Proceedings of the Pakistan Academy of Sciences: A Physical and Computational Sciences 61(2): 153-165 (2024)

ISSN (Print): 2518-4245; ISSN (Online): 2518-4253
http://doi.org/10.53560/PPASA(61-2)870

Pakistan Academy of Sciences

Research Article

Sentiment Analysis using Bidirectional Encoder Representations from Transformers

Adil Rehman[1], Khushal Das[2], Kamlish[3*], and Fazeel Abid[4]

[1]Department of Information Systems, Dr. Hasan Murad School of Management, University of Management and Technology, Lahore, Pakistan
[2]Department of Computer Science, Modelling, Electronics and Systems Engineering, University of Calabria, Rende, Italy
[3]Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Islamabad, 44000, Pakistan
[4]Department of Information Technology and Computer Science, University of Lahore, Pakistan

Abstract: In the contemporary digital landscape, a significant volume of data is generated through social networks such as Twitter, Facebook, and Instagram. This study presents a method for extracting sentiments from Twitter, focusing on two sentiment-based datasets: the Twitter and emotional sentiments datasets. After extraction and preprocessing, we employed three deep learning models: Recurrent Neural Networks (RNNs), Bidirectional Long Short-Term Memory (BiLSTM), and a pre-trained Bidirectional Encoder Representations from Transformers (BERT) model. We introduced Se-BERT, a model designed for emotional sentiment analysis. Our experiments showed that Se-BERT achieved accuracy levels of 97.29% for tweet sentiments (positive and negative) and 86.77% for emotional sentiments (joy, sadness, love, fear, anger, surprise). These results demonstrate that Se-BERT outperforms RNN and BiLSTM in terms of accuracy for sentiment analysis, thereby significantly enhancing information retrieval and providing a deeper understanding of user behaviour.

Keywords: Sentiment Analysis, Tweets Dataset, BiLSTM, BERT Model.

1. INTRODUCTION

The internet is crucial in modern life, enabling information retrieval, communication, and business activities. It features human-created content—like blogs, stories, and tweets, which provide expressive platforms for sharing thoughts—and machine-generated content, such as automated news and chatbots [1]. While human-generated content fosters personal expression and connections, machine-generated content is often seen as impersonal, lacking the human touch, and unable to comprehend emotions [2]. Differentiating between these two types of content can be challenging, especially in online interactions requiring vigilance. Additionally, managing online content to ensure safety while protecting freedom of expression presents ongoing challenges, as it involves finding the right balance between oversight and openness. Despite challenges, the internet remains a powerful connection, information-sharing, and access tool. It's essential to recognise the limitations of machine-generated content and approach online interactions with caution and critical thinking, especially as technology becomes increasingly embedded in our daily lives [3].

Understanding people's feelings is essential for fostering positive outcomes and mitigating negative ones. Sentiment analysis, a branch of Natural Language Processing (NLP), automates the detection of emotions in text, categorising them as positive or negative and even identifying specific emotions like joy or sadness [4]. This process involves converting text into a numeric vector format, using American Standard

Received: March 2024; Revised: May 2024; Accepted: June 2024
* Corresponding Author: Kamlish <kamlish@outlook.com>

Code for Information Interchange (ASCII) or Unicode, which encapsulates the expressive information contained within the text. However, the challenge lies in the imperfections of models due to insufficient training with diverse datasets, impacting optimal performance [5]. Our study employs neural networks, including Recurrent Neural Networks (RNN) and advanced models like Bidirectional Long Short-Term Memory (BiLSTM) and Sentiment-Bidirectional Encoder Representations from Transformers (Se-BERT), to analyse textual data and extract insights on how people feel about various topics or products [2]. This research introduces two primary approaches for sentiment analysis: emotion detection and NLP-based techniques.

Further, sentiment analysis has garnered significant attention due to its diverse applications in financial markets, social media, and human-computer interaction. Financial Sentiment Analysis (FSA) has evolved with two main research streams focusing on developing advanced techniques and applying them to market applications, highlighting the importance of understanding market sentiment and its relationship with investor sentiment [6]. Similarly, the advent of Web 4.0 has further expanded the applications of sentiment analysis across various domains, utilising machine learning and deep learning approaches to analyse users' emotions and generate insights for businesses, governments, and researchers [7]. Moreover, emotion detection has become crucial for enhancing human-computer interactions, with machine learning techniques such as SVM and Naive Bayes being widely used to analyse textual data and evaluate performance based on accuracy [8]. Our study introduces Se-BERT, a model designed for sentiment analysis using BERT. By providing significant insights into user behaviours and enhancing information retrieval capabilities, Se-BERT performs better in analysing tweets and emotional sentiments. By integrating the methodologies and insights from these seminal works, our research contributes to the growing body of knowledge in sentiment analysis and its practical applications in various fields.

Emotion Detection: This sentiment analysis method uses machine learning algorithms to identify and analyse emotions in text, categorising them into emotions like happiness, sadness, anger, or fear. It's beneficial for analysing unfiltered expressions on social media [9]. Emotion detection in texts, like tweets about a brand, identifies sentiment—positive, negative, or neutral—and guides marketing and product strategies.

Sentiment Analysis using NLP: This approach uses NLP techniques to categorise and assess text sentiment. NLP, a subfield of AI, analyses interactions between computers and human language [10]. Part-of-speech tagging, dependency parsing, and named entity recognition help identify critical phrases and sentiments in texts. For instance, analysing a product review could reveal critical features and associated feelings, which are helpful for marketing insights.

Advanced methods like Long Short-Term Memory (LSTM) with attention mechanisms enhance complex data analysis, such as noisy commercial time-series data [11]. These models focus on important textual information and integrate emotional data, helping analyse user personality and sentiment on platforms like social media [12]. This approach is crucial across various fields, from market analysis to customer feedback on social media, where understanding sentiment can guide business strategies and product development [13]. Like all machine learning models, sentiment categorisation models require data to be transformed into a standardised vector format, using ASCII or Unicode, to encapsulate textual meaning. This transformation is crucial for models like BERT, which rely on pre-trained language models for fine-grained sentiment analysis [14]. These NLP models and deep learning architectures have advanced significantly, enabling text conversion into machine-readable vector forms. These vectors represent text in an n-dimensional space, facilitating machine understanding through word embeddings [15]. In the stock market, sentiment analysis is increasingly used to gauge investor sentiment through online reviews and communications, with studies showing the BERT model outperforming the LSTM and SVM models [11]. During the COVID-19 lockdowns, sentiment analysis on Twitter became crucial for understanding political and social dynamics. In Nepal, sentiment analysis has been applied using multiple languages, with LSTM models providing the best results for analysing text and user personality from various social media sources [16]. Further development in this field includes an attention-based LSTM model that integrates sentiment and attention data to predict the personality traits of users based on their online interactions [12].

In a study, Jin *et al*. [11] predicted the final stock market price by incorporating stockholders' sentiments and utilizing empirical modal decomposition (EMD). They also enhanced the LSTM model with an attention mechanism to improve performance and reduce time delays in stock market predictions. The primary contribution lies in predicting prices, which directly influences the stock market by analyzing the sequence of stock prices and closing prices, and improving the LSTM mechanism for better accuracy. Chen *et al*. [17] introduced a novel approach to extracting target expressions from opinionated phrases. They employed a Bidirectional LSTM combined with a Conditional Random Field (CRF) layer (BiLSTM-CRF) for classification. The key research challenge is developing a deep learning model for document-level classification. Rhanoui *et al*. [18] addressed this by implementing a BiLSTM model that combines convolutional and bidirectional recurrent neural networks, using Doc2vec for document-level sentiment analysis. Basiri *et al*. [19] further explored deep learning sentiment analysis by investigating both BiLSTM and Bi-GRU architectures, achieving a maximum accuracy of 90%. Building on these efforts, Li *et al*. [20] advanced the field by leveraging the BERT pre-trained language model for End-to-End Aspect-Based Sentiment Analysis (E2E-ABSA), where BERT outperformed other state-of-the-art-models. Additionally, integrating Word2Vec and Glove embeddings with an LSTM-CRF model, alongside BERT, resulted in even better performance.

BERT, a pre-trained bidirectional and unsupervised language model developed by Google, serves as the foundation for contextual word representations in ABSA tasks. Hoang *et al*. [21] demonstrated that BERT outperformed previous single-sentence classification models. BERT classifies sentiments as positive, negative, neutral, or conflict using the Fine-grained Stanford Sentiment Treebank dataset [14]. Its transformer-based architecture and fixed-size word vector representations enable fine-grained sentiment classification. In Chinese stock reviews, BERT has been applied to predict sentiments from various online stockholder evaluations [22]. Despite the challenges of expanding labelled datasets for Chinese stock classifications, BERT has demonstrated improvements through fine-tuning with a BERT + FC model, which shows high efficacy and potential for broad application. BERT has also been employed to detect bots on social media by analysing sentiment features [23]. Its integration with glove for word embedding and comparison against models like Random Forest, SVM, and logistic regression has yielded superior results in identifying bot-driven content. The use of BERT extends to analysing tweets related to COVID-19 and identifying sentiments from global and India-specific Twitter data [15]. This application has shown enhanced performance in sentiment analysis from tweets. Additional points on sentiment analysis include its applications in financial predictions, social media monitoring, customer service, and healthcare. Challenges involve dealing with sarcasm and figurative language, requiring advanced detection techniques. The growth of multilingual sentiment analysis necessitates language-specific training [24]. Ethical concerns about privacy, data usage, and biases in machine learning models are critical considerations in sentiment analysis deployment.

Understanding how people feel in written text, known as sentiment analysis, doesn't always work the same for every type of writing. To improve it, we can try different ways to use different sets of writing. It's not just about the techniques we use; it's also about getting the text ready for analysis. We should test different models and methods to make sentiment analysis work well. We should also clean up the text, like removing common words and making everything lowercase. Other tricks, like picking out the most important words or creating new features from the text, can also help. Different types of writing, like social media posts or news articles, might need special treatment. Words can mean different things in different places, so we might need unique models for each type. Ultimately, improving sentiment analysis means trying many things and seeing what works best for different kinds of writing. We implemented three baseline models, RNN, BiLSTM, and BERT, and did a comparative analysis using two sentiment analysis-based datasets of 800000 and 16000 rows simultaneously. Later, we Proposed a Se- BERT model based on CNN layer and BERT embeddings and achieved the best results compared to baseline models.

Despite numerous studies focusing on sentiment analysis using various models, there remains a gap in achieving optimal accuracy and efficiency, particularly with large and diverse datasets. Se-BERT, derived from BERT, outperforms RNN and BiLSTM using advanced deep learning. Se-BERT excels in analysing tweet and emotional sentiment datasets, with impressive accuracy of 97.29% and 86.77%, effectively capturing nuanced emotional expressions in social media data.

2. METHODOLOGY

We have divided the proposed methodology into two parts: data description and proposed architecture. The data description and architecture are described below.

I. **Data Description:** It's all about datasets, along with the diagrams and tables related to them.
II. **Se-BERT Model - Architecture:** It's all about BERT and the parameters used in code and architecture.

BERT is a Bidirectional Encoder Representation from Transformers and is based on transformers. Transformers are designed to proceed with sequential input data such as translation or text summarisation, and transformers are encoder and decoder stacks. BERT, which Google's pre-trained model introduces. BERT is an encoder stack of transformers with two variations: BERT base (12 layers) and BERT large (24 layers). Since its beginnings, it has completed various NLP tasks with state-of-the-art accuracy. BERT's architecture can better understand the language, learn variation in data forms, and perform well crossways a variation of NLP tasks because it was trained on a vast text corpus. Due to its bidirectional nature, BERT collects data from both the left and right sides of a token's context through the training phase. BERT pre-training model based on transformer encoders is a series of layers used in high dimensional representation of an input sequence. The main feature of the Transformer encoder is the attention mechanism, which allows the model to attend to different parts of the input data when generating the representation. The attention mechanism allows the model to capture the relationships between other parts of the input data and to use this information to develop a more accurate picture. Each Transformer encoder layer has a self-attention mechanism to assess input relationships and a feed-forward network to process and produce the final representation.The output of one layer can be used as input to the next layer, allowing the model to create deeper and more complex expressions. BERT tokenizers are used in the preprocessing steps of transformer encoders. This tokenisation process helps to confirm that the input data is constant with the pre-training and fine-tuning of the BERT model, as discussed in Figure 1.

Due to the large volume of text it was trained on, BERT has shown the ability to perform well on NLP tasks. For tasks such as Se-BERT, performance has demonstrated the ability to increase with the assistance of advanced training on Twitter text, termed post-training. To accomplish the Se-BERT challenge, the post-training created Se-BERT as a question-and-answer exercise using the machine reading comprehension technique known as review reading comprehension. Solving Se-BERT as a classification task performed sequentially using the

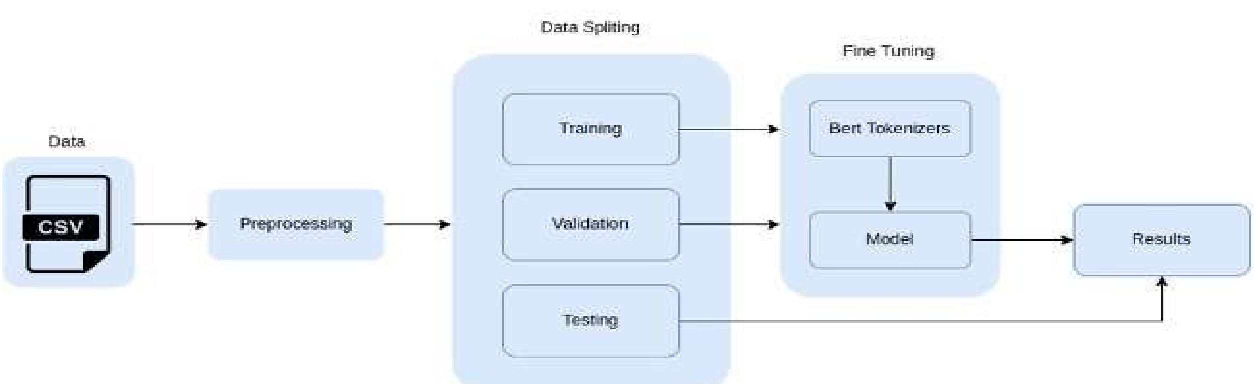

Fig. 1. Illustration of the Se-BERT model methodology, outlining data preprocessing, model architecture, and training phases.

BERT model by building auxiliary sentences has improved the results of the state-of-the-art model initialised as single-sentence classification. The architecture of transformers is based on attention mechanics; it determines the essential classifications for each computing step. In addition to projecting the input to a higher dimensional space vector, the encoder uses significant keywords as extra input to the decoder. It improves the decoder since it better understands the crucial sequences and the keywords that offer the context for the phrase. The BERT Tokenizer, a word tokenizer that turns words into tokens, is used first to handle the BERT model's text input. It yields a set of tokens, each of which denotes a word, along with two extra tokens: the set's opening addition is the classifier token [CLS]. When two groups of sentences are compared using BERT. Three distinct embedding layers with the exact dimensions are used to process this collection of tokens later. The sum of these layers is then sent to the encoder layer. Layers (Token Embedding Layer, Segment Embedding Layer, and Position Embedding) As shown in Figure 2, this is the complete architecture of the BERT Model. Still, using the BERT model, we need to apply for the single sentence sentiment analysis.

We have categorised the methodology into the data description and the proposed model architecture. Below is an explanation of the data definition and model architecture.

2.1. Data Description

We utilized two distinct sentiment analysis datasets for our implementation. The first dataset, the Sentiment Tweets dataset, contains 800,000 tweet records. The second dataset, the Emotional Sentiment dataset, includes 16,000 records categorized into six different emotions: fear, anger, surprise, sadness, love, and joy, as shown in Figure 3. By selecting the datasets provided by KanAnova [25] and Praveen [26], we implemented three models and analysed the results using the Se-BERT model.

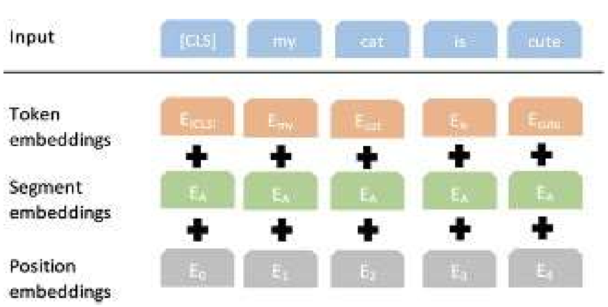

Fig. 2. Representation of BERT model inputs, including token, segment, and positional embeddings.

2.2. Baseline Methods

2.2.1. Recurrent neural networks (RNN)

RNNs are an essential subset of neural networks usually utilised in natural language processing. They belong to a type of neural network that, while maintaining hidden states, permits the use of prior outputs as inputs. RNN is a conceptual part of memory and stores all textual-related data. RNNs are appropriately titled because they constantly complete the same task for every input in a sequence, with the results dependent on earlier calculations. In the RNN diagram, the input layer' x' is sent to the neural network and passed to the middle layer 'h', which contains multiple hidden layers. Then, it gets a response in the output layer 'y'. In this diagram, A, B, and C are the parameters used to improve the result of the RNN model, and 't' is the current state used in hidden layers to fetch back to improve results on behalf of recent developments, as shown in Figure 4. We implemented an RNN using two datasets. For the first sentiment analysis dataset, which contains 800,000 sentiment records labelled as positive or negative, we pre-processed the data by removing stop words from the sentences. We then analysed the labelled data using a word cloud. After this analysis, we implemented the Word2Vec

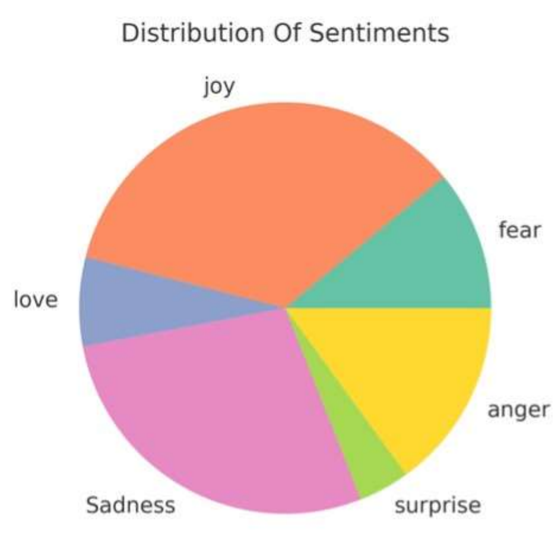

Fig. 3. Distribution of different emotional sentiments in the emotions' dataset.

model with 100 embedding layers. We used various hyperparameters, including the TensorFlow tokenizer with a vocabulary length of 60,000 and a maximum sentence length of 60. For training the RNN model, we used the Adam optimizer, which is known for its effectiveness, and for the loss function, we used the mean squared error method.

2.2.2. *BiLSTM model*

A bidirectional Long Short-Term Memory (BiLSTM) is a type of RNN and classification processing model that includes two LSTMs, one receiving input data forward and the other receiving it backwards. By successfully expanding the network's information pool, BiLSTM improves the contextual availability of the algorithm. It works with two instructions since it also works with two hidden layers, so this is a central point of disagreement of LSTM [16], but BiLSTM has proved to be of good result in NLP. However, according to BiLSTM, it contains two directions: one takes the input to the forward direction, and the other takes the backward direction, as shown in Figure 5. BiLSTM [28] improves the availability of information and context availability to the algorithms. We implemented BiLSTM with two datasets of sentiment analysis and emotional sentiment analysis. In the first dataset of sentiment analysis, we analysed the data with sentiments labels of positive and negative. After conducting research, we implemented the Word2Vec model and initialised the embedding matrix with 100 embedding dimensions. We used different hyperparameters, including the TensorFlow tokenizer with a vocabulary length of 60,000, a maximum sentence length of 60, a batch size of 1024, and ten epochs to train BiLSTM model. we used Adam optimiser, known for its effectiveness, during the training process. For the loss function, we use the binary cross-entropy method.

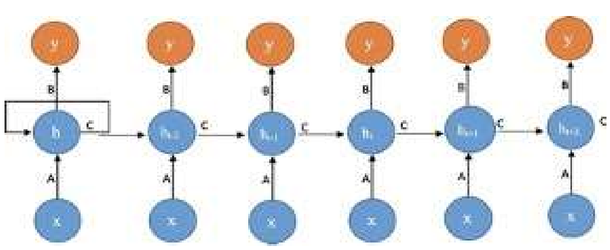

Fig. 4. The RNN model's structure shows input, hidden, and output layers.

2.3. Se-BERT Model Architecture

In the Se-BERT model, we used BERT transformer encoders and a 1D convolution layer (Conv1D) to analyse the best results. BERT is a Bidirectional Encoder Representation from Transformers and is based on transformers. Transformers are designed to proceed with sequential input data such as translation or text summarization. BERT pre-trained model, which is introduced by Google BERT Large. The model usages 24 layers, the hidden size is 1024, BERT's overall constraint size is 110 MB, Adam's optimiser is employed, and 2E-5 is used as the learning rate., the batch size of training and test set is 200. The epochs for training data are used as 6. BERT tokenizer is used to convert words into tokens. Then, we set the maximum length of the sentence to 60 according to our dataset sentence full length, and for the loss, we applied the entropy Loss function. Eight hours were used to train our data using Kaggle GPU, and we achieved 94. A Conv-1D, known as 1D CNN, is a convolutional neural network that operates on one-dimensional input data, such as a time series or a sequence of words. It is like a 2D CNN, commonly used for image processing tasks, but the main difference is that a 1D CNN linearly processes data. In a 1D CNN, the input data is run through several convolutional layers, each applying different filters. These filters are designed to detect specific patterns or features in the input data, such as changes in amplitude or frequency. The output of each convolutional layer is then passed through a pooling layer, which reduces the data's dimensionality and helps extract the most essential features. 1D CNNs can be used for various tasks, including natural language processing, speech recognition, and time series forecasting. They are particularly well suited to working with sequential data, where the order of the data is essential.

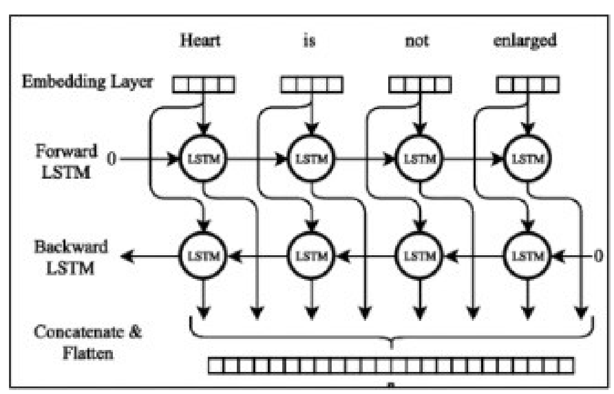

Fig. 5. BiLSTM model architecture [27].

The goal of an emotional dataset is to analyse a user's emotions based on their communication and tweets, particularly in the context of their interactions with other users. To achieve this, we used a BERT model combined with an Adam optimiser, which helps improve the training process's efficiency. One of the critical parameters we set was the maximum length of the input sequence, which we limited to 130 characters. It helped ensure that the model could process the input data quickly and accurately. We also set the learning rate to $5e^{-5}$, which helped ensure the model could learn from the input data effectively. To train the model, we used a loss function called categorical cross-entropy. This loss function is commonly used in multiclass classification tasks like this one, where we try to predict one of several emotional categories. We could ensure the model made accurate predictions by minimising the cross-entropy loss. Overall, the model's performance was excellent, with an accuracy of 97.29%. It means that the model could correctly predict the emotional category of the input data in nearly all cases. This level of accuracy is awe-inspiring, given the complexity of the dynamic dataset and the wide range of emotions it covers. One of the key advantages of using a BERT model for this type of task is its ability to analyse the context of the input data. Unlike other models that analyse individual words or phrases, BERT can consider the entire context of a sentence or communication, which helps improve the accuracy of predictions.

In addition to its accuracy, the BERT model is also highly flexible and can be adapted to work with a wide range of input data types. It is ideal for analysing emotions in various contexts, including social media, email communications, and more. Overall, the results of our analysis demonstrate the power and flexibility of the BERT model for analysing emotions in complex datasets. We accurately predicted emotional categories from user communications and tweets using this model with effective optimisation techniques and loss functions. Using the e-BERT model implementation, deep learning layers are used to achieve better performance; the architecture of my created model is defined below, as shown in Figure 6. One of the significant benefits of using a particular model for sentiment categorisation is that it can utilise both pre-labelled and unlabelled data, thanks to the capabilities of a tool called BERT. It is essential for social media networks, which generate vast text data every second. However, manually labelling this data is an incredibly time-consuming task that requires a significant amount of labour. To break it down further, sentiment categorisation is a task in which we try to determine the emotional tone of a text - for example, whether it expresses positive or negative sentiment. This task is becoming increasingly important as the amount of text data generated on social media platforms grows. One of the most popular tools used for sentiment categorisation is a model called BERT (Bidirectional Encoder Representations from Transformers). One of the key advantages of using BERT is that it can use pre-labelled and unlabelled data. This is significant because labelled data is data humans manually tag to indicate their sentiment, whereas unlabelled data has yet to be classified.

The ability to use unlabelled data is essential for social media networks, which generate vast amounts of data every second. Manually tagging this data is an incredibly labour-intensive operation that would be nearly impossible to keep up with. However, using BERT makes it possible to categorise this data automatically without manual labelling. The reason BERT can do this is because it is what is known as a pre-trained model. It means

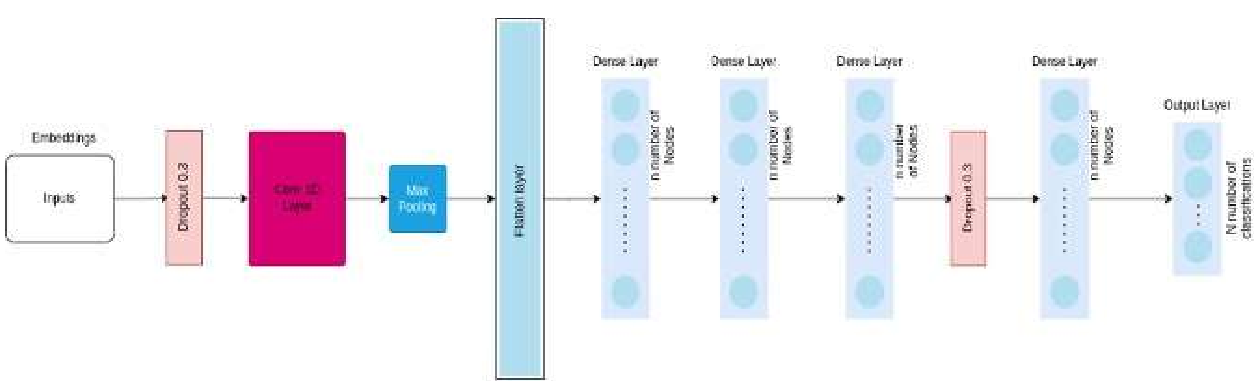

Fig. 6. Architecture of the Se-BERT model, combining BERT transformer encoders with a Conv1D layer.

it has already been trained on a large amount of text data and has learned to recognise specific patterns and features. When it encounters new, unlabelled data, it can use this existing knowledge to categorise it based on how it sees it. Pre-labelled data remains invaluable, as it provides a clear signal for the model to learn from. The ability to use unlabelled data sets BERT apart, making it a powerful tool for sentiment categorization. Automatically categorising sentiment in real-time text data is increasingly important for social media networks. BERT's ability to use pre-labelled and unlabelled data is a handy tool for this task. While manual data labelling is still essential, unlabelled data can significantly speed up the process and make it more efficient.

3. RESULTS AND DISCUSSION

In our sentiment analysis, we used two datasets with three deep learning models: RNN, BiLSTM, and BERT's pre-trained model. Among these, we achieved the best performance with the BERT base model, which is a part of NLP (Natural Language Processing). We used RNN, BiLSTM, and BERT pre-trained models as benchmarks and then initialized the Se-BERT model. For sentiment classification, the model takes a text input from natural language processing and outputs labels in an encoded numerical format (e.g., 0, 1). To achieve optimal performance, we applied multiple models and preprocessing techniques. We utilized the RNN and BiLSTM [16] models, with BiLSTM being an advanced version of the RNN deep learning model, and also employed the pre-trained BERT model (Bidirectional Encoder Representations from Transformers) for sentiment predictions. For the preprocessing and execution of baseline models, we evaluated the results using classification reports that included metrics such as Precision, Recall, and F1-score.

Precision is a metric that measures the accuracy of the model's predictions. It is defined as the ratio of correct sentiment predictions (True Positives, TP) to the total number of sentiment predictions (True Positives + False Positives, TP + FP) [29].

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

Recall computes the completeness of a model, which is defined as the ratio of correct sentiment prediction (TP) and the total number of actual sentiment predictions (TP + FN).

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The F-score (the F1-score or F-measure) represents the harmonic average of precision and recall and is usually used to optimise a model towards either precision or recall.

$$Recall = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

3.1. Baseline Methods

3.1.1. Recurrent neural networks (RNN)

After implementing the model, we achieved 82% accuracy on the tweet sentiment dataset. The sentiment analysis classification report for the RNN model's precision, recall, and F1-score is presented in Figure 7, and all results are displayed in Table 1. Precision measures the accuracy of the model, defined as the proportion of correct sentiment predictions (True Positives, TP) to the total number of sentiment predictions (True Positives + False Positives, TP + FP). The second dataset belongs to users' emotions, which depend on their behaviour.

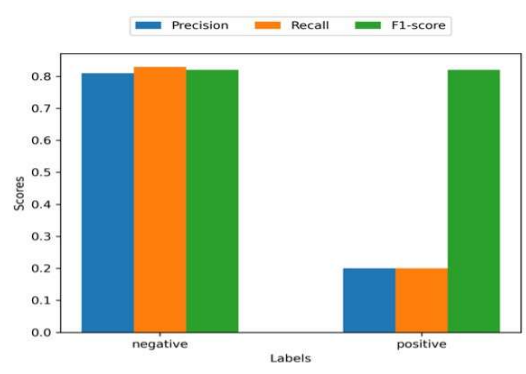

Fig. 7. Classification report for RNN on tweets sentiment dataset, showing precision, recall, and F1-score

Table 1. RNN classification report on tweets sentiments.

Parameters	Precision	Recall	F1-score	Support
Negative	0.81	0.83	0.82	39989
Positive	0.20	0.20	0.82	40011
Accuracy	---	---	0.82	80000
Macro avg	0.07	0.17	0.82	80000
Weighted avg	0.13	0.34	0.82	80000

We analysed this emotional dataset. We received six emotions as a label for the user's sentences. We used glove embedding to initialise the embedding metrics with 300 embedding dimensions. We implemented a categorical cross-entropy loss function with Adam optimiser using this method. After implementing the RNN model, we achieved 73% accuracy in the emotional sentiment analysis dataset. The emotional sentiment analysis classification report for the RNN model's precision, recall, and F1-score is presented in Figure 8, with the complete details shown in Table 2.

3.1.2. BiLSTM model

BiLSTM, an advanced form of RNN, proceeds bidirectionally forward and backwards and performs well, then generates results that we achieve better than RNN, with 84.52% accuracy in the tweets sentiments dataset. We show sentiment analysis results of the BiLSTM model (precision, recall, and F1-score) in Figure 9 diagram and Table 3. In the second dataset, which belongs to users' emotions and depends on their behaviour, we analysed the BiLSTM model with the emotional sentiment dataset. We received six emotions as a label for the user's sentences. We used glove embedding to initialise the embedding metrics with 300 embedding dimensions. This method utilised the Adam optimiser with a 0.001 learning rate and categorical cross-entropy loss function. BiLSTM, which is an advanced form of RNN, proceeds bidirectional forward and backward and performs well, then generates good results that we achieve better than RNN, that is, 92.15% accuracy in the dataset of emotional sentiments, and we show the impact of BiLSTM model precision, recall, and F1-score in Figure 10 and Table 4. For contextual word representations, we implemented word embedding to translate text into the machine-readable vectorised format using Word2Vec embeddings; we used bidirectional LSTMs and Word2Vec embeddings and pre-trained glove

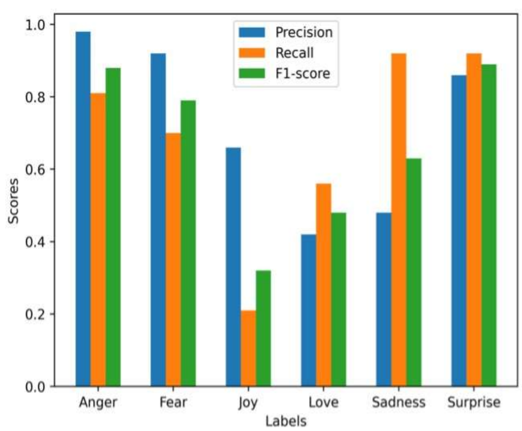

Fig. 8. Classification report for RNN on emotions sentiment dataset, showing precision, recall, and F1-score for each emotion.

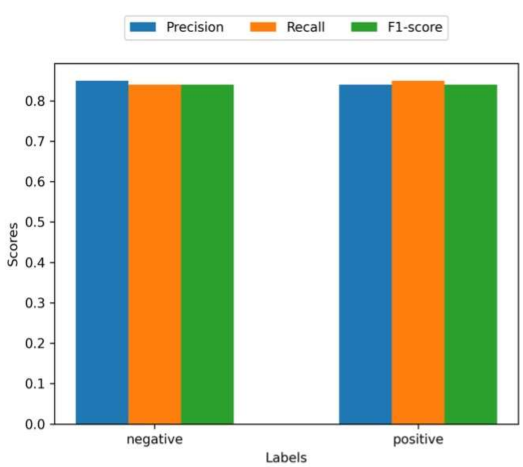

Fig. 9. Classification report for BiLSTM on tweets sentiment dataset, showing precision, recall, and F1-score.

Table 2. RNN classification report on emotions sentiments dataset.

Parameters	Precision	Recall	F1-score	Support
Anger	0.98	0.81	0.88	581
Fear	0.92	0.70	0.79	275
Joy	0.66	0.21	0.32	159
Love	0.42	0.56	0.48	66
Sadness	0.48	0.92	0.63	224
Surprise	0.86	0.92	0.89	695
Micro avg	0.79	0.79	0.79	2000
Macro avg	0.72	0.69	0.67	2000
Weighted avg	0.83	0.79	0.79	2000
Samples avg	0.79	0.79	0.79	2000

Table 3. BiLSTM classification report on tweets sentiments dataset.

Parameters	Precision	Recall	F1-score	Support
Negative	0.85	0.84	0.84	39989
Positive	0.84	0.85	0.84	40011
Macro avg	0.84	0.84	0.84	80000
Weighted avg	0.84	0.84	0.84	80000

Table 4. BiLSTM classification report on emotions sentiments dataset.

Parameters	Precision	Recall	F1-score	Support
Anger	0.97	0.96	0.97	581
Fear	0.94	0.93	0.93	275
Joy	0.74	0.92	0.82	159
Love	0.79	0.74	0.77	66
Sadness	0.88	0.90	0.89	224
Surprise	0.96	0.91	0.94	695
Macro avg	0.92	0.92	0.92	2000
Macro avg	0.88	0.90	0.89	2000
Weighted avg	0.93	0.92	0.92	2000
Sample avg	0.92	0.92	0.92	2000

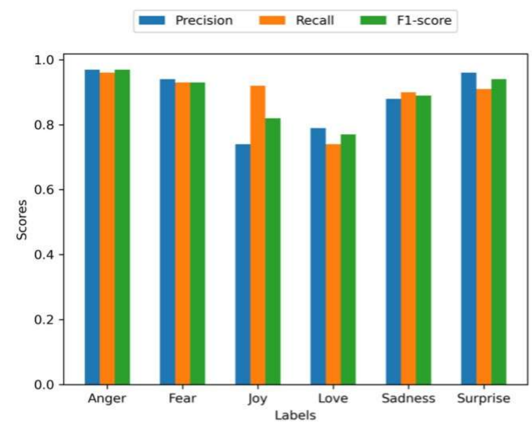

Fig. 10. Classification report for BiLSTM on emotions sentiment dataset, showing precision, recall, and F1-score for each emotion.

embeddings. We trained BiLSTM with embedding Matrixes are introduced and discussed by Yue and Li in [30]. We distributed data into categories and apply preprocessing to remove special characters or non-readable content. We don't need to clean tweets' emotion-based data and remove non-usable content. To achieve the best result, we used Adam Optimizer, which performs better, and for the loss function, we used binary cross-entropy loss function. Using these techniques, we approach our model accuracy at 84.52%, which is better. After applying BiLSTM, we used the BERT model based on the transformers [31] and achieved the best performance compared to BiLSTM.

3.2. Experimental Setup

3.2.1. Se-BERT setup

Every productivity element is connected to every input element by the BERT model, which is utilised for sentiment analysis. BERT is an established and pre-trained model based on transformers. BERT is the encoded process of transformer architecture. Pre-trained language models provide contextual data into tokens, representing the existence and representation of words from training data. BERT is considered the word from both the right and left side simultaneously. Although the idea is modest, it expands outcomes at numerous NLP tasks such as user reviews, sentiment examination, and question-and-answering systems when order is compared separately to training left and right. The input representation of the BERT can signify a sentence in a sequence of tokens [32]. The input representation of a particular token is created by adding up its relevant segments, tokens, and positional embeddings. For the implementation of classification tasks, the first word of each sequence is an exclusive classification embedding [33]. The Se-BERT model performance improves in both datasets, and the results are better than BiLSTM [28].

In sentiment analysis, accuracy measures the percentage of correctly classified positive tweets, while recall indicates the proportion of actual positives identified. The F score combines accuracy and recall, balancing precision and recall providing a comprehensive view of model performance. By analysing accuracy, recall, and F- score, we can better understand the strengths and weaknesses of a sentiment analysis model and make informed decisions about how to improve its performance. We trained three models in two sentiment analysis datasets for all the research and outcomes. We expected the models to improve the results after combining our dataset because there would be more features and emotional categories. The state-of-the-art model is BERT and uses transformers, which transform words into tokens and tokens captured by the corpus to find sentiments in a text. On behalf of both models, RNN and BiLSTM, we achieve the best result by using the BERT model, which is based on transformers and converts words into the embedding form [30].

Using the BERT model, we applied a deep learning technique to add three dense layers with one dropout proposed as a Se-BERT model. We achieved 86.77% and 97.29% accuracy in two datasets of tweets and emotional sentiments, respectively.

BERT uses positional word embeddings to generate distinct word embeddings for each word, depending on its position in the text, in contrast to the prior best models that generate one vector for each word [34]. Another factor could be using sentence-pair classification instead of the previous top models. Which relied on single sentence sorting to identify which feature was present in the text. These results are found using GPU, and we achieve our best result with 80,000 records of a dataset in 5 hours with three epochs using the Se-BERT model. The classification report is shown in Table 5 and Table 6. We show tweets' sentiment analysis results of the Se- BERT model precision, recall, and F1-score in Figure 11 and Table 7.

Table 5. Datasets accuracy on all models.

Datasets	Epochs	Models	Accuracy	Support
Tweets dataset	10	RNN	82%	581
	10	BiLSTM	84.52%	275
	6	Se-BERT	82%	159
Emotional Dataset	10	RNN	34%	66
	15	BiLSTM	92.15%	224
	20	Se-BERT	94.29%	695

Table 6. Se-BERT model classification report on emotional sentiments.

Parameters	Precision	Recall	F1-score	Support
Anger	0.92	0.94	0.93	813
Fear	0.90	0.88	0.89	712
Joy	0.96	0.95	0.96	2028
Love	0.86	0.86	0.86	492
Sadness	0.97	0.96	0.97	1739
Surprise	0.77	0.87	0.82	216
Accuracy	---	---	0.94	6000
Macro avg	0.90	0.91	0.90	6000
Weighted avg	0.94	0.94	0.94	6000

Table 7. Se-BERT classification report on tweets sentiments dataset.

Parameters	Precision	Recall	F1-score	Support
Negative	0.82	0.82	0.82	39989
Positive	0.82	0.81	0.82	40011
Accuracy	---	---	0.82	80000
Macro avg	0.82	0.82	0.82	80000
Weighted avg	0.82	0.82	0.82	80000

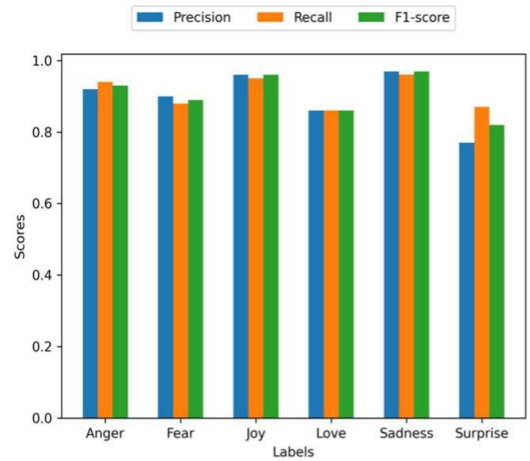

Fig. 11. The Classification report for Se-BERT on the sentiment's dataset shows precision, recall, and F1-score.

4. CONCLUSIONS

In conclusion, this study introduces Se-BERT, a novel sentiment analysis model derived from BERT. It evaluates performance against established RNN and BiLSTM models across emotional and tweet sentiment datasets. Se-BERT exhibits exceptional accuracy, scoring 97.29% for emotional sentiments and 94.84% for tweet sentiments. Sentiment analysis, a crucial aspect of natural language processing, seeks to identify and understand sentiments expressed in text. Machine learning models like RNN, BiLSTM, and BERT have propelled sentiment analysis research, and Se-BERT contributes as a specialised variant of the BERT model for enhanced sentiment analysis. The evaluation involves two datasets: tweets sentiments and emotional sentiments. Se-BERT surpasses RNN and BiLSTM, achieving a final score of 97.29% on the emotional sentiment dataset and 94.84% on the tweet sentiments dataset. These results underscore Se-BERT's effectiveness in deciphering nuanced emotional expressions in social media data. Beyond academic interest, sentiment analysis finds practical applications in social media monitoring, customer feedback analysis, and market research. With its capacity for training on extensive datasets and fine-tuning for specific tasks, Se-BERT emerges as a robust tool for these applications. Understanding user behaviour on platforms like Twitter is crucial, and sentiment analysis of social media data provides valuable insights into users' emotions, opinions, and attitudes. This research's comparative analysis of RNN, BiLSTM, and pre-trained BERT guides researchers and practitioners in selecting appropriate models. Se-BERT not only outperforms existing models but also offers a highly accurate method for emotional sentiment analysis. Future research could explore sentiment analysis applications in different contexts and develop new models for improved accuracy and efficiency. This study significantly advances sentiment analysis methodologies and their practical.

5. CONFLICT OF INTEREST

The authors declare no conflict of interest.

6. REFERENCES

1. E. Adamopoulou, and L. Moussiades. Chatbots: History, technology, and applications. *Machine Learning with Applications* 2: 100006 (2020).
2. K. Machova, M. Szaboova, J. Paralic, and J. Micko. Detection of emotion by text analysis using machine learning. *Frontiers in Psychology* 14: 1190326 (2023).
3. E. Cambria, and B. White. Jumping NLP curves: A Review of Natural Language Processing Research. *IEEE Computational Intelligence Magazine* 9(2): 48-57 (2014).
4. J.C. Pereira-Kohatsu, L. Quijano-Sanchez, F. Liberatore, and M. Camacho-Collados. Detecting and Monitoring Hate Speech in Twitter. *Sensors* 19(21): 4654 (2019).
5. F.M. Shiri, T. Perumal, N. Mustapha, and R. Mohamed. A Comprehensive Overview and Comparative Analysis on Deep Learning Models: CNN, RNN, LSTM, GRU. *Preprint ArXiv1*: 2305.17473 (2023).
6. K. Du, F. Xing, R. Mao, and E. Cambria. Financial Sentiment Analysis: Techniques and Applications. *ACM Computing Surveys* 56(9): 1-42 (2024).
7. U. Singh, K. Abhishek, and H.K. Azad. A Survey of Cutting-edge Multimodal Sentiment Analysis. *ACM Computing Surveys* 56(9): 1-36 (2024).
8. A. Alslaity, and R. Orji. Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. *Behaviour and Information Technology* 43(1): 139-164 (2024).
9. C.R. Sugimoto, S. Work, V. Lariviere, and S. Haustein. Scholarly use of social media and altmetrics: A review of the literature. *Journal of the Association for Information Science and Technology* 68(9): 2037-2062 (2017).
10. J.P. Bharadiya. A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. *European Journal of Technology* 7(1): 58-66 (2023).
11. Z. Jin, Y. Yang, and Y. Liu. Stock closing price prediction based on sentiment analysis and LSTM. *Neural Computing and Applications* 32: 9713-9729 (2020).
12. J. Zhao, Z. Dalin, Y. Xiao, L. Che, and M. Wang. User personality prediction based on topic preference and sentiment analysis using LSTM model. *Pattern Recognition Letters* 138: 397-402 (2020).
13. M. Li, W. Li, F. Wang, X. Jia, and G. Rui. Applying BERT to analyse investor sentiment in stock market. *Neural Computing and Applications* 33: 4663-4676 (2021).
14. M. Munikar, S. Shakya, and A. Shrestha. Fine-grained Sentiment Classification Using BERT. *Preprint ArXiv1*: 1910.03474 (2019).
15. M. Singh, A.K. Jakhar, and S. Pandey. Sentiment analysis on the impact of coronavirus in social life using the BERT model. *Social Network Analysis and Mining* 11: 33 (2021).
16. M. Tripathi. Sentiment Analysis of Nepali COVID19 Tweets Using NB SVM and LSTM. *Journal of Artificial Intelligence* 3: 151-168 (2021).
17. T. Chen, R. Xu, Y. He, and X. Wang. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems with Applications* 72: 221-230 (2017).
18. M. Rhanoui, M. Mikram, S. Yousfi, and S. Barzali. A CNN-BiLSTM model for Document-Level Sentiment Analysis. *Machine Learning and Knowledge Extraction* 1(3): 832-847 (2019).
19. M.E. Basiri, S. Nemati, M. Abdar, E. Cambria, and U.R. Acharya. ABCDM: An Attention-based Bidirectional CNN-RNN Deep Model for sentiment analysis. *Future Generation Computer Systems* 115: 279-294 (2021).
20. X. Li, L. Bing, W. Zhang, and W. Lam. Exploiting BERT for End-to-End Aspect-based Sentiment Analysis. *Preprint ArXiv*2: 1910.00883 (2019).
21. M. Hoang, O.A. Bihorac, and J. Rouces. Aspect-based sentiment analysis using bert. *Proceedings of the 22nd Nordic Conference on Computational Linguistics, Turku, Finland* (September 30 – October 02, 2019) pp. 187-196 (2019).
22. M. Li, L. Chen, J. Zhao, and Q. Li. Sentiment analysis of Chinese stock reviews based on BERT model. *Applied Intelligence* 51: 5016–5024 (2021).
23. M. Heidari, and J.H. Jones. Using BERT to Extract Topic-Independent Sentiment Features for Social Media Bot Detection. *2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New*

York, NY, USA, pp. 0542-0547 (2020).

24. E.F. Can, A. Ezen-Can, and F. Can. Multilingual Sentiment Analysis: An RNN-Based Framework for Limited Data. *Preprint ArXiv*: 1806.04511 (2018).
25. M.M. KazAnova. Sentiment140 dataset with 1.6 million tweets. *Kaggle*: (2017). https://www.kaggle.com/datasets/kazanova/sentiment140 (accessed 22 July 2023)
26. P. Praveen. Emotions dataset for NLP. *Kaggle*: (2020). https://www.kaggle.com/datasets/praveengovi/emotions-dataset-for-nlp (accessed 22 July 2023)
27. S. Cornegruta, R. Bakewell, S. Withey, and G. Montana. Modelling Radiological Language with Bidirectional Long Short-Term Memory Networks. *Preprint ArXiv*:1609.08409 (2016).
28. G. Xu, Y. Meng, X. Qiu, Z. Yu, and X. Wu. Sentiment Analysis of Comment Texts based on BiLSTM. *IEEE Access* 7: 51522–51532 (2019).
29. K. Das, M. Shehryar, F. Abid, M. Ashraf, M. Adil, S. Inam, E.U. Haq, and H. Mushtaq. Impact of using e-learning tools on Student's Psychological Health during covid-19. *VFAST Transactions on Software Engineering* 9(3): 120-127 (2021).
30. W. Yue, and L. Li. Sentiment Analysis using Word2vec-CNN-BiLSTM Classification. *Seventh International Conference on Social Networks Analysis, Management and Security (SNAMS), Paris, France* (December 14–16, 2020) pp. 1-5 (2020).
31. L. Zhao, L. Li, X. Zheng, and J. Zhang. A BERT based Sentiment Analysis and Key Entity Detection Approach for Online Financial Texts. *Preprint ArXiv*: 2001.05326 (2020).
32. Y. Song, J. Wang, Z. Liang, Z. Liu, and T. Jiang. Utilizing BERT Intermediate Layers for Aspect Based Sentiment Analysis and Natural Language Inference. *Preprint ArXiv*: 2002.04815 (2020).
33. A. Karimi, L. Rossi, and A. Prati. Adversarial Training for Aspect-Based Sentiment Analysis with BERT. *Preprint ArXiv*: 2001-11316 (2021).
34. S.M. Rezaeinia, R. Rahmani, A. Ghodsi, and H. Veisi. Sentiment analysis based on improved pre-trained word embeddings. *Expert Systems with Applications* 117: 139–147 (2019).