

Article

A BERT framework to sentiment analysis of tweets

Abayomi Bello ¹, Sin-Chun Ng ¹ and Man-Fai Leung ^{1,*}

¹ School of Computing and Information Science, Faculty of Science and Engineering, Anglia Ruskin University, Cambridge, U.K.

* Correspondence: man-fai.leung@aru.ac.uk

Abstract: Sentiment analysis has been widely used in Microblogging sites like Twitter in recent decades, where millions of users express their opinions and thoughts because of its short and simple manner of expression. Several studies reveal the state of sentiment which does not express sentiment based on the user context because of different length and ambiguous emotional information. Hence, this study propose text classification with the use of Bidirectional Encoder Representation from Transformer (BERT) for natural language processing with other variant. The experimental findings demonstrate that the combination of BERT with CNN, BERT with RNN, BERT with BiLSTM performs well in terms of accuracy rate, precision rate, recall rate, and F1-score compare to when it was used with Word2vec and when it was used with no variant.

Keywords: sentiment analysis; deep learning; tweets; BERT; LSTM

1. Introduction

It is obvious that the emergence of real-time information networking platforms like Twitter has led to the development of an unmatched public collection of viewpoints about all relevant worldwide entity thereby interfering and affecting human lifestyle [1]. Twitter may be a great platform for opinion generation and presentation, but it also presents new and unique obstacles, and the process would be incomplete without capable tools for assessing those thoughts to speed up their consumption. The best approach over time has proven to be using sentiment analysis tools to identify individual attitudes and emotions [2].

Sentiment analysis, which is additionally referred to as subjective investigation or artificial intelligence of emotions) is a natural language processing (NLP) technique for extracting information patterns and key characteristics from a large body of text. It analyses the thoughts, attitudes, viewpoints, opinions, convictions, remarks, requests, inquiries, and interests expressed by the author based on feelings not reason in the form of texts, with entities like service, issue, person, product, event, object, organisations and their attributes. It identifies the author's overall attitudes toward a text, which could be anything from blog posts to product reviews to online forums to speeches to data from databases to social media documents [3].

The need for Natural Language Processing (NLP) arouse as a result of the need for computer to understand spoken and written language of humans. This brought about bag of words (BOW) which uses N-grams but the contextual meaning of words are ignored with BoW models, after which the word embedding was developed to overcome this issue which always takes words similar in meaning as similar context but this model always rely on large vocabulary and high computational power. [30] Word2Vec was then developed which generate only one vector embedding for each word but also has a shortcoming by considering left or right context. The transformer model which was introduced by google in 2018 then solves the aforementioned problem and enhance language processing.

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Last-name

Received: date

Accepted: date

Published: date

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

It helped to overcome the problem of transfer learning and has recorded great achievement on Natural Language Processing tasks such as Named Entity Recognition (NER), Question Answering and Sentiment Analysis. The BERT has been pre-trained on large corpus of English data which acts like a benchmark and helps solve similar problem. The BERT has transformer encoder layers enhanced with a self-attention mechanism.

This study seeks to produce an approach that can identify the opinion, attitude of a writer in a tweet according to context. The pre trained transformer BERT and Word2Vec was used with the Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), Bidirectional Long Short Term Memory (BiLSTM) and experiment was carried out and the Proposed BERT obtained a state of the art performance.

This study contributions includes proposing the BERT for NLP that can identify sentiment in tweets according to three categories (Positive, Negative and Neutral) based on the context of the writer. Our approach is distinct from similar studies because:

- We have trained our model using six different dataset which is a combination of different tweets.
- We will be combining knowledge embedded in the pre trained Bidirectional transformer (BERT) with deep learning classifier to detect sentiment (Positive, Negative or Neutral) other than just using machine learning classifier.
- The proposed BERT will dynamically generate vector according to the word context and when placed into deep learning classifiers such as CNN, RNN or BiLSTM to predict output, achieves an accuracy of 93% and F-measure of 95%.

2. Literature Review

The development of the internet has changed how people now express their ideas and thoughts, according to Kepios, there are 4.74 billion social media users around the world equating to 59.3 percent of the total global population. For context, the data suggest that more than 75 percent of eligible global population now uses social media [33]. Further research shows that a typical social media users actively uses or visits an average of 7.2 million different social platforms each month and spends close to 2 and half hours per day using social media. The world spends more than 10 billion hours using social platforms each day which is equivalent of nearly 1.2 million years of human existence. Additionally, social media gives businesses a chance by offering them a platform to engage with their customers for advertising. People heavily rely on user-generated content from the internet when making decisions. Social networking services like Twitter is a valuable sources of information to find out what happened or what is happening in a geographic area [34]. Microblogs have become an important origin of information regarding events happening in a location during a period of time [35]. Twitter been one of the most used platforms with easy access to the tweet with connection to the API and having a maximum length of 280 characters makes it suitable to effectively monitors emotions, sentiments, opinions, and attitudes of different subject matter.

Artificial Intelligence, AI is the art and science of building intelligent machines, particularly smart computer programmes. Furthermore, AI can be defined as the imitation or reproduction of cognitive functions by computer systems that can reason logically and act in ways resembling those of humans. The subject gained popularity as a subject in academic literature after the 1950s. Various industries use AI, including communication, IT, healthcare, agriculture, logistics, education, and aviation. Artificial intelligence-powered chatbots significantly improve customer service [20].

In recent years, natural language processing (NLP) has drawn a lot of interest for its ability to computationally represent and analyse human language. It has expanded the range of industries in which it is used, including machine translation, email spam detection, information extraction, summarization, and medical. [11]. It makes interactions between people and computers simple and effective using computational linguistics and machine learning. NLP systems can output written texts or processed speech from inputs like text, images, or speech [7].

Neural network is a subset of machine learning with numerous applications such as compressed image reconstruction [14], asset allocation [21, 22], non-negative matrix factorization [23, 24] and model predictive control [27]. With the development of transfer learning G.E. Hinton firstly introduced the concept of deep learning in 2006, and it is simply extracting features from raw data with the help of using layers [9]. The human brain affects neural networks, which are made up of numerous neurons and form an amazing network. Deep learning networks can be used to teach both supervised and unsupervised categories [17, 26, 28, 29]. CNN (convolutional neural networks), RNN (recurrent neural networks), recursive neural networks, and many other networks with more than three layers are considered as deep learning approaches. Text creation, vector representation, word representation estimation, sentence classification, phrase modelling, feature presentation and emotion recognition benefit greatly from neural networks [18, 25, 27].

Additionally, the term deep learning has gained popularity among computer scientists to refer to pattern-recognition algorithms that enable computers to learn on their own, leading to improvements in speech and image recognition as well as more precise translation software. In addition to a deeper focus on context, thought, and abstraction, deep learning can also refer to knowledge that is less surface-level and more contemplative and abstract [6].

2.1 Sentiment Analysis based on Machine Learning Approach

Suhasini et al.[30] was able to identify emotions in twitter using supervised learning. K-nearest neighbor (KNN) and Naive Bayes (NB) were the two algorithms compared and the study shows that Naives Bayes outperformed the K-nearest neighbor.

Jayakody et al. [8] collected data from twitter post based on product review, then analyze using the Support Vector machine (SVM), logical regression and K-nearest neighbor machine learning algorithm and count vectorizer and term frequency-inverse document frequency mechanisms for converting text into vectors for the data to be inputted into the machine learning model. The highest accuracy score was achieved by Logistic Regression with count vectorizer with an accuracy rate of 88.26%.

Bhagat et al [31] used a hybrid approach of Naive Bayes and K-Nearest Neighbor to divide tweets into three classes: Positive, Negative and Neutral and they achieved a better accuracy than the random forest.

2.2 Sentiment Analysis based on Deep Learning and BERT Approach

Chiorrini et al [12] proposed two BERT based approaches for text classification: BERT-base and cased BERT-base. They gathered information from microblogging sites, particularly twitter. Two separate datasets were employed in their experiment and they were used for sentiment analysis and emotion recognition. The proposed model gives an accuracy of 92%. They emphasised that BERT produces positive outcome for text classification.

Huang et al [32] presented a model for text classification where he used the Deep Convolutional Neural Network, Bidirectional Gated Recurrent. This model was based on BERT. Two dataset (CCERT email and movie comment) are used. The result gives an accuracy of 92.66% on CCERT while 91.89% on the movie dataset.

The researcher in the study [13], represented a seven-layer framework to analyse the feelings of sentences. This framework was based on CNN and Word2vec to calculate vector representation and SA, respectively. Google has proposed the use of Word2vec. To improve the correctness and generalizability of the suggested model, the researcher employed the dropout technology, normalisation, and parametric rectified linear unit (PReLU). The framework was validated using a data set from rottentomatoes.com that contains a corpus of movie review extracts with five labels: positive, slightly positive, neutral, negative, and somewhat negative. Compared to earlier models, such as Matrix-Vector

recursive neural network (MV-RNN) and recursive neural network, the suggested model outperformed the previous with an accuracy of 45.5%.

3. Materials and Methods

This section provide a concise and precise description of the experimental results, their interpretation, as well as the experimental conclusions that can be drawn.

3.1. Data Collection and Preprocessing

Six datasets were used and they were tweets collected from Kaggle. The six datasets were combined using Python's concatenating function. There are 2 columns and 212661 rows in the dataset altogether. The null values were eliminated, and mapping was completed. 1.0 was assigned to Positive, 0.0 to Neutral, and -1.0 to Negative. Special characters, punctuation, numbers, symbols, and hashtags were removed from the model dataset as shown in Table 1 below. The dataset sentiment contains 71,658 Neutrals, 85231 Positives, and 55,772 Negatives and has a percentage of 40.1% to be Positive, 33.7% Neutral, and 26.2% Negative as shown in Table 2 below.

Table 1. Dataset Composition.

Dataset Name	Rows by Columns	Source
Twitter_Data.csv	162,980 * 2	Kaggle.com
Apple-twitter-sentiment-texts.csv	1,630 * 2	Kaggle.com
FinalSentimentdata2.csv	3,090 * 2	Kaggle.com
Tweets.csv	14,640 * 2	Kaggle.com
Train.csv	27,481 * 4	Kaggle.com
Test.csv	3,534 * 3	Kaggle.com
Model Dataset after Concatenating	213,355 * 2	

Table 2. Sentiment count.

Sentiment	Sentiment Count	Sentiment Percentage
Negative	55,717	26.2%
Neutral	71,658	33.7%
Positive	85,167	40.1%

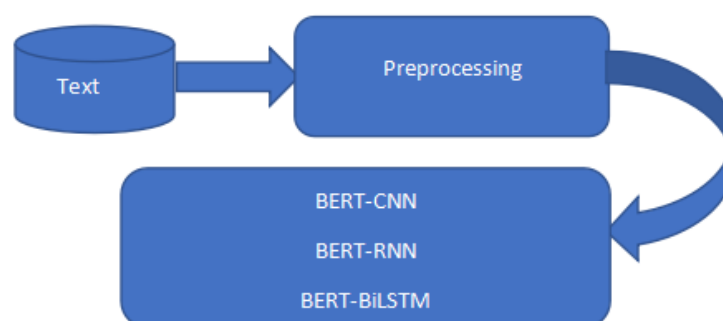
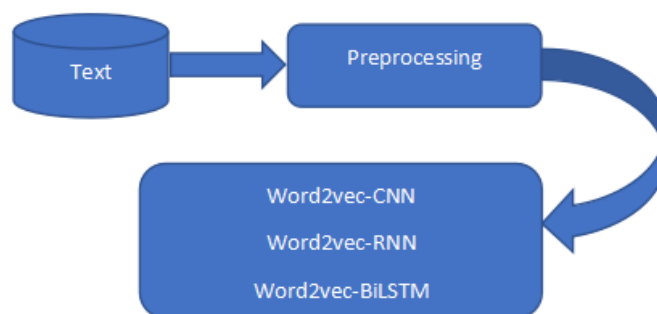
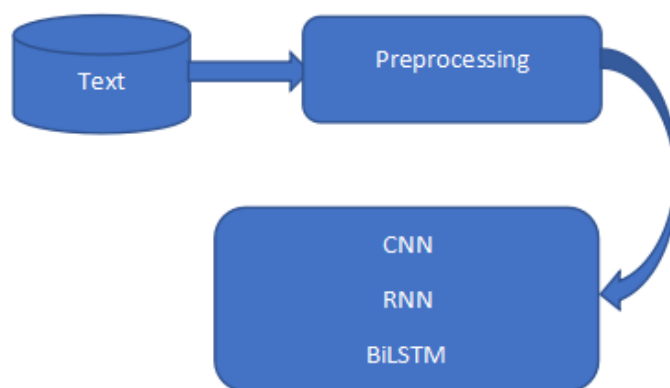


Figure 1: BERT based Architecture**Figure 2** Word2Vec based Architecture**Figure 3** CNN, RNN, BiLSTM Architecture

3.2. Bidirectional Encoder Representations from Transformers (BERT)

BERT is a machine learning pre-trained model developed by Google AI in 2018 for natural language understanding tasks to assist machines in comprehending the context of phrases [10]. BERT employs transfer learning, and the architecture is based on the transformer model. Transfer learning entails training a model for a broad task and then applying the knowledge gained to fine-tune BERT for a new task [16]. BERT has been trained on two tasks: masked language tasks, in which sentences are fed to the model and some words are masked or hidden for the model, and the model attempts to predict these hidden words, and unmasked language tasks, in which sentences are fed to the model and some words are masked or hidden for the model, and unmasked language tasks, in the other task is sentence prediction where pair of sentences are fed to the model each round and the model need to predict whether one sentence is followed by the other or not. BERT has been trained over a large dataset for these two tasks. The dataset contained all English Wikipedia and 11,038 books. BERT uses an encoder from a transformer model, a type of neural network that takes a sentence as input to the model. Then each word of the sentence is tokenised, and these tokenised words are fed to the BERT model. BERT output is a vector representation for each tokenised word.

Using encoders from Transformer enables BERT to have a better context understanding than traditional neural networks such as LSTM or RNN since the encoder process all inputs, which is the whole sentence, simultaneously so when building a context for a

word, BERT will take into account the inputs before it and also the inputs after the word, while the LSTM or RNN process the input taking in account only the prior inputs, and that will be reflected on the output vector value for the word, so the word “python” in the two sentences (I just started learning Python) and (Python are found in part of Africa and Asia) would have the same vector value -and as a result the same meaning- when using LSTM or RNN, on the other hand, it would have two different vectors using BERT, so as a result, using BERT will in most cases give us better performance than using the traditional machine learning algorithms.

3.3. Word Embedding

Computers are programmed to operate in numbers. It worked on Bits which are zeros and ones. The question now is what happens when the software or a task must process a word? This word needs to be given to the computer as the only thing it can understand is numbers which means it needs to be broken down into bits (zeros and ones). The most straightforward approach in NLP is to create a vocabulary with many terms and assign a number to each word in the vocabulary [19]. Word embedding analysis is a natural language processing method in which a neural network is trained using machine learning to anticipate the contexts in which words are employed [4].

3.4. RNN

RNN is a type of artificial neural network that identify patterns in data and utilize them to anticipate the following most likely outcome. It operates on the tenet that each layer's output is saved and fed back into the system's input in order to forecast that layer's output.

Table 3. Parameter Setting of RNN.

Parameter	Values
Learning Rate	0.001
Epoch	10
Optimizer	Adam
Batch Size	128
Activation	Softmax
Loss	categorical_crossentropy

3.5. CNN

CNN is formed of different layers of neurons. This works best on images, when an image is entered, each layer of the network creates a number of activation that are passed on to the following layer. Typically, the first layer extracts fundamental features like edges that run horizontally or diagonally. The following layer receives this output and detects more intricate features like corners or multiple edges. The network may recognise increasingly more complex elements, including objects, faces, etc., as we go further into it.

Table 4. Parameter Setting of CNN.

Parameter	Values
Epoch	10
Optimizer	Adam
Batch Size	128

Activation	Softmax
Loss	categorical_crossentropy

3.6 BiLSTM

A Bidirectional Long-Short Term Memory, or BiLSTM, was employed since it is a model in which processing is done in order. It comprises dual LSTMs, one is open to input in one direction, while the other is open to input in another. A Fully Connected neural network is made up of multiple fully connected layers that link all of the neurons in each layer to all of the neurons in the other layer [10].

Table 5. Parameter Setting of BiLSTM.

Parameter	Values
Epoch	10
Optimizer	Adam
Batch Size	128
Activation	Softmax
Loss	categorical_crossentropy

3.7. Word2Vec

For each word, the Word2Vec embedding approach only offers a single, independent embedding vector. Only one vector per word is saved by Word2vec in the output model. Although Word2vec is trained using contextual neighbours, a downstream NLP task uses it without context. Since the representation is only kept as one vector per word! Therefore, stagnant in use. This restricts the ability to understand a word's meaning across two contexts and used in two different situations for example. "river bank" and "bank deposit", "apple macbook" and "apple as a fruit" or "Python" as a programming language and "python" as a snake.

Table 6. Parameter Setting of Word2vec with CNN, RNN, BiLSTM.

Parameter	Values
Learning Rate	1e-5
Epoch	10
Optimizer	Adam
Batch Size	32
Activation	Softmax
Loss	categorical_crossentropy

3.8. BERT

Natural Language Processing (NLP) pre-trained model named Bidirectional Encoder Representations from Transformers (BERT) which uses transfer learning and has its architecture based on transformer model. BERT was developed by Google AI in 2018 to help machines understand the context of sentences [4].

BERT has been trained on two tasks: masked language tasks, in which sentences are fed to the model and some words are masked or hidden for the model, and the model attempts to predict these hidden words, and unmasked language tasks, in which sentences are fed to the model and some words are masked or hidden for the model, and unmasked language tasks, in the other task is sentence prediction where pair of sentences are fed to the model each round and the model need to predict whether one sentence is followed by the other or not. BERT has been trained over a large dataset for these two tasks. The dataset contained all English Wikipedia and 11,038 books. BERT uses an encoder from a transformer model, a type of neural network that takes a sentence as input to the model. Then each word of the sentence is tokenised, and these tokenised words are fed to the BERT model. BERT output is a vector representation for each tokenised word.

Using encoders from Transformer enables BERT to have a better context understanding than traditional neural networks such as LSTM or RNN since the encoder process all inputs, which is the whole sentence, simultaneously so when building a context for a word, BERT will take into account the inputs before it and also the inputs after the word, while the LSTM or RNN process the input taking in account only the prior inputs, and that will be reflected on the output vector value for the word, so the word “python” in the two sentences (I just started learning Python) and (Python are found in part of Africa and Asia) would have the same vector value -and as a result the same meaning- when using LSTM or RNN, on the other hand, it would have two different vectors using BERT, so as a result, using BERT will in most cases give us better performance than using the traditional machine learning algorithms. The BERT comes in two forms which are the BERT base and the BERT large. The BERT base consist of 12 encoders with a hidden size of 768 and the BERT large has 24 encoders with hidden size of 1024. The study employed the BERT base.

Table 6. Parameter setting for BERT with CNN, RNN, BiLSTM

Parameter	Values
Learning Rate	1e-5
Epoch	10
Optimizer	Adam
Batch Size	128
Activation	Softmax
Loss	sparse_categorical_crossentropy

4. Results and Discussion

4.1. Performance indicators

Precision is a measure of correctness that explains how many total positive predictions are positive. It is determined by dividing the total number of predicted positives by the total number of classified positives. The precision level should be high for a well-performed model. Precision is defined as follows:

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FP}), \quad (1)$$

where TP is true positive and FP is false positive.

A recall is the ratio of all positively classified classes that were correctly identified to all positively classified classes or the number of classes with a positive outcome that are correctly predicted. A good model should have a high recall rate. Recall is defined as follows:

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

where FN is false negative.

A high F1-score indicates high precision and recall because the score contains information about these two variables. It is defined as follows:

$$\text{F1} = (2 * \text{Precision} * \text{Recall})/(\text{Precision}+\text{Recall}). \quad (3)$$

Mean absolute error describes the discrepancy between actual and anticipated values. As the value goes down, the model's performance gets better. An ideal predictor of the outputs is a model with a mean absolute error of zero.

The average of the square of the difference between the data's original and anticipated values is used to calculate the mean square error. As the value goes down, the model's performance gets better.

The standard deviation of the errors that result from making a prediction on a dataset is known as the root mean square error. However, when determining the model's accuracy, the value's root is considered. As the value goes down, the model's performance gets better.

4.2. Experimental results

As shown in Table 7, the summary of the accuracy, precision, recall and F1 score of all the model considered in the study

Table 7. Model performance summary.

Models	Accuracy	P	P	P	R	R	R	F1	F1	F1
		NEG	NEU	POS	NEG	NEU	POS	NEG	NEU	POS
CNN	89%	0.88	0.88	0.90	0.81	0.90	0.93	0.84	0.89	0.91
RNN	90%	0.89	0.89	0.92	0.86	0.90	0.94	0.87	0.90	0.93
BiLSTM	90%	0.87	0.89	0.93	0.88	0.90	0.92	0.88	0.89	0.93
Word2Vec-CNN	57%	0.50	0.60	0.60	0.45	0.53	0.70	0.47	0.56	0.64
Word2Vec-RNN	48%	0.50	0.47	0.48	0.30	0.40	0.67	0.37	0.43	0.56
Word2Vec-BiLSTM	55%	0.50	0.56	0.56	0.26	0.61	0.71	0.34	0.58	0.63
BERT-CNN	93%	0.92	0.93	0.95	0.92	0.92	0.95	0.92	0.92	0.95
BERT-RNN	93%	0.90	0.93	0.95	0.93	0.91	0.95	0.92	0.92	0.95
BERT-BiLSTM	93%	0.91	0.93	0.96	0.93	0.92	0.95	0.92	0.92	0.95

As shown in Table 8, below the Mean absolute error, mean squared error and root mean square error of all the model considered in the study.

Table 8. Error evaluation.

Models	Mean Absolute Error	Mean Squared Error	Root Mean Square Error
CNN	0.1531	0.2305	0.4801
RNN	0.1253	0.1815	0.4260
BiLSTM	0.1242	0.1764	0.4200

Word2Vec-CNN	0.2822	0.2822	0.5313
Word2Vec-RNN	0.3456	0.3456	0.5879
Word2Vec-BiLSTM	0.2963	0.2963	0.5443
BERT-CNN	0.0796	0.1078	0.3283
BERT-RNN	0.0825	0.1133	0.3366
BERT-BiLSTM	0.0824	0.1128	0.3358

The confusion matrix of CNN indicates that 9,261 out of 11,389 negative sentiment were correctly classified, while 12,673 out of 14,158 neutral sentiment were correctly classified and 15,710 out of 16,962 positive sentiment were correctly classified as shown in Figure 4 below.

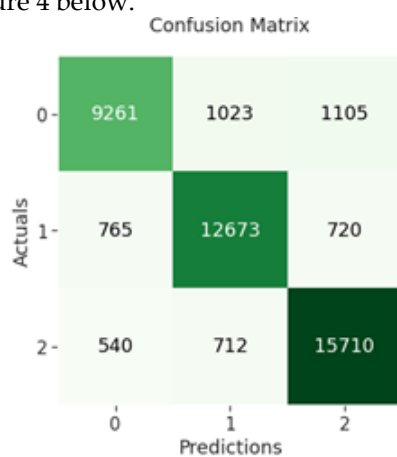


Figure 4. CNN confusion matrix.

The confusion matrix of CNN indicates that 9,751 out of 11,389 negative sentiment were correctly classified, while 12,737 out of 14,158 neutral sentiment were correctly classified and 15,892 out of 16,962 positive sentiment were correctly classified as shown in Figure 5 below.

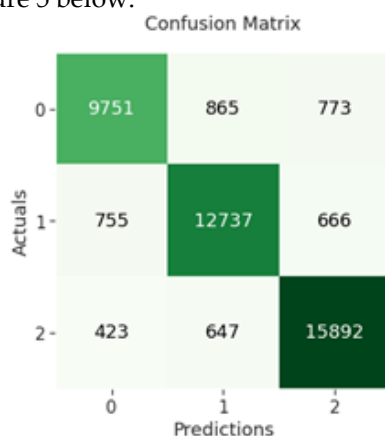


Figure 5. RNN confusion matrix.

The confusion matrix of BiLSTM indicates that 9,847 out of 11,221 negative sentiment were correctly classified, while 12,816 out of 14,272 neutral sentiment were correctly classified and 15,700 out of 17,040 positive sentiment were correctly classified as shown in Figure 6 below.

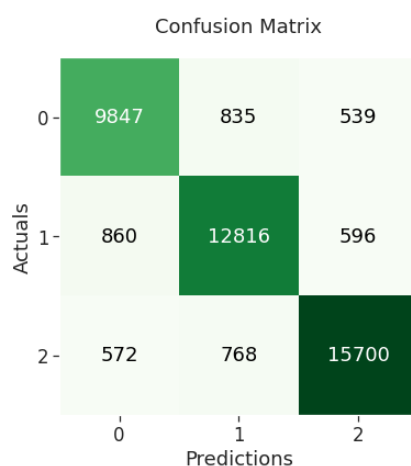


Figure 6 BiLSTM confusion matrix

The confusion matrix of Word2Vec-BiLSTM indicates that 2,901 out of 11,132 negative sentiment were correctly classified, while 8,715 out of 14,393 neutral sentiment were correctly classified and 12,000 out of 16,984 positive sentiment were correctly classified as shown in Figure 7 below.

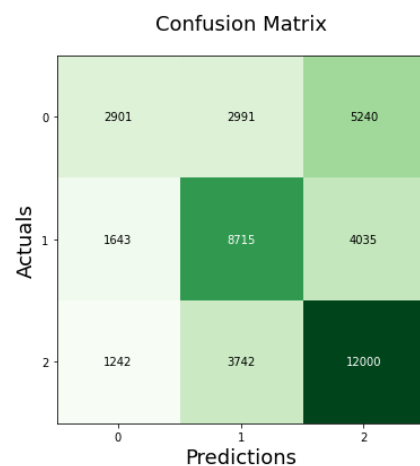


Figure 7 Word2Vec-BiLSTM confusion matrix

The confusion matrix of BERT-CNN indicates that 10,348 out of 11,221 negative sentiment were correctly classified, while 13,145 out of 14,272 neutral sentiment were correctly classified and 16,254 out of 17,040 positive sentiment were correctly classified as shown in Figure 8 below.

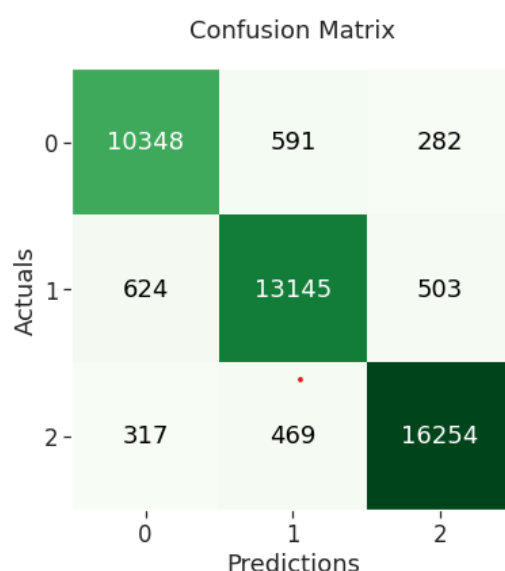


Figure 8 BERT-CNN confusion matrix

The confusion matrix of BERT-RNN indicates that 10,478 out of 11,221 negative sentiment were correctly classified, while 13,003 out of 14,272 neutral sentiment were correctly classified and 16,199 out of 17,040 positive sentiment were correctly classified as shown in Figure 9 below.

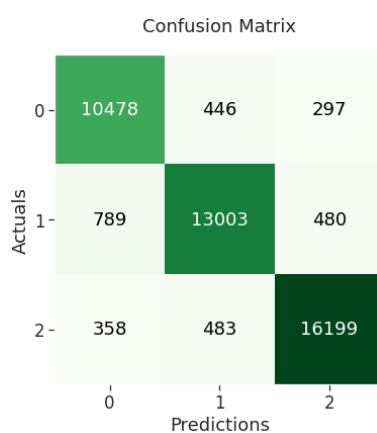


Figure 9 BERT-RNN confusion matrix

The confusion matrix of BERT-BiLSTM indicates that 10,389 out of 11,221 negative sentiment were correctly classified, while 13,097 out of 14,272 neutral sentiment were correctly classified and 16,183 out of 17,040 positive sentiment were correctly classified as shown in Figure 10 below.

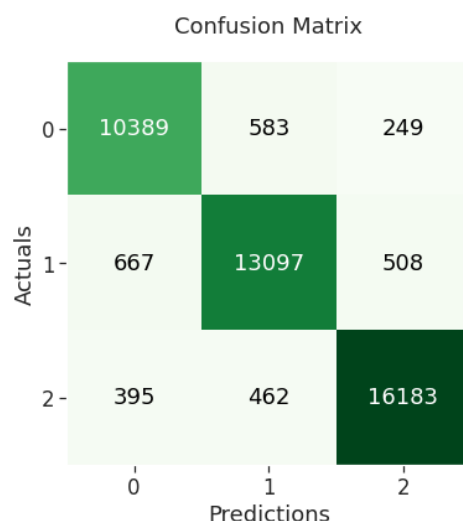


Figure 10 BERT-BiLSTM confusion matrix

5. Conclusions

The traditional approach of natural language processing (NLP) with the use of Word2Vec, CNN, RNN and BiLSTM has few limitation of not capturing the deeper context of the word. The BERT has more understanding than traditional since the encoder process all inputs, which is the whole sentence, simultaneously so when building a context for a word, BERT will take into account the inputs before it and also the inputs after the word, while the Word2Vec restricts the ability to understand a word's meaning across two contexts and used in two different situations which in turn will be reflected on the output vector value for the word.

The combination of the transformer model BERT with CNN, RNN, BiLSTM gives a state of the art performance in terms of the accuracy, recall, and precision.

Further works can be carried out to analyse sentiment on more data that is not sourced online because some information shared online can be shared by tourist or people with little or no understanding about the subject matter, also instead of sentiment, emotion like happy, sad, surprise can also be understudy. Other transformer model like RoBERTa can also be further investigated in further study.

Supplementary Materials: Not applicable.

Author Contributions: Conceptualization, S.-C.N. and M.-F.L.; methodology, A.B.; software, A.B.; validation, M.-F.L., S.-C.N. and A.B.; formal analysis, M.-F.L.; investigation, S.-C.N.; resources, A.B.; data curation, A.B.; writing—original draft preparation, A.B.; writing—review and editing, S.-C.N. and M.-F.L.; visualization, A.B.; supervision, S.-C.N.; project administration, S.-C.N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found in the following URLs:

https://www.kaggle.com/datasets/saurabhshahane/twitter-sentiment-dataset?select=Twitter_Data.csv

<https://www.kaggle.com/datasets/seriousran/appletwittersentimenttexts?select=apple-twitter-sentiment-texts.csv>

<https://www.kaggle.com/datasets/surajkum1198/twitterdata?select=finalSentimentdata2.csv>

<https://www.kaggle.com/datasets/yasserh/twitter-tweets-sentiment-dataset?select=Tweets.csv>

<https://www.kaggle.com/code/toygarr/contextual-model-and-crawling-for-real-tweets/data?select=train.csv>

<https://www.kaggle.com/code/toygarr/contextual-model-and-crawling-for-real-tweets/data?select=test.csv>.

Acknowledgments: The authors would also like to thank the anonymous reviewers for their insightful comments.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. J. P. U. S. D. Jayakody and B. T. G. S. Kumara, "Sentiment analysis on product reviews on twitter using Machine Learning Approaches," 2021 International Conference on Decision Aid Sciences and Application (DASA), 2021, pp. 1056-1061
2. Pham, T., Vo, D., Li, F., Baker, K., Han, B., Lindsay, L., Pashna, M. and Rowley, R., 2020. Natural language processing for analysis of student online sentiment in a postgraduate program. *Pacific Journal of Technology Enhanced Learning*, 2(2), pp.15-30.
3. Lamba, M. and Madhusudhan, M., 2021. Sentiment Analysis. *Text Mining for Information Professionals*. pp.191-211.
4. Devlin J., M.-W. Chang, K. Lee and K. Toutanova, 2018 BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding.
5. Durrheim, K., Schuld, M., Mafunda, M. and Mazibuko, S., 2022. Using word embeddings to investigate cultural biases. *British Journal of Social Psychology*.
6. Halbert, J., and Kaser, L. 20018. Deep learning: Inquiring communities of practice. Education Canada.
7. Jain, A., Kulkarni, G. and Shah, V., 2018. Natural Language Processing. *International Journal of Computer Sciences and Engineering*. 6(1), pp.161-167.
8. Jayakody, J.P.U.S.D. and Kumara, B.T.G.S. (2021) "Sentiment analysis on product reviews on Twitter using machine learning approaches," 2021 International Conference on Decision Aid Sciences and Application (DASA) [Preprint]. Available at: <https://doi.org/10.1109/dasa53625.2021.9682291>.
9. Mishra, R.K., Reddy, G.Y.S. and Pathak, H. (2021) The understanding of Deep Learning: A Comprehensive Review, *Mathematical Problems in Engineering*. Hindawi. Available at: <https://www.hindawi.com/journals/mpe/2021/5548884/> (Accessed: December 15, 2022).
10. Mahajan, P., (2022) Fully Connected vs Convolutional Neural Networks. [online] Medium.
11. Khurana, D., Koli, A., Khatter, K. et al. Natural language processing: state of the art, current trends and challenges. *Multimed Tools Appl* (2022). <https://doi.org/10.1007/s11042-022-13428-4>
12. Chiorrini, A., Diamantini, C., Mircoli, A., and Potena D., "Emotion and sentiment analysis of tweets using BERT," in EDBT/ICDT Workshops, Nicosia, Cyprus, 2021.
13. Dai, C., Che, H. and Leung, M. F. A neurodynamic optimization approach for L1 minimization with application to compressed image reconstruction. *International Journal on Artificial Intelligence Tools*. 2021, 30(01), 2140007.
14. Sharma, N. and Jain, V., 2020. Evaluation and Summarization of Student Feedback Using Sentiment Analysis. *Advances in Intelligent Systems and Computing*, pp.385-396.
15. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A., Kaiser, L. and Polosukhin, I., 2017. Attention is all you need.
16. Vateekul, P. and Koomsubha, T., 2019. A study of sentiment analysis using deep learning techniques on Thai Twitter data. 2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE).
17. Zhang, L., Wang, S. and Liu, B., 2018. Deep learning for sentiment analysis: A survey. *WIREs Data Mining and Knowledge Discovery*. 8(4).
18. Zorn, J., 2020. Deep Learning for NLP: Word Embeddings. [online] Medium. Available at: <https://towardsdatascience.com/deep-learning-for-nlp-word-embeddings-4f5c90bcdab5>.

19. Buhalis, R. and Leung, R., 2018. Smart hospitality — Interconnectivity and interoperability towards an ecosystem. *International Journal of Hospitality Management*, [e-journal] 71, pp.41-50.
20. Leung, M. F., Wang, J. and Che, H. Cardinality-constrained portfolio selection via two-timescale duplex neurodynamic optimization. *Neural Networks*. 2022, 153, 399-410.
21. Leung, M. F., Wang, J. and Li, D. Decentralized robust portfolio optimization based on cooperative-competitive multiagent systems. *IEEE Transactions on Cybernetics*. 2022, 52(12), pp. 12785-12794. DOI: 10.1109/TCYB.2021.3088884.
22. Chen, K., Che, H., Li, X. and Leung, M. F. Graph non-negative matrix factorization with alternative smoothed L0 regularizations. *Neural Computing and Applications*. 2022, 1-15. DOI: 10.1007/s00521-022-07200-w.
23. Che, H.; Wang, J.; Cichocki, A. Bicriteria sparse nonnegative matrix factorization via two-timescale duplex neurodynamic optimization. *IEEE Transactions on Neural Networks and Learning Systems*. In press. DOI: 10.1109/TNNLS.2021.3125457.
24. Chakravarthi, B., Ng, S. C., Ezilarasan M. R. and Leung, M. F. EEG-based emotion recognition using hybrid CNN and LSTM classification. *Frontiers in Computational Neuroscience*. 2022, 16, 1019776. DOI:10.3389/fncom.2022.1019776
25. Cao, W., Zhang, Z., Liu, C., Li, R., Jiao, Q., Yu, Z. and Wong, H. S. Unsupervised discriminative feature learning via finding a clustering-friendly embedding space. *Pattern Recognition*. 2022, 129, 108768.
26. Wang, J., Wang, J. and Han, Q. L. Neurodynamics-based model predictive control of continuous-time under-actuated mechatronic systems. *IEEE/ASME Transactions on Mechatronics*. 2021, 26(1), 311-322.
27. Cao, W., Qian, S., Wu, S. and Wong, H. S. Unsupervised multi-task learning with hierarchical data structure. *Pattern Recognition*. 2019, 86, 248-264.
28. Duan, Y., Chen, N., Bashir, A. K., Alshehri, M. D., Liu, L., Zhang, P. and Yu, K. A Web Knowledge-Driven Multimodal Retrieval Method in Computational Social Systems: Unsupervised and Robust Graph Convolutional Hashing. *IEEE Transactions on Computational Social Systems*. In press. DOI: 10.1109/TCSS.2022.3216621.
29. Mikolov T, Grave E, Bojanowski P, et al. Advances in pre-training distributed word representations. *ArXiv:1712.09405 [Cs]*. December 26, 2017.
30. Suhasini, M., and Srinivasu, B., Emotion detection framework for twitter data using supervised classifiers, in *Data Engineering and Communication Technology*, Singapore, Springer, pp. 565–576, 2020.
31. Bhagat, C., and Mane, D., "Text categorization using sentiment analysis," in *Proceeding of Int. Conf. on Computational Science and Applications*, Singapore, Springer, 2020
32. H. Huang, X. Y. Jing, F. Wu, Y. F. Yao et al., "DCNN-Bigru text classification model based on BERT embedding," in 2019 IEEE Int. Conferences on Ubiquitous Computing & Communications (IUCC) and Data Science and Computational Intelligence (DSCI) and Smart Computing, Networking and Services (SmartCNSShenYang, China, pp. 632–637, 2019.
33. *Kepios*. Available at: <https://kepios.com/> (Accessed: December 25, 2022).
34. C. Comito, D. Falcone and D. Talia, "A Peak Detection Method to Uncover Events from Social Media," 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2017, pp. 459-467, doi: 10.1109/DSAA.2017.69.
35. C. Comito, C. Pizzuti and N. Procopio, "Online Clustering for Topic Detection in Social Data Streams," 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI), 2016, pp. 362-369, doi: 10.1109/ICTAI.2016.0062.