© August 2022 | IJIRT | Volume 9 Issue 3 | ISSN: 2349-6002

Censored Regressive ZIJDENBOS Indexed Convolutional Deep Belief Neural Network Learning for Sentiment Classification

Dr. Shubha. S

Assistant Professor, Department of Computer Science, Government First Grade College, Malleshwaram, Bangalore

***Abstract*—Opinion Mining or Sentiment Analysis is a process of classifying the user's opinion namely positive, negative or neutral about a topic or a product. Traditional sentiment analysis focuses on extracting opinion polarization at a coarse level. Recently, the conventional machine learning technique is analyzed by regarding the customer opinion about the products automatically from online reviews. But it is ineffective and still not possible for evaluating sentiment analysis through enormous data and online processing requirements. In order to improve the Sentiment Classification, a novel Censored Regressive Zijdenbos Indexed Convolutional Deep Belief Network Learning Classification (CRZICDBNLC) technique is introduced. The Convolutional Deep Belief Neural Network Learning Classification technique includes different layers for analyzing the given reviews with different processes namely preprocessing, feature extraction and classification. First the reviews are collected from the dataset and are given to the input layer of Convolutional Deep Belief Network Learning Classifier. Later the reviews are given to the first hidden layer of deep learning where the preprocessing is carried out by removing the stopwords and stem words. Then the Feature extraction process is performed in the second hidden layer using Censored Regression. Finally, the classification is done at the third hidden layer for finding the user's opinion using the Zijdenbos similarity Index. In this way, accurate Sentiment Classification is performed with higher accuracy. Experimental assessment is carried out with various parameters such as accuracy, precision, recall, F-measure and computational time with respect to a number of reviews. The quantitatively discussed result verifies that the proposed CRZICDBNLC technique achieves higher accuracy and minimal computation time as compared to the conventional methods.**

***Index Terms*—Censored Regression-based feature extraction, Convolutional Deep Belief Network, Preprocessing, Opinion mining or Sentiment Analysis, Zijdenbos similarity index-based classification.**

I. INTRODUCTION

With the excessive development of online reviews, sentiment classification has become a fundamental issue in recent years. The main argument of sentiment classification is to identify the equivalent sentiment polarity (e.g., positive, neutral, or negative) towards certain things or products. Various techniques have been introduced to build sentiment classifiers, but the classification performances of some existing methods are not good enough.

A single-layered Convolutional Neural Network (CNN) model with a fastText embedding technique (fastText+CNN) model was introduced in [1] for sentiment classification. The accuracy of binary sentiment classification but the time consumption of the sentiment classification was not minimized. Sentiment classification based on the attention mechanism and the bidirectional long short-term memory network (SC-ABiLSTM) was developed in [2]. However, the designed mechanism failed to improve the accuracy of sentiment classification for large-scale text to predict the sentiments.

A Global and Local Dependency Guided Graph Convolutional Networks (GL-GCN) was developed in [3] for sentiment classification. However, the designed proposed method achieves superior performance, but the performance of time complexity was not minimized. An SSentiA (Self-supervised Sentiment Analyzer) that integrates a machine learning classifier was developed in [4] for sentiment classification from unlabeled data, but the designed classifier considerably increases the performance of sentiment classification. An enhanced hybrid feature selection method was introduced in [5] to increase the sentiment classification. However, the designed feature selection technique was not tested on a larger dataset to measure its efficiency and effectiveness.

A multi-task learning model based on a multi-scale Convolutional Neural Network (CNN) was developed in [6] for sentiment classification. However, the design failed to enhance the performance of sentiment classification for multi-task learning. An Intelligent Hybrid Feature Selection for Sentiment Analysis (IHFSSA) based on ensemble learning techniques was developed in [7]. But it failed to incorporate deep learning for enhancing the Sentiment classification.

A Novel Hybrid Deep Learning model was designed in [8] based on feature extraction for sentiment classification. However, the designed hybrid model has not improved the performance of sentiment classification. An Aspect-Based Opinion Mining (ABOM) based on the deep learning model was introduced in [9] to increase the accuracy of the recommendation process**.** But it failed to minimize the time consumption for large-scale data analysis. A new Deep Learning model for fine-grained aspect-based Opinion Mining technique was introduced in [10]. But the preprocessing and feature extraction was not performed.

A. Proposal Contribution

The issues reviewed by the above-said literature are overcome by introducing a novel CRZICDBNLC technique with the following contribution-

- To improve the accuracy of sentiment classification, the CRZICDBNLC technique is designed based on the three steps namely preprocessing, feature extraction and classification.
- To minimize the computational time of sentiment classification, the CRZICDBNLC technique performs the preprocessing step as well as feature extraction. In preprocessing step, textblop tokenizer, Spacy stopword removal technique, Paice/Husk stemmer is employed to obtain the words. This process extracts the main words for classification.
- Censored Regression is employed to perform the feature extraction by analyzing the words in the hidden layer of deep learning. This process extracts the most important words for classification.
- Finally, the Zijdenbos Similarity Index is applied for customers review classification based on the similarity measure. The Zijdenbos similarity coefficient increases the classification accuracy.
- At last, extensive taxing is conducted to estimate the performance of CRZICDBNLC technique and other related works. The experimental result shows that the CRZICDBNLC technique is analyzed with the various performance metrics with a number of reviews.

B. Organization of the paper

The paper is arranged into different sections. Section 2 discusses the literature review of Sentiment Classification. Section 3 describes proposal methodology. Section 4 briefly describes the proposed CRZICDBNLC technique with a neat diagram. Section 5 provides information on the experimental settings with the dataset. In section 6, the test outcomes and comparative analysis are presented using various parameters. Finally, section 7 concludes the paper.

II. LITERATURE REVIEW

Entity-Sensitive Attention and Fusion Network (ESAFN) was introduced in [11] for Multimodal Sentiment Analysis. However, the designed end-to-end entity-level multimodal sentiment analysis was a promising direction. A hybrid rule-based method was introduced in [12] to enhance the performance of semantic representation. However, the designed method failed to employ semantic rules for validating the efficacy of deep semantic information.

An interactive attributes attention model was developed in [13] to improve the sentiment classification performance for customer reviews. However, it failed to consider the performance of the proposed model with implicit features. A novel interactive model was designed in [14] for both local and global interactions between users and products. However, the designed model failed to show the efficiency of the proposed model with several strong baselines.

Deep Learning Modified Neural Network (DLMNN) technique was developed in [15] for sentiment analysis of online product reviews. However, the designed system failed to improve the performance using a hybridization algorithm.

Attention-based Word Embeddings with Artificial Bee Colony Algorithm were designed in [16] for sentiment classification. But the performance of accuracy using sentiment classification was not improved. An enhanced adversarial learning technique was introduced in [17] for sentiment classification. However, the designed technique failed to consider some other attention mechanisms in a neural network for the task of sentiment classification.

A contextual information sentiment model was designed in [18] for recommender systems. However, the designed model failed to improve the performance of recommendation content and quality. A Graph Domain Adversarial Transfer Network (GDATN) was introduced in [19] to detect the sentiment label of unlabeled target domain data. However, the design of the deep model was not studied for an in-depth analysis. A modified Genetic Algorithm with the Wrapper approaches (GAWA) was developed in [20] for significant features to improve the sentiment classification. However, the designed algorithm failed to implement with multiple datasets to select the best features for enhancing the classification performance.

III. PROPOSAL METHODOLOGY

Sentiment analysis or opinion mining is the process of extracting useful information from various sources. Several different techniques have been widely employed with different classifiers to get the analysis of sentiment. However, the classifiers are still suffering from low accuracies, mainly due to the deficiency for Sentiment analysis. Therefore, a novel CRZICDBNLC technique is introduced with a neat diagram for Sentiment classification with higher accuracy. The CRZICDBNLC technique uses the Convolutional Deep Belief Network Learning for accurate sentiment Classification.

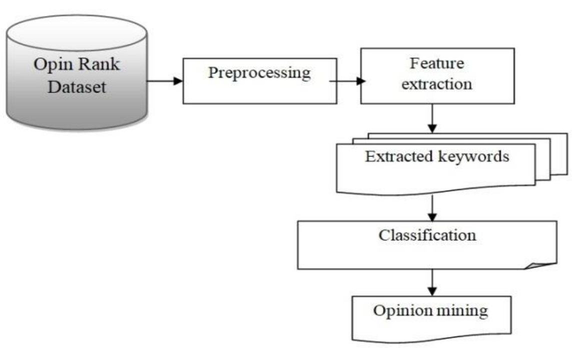

Figure 1. Architecture diagram of CRZICDBNLC technique

Figure 1 illustrates the architecture diagram of the CRZICDBNLC technique. The proposed CRZICDBNLC technique includes three different processes namely preprocessing, feature extraction and classification includes positive, negative, or neutral opinion. This in turn improves the classification accuracy. The different processes of the proposed CRZICDBNLC technique are discussed in the following subsections.

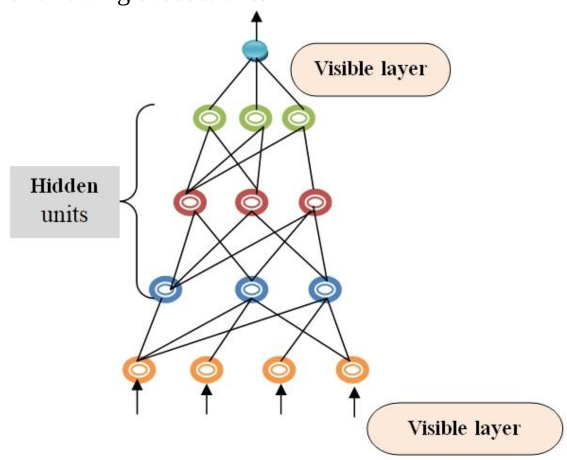

Figure 2. Schematic construction of Convolutional Deep Belief Network

Figure 2 illustrates the schematic construction of the Convolutional Deep Belief Neural Network. The Convolutional Deep Belief Network is a type of deep artificial neural network that consists of many layers that are stacked together. The network is composed of multiple layers of latent variables ("hidden units"), with connections between the layers but not between units within each layer. The network also consists of visible layers (i.e. input and output layer) that are connected in a feed-forward manner with adjustable weights. The input layer receives a number of reviews$r_1, r_2, r_3, \ldots r_n$. The input layer x(t) of the network is expressed as given below,

$$x(t) = b + [\sum_{i=1}^{n} R_i(t) * \ \omega_0]$$

Where, $R_i(t)$ indicates a review, ‘ b’ denotes a bias that stored the value is ‘1’, ‘w_0’ represents the weight at the input layer. Then the input is transferred into the first hidden layer where the relevant preprocessing is performed.

A. Preprocessing

Pre-processing is the process of cleaning and organizing the text for an accurate classification process. The online reviews usually contain many words in the text and this irrelevant word increases the dimensionality of the problem and hence makes the classification more difficult. Therefore, it also increases the computational complexity of the classification process. The proposed CRZICDBNLC technique first performs the Pre-processing that includes different processes namely Tokenization, stop words removal and stemming, etc.

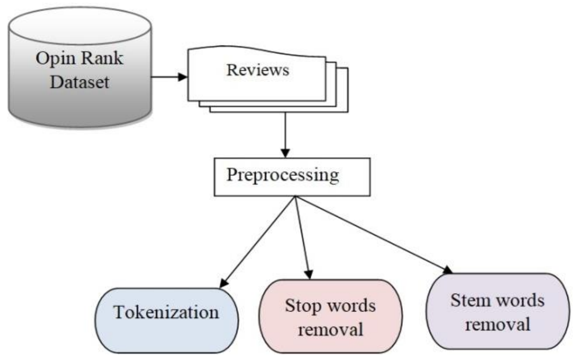

Figure 3. Block diagram of Pre-processing

Figure 3 illustrates the block diagram of review pre-processing. Tokenization breaks the raw text into many words, called tokens. These generated tokens are used for understanding the context. The proposed technique uses the textblop tokenizer for splitting the text into words using punctuation and spaces inside the square bracket.

$$R_1 \rightarrow ['w_1',\ 'w_2',\ 'w_3', \ldots\ 'w_k'] \quad --- (1)$$

Where R indicates a review is partitioned into a number of tokens called words $w_1, w_2, w_3, \ldots\ldots, w_k$ using textblop tokenizer.

B. Spacy stopword removal technique

Stop word is a word used for connecting high discriminating power words while forming a sentence. The stop words have no meaning. The proposed CRZICDBNLC technique uses the Spacy stopword removal to improve the performance of the review classification. These kinds of words are filtered before processing. The most commonly used stop words in the sentiment classification, "are", "the", "a", "an", "in, "and", "our", "this", and so on.

C. The Paice/Husk stemmer

Word stemming is a process used to extract the root words by removing the suffixes from a word.

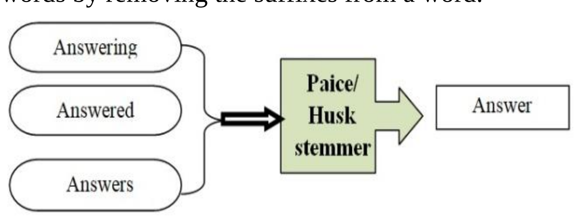

Figure 4. Example of Paice/Husk stemmer

Figure 4 demonstrates the example of Paice/Husk stemmer. By applying the Paice/Husk stemmer, the word ends with 'ing', 'ed', 's', which are called suffixes are removed from the review and we obtain the root word answer.

D.Censored regressive Feature extraction

Feature extraction is the process of selecting the relevant features for classification task in sentiment analysis, to select informative and relevant features for producing accurate classifying results and minimize time complexity. The proposed technique uses the censored regression models which are a class of models in which the relevant features (i.e. words) are selected based on a certain threshold in the second hidden layer. The Regression function measures the distance between the features.

$$D = \|w_i - w_j\| \quad (2)$$

In equation (2), D denotes distance between the features w_i, w_j. The Regression function sets the threshold function to find the minimum distance between the features. Therefore, the relevant features are formulated as given below,

$$Z = \begin{cases} D > t\ ; irrelevant\ features \\ D < t\ ; relevant\ features \end{cases} \quad (3)$$

Where Z indicates an output function, D the distance, and t the threshold. The proposed technique finds the relevant features (i.e. words) and other features are removed. The extracted relevant features are given to the third hidden layer.

IV. ZIJDENBOS SIMILARITY INDEXIVE SENTIMENT CLASSIFICATION

The sentiment classification is performed in the third hidden layer with the extracted words. The Zijdenbos similarity index is applied to measure the similarity between the sentiment words. The similarity is estimated as given below,

$$R\left(w_i, w_j\right) = 2 * \frac{|w_i \cap w_j|}{|w_i| + |w_j|} \quad ---- \quad (4)$$

Where, $R(w_i,w_j)$ denotes a Zijdenbos similarity coefficient, ∩ denotes a mutual dependence between the sentiment words w_i, w_j. Zijdenbos similarity coefficient provides the output value between 0 and 1.

$$Q = \begin{cases} R\left(w_i, w_j\right) > 0.5 \ ; & Positive\ opinion \\ R\left(w_i, w_j\right) = 0.5 \ ; & Neutral\ opinion \\ R\left(w_i, w_j\right) < 0.5 \ ; & Negative\ opinion \end{cases} \quad (5)$$

Where Q indicates a Zijdenbos similarity index. Here, if the similarity $R(w_i,w_j)$ is 0.5 then the review is classified as neutral. The similarity value $R(w_i,w_j)$ is greater than the 0.5 is called a positive opinion. The similarity value is lesser than 0.5 is called a negative opinion. Based on a similarity measure, the sentiments are accurately classified at the output layer. The CRZICDBNLC algorithmic process of the is described as given below,

Algorithm 1: Censored Regressive Zijdenbos Indexed Convolutional Deep Belief Network Learning Classification

Input: Dataset, number of reviews $R_1,R_2,R_3,....R_k$.

Output: Increase the accuracy

Begin
1. Number of reviews $R_1,R_2,R_3,....R_k$ taken as input at the input layer
2. For each review R_1 // hidden unit 1
   a. Apply textblop tokenizer to partition the review into words
   b. Perform Spacy stopword removal
   c. Apply Paice/Husk stemmer to remove the stem words
   End for
3. Apply Censored regression // hidden unit 2
4. if $(D < t)$ then
   Extract the relevant features
   else
   Extract the irreverent features
   End if
5. For each feature with extracted features // hidden unit 2
   Compute similarity $R(w_i,w_j)$
   If $(R(w_i{\cdot}w_j)>0.5)$ then
   the review is classified as a positive opinion
   else if $(R(w_i{\cdot}w_j)=0.5)$ then
   the review is classified as the neutral opinion
   else if $(R(w_i{\cdot}w_j)<0.5)$ then
   the review is classified as a negative opinion
   end if
   Obtain classification results at the output layer
   End for

End

Algorithm 1 explains the process of opinion classification to increase accuracy with minimal time consumption. Initially, the number of reviews is sent to the input layer. After that, the preprocessing is carried out in the first hidden layer for performing tokenization, stopword removal and stemming to reduce the time consumption. Next, feature extraction is performed in the second hidden layer using Censored Regression. In the third hidden layer, classification is performed using the Zijdenbos similarity Index for finding the positive, negative and neutral opinions. If the similarity is greater than 0.5 then the Reviews are classified as positive opinion. If the similarity is lesser than 0.5 then the Reviews are classified as negative opinions. And if the similarity is equal to 0.5 the Reviews are classified as neutral opinion. Finally, the classification results are obtained at the output layer with higher accuracy.

V. EXPERIMENTAL SETUP

In this section, experimental evaluation of the proposed CRZICDBNLC technique and existing fastText+CNN model [1] and SC-ABiLSTM [2] are implemented in Java using OpinRank Review Dataset Data Set taken from UCI Machine Learning Repositoryhttps://archive.ics.uci.edu/ml/datasets/opinrank+review+dataset . The data set comprises user reviews of cars and hotels collected from TripAdvisor (~259,000 reviews) and Edmunds (~42,230 reviews). Full reviews of 140-250 cars are collected for the years 2007, 2008, and 2009. The Total numbers of reviews are 42,230. Full reviews of hotels are collected from 10 different cities (Dubai, Beijing, London, New York City, New Delhi, San Francisco, Shanghai, Montreal, Las Vegas, Chicago). The total numbers of reviews are 259,000. From the reviews, 10,000-1,00,000 reviews are considered for conducting the experiment.

VI. PERFORMANCE ANALYSIS

The experimental results of the proposed CRZICDBNLC technique and existing fastText+CNN model [1] and SC-ABiLSTM [2] are discussed based on certain parameters such as accuracy, precision, recall, F-measure, and prediction time with respect to a number of reviews. The effectiveness and efficiency of the proposed and existing methods are discussed using tables and graphical representation.

A. Performance of Accuracy

Accuracy is measured as the number of reviews that are correctly classified from the total number of reviews. The overall accuracy is measured as given below,

$$Acc = \left[\frac{t_{pos}+t_{neg}}{t_{pos}+t_{neg}+f_{pos}+f_{neg}}\right] * 100 \quad \text{-------} (6)$$

Where Acc denotes accuracy, t_{pos} represents the true positive i.e. number of reviews correctly classified, t_{neg} indicates true negative, f_{pos} symbolizes the false positive, f_{neg} denotes a false negative. The accuracy is measured in percentage (%).

Table 1. Accuracy

Number of reviews	Accuracy (%)		
	CRZICDBNLC	fastText + CNN model	SC-ABiLSTM
10000	95	89	85
20000	96.75	92	86.5
30000	97.33	92	88.33
40000	96.5	93.25	90
50000	95.8	92.2	90.4
60000	96	92.5	91.16
70000	96.14	92.85	91
80000	96.25	93	92.12
90000	96.11	92.22	91.11
100000	96.1	91.5	90.5

Table 1 describes the experimental results of accuracy along with the number of reviews taken in the ranges from 10,000 to 1,00,000 from the dataset. The obtained results of prediction accuracy using CRZICDBNLC are compared to the two existing classification techniques namely the fastText+CNN model [1] and SC-ABiLSTM [2]. According to the observed results, the presented CRZICDBNLC efficiently achieves higher prediction accuracy than the other classification schemes. Let us consider 10,000 reviews for conducting the experiments in the first iteration. By applying the CRZICDBNLC, the observed accuracy is 95% whereas the accuracy percentage of the existing [1] and [2] are 89% and 85% respectively. The performance of the proposed CRZICDBNLC is evaluated with the other existing methods. Finally, the average of comparison results designates that the accuracy of the proposed CRZICDBNLC technique is increased by 5% when compared to [1] and 7% when compared to [2].

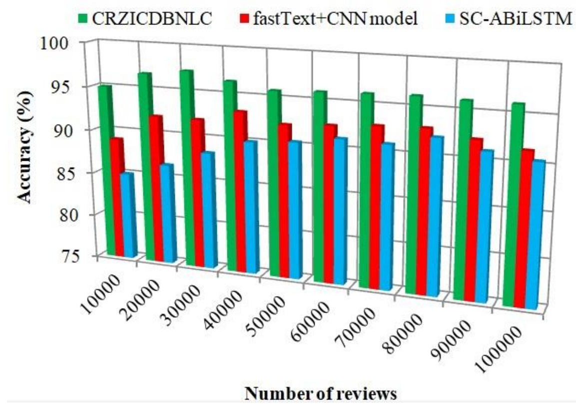

Figure 5. Performance results of accuracy

Figure 5 demonstrates the performance analysis of accuracy along with the number of reviews in the ranges from 10,000 to 1,00,000. The reviews are taken in x-axis and the accuracy is observed at y-axis. From the graph we see that the accuracy of the proposed CRZICDBNLC has the ability for increasing prediction accuracy. The reason for this improvement is to apply the Convolutional Deep Belief Network Learning Classification. The classification is completed at the third hidden layer for finding the user’s opinion using the Zijdenbos similarity Index with the extracted keywords. Based on the similarity values, the reviews were correctly classified as positive, negative and neutral with higher accuracy.

B. Performance of Precision

Precision is estimated as the ratio of true positives and the sum of true positive and false positives. The precision is calculated as given below,

$$Prc = \left[\frac{t_{pos}}{t_{pos}+f_{pos}}\right] * 100 \qquad (7)$$

Where Prcdenotes precision, t_{pos} represents the true positive i.e. number of reviews correctly classified, f_{pos} indicates the false positive. The precision is measured in percentage (%).

Table 2 precision

Number of reviews	Precision (%)		
	CRZICDBNLC	fastText + CNN model	SC-ABiLSTM
10000	96.77	93.10	90.36
20000	97.91	94.56	90.28
30000	98.26	94.94	92.56
40000	97.37	95.94	93.66
50000	96.86	94.83	93.88
60000	97.21	95.17	94.56
70000	97.46	95.41	94.84
80000	97.65	95.71	95.25
90000	97.67	95.43	94.53
100000	97.70	95.10	94.50

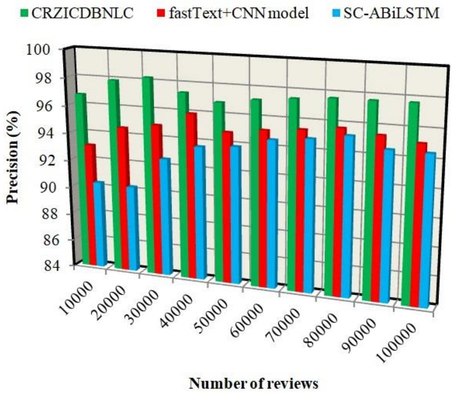

Figure 6. Performance results of precision

Table 2 and figure 6 represent the comparative analysis of precision using three methods namely CRZICDBNLC and two existing classification techniques namely fastText+CNN model [1] and SC-ABiLSTM [2]. The observed results demonstrate that the CRZICDBNLC technique increases the performance of precision than the existing methods. Figure 6 summarizes the overall precision measures of the three algorithms. Under the presence of 10000 reviews for conducting the experiments, the precision of the CRZICDBNLC technique was found to be 95.60% of precision and the results of the existing [1], [2] are 93.10% and 90.36%. From the observed results, we see that the CRZICDBNLC technique offered better precision results. The comparison of ten results indicates that the precision of CRZICDBNLC is significantly improved by 3% and 4% when compared to [1] and [2] existing methods. The reason for this considerable improvement is due to the performance of the deep classification based on the Zijdenbos similarity Index. The deep learning technique minimizes the error and identifies the different review classification results. Therefore, the CRZICDBNLC technique increases the true positives and minimizes the false positives.

C. Performance of Recall

The recall is defined as the ratio of reviews that are correctly classified to the total number of reviews. The recall is measured using the given formula,

$$Rec = \left[\frac{t_{pos}}{t_{pos}+f_{neg}}\right] * 100 \qquad (8)$$

Where, Rec denotes recall, t_{pos} represents the true positive, f_{neg} denotes the false negative. The recall is measured in percentage (%).

Table 3 Recall

Number of reviews	Recall (%)		
	CRZICDBNLC	fastText+ CNN model	SC-ABiLSTM
10000	97.82	94.18	91.46
20000	98.68	96.66	94.04
30000	98.95	96.33	94.31
40000	98.93	96.73	95.23
50000	98.72	96.71	95.55
60000	98.58	96.73	95.77
70000	98.49	96.89	95.29
80000	98.42	96.74	96.16
90000	98.24	96.13	95.69
100000	98.21	95.62	95.02

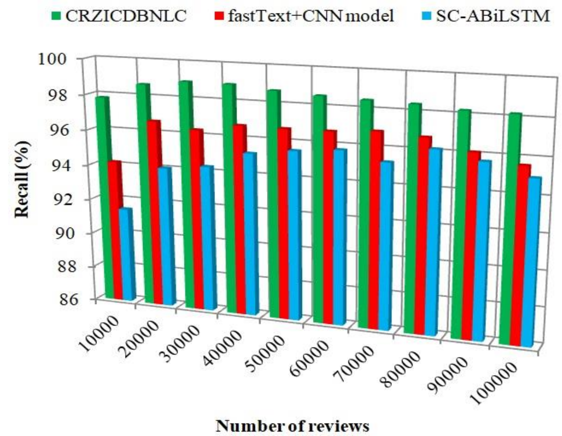

Figure 7 Performance results of the recall

Table 3 and figure 7 illustrate the performance outcomes of recall with respect to a number of reviews collected from the dataset. The performance of recall is measured using three different methods CRZICDBNLC and two existing classification techniques namely the fastText+CNN model [1] and SC-ABiLSTM [2] with a varied number of reviews. We considered 10000 reviews for conducting the experiment, the proposed CRZICDBNLC technique attains 97.82% of recall values whereas the fastText+CNN model [1] and SC-ABiLSTM [2] achieves 94.18% and 91.46% respectively. Similarly, other results are observed for different counts of the input reviews. After obtaining the ten results, the performance results of the proposed CRZICDBNLC technique were compared with the existing methods [1] and [2]. The overall comparison results shows that the recall of the CRZICDBNLC technique is better and has increased by 2% and 4% when compared to conventional methods [1], [2] respectively.

D. Performance of F-measure

The F-measure is measured as the mean of precision as well as recall. It is formulated as given below,

$$f - mes = \left[2 * \frac{Prc * Rec}{Prc + Rec}\right] * 100 \qquad (9)$$

Where f – mes denotes an F-measure Prc denotes precision, Rec denotes a recall. F-measure is measured in terms of percentage (%).

Table 4 F-measure

Number of reviews	F-measure (%)		
	CRZICDBNLC	fastText+ CNN model	SC-ABiLSTM
10000	97.29	93.63	90.90
20000	98.29	95.59	92.12
30000	98.60	95.62	93.42
40000	98.14	96.33	94.43
50000	97.78	95.76	94.70
60000	97.89	95.94	95.16
70000	97.97	96.14	95.06
80000	98.03	96.22	95.70
90000	97.95	95.77	95.10
100000	97.95	95.35	94.75

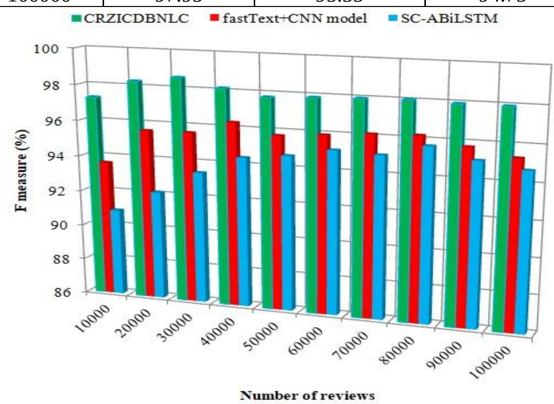

Figure 8. Performance results of the recall

Performance results of the F-measure for varying numbers of reviews are depicted in table 4 and figure 8. The number of reviews about car and hotels are collected from the OpinRank dataset. The observed results designate that the F-measure of the proposed CRZICDBNLC technique is increased when compared to the conventional methods. This is confirmed through mathematical analysis. Let us consider the 10000 reviews. The F-measure of the CRZICDBNLC technique is 97.29% whereas the F-measure of exiting methods namely fastText+CNN model [1] and SC-ABiLSTM [2] are 93.56% and 90.90% respectively. Based on the observed results, the proposed CRZICDBNLC technique enhances the F-measure by 2% and 4% when compared to [1] and [2] existing methods.

E. Performance of computational time

It is defined as the amount of time consumed by the algorithm to perform the online review classification. Therefore, the overall time consumption is formulated as follows,

$$Ct = n * time\ [COR] \quad \text{- - - - - - - - - (10)}$$

Where Ct denotes a computational time, n denotes number of reviews, COR indicates a classification of each one of reviews. The time is measured in terms of milliseconds (ms).

Table 5 Computational time

Number of reviews	Computational time (ms)		
	CRZICDBNLC	fastText+CNN model	SC-ABiLSTM
10000	60	70	80
20000	66	74	78
30000	70.5	78	84
40000	76	80	88
50000	85	90	100
60000	90	96	102
70000	94.5	98	105
80000	100	104	112
90000	104.4	108	117
100000	110	120	125

Table 5 shows the experimental results of computational time of review classification using CRZICDBNLC technique, fastText+CNN model [1], and SC-ABiLSTM [2]. Among three different methods, the proposed CRZICDBNLC technique consumes lesser time than the conventional [1] and [2] methods. This is proved through the statistical measure. Consider 10000 reviews for classification, the proposed CRZICDBNLC technique consumes the $60ms$ of time consumption whereas the time consumption of existing [1] and [2] is 70ms and 80ms respectively. Likewise, ten various results are observed and the proposed CRZICDBNLC minimizes the computational time by 7% and 14% when compared to the existing fastText+CNN model [1] and SC-ABiLSTM [2] respectively.

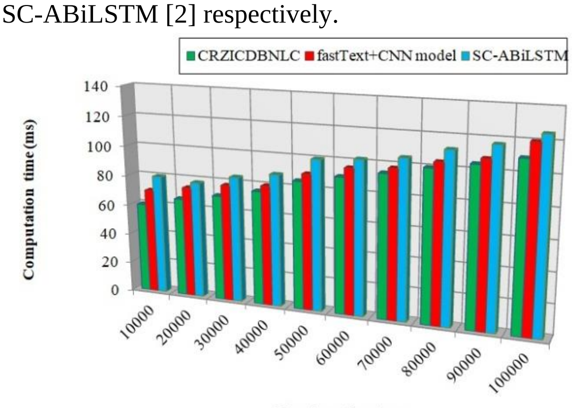

Figure 9. Performance results of computational time

Figure 9 shows the performance of Computational time for a different number of reviews in the ranges from 10000 to 100000. The observed results confirm that the CRZICDBNLC technique is significantly minimized than the conventional methods fastText+CNN model [1] and SC-ABiLSTM [2]. As seen in figure 9, the computational time of all the methods gets increased while increasing the number of reviews. However, the proposed CRZICDBNLC technique consumes less Computational time. This is due to the application of review preprocessing and feature extraction. Initially, number of reviews are collected and applied the preprocessing in the first hidden layer of deep learning. First, the review is portioned into a number of words using textblop tokenizer. Then the stop words are removed by applying a Spacy technique. Followed by Paice/Husk stemmer s applied to remove the stem words. After that, the Feature extraction process is performed in the second hidden layer using Censored Regression. Based on extracted features, the classification is performed with minimum time consumption.

VII. CONCLUSION

In this paper, a deep learning model based on Opinion Mining called the CRZICDBNLC technique is introduced for improving the accuracy and minimizing time consumption. Sentiment classification is a well-known mining technique to show user reviews about certain things. Initially, the review preprocessing is carried out in Convolutional Deep Belief Network Learning Classification to remove the redundant words from the reviews and thus the classification performance gets increased with minimum time consumption. Later, the feature extraction process is carried out using Censored Regression. Finally, the classification is performed in Deep Learning to categorize the review into Positive, Negative and Neutral. This helps for increasing the classification accuracy. The experimental evaluation is carried out with different metrics such as accuracy, precision, recall, F-measure, and computational time. The observed results show that the CRZICDBNLC technique increases the accuracy, precision, recall, F-measure and minimizes the time consumption when compared to state-of-the-art works.

REFERENCES

[1] Isnaini Nurul Khasanah, “Sentiment Classification Using fastText Embedding and Deep Learning Model”, Procedia Computer Science, Elsevier, Volume 189, 2021, Pages 343-350

[2] Peng Wu, Xiaotong Li, Chen Ling, Shengchun Ding, Si Shen, “Sentiment classification using attention mechanism and bidirectional long short-term memory network”, Applied Soft Computing, Elsevier, Volume 112, 2021, Pages 1-13

[3] Xiaofei Zhu, Ling Zhu, Jiafeng Guo, Shangsong Liang, Stefan Dietze, “GL-GCN: Global and Local Dependency Guided Graph Convolutional Networks for aspect-based sentiment classification”, Expert Systems with Applications, Elsevier, Volume 186, 2021, Pages 1-11

[4] Salim Sazzed and Sampath Jayarathna, “SSentiA: A Self-supervised Sentiment Analyzer for classification from unlabeled data”, Machine Learning with Applications, Elsevier, Volume 4 2021, Pages 1-14

[5] Nur Syafiqah Mohd Nafis and Suryanti Awang, “An Enhanced Hybrid Feature Selection Technique Using Term Frequency-Inverse Document Frequency and Support Vector Machine-Recursive Feature Elimination for Sentiment Classification”, IEEE Access, Volume 9, 2021, Pages 52177 – 52192

[6] Ning Jin, Jiaxian Wu, Xiang Ma, Ke Yan, Yuchang Mo, “Multi-Task Learning Model Based on Multi-Scale CNN and LSTM for Sentiment Classification”, IEEE Access, Volume 8, 2020, Pages 77060 – 77072

[7] Jawad Khan, Aftab Alam, Youngmoon Lee, “Intelligent Hybrid Feature Selection for Textual Sentiment Classification”, IEEE Access, Volume 9, 2021, Pages 140590 – 140608

[8] Mehmet Umut Salur and Ilhan Aydin, “A Novel Hybrid Deep Learning Model for Sentiment Classification”, IEEE Access, Volume 8, 2020, Pages 58080 – 58093

[9] Aminu Da’u, Naomie Salim, Idris Rabiu, Akram Osman, “Recommendation System Exploiting Aspect-based Opinion Mining with Deep Learning Method”, Information Sciences, Elsevier, Volume 512, February 2020, Pages 1279-1292

[10] Ahmed R. Abas; Ibrahim El-Henawy; Hossam Mohamed; Amr Abdellatif, “Deep Learning Model for Fine-Grained Aspect-Based Opinion Mining”, IEEE Access, Volume 8, 2020, Pages 128845 – 128855

[11] Bingkun Wang, Weina He, Zhen Yang, Shufeng Xiong, "An Unsupervised Sentiment Classification Method Based on Multi-Level Fuzzy Computing and Multi-Criteria Fusion", IEEE Access, Volume 8, 2020, Pages 145422 – 14543

[12] Xuelian Li; Bi Wang; Lixin Li; Zhiqiang Gao; Qian Liu; Hancheng Xu; Lanting Fang, "Deep2s: Improving Aspect Extraction in Opinion Mining with Deep Semantic Representation", IEEE Access, Volume 8, 2020, Pages 104026 – 104038

[13] You Zhang, Jin Wang, Xuejie Zhang "Personalized sentiment classification of customer reviews via an interactive attributes attention model", Knowledge-Based Systems, Elsevier, Volume 226, 2021, Pages 1-12

[14] Xiabing Zhou, Zhongqing Wang, Min Zhou, Qifa Wang, Shoushan Li, Min Zhang & Guodong Zhou, "Sentiment classification via user and product interactive modeling", Science China Information Sciences, Volume 64, 2021, Pages 1-13

[15] P. Sasikala & L. Mary Immaculate Sheela, "Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS", Journal of Big Data, Volume 7, 2020, Pages 1-20

[16] Ming Zhang, Vasile Palade, Yan Wang, Zhicheng Ji, "Attention-based Word Embeddings using Artificial Bee Colony Algorithm for Aspect-level Sentiment Classification", Information Sciences, Volume 545, 2021, Pages 713-738

[17] Yueshen Xu, Lei Li, Honghao Gao, Lei Hei, Rui Li, YihaoWang, "Sentiment classification with adversarial learning and attention mechanism", Computational Intelligence, Wiley, Volume 37, Issue 2, 2020, Pages 774-798

[18] Nurul Aida Osman, Shahrul Azman Mohd Noah, Mohammad Darwich, Masnizah Mohd, "Integrating contextual sentiment analysis in collaborative recommender systems", PLoS ONE, Volume 16, Issue 3, 2021, Pages 1-21

[19] Hengliang Tang, Yuan Mi, Fei Xue, Yang Cao, "Graph Domain Adversarial Transfer Network for Cross-Domain Sentiment Classification", IEEE Access, Volume 9, 2021, Pages 33051 – 33060

[20] Abdur Rasool, Ran Tao, Marjan Kamyab, Shoaib Hayat, "GAWA–A Feature Selection Method for Hybrid Sentiment Classification", IEEE Access, Volume 8, 2020, Pages 191850 - 191861