Customized BERT with Convolution Model: A New Heuristic Enabled Encoder for Twitter Sentiment Analysis

Fatima-ezzahra LAGRARI[1], Youssfi ELKETTANI[2]

Department of Mathematics, Ibn Tofail University of Science

Kenitra, Morocco

Abstract—**The Twitter messaging service has turned out to be a domain for news consumers and patrons to convey their sentiments. Capturing these emotions or sentiments in an accurate manner remains a major challenge for analysts. Moreover, the Twitter data include both spam and authentic contents that often affects accurate sentiment categorization. This paper introduces a new customized BERT (Bidirectional Encoder Representations from Transformers) based sentiment classification. The proposed work consists on pre-processing and tokenization step followed by a customized BERT based classification via optimization concept. Initially, the collected raw tweets are pre-processed via "stop word removal, stemming and blank space removal". Prevailing semantic words are acquired, from which the tokens (meaningful words) are extracted in the tokenization phase. Subsequently, these extracted tokens will be subjected to classification via optimized BERT, which weights and biases are optimally tuned by Standard Lion Algorithm (LA). In addition, the maximum sequence length of BERT encoder is updated with standard LA. Finally, the performance of the proposed work is compared over other state-of-the-art models with respect to different performance measures.**

Keywords—Twitter data; sentiment analysis; tokenization; optimized BERT; Lion Algorithm

I. INTRODUCTION

The Internet has become a platform for online learning, exchanging ideas and sharing opinions. Social media like Twitter, Facebook, Google+ can be referred to the group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content", as defined by Kaplan and Haenlein [9,8].

After the introduction of social media, the globe is entirely connected and hence aids users to exchange their information at any instance of time with lower cost and lower delivery time. In real-time, Twitter is a renowned "social micro-blogging service" that permits the users to post their opinions in the form of shorter messages within 140 characters or less and these short messages are referred as tweets [10-14]. In Twitter, training data are typically obtained by either assuming that tweets' polarities (positive, negative, neutral) can be inferred using emotions or by taking consensus from the results returned by the sentiment detection websites. Sentimental Analysis deals with getting know the real opinion/voice of people on specific product, services, organization, movies, news, events, issues and their attributes.

Twitter sentiment analysis has attracted much attention due to the rapid growth in Twitter's popularity as a platform for people to express their opinions towards a great variety of topics [15-17]. Approaches to Twitter sentiment analysis tend to focus on the identification of sentiment of individual tweets (tweet-level sentiment detection). Broadly speaking, existing work on tweet-level sentiment detection follows two main approaches, namely machine learning and lexicon-based approach. The supervised learning and unsupervised learning are the two categories of the machine learning approach. Sentiment classification using machine learning approach consists of two steps: feature extraction and classification with algorithms. Supervised learning approaches require training data for sentiment classifier learning, which is more computationally complex [18-21]. The conventional techniques on the Twitter sentiment analysis comprise of supervised learning schemes and dictionary-oriented techniques for sentiment classification. However, a most important challenge regarding the machine learning scheme is the selection of features that lead to minimal sparsity. The two main challenges of sentiment analysis are: (1) tweets are generally written in informal language (2) short messages show limited cues about sentiment and (3) acronyms and abbreviations are extensively used on Twitter [25,23].

Moreover, ANN (Artificial Neural Network) model performs better in most of the experiments while comparing to Fuzzy logic. ANN for the purpose of classification of sentiments helps to gain the accuracy in terms of correlations and dependencies [22,24]. The optimization algorithms have undergone various improvements in terms of many factors. One among them is by introducing adaptive operators or adaptive functions [26-29].

The major contribution of this research work consists of:

- An optimized BERT (framework, whose maximum sequence length of encoder is updated by the renowned standard LA.

- Further, the weight and bias of BERT framework are fine-tuned by the LA that ensures the prediction accuracy.

The rest of the paper is organized as follows: the recent works in sentiment analysis are discussed in Section 2. The pre-processing and tokenization steps are depicted in Section 3. Further, in Section 4 the optimized BERT for sentiment classification with Lion Algorithm: objective function and solution encoding are presented. The resultant acquired with the presented work is discussed in Section 5. A strong conclusion of this research work is provided in Section 6.

II. LITERATURE REVIEW

A. Related Works

In 2018, Jianqiang *et al*. [1] have introduced word embedding using unsupervised learning on large Twitter corpora. Further, in between the tweet and the word there is a co-occurrence statistical character and in supplement the latent contextual semantic relationships are also present. The sentiment feature set was formed by word sentiment polarity score features as well as the n-grams features with the aid of the word embedding. Finally, the sentiment classification labels were trained and determined in the deep convolution neural network, which intakes the sentiment feature set of tweets as input.

In 2020, Phan *et al*. [2] have introduced a novel approach for sentiment analysis from the Twitter data. This approach was developed on the basis of a "feature ensemble model" that had encapsulated the fuzzy sentiment, which had considered the elements like the "lexical, word-type, semantic, position, and sentiment polarity of the words".

In 2019, Iqbal *et al*. [3] have constructed a novel integrated framework for Twitter sentiment analysis. The authors have introduced a novel GA (Genetic algorithm) with the intention of enhancing the scalability of the classifier by means of reducing the feature dimensions. The evaluation of the proposed model was made over the existing feature reduction approaches like the PCA (Principal component analysis) and LSA (Latent Semantic Analysis).

In 2020, Ruz *et al*. [4] have developed a new Sentiment analysis approach based on Bayesian network classifiers. The authors used the Bayes factor approach in order to curtail the edges automatically during the training mechanism. The evaluation of the proposed approach was made on two Spanish datasets: The 2010 Chilean earthquake and the 2017 Catalan independence referendum. The resultant of the evaluation had exhibited the effectiveness of the presented work over the existing works.

In 2020, Nagamanjula and Pethalakshmi [5] have developed LAN^2FIS for opinion mining and sentiment analysis. Here, the features were selected from the data (public tweets) using a bi-objective optimization (minimum redundancy and maximum relevancy). Further, with the intention of solving the issues regarding the computation time, they have implemented the proposed framework in a "parallel and distributed way" with the aid of the "Hadoop framework with the MongoDB database".

In 2020, Ombabi *et al*. [6] have introduced a novel deep learning model on the basis of the one-layer CNN architecture for more efficient Arabic language sentiment analysis. The authors have extracted the local feature using the one-layer CNN architecture and the long-term dependencies were maintained with two layers LSTM (Long Short-Term Memory). The final classification resultant was acquired from SVM (Support Vector Machine), which intakes the resultant from LSTM and CNN.

In 2017, Pandey *et al*. [7] have developed a novel met heuristic method (CSK) for efficient Sentiment analysis and this approach was based on the "K-means" and Cuckoo Search. From the Twitter dataset, the sentimental contents having the optimum cluster-heads were explored with the proposed method. They have compared the proposed work over the existing models and the resultant had exhibited the efficiency of the proposed model.

In 2019, Abid *et al*. [8] have constructed joint architecture by means of combing the CNN and the RNN (Recurrent Neural Network) for Sentiment analysis. Initially, the "long-term dependencies" were captured with the RNN, they were captured with the help of the CNN global average pooling layer, the syntax and vocabulary. Based sentiments issues were solved with the GloVe (Global Vectors) based word embedding method. The resultant of the model had exhibited a higher performance with this slight hyper parameter tuning.

B. Review

Table I gives a summary of the related works presented in the literature section in terms of features and challenges.

At first, DCNN was introduced in [1], which offers a better product model and it also includes improved purchase decisions. However, there was no consideration of positive and negative opinion words. Fuzzy approach was exploited in [2] that fuses more online reviews and it also offers better ranking on products, but it needs more convenient purchase decisions. In addition, GA was deployed in [3,33] that avoid the redundant outcomes and it offers improved accuracy. Anyhow, it requires automatic syntactic rule extraction. Likewise, Bayesian network was suggested in [4] that offers improved analysis of service and products and it also concerns on better prediction on sentiments, however, it needs implementation of feedback loop during the training process. Likewise, LAN^2FIS was exploited in [5], which deals rapidly with consumer reviews and it is more effective but requires more application studies. Further, CNN was exploited in [6] that eliminates noise and it offers a better classification of sentiments, anyhow it was not much adaptable to all languages. CSK was implemented in [7], it offers better sensitivity and it also offers improved accuracy, but it will be interesting to include more attributes. At last, CNN and RNN was suggested in [8] that provides timely responses and it also recognizes negative reviews. However, it necessitates additional contextual factors. These limitations have to be considered for improving the sentiment analysis currently, and in future as well.

TABLE I. FEATURES ANS CHALLENGES OF EXISTING SENTIMENT ANALYSIS APPROACH

Author	Approach	Features	Challenges
Jianqiang et al. [1]	DCNN	Improved purchase decisions Better product model	No consideration on positive and negative opinion words
Phan et al. [2]	Fuzzy	Fuses more online reviews. Better ranking on products	Needs more convenient purchase decisions
Iqbal et al. [3]	GA	Avoids the redundant outcomes Better accuracy	Requires automatic syntactic rule extraction
Ruz et al. [4]	Bayesian network	Improved analysis on service and products. Better prediction on sentiments	Needs implementation of feedback loop during training process
Nagamanjula et al. [5]	LAN2FIS	Deals rapidly with consumer reviews. More effective.	Requires more application studies
Ombabi et al. [6]	CNN	Eliminates noise. Better classification of sentiments	It was not much adaptable to all languages
Pandey et al. [7]	CSK	Better sensitivity Improved prediction accuracy	Need to concern on more attributes
Abid et al. [8]	CNN and RNN	Enhanced timely responses. Recognizes negative reviews	Necessitates additional contextual factors
Lagrari et al. [33]	GA and random forests	Avoids the redundant outcomes. Better accuracy	Computationally expensive Difficulty of designing objective function
Elkhechafi et al. [35]	GA and Firfly algorithm	Best Success rate Good ability to deal with multimodality	efficient in solving optimization problem Better in terms of time

III. PROPOSED TWITTER SENTIMENT ANALYSIS MODEL

A novel sentiment classification approach is developed for accurate detection of the sentiments from the Twitter data D. The proposed model encapsulates two major steps: "(a) Pre-processing and tokenization, (b) classification". The diagrammatic representation of the presented work is illustrated in Fig. 1. Initially, the raw data are subjected to pre-processing that includes three different steps like "stop word removal, stemming, blank space removal". The pre-processed words $D_{pre-process}$ are subjected to tokenization, in which the stream of words is broken into symbols, words and other meaningful elements referred as "tokens". At the end of tokenization, only specific meaningful words are selected. These tokenized words are denoted as D_{Token}, which is classified via optimized BERT framework [32]. As a major contribution, the weight and bias of the BERT framework is optimized using the standard LA (Lion Algorithm) [31]. In addition, to make the proposed work applicable for huge datasets, the proposed optimized BERT is customized by updating the maximum sequence length of BERT encoder by standard LA. Finally, the optimized BERT framework generates the classified results such as positive, negative or neutral sentiment.
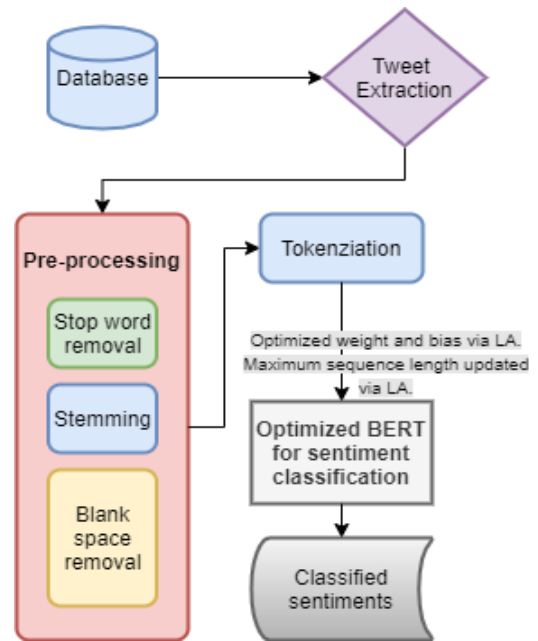

Fig. 1. Proposed Sentiment Analysis Framework.

IV. PRE-PROCESSNIG AND TOKENIZATION

A. Pre-Processing

Initially, the raw tweets D are collected from three standard databases (see the experimental section).

The data-processing is a crucial step that is applied to any of the collected data before embedding it with sentiment extraction approach. In general, the data pre-processing permits generating text classification via higher quality as well as to diminish the computational complexity. In this research work, the pre-processing step consists of stemming, stop words removal and blank space removal.

1) Stemming: It is the mechanism of supplanting words with their stems, or roots. For the BOW (bag of words), the dimensions are lessened during the mapping of the root-related words into a unit word. For illustration, the words "reading, read and reader" are the root-related words and they get mapped into a single word "read". Apart from this, during the application of the stemming, the bias might get increased. As a resultant, the over-stemming (i.e. "experiment" and "experience" gets mapped into "exper") and under-stemming ("adhere" and "adhesion" gets merged) errors might occur. Over stemming brings down accuracy and under-stemming brings down recall.

2) Stop-words removal: In a sentence, the connecting function between the words is given by the stop words. These stop words add meaning to the document during the construction of the Natural Language Processing model or text data assessment. The most commonly utilized stop-words are "the", "is", "at", "which" and "on". Further, before performing the classification, the removal of the stop-words takes place as they are more frequent and do not influence the sentence's final sentiment.

3) Blank space removal: Since the blank space increases the dimensionality between the words, they are to be rejected. Once the blank space or extra whitespace or tab spaces are identified in the sentence, they are removed and replaced by a whitespace. In addition, the "Twitter hashtags, retweets, word capitalization, word lengthening, question marks, presence of web addresses in tweets, exclamation marks, internet emoticons and internet shorthand/slangs" are also removed. At the end of pre-processing, extraction of certain keywords takes place. The extracted keywords are subjected to further processing. The pre-processed words $D_{pre-process}$ are subjected to tokenization.

B. Tokenization

In general, the tokenization is the mechanism of creating a BOW from $D_{pre-process}$. The breaking of the approaching string into comprising words and different components. The singular words can be distinguished with normal separator like whitespace; anyway, different symbols can likewise be utilized. Tokenization of web-based social networking information is significantly more troublesome than tokenization of the overall content since it contains various emojis, URL links, contractions that can't be effectively isolated as entirety substances. The consolidation of the accompanying words into "phrases or n-grams" is the overall practice and it can be "unigrams, bigrams, trigrams, and so on". In general, a single word is said to be a Unigrams, while assortments of two neighbouring words in a text is said to be bi- grams and trigrams are assortments of three neighbouring words. N-grams based tokenization technique can diminish predisposition, yet may increment factual inadequacy. It has been demonstrated that the utilization of n-grams can improve the quality of text characterization. At the end of tokenization, only specific meaningful words are selected. These tokenized words are denoted as D_{Token}, which is classified via optimized BERT framework.

V. OPTIMIZED BERT FOR SENTIMENT CLASSIFICATION WITH LION ALGORITHM THE TEMPLATE

A. Optimized BERT Framework

BERT is referred as "Bidirectional Encoder Representations from Transformers". This approach was developed in [32] with the objective of pre-training the deep bidirectional representations that was utilized to create the NLP from unlabelled texts. This was done in all layers by means of conditioning both the left and right context. Typically, the BERT framework encloses three major parts: Input layer, BERT encoder and output layer. The BERT framework is illustrated in Fig. 2.

1) Input Layer: The input layer is fed with D_{Token} that has K count of words. This is denoted as $D_{Token(1:K)} = D_{Token(1)}, D_{Token(2)}, .., D_{Token(i)}, ..., D_{Token(K)}$,in which $D_{Token(i)}$ is the ith word of the tokenized input sequence and it is $1 \leq D_{Token(i)} \leq K$. In one token sequence, the input sequence can be represented either be a couple of text sequence or a unit text sequence. The first token is always the "CLS"

which encapsulates the classification embedding. In addition, the segments are separated with special token "SEP".

2) Proposed BERT encoder: It is a "multi-layer bidirectional Transformer encoder" with 12 transfer blocks and the maximum sequence length of 512 tokens (pre-trained). The output from the encoder is the representations of the sequence and it can be a hidden state vector or the "hidden state vector's time-step sequence". Here, the final "hidden state vector" is utilized in this research work and the standard LA is deployed here to predict the best sequence token among the maximum sequence count. Moreover, the maximum sequence count is pre-trained and it couldn't be utilized for huge datasets. Thus, to make the sentiment classification applicable for huge datasets, the maximum sequence count of the BERT encoder is updated with standard LA.

3) Output layer: It is a simple "softmax classifier" that is embedded at the top of the Proposed BERT encoder. This helps in predicting the probability of the labels C in the D_{Token}. This is mathematically expressed in Eq. (1), in which H is the final hidden state and T is the task-specific parameter matrix.

$$P(C|H) = \text{Soft max}(T.H) \tag{1}$$

On the other hand, during the training stage, the weight as well as bias is fixed and pre-trained, since BERT is a "pre-trained model". But a natural question has arisen, whether the pre-trained bias could be proficient in processing natural languages of any data scale. This is bit complex with the pre-trained bias as well as pre-trained weights, since the datasets of Twitter is bulky. Thus, in this research work, the bias (B) and the weight (W) of BERT will be trained with the standard LA.
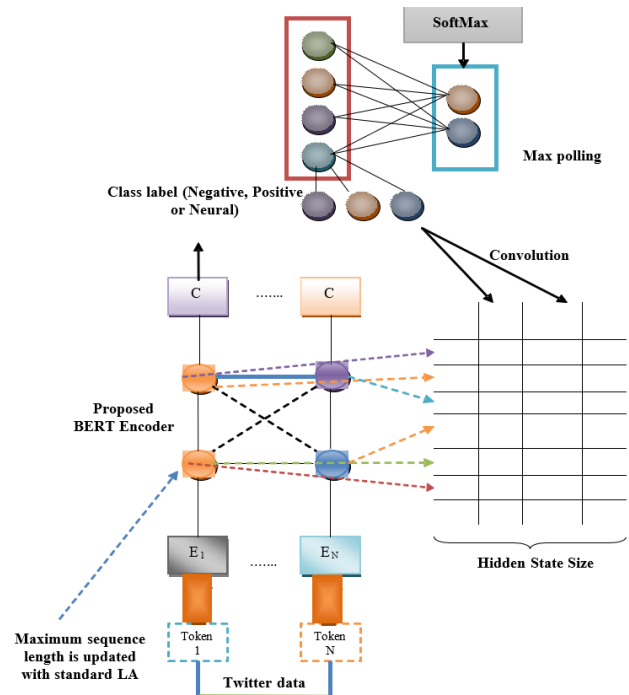

Fig. 2.   Proposed Bert Framework.

B. Objective Function and Solution Encoding

As mentioned above, the weight (W) and the bias (B) of the BERT model is fine-tuned by LA model. The input solution to the algorithm is shown in Fig. 3. Moreover, the objective function defined in the work is enhancing accuracy (Acc), which is expressed in Eq. (2).

$$Ob = max(Acc) \tag{2}$$

C. Standrad LA

LA is a natural inspired optimization algorithm that was developed on the basis of the unique social behaviour of the lions, particularly, terrestrial defence and territorial takeover. In between the nomadic as well as residual males occurs the terrestrial defence, while the terrestrial take over exists between the old territorial and the new territorial males. The steps followed in the standard LA are described below [31].

1) Step 1 - Pride Generation and fitness evaluation: In this step, the initialization of the pride's male territorial lion X^{male} and female territorial lion X^{female} and nomad lion X^{nomad} take place. The arbitrary solutions for X^{male}, X^{female} and X_1^{nomad} are termed as P. In addition $fn(X^{mfemale})$, $fn(X^{male})$ and $fn(X_1^{nomad})$ refers to the fitness of X^{male}, X^{female} and X_1^{nomad}. During the initialization, the reference fitness is set as $fn^{ref} = fn(X^{male})$ and the generation count is fixed as $N_g = 0$, which is described at the termination step.

2) Step 2 - Fertility Evaluation: X^{male} and X^{female} are used for fertility and this fertility evaluation benefits the solutions to get away from local optima as well as convergence issues. The updated female lion is denoted by $X^{female+}$, which is ensured by the "sterility rate" S_r that increases by 1 at the end of the crossover. The updated female lion is achieved with X^{female} and the random integer p that is within the interval $[1, P]$. This is expressed in Eq. (3).

$$\{X^{female+} = min[X_p^{max}, max(X_p^{min}, \lambda_p)]\} \tag{3}$$

Moreover, the female renewal function λ and the random integers ra_1 and ra_2 are generated within the interval $[0, 1]$.

$$\lambda_p = \{X_p^{female} + [0.1ra_2 - 0.05(X_p^{male} - ra_1 X_p^{female})]\} \tag{4}$$

3) Step 3 - Mating and Cub growth: X^{male} and X^{female} goes through crossover and mutation operation. Among them, the crossover is performed initially and it is based on the littering rate of the lion. At the end of crossover and mutation, X^{cubs} with X^{m_cub} male cub and X^{f_cub} female cub are produced. On the basis of the fitness, X^{cubs} are formed. Further, "Cub growth function is a local solution search function" for the male and female cubs. The random mutation rate G_r approves this mechanism. The previous X^{m_cub} and X^{f_cub} is replaced by the mutated X^{m_cub} and X^{f_cub}, only if X^{m_cub} and X^{f_cub} is better when compared to the existed mutation.

4) Step 4 - Territorial Defence and takeover: With the aid of search space, the identification of the territorial defence takes place. This can be given as "nomad coalition, pride and survival fight". In general, the territorial takeover is said to be

the mechanism of providing territory to the male as well as female cubs as they become matured and stronger. More particularly, terrestrial take over occurs only, when the age of the cub (A^{cub}) is greater than or equal to the maturity age. The mathematical equation corresponding to the selection of X^{e_nomad} are depicted in Eq. (5), Eq. (6) and (7), respectively.

$$fn(X^{e\_nomad}) < fn(X^{male}) \tag{5}$$

$$fn(X^{e\_nomad}) < fn(X^{m\_cub}) \tag{6}$$

$$fn(X^{e\_nomad}) < fn(X^{f\_cub}) \tag{7}$$

5) Step 5 – Termination: The algorithm terminates when the count of fitness goes beyond the limit. This is expressed via two conditions as per Eq. (8) and Eq. (9), respectively.

$$N_g > N_{g(max)} \tag{8}$$

$$\lfloor fn(X^{male}) - fn(X^{optimal}) \rfloor \leq T_e \tag{9}$$

The error threshold is specified as T_e, and "maximum count of the generations" is represented and the target minimum is depicted as $N_{g(max)}$ and $f(X^{optimal})$. The flow chart of standard LA is shown in Fig. 4.
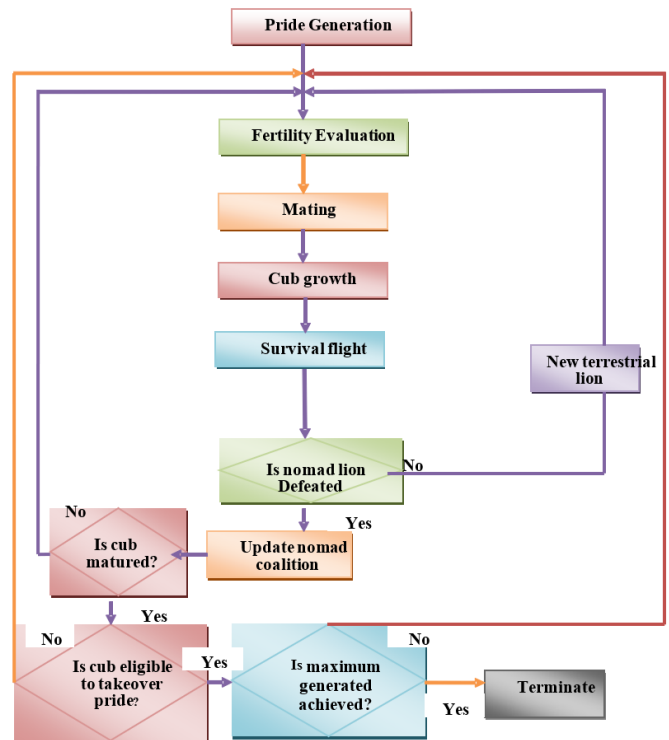

Fig. 3.   Solution Encoding.

Fig. 4.   Flowchart of Standard LA.

VI. RESULTS AND DISCUSSIONS

A. Experiments

The proposed sentiment classification with optimized BERT was implemented in Python and the corresponding outcomes acquired are noted. The experimentation was carried out using three Datasets:

Dataset 1 (Brands and Product Emotions[1]) contains 9094 rows with 3 variables where contributors evaluated tweets about multiple brands and products. The crowd was asked if the tweet expressed positive, negative, or no emotion towards a brand and/or product. If some emotion was expressed, they were also asked to say which brand or product was the target of that emotion.

Dataset 2 (TWCS for Customer Support on Twitter[2]) is a large (3 million tweets), modern corpus of tweets and replies to aid innovation in natural language understanding and conversational models, and for study of modern customer support practices and impact.

Dataset 3 is called sentiment 120 Dataset, which contains 1,600,000 tweets extracted using the Twitter API. The tweets have been annotated (0 = negative, 4 = positive) and they can be used to detect sentiments [34].

Further, the proposed optimized BERT model was compared over the existing models like IB-K[3], NB [9], SMO [3], Bayesian Net [4], jRip [3], j48 [3], PART [3], CNN[1] and BERT [30] in terms of specificity (true negative rate), sensitivity (True positive rate), accuracy, precision (Positive Predictive Value), FPR (False positive rate), FNR (False negative rate), FOR (False omission rate), NPV (Negative predictive value), FDR (False discovery rate), F1-Score and MCC (Matthews correlation coefficient), respectively. This evaluation is done by varying the Training Percentage (TP).

B. Analysis on Dataset 1 (Brands and Product Emotions): Performance and Error

The Dataset is evaluated in terms of positive measures, negative and other measures. The resultants are graphically exhibited in Fig. 5, Fig. 6 and Fig. 7, respectively. The positive measures like "Accuracy, sensitivity, balancing accuracy" and precision are shown in Fig. 5. It is observed that the proposed work attains accurate results when compared over other conventional models, which ensures the fulfilment of the objective defined in this work. On observing the accuracy of the presented work at TP = 90 is 92.2, which is 15.9%, 14.4%, 11%, 10.3%, 9.46%, 5.85%, 3.15%, 1.67% and 0.75% better than the existing approaches like IB-K, NB, SMO, Bayesian Net, jRip, j48, PART, CNN and BERT, respectively.

In case of sensitivity, the maximum sensitivity is recorded by the presented work for every variation in TP Fig. 5(b). Among the sensitivity of the presented work, the maximal sensitivity of 91.2 is recorded at TP=90. Moreover, the balancing accuracy of the presented work is higher at TP=90

and it is 22%, 17.8%, 14.2%, 13%, 8.87%, 7.47%, 4.4%, 2.2% and 0.757% better than the existing models like IB-K, NB, SMO, Bayesian Net, jRip, j48, PART, CNN and BERT, respectively. In addition, the precision of the presented work is higher than all the existing works as per Fig. 5(d). The highest values recorded by the presented work at TP= 40, TP=50, TP=60, TP=70, TP=80 and TP=90 are 74.4, 81.6, 81.8, 86.8, 93.2 and 94, respectively.

On the other hand, the negative measures like FNR, FPR, and FDR and FOR also help in exhibiting the enhancement level of the presented work. The lower the error measures, higher the accuracy of the classification. The FNR (in Fig. 6(a)) of the presented work is lower at every variation in TP. At TP=90, the FNR of the presented work is 78.5 and it is 65.1%,62.97%,56.3%,53.82%,50.93%,40.53%,21.5%,16.4% and 8.18% better than IB-K, NB, SMO, Bayesian Net, jRip, j48, PART, CNN and BERT, respectively. Then, in case of FPR, the lowest value is recorded by the presented work as 76.2 at TP=40% and in all other variation in TP's also the presented work records the lowest value. In addition, the FDR and FOR of the presented work is lower for every variation in TP. The lowest FDR is recorded by the presented work at TP=60 (10.9).

In addition, the other measures like NPV, PPV, MCC and F1-Score of the concern database is shown in Fig. 7. All these measures exhibit higher performance with the presented work, while compared with the existing one. The NPV of the presented work is higher at TP=90 and it is 22%, 17.8%, 14.2%, 13%, 8.87%, 7.47%, 4.4%, 2.22% and 7.57% better than IB-K, NB, SMO, Bayesian Net, jRip, j48, PART, CNN and BERT, respectively. Thus, as a whole the presented work shows the highest positive performance and lowest negative performance, which makes it much suitable for sentiment classification.

The overall training error performance of the presented work over the existing work is shown in Table II. The overall error performance of the presented work is lower at TP=90 and it is 75.1%, 73.5%, 67.7%, 66%, 64.3%, 52.8%, 37.4%, 24.1% and 12.6% better than existing IB-K, NB, SMO, Bayesian Net, jRip, j48, PART, CNN and BERT, respectively.

C. Analysis on Customer Support on Twitter Dataset (TWCS)

For a matter of clarity, we present here only the overall training error performance of the proposed work over the existing work is shown in Table III. Here, the presented work shows the lowest performance, while compared to the existing works. The lowest performance is revealed by the presented work at TP=90.

D. Analysis on Sentiment 120 Dataset

The overall training performance of the presented work over the existing work is tabulated in Table IV. The lowest error performance is recorded by the presented work, while compared to the existent works. At TP= 90, the presented work shows the lowest error value is 5.72, which is 68.75, 64.4%, 60.2%, 51.95, 43.8%, 40%, 39.4%, 25.7% and 12.75 better than the existing models like IB-K, NB, SMO, Bayesian Net, jRip, j48, PART, CNN and BERT, respectively.
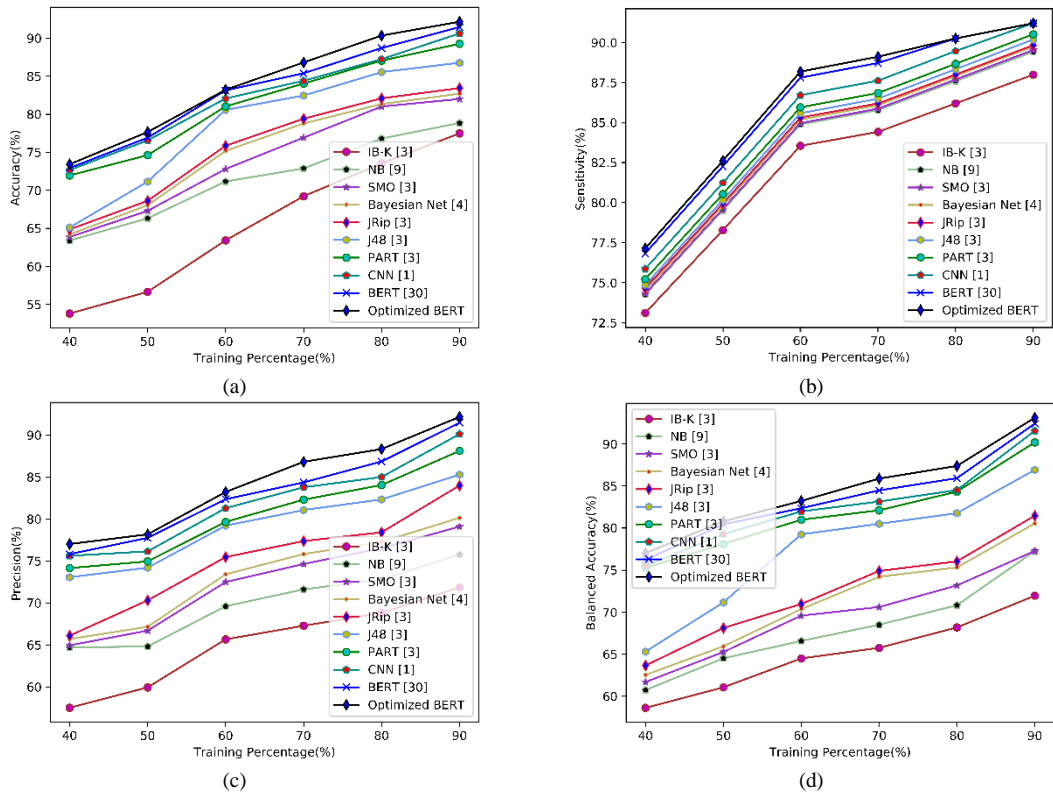
---

[1] https://data.world/crowdflower/brands-and-product-emotions
[2] https://www.kaggle.com/thoughtvector/customer-support-on-twitter

Fig. 5. Performance Analysis of Presented Work over Existing Approaches in Terms of Positive Measures for Brands and Product Emotions Dataset Showing (a) Accuracy, (b) Sensitivity, (c) Precision and (d) Balanced Accuracy.
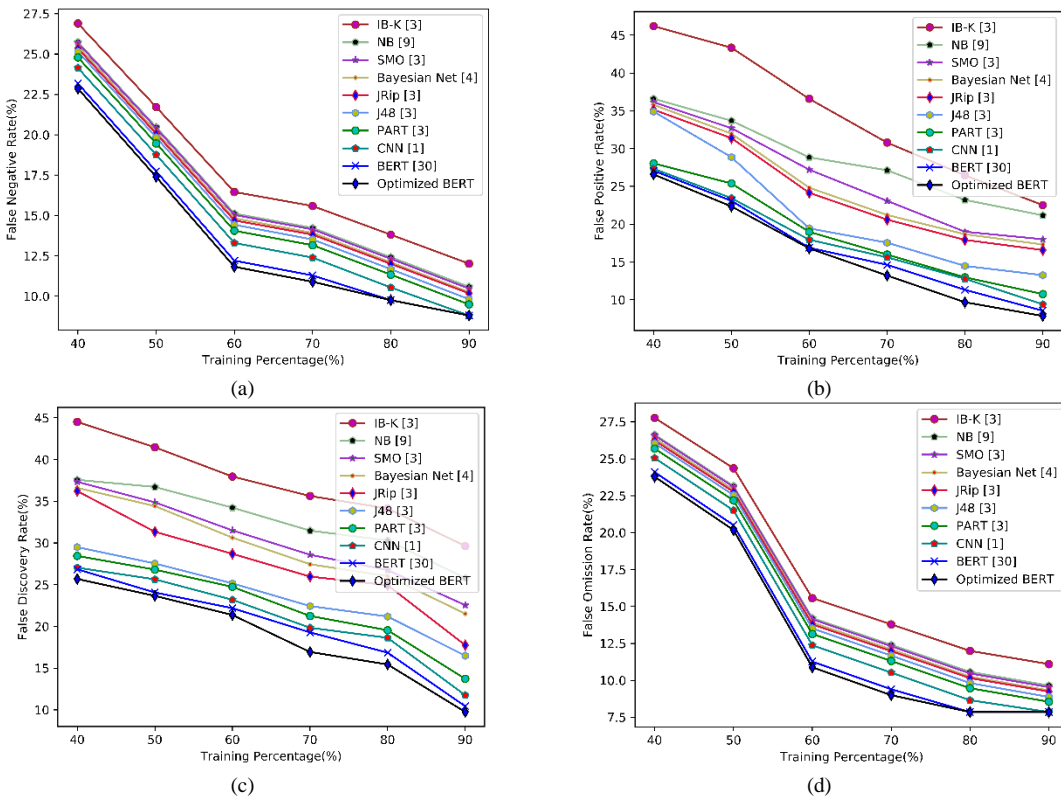

Fig. 6. Performance Analysis of Presented Work over Existing Approaches in Terms of Negative Measures for Brands and Product Emotions Dataset Showing (a) FNR, (b) FPR, (c) FDR and (d) FOR.
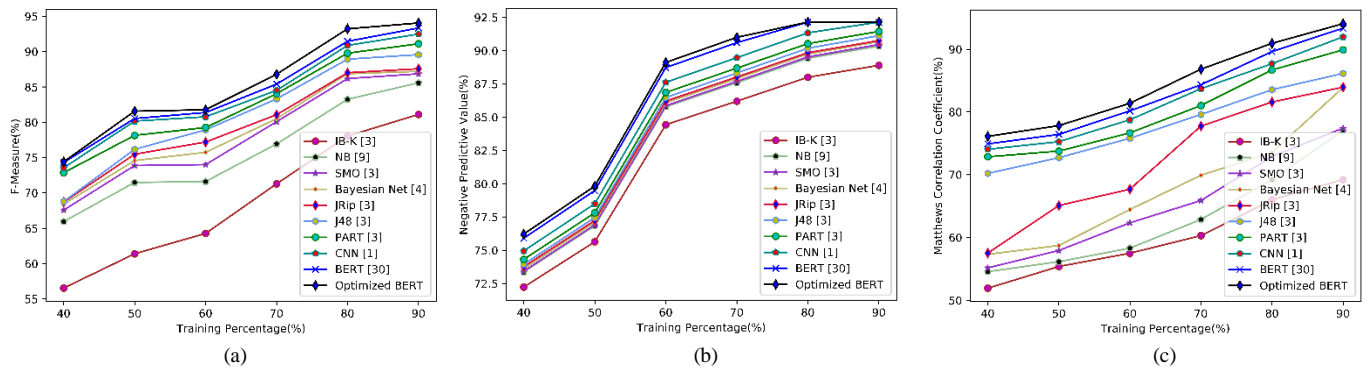
Fig. 7. Performance Analysis of Presented Work over Existing Approaches in Terms of other Measures for Brands and Product Emotions Dataset (a) F1-Score, (b) NPV, (c) MCC.

TABLE II. OVERALL TRAINING ERROR PERFORMANCE EVALUATION OF BRANDS AND PRODUCT EMOTIONS DATASET (%)

Methods	TP=40	TP=50	TP=60	TP=70	TP=80	TP=90
IB-K [3]	34.4	33.3	30.3	25.6	23.4	20.1
NB [9]	22.7	22	21.8	21.6	20	18.7
SMO [3]	22.1	20.8	20	17.3	15.6	1.55
Bayesian Net [4]	21.7	20	17.4	15.3	15.3	14.7
jRip [3]	20.9	19.3	16.6	14.6	14.5	14
j48 [3]	20.6	16.3	11.5	11.3	10.9	10.6
PART [3]	12.2	12.2	11	9.65	9.34	7.99
CNN [1]	11.4	9.93	9.82	9.26	9.13	6.59
BERT [30]	11.1	9.51	8.65	8.19	7.62	5.72
Optimized BERT	**10.5**	**8.64**	**8.56**	**6.67**	**5.91**	**5.00**

TABLE III. OVERALL TRAINING ERROR PERFORMANCE OF TWCS DATASET (%)

Methods	TP=40	TP=50	TP=60	TP=70	TP=80	TP=90
IB-K [3]	33.8	31.2	29.8	25.2	21.1	18.9
NB [9]	21.6	20.6	19.9	19.5	17.6	11.8
SMO [3]	21.3	19.1	18.6	18.2	12.2	11.5
Bayesian Net [4]	20.8	18.1	18	15.5	12	10.9
jRip [3]	20.6	16.7	15.6	14.6	11.6	10.8
j48 [3]	16.3	14.8	13.9	13.6	11.4	9.92
PART [3]	16.2	14.2	12.5	12.3	10	9.87
CNN [1]	15.6	13.9	12.2	12.1	9.77	9.01
BERT [30]	14.3	13.4	11.9	10.7	9.63	8.07
Optimized BERT	**12.6**	**12.1**	**10.9**	**10.4**	**9.18**	**7.68**

Both positive and negative measurements of optimized BERT on the three datasets gives a better result over the existing approaches.

For future work, we plan to take into consideration an important aspect in sentiment analysis which is emoticons that can reflect the mood of the writer. Another aspect could be the comparison of the training and execution time over the existing approaches.

TABLE IV. OVERALL TRAINING ERROR ANALYSIS OF SENTIMENT 120 DATASET (%)

Methods	TP=40	TP=50	TP=60	TP=70	TP=80	TP=90
IB-K [3]	27.7	25.8	24.3	23.8	22.9	18.3
NB [9]	21.3	20.1	19.5	19.4	16.7	16.1
SMO [3]	19.7	18.6	17.4	16.7	15	14.4
Bayesian Net [4]	16.2	16	13.6	12.9	12.4	11.9
jRip [3]	16	15	12.7	12.4	10.2	10.2
j48 [3]	15.9	14.9	12.3	11.7	9.90	9.57
PART [3]	15.5	12	11.6	11.3	9.49	9.44
CNN [1]	14.3	11.9	10.5	10.1	8.64	7.71
Optimized BERT	**12.9**	**11.5**	**9.99**	**9.56**	**8.34**	**6.55**

VII. CONCLUSION

A new customized BERT based sentiment classification was introduced in this research work. The proposed work includes two major phases: pre-processing and tokenization, and Customized BERT based classification via optimization concept. The data collected was pre-processed with "based classification via optimization concept", which was then tokenized. Prevailing semantic words were acquired, from which the tokens (meaningful words) were extracted in the tokenization phase. The optimized BERT was introduced for classifying the tokens. In the optimized BERT, the weight and biases are optimally tuned by Standard LA. In addition, the maximum sequence length of BERT encoder was updated with standard LA. It is observed that the proposed work attains accurate results when compared over other conventional models, which ensures the fulfilment of objective defined in this work. On observing the accuracy of the presented work for Brands and Product Emotions dataset at TP =90 is 92.2, which is 15.9%, 14.4%, 11%, 10.3%, 9.46%, 5.85%, 3.15%, 1.67% and 0.75% better than the existing approaches like IB-K, NB, SMO, Bayesian Net, jRip, j48, PART, CNN and BERT, respectively.

REFERENCES

[1] Jianqiang, Z., Xiaolin, G., & Xuejun, Z. (2018). Deep convolution neural networks for twitter sentiment analysis. IEEE Access, 6, 23253-23260.

[2] Phan, H. T., Tran, V. C., Nguyen, N. T., & Hwang, D. (2020). Improving the performance of sentiment analysis of tweets containing

fuzzy sentiment using the feature ensemble model. IEEE Access, 8, 14630-14641.

[3] Iqbal, F., Hashmi, J. M., Fung, B. C., Batool, R., Khattak, A. M., Aleem, S., & Hung, P. C. (2019). A hybrid framework for sentiment analysis using genetic algorithm based feature reduction. IEEE Access, 7, 14637-14652.

[4] Ruz, G. A., Henríquez, P. A., & Mascareño, A. (2020). Sentiment analysis of Twitter data during critical events through Bayesian networks classifiers. Future Generation Computer Systems, 106, 92-104.

[5] Nagamanjula, R., & Pethalakshmi, A. (2020). A novel framework based on bi-objective optimization and LAN2FIS for Twitter sentiment analysis. Social Netw. Analys. Mining, 10(1), 34.

[6] Ombabi, A. H., Ouarda, W., & Alimi, A. M. (2020). Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks. Social Network Analysis and Mining, 10(1), 1-13.

[7] Pandey, A. C., Rajpoot, D. S., & Saraswat, M. (2017). Twitter sentiment analysis using hybrid cuckoo search method. Information Processing & Management, 53(4), 764-779.

[8] Abid, F., Alam, M., Yasir, M., & Li, C. (2019). Sentiment analysis through recurrent variants latterly on convolutional neural network of Twitter. Future Generation Computer Systems, 95, 292-308.

[9] Sailunaz, K., & Alhajj, R. (2019). Emotion and sentiment analysis from Twitter text. Journal of Computational Science, 36, 101003.

[10] Fang, Y., Tan, H., & Zhang, J. (2018). Multi-strategy sentiment analysis of consumer reviews based on semantic fuzziness. Ieee Access, 6, 20625-20631.

[11] Ahmad, S., Asghar, M. Z., Alotaibi, F. M., & Awan, I. (2019). Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Human-centric Computing and Information Sciences, 9(1), 24.

[12] Gutiérrez-Batista, K., Campaña, J. R., Vila, M. A., & Martin-Bautista, M. J. (2018, June). Fuzzy Analysis of Sentiment Terms for Topic Detection Process in Social Networks. In International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (pp. 3-14). Springer, Cham.

[13] Wang, D., Li, J., Xu, K., & Wu, Y. (2017). Sentiment community detection: exploring sentiments and relationships in social networks. Electronic Commerce Research, 17(1), 103-132.

[14] del Pilar Salas-Zárate, M., Medina-Moreira, J., Álvarez-Sagubay, P. J., Lagos-Ortiz, K., Paredes-Valverde, M. A., & Valencia-García, R. (2016, November). Sentiment analysis and trend detection in Twitter. In International Conference on Technologies and Innovation (pp. 63-76). Springer, Cham.

[15] Sánchez-Rada, J. F., & Iglesias, C. A. (2019). Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. Information Fusion, 52, 344-356.

[16] Liu, N., & Shen, B. (2020). Aspect-based sentiment analysis with gated alternate neural network. Knowledge-Based Systems, 188, 105010.

[17] Ahuja, R., Chug, A., Kohli, S., Gupta, S., & Ahuja, P. (2019). The impact of features extraction on the sentiment analysis. Procedia Computer Science, 152, 341-348.

[18] Ren, R., Wu, D. D., & Liu, T. (2018). Forecasting stock market movement direction using sentiment analysis and support vector machine. IEEE Systems Journal, 13(1), 760-770

[19] Tang, F., Fu, L., Yao, B., & Xu, W. (2019). Aspect based fine-grained sentiment analysis for online reviews. Information Sciences, 488, 190-204.

[20] Geetha, M., Singha, P., & Sinha, S. (2017). Relationship between customer sentiment and online customer ratings for hotels-An empirical analysis. Tourism Management, 61, 43-54.

[21] Qiu, J., Liu, C., Li, Y., & Lin, Z. (2018). Leveraging sentiment analysis at the aspects level to predict ratings of reviews. Information Sciences, 451, 295-309.

[22] López, M., Valdivia, A., Martínez-Cámara, E., Luzón, M. V., & Herrera, F. (2019). E2SAM: evolutionary ensemble of sentiment analysis methods for domain adaptation. Information Sciences, 480, 273-286.

[23] Ibrahim, N. F., & Wang, X. (2019). Decoding the sentiment dynamics of online retailing customers: Time series analysis of social media. Computers in Human Behavior, 96, 32-45.

[24] Xiong, S., Wang, K., Ji, D., & Wang, B. (2018). A short text sentiment-topic model for product reviews. Neurocomputing, 297, 94-102.

[25] Chen, M. H., Chen, W. F., & Ku, L. W. (2018). Application of Sentiment Analysis to Language Learning. IEEE Access, 6, 24433-24442.

[26] Rajakumar, B. R. (2013). Impact of static and adaptive mutation techniques on the performance of Genetic Algorithm. International Journal of Hybrid Intelligent Systems, 10(1), 11-22.

[27] Rajakumar, B. R. (2013). Static and adaptive mutation techniques for genetic algorithm: a systematic comparative analysis. International Journal of Computational Science and Engineering, 8(2), 180-193.

[28] Swamy, S. M., Rajakumar, B. R., & Valarmathi, I. R. (2013). Design of hybrid wind and photovoltaic power system using opposition-based genetic algorithm with Cauchy mutation.

[29] Rajakumar, B. R., & George, A. (2013). APOGA: An adaptive population pool size based genetic algorithm. AASRI Procedia, 4, 288-296.

[30] Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019, October). How to fine-tune bert for text classification?. In China National Conference on Chinese Computational Linguistics (pp. 194-206). Springer, Cham.

[31] Boothalingam, R. (2018). Optimization using lion algorithm: a biological inspiration from lion's social behavior. Evolutionary Intelligence, 11(1-2), 31-52.

[32] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[33] Lagrari, F. E., Ziyati, H., & El Kettani, Y. (2018, July). An Efficient Model of Text Categorization Based on Feature Selection and Random Forests: Case for Business Documents. In International Conference on Advanced Intelligent Systems for Sustainable Development (pp. 465-476). Springer, Cham.

[34] Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N project report, Stanford, 1(12), 2009.

[35] Elkhechafi, M., Hachimi, H., & Elkettani, Y. (2017). A new hybrid firefly with genetic algorithm for global optimization. Int. J. Manag. Appl. Sci, 3, 47-51.