep Learning

Ghazi A, Fatih Ö

***Abstract:** The whole world is changing rapidly with current innovations, using the Internet, has become a fundamental requirement in people's lives. Nowadays, a massive amount of data made by social networks based on daily user activities. Gathering and analyzing people's opinions are crucial for business applications when they are extracted and analyzed accurately. This data helps the corporations to improve product quality and provide better customer service. But manually analyzing opinions is an impossible task because the content is unorganized. For this reason, we applied sentiment analysis that is the process of extracting and analyzing the unorganized data automatically. The primary steps to perform sentiment analysis include data collection, pre-processing, word embedding, sentiment detection, and classification using deep learning techniques. This work focused on the binary classification of sentiments for three product reviews of fast-food restaurants. Twitter is chosen as the source of data to perform analysis. All tweets were collected automatically by using Tweepy. The experimented dataset divided into half of the positive and half of the negative tweets. In this paper, three deep learning techniques implemented, which are Convolutional Neural Network (CNN), Bi-Directional Long Short-Term Memory (Bi-LSTM), and CNN-Bi-LSTM, The performance of each of them measured and compared in terms of accuracy, precision, recall, and F1 score Finally, Bi-LSTM scored the highest performance in all metrics compared to the two other techniques.*

Keywords*: Sentiment Analysis, CNN, Bi-LSTM, NLP (Natural Language Processing)*

I. INTRODUCTION

Internet usage has become a fundamental requirement in people's lives, as they can buy and sell things or services online. Nowadays, social networking services provide a simple form of communication that permits users to exchange information and opinions directly with each other, or on a public platform, and they become the most significant resources for collecting information about people's feelings and sentiments on various topics. Thus, if someone needs to buy an item, there is no need to ask their friends and family for opinions on the products. Presently multiple user surveys are available in public web forums. Gathering and analyzing people's opinions are crucial, especially when they are extracted and analyzed appropriately.

Revised Manuscript Received on May 30, 2020.
***** Correspondence Author
Ghazi A*, Software Engineering, Firat University, Elazig, Turkey. ghazi.abdalla89@gmail.com
Fatih Ö, Assistant Professor, Software Engineering, Firat University, Elazig, Turkey. fatihozyurt@firat.edu.tr

Manual extraction and analysis of opinion is an impossible task because the content is disorganized and written in natural language. Sentiment analysis can be used to analyze opinions automatically, that usually modeled as a text classification problem. Text classification is a crucial task for natural language processing that can be performed in many applications that use understanding the natural language to determine the purpose and meaning behind the text and apply it to resolve multiple issues [1]. The ambiguity of the texts makes NLP extremely difficult in which a phrase or statement is not explicitly determined, and like most other languages, ambiguity, is broad in the English language [2].

Sentiment analysis is the classification of sentiments within text data using text analysis techniques. It can be used to change the public view of peoples automatically from unstructured data into structured data about brands, items, services, and politics. This data can be considerably helpful for business applications to revise marketing strategy by understanding the customer feelings on products also, sentiment analysis of people’s comments on social media sites can easily indicate whether consumers are satisfied with the products or not. It can help companies and corporations to get feedback from target consumers to identify their strengths and weaknesses also to know exactly how to raise the quality of their items or services [3].

This paper involves four sections organized in the following manner: Section 1 presents a summary of deep learning and how deep learning works. Section 2 shows the methodology of our work. Section 3 determines the classification techniques used for this work. Section 4 is the result and evaluation.

II. DEEP LEARNING

Deep learning is a specific part of machine learning in artificial intelligence (AI) and consists of algorithms that allow the software to train itself to perform tasks by exposing multilayer neural networks to massive amounts of data [4]. Recently, deep learning algorithms have given effective performance in natural language processing uses, comprising sentiment analysis over multiple datasets [5]. The greatest value of deep learning is that we do not need to manually extract features, instead of that, they take word embedding, as input which containing context information, and the middle layers of the neural network learn the features during the training phase by themselves. Words are expressed in the high dimensional vector and feature extraction performed by the neural network [6].
The main reason deep learning starts very rapidly due to provide superior performance on various issues also makes problem-solving much easier because it is fully automatic [7].

The deep neural network is about assigning inputs to targets through a deep chain of simple data transformations (layers), and such layers are learning by observing many samples of input and targets.
Transformation, performed through a layer that parameterized by own weights, also termed parameters. Learning layers means discovering a series of values for the weights of all layers in the network in such a way the network will precisely set the input samples for the targets associated with them. The deep neural network can include many million parameters, and getting the right value for each parameter looks like a dispiriting duty because changing the value of the individual parameter will influence the behavior of all other parameters, for this purpose loss function also called objective function will be used to computes a distance score by comparing the prediction of the network and the real object to estimate how far the predicted output is from the real object. After computing the distance score between the predictions of the network and the real target by using a loss function, this score is utilized as a feedback sign to slightly improve the value of weights, in a way that will reduce the loss score, this improvement is a function of the optimizer that performs what's called the Backpropagation algorithm. In the Backpropagation algorithm, at first weights of the network are appointed with random values, hence the network only performs a sequence of arbitrary shifts. Generally, the results are ideally far from what it should be, so the loss score is too high. But, in each case the network handles, the weights are slightly modified in the right trend, and the loss score decreases, that's the training loop that iterated enough times, giving weight values that reduce the loss function. The lowest loss network is the network where the outputs are closest to the targets.

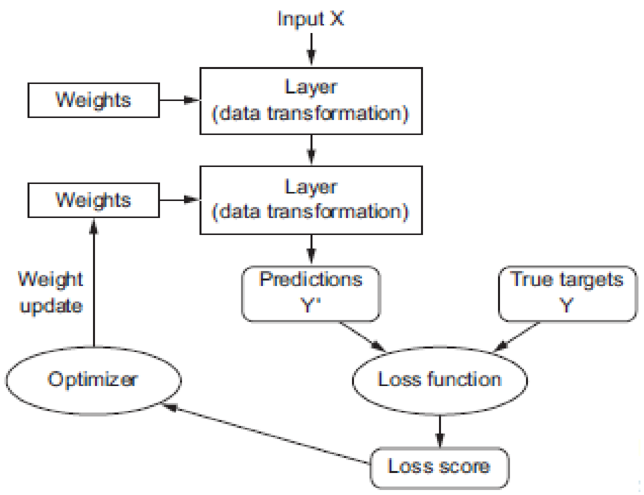

Figure I: Deep Learning Process [4]

III. METHODOLOGY

The primary steps to perform sentiment analysis include data collection, pre-processing, word embedding, sentiment detection, and classification using deep learning techniques. In this work, all data gathered automatically by using Tweepy. In the pre-processing phase, the data transformed into a normal text, which leads it more sensible to the machines compared with the prior form. In the next stage, word embedding performed to represent words in a numeric form. Finally, to classify sentiments, in the training data, each review related to a class label, this data is transferred to classifiers to train and learn, then test reviews are provided to the model, and classification is performed through these trained classifiers, finally, reviews classified into the positive or negative. In this paper, three deep learning techniques implemented which are Bi-LSTM, CNN, and CNN-Bi-LSTM.

A. Data Collection

Data collection is the process of accumulating data in many diverse sources. When deep learning grows popular, training data is much needed for good performance [8]. Data collection mainly comprises of data acquisition, data labeling, and enhancement of present data. The data acquisition purpose is to get datasets that deep learning models can be able to train on it, and it includes three methods: data discovery, data augmentation, and data generation. Data discovery is essential when it needs to share or quest for new datasets and more datasets can be accessed on the Web [9]. Data augmentation is an extension of the data discovery as present datasets are improved by appending more exterior data. Data generation utilized when no external data set is available. In this work, all tweets are written in English and collected by using Twitter API, for this reason, Tweepy which is a python library was used to connect with a Twitter Application Programming Interface (API) to extract real-time data from Twitter, and it installed by using pip command: pip install Tweepy. The data originally stored in CSV file format with two columns which are, review and sentiment.

B. Pre-Processing

Pre-processing is a very significant step to convert the text in a human language into a machine-readable form for more processing, and it affects the efficiency of other steps. The pre-processing step aims to make the data more machines readable to reduce ambiguity in feature extraction. In this work, some steps were used to normalize the text which is, converting the upper case to lower case, remove duplicate text, stop words, numbers, multiple spaces, special characters, a single character, punctuation marks, URLs, Html, mention, and hashtags. Also performing lemmatization for words, it is the process of substituting words with a stem or base words to decrease inflectional structure to a typical root structure, and expansion of slangs and abbreviations.

Table I: Example of acronym words and expansions

Acronym	Expansion words
cwmaos	Coffee with milk and one sugar
evry1	Everyone
gj	Good job

C. Word Embedding

Word embedding is a main deep learning techniques used to the numeric representations of words that are useful to solve the problem in natural language processing. Neural networks in NLP, do not receive raw words as input since the networks can only understand the numbers.

Retrieval Number: G5814059720/2020©BEIESP
DOI: 10.35940/ijitee.G5814.059720
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Hence, words should transform into feature vectors, or word embedding's [10]. The word vectors can be learned by feeding a large group of raw text into a network and training it for a sufficient amount of time. After training word embedding, it used to extract similarities between words or other relationships.

This method has gotten a great deal of consideration in the text, including sentiment analysis due to its capabilities to take the syntactic and semantic similarities between words. For example, vectors for the words food and rice will have higher similarities than the rice and car vectors. Recently, different strategies developed to create meaningful models that can learn word embedding's from huge texts. The most popular methods are word2vec [6, 10] and global vectors (Glove) [11]. At present both of the methods are among the most reliable and useful word embedding methods that can turn words into meaningful vectors.

D. Sentiment Classification

In this work, after performing word embedding and creating a feature vector classification is done using CNN, CNN-Bi-LSTM, and Bi-LSTM, then the performance of their results are compared.

IV. CLASSIFICATION TECHNIQUES

In this part, we explore some various techniques used to sentiment analysis. They are as follows:

A. Long short-term memory (LSTM)

LSTM is a specific kind of RNN that functions are more sophisticated, and it learns to manage the flow of information [12]. The standard RNN has an issue of gradient vanishing or exploding. To conquer these issues, an LSTM intended by Hochreiter and Schmidhuber [13]. LSTM involved a memory cell, input gate, output gate, and a forget gate. Data can be saved, read, or write from cell-like information in a computer's memory [14]. The cell makes decisions about what to read, write, or erase through opened and closed gates. These gates work on the signals they receive and pass or block data due to its strength or weakness. This division of responsibilities enables the network model to retain information for long periods [15].

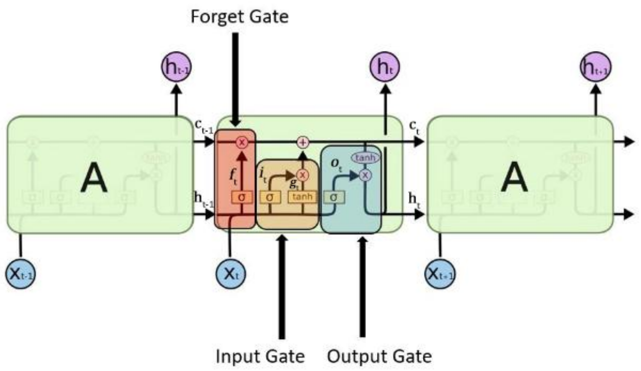

Figure II: LSTM with its gates

B. Bidirectional long short-term memory (Bi-LSTM)

Bidirectional LSTMs are an extension of standard LSTMs that can enhance the efficiency of the model in sequencing classification issues. Bi-LSTMs train on two LSTMs rather than one LSTM on the input sequence, where all time steps of the input sequence are available. The first one passed on the input sequence without modification, and the other one passed on a reversed copy of the input sequence, and it connects them to the same output, thus at each time step, the networks can have backward and forward information about the sequences. This extra setting adds to the network and enhances the accuracy of the network. Bi-LSTMs are mainly useful when an input context is needed. For example, in sentiment analysis, performance can be improved by knowing the words before and after existing words.

C. Convolutional neural network (CNN)

CNN is a special kind of neural network originally intended to computer vision and exploits layers with convolving filters that implemented to local features. It broadly applied in different applications such as NLP, speech processing, and computer vision. The network contains neurons with weights and biases that are modified based on training data by some learning algorithm, and it has local receptive fields, which are small regions of neurons in the input layer connected to the neurons of the hidden layer. The CNN structure consists of convolution layers, a pooling layer, and one or more fully connected layers [16]. 1d-CNN, first intended by Kim, works with patterns in one dimension and tends to be useful in natural language processing, it receives sentences of different lengths as input and provides fixed-length vectors as output [17]. The maximum sentence length processed by the network, the longer sentence is cut and the shorter sentence filled with zero vectors. Next, the dropout regularization utilized to manage over-fitting.

V. RESULT AND EVALUATION

In this part, we discuss the results obtained through CNN, CNN-Bi-LSTM, and Bi-LSTM, all trained on 200k Twitter datasets that were created by us, and the results are compared based on metrics like accuracy, precision, recall, and f1 score.

In building any deep learning model, one of the primary tasks is to evaluate its performance; the performance of each technique used in this work is measured, by computing different metrics and the ultimate purpose behind working with different metrics is to understand how well a deep learning model is going to perform on unseen data. In this work, the following metrics are used: Accuracy is the proportion of the accurately analyzed samples to the total number of samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Precision is the number of accurate positive analyzed samples to the number of predicted positive results by the classifier.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

Recall is the number of accurate positive analyzed samples to the number of all relevant samples.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

F1 score computes precision and recall of the test to calculate the score.

$$F1 = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

In the above equations, TP is the true positive and predicted correctly, FP is the false positive and predicted incorrectly, TN is the true negative and predicted correctly, FN is the false negative and predicted incorrectly.

Table 2 illustrates the results and performance comparisons of the models, in this work, the experimented dataset split into training, validation, and testing set. The training set was given 80% of the dataset, while the validation and testing set were each given 10% of the dataset. Models trained on 162,000 records, validate on 18,000 records, and tested on 20,000 records.

Table II: Results and performance comparisons of the models

Techniques	Accuracy	Precision	recall	F1 score
CNN	89.0	89.0	89.0	89.0
CNN-Bi-LSTM	89.36	89.5	89.5	89.0
Bi-LSTM	**90.3**	**90.5**	**90.5**	**90.0**

In this work, the result of each technique achieved in the following configurations: each of the models configured with a dropout layer to restrict the neural network from memorizing the training set, which is useful to prevent the overfitting. The models compiled with the Adam optimizer with the batch size of 128 for 15 epochs, the output layer in all models is a fully-connected dense layer with sigmoid activation that makes a binary prediction. In the CNN model; the network has a three-layer of 1d-CNN that all layer implemented with 64 filters and a kernel size of 1, 2, 3 respectively, after each layer a max-pooling layer with 2 pooling filter size is applied that selects the value with the highest weight only and ignores the rest values which significantly enhance the results of the convolutional layer and reduces the input to the next layer. Also, it has a one flattens layer that transforms a two-dimensional matrix of features into a vector that can be fed into the output layer. In the CNN-Bi-LSTM model; the network has one layer of 1d-CNN with 64 filters and a kernel size of 1, after that a max-pooling layer with 2 pooling filter sizes and it has a two-layer of Bi-LSTM that each layer has 64 neurons. In the Bi-LSTM model, the network has a three-layer of Bi-LSTM that each layer has 64 neurons. Finally, Bi-LSTM showed the best performance in all metrics and achieved the highest accuracy of 90.3 %, and it tested on 20k dataset it predicted 18060 tweets correctly and 1940 tweets incorrectly.

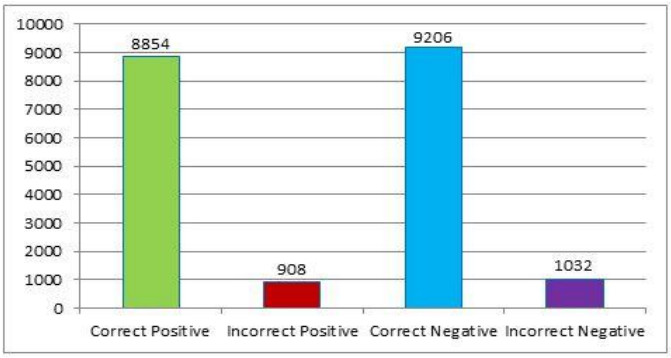

Figure III: Best model statistics of correct and incorrect predictions

VI. CONCLUSION

Sentiment analysis is the application that many companies use to boost their advancement. Most business organizations consider that the success of their business depends only on customer satisfaction. It presents a significant role in the research field of text mining, which helps to extract and analyze a vast collection of unorganized data collected on the web. It applies machine learning and deep learning algorithms that integrate with text mining to get valuable information from unorganized data, and this information has many advantages for business applications to improve product quality, revise marketing strategy, provide better customer service, and identify customer reviews in which helps the company to identify its strengths and weaknesses. All of which add up to boost sales and revenue. This work focused on analyzing sentiments of product reviews into positive or negative using sentiment analysis techniques. We used CNN, CNN-Bi-LSTM, and Bi-LSTM in the analysis process, also word2vec technique was used for word embedding. In this work, tweets, collected automatically from Twitter using Tweepy and all the techniques experimented on the 200k dataset that divided into half of the positive and half of the negative tweets. In conclusion, the results showed that CNN takes less time to process than Bi-LSTM, but performance is not up to the mark. Bi-LSTM achieved better performance compared to CNN, and CNN-Bi-LSTM especially when the volume of the data, is huge because it builds the relationship between hidden vectors at each time step, that's why performed the task much slower, finally, Bi-LSTM achieved the highest accuracy of 90.3%.

REFERENCES

1. S. Xu, H. Liang and T. Baldwin, "Unimelb at semeval-2016 tasks 4a and 4b: An ensemble of neural networks and a word2vec based model for sentiment classification," in *Proceedings of the 10th international workshop on semantic evaluation (SemEval-2016)*, 2016.

2. H. Lakkaraju, R. Socher and C. Manning, "Aspect specific sentiment analysis using hierarchical deep learning," in *NIPS Workshop on deep learning and representation learning*, 2014.
3. H. Binali, V. Potdar and C. Wu, "A state of the art opinion mining and its application domains," in *2009 IEEE International Conference on Industrial Technology*, 2009.
4. F. Chollet, Deep Learning with Python, Manning Publications Co., 2018.
5. R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu and P. Kuksa, "Natural language processing (almost) from scratch," *Journal of machine learning research,* vol. 12, pp. 2493-2537, 2011.
6. T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in neural information processing systems*, 2013.
7. F. Özyurt, "Efficient deep feature selection for remote sensing image recognition with fused deep learning architectures." The Journal of Supercomputing (2019): 1-19.
8. I. Goodfellow, Y. Bengio and A. Courville, Deep learning, MIT press, 2016.
9. I. G. Terrizzano, P. M. Schwarz, M. Roth and J. E. Colino, "Data Wrangling: The Challenging Yourney from the Wild to the Lake," in *CIDR*, 2015.
10. T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient estimation of word representations in vector space," *arXiv preprint arXiv:1301.3781,* 2013.
11. J. Pennington, R. Socher and C. D. Manning, "Glove: Global vectors for word representation," in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014.
12. S. Hochreiter, "The vanishing gradient problem during learning recurrent neural nets and problem solutions," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* vol. 6, no. 02, pp. 107-116, 1998.
13. S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.
14. H. Strobelt, S. Gehrmann, H. Pfister and A. M. Rush, "Lstmvis: A tool for visual analysis of hidden state dynamics in recurrent neural networks," *IEEE transactions on visualization and computer graphics,* vol. 24, no. 1, pp. 667-676, 2017.
15. X. Li and X. Wu, "Constructing long short-term memory based deep recurrent neural networks for large vocabulary speech recognition," in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.
16. D. Ciregan, U. Meier and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE conference on computer vision and pattern recognition*, 2012.
17. Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882,* 2014.

AUTHORS PROFILE

Ghazi A, he obtained his B.Sc. in Computer Science from University of Sulaimani, Iraq, in 2012. Currently pursuing his Master degree in Software Engineering. He has 3 years of teaching experience as an Asst. Lecturer. His teaching includes Data Structure, Object Oriented Programming, and Graphics with C#. He has published 1 research paper in “2019 7th International Symposium on Digital Forensics and Security (ISDFS)”. His interest is in Programming language and deep learning.

Fatih Özyurt received bachelor's and master's degrees in computer engineering from Eastern Mediterranean University in Cyprus, and Fatih University in Istanbul 2011 and 2014, respectively. In 2019, he graduated from Ph.D. degree in Software Engineering at Firat University in Elazig. He works as Assistant Professor at the Department of Software Engineering at Firat University. He has more than 30 international conference and journal paper. His research interests include pattern recognition, artificial neural networks, image processing, deep learning, and information security.