Social Network Analysis and Mining (2022) 12:55
https://doi.org/10.1007/s13278-022-00882-z

ORIGINAL ARTICLE

Detection of extreme sentiments on social networks with BERT

M. Luqman Jamil[1] · Sebastião Pais[1,2,4] · João Cordeiro[1,3] · Gaël Dias[4]

Received: 27 November 2021 / Revised: 18 April 2022 / Accepted: 20 April 2022

Abstract
Online social networking platforms allow people to freely express their ideas, opinions, and emotions negatively or positively. Previous studies have examined sentiments on these platforms to study their behavior in different contexts and purposes. The mechanism of collecting public opinion information has attracted researchers to automatically classify the polarity of public opinions based on the use of concise language in messages, such as tweets, by analyzing social media data. In this paper, we extend the preceding work where an unsupervised approach to automatically detect extreme opinions/posts in social networks is proposed. The performance of the proposed approach is evaluated on five different social network and media datasets. In this work, we use a semi-supervised approach known as BERT to reevaluate the accuracy of our prior approach and the obtained classified dataset. The experiment proves that in these datasets, posts that were previously classified as negative or positive extreme are extremely negative or positive in many cases while using BERT. Furthermore, BERT shows the capability to classify the extreme sentiments when fine-tuned with an appropriate extreme sentiments dataset.

Keywords BERT · Sentiment Analysis · Extreme Sentiment Analysis · Violent Extremism · Social Media · Social Networks

1 Introduction

Online social networks, such as Facebook, Twitter, Tumblr, and YouTube, have become a de-facto platform for hundreds of millions of Internet users to establish and maintain interpersonal relationships. In recent years, the emergence of microblogging services has greatly influenced the way people think, communicate, behave, learn, and conduct business. These popular social platforms are new forms of blogging that facilitate communication between people. By writing posts, sharing articles, videos, links, or tweeting messages, people make their own opinions, ideas, and thoughts, in a constructive or destructive tone (Persia and D'Auria 2017).

Most of the information published on social networks is harmless. It represents casual, conventional, or expressive crowds, as well as noisy data. However, any collection of tweets or posts with a focus on emerging controversial topics can pose a potential threat to individuals and society (Becker et al. 2011). Researchers and policymakers are still trying to discover the rise of violent extremism among people and take appropriate measures to prevent it. For example, the work of Krumm (2013) shows that the use of specifically radicalized language by people acting and protesting on social media may lead to violent extremism. In addition, terrorist organizations use social networks to study human sentiments by accessing uncensored content for collecting information of public views, by monitoring data from social networks, and automatically classifying the polarity of public sentiments upon using the concise language in posts or/and tweets. This

M. Luqman Jamil, Sebastião Pais, João Cordeiro and Gaël Dias have contributed equally to this work.

✉ Sebastião Pais
sebastiao@di.ubi.pt

M. Luqman Jamil
luqman.jamil@ubi.pt

João Cordeiro
jpcc@ubi.pt

Gaël Dias
gael.dias@unicaen.fr

[1] Department of Computer Science, University of Beira Interior, Covilhã, Portugal

[2] NOVA LINCS, New University of Lisboa, Lisboa, Portugal

[3] LIAAD, INESC TEC, Porto, Portugal

[4] GREYC, University of Caen Normandie, Caen, France

Published online: 22 May 2022

allows violent extremists to increase recruitment by being able to establish personal relationships with a global audience to their advantage (Scanlon and Gerber 2014).

Apart from terrorism, far-right extremism has been on the rise globally. It is important to track this trend to understand white supremacist movements, especially regarding their presence and their traces on the landscape. The study Dixit and Miller (2022) highlights the danger of white supremacist violence and racial injustices in the United States and its link with historical events. In Germany, the number of right-wing extremist crimes committed increased sharply in 2020 to a two-decade high which accounted for 23,064 criminal offenses (Politico 2021). This significant increase led Germany to see right-wing extremism as a top security threat (VOA 2020).

Early possible detection of extremist sentiments can help authorities identify and monitor the suspects to take appropriate measures for preventing any viable attacks. A prime example of such a case can be the 18-year Portuguese student who planned to carry out a “massacre” in his university. The main target of this planned attack was one of the professors at the university along with the indiscriminate killing of other people at the university. The attack was stopped before happening thanks to a member of the dark web group where the alleged attacker shared his intention, stating the imminent details to carry out the attack. The member of that group alerted the FBI, which on the same day alerted Portuguese authorities. The suspect was traced using his online footprint and detained a few days before the intended day of the attack (DN 2022). This incident shows the need for counter-extremism mechanisms for detecting such sentiments on social networks.

1.1 Problem definition

The detection and classification of extreme sentiments is a specific type of sentiment analysis. These sentiments are highly negative or positive about a particular subject, object, or person (Tanoli and Pais 2020). In a more general form, an extreme sentiment can be seen described as the worst or best opinion, judgment, or evaluation formed in one’s mind about a particular thing or person. Extremists use two forms of narratives; positive extreme for inciting people to carry out attacks or pursue them to believe in their agenda and Negative Extreme for spreading hate and anger to spread austerity. It is important to note that extreme feelings have a great distance (score) from neutral feelings. This paper considers extreme positive or extreme negative sentiment as extreme sentiment if it is a personal opinion and aims to define a strategy to spot them on social networks.

1.2 Contribution

In our recent work of Pais et al. (2020), an unsupervised approach has been used to detect extreme sentiments for the text obtained from social platforms. In this paper, we propose to use pre-trained BERT Devlin et al. (2019) to validate the efficacy of our prior approach for detecting extreme sentiments. BERT is a deep learning (Goodfellow et al. 2016) model based on a semi-supervised approach, pre-trained on Wikipedia and Book Corpus, a dataset containing more than 10,000 books of different genres. In addition, we use *Transfer Learning* (Ruder et al. 2019) which is typically done for tasks where the dataset contains too little data to train a full-scale model from scratch. The model is then fine-tuned, where the basic model is unpacked and re-trained on the new data with a very low learning rate. This can potentially lead to significant improvements by gradually adapting pre-trained features to the new data. We use BERT to validate the classified dataset acquired using the previously proposed unsupervised approach (Pais et al. 2020). This work confirms that the previously classified extreme events are indeed extreme events in most cases, hence reinforcing the validity of the approach.

The rest of the paper is structured as follows: The state-of-the-art and related works are briefly reviewed in Sect. 2, and the BERT approach is detailed in Sect. 3. The experimental setup is explained in Sect. 4, and the results and related discussions are presented in Sect. 5. Finally, the paper is concluded in Sect. 6 with some future research directions.

2 Related work

2.1 Detection and classification of extremism in social networks

The authors of Ahmad et al. (2019) proposed a binary classification task to detect extremist affiliation. The work focuses on machine learning classifiers, i.e., random forest, support vector machine (SVM), K-nearest neighbors (KNN), Naive Bayes, and deep learning. In this work, sentiment-based extremist classification techniques have been applied to user tweets which work in three modules: (i) collection of user tweets, (ii) preprocessing, and (iii) classification into extremist and non-extremist classes using various deep learning-based sentiment models, namely long short-term memory (LSTM) + convolutional neural networks (CNN/ConvNet), FastText and gated recurrent units (GRU). Although the results show the efficiency in terms of improved precision, recall, F-measure, and accuracy but contain some limitations. For example, the classification is limited to binary class rather than multi-class classification. There is also a lack of automation for streamlining the process of crawling, cleaning, and storing Twitter content.

The work in Kaur et al. (2019) uses a deep learning approach for automatic detection of extremism. The data collected by the authors are divided into radical, non-radical, and irrelevant through the use of relevant annotators. Word2Vec has been utilized in this work to generate word embeddings from data. Authors have used LSTM to detect extremism and classify the data as radical, not radical, and irrelevant in the context of India. This work uses specialized annotators to label the data and the labeling of text is based on the characteristics specified by the authors like abuse of Indian military personnel, anti-national discourse, endorsing terrorism/terrorists, and inciting others. The authors use different machine learning algorithms such as random forest, SVM, and Max Entropy for detecting radical content over online media. The proposed approach achieved precision of 85.9%, and it can be extended using an additional layer of CNN for precise identification of required features.

The work in Jaki and Smedt (2019) detects hatred and right-wing extremism in German Twitter users. The authors have identified several dehumanizing catchphrases used by right-wing extremists. The study classified and collected tweets as hate or non-hate for automatic detection. For the purpose of training the model, the authors used tweets in German and English. The study uses character trigrams as a method for feature extraction. Various features such as emojis, unigrams, bigrams, punctuation marks are also taken into account. The authors also tested their models in various unknown samples, some of which are marked manually by experts. These unknown samples were collected from various sources such as some random articles, Wikipedia pages in German, and German far-right conspiracy websites. The analysis of German right-wing tweets demonstrates how the insight of hate data can be employed for the development of automatic detection systems.

2.2 Sentiment analysis approaches

Sentiment analysis (SA) is a sub-field of *natural language processing (NLP)* mainly concerned with an effective way to determine the polarity of a text and emotion recognition, i.e., the prediction of whether the opinion expressed in a text is positive, negative, or neutral (Cambria 2016). These analyses are a powerful tool for deriving insights from large amounts of opinion-based data, such as social media posts and product reviews. SA is a proficient area for researchers, especially in the context of social media activity. In general, sentiment analysis systems fall into two categories: knowledge-based systems and statistics-based systems. Earlier knowledge-based approaches were the most popular among researchers to identify sentiment polarity in texts. However, nowadays researchers are increasingly using statistically-based approaches with an emphasis on supervised statistical methods (Cambria et al. 2018).

Wagh et al. (2018) designed a general sentiment classification to analyze whether a data label is available or unavailable in the target domain. The study analyzed the public dataset of four million tweets from Stanford University to predict the sentiment polarity in user opinions. SA using Hadoop, which rapidly runs large datasets on a real-time Hadoop cluster, was presented by Mane et al. (2014). It is a platform designed to solve large, unstructured, and complex big data problems using the divide-and-conquer approach to data processing. The study used a number-based approach to scale statements into several classes that assigned an appropriate range of different sentiments. *SENTA* (Bouazizi and Ohtsuki 2017) is an SA tool that offers many features to the end-user. The authors collect texts from Twitter and use SENTA to perform multi-class SA on the texts. Most of these approaches are supervised methods; our research focuses on an unsupervised and language-independent methodology to detect extreme sentiments on social media platforms.

Authors in Liang et al. (2022) proposed a graph convolutional network that is built on SenticNet (Cambria and Hussain 2015) to exploit the affective dependencies of the sentence according to the specific aspect which is called Sentic GCN. It uses a novel solution to build graph neural networks via integrating the affective knowledge from SenticNet in order to strengthen the dependency graphs of sentences. This novel affective enhanced graph model considers both the reliance of contextual words and aspect words and the affective information between opinion words. It first make graphs over dependency trees of sentences and then it enhance the word dependencies of each sentence by incorporating affective information into the graph and highlighting the specific aspect via leveraging the connection between aspect words and contextual affective words.

2.3 Sentiment-based Lexicons

SenticNet 5 (Cambria et al. 2018) encodes denotative and connotative information commonly associated with real objects, actions, events, and people. It avoids the indiscriminate use of keywords and word co-occurrences and instead relies on the implicit meaning associated with common sense concepts. Unlike purely syntactic techniques, *SenticNet 5* can detect subtly expressed emotions by analyzing multi-word utterances that do not explicitly express emotion but are associated with concepts that do. Here are two examples from the *SenticNet 5* dataset: *favorite* scores 0.87 (positive), *worry* scores − 0.93 (negative).

SenticNet 6 (Cambria et al. 2020 is a new version of SenticNet (Cambria and Hussain 2015), which is a semantic resource for concept-level sentiment analysis available for public use. It utilizes an ensemble of graph mining and multi-dimensional scaling to connect the conceptual and affective difference between word-level natural language data and the concept-level opinions and sentiments expressed by them. SenticNet 6 combines logical reasoning within deep learning architectures and it is created using a method for knowledge representation that is both top-down and bottom-up. The top-down approach is for the fact that it make use of symbolic models (i.e., logic and semantic networks) to encode meaning while bottom-up approach uses subsymbolic methods (i.e., biLSTM and BERT) to completely learn syntactic patterns from data. It deconstructs multi-word expressions into primitives and superprimitives, hence eliminating the need to build a lexicon that assigns polarity to thousands of words and multi-word expressions. The required thing needed is the polarity of superprimitives. For example, expressions like grow_profit, enhance_reward or intensify_benefit are all generalized as INCREASE(GAIN) and, hence, classified as positive.

SentiWordNet 3.0 (Baccianella et al. 2010) was developed by automatically scoring all *WordNet* synsets with the terms "positivity," "negativity," and "neutrality." Each synset has three numerical scores identifying the terms as positive, negative, and objective (i.e., neutral), e.g., *majestic* score 0.75 (positive term), and *invalid* score 0.75 (negative). The authors of *SentiWordNet 3.0* suggest using their tool as the basis for the development of extremism lexical resource, a comprehensive lexical resource to be used to support sentiment classification and opinion mining applications (Pang and Lee 2008).

2.4 Sentiment analysis datasets

SA development and tuning require sizeable labeled training datasets, also known as the SA training dataset. The first step in developing the analysis requires an SA dataset with thousands of statements already labeled as positive, negative, or neutral. Finding training data is difficult because a human expert must determine and label the polarity of each statement in the training data. Using already available training data reduces the time and effort required to develop a new dataset. The work Friedrich et al. (2015) uses *Sentiment140* (Go et al. 2009) and *SentiStrength* (Thelwall et al. 2012) on a prominent representative set of research articles, explicitly applying some techniques to sentiment analysis of articles circulating on Twitter. The dataset consists of two comma-separated values (CSV) files: one for testing and one for training. *Sentiment140* provides a sentiment value of the tweet on a scale of 0–4, where 0 = negative, 2 = neutral, and 4 = positive. The values have been converted into three sentiment categories: positive, negative, and neutral for better comparison. In our work, we chose the test file for the evaluation of our system.

Authors in Vadicamo et al. (2017) use Twitter for Sentiment Analysis (T4SA) (Vadicamo et al. 2017) visual dataset, which contains text and multimedia data to examine user sentiment. The authors collected Twitter data via a continuous tracker for six months and used it for a visual assessment of SA. The study of Smeureanu and Bucur (2012), which aims to detect users' opinions on movie reviews using the RT-polarity (Pang and Lee (2005 dataset, classified 2000 comments into two different categories. In general, the comments mainly consist of sentences. The authors classify user sentiments at the sentence level and then classify all comments as opinions. The resulting collection consists of two files: one for each set of 5331 positive and negative opinions.

TurntoIslam AZSecure-data (2013) and *Ansar1* University of Arizona Artificial Intelligence (2013), both with posts, are organized into threads that generally indicate the topic being discussed and focus on extremist (e.g., jihad) and general Islamic religious discussions. Each post contains detailed metadata, e.g., date and member name. As advertised on the forum, this is an English-language forum aimed at correcting common misconceptions about Islam. Radical participants also occasionally express support for fundamentalist militant groups. These two corpora will help us understand whether our approach works well in extremist religious discourse (e.g., jihadist) and general Islamic discourse.

Although a large number of approaches exist and few studies have offered an explicit comparison between SA techniques, the work of Gonçalves et al. (2013) shows comparisons of eight popular SA methods in terms of coverage and agreement. Ribeiro et al. (2016) present a sentence-level comparison of twenty-four popular sentiment analysis methods, based on a benchmark of eighteen labeled datasets. Performance was evaluated on two sentiment classification tasks: negative vs positive and three classes, namely negative, neutral and positive. However, these studies never compare the effectiveness of sentiment analysis methods or sentiment lexicons on the specific task of identifying extreme sentiments, i.e., extremely positive and extremely negative sentiments. To the best of our knowledge, the present work is one of the few direct attempts to identify extreme sentiments, i.e., extremely positive and/or extremely negative sentiments on social platforms, using BERT (Devlin et al. 2019).

3 Methodology

The objective of this work is to validate the extended lexicon which classified extreme posts using the proposed method called “ExtremeSentiLex.” To carry out this task, we utilize deep learning transformer-based model introduced by Google, known as BERT. Section 3.1 briefly introduces the previously proposed approach, Sects. 3.2 and 3.3 briefly explain the working of BERT and its application for our use case. The process flow is given in Fig. 1.

3.1 ExtremeSentiLex

In our previous work Pais et al. (2020), an unsupervised approach for automatic detection of people’s extreme sentiments on social networks was proposed. The approach was based on two steps: (1) extreme sentiment generator (ESG)—we automatically build a standard lexicon consisting of extreme positive and negative sentiment terms, and extend that same lexicon with a method based on word embeddings; (2) extreme sentiment classifier (ESC)—to validate the lexicon, using an unsupervised approach for automatic detection of extreme sentiments. We further evaluated our performance on five different social networks and media datasets (Sect. 2.4).

We designed and developed a prototype system composed of two components, i.e., ESG and ESC. ESG, based on statistical methods, is applied on *SentiWordNet 3.0* and *SenticNet 5* to generate a standard lexical resource known as ExtremeSentiLex that contains only extreme positive and negative terms. Additionally, we extend this new lexicon with new terms through the word embedding method (Mikolov et al. (2013, in order to study the behavior of our tools when tested with more terms. Antiextremism agencies can also use these lexical resources to find extreme opinion(s) on social networks to counter violent extremism. We embed the lexicons in the ESC and run them on the compilation of five different datasets, constituted of social network and media posts. The purpose of this experiment is to assess the performance of the proposed tool, and this evaluation will validate the proposed hypothesis that the ESC finds posts with extremely negative and positive sentiments in these datasets. To obtain more objective results, we use a confusion matrix to calculate recall, precision, F1-score, and accuracy to check the performance of the ESC.

In the preceding work, we presented and discussed the initial results of each dataset individually in the tables. The arrangements for these tables are different, according to each dataset itself from the original settings. In our case, **P**—positive, **N**—negative and **neutral** are the original polarity of the posts, **EP** are posts classified as extreme positive, **EN** means posts classified as extreme negative, and **E+INC** are posts classified as non-extreme or inconclusive.

We concluded that the extended lexicon detects more extreme posts. There is an approximately 2–5% increase in each category for RT-polarity, Sentiment140, TurntoIslam, and Ansar1 datasets. The most significant increase is observed in the T4SA data, from approximately 22 to 24% for the total number of the extreme and total number of positive extreme and 1% of total negative extreme. It is also observed that by extending the original lexicon with related terms, the proposed tool identified more extreme posts. It is expected since social media posts tend to be short, so a more extensive lexicon has a higher probability of detecting extreme sentiments in these short texts. The results obtained by using an extended lexicon can be seen in Table 1.

3.2 Bidirectional encoder representations from transformers (BERT)

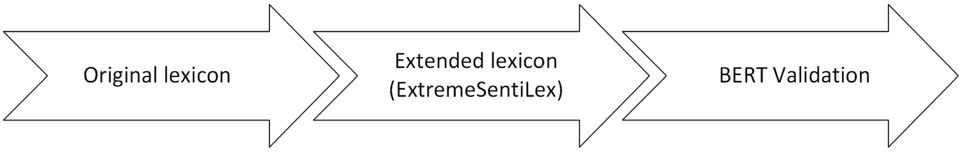
BERT is designed to pre-train deep bidirectional representations from the unlabeled text by jointly conditioning the left and right contexts. Its pre-trained model acts as the brain, which can then learn and adjust to the increasingly large resources of discoverable content and queries and can be

Table 1 Results obtained using the extended lexicon

	Datasets		
	RT-polarity	Sentiment140	T4SA
Recall$_{EP}$	**92%**	**97%**	**98%**
Recall $_{EN}$	41%	45%	43%
Precision $_{EP}$	64%	64%	**81%**
Precision $_{EN}$	81%	**93%**	**89%**
F1-score $_{EP}$	75%	**77%**	**88%**
F1-score $_{EN}$	54%	60%	58%
Accuracy	67%	71%	**82%**

Bold is to highlight the most relevant results

Fig. 1 Context of this work with regard to the previous approach



fine-tuned to the user's specifications. This process is called transfer learning. The pre-trained BERT model can be fine-tuned with a single additional output layer to create state-of-the-art models for different NLP problems. It encodes context bidirectionally and requires minimal architectural changes for a wide range of NLP tasks (Devlin et al. 2019). Using a pre-trained transformer encoder, BERT can represent any token based on its bidirectional context.

BERT is pre-trained on a massive corpus of unlabeled text, including Wikipedia (2,500 million words) and Book Corpus (800 million words). This pre-training step is half of the magic behind BERT's success. As the model is trained on a large text corpus, the model starts to pick up a more profound and intimate understanding of how the language works. This knowledge is the backbone that is useful for almost any NLP task. The most helpful feature of BERT is fine-tuning, whereby by adding just some of the additional output layers, we can create state-of-the-art models for various NLP tasks. BERT is currently being used by Google to optimize the interpretation of search engine queries. Initially, it was limited to the English language, but by December 2019, the model had already been rolled out in over 70 languages. BERT performs exceptionally well on various NLP and sequence-to-sequence-based language generation-related tasks such as *question answering*, *abstract summarization*, *sentence prediction*, *conversational response generation*, *polysemy* and *coreference* (words that sound or look the same but have distinct meanings) *resolution*, *word sense disambiguation*, *natural language inference*, and *sentiment classification* (text classification).

3.3 Fine-tuning BERT for text classification

The BERT-base model incorporates an encoder with 12 transformer blocks, 12 self-attention heads, and 768 units of hidden embedding parameters, a sequence of hidden states of the last layer of the model. The original BERT achieved state-of-the-art results on eleven NLP tasks. However, we are only interested in its classification task. BERT has two versions of different model sizes (Devlin et al. 2019). The base model (BERT-base) uses 12 layers (transformer encoder blocks) with 768 hidden units (hidden size) and 12 self-attention heads. The large model (BERT-large) uses 24 layers with 1024 hidden units and 16 self-attention heads. Notably, the former has 110 million parameters, while the latter has 340 million parameters. In our work, here presented, we have fine-tuned our model on a pre-trained BERT-base, using 12 layers, 768 hidden units, and 12 self-attention heads. BERT takes an input of a sequence of up to 512 tokens and outputs the sequence representation. The sequence has one or two segments, where the first token of the sequence is always [CLS] and contains the particular classification embedding, and another special token [SEP] is used to separate the segments. BERT picks the final hidden state h of the first token [CLS] for text classification tasks to represent the complete sequence. In order to get the predicted probabilities from the trained model, a softmax classifier is added to the top of the BERT model.

Firstly, the dataset is vectorized for feeding it to the classifier since it is originally in text format. Different models are available for vectorizing text but BERT learns contextual-embedding rather than learning context-free, such as in the case of Word2Vec. It performs tokenization using WordPiece (Wu et al. 2016) method. In addition to [CLS] and [SEP], it adds a new token called [PAD], to make the length of all sentences equal to the specified sequence length required by the model, and an attention mask is introduced to tell the model about [PAD] tokens. These are then used to input the model to obtain vector representation of each token. Since the base model has 12 layers of encoders, tokens are fed into the first encoder, and the output of the first encoder is then given as input for the second encoder, and so on until the last encoder. The last encoder, which is encoder 12 returns the embeddings for all tokens in the sentences. The representation size of each token is 768 in BERT-base model. This phenomenon is shown in Fig. 2.

For single text classification applications, the BERT representation of the special classification token "[CLS]" encodes information about the entire input text string. The single input text representation is fed into a small multilayer perceptron (MLP) consisting of fully connected (dense) layers to produce the distribution of all discrete label values (Zhang et al. 2020).

4 Experimental setup

In this experiment, BERT is used which is a transformer-based machine learning technique for NLP. The datasets classified in previous work (Pais et al. 2020) are used for training, while the target variable is polarity, a class created by using the ExtremeSentiLex. The experiment is conducted on all five datasets, and at the end, all five datasets are combined to check the overall performance of the proposed method and BERT. This comes in handy as the type of data can be a primary driver in determining the classification. So, the mixing of diverse data challenges the model and allows space for better insights.

4.1 Loading data

The classified dataset is loaded from ExtremeSentiLex containing six files. All of these files for our experiment except the Sentiment140 training file because of missing the extreme negative class in polarity. The files are in text format that are then loaded in pandas data frames. The primary

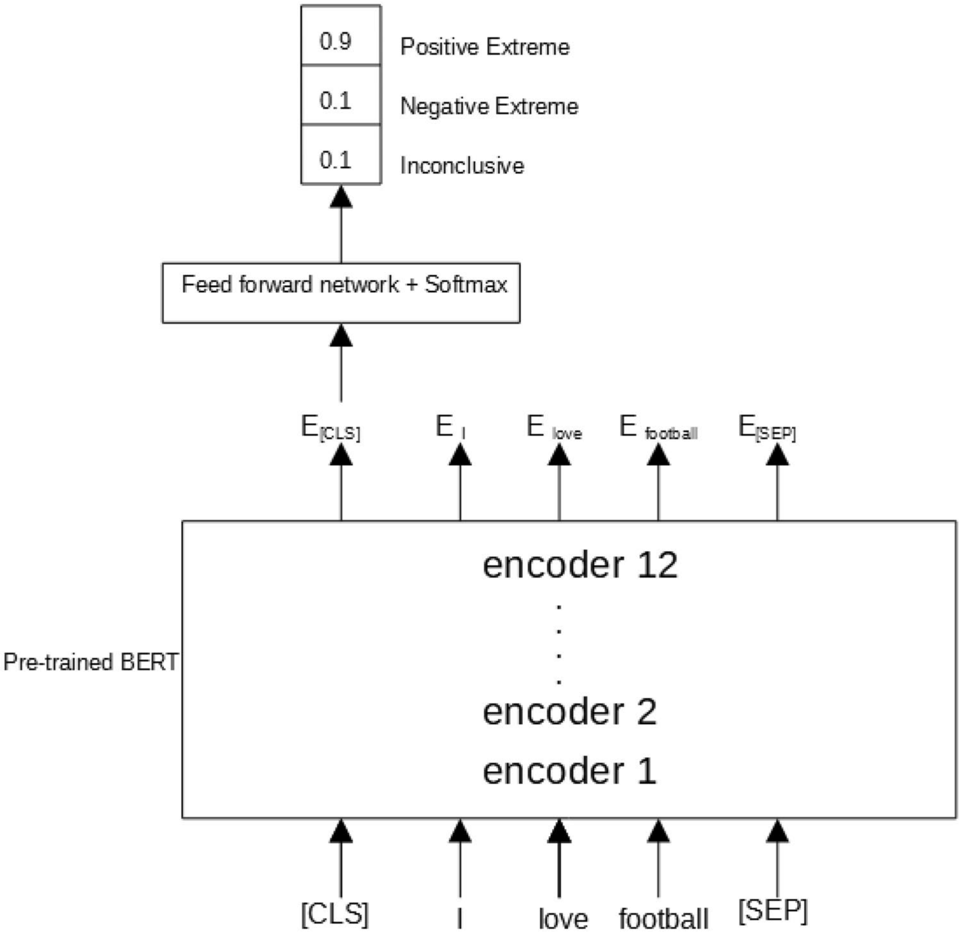


Fig. 2 Tokens are embedded using 12 encoders in the BERT-base model and fed into a feed-forward network and softmax function to get the classification probabilities

data we need are in two columns: "message" and "polarity." Other columns are dropped as they are not needed in the experiment.

4.2 Preprocessing

In preprocessing, one-hot encoding is used for categorical polarity data. Another extra step is necessary in the case of dataset file T4SA. This file is encoded and contains a byte string. We decode the file and clean the text containing special characters.

4.3 Text split for longer text

BERT has a limitation of 512 max length for input characters. It means that the input data cannot be longer than 512 for training. Extreme *Ansar1* contains 15325 characters long input feature, and extreme *TurntoIslam* has a maximum text length of 10034 characters. These data cannot be used directly as input for the training BERT model. The simple and rough way is to truncate directly, take the initial part up to 512 characters max, and discard the rest. Although this simple naive method is effective in many cases, a complex method tackles this issue based on the HIERARCHICAL (cascade) idea, which divides the longer text into smaller chunks and feeds them into the base model (Pappagari et al. 2019). Any text larger than 450 characters is divided into 500 characters with an overlap length of 50 words. These split data keep the same polarity class as its source. It is then given as input for training the BERT model.

4.4 Sampling

The dataset of our experiment is highly unbalanced, with high inconclusive and positive extreme polarity outnumbering the small class of extreme negative terms. Training without balancing the data will cause inaccurate results, i.e., predicting well for inconclusive and positive extreme while poorly for negative extreme. Negative extreme happens to have the essential kind of terms which we cannot ignore. Therefore, undersampling and a mix of under and oversampling techniques are used to get the balanced dataset with an equal number of polarity classes. The sampling technique differs for each dataset, and it is explained in Sect. 5 under the heading of each dataset.

4.5 Train-test split

After sampling and balancing the dataset, we use 80% of the data for training while keeping 10% for validating the training and 10% for testing the model.

4.6 Training and validation

We use the "bert-base-cased" pre-trained model as most of our data are in English. The model runs six epochs and a different number of batches for training with each dataset. The batch number varies as the higher number for batch size will cause GPU memory constraints. To overcome this limitation, we keep batch size around 6, 8, or 10 depending on the training data size and use validation data to analyze the training accuracy and loss. The hyper-tuning parameters include optimizer and learning rate. AdamW optimizer is used in the experiment which was also used in the original BERT model for pre-training. The learning rate used in the model is 2e-5 which is necessary to make BERT overcome the catastrophic forgetting problem. However, higher learning rates can be used such as 4e-4 which can cause the failure of converge on the training set (Sun et al. 2019). A smaller learning rate may allow the model to learn a more optimal or even globally optimal set of weights.

4.7 Testing

After training is complete, we run the test data on the trained model to check how well the model performs. The minimum difference between validation and test accuracy reflects the overall accuracy of our model. The results of each dataset are discussed in Sect. 5.

4.8 Performance metrics

We will use common performance metrics for evaluating our performing model. It includes accuracy, confusion matrix, *F1-score*, precision, and recall. As usual, *accuracy* is the fraction of correct predictions, the number of hits divided by the total number of predictions. The *confusion matrix*, also known as *error matrix*, assesses the classification accuracy by calculating the confusion matrix with each row corresponding to the true class. It is displayed in a table layout that allows visualization of the efficacy of a classification algorithm. *Precision* is the ability of the classifier to not predict the false label or value. *The recall* is the ability of the classifier to find all the positive samples. It can be also referred to as the fraction of the relevant labels that is successfully predicted. *The F1-score* also referred to as *balanced F-score* or *F-measure*, is the weighted average of the precision and recall. It achieves its best value at 1 and the worst at 0. The *F-score* is also used for calculating classification problems with more than two classes which is also called multi-class classification. These two classes are called micro-averaging and macro-averaging. The final score is obtained by micro-averaging which is biased by class frequency, whereas macro-averaging takes all classes as equally important. Another type of *F1-score* is the weighted average.

There are three types of averages namely micro, macro and weighted. Micro-average evaluates metrics globally by computing the total true positives, false negatives, and false positives. Macro-average tallies metrics for each label and finds their unweighted mean. The imbalanced labels are not taken into account. Weighted average determines metrics for each label. It finds their average weighted by support. This changes the macro average to reckon unbalanced label which can lead to F-score that is distinct from precision and recall.

Support is the number of actual occurrences of the class in the specified dataset. Unbalanced support in the training data may indicate structural weaknesses in the scores of the reported classifiers and could indicate the need for stratified sampling or re-balancing. The *support* does not variate between the models but rather diagnoses the evaluation process.

Table 2 Extreme posts detected from datasets using the extended lexicon

	Datasets				
	RT-polarity	Sentiment140	T4SA	TurntoIslam	Ansar1
Total of extreme	2518 (≈24%)	63 (≈13%)	423689 (≈36%)	120644 (≈36%)	12002 (≈41%)
Extreme positive	1928 (≈18%)	49 (≈10%)	372090 (≈32%)	110658 (≈33%)	10534 (≈36%)
Extreme negative	590 (≈6%)	14 (≈3%)	51599 (≈4%)	9986 (≈3%)	1468 (≈5%)
Total	10662 (≈100%)	497 (≈100%)	1179957 (≈100%)	335328 (≈100%)	29492 (≈100%)

5 Results and discussion

In this section, the results obtained for the detection of extreme sentiments by using BERT are presented. Apart from using the five datasets (RT-polarity, T4SA, Sentiment140, TurntoIslam, and Ansar1), we have also combined these datasets to build a new dataset, comb_all to analyze the overall performance of the proposed model. Hence, six experimental results and findings are outlined for each dataset.

The datasets used in this experiment are acquired from our previous work, which we refer to as the extended lexicon (Pais et al. 2020). The comprehensive lexicon data are highly unbalanced, as shown in Table 2. The number of negative extremes is very low compared to inconclusive and positive extremes. The split of these data into train, validation, and test sets causes further isolation, which results in higher accuracy of dominant classes while ignoring the minor class, which is negative extreme. To tackle this issue, we keep the same number from each class and use sampling techniques to increase the number of records to feed the model.

Table 3 Classification report of RT-polarity dataset

	RT-polarity			
	Precision	Recall	F1-score	Support
Inconclusive	0.63	0.71	0.67	31
Positive extreme	0.64	0.48	0.55	29
Negative extreme	0.74	0.84	0.79	25
Accuracy			0.67	85
Macro avg	0.67	0.68	0.67	85
Weighted avg	0.67	0.67	0.66	85

5.1 RT-polarity

This is a dataset of classified movies, containing the polarity of tweets calculated in the first phase (Sect. 3). It is highly skewed toward the *inconclusive* and the *positive extreme* classes, with a tiny percentage of *negative extreme*. The total records are 26108 for inconclusive, 5596 for positive Extreme, and 282 for negative extreme. The undersampling technique is used to prepare data for training. The equal number of data is taken randomly from the class of inconclusive and positive extreme with regard to the negative extreme which implies 282 entries for each class. The train set contains 20% of data, while 80% of the data is used for the test and validation set. After running three epochs of training, the validation accuracy of the model becomes flat at 67% approximately. The test accuracy is the same as the training accuracy which shows the model is performing well.

The model gives the following classification report for RT-polarity. The overall accuracy of the model is 67%. Table 3 also highlights the precision, recall, and F1-score for our target classes.

The use of a *confusion matrix* allows us to have a better view of the efficacy of our model, revealing specifically the type of errors being committed. It shows the values of true positive, false positive and vice versa for each class (Fig. 3).

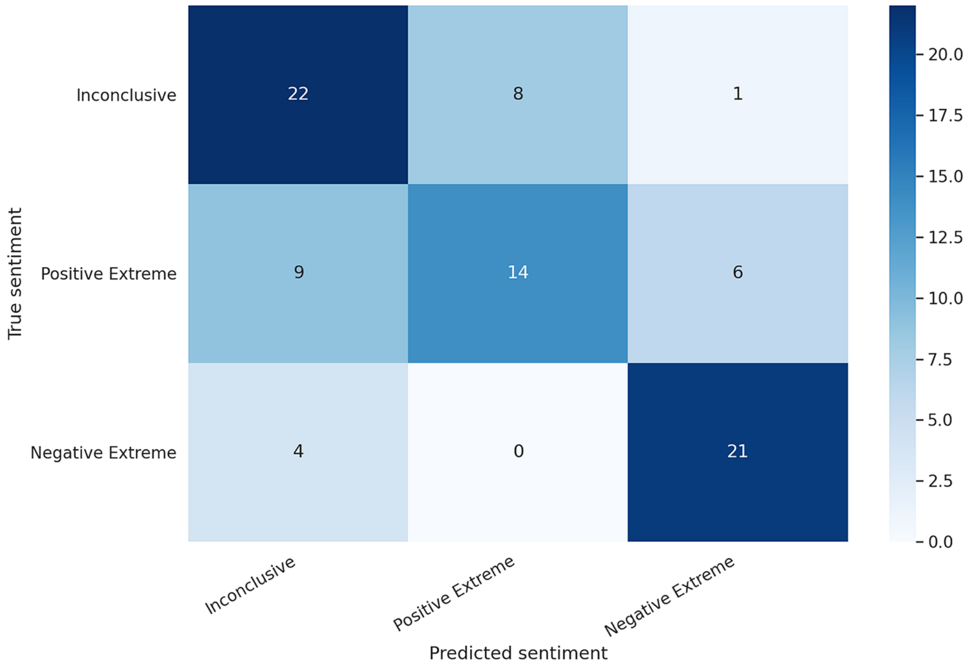


Fig. 3 Confusion matrix of RT-polarity dataset

5.2 T4SA

The next dataset we use is T4SA (Twitter for sentiment analysis). The text in the T4SA dataset is in byte string format, and we have applied a special functionality to clean and decode the text in order to finalize it for training as it contains emojis. The number of entries is 3134881 for inconclusive, 293568 and positive extreme and 24114 for negative extreme. As the negative extreme is the lowest class with 24114 entries, the same number of entries are randomly selected for the inconclusive and positive extreme class while using the undersampling technique. The train set contains 20% of the data, while 80% of the data is used for the test and validation sets. Sentiment analysis is widely used on Twitter datasets because of its usefulness. Therefore, the BERT model gives 99% training, 98% validation, and 98% test accuracy after running six epochs, supposedly because BERT is pre-trained on a large corpus of unlabeled text. The classification report on the T4SA dataset is given in Table 4.

Table 4 Classification report of T4SA dataset

	T4SA			
	Precision	Recall	F1-score	Support
Inconclusive	0.98	0.98	0.98	2381
Positive extreme	0.99	0.99	0.99	2484
Negative extreme	0.99	0.99	0.99	2370
Accuracy			0.99	7235
Macro avg	0.99	0.99	0.99	7235
Weighted avg	0.99	0.99	0.99	7235

The performance of the classification model onset of test data for which the actual values are known is shown in the confusion matrix (Fig. 4).

5.3 Sentiment140

The sentiment140 data contain two files. The sentiment140_test is used for creating comb_all dataset, and the sentiment140_train file is discarded because of the absence of "extreme negative" class in polarity. The text, which is the training variable, is the message that needs preprocessing. The sentiment_train140 dataset is tiny, and the number of entries is 1808 inconclusive, 168 for positive extreme and 12 for negative extreme. Since the negative extreme class is extremely small with 12 entries in total, the probability of data leakage and overfitting is high regardless of sampling technique used. Therefore, the experiment is not done on this dataset; rather, it is experimented indirectly by combining it with other datasets in comb_all.

5.4 TurntoIslam

The TurntoIslam dataset contains two extra classes of our target variable, "polarity," similar to Ansar1. Since the main goal of our work is to find extreme positive and extreme negative terms, we regard the negative non-extreme and positive non-extreme as inconclusive or neutral. The number of records is 690902 for inconclusive, 302178 for positive extreme, 3828 for positive non-extreme, 3574 for

Fig. 4 Confusion matrix of T4SA dataset

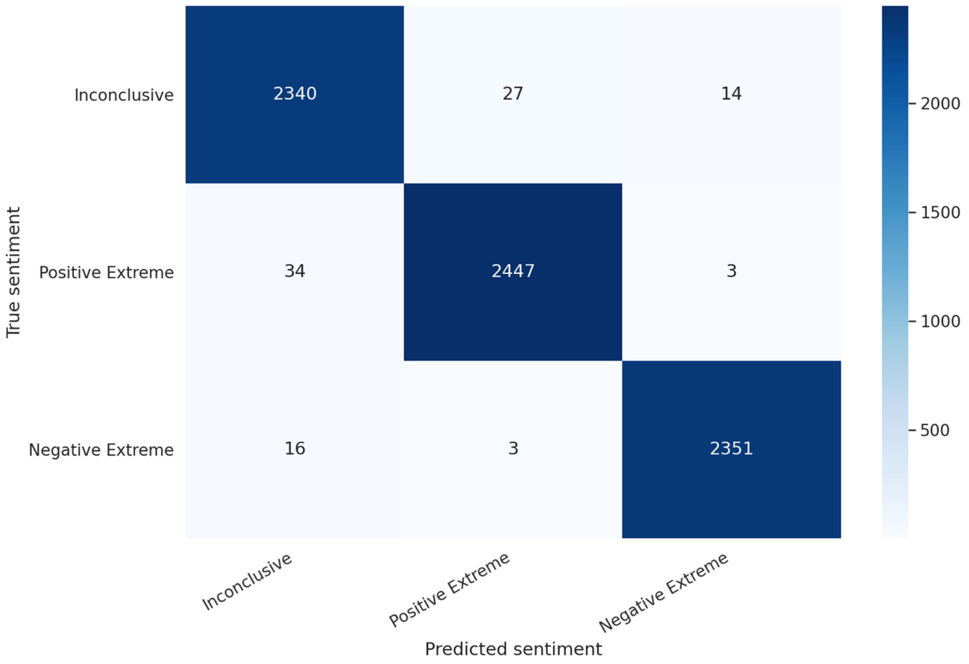


Table 5 Classification report of TurntoIslam dataset

	TurntoIslam			
	Precision	Recall	F1-score	Support
Inconclusive	0.96	0.94	0.95	282
Positive extreme	0.78	0.71	0.75	276
Negative extreme	0.94	0.85	0.89	329
Positive non-extreme	0.62	0.74	0.67	276
Negative non-extreme	0.68	0.70	0.69	286
Accuracy			0.79	1449
Macro avg	0.80	0.79	0.79	1449
Weighted avg	0.80	0.79	0.79	1449

negative extreme, and 2501 for negative non-extreme. The smallest class of dataset is negative non-extreme with 2501 entries; therefore, we randomly take the same number of entries from each class using the undersampling technique. The text used to train the model is long, so chunks of 500 words are made if the text is more than 450 characters, and the polarity values are maintained for each chunk. These data are then used for training the model. The train set contains 20% of the data, while the rest of the data are used as a test and validation set. Although the training accuracy increases with training epochs, the validation accuracy remains consistent, around 77%. The accuracy we get on the test set is 79% which exhibits little difference from the validation accuracy. The classification report of the dataset TurntoIslam shows the overall accuracy we obtained is 79% as given in Table 5.

The confusion matrix of TurntoIslam given in Fig. 5 indicates that the model works well for the negative extreme, which is the most important class of our dataset.

5.5 Ansar1

Ansar1 is a dark web forum that contains mixed language discussions on forum. Ansar1 dataset yields lower accuracy compared to other datasets and it can be attributed to the presence of foreign languages other than the English language in sentences. While the central part consists of English, the mixture of another language, especially Arabic, can influence the model, as we are using the BERT pre-trained model for English. Although multilingual pre-trained models are available for BERT, they do not provide better results with the type of data we are using. We split the more extensive texts into smaller chunks to feed the

Table 6 Classification report of Ansar1 dataset

	Ansar1			
	Precision	Recall	F1-score	Support
Inconclusive	0.82	0.82	0.82	74
Positive extreme	0.56	0.58	0.57	52
Negative extreme	0.47	0.46	0.47	61
Positive non-extreme	0.44	0.50	0.47	50
Negative non-extreme	0.51	0.45	0.48	62
Accuracy			0.58	299
Macro avg	0.56	0.56	0.56	299
Weighted avg	0.58	0.58	0.58	299

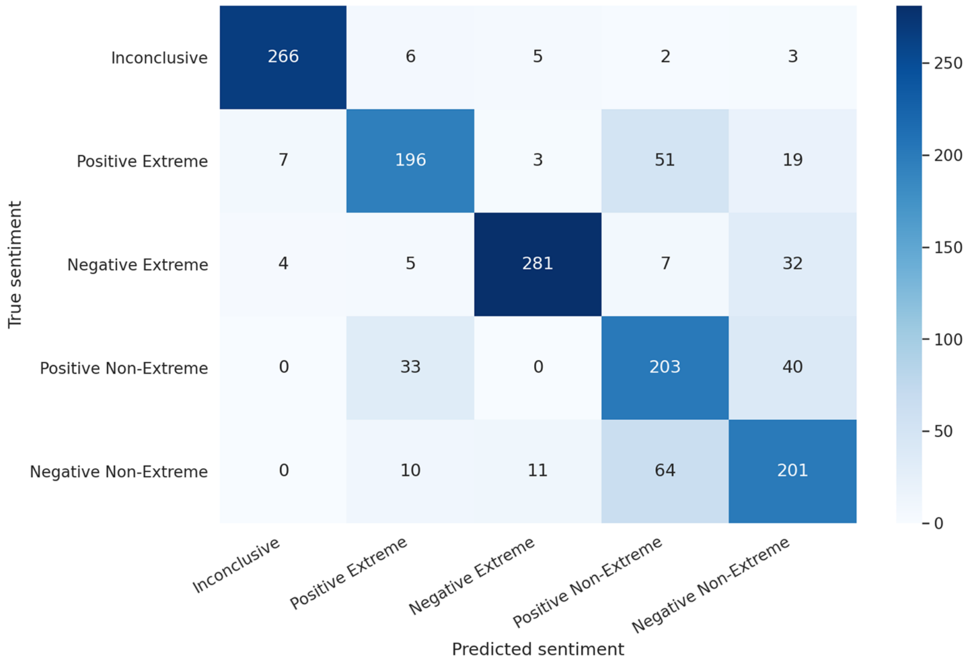
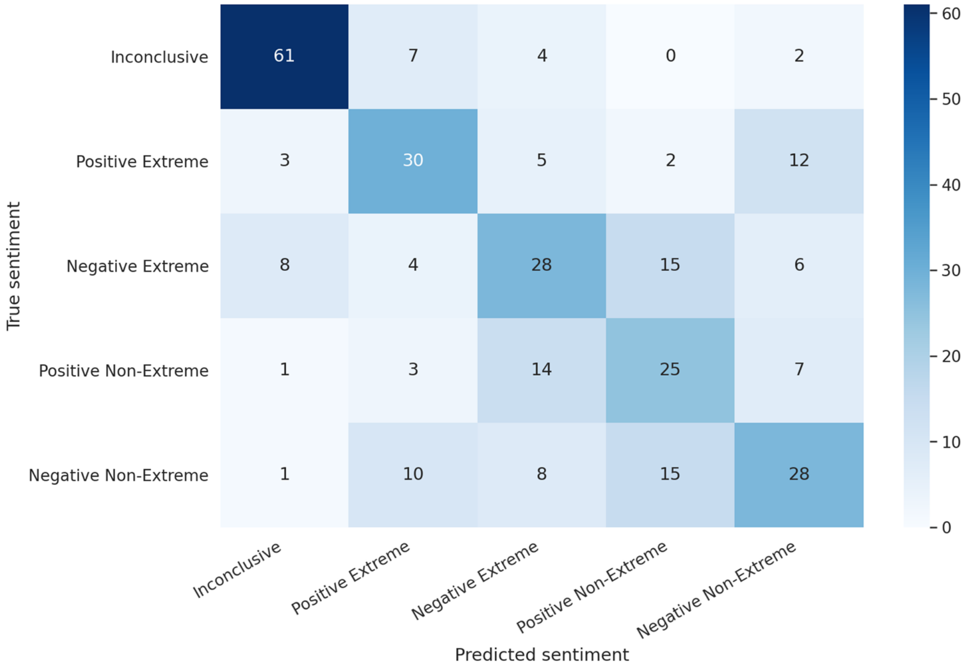


Fig. 5 Confusion matrix of TurntoIslam dataset

Fig. 6 Confusion matrix of Ansar1 dataset



BERT model. The number of entries is 65301 for inconclusive, 21334 for positive extreme, 796 for negative extreme, 520 for positive non-extreme, and 491 for negative non-extreme. The smallest class is 491 for the negative non-extreme class; therefore, each class is reduced to the same size of 491 using the undersampling technique. These data are then split 20% for training and the rest for validation and testing. Although the training accuracy exceeds 80%, the validation accuracy remains between 56 and 58% while running epochs of our model. The accuracy we get on the test set is 57% which is an approximate value of our validation accuracy. The classification report for the Ansar1 data set is shown in Table 6.

Figure 6 gives the comparison of results for each class in the form of confusion matrix.

5.6 comb_all

The last dataset for the experiment is made by combining all the previous five datasets we tested on. All datasets are concatenated based on the common classes of polarity present in all datasets to preserve the data consistency, which is "inconclusive, positive extreme and negative extreme." This dataset is referred to as comb_all. The sampling technique is not required here as the data used to create this dataset are already passed through the stages of the sampling phase. A total of 20% data are used for the training set, and the rest of the 80% data are used for test and validation set. This combination of data into a new dataset provides a good insight into the efficacy of using BERT, which seems to perform well. It proves the reliability of BERT for our tasks and related ones. Also, it verifies our approach to classify into extreme positive and extreme negative classes. The training accuracy of the model reaches around 98%, while the validation accuracy of our model is around 95%. The accuracy we get on the test set is also 95%, which confirms the validation accuracy. The classification report of comb_all is shown in Table 7.

Table 7 Classification report of comb_all dataset

	comb_all			
	Precision	Recall	F1-score	Support
Inconclusive	0.94	0.94	0.94	2904
Positive extreme	0.95	0.95	0.95	3110
Negative extreme	0.96	0.97	0.96	3044
Accuracy			0.95	9058
Macro avg	0.95	0.95	0.95	9058
Weighted avg	0.95	0.95	0.95	9058

Figure 7 shows the confusion matrix of our model, which achieves good results for extreme values, especially for *negative extreme*.

For a comprehensive analysis and visualization of the difference between the original lexicon and the expanded one, using BERT, we can look at the results presented in Table 8, for each dataset. The results are also shown in form of graphs in Figs. 8, 9, and 10. However, the Sentiment140 dataset results are presented here only for the sake of comparison and to highlight the issue of data

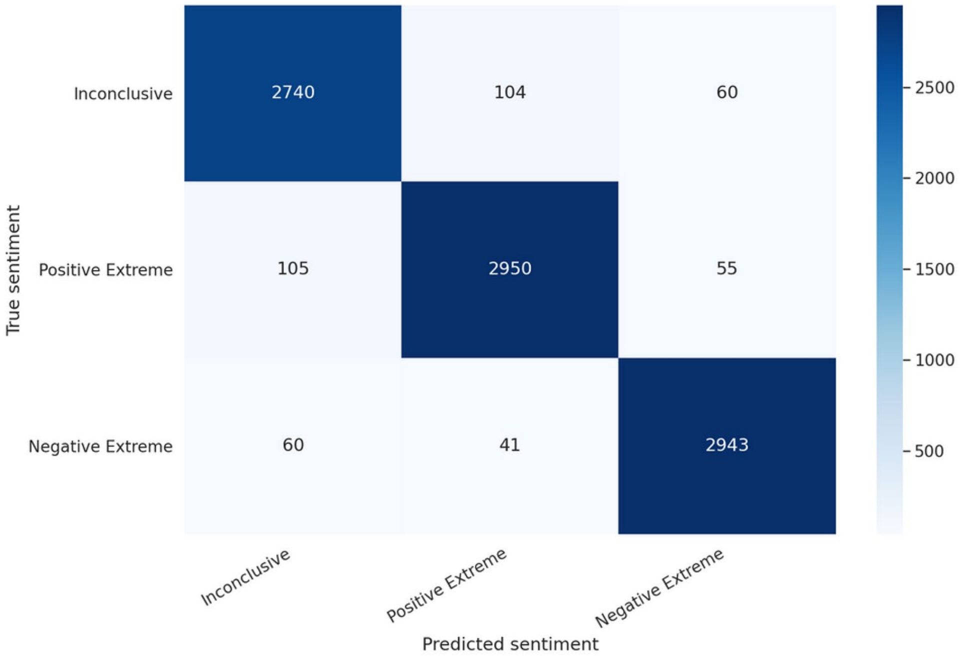


Fig. 7 Confusion matrix of comb_all dataset

Table 8 Results obtained using BERT

	Datasets		
	RT-polarity	Sentiment140	T4SA
Recall $_{EP}$	48%	80%	**99%**
Recall $_{EN}$	**84%**	**100%**	**99%**
Precision $_{EP}$	64%	**80%**	**99%**
Precision $_{EN}$	74%	80%	**99%**
F1-score $_{EP}$	55%	**80%**	**99%**
F1-score $_{EN}$	**79%**	**89%**	**99%**
Accuracy	67%	**88%**	**99%**

Bold is to highlight the most relevant results

Table 9 Results obtained using the extended lexicon

	Datasets		
	RT-polarity	Sentiment140	T4SA
Recall $_{EP}$	**92%**	**97%**	**98%**
Recall $_{EN}$	41%	45%	43%
Precision $_{EP}$	64%	64%	**81%**
Precision $_{EN}$	**81%**	93%	**89%**
F1-score $_{EP}$	**75%**	77%	88%
F1-score $_{EN}$	54%	60%	58%
Accuracy	**68%**	71%	82%

Bold is to highlight the most relevant results

leakage and overfitting caused by small and imbalanced data (Table 9).

In terms of overall accuracy, BERT shows promising results for the datasets RT polarity, Sentiment140 and T4SA. Mainly, it successfully classifies the Extreme Negative (EN) after fine-tuning the base model of BERT as seen in Table 8. It yields higher result values for the dataset T4SA which can be because of being a massive training data and the fact that it contains common topics and themes. In regard to extreme negative sentiments, it shows limited ability to classify the sentiments. Our proposed method outperforms BERT in certain cases and highlights the limitation of BERT for detecting extreme sentiments. BERT is one of the top models for the purpose of general sentiment analysis, while for very specific tasks of extreme sentiments, it requires further enhancements or other models on top of the base BERT model for better detection.

6 Conclusion and future work

In this paper, we validated an unsupervised and language-independent approach for detecting people's extreme sentiments on social media platforms using BERT. The proposed approach is based on defining extreme polarity for terms and generating extreme sentiment lexicon by relying upon two standard lexical resources, i.e., *SentiWordNet 3.0* and *SenticNet 5*. This work provides a standard lexicon consisting of extreme positive and negative terms polarity. We implemented a prototype system with two different components *ESG* and *ESC*. The proposed system experimented on five social networks and media data lexicons to analyze its accuracy, effectiveness, and

Fig. 8 Comparison between results of RT-polarity

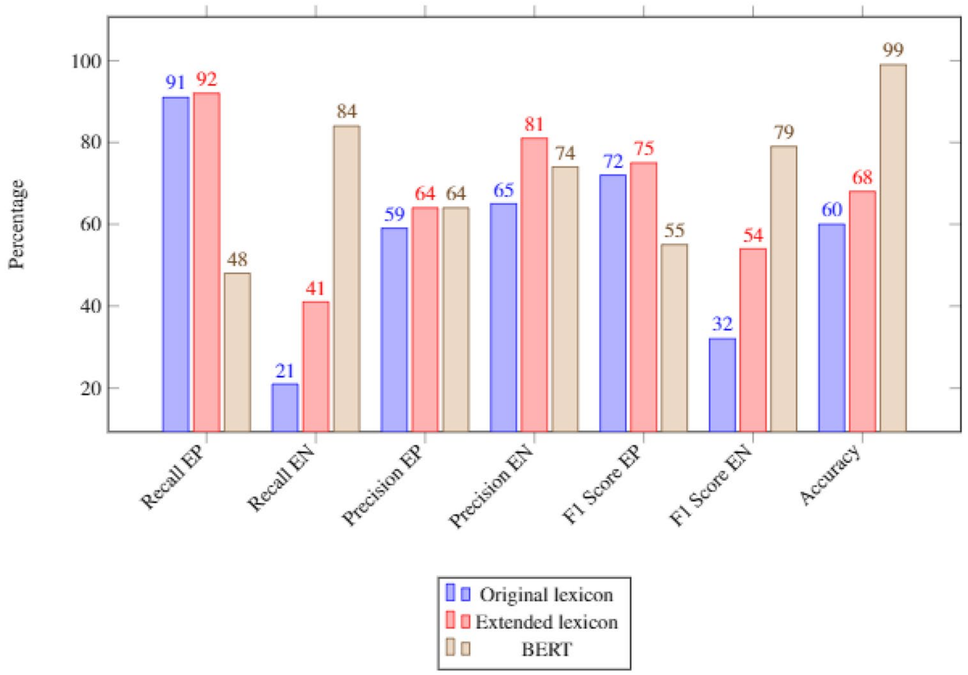
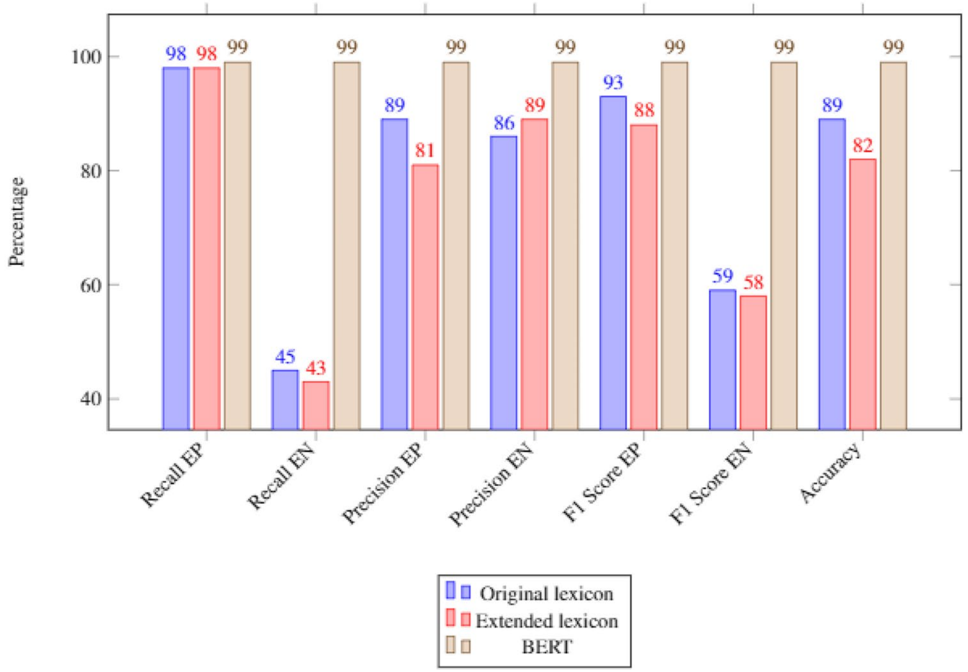


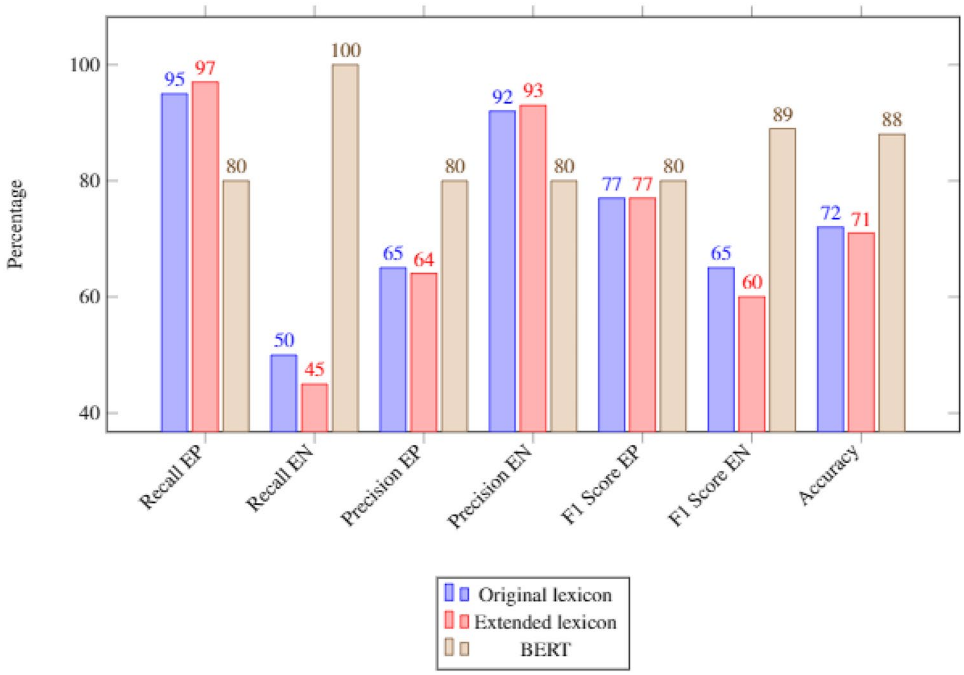
Fig. 9 Comparison between results of T4SA



efficiency. Moreover, word embeddings have also been utilized to extend the lexicon to analyze the improvement in the system's performance. The obtained results are promising and encouraging, as the system shows excellent improvement using the extended lexicon. This standard lexicon can also be helpful for other researchers to exploit it for SA studies and anti-extremism authorities, allowing them to identify and prevent violent extremism early. Furthermore, the results acquired using BERT (semi-supervised approach) versus our proposed unsupervised approach reflect two takeaways, the reliability of the proposed approach for detecting and classifying extreme sentiments, and the effectiveness of the BERT model for learning extreme polarity.

As an extension of this research, we plan to deal with the deficiencies of the proposed model, especially the weakness in detecting the opposites. It tends to ignore the negations while only picking the emotion/polarity. For this, we will

Fig. 10 Comparison between results of Sentiment140



apply linguistic tools in our approach, for example, to detect negation (*he is happy* is the opposite of *the is not happy*), to detect expressions with intensifiers (e.g., *he likes a lot*). For future research, we also plan to enhance our approach using NLP techniques to detect radical elements on social media and networks. A radical event has some specific features being quite different from the identification of extremism, as, for example, a radical behavior does not imply the manifestation of extreme sentiments.

Acknowledgements This work was supported by National Founding from the FCT-Fundacão para a Ciência e a Tecnologia, through the MOVES Project-PTD/EEI-AUT/28918/2017 and by operation Centro-01-0145-FEDER-000019-C4-Centro de Competências em Cloud Computing, co-financed by the European Regional Development Fund (ERDF) through the Programa Operacional Regional do Centro (Centro 2020), in the scope of the Sistema de Apoio à Investigação Científica e Tecnológica.

Funding This work was supported by National Founding from the FCT Fundacão para a Ciência e a Tecnologia, through the MOVES Project- PTDC/EEI-AUT/28918/2017, and by Operação Centro-01-0145-FEDER-000019-C4-Centro de Competências em Cloud Computing, co-financed by the Programa Operacional Regional do Centro (CENTRO 2020), through the Sistema de Apoio à Investigação Científica e Tecnológica-Programas Integrados de IC&DT.

Availability of data and material In the final version will be available data and material.

Code availability In the final version will be available code.

Declaration

Conflict of interest The authors declare that they have no known competing financial interests or personal relationships or conflicts of interest that could have appeared to influence the work reported in this paper.

Consent for publication The author's consent publication.

Consent to participate The author's consent participates.

Ethical approval The authors declare that they have no known ethics issue that could have appeared to influence the work reported in this paper.

References

Ahmad S, Asghar MZ, Alotaibi FM, Awan I (2019) Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. HCIS 9(1):24

AZSecure-data (2013) Turn to Islam forum dataset. University of Arizona Artificial Intelligence Lab, AZSecure-data. https://www.azsecure-data.org/dark-web-forums.html

Baccianella S, Esuli A, Sebastiani F (2010) Sentiwordnet 3.0: an enhanced lexical resource for sentiment analysis and opinion mining. In: LREC, vol. 10, pp 2200–2204

Becker H, Naaman M, Gravano L (2011) Selecting quality twitter content for events. In: Fifth international AAAI conference on weblogs and social media

Bouazizi M, Ohtsuki T (2017) A pattern-based approach for multi-class sentiment analysis in twitter. IEEE Access 5:20617–20639

Cambria E (2016) Affective computing and sentiment analysis. IEEE Intell Syst 31(2):102–107. https://doi.org/10.1109/MIS.2016.31

Cambria E, Poria S, Hazarika D, Kwok K (2018) Senticnet 5: discovering conceptual primitives for sentiment analysis by means of context embeddings. In: Thirty-second AAAI conference on artificial intelligence

Cambria E, Hussain A (2015) Senticnet. In: Sentic computing. Springer, pp 23–71

Cambria E, Li Y, Xing FZ, Poria S, Kwok K (2020) Senticnet 6: ensemble application of symbolic and subsymbolic AI for sentiment analysis. CIKM '20. Association for Computing Machinery, New York, NY, USA, pp 105–114 . https://doi.org/10.1145/3340531.3412003

DN (2022) PJ deteve estudante que preparava atentado na Faculdade de Ciências. https://www.dn.pt/sociedade/pj-deteve-jovem-com-planos-e-armas-para-atentado-numa-universidade-de-lisboa-14578423.html

Devlin J, Chang MW, Lee K, Toutanova K (2019) Bert: pre-training of deep bidirectional transformers for language understanding. In: NAACL

Dixit P, Miller K (2022) Erasing historical violence from the study of violent extremism: memorialization of white supremacy at stone mountain, united states. Crit Stud Terror. https://doi.org/10.1080/17539153.2022.2032550

Friedrich N, Bowman TD, Stock WG, Haustein S (2015) Adapting sentiment analysis for tweets linking to scientific papers. arXiv preprint arXiv:1507.01967

Go A, Bhayani R, Huang L (2009) Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, vol 1, p 2009

Gonçalves P, Araújo M, Benevenuto F, Cha M (2013) Comparing and combining sentiment analysis methods. In: Proceedings of the first ACM conference on online social networks, pp 27–38

Goodfellow I, Bengio Y, Courville A (2016) Deep learning. http://www.deeplearningbook.org

Jaki S, Smedt TD (2019) Right-wing german hate speech on twitter: analysis and automatic detection. CoRR. arXiv:1910.07518

Kaur A, Saini JK, Bansal D (2019) Detecting radical text over online media using deep learning. CoRR. arXiv:1907.12368

Krumm JS (2013) Influence of social media on crowd behavior and the operational environment. Technical report, Army Command and Geberal Staff College Fort Leavenworth KS School of …

Liang B, Su H, Gui L, Cambria E, Xu R (2022) Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks. Knowl Based Syst. https://doi.org/10.1016/j.knosys.2021.107643

Mane SB, Sawant Y, Kazi S, Shinde V (2014) Real time sentiment analysis of twitter data using hadoop. IJCSIT Int J Comput Sci Inf Technol 5(3):3098–3100

Mikolov T, Chen K, Corrado G, Dean J (2013) Efficient estimation of word representations in vector space

Pais S, Tanoli IK, Albardeiro M, Cordeiro J (2020) Unsupervised approach to detect extreme sentiments on social networks. In: 2020 IEEE/ACM international conference on advances in social networks analysis and mining (ASONAM), pp 651–658 . https://doi.org/10.1109/ASONAM49781.2020.9381420

Pang B, Lee L et al (2008) Opinion mining and sentiment analysis. Found Trends® Inf Retr 2(1–2):1–135

Pang B, Lee L (2005) Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales. In: Proceedings of the ACL

Pappagari R, Żelasko P, Villalba J, Carmiel Y, Dehak N (2019) Hierarchical transformers for long document classification. Automatic speech recognition and understanding workshop. arXiv:1910.10781v1

Persia F, D'Auria D (2017) A survey of online social networks: challenges and opportunities. In: 2017 IEEE International Conference on Information Reuse and Integration (IRI), pp 614–620 . IEEE

Politico (2021) Germany records highest level of right-wing extremist crime in 20 years. https://www.politico.eu/article/germany-records-highest-level-of-right-wing-extremist-crimes-in-20-years/

Ribeiro FN, Araújo M, Gonçalves P, Gonçalves MA, Benevenuto F (2016) Sentibench-a benchmark comparison of state-of-the-practice sentiment analysis methods. EPJ Data Sci 5(1):1–29

Ruder S, Peters ME, Swayamdipta S, Wolf T (2019) Transfer learning in natural language processing. In: Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: tutorials, pp 15–18

Scanlon JR, Gerber MS (2014) Automatic detection of cyber-recruitment by violent extremists. Secur Inf 3(1):5

Smeureanu I, Bucur C et al (2012) Applying supervised opinion mining techniques on online user reviews. Inf Econ 16(2):81–91

Sun C, Qiu X, Xu Y, Huang X (2019) How to fine-tune bert for text classification?. In: CCL

Tanoli IK, Pais S (2020) A lexicon based approach to detect extreme sentiments

Thelwall M, Buckley K, Paltoglou G (2012) Sentiment strength detection for the social web. J Am Soc Inf Sci Technol 63(1):163–173. https://doi.org/10.1002/asi.21662

University of Arizona Artificial Intelligence Lab, D.H.C. AZSecure-data: Ansar1 Forum Dataset (2013)

VOA (2020) Germany sees right-wing extremism as top security threat. https://www.voanews.com/a/europe_germany-sees-right-wing-extremism-top-security-threat/6192511.html

Vadicamo L, Carrara F, Cimino A, Cresci S, Dell'Orletta F, Falchi F, Tesconi M (2017) Cross-media learning for image sentiment analysis in the wild. In: Proceedings of the IEEE international conference on computer vision, pp 308–317

Vadicamo L, Carrara F, Cimino A, Cresci S, Dell'Orletta F, Falchi F, Tesconi M (2017) Cross-media learning for image sentiment analysis in the wild. In: 2017 IEEE international conference on computer vision workshops (ICCVW), pp 308–317. https://doi.org/10.1109/ICCVW.2017.45

Wagh B, Shinde J, Kale P (2018) A twitter sentiment analysis using NLTK and machine learning techniques. Int J Emerg Res Manag Technol 6(12):37–44

Wu Y, Schuster M, Chen Z, Le QV, Norouzi M, Macherey W, Krikun M, Cao Y, Gao Q, Macherey K, et al (2016) Google's neural machine translation system: bridging the gap between human and machine translation. arXiv preprint arXiv:1609.08144

Zhang A, Lipton ZC, Li M, Smola AJ (2020) Dive into deep learning. https://d2l.ai

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.