Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN IMPROVED SENTIMENT CLASSIFICATION MODEL USING BERT CLASSIFICATION WITH RANGER ADABELIEF OPTIMIZER

[1]SHIRAMSHETTY GOUTHAMI , [2] DR . NAGARATNA P . HEGDE

[1] Research Scholar ,Computer Science and Engineering, Osmania University , Telangana, India.

[2]Professor, Computer Science and Engineering, Vasavi College of Engineering ,Telangana, India

[1] gouthami.shiramshetty@gmail.com, [2] nagaratnaph@staff.vce.ac.in

ABSTRACT

Sentiment analysis is a useful technique for extracting opinions from unstructured data that includes language like product and movie reviews. It can be used to gather customer feedback, evaluate brand perception, and conduct market research. In the field of Natural Language Processing (NLP), sentiment analysis of Twitter data has recently emerged as a popular topic. Twitter data sets are compiled using the Twitter API and contain real-time tweets about various topics. Automated text analysis is used to determine public opinion on specific issues. Although many machine learning algorithms are utilized for this purpose, previous approaches have failed to enhance the accuracy of categorization and classification as per industry requirements. To address this issue, a Long Short-Term Memory (LSTM) framework was introduced. This approach was applied to analyze airline customer feedback data from Twitter, and it resulted in improved classification performance. However, the LSTM framework achieve better performance for small and medium size datasets only. The performance degraded when applicable for large amount of data. To address the limitation of LSTM framework, in this research work proposed Bidirectional Encoder Representations from Transformers (BERT) classification and the Ranger AdaBelief optimizer. The BERT classifier is used for classification, and Ranger AdaBelief is employed to optimize loss during the classification process. The proposed model achieved better results than previous research, with an accuracy of 92.54%, precision of 90.15%, recall of 90.24%, and F1-score of 91.14%. The BERT classification with Ranger AdaBelief optimizer significantly outperformed previous approaches for sentiment analysis classification.

Keywords: *Data Pre-process, BERT Classifier, Ranger AdaBelief Optimizer, Sentiment Analysis, Error Rate Analysis*

1. INTRODUCTION

The widespread use of the internet has led to the rise of social media as an integral part of our lives. Among the different social media platforms, Twitter is the most popular and allows users to share their thoughts and ideas via tweets. The platform has become a primary source for gathering real-time numerical information on public sentiment. Natural Language Processing (NLP) is a tool used in sentiment analysis to determine the emotions conveyed in tweets. The goal of sentiment analysis is to determine how Twitter users feel about various topics. Sentiment analysis is an automatic way to identify public opinion by analyzing the polarities of textual information. To categorize the sentiment of tweets collected from Twitter, sentiment analysis is employed. Earlier studies utilized machine learning for categorizing tweets, but the effectiveness of machine learning was limited by the reputation of the data in huge datasets. Recent studies have utilized deep learning classifiers to increase classification accuracy. Prior to creating a deep learning model, relevant information is gathered from Twitter and redundant data is removed using pre-processing methods. The ability to extract and categorize public opinion on various topics has recently drawn the attention of researchers, resulting in academic and industrial applications. The polarity derived from sentiment analysis is valuable to numerous product manufacturing industries as it helps them to understand public opinion about their products. This research utilizes the Twitter API dataset to examine the sentiment of tweets using BERT classification with the Ranger AdaBelief optimizer to fine-tune the parameters used in classifying sentiment. The main contributions of this research are introducing an improved classification model using BERT classifier and Ranger AdaBelief optimizer to analyze the sentiment of tweets, and analyzing the error rate produced by the BERT classification model to evaluate its efficiency. Additionally, the feature extraction process is performed using the TF-IDF method.

The rest of the portion present in the paper is organized in the following way: Section 2 discussed the recent researches on sentiment analysis, section 3 provides the proposed work of this paper. The results and analysis is discussed in section 4 and at last, section 5 presents the overall summary of the paper.

2. RELATED WORKS

Chetanpal Singh et al. [16] have introduced a deep learning approach for evaluating the sentiment of Tweets regarding the reviews of COVID-19. The deep learning approach utilized Long-Short Term Model and Recurrent Neural Network (LSTM-RNN). Moreover, the algorithm make use of enhanced feature weighted layers by attention mechanism. The input data was obtained from publicly availed Kaggle dataset and sentiment was analyzed based on four classes such as sad, joy, fear and anger. The adopted deep learning approach considered the complexities in analyzing the textual data and helps to attain better accuracy. However, the classification of sentiment is not detected for the specified topic labels.

Singh et al. [17] employed the Bidirectional Encoder Representations from Transformers (BERT) model to perform sentiment analysis on Twitter data related to the impact of COVID-19, classified by the location of the tweet authors in India or other parts of the world. The tweets were collected during a time when positive tweets about the virus were scarce. The TextBlob method was used to predict tweet polarities and subjectivities, and the resulting emotion analysis using BERT can aid government decision-making in future pandemics. However, the BERT model was limited to classifying COVID-19 impacts in the specified regions.

Meanwhile, Gandhi et al. [18] developed a sentiment analysis model for movie reviews on Twitter using Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM). They used the IMDB dataset as input and evaluated the effectiveness of the CNN-LSTM model using both stop words and word2vec methods. The stop word method was used to remove common words, while the word2vec method was used to convert the tweets into vector formats with various dimensions. The LSTM outperformed the CNN in terms of validation accuracy, and the sigmoid activation function was used to transfer input and output values for movie tweets. However, the model was not suitable for analyzing complex features to improve results from text data.

Sergiu Cosmin Nistor et al [19] have developed a system to analyze the Twitter sentiments using Recurrent Neural Network (RNN). The developed solution converts the text into numerical representations. After this, the stage of pre-processing takes place at the side of input and the information was transferred to the recurrent layers present in the network. Moreover, the attention based mechanism was applied and finally the feed forward layer was used to classify the sentiment scores. The RNN performs sentiment analysis using memory tape which stores the data and interact with the surroundings. However, the model obtains biasing possibility in the dataset which may affects the overall performance of the model.

Maryum Bibi et al [20] have introduced an ensemble unsupervised framework incorporating the concept based hierarchical clustering to classify the sentiment obtained from

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Twitter data. The Twitter data was provided as input for the concept level module then pre-processing takes place. The model analyzes the sentiment of the labeled tweets and examine the tweets to clustering approaches for which sentiment labels could not be discovered. The features of given tweets were represented using TF-IDF and Boolean methods. The limited knowledge base of concept based approach leads to improper classification.

A. Naresh and P. Venkata Krishna [21] have introduced Sequential Minimal Based Optimization Decision Tree (SMODT) approach to classify the Twitter data. The process undergone three stages, the data was gathered and pre-processed in first stage and in second stage optimization takes place by extracting the required features. In final stage, the trained data was classified and polarity was evaluated. Since SMODT is a hybrid approach, it effectively optimized the total time for training and provides better performance during classification. However, the SMODT approach was only suitable for datasets with more number of data.

Atheer S. Alhassun and Murad A. Rassam [22] have introduced a combined framework using deep learning approach. The combined framework is a combination of text-based data with convolutional neural network and metadata with simple neural network. The output obtained from both the models were integrated to detect the spam in Twitter platform. The combined framework utilized Convolutional Neural Network (CNN) with five layers to detect the unsolidity in the Twitter data. However, the detection accuracy for unstandardized dataset was not enough in minimum training period.

3. SENTIMENT CLASSIFICATION USING BERT AND RANGER ADABELIEF OPTIMIZER

3.1 Problem Statement

Automated classification of opinions, views, and emotions in user feedback data, known as sentiment analysis, is a Natural Language Processing (NLP) technique. While many machine learning algorithms have been used for this purpose, previous approaches have not met industry requirements for accurate categorization and classification. To improve performance, a new approach was introduced using a novel TF-IDF algorithm and Long Short-Term Memory (LSTM) framework to analyze airline customer feedback data from Twitter. This approach improved classification performance, but the LSTM framework only performed well with small and medium datasets and had reduced performance with large datasets. To overcome this limitation, a new BERT framework based on Range Adaboost Optimizer was proposed and applied to Twitter data classification, resulting in significant improvements.

The process of analyzing sentiment on the social platforms such as Twitter is regarded as an important method for gathering data about the public's emotions in day-to-day applications. The suggested research assists in extracting the mood of tweets produced by users of Twitter in different conditions, and the proposed approach can distinguish emotion from text. The Deep Learning approach is used to analyze sentiment, which aids in the model's capacity to learn automatically.The block diagram for analysis of sentiment from tweets based on deep learning approach is specified in Figure 1 as follows,

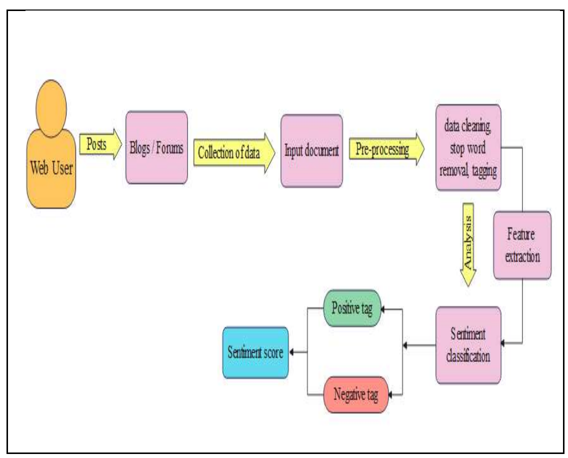

Figure 1. Diagrammatical representation of process involved in classification of sentiment

An improved deep learning technique for assessing tweet's sentiment acquired from Twitter API is utilized to separate emotion of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

human, with sentence sentiments classified as positive or negative. Because of the informal language used in tweets, doing sentiment analysis is a critical task. This leads to the provision of inventive and diverse perspectives. The proposed sentiment classification model undergoes different stages such as (i) Collection of data (ii) data pre-processing (iii) extraction of features and (iv) classification of sentiments. The below subsections describes the fore mentioned process involves in classifying the sentiment of the tweets.

3.2 Collection of Data

The data collection is the process of accumulating and computing the information from various online real world platforms. The data is collected from various sources and the process of analyzation and classification takes place on the basis of user interests. In this research the data is gathered from Twitter API dataset, which is a well-known publicly accessible database.

3.2.1 Twitter Dataset: Twitter dataset contains tweets about individuals, spaces, direct messages, lists, trends, media, and locations. Twitter is one of the publicly accessible datasets that contributes to real-world applications. In given dataset contain approximately 60000 records. Furthermore, the Twitter dataset includes tweets with extended phrases, stop words, misspelled words, and so on. To eliminate them, a pre-processing approach must be used to facilitate the categorization process.

3.3 Data Pre-processing

The data pre-processing is a significant part in cleaning text data which helps to examining the tweet' s sentiment. In general, data regarding the real time applications consist of colloquial language, prepositions, and punctuation marks. The presence of these words makes tweet unsuitable for direct analysis so, the normalization must be performed to improve the format of the data. Since, the real time data contains fragments and lack of precise data formats, it leads to inaccuracies during the process of classification. So, the pre-processing acts as a significant approach to resolve the issues arisen during the process of sentiment classification. Tweets can be pre-processed based on two techniques such as ordinary text pre-processing and Twitter specified pre-processing techniques. The ordinary text pre-processing method takes place using Natural Language Tool Kit (NLTK) library and Tweet Arc Natural Language Processing (NLP) method is used to process the tweets. The following pre-processing methods are utilized in this research to enhance the quality of raw data, (i)tokenization, (ii) Parts of tagging and (iii) Stop word removal.

3.3.1Tokenization

The boundary detection is an essential step in processing the texts and the boundaries are identified using spaces and punctuations. The technique involved in splitting of sentence and detecting the individual entities of the sentence is known as tokenization. The process of tokenization has its main significance in removal of URL and hashtags in tweets. In a normal technique of classification, the tokenization takes place in the URLs and the separation of '#' takes place. Moreover, the hashtag detection helps in processing various applications like opinion mining, detection of event and so on.

3.3.2 Part of Speech (POS) Tagging:

The POS tagging is a significant part in processing the tweets and the class of the word is detected on the basis of word's position. The Twitter specified preprocessing helps to detect the Emoticons, Re Tweet (RT) tags which is generally present in normal tweets. But the normal preprocessing methods will not consider the parameters while processing.

3.3.3 Removal of Stop Words:

The stop word removal eliminates the frequently utilized words which is considered as insignificant during the tweet analysis. These stop words must be neglected from the tweets before the process of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

classification and the presence of stop words will affect the accuracy of the model. This process minimizes the count of total words before entering the classification model.

3.4 TF-IDF Feature Extraction

The pre-processed output is provided as input for the process of feature extraction. This research uses Term Frequency – Inverse Document Frequency (TF-IDF) to extract the features, it receives the pre-processed data and ranks the effective oneto extract the key element from tweets. The value of TF is measured using the equation 1 as follows,

$$T = \frac{N_S}{TN} \quad (1)$$

Where the entire number of terms is represented as s and the terms present in text is denoted as TN. The significant terms present in the tweets are measured using IDF and it is evaluated using the equation (2) as follows,

$$IDF = log_e\left(\frac{ND}{TD}\right) \quad (2)$$

The total number of words in the document is represented as TD and the number of documents (tweet) is represented as ND. The weightage of the term is computed using equation (3) as follows:

$$TF - IDF(s, DC) = TF(s, DC) \times IDF\ (s) \quad (3)$$

Where the document is represented as DC and the count of terms are represented as s.

3.4.1 Sentiment Classification Using BERT model and Range AdaBelief Optimizer for Loss Optimization

The extracted features are utilized as input for the process of classification. This research utilized Bi-directional Encoding Representation for Transformer (BERT) model for classifying sentiment present in tweets andloss is optimized using Ranger AdaBelief optimizer.

3.5 BERT Model

The BERT model effectively classifies the sentiment present in the tweets and it analyzes the meaning of the word in the tweet which depends on the remaining words present in the tweets. The BERT provides all the input and handle the dependencies of the related words present in the tweet. The BERT model is categorized into two types such as BERT-large model and BERT-base model. The base model of BERT consists of 12 transformer encoders and the large of BERT consist of 24 transformer encoders.The classification of text BERT utilize the hidden state of the first token and an ordinary softmax classifier is included in top of BERT to determine the label, which is represented in equation (4) as follows:

p(c|h)=softmax (Wh) (4)

Where h is the hidden state and the predicted label is denoted as W.

To make the BERT classifier effective for classification the following factors are considered,

1. The tweets with maximum length of 512 words can be processed with BERT classification

2. The second factor relies on selection of layers, BERT classifier have embedding layer, an encoding layer and pooling layer.

3. The optimization loss occurs with the BERT classification model, so a better optimizer with an appropriate learning rate is required to optimize the loss functions.

The optimization of loss takes place using Ranger AdaBelief optimizer to get the desired classification accuracy for the model. The collected data is categorized into training and testing sets using train-test split and the training set is transformed into corresponding tensors of the model.The size is defined to generate the tensors and fine-tune the BERT model using Ranger AdaBelief optimizer. Ranger AdaBelief optimizer helps to optimize the loss and helps to compute the classification accuracy. Before optimizing the loss function, the output from the BERT classifier is encoded to convert the text into digital vectors.

Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.5.1 Range AdaBelief Optimizer

Range AdaBelief optimizer is the combination of AdaBelief optimizer, gradient centralization and look ahead optimizer. These three optimizers are combined by tuning their hyper parameters and helps in the process of classification.

The AdaBelief optimizer is modified from Adam optimizer without including additional parameters. The AdaBelief optimizer have faster convergence, generalization and stability during the time of training. The perception of AdaBelief is based on step size adaption based on the belief on direction of the gradient. The Exponential Moving Average (EMA) of the noisy gradient is predicted for the next step of the gradient. If the notified gradient deviates from the predicted range, distrust is made in the present observation and proceeded with small step. But, when the observed gradient is nearer to the predicted value, the trust is made to take larger steps. AdaBelief performs with other methods with high convergence rate and provides better classification accuracy. The modification was pretended in Adam optimizer to develop AdaBelief optimizer. The observed gradient at step t is denoted as g_t and EMA is represented as $m_{t.}$ The EMA value of g_t^2 and $(g_t - m_t)^2$ is represented as v_t and g_t respectively. When the variation occurs among g_t and $m_{t,}$ a weak belief was made and small step was taken. When there is no variation among g_t and m_t, a strong belief is made. The overall equation obtain from AdaBelief optimizer is represented in equation (4) as follows:

$$\Delta\theta_t^{AdaBelief} = {}^{-\alpha m_t}\Big/_{\sqrt{s_t}} \quad (4)$$

Where the observed gradient at step t is denoted as g_t and EMA is represented as $m_{t.}$

AdaBelief optimizer provides better intuition for high dimensional classes. The equation $f(x,y) = |x| + |y|$ is considered as loss function and the gradient present in each axis lies among the range of $\{1, -1\}$. When the initializing point is nearer to x −axis, the oscillation occurs in the direction of y axis and the initializing point is increased in x −axis.

When the algorithm performs for long time range, t becomes large and EMA becomes smaller. The formulations for $m_t$andv_t is represented the equations (5) and (6) respectively,

$$m_t = EMA(g_{0,}g_{1,\dots,}g_{t)} \approx E(gt), m_{t,x} \approx E(gt,x) = 1, m_{t,y} \approx E(gt,y) = 0 \quad (5)$$

$$v_t = EMA\left(g_0^2, g_1^2, \dots, g_t^2\right) \approx E(gt), v_{t,x} \approx E\left(g_{t,x}^2\right) = 1, v_{t,y} \approx E\left(g_{t,y}^2\right) = 1 \quad (6)$$

The bias correction is proceeded to reduce the error among EMA and g_t which is represented in equation (7) as follows:

$$s_t = EMA((g_0 - m_0)^2, \dots, (g_t - m_t)^2) \approx E[(g_t - Eg_t)^2] = Varg_{t,}s_{t,x} \approx 0, s_{t,y} \approx 1 \quad (7)$$

AdaBelief optimizer considers both magnitude and sign of g_t which helps to take larger step in x direction and small step in the direction of y axis. This state makes the AdaBelief optimizer ideal and better in classification of sentiments.

3.5.2 Formulation of Gradient Centralization

The gradient obtained from backward propagation is obtained by centralizing the gradient which is represented in equation (8) as follows:

$$\emptyset_{GC}\left(\nabla_{w_i}L\right) = \nabla_{w_i}L - \mu\nabla_{w_i}L \quad (8)$$

Where the operator of gradient centralization is represented as $\emptyset_{GC}$, the weighted vector is represented as w_i and the value of $\mu\nabla_{w_i}L = \frac{1}{M}\sum_{j=1}^{M}\nabla_{w_i}L$.

The gradient centralization is performed to evaluate the mean value of vectors present in the weighted matrix. The formulation of matrix is represented in equation (9) as follows:

$\emptyset_{GC}(\nabla_w L) = \mathrm{P}\,\nabla_w L,\ P = I - ee^T$
(9)

Where P denotes the physical value of the gradient. Practically, the mean value present in weighted vector can be removed using gradient centralized operator.

3.5.3 Look Ahead Optimizer

The quick weights updates produce rapid progress along the low curvature directions when oscillating in the high curvature directions. Through parameter interpolation, the slow weights assist smooth out the oscillations. The Look ahead optimizer utilized fast and slow weights to enhance learning in high curvature directions, decreases variation, and allows Look ahead to converge quickly in practice.

Slow weighted route: The slow weights are obtained from EMA of final weights present in the inner loop which is represented in equation (10) as follows:

$$\emptyset_{t+1} = \emptyset_t + \alpha(\theta_{t,k} - \emptyset_t) \qquad (10)$$

Where $\emptyset_t$ is the weighted parameter at time t, the slow weighted step size is represented as α and the period of synchronization is represented as α.

Fast weighted route: The fast weighted route considers the selection of optimizer present in the inner loop of the weighted function. The fast weighted route is upgraded using the equation (11) as follows,

$\emptyset_{t,i+1} = \emptyset_{t,i} + A\big(L, \emptyset_{t,i-1,} d\big).$
(11)

Where the optimization algorithm is represented as A, the objective function is denoted as L and the d is training samples of mini batch.

To achieve fast convergence and stability during training, this study proposes using the Ranger AdaBelief optimizer. This optimizer combines the AdaBelief optimizer with gradient centralization, which ensures that the gradient vectors have a mean value of zero, and the look ahead optimizer, which speeds up the iteration rate. The combined use of these optimizers enhances the efficiency of the sentiment classification process, which is performed using the BERT classifier. The sentiment score is determined by classifying the data into different categories based on the user's interest. The output from the Ranger AdaBelief optimizer helps to optimize the loss and improve the effectiveness of the BERT classifier. As a result, the BERT classifier can accurately classify tweets as positive or negative based on the user's interest.

4. RESULTS AND ANALYSIS

The overall performance of the proposed BERT with AdaBelief optimizer is evaluated in this section and comparison is performed with existing techniques.

4.1 Experimental Setup

The proposed BERT with AdaBelief optimizer model is implemented on Anaconda Navigator 3.5.2.0 (64-bit) Python 3.7 and system specification with windows 10 (64 bit), intel core i7 processor and 16GB RAM. Tensor flow, keras, Sklearn frame works libraries are used and the framework is plotted using Matplot library. The experimentation is accomplished for 1000 tweets gathered from different tweets and tweet's polarization is evaluated. The BERT classifier is used for classification and AdaBelief optimizer is used for loss optimization. The below Table.1 shows the outcome obtained for various tweets utilizing Ranger AdaBelief method.

Table.1 Results For Sample Tweets On Different Topics Using BERT With Adabelief Optimizer

S NO	Tweets	Polarities
1	RT @joshjeje2: You will be surprised to find out that Jennifer Bamuturaki the CEO of Uganda Airlines only earns 20million and the rest of t…	Positive
2	American Airlines to purchase Boom Supersonic Overture aircraft.	Neutral

Journal of Theoretical and Applied Information Technology
30th June 2023. Vol.101. No 12

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

	https://t.co/gQJPMI1RZ6 Flights to America are go… https://t.co/k4u9jPA9vL	
3	@brianmixologist @JoelSsenyonyi @UG_Airlines @JonahRuhima @eyaru_levi @wadadamichael @TheOtheGuyHere @BandiVan… https://t.co/ipw9hwnEOO	Neutral
4	#Swissport is celebrating 15 years of #cargohandling with Asiana Airlines at @BrusselsAirport! In 2007, we began pr… https://t.co/1RTJLcZw61	Neutral
5	#ICYMI: With more than 64,000 employees, International Airlines Group (IAG) is considered one of the world's larges… https://t.co/J8eCq4a0uM	Positive
6	@AmericanAir has signed an agreement with @boomaero to purchase 20 of the firm's Overture aircraft.… https://t.co/z08z5IlT6h	Negative

4.2 Performance Metrics

The overall efficiency of Range AdaBelief technique is evaluated based on the performance metrics stated as follows:

Accuracy: It is stated as the total count of properly categorized data to total count of data.

$$Accuracy = \frac{TN + TP}{TN + FP + TP + FN}$$

Where, TNdenotes true negative, TPdenotes true positive, FPdenotes false positive and FNdenotes false negative.
Precision: It is the total number of truly positive values to the total number of positives.

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

Recall: It is defined as the ratio among the total count of Positives that is properly categorized as Positive to total samples of positive and negative values.

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

F1-score: It is evaluated by using both recall and precision. when values of precision and recall are 1, the F1 score will be one.

$$F1\ score = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

4.3 Performance Analysis

The performance of the suggested BERT classifier with Ranger AdaBelief optimizer is related with the performance of existing classifiers. The model's overall performance is relatively higher when analysed with the classifiers such as SVM, ANN (MLP) and LSTM classifiers. The obtained outcome shows that the accuracy of the suggested technique is 93% that is greater than the other classification models. Thus, overall performance of the proposed classifier is relatively better than the existing models. The overall performance of the BERT with AdaBelief optimizer is represented as tabular and graphical representation in Table.1and Figure.2 correspondingly.

Table.2 Analysis Of BERT With Ranger Adabelief With Existing Models

Classifiers	Accuracy	Precision	Recall	F1 Measure
SVM	68	63	66	64
ANN(MLP)	66	73	66	69
LSTM	80	81	74	77
BERT with Range AdaBelief	92	90	91	91

The results from table 1 shows that the proposed BERT classifier achieved better classification accuracy of 93%. The better result is due to the presence of Ranger AdaBelief optimizer which effectively optimize the loss functions. AdaBelief optimizer provides better intuition for high dimensional classes and helps to classify the sentiment of the tweets effectively. The graphical representation for performance of proposed classifier with existing model is represented in figure 2 as follows,

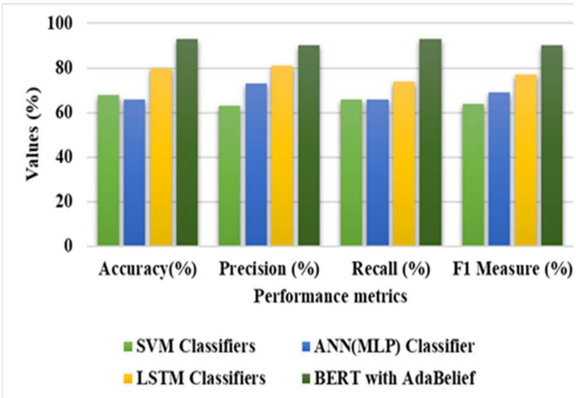

Figure 2. Graphical Representation To Evaluate The Performance Of Classifier.

4.4 Error Rate Analysis

It is defined as the process of detecting the error and analyzing imprecise predictions to study about the low performing area of the model. The error rate due to Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) is represented in table 2, The error rate obtained by the proposed model is comparatively low when compared with the existing classifiers.

MAE: It is stated as a measure of error amongbalanced opinions which explicit the same value and it is evaluated using the formula represented as follows,

$$MAE = \frac{\sum_i^n |y_i - x_i|}{n}$$

Where y_i is value obtained from prediction, x_i is true value, n is total count of values

MSE: It is statedas average differenceamong the definite and expected value and it can be evaluated utilizing formula represented as follows,

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

Where, y_i is value obtained from observation, $\hat{y}_i$ is value obtained from prediction, n is the number of data.

RMSE: It can be stated as the standard deviation which ariseswhile predicting the errors. RMSE can be computed using the formula provided below,

$$RMSD = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}}$$

Where, N is the number of not missed data values, actual value is x_i, the assessed value is $\hat{x}_i$.

Table 2. Comparison of error rate with distinct classifier

Classifiers	MAE	MSE	RMSE
SVM	0.42	0.58	0.76
ANN(MLP)	0.45	0.68	0.82
LSTM	0.21	0.25	0.50
BERT with Range AdaBelief	0.08	0.11	0.39

The results from Table 2 shows that the proposed BERT classifier obtains minimal error rate when compared with other distinct classifiers. Moreover, the process of fine tuning helps BERT classifier to achieve minimal rate of error. The graphical representation for analysis of error is represented in Figure 3 as follows,

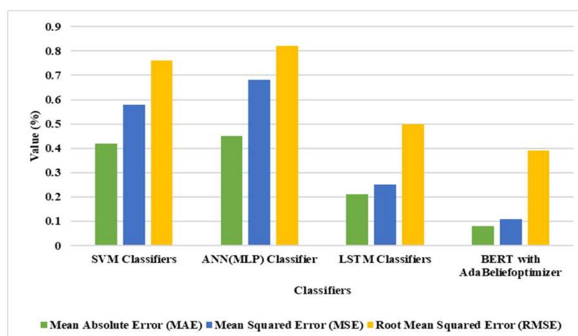

Figure 3. Graphical Representation For Error Analysis

4.5 Comparative analysis

Table 3. Comparative Table for Proposed BERT and LR Stated Algorithms

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

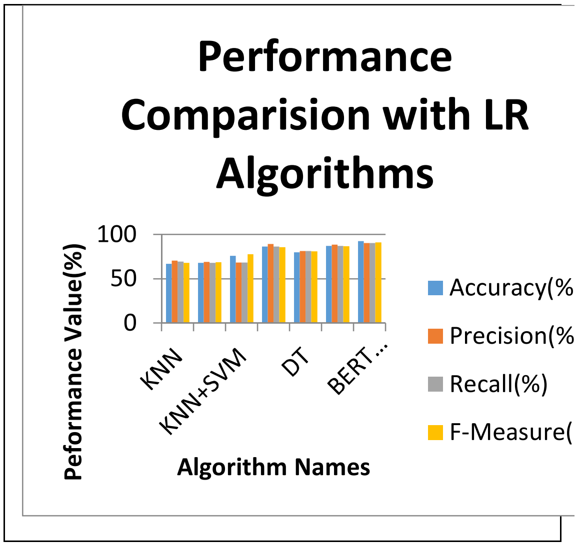

Figure 4. Graphical Representation For LR Comparison

The section provides the comparison of the proposed BERT Classification with Ranger AdaBelief Optimizer with existing methodologies discussed in related works. The comparison is made with literature review stated classifiers such as K-Nearest Neighbor (KNN) classifier, Support Vector Machine (SVM), hybrid of KNN and SVM, Sequential Minimal Optimization (SMO), Decision Tree (DT) and Convolutional Neural Network (CNN) classifier. The table 3 shows the result of the proposed BERT classification with Range AdaBelief optimizer performs better in overall metrics and achieved classification accuracy of 93.3%.

Table 4. Comparative Table For Proposed BERT And LR Stated Algorithms

Classifiers	Accuracy	Precision	Recall	F1-Score
KNN	67.0	70.5	69.3	67.9
SVM	68.0	69.0	68.1	68.7
KNN+ SVM	76.0	68.45	68.14	77.56
SMO	86.23	89.2	86.2	85.8
DT	80.0	81.4	81.4	80.95
CN	87.24	88.63	87.12	86.67
BERT with Range AdaBelief	92.54	90.15	90.24	91.14

The comparison among the classifiers are

Classifiers	Accuracy	Precision	Recall	F1-Score
LSTM	85.54%	80.15%	80.24%	81.14 %
BERT	92.54%	90.15%,	90.24%	91.14 %

evaluated based on accuracy, precision, recall and F-measure. The results show that the proposed BERT classification with Range AdaBelief optimizer obtains better results in overall evaluation. The better results are due to the presence of Range AdaBelief which performs loss optimization occurred in the process of classification using BERT classifier.

A comparative analysis was conducted on the performance of previous research work such as LSTM, and BERT.

The performance of LSTM improved in terms of accuracy (85.54%), precision (80.15%), F1-Score (81.14%), and recall (80.24%). However, it should be noted that the LSTM performed better on small and medium-sized datasets, but its performance degraded when applied to large datasets.

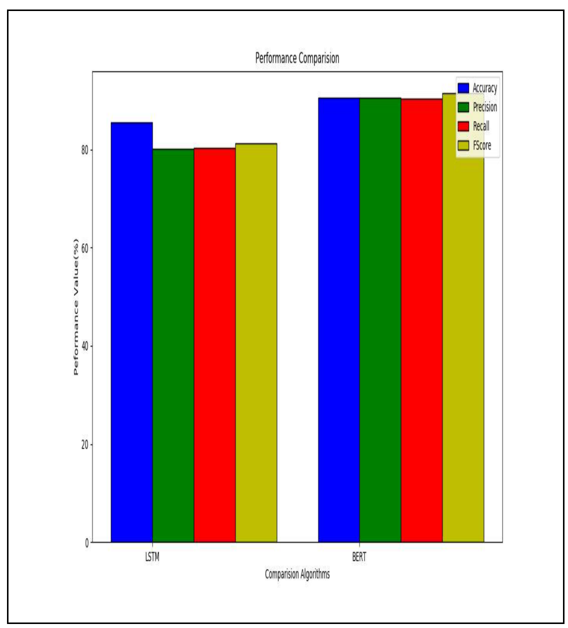

Figure 5. Graphical Representation for BERT Performance

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

The performance of BERT outperformed than LSTM algorithms with an accuracy of 92.54%, precision of 90.15%, F1-Score of 91.14%, and recall of 90.24%. BERT was found to be effective in handling complex features and outperformed the other two algorithms in the task of sentiment analysis. Overall, BERT showed the highest performance in terms of accuracy and other evaluation metrics, followed by LSTM.

5. CONCLUSION:

This research proposes an improved sentiment classification model that uses the BERT classification algorithm with the Ranger AdaBelief optimizer. The BERT model effectively classifies the sentiment in tweets by analyzing the meaning of words and handling the dependencies of related words in the tweet. The Ranger AdaBelief optimizer, which combines AdaBelief, gradient centralization, and look-ahead optimizers, is utilized in conjunction with TF-IDF feature vectors to facilitate the categorization process. The proposed model performs sentence-level analysis to detect the sentiment of tweets and categorizes them as positive or negative based on the polarity of the sentences. The experimental results demonstrate that the proposed BERT classification model with the Ranger AdaBelief optimizer achieves a higher classification accuracy of 92.54% compared to existing classification models.

Future work includes applying the proposed model to different social media platforms, such as Facebook and Instagram, to evaluate its effectiveness.

REFERENCES

[1] Basarslan, Muhammet Sinan, and Fatih Kayaalp. "Sentiment analysis with machine learning methods on social media." (2020).

[2] Mathur, Amrita, Purnima Kubde, and Sonali Vaidya. "Emotional analysis using twitter data during pandemic situation: Covid-19." IEEE Transactions on Electronics Systems, pp. 845-848. IEEE, 2020.

[3] Onan, Aytuğ. "Sentiment analysis on massive open online course evaluations: a text mining and deep learning approach." Computer Applications in Engineering Education 29, no. 3 (2021): 572-589.

[4] Jianqiang, Zhao, Gui Xiaolin, and Zhang Xuejun. "Deep convolution neural networks for twitter sentiment analysis." IEEE access 6 (2018): 23253-23260.

[5] Villavicencio, Charlyn, Julio Jerison Macrohon, X. Alphonse Inbaraj, Jyh-Horng Jeng, and Jer-Guang Hsieh. "Twitter sentiment analysis towards covid-19 vaccines in the Philippines using naïve bayes." Information 12, no. 5 (2021): 204.

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

[7] Samuel, Jim, GG Md Nawaz Ali, Md Mokhlesur Rahman, Ek Esawi, and Yana Samuel. "Covid-19 public sentiment insights and machine learning for tweets classification." Information 11, no. 6 (2020): 314.

[8] Trivedi, Shrawan Kumar, and Amrinder Singh. "Twitter sentiment analysis of app based online food delivery companies." Global Knowledge, Memory and Communication (2021)

[9] AlBadani, Barakat, Ronghua Shi, and Jian Dong. "A novel machine learning approach for sentiment analysis on Twitter incorporating the universal language model fine-tuning and SVM." Applied System Innovation 5, no. 1 (2022): 13.

[10] Valle-Cruz, David, Vanessa Fernandez-Cortez, Asdrúbal López-Chau, and Rodrigo Sandoval-Almazán. "Does twitter affect stock market decisions? financial sentiment analysis during pandemics: A comparative study of the h1n1 and the covid-19 periods." Cognitive computation 14, no. 1 (2022): 372-387.

[11] Naseem, Usman, Imran Razzak, Matloob Khushi, Peter W. Eklund, and Jinman Kim. "COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis." IEEE Transactions on Computational Social Systems 8, no. 4 (2021): 1003-1015.

[12] Srivastava, Ankit, Vijendra Singh, and Gurdeep Singh Drall. "Sentiment analysis of twitter data: A hybrid approach." International Journal of Healthcare Information Systems and Informatics (IJHISI) 14, no. 2 (2019): 1-16.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[13] Rodrigues, Anisha P., and Niranjan N. Chiplunkar. "A new big data approach for topic classification and sentiment analysis of Twitter data." Evolutionary Intelligence (2019): 1-11.

[14] López-Chau, Asdrúbal, David Valle-Cruz, and Rodrigo Sandoval-Almazán. "Sentiment analysis of Twitter data through machine learning techniques." In Software engineering in the era of cloud computing, pp. 185-209. Springer, Cham, 2020.

[15] Zainuddin, Nurulhuda, Ali Selamat, and Roliana Ibrahim. "Hybrid sentiment classification on twitter aspect-based sentiment analysis." Applied Intelligence 48, no. 5 (2018): 1218-1232.

[16] Singh, Chetanpal, Tasadduq Imam, Santoso Wibowo, and Srimannarayana Grandhi. "A deep learning approach for sentiment analysis of COVID-19 reviews." *Applied Sciences* 12, no. 8 (2022): 3709.

[17] Singh, Mrityunjay, Amit Kumar Jakhar, and Shivam Pandey. "Sentiment analysis on the impact of coronavirus in social life using the BERT model." Social Network Analysis and Mining 11, no. 1 (2021): 33.

[18] Gandhi, Usha Devi, Priyan Malarvizhi Kumar, Gokulnath Chandra Babu, and Gayathri Karthick. "Sentiment analysis on twitter data by using convolutional neural network (CNN) and long short-term memory (LSTM)." Wireless Personal Communications (2021): 1-10.

[19] Nistor, Sergiu Cosmin, Mircea Moca, Darie Moldovan, Delia Beatrice Oprean, and Răzvan Liviu Nistor. "Building a twitter sentiment analysis system with recurrent neural networks." Sensors 21, no. 7 (2021): 2266.

[20] Bibi, Maryum, Wajid Arshad Abbasi, Wajid Aziz, Sundus Khalil, Mueen Uddin, Celestine Iwendi, and Thippa Reddy Gadekallu. "A novel unsupervised ensemble framework using concept-based linguistic methods and machine learning for twitter sentiment analysis." Pattern Recognition Letters 158 (2022): 80-86.

[21] Naresh, A., and P. Venkata Krishna. "An efficient approach for sentiment analysis using machine learning algorithm." Evolutionary intelligence 14 (2021): 725-731.

[22] Alhassun, Atheer S., and Murad A. Rassam. "A combined text-based and metadata-based deep-learning framework for the detection of spam accounts on the social media platform twitter." Processes 10, no. 3 (2022): 439.

[23] Sayeedunnisa, S., Nagaratna Parameshwar Hegde, and Khaleel Ur Rahman Khan. "Sarcasm Detection: A Contemporary Research Affirmation of Recent Literature." *ECS Transactions* 107.1 (2022): 2419. "

[24] Sayeedunnisa, S., Nagaratna Parameshwar Hegde, and Khaleel Ur Rahman Khan. "Sarcasm Detection: A Contemporary Research Affirmation of Recent Literature." *ECS Transactions* 107.1 (2022): 2419. "

[25] Sayeedunnisa, S. F., Hegde, N. P., & Khan, K. U. R. (2021). Feature Selection by Associativity for Sentiment Analysis. In Smart Computing Techniques and Applications (pp. 423-430). Springer, Singapore.