Journal of Smart Internet of Things (JSIoT)
VOL 2023, No.02 | 84-104 | 2023
DOI: 10.2478/jsiot-2023-0012

Article

Computational Intelligence-based Data Analytics for Sentiment Classification on Product Reviews

Ramy Riad Al-Fatlawy[1]

[1] Computer Engineering, Artificial Intelligence and Robotics department, Islamic Azad University (Khorasgan), Isfahan, Iran

*Corresponding author: Ramy Riad Al-Fatlawy (email: 39917541044049@khuisf.ac.ir)

Received 17 August 2023; Accepted 13 September 2023; Issue Published 15 December 2023

Abstract: Computational Intelligence (CI) is a sub-field of Artificial Intelligence (AI) that concentrates on the development of intelligent techniques and models to solve complex problems. When it comes to data analytics, CI techniques can be highly effective in handling large volumes of data, extracting meaningful patterns, and making accurate predictions. An increasing number of online reviews were being posted on the Internet every day with the rapid growth of Electronic Commerce (E-Commerce) and social networks. In the fastest growing research area, Sentiment Analysis (SA) is one among them that helps consumers in making better decisions relating to purchases through proper analysis and understanding of shared sentiments from social media and the web. In recent times, several approaches were modelled for acquiring insights from such datasets. But still, there comes a problem in managing text of large size; hence, precise polarity recognition of consumer reviews was an exciting and ongoing issue. This article introduces a novel Sentiment Analysis on Product Reviews using Enhanced Grasshopper Optimization with Deep Learning (SAPR-EGODL) approach. The objective of the SAPR-EGODL approach lies in the identification and classification of different sentiment types that exist in product reviews. At the initial stage, data pre-processing takes place which transforms the product review data into meaningful data. Next, the SAPR-EGODL technique employs Multi-Head Attention-based Bidirectional Long Short Term (MHABLSTM) technique for sentiment classification. In this study, the EGO model is exploited for improving the classification accomplishment of the MHABLSTM technique. A sequence of simulations was accomplished on different datasets for examining the advanced sentiment classification results of the SAPR-EGODL technique. The comprehensive relative research exhibited the promising accomplishment of the SAPR-EGODL model compared to current techniques.

Keywords: Data analytics; Sentiment analysis; Product reviews; Product/Service development; Deep learning; Enhanced grasshopper optimization algorithm

Copyright © 2023 Journal of Smart Internet of Things (JSIoT) published by Future Science for Digital Publishing and Sciendo . This is an open access article license CC BY (https://creativecommons.org/licenses/by/4.0/)

1. Introduction

SA is a common application of data analytics in understanding the opinions and emotions expressed in product reviews. It involves using computational techniques to analyze text data and determine the sentiment associated with it, whether it's positive, negative, or neutral. Value co-creation refers to a dynamic procedure where organizations meet the needs of customers and realize value sharing amongst shareholders by using value exchange and resource integration [1]. The effective benefits of value co-creation to society and business have proved successful and effective in endorsing sustainability. The main aim of value co-creation was to fortify the communication of data and resources with other shareholders in service or product expansion [2]. Digital transformation is changing society and the way of communication between customers and businesses. An increasing number of online reviews were posted on the Internet every day with the fast progression of e-commerce and social networks [4]. The user-generated context in sharing opinions and ideas is becoming a significant source of service or product development. In supporting service or product design and development, information technologies had a great contribution to creating value.

SA includes inspecting the thoughts, opinions, attitudes, emotions, and perceptions of persons shared on social media [4]. Particularly, SA intends in classifying a specific written text as a negative, neutral, or positive sentiment. Three key methods indulged in SA were hybrid, machine learning-related, and lexicon related. The lexicon-related method can be classified into 2 methods they are dictionary and corpus-based. In a dictionary-related method, the sentiment classification can be done through a dictionary of terms identified in SentiWordNet and WordNet [5]. Conversely, the corpus-related analysis methods rely on the statistical analysis of contents by using methods linked to the Conditional Random Field (CRF), k-nearest neighbours, and Hidden Markov Models (HMM). Hence, the corpus-related method does not depend on a predefined dictionary [6]. Furthermore, Machine Learning (ML) related SA methods include Deep Learning (DL) and conventional methods. But the hybrid method compiles lexicon and ML methods.

Such conventional methods of SA are well-known methods because of their potential outcomes; but, one disadvantage of such methods was that they include feature engineering [7]. Hence, the DL method has been presented due to its ability to learn written text without the necessity to execute manual feature engineering, henceforth outpacing other techniques of SA. Numerous researchers have previously explored SA and its problems [8]. Owing to its influence and relevance in forming numerous cutting-edge apps, the author has selected SA of user evaluations by making use of DL [9]. SA using classification addresses the aforementioned problems through the extraction of subjective data from given text like customer reviews. Owing to their successes and popularity, DL techniques are becoming practical for attaining reasonable performance. The online review classification through a DL technique defines the complete semanticization of customer feedback by accurate class into negative and positive emotions [10]. The data used was a compilation of accessory products and Amazon cell reviews acquired from the Snap data.

This article introduces a novel Sentiment Analysis on Product Reviews using Enhanced Grasshopper Optimization with Deep Learning (SAPR-EGODL) approach. At the initial stage, data pre-processing takes place which transforms the product review data into meaningful data. Next, the SAPR-EGODL technique employs Multi-Head Attention-based Bidirectional Long Short Term (MHABLSTM) technique for sentiment classification. In this study, the EGO model is exploited for improving the classification accomplishment of the MHABLSTM technique. A sequence of simulations was accomplished on different datasets for examining the advanced sentiment classification results of the SAPR-EGODL technique.

2. Related Works

Krishnan et al. [11] suggest a novel SA of product review dependent upon intelligent approaches. Primarily, the MongoDB recorded tweets can be exposed to pre-processed stages such as blank space extraction, stop word extraction, and stemming. Additionally, the keywords can be extracted in the pre-processing tweets. Based on keywords extraction, the recent semantic words were extracted, and then the sentiment keyword are classified. The classifier of extracting features was completed utilizing a DL classification termed DBN whereas the optimized activation function was completed by WD-WOA. Bhuvaneshwari et al. [12] establish the Bi-LSTM Self Attention-based CNN (BAC) approach for the subjectivity classifier of reviews. During this method, the authors utilize pre-training word embedded for reducing the dimensional of text representations that avoid data sparsity problems. The authors also carry out an attention process for capturing n-gram factors and concentration on the vital data in the context by setting distinctly weighted betwixt word and sentence. The presented method utilizes CNN and Bi-LSTM for learned classifier features and capturing the semantic and context data that is vital in defining the sentiment divisions.

In [13], a new automated method was presented in this document to effectual online product SA. Afterwards gathering the multi-modal data in the Amazon sites, the image and data normalized systems can be utilized to the optimum understanding of gathered data. At last, the Random Evolutionary WOA (REWOA) and DBN systems are combined for sentiment classifier and feature vector optimization. Liu et al. [14] present the DL approach of Bert-BiGRU-Softmax with hybrid masking, extracting attention and review method that executes sentiment Bert method as input layer for extracting multi-dimension product features in e-commerce review, and Softmax with attention process as the outcome state for classifying the positive/negative suggestion.

In [15], a new system was projected that employs feature-level sentiment recognition that concentrates on the item features. This work is executing and testing on the customer reviews of Amazon (crawled data) whereas feature terms can be recognized initially in all the reviews. The method applies pre-processed functions namely tokenization, casing, stemming, and stop-word extraction on the database for extracting significant data and lastly provides a rank to their classifier in negative or positive reviews. The authors [16] utilized a DL Modified NN (DLMNN), this method was presented intended at SA of online product reviews; also, using Improved ANFIS (IANFIS), this system was projected intended at future predictive of online products for trouncing the above-mentioned problems. IANFIS applies a weighted feature and classifier on the products to upcoming predictive. In [17], the authors project a DL-based method for SA on Twitter posts. During the projected method, three-word embedded-based systems (like fastText, GloVe, and word2vec) and CNN are employed.

3. The Proposed Model

In this article, a new SAPR-EGODL method for sentiment classification on online product reviews is developed. The major aim of the SAPR-EGODL method exists in the identification and classification of different sentiment kinds present in the review of the product. In this developed SAPR-EGODL model, three different processes are utilized called data pre-processing, sentiment classification using the MHABLSTM model, and EGOA-based hyperparameter tuning. Fig. 1 demonstrates the block diagram of the SAPR-EGODL model.

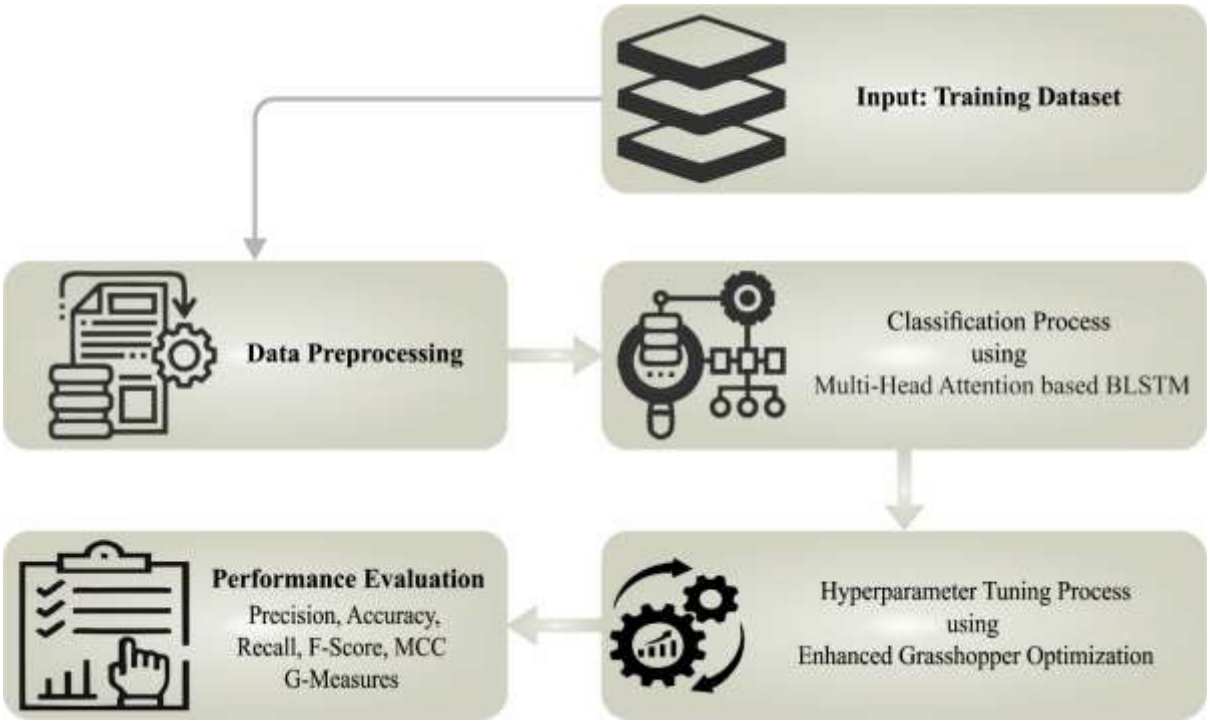


Figure 1: Block diagram of SAPR-EGODL methodology

3.1 Data Pre-processing and Word Embedding

Firstly, data pre-processing takes place which transforms the product review data into meaningful data [18]. The partitioned dataset from the text data has been preprocessed using stemming and eliminating stop words. Preprocessing is a major operation utilized to easily arrange different datasets and provide efficient outcomes by enhancing representation. The data comprises unwanted words and phrases that affect the operation. Firstly, the textual dataset is accumulated in the relational dataset and every discrete review is separated into sentences and an ensemble of sentences. The stop word removal is accomplished for enhancing the efficiency of classifying the text. Now, the stop word removal and stemming refined the dataset.

Removal of stop words: This is a procedure that eliminates words with lesser illustrative values for the dataset. Few non-depicted words comprise articles and pronouns. While estimating data, a few words aren't meant for the textual content, and eliminating these unnecessary words is imperative. This technique is named stop word removal. Specific word involves conjunctions, prepositions, and articles, constantly appearing and are termed stop words. For the removal of these stop words, the essential approach is applied for removing unwanted words through vocabulary since the vector space size doesn't provide any sense. The stop word indicates the word that doesn't contain any dataset. It is a procedure utilized for eliminating stop words from a larger review set. The removal of stop words is utilized to improve processing.

Stemming: This process is exploited to translate words to stem form. Consequently, the crucial model is utilized for minimizing words. It is the process of linguistic normalization where fewer words are decreased. Furthermore, it is the process utilized for retrieving data to describe the process of decreasing unwanted words to the root form and word stem. Words like connected, connection, and, connecting are decreased to connect.

$$M_l = \{Q_i, 1 \leq \eta \leq P_k\} \quad (1)$$

Whereas Q_i represents overall words existing in the textual dataset from the database. The

preprocessed result produced from preprocessing is formulated by M_l which is inputted to the feature extraction stage. For the word embedding process, the Glove technique is employed in this work.

3.2 Sentiment Classification using MHABLSTM Model

For accurate recognition and classification of sentiments, the MHABLSTM model is utilized in this work. LSTM extends RNN with a memory cell, instead of a recurrent unit, to ease the temporal relationship learning, and store output data on a long timescale [19].

$$i_t = \sigma_i(W_{ai}a_t + W_{hi}h_{t-1} + W_{ci}c_{t-1} + b_i) \quad (2)$$

$$f_t = \sigma\left(W_{af}a_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f\right) \quad (3)$$

$$c_t = f_t c_{t-1} + i_t \sigma_c(W_{ac}a_t + W_{hc}h_{t-1} + b_c) \quad (4)$$

$$o_t = \sigma_0(W_{ao}a_t + W_{ho}h_{t-1} + W_{co}c_t + b_0) \quad (5)$$

$$h_t = o_t \sigma_h(c_t) \quad (6)$$

From the expression, c, i, f, & o, indicates the cell activation vector, input, forget, and output gates, and the corresponding size as vector h defines the hidden value. σ denotes nonlinear function. $W_{ai}, W_{hi}, W_{ci}, W_{af}, W_{hf}, W_{cf}\ W_{ac}, W_{hc}, W_{ao}, W_{ho}$, & W_{co} shows weight matrices. b_j, b_f, b_c, & b_0 indicate bias vector. Fig. 2 demonstrates the architecture of BiLSTM:

$$s_t = softmax(0_{last} \times (0_{all} \times W_t)^H), 0_{last} \in R^{B,1,Z} \quad (7)$$

$$o_t = s_t \times 0_{all}, 0_{all} \in R^{B,T,Z}, s_t \in R^{B,1,T} \quad (8)$$

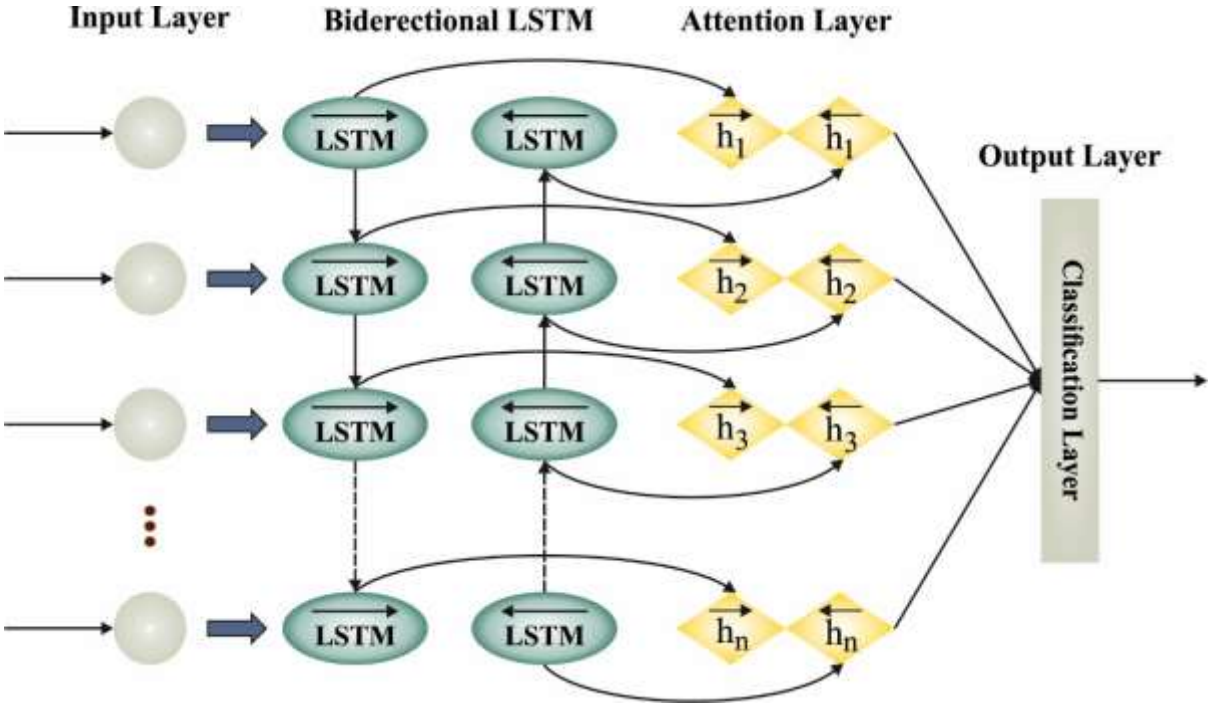


Figure 2: Structure of BiLSTM

From the expression, 0_{last} indicates the last time output, 0_{all} represents the time output, B denotes the size of the batch, T represents the time steps number, and Z indicates the dimension of the feature. W_t shows the parameter matrices, and H indicates the transpose operator:

$$K_i = W_{i,k} \times 0_{all} + b_{i,k}, K_i \in R^{B,T,\frac{Z}{n}}, W_{i,k} \in R^{Z,\frac{Z}{n}}, b_{i,k} \in R^{\frac{Z}{n}} \quad (9)$$

$$V_i = W_{i,v} \times 0_{all} + b_{i,v}, V_i \in R^{B,T,\frac{Z}{n}}, W_{i,v} \in R^{Z,\frac{Z}{n}}, b_{i,v} \in R^{\frac{Z}{n}} \quad (10)$$

$$Q_j = W_{i,q} \times 0_{last} + b_{i,q}, Q_j \in R^{B,1,\frac{Z}{n}}, W_{i,q} \in R^{Z,\frac{Z}{n}}, b_{i,q} \in R^{\frac{Z}{n}} \quad (11)$$

Now, $K, V, and\ Q$ represent the key, value, and query. n shows the attention heads number and b signifies bias is evaluated by the following expression:

$$s_i = softmax\left(Q_i \times K_i^H\right), s_i \in R^{B,1,T} \quad (12)$$

$$context_i = s_i \times V_i, context_i \in R^{B,1,\frac{Z}{n}} \quad (13)$$

$$CV = Concat([context_1, \dots, context_n]), CV \in R^{B,1,Z} \quad (14)$$

Now, $context_i$ represents the reduction dimension context vector from subspace and s_i depicts the attention score of the multi-head time dimension.

3.3 Hyperparameter Tuning using EGO Algorithm

In this study, the design of the EGO method assists in improving the classification accomplishment of the MHABLSTM method. GOA emulates the grasshopper insect behaviour. These grasshoppers affect crop productivity and agriculture, and the life cycle encompasses adulthood, egg, and nymph [20]. In the nymph phase, the major features contain the activities of moving and jumping in the rolling cylinder (with slower motion and smaller steps). In the adulthood stage, grasshoppers travel a long distance in swarming (with abrupt movement and long-range). This behaviour is scientifically expressed by taking into account the grasshopper potion (x_i):

$$x_i = S_i + G_i + A_i, i = 1,2, \dots, N \quad (15)$$

In Eq. (15), S_i indicates social interaction of i-th grasshopper:

$$S_i = \sum_{j=1, i \neq j}^{N} s\left(d_{ij}\right)\widehat{d_{ij}}, d_{ij} = \left|x_i - x_j\right| \quad (16)$$

Here d_{ij} shows the distance amongst $i\text{-}th$ and $j\text{-}th$ Grasshoppers; s signifies the strength of social force function:

$$s(y) = fe^{\frac{-y}{1}} - e^{-y} \quad (17)$$

From the expression, G_i and A_i correspondingly indicate the gravity force and wind advection for $i\text{-}th$ grasshopper, l and f signify the attractive length scale and intensity of attraction:

$$G_i = -g\hat{e}_g, A_i = u\hat{e}_w \quad (18)$$

Now, e_w and e_g signify the unity vector to the direction of the wind and the earth's centre, and g and u correspondingly characterize the gravitational constant and constant drift:

$$x_i = c\left(\sum_{j=1, i \neq j}^{N} c\frac{u-1}{2} s\left(\left|x_j - x_i\right|\right)\frac{x_j - x_i}{d_{ij}}\right) + \widehat{T_d} \quad (19)$$

Here l and u signify the lower and upper boundaries of the searching region; T_d portrays the optimum solution values, and s is represented in Eq. (17):

$$c = c_{\max} - t\frac{c_{\max} - c_{\min}}{t_{\max}} \quad (20)$$

In Eq. (20), $c_{\max}$ and $c_{\min}$ indicate the maximal value (corresponding to 1) and minimal value (corresponding to 0.00001) of c; t indicates the present iteration and $t_{\max}$ epitomizes the maximal

iteration amount. The pseudocode of GOA is portrayed in Algorithm 1.

Algorithm 1: Pseudocode of GOA

Initialize the parameter values like population size(N), c_{max}, c_{min} and maximal amount of iteration ($t_{\max}$)
Randomly produce a population (X)
Set the current iteration $t = 1$
while ($t < t_{\max}$) do
    Estimate the fitness function f
    Choice of the optimal solution $\hat{T}_d$
    Upgrade the value of c based on Eq. (20)
    for $i = 1{:}N$ do
        Normalize the distance between the solutions in X.
        Upgrade $x_i \in X$ by Eq. (19)
    end for
        $t = t + 1$
end while
Return $\hat{T}_d$.
where t

In this work, the OBL technique has been exploited to describe the opposite solution for a present solution, and later applies the fitness function (f) values to define if the opposite is superior to the existing solution. The major description of OBL is taking the opposite value $\overline{x}$ to the real value $x \in [u$ into account:

$$\overline{x} = u + l - x \tag{21}$$

This description is a generalization to n-dimensional based on the succeeding equation:

$$\overline{x} = u_i + l_i - x_i, i = 1,2, \dots, N \tag{22}$$

In Eq. (22), $\overline{x} \in R^n$ denotes the opposite vector in real vector $x \in R^n$. Additionally, the 2 solutions (x and $\overline{x}$) are evaluated, and the optimal solution is saved, but the other was eliminated by relating the fitness function. For instance, if $f(x) \leq f(\overline{x})$ (to minimized), x, and $\overline{x}$ are stored. The EGO approach derives an FF for accomplishing better classification accuracy. It defines a positive integer to describe the improved candidate solutions accomplishment. The classification error rate reduction is considered as the fitness function.

$$fitness(x_i) = ClassifierErrorRate(x_i)$$
$$= \frac{number\ of\ misclassified\ samples}{Total\ number\ of\ samples} * 100 \tag{23}$$

4. Performance Validation

The performance evaluation of the SAPR-EGODL approach is tested by implementing two datasets: Amazon-Fine-Food Reviews (AFFR) and Amazon Products (AP). Table 1 illustrates the dataset description.

Table 1 Dataset Description

Class	**Dataset**	
	Amazon-Fine-Food-Reviews	**Amazon-Products**
Positive	21784	13251

Negative	3216	749
Total Number of Instances	**25000**	**14000**

The confusion matrix of the SAPR-EGODL model on the enforced AFFR dataset is depicted in Fig. 3. The figure depicted that the SAPR-EGODL model has properly categorized the positive and negative reviews under all cases.

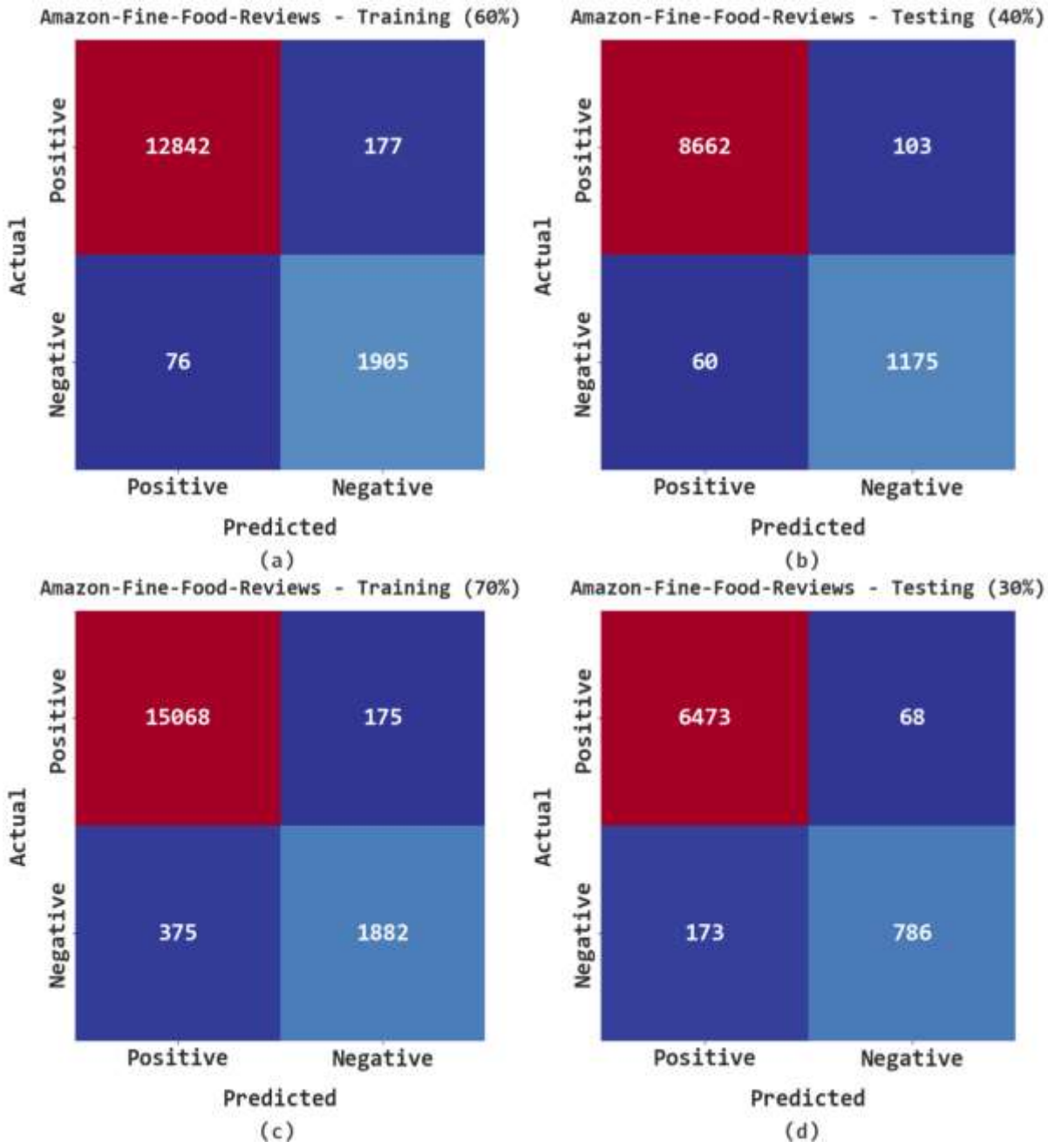


Figure 3: Confusion matrix of SAPR-EGODL technique under AFFR dataset TR and TS datasets of (a-b) 60:40 and (c-d) 70:30

Table 2 highlights the complete sentiment classification outputs of the SAPR-EGODL method on the AFFR dataset. The outcome shows that the SAPR-EGODL method has classified the positive

and negative reviews. As a sample, on 60% of TR data, the SAPR-EGODL model has shown an effectual outcome with an average $accu_y$ of 97.40%, $prec_n$ of 95.46%, $reca_l$ of 97.40%, F_{score} of 96.40%, MCC of 92.84%, and $G_{measure}$ of 96.41%. Furthermore, on 40% of TS data, the SAPR-EGODL approach has demonstrated effective outcomes with average $accu_y$ of 96.98%, $prec_n$ of 95.63%, $reca_l$ of 96.98%, F_{score} of 96.29%, MCC of 92.60%, and $G_{measure}$ of 96.30%. Additionally, on 70% of TR data, the SAPR-EGODL approach has illustrated effective outcomes with average $accu_y$ of 91.12%, $prec_n$ of 94.53%, $reca_l$ of 91.12%, F_{score} of 92.73%, MCC of 85.58%, and $G_{measure}$ of 92.78%.

Table 2 Sentiment classification outcome of SAPR-EGODL methodology under the AFFR dataset

Amazon-Fine-Food-Reviews						
Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	MCC	$G_{measure}$
Training Phase (60%)						
Positive	98.64	99.41	98.64	99.02	92.84	99.03
Negative	96.16	91.50	96.16	93.77	92.84	93.80
Average	**97.40**	**95.46**	**97.40**	**96.40**	**92.84**	**96.41**
Testing Phase (40%)						
Positive	98.82	99.31	98.82	99.07	92.60	99.07
Negative	95.14	91.94	95.14	93.51	92.60	93.53
Average	**96.98**	**95.63**	**96.98**	**96.29**	**92.60**	**96.30**
Training Phase (70%)						
Positive	98.85	97.57	98.85	98.21	85.58	98.21
Negative	83.39	91.49	83.39	87.25	85.58	87.34
Average	**91.12**	**94.53**	**91.12**	**92.73**	**85.58**	**92.78**
Testing Phase (30%)						
Positive	98.96	97.40	98.96	98.17	85.07	98.18
Negative	81.96	92.04	81.96	86.71	85.07	86.85
Average	**90.46**	**94.72**	**90.46**	**92.44**	**85.07**	**92.51**

The TACC and VACC values of the SAPR-EGODL approach under AFFR dataset accomplishment in Fig. 4. The figure depicts that the SAPR-EGODL approach has illustrated advanced accomplishment with TACC and VACC values. Note that the SAPR-EGODL model has attained the greatest TACC outputs.

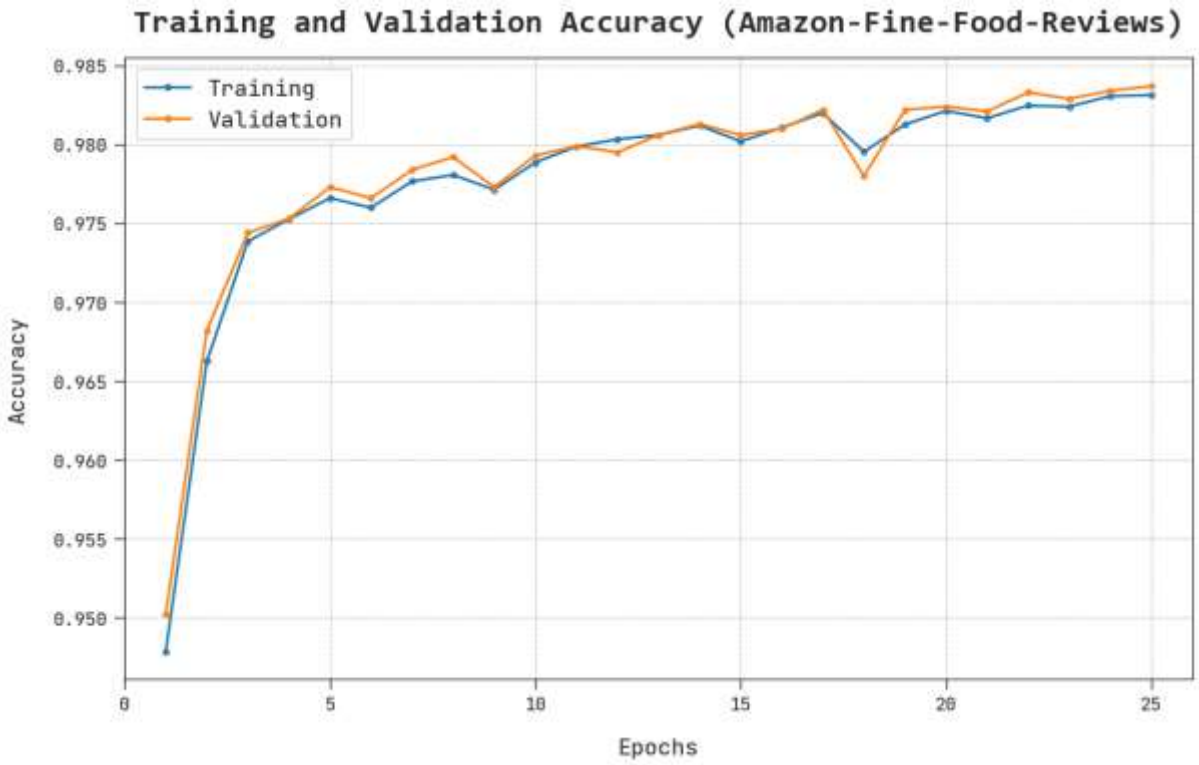


Figure 4: TACC and VACCC evaluation of SAPR-EGODL methodology under the AFFR dataset

The TLS and VLS values of the SAPR-EGODL methodology are examined under AFFR dataset accomplishment in Fig. 5. The figure depicts that the SAPR-EGODL methodology has illustrated advanced accomplishment with lower TLS and VLS values. Note that the SAPR-EGODL model has given an outcome in lessened VLS outputs.

A precise precision-recall examination of the SAPR-EGODL method under the AFFR dataset is described in Fig. 6. The figure showed that the SAPR-EGODL method has given output in advanced precision-recall values under the overall classes.

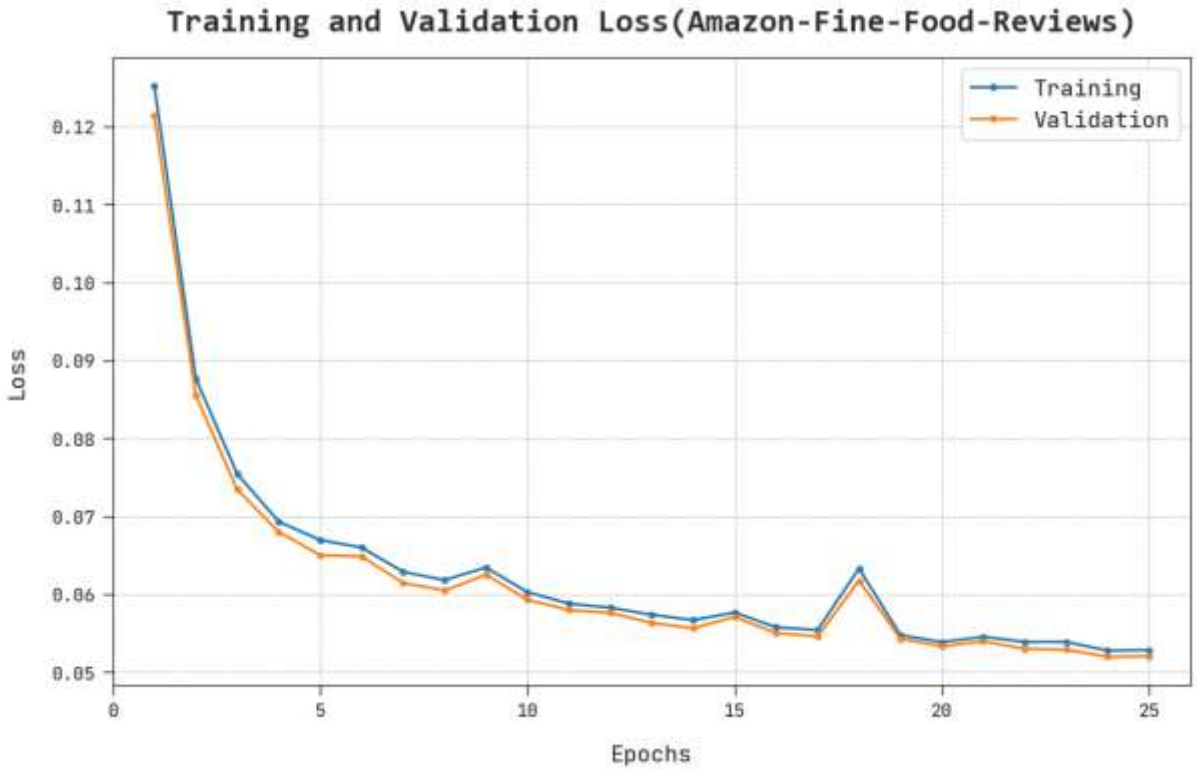


Figure 5: TLS and VLS evaluation of SAPR-EGODL methodology under the AFFR dataset

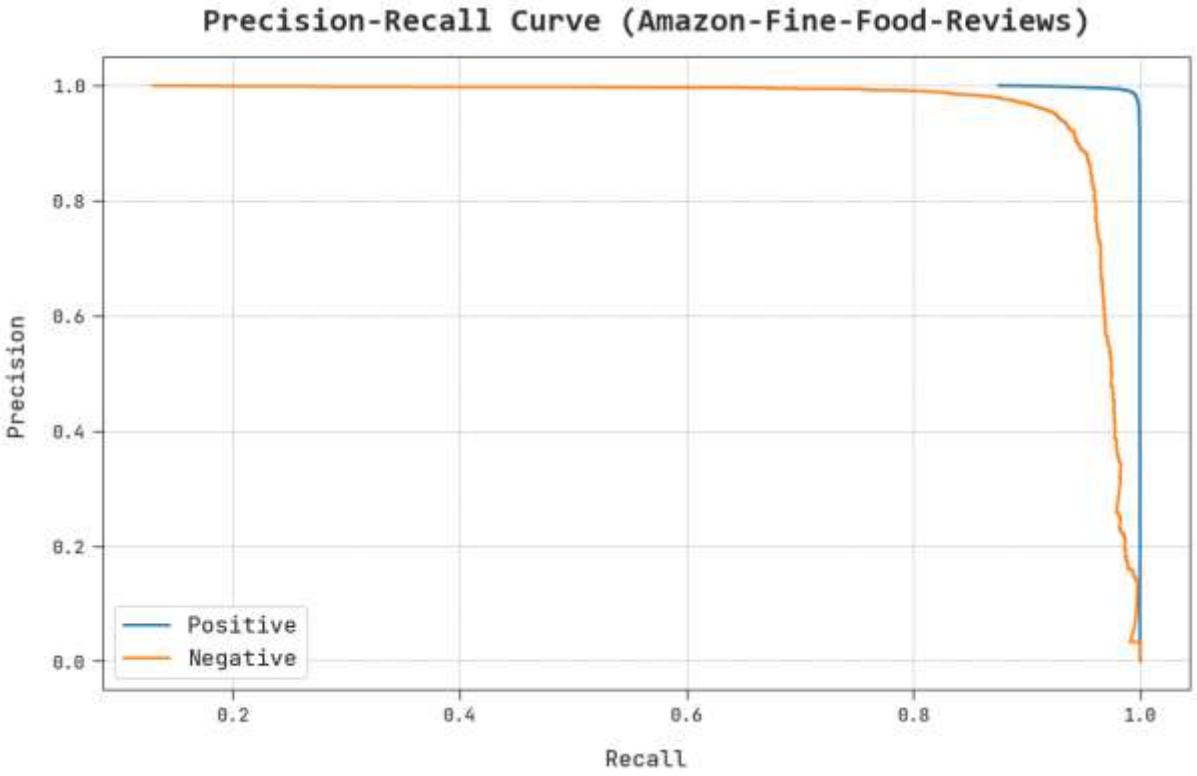


Figure 6: Precision-recall evaluation of SAPR-EGODL methodology under AFFR dataset

The confusion matrix of the SAPR-EGODL model on the applied Amazon products dataset is

demonstrated in Fig. 7. The figure depicts that the SAPR-EGODL model has appropriately categorized the positive and negative reviews under all cases.

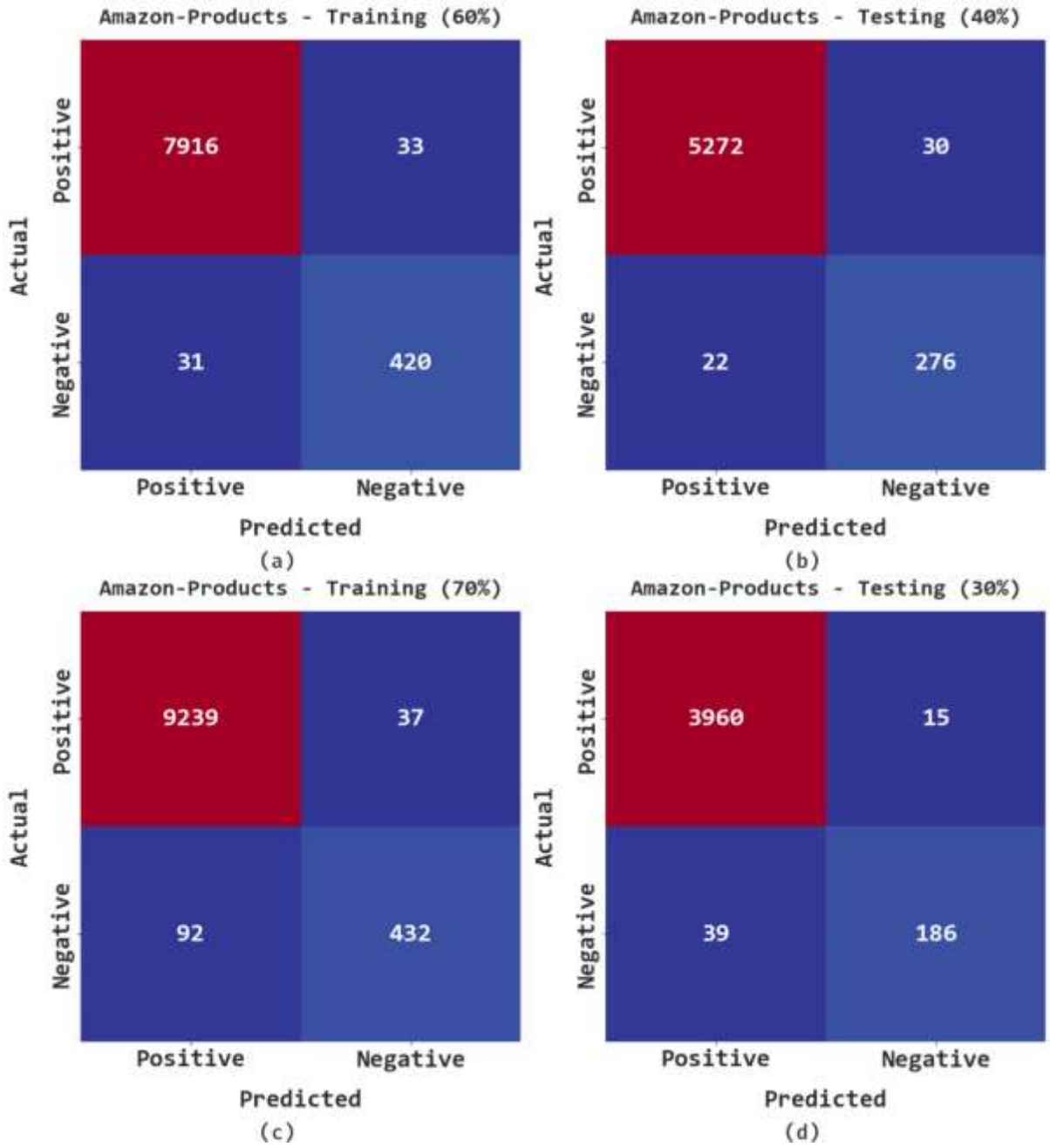


Figure 7: Confusion matrix of SAPR-EGODL methodology under Amazon product dataset TR and TS datasets of (a-b) 60:40 and (c-d) 70:30

Table 3 highlights the complete sentiment classification outputs of the SAPR-EGODL model on the Amazon products dataset. The outcome shows that the SAPR-EGODL model has classified the positive and negative reviews. As a sample, on 60% of TR data, the SAPR-EGODL approach has illustrated the effectual outputs with an average $accu_y$ of 96.36%, $prec_n$ of 96.16%, $reca_l$ of 96.36%, F_{score} of 96.26%, MCC of 92.52%, and $G_{measure}$ of 96.26%. Furthermore, on 40% of TS data, the SAPR-EGODL approach has illustrated effective outcomes with average $accu_y$ of

96.03%, $prec_n$ of 94.89%, $reca_l$ of 96.03%, F_{score} of 95.45%, MCC of 90.91%, and $G_{measure}$ of 95.45%. Additionally, on 70% of TR data, the SAPR-EGODL technique has demonstrated effective outcomes with average $accu_y$ of 91.02%, $prec_n$ of 95.56%, $reca_l$ of 91.02%, F_{score} of 93.16%, MCC of 86.47%, and $G_{measure}$ of 93.22%.

Table 3 Sentiment classification output of SAPR-EGODL methodology under Amazon products dataset

Amazon-Products						
Class	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}	**MCC**	$G_{measure}$
Training Phase (60%)						
Positive	99.58	99.61	99.58	99.60	92.52	99.60
Negative	93.13	92.72	93.13	92.92	92.52	92.92
Average	**96.36**	**96.16**	**96.36**	**96.26**	**92.52**	**96.26**
Testing Phase (40%)						
Positive	99.43	99.58	99.43	99.51	90.91	99.51
Negative	92.62	90.20	92.62	91.39	90.91	91.40
Average	**96.03**	**94.89**	**96.03**	**95.45**	**90.91**	**95.45**
Training Phase (70%)						
Positive	99.60	99.01	99.60	99.31	86.47	99.31
Negative	82.44	92.11	82.44	87.01	86.47	87.14
Average	**91.02**	**95.56**	**91.02**	**93.16**	**86.47**	**93.22**
Testing Phase (30%)						
Positive	99.62	99.02	99.62	99.32	86.80	99.32
Negative	82.67	92.54	82.67	87.32	86.80	87.46
Average	**91.14**	**95.78**	**91.14**	**93.32**	**86.80**	**93.39**

The TACC and VACC values of the SAPR-EGODL approach under Amazon products dataset accomplishment in Fig. 8. The figure depicts that the SAPR-EGODL approach has exhibited maximum accomplishment with enhanced TACC and VACC values. Note that the SAPR-EGODL model has attained the highest TACC outputs.

The TLS and VLS values of the SAPR-EGODL model are examined under Amazon products dataset accomplishment in Fig. 9. The figure illustrates that the SAPR-EGODL approach has shown enhanced accomplishment with minimum TLS and VLS values. The SAPR-EGODL model has attained the lowest VLS outputs.

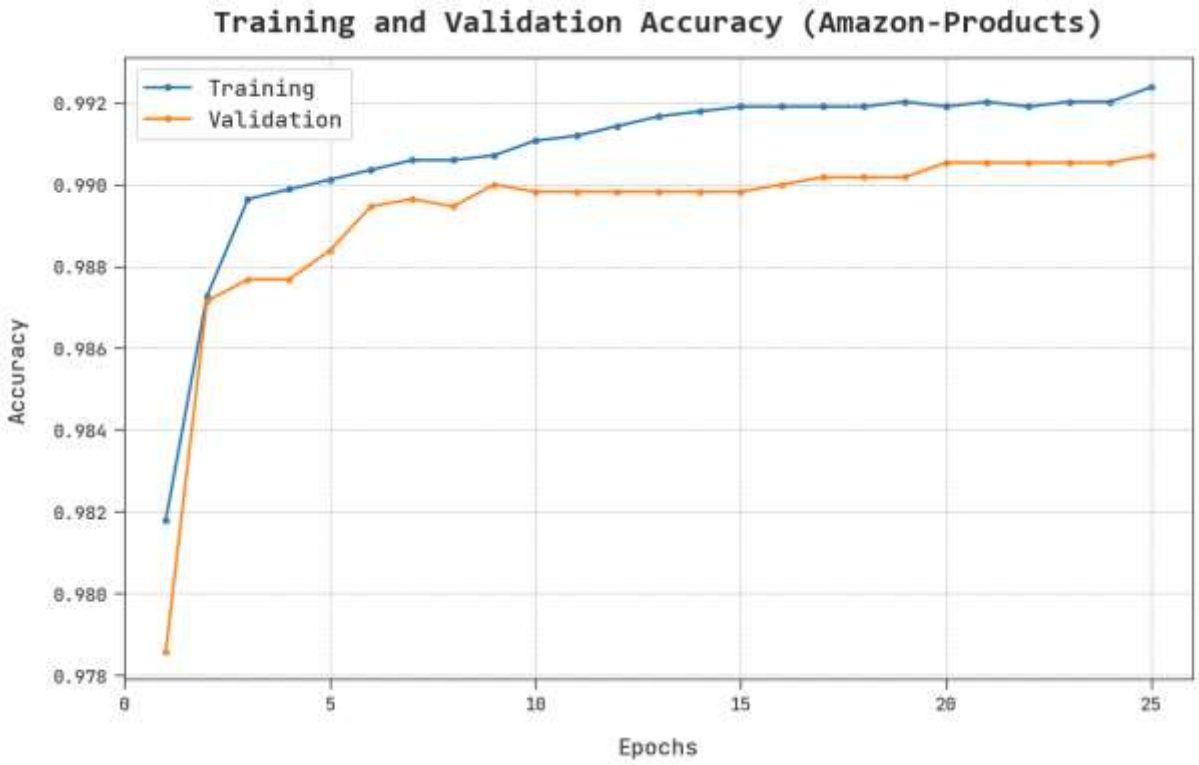


Figure 8: TACC and VACCC evaluation of SAPR-EGODL methodology under Amazon products dataset

A precise precision-recall analysis of the SAPR-EGODL technique under the Amazon products dataset is depicted in Fig. 10. The figure specified that the SAPR-EGODL technique has obtained enhanced precision-recall values under the overall classes.

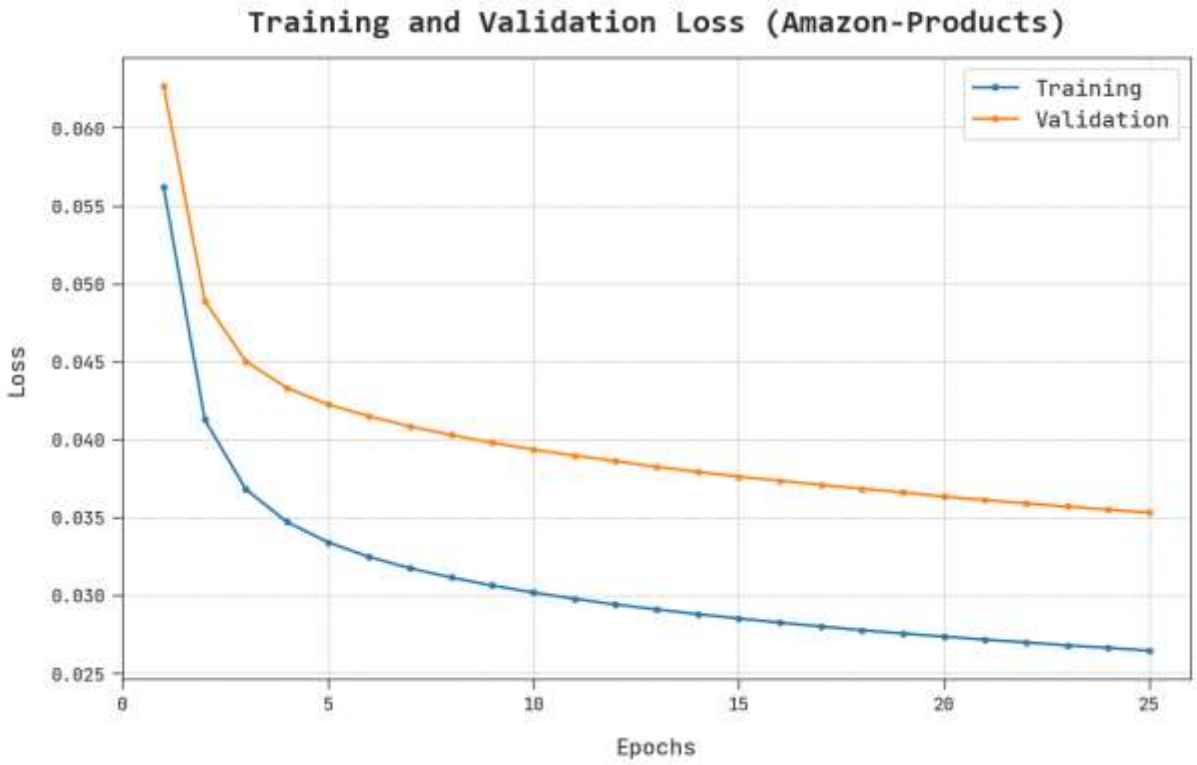


Figure 9: TLS and VLS evaluation of SAPR-EGODL methodology under Amazon products dataset

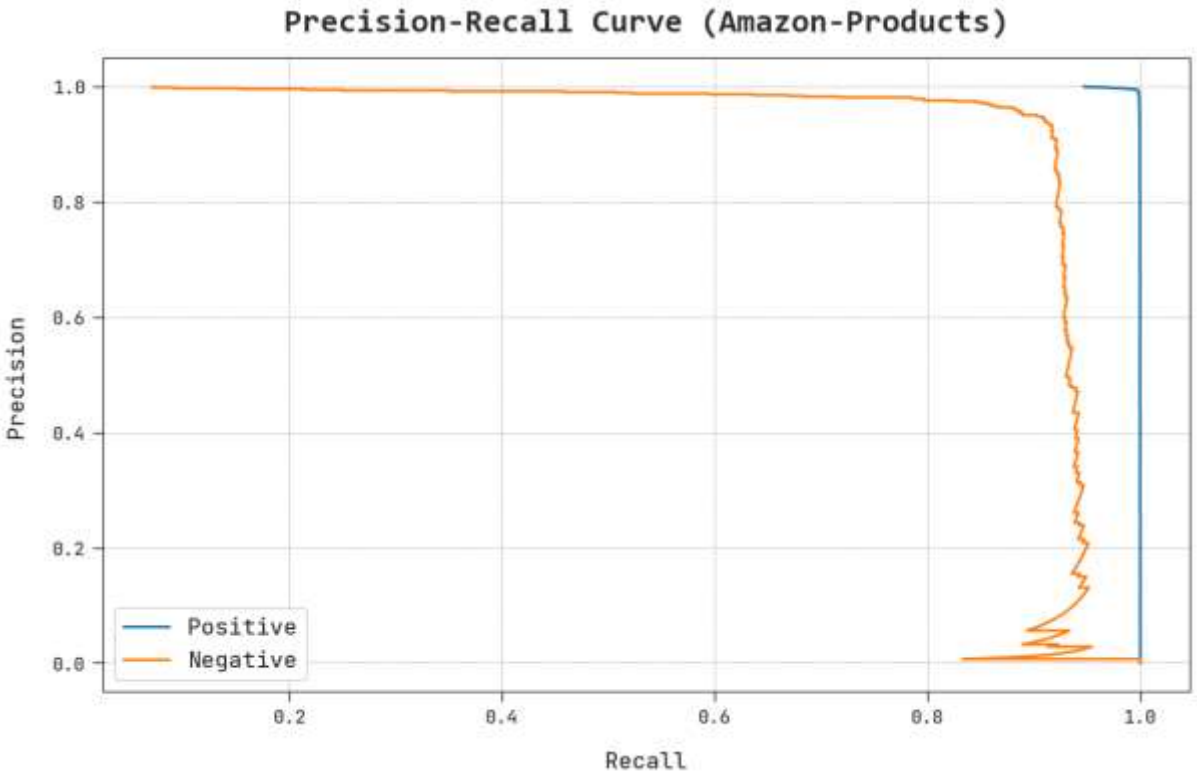


Figure 10: Precision-recall evaluation of SAPR-EGODL methodology under Amazon products dataset

Table 4 portrays a complete comparison results of the SAPR-EGODL approach on the AFFR dataset [21]. Fig. 11 offers a brief $accu_y$ and F_{score} analysis of the SAPR-EGODL approach with current methods. The outputs indicated the enhanced accomplishment of the SAPR-EGODL method over other models. Based on $accu_y$, the SAPR-EGODL method has reached an increased $accu_y$ of 97.40% while the LSTM, LSTM-ReLU, deep LSTM, ensemble, GoogleNet, and custom DLL models have resulted in reduced $accu_y$ of 92.05%, 89.18%, 92.65%, 92.75%, 91.41%, and 88.14% respectively. Meanwhile, based on F_{score}, the SAPR-EGODL model has obtained a better F_{score} of 96.40% whereas the LSTM, LSTM-ReLU, deep LSTM, ensemble, GoogleNet, and custom DLL models have resulted in decreased F_{score} of 89.67%, 92.52%, 91.89%, 89.95%, 89.67%, and 91.59% respectively.

Table 4 Relative evaluation of SAPR-EGODL methodology with current models under the AFFR dataset

Amazon-Fine-Food-Reviews				
Methods	$Accu_y$	$Prec_n$	$Reca_l$	F_{score}
SAPR-EGODL	97.40	95.46	97.40	96.40
LSTM	92.05	90.73	93.02	89.67
LSTM-ReLU	89.18	91.52	92.84	92.52
Deep LSTM	92.65	93.64	90.31	91.89
Ensemble Model	92.75	91.01	93.78	89.95
GoogleNet	91.41	88.36	92.47	89.67
Custom DLL Model	88.14	89.89	93.93	91.59

Fig. 12 exhibits the detailed $prec_n$ and $reca_l$ analysis of the SAPR-EGODL with recent techniques. The outcomes displayed the enhanced performance of the SAPR-EGODL over other techniques. Based on $prec_n$, the SAPR-EGODL technique has attained an increased $prec_n$ of 95.46% while the LSTM, LSTM-ReLU, deep LSTM, ensemble, GoogleNet, and custom DLL techniques have resulted in reduced $prec_n$ of 90.73%, 91.52%, 93.64%, 91.01%, 88.36%, and 89.89% respectively. Meanwhile, based on $reca_l$, the SAPR-EGODL model has attained an increased $reca_l$ of 97.40% while the LSTM, LSTM-ReLU, deep LSTM, ensemble, GoogleNet, and custom DLL models have resulted in a reduced $reca_l$ of 93.02%, 92.84%, 90.31%, 93.78%, 92.47%, and 93.93% respectively.

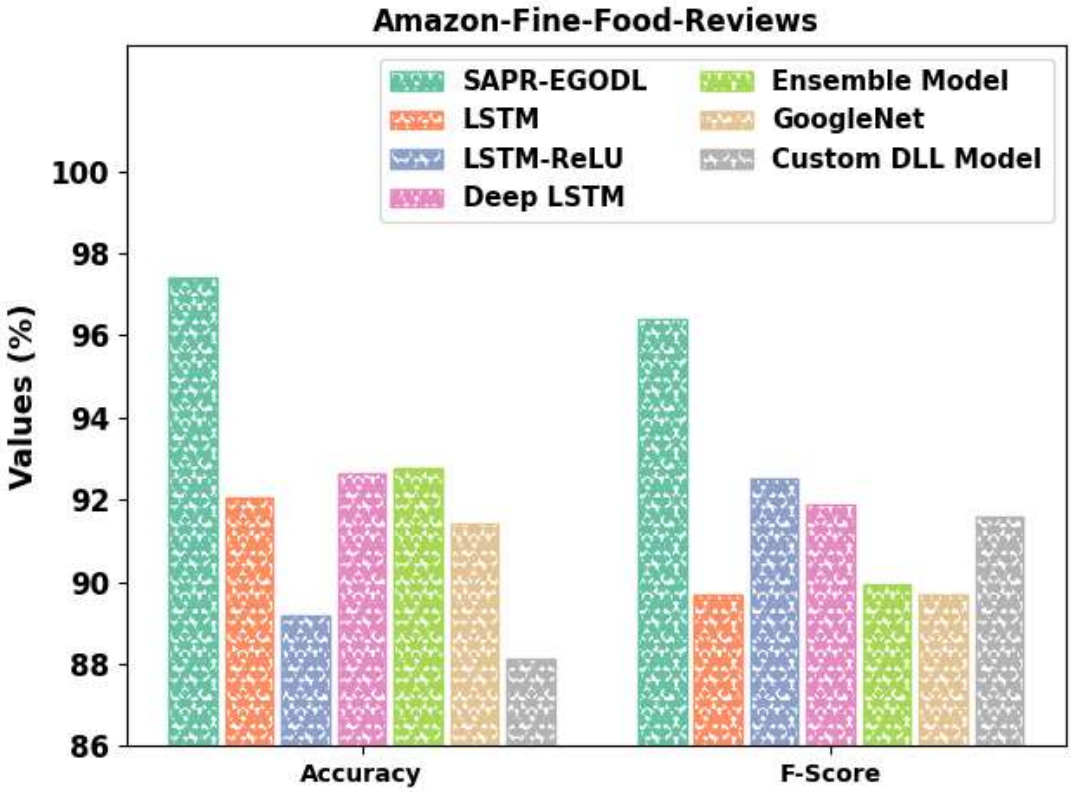


Figure 11: $Accu_y$ and F_{score} evaluation of SAPR-EGODL methodology under the AFFR dataset

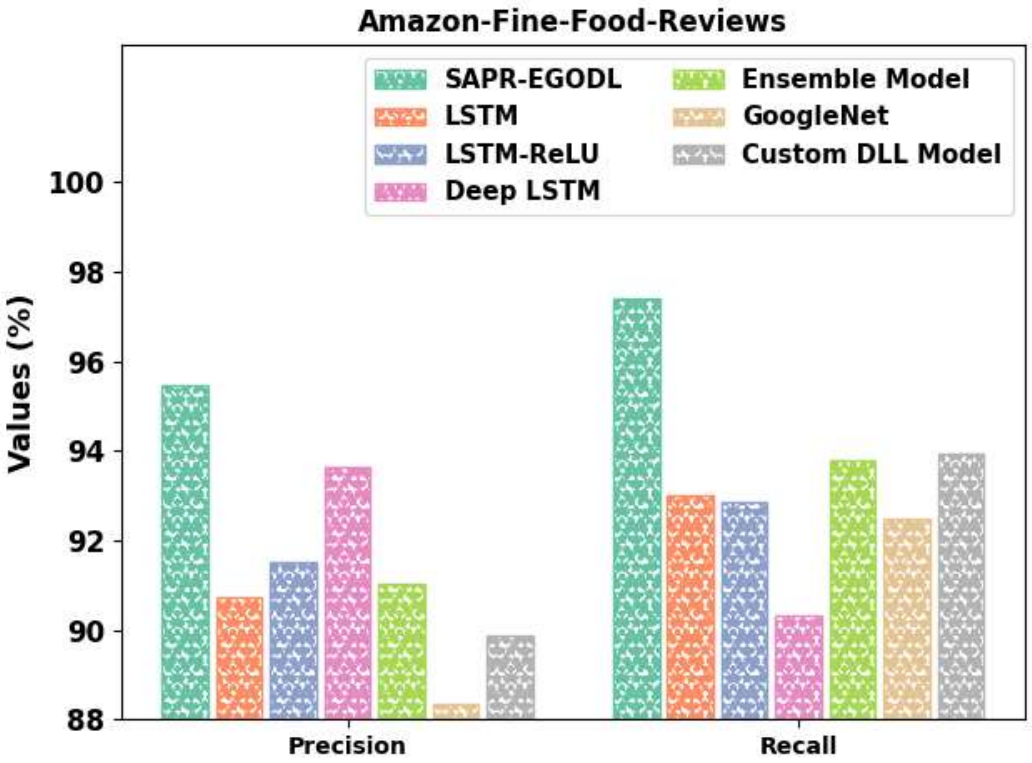


Figure 12: $Prec_n$ and $reca_l$ evaluation of SAPR-EGODL methodology under AFFR dataset

Table 5 shows an overall comparison of outputs of the SAPR-EGODL method on the Amazon dataset. Fig. 13 presents a detailed $accu_y$ and F_{score} analysis of the SAPR-EGODL method with recent methods. The outputs exhibited the enhanced accomplishment of the SAPR-EGODL over other methods. Based on $accu_y$, the SAPR-EGODL method has reached an increased $accu_y$ of 96.36% while the LSTM, LSTM-ReLU, deep LSTM, ensemble, GoogleNet, and custom DLL methods have resulted in reduced $accu_y$ of 90.72%, 89.79%, 90.80%, 91.84%, 88.15%, and 90.28% respectively. Meanwhile, based on F_{score}, the SAPR-EGODL method has attained an increased F_{score} of 96.26% while the LSTM, LSTM-ReLU, deep LSTM, ensemble, GoogleNet, and custom DLL approaches have resulted in a reduced F_{score} of 88.39%, 91.70%, 88.60%, 88.40%, 90.55%, and 91.80% respectively.

Table 5 Relative evaluation of SAPR-EGODL methodology with current methods under the Amazon products dataset

Amazon-Products				
Methodology	**Accuracy**	**Precision**	**Recall**	**F-Score**
SAPR-EGODL	96.36	96.16	96.36	96.26
LSTM	90.72	88.06	88.53	88.39
LSTM-ReLU	89.79	90.22	91.51	91.70
Deep LSTM	90.80	89.64	93.19	88.60
Ensemble Model	91.84	91.64	89.45	88.40

GoogleNet	88.15	90.22	89.33	90.55
Custom DLL Model	90.28	90.15	92.22	91.80

Fig. 14 portrays the detailed $prec_n$ and $reca_l$ analysis of the SAPR-EGODL with current approaches. The outputs exhibited the improved accomplishment of the SAPR-EGODL over other approaches. Based on $prec_n$, the SAPR-EGODL method has gained an increased $prec_n$ of 96.16% while the LSTM, LSTM-ReLU, deep LSTM, ensemble, GoogleNet, and custom DLL methods have resulted in reduced $prec_n$ of 88.06%, 90.22%,89.64%, 91.64%, 90.22%, and 90.15% correspondingly.

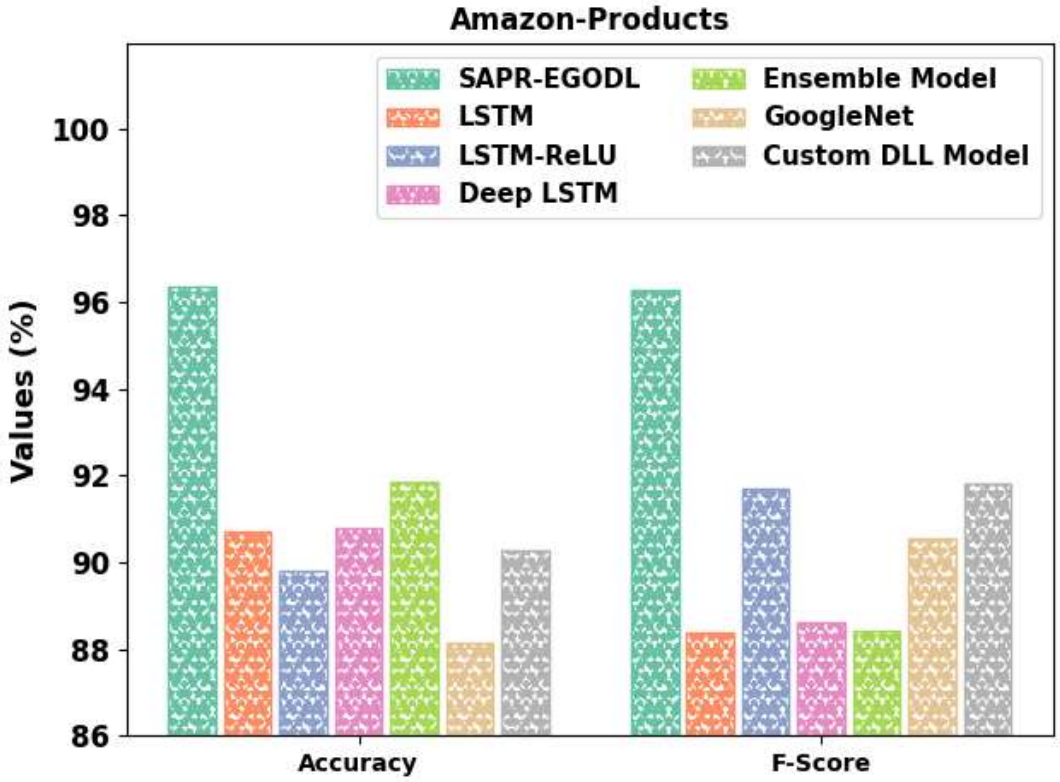


Figure 13: $Accu_y$ and F_{score} evaluation of SAPR-EGODL methodology under Amazon products dataset

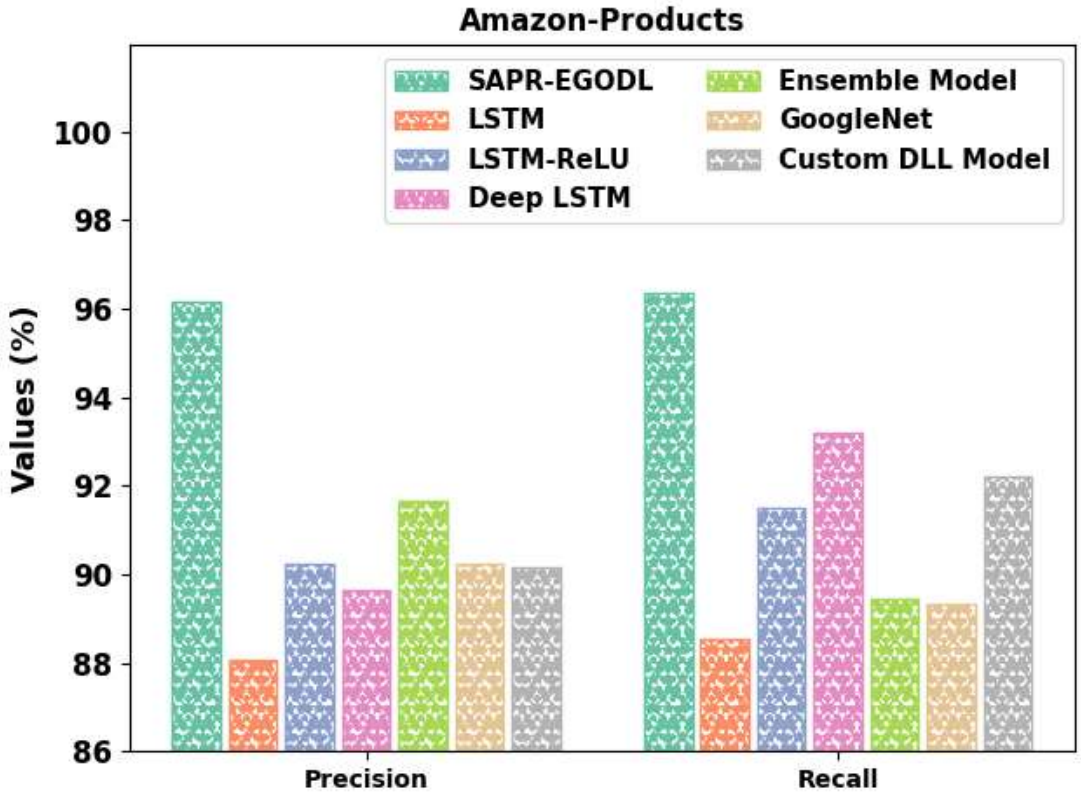


Figure 14: $Prec_n$ and $reca_l$ analysis of SAPR-EGODL approach under Amazon products dataset

Meanwhile, based on $reca_l$, the SAPR-EGODL method has attained an increased $reca_l$ of 96.36% while the LSTM, LSTM-ReLU, deep LSTM, ensemble, GoogleNet, and custom DLL methods have resulted in a reduced $reca_l$ of 88.53%, 91.51%, 93.19%, 89.45%, 89.33%, and 92.22% correspondingly. Thus, the SAPR-EGODL model can be implemented for accurate sentiment classification on product reviews.

5. CONCLUSIONS

In this article, a new SAPR-EGODL approach for sentiment classification on online product reviews is developed. The objective of the SAPR-EGODL approach lies in the identification and classification of different sentiment types that exist in product reviews. At the initial stage, data pre-processing takes place which transforms the product review data into meaningful data. Next, the SAPR-EGODL technique employs Multi-Head Attention-based Bidirectional Long Short Term (MHABLSTM) technique for sentiment classification. In this study, the EGO model is exploited for improving the classification accomplishment of the MHABLSTM technique. A sequence of simulations was accomplished on different datasets for examining the advanced sentiment classification results of the SAPR-EGODL technique. The comprehensive relative research exhibited the promising accomplishment of the SAPR-EGODL model compared to current techniques. In the future, the performance of the SAPR-EGODL algorithm can be boosted by an ensemble of fusion-based DL models.

References

[1] Bhateja, V., Yang, X.S., Lin, J.C.W., & Das, R. (2023). Evolution in Computational Intelligence. Evolution.
[2] Diekson, Z.A., Prakoso, M.R.B., Putra, M.S.Q., Syaputra, M.S.A.F., Achmad, S., & Sutoyo, R. (2023).

Sentiment analysis for customer review: Case study of Traveloka. Procedia Computer Science, 216, 682-690.
[3] Taherdoost, H., & Madanchian, M. (2023). Artificial intelligence and sentiment analysis: A review in competitive research. Computers, 12(2), 37.
[4] Mutinda, J., Mwangi, W., & Okeyo, G. (2023). Sentiment analysis of text reviews using lexicon-enhanced bert embedding (LeBERT) model with convolutional neural network. Applied Sciences, 13(3), 1445.
[5] Patruni, M.R., Angadi, A., Gorripati, S.K., & Saraswathi, P. (2023). Artificial Intelligence Techniques in Text and Sentiment Analysis. In Handbook of Research on Artificial Intelligence Applications in Literary Works and Social Media (pp. 171-191). IGI Global.
[6] Do, H.H., Prasad, P.W.C., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. Expert systems with applications, 118, 272-299.
[7] Haque, T.U., Saber, N.N., & Shah, F.M. (2018, May). Sentiment analysis on large scale Amazon product reviews. In 2018 IEEE international conference on innovative research and development (ICIRD) (pp. 1-6). IEEE.
[8] Fauzi, M.A. (2019). Word2Vec model for sentiment analysis of product reviews in Indonesian language. International Journal of Electrical and Computer Engineering, 9(1), 525.
[9] Verma, P., Dumka, A., Bhardwaj, A., & Ashok, A. (2022). Product Review-Based Customer Sentiment Analysis Using an Ensemble of mRMR and Forest Optimization Algorithm (FOA). International Journal of Applied Metaheuristic Computing (IJAMC), 13(1), 1-21.
[10] Sindhura, S., Praveen, S.P., Safali, M.A., & Rao, N. (2021, September). Sentiment Analysis for Product Reviews Based on Weakly-Supervised Deep Embedding. In 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA) (pp. 999-1004). IEEE.
[11] Krishnan, H., Elayidom, M.S., & Santhanakrishnan, T. (2022). Sentiment analysis of product reviews using weighted distance-based whale optimisation assisted deep belief network. International Journal of Business Information Systems, 39(2), 241-277.
[12] Bhuvaneshwari, P., Rao, A.N., Robinson, Y.H., & Thippeswamy, M.N. (2022). Sentiment analysis for user reviews using Bi-LSTM self-attention based CNN model. Multimedia Tools and Applications, 81(9), 12405-12419.
[13] Mehbodniya, A., Rao, M.V., David, L.G., Nigel, K.G.J., & Vennam, P. (2022). Online product sentiment analysis using random evolutionary whale optimization algorithm and deep belief network. Pattern Recognition Letters, 159, 1-8.
[14] Liu, Y., Lu, J., Yang, J., & Mao, F. (2020). Sentiment analysis for e-commerce product reviews by deep learning model of Bert-BiGRU-Softmax. Mathematical Biosciences and Engineering, 17(6), 7819-7837.
[15] Nandal, N., Tanwar, R., & Pruthi, J. (2020). Machine learning based aspect level sentiment analysis for Amazon products. Spatial Information Research, 28(5), 601-607.
[16] Sasikala, P., & Mary Immaculate Sheela, L. (2020). Sentiment analysis of online product reviews using DLMNN and future prediction of online product using IANFIS. Journal of Big Data, 7(1), 1-20.
[17] Onan, A. (2019, August). Deep learning based sentiment analysis on product reviews on Twitter. In International Conference on Big Data Innovations and Applications (pp. 80-91). Springer, Cham.
[18] Ferreira, R., Freitas, F., de Souza Cabral, L., Lins, R.D., Lima, R., França, G., Simskez, S.J., & Favaro, L. (2013, November). A four dimension graph model for automatic text summarization. In 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) (Vol. 1, pp. 389-396). IEEE.
[19] Kumar, A., Narapareddy, V.T., Srikanth, V.A., Malapati, A., & Neti, L.B.M. (2020). Sarcasm detection using multi-head attention based bidirectional LSTM. Ieee Access, 8, 6388-6397
[20] Ewees, A.A., Abd Elaziz, M., & Houssein, E.H. (2018). Improved grasshopper optimization algorithm using opposition-based learning. Expert Systems with Applications, 112, 156-172.
[21] Iqbal, A., Amin, R., Iqbal, J., Alroobaea, R., Binmahfoudh, A., & Hussain, M. (2022). Sentiment Analysis of Consumer Reviews Using Deep Learning. Sustainability, 14(17), 10844.