ERF-XGB: Ensemble Random Forest-Based XG Boost for Accurate Prediction and Classification of E-Commerce Product Review

4 authors, including:

Daniyal Alghazzawi
King Abdulaziz University
206 PUBLICATIONS 2,854 CITATIONS

SEE PROFILE

Sahar Badri
King Abdulaziz University
13 PUBLICATIONS 81 CITATIONS

SEE PROFILE

Syed Hamid Hasan
King Abdulaziz University
56 PUBLICATIONS 218 CITATIONS

SEE PROFILE

Article

ERF-XGB: Ensemble Random Forest-Based XG Boost for Accurate Prediction and Classification of E-Commerce Product Review

Daniyal M. Alghazzawi [ID], Anser Ghazal Ali Alquraishee, Sahar K. Badri [ID] and Syed Hamid Hasan *[ID]

Department of Information Systems, College of Computer Sciences and Information Technology, King Abdulaziz University, Jeddah 80200, Saudi Arabia; dghazzawi@kau.edu.sa (D.M.A.); agali@kau.edu.sa (A.G.A.A.); skbadri@kau.edu.sa (S.K.B.)
* Correspondence: shhasan@kau.edu.sa

Abstract: Recently, the concept of e-commerce product review evaluation has become a research topic of significant interest in sentiment analysis. The sentiment polarity estimation of product reviews is a great way to obtain a buyer's opinion on products. It offers significant advantages for online shopping customers to evaluate the service and product qualities of the purchased products. However, the issues related to polysemy, disambiguation, and word dimension mapping create prediction problems in analyzing online reviews. In order to address such issues and enhance the sentiment polarity classification, this paper proposes a new sentiment analysis model, the Ensemble Random Forest-based XG boost (ERF-XGB) approach, for the accurate binary classification of online e-commerce product review sentiments. Two different Internet Movie Database (IMDB) datasets and the Chinese Emotional Corpus (ChnSentiCorp) dataset are used for estimating online reviews. First, the datasets are preprocessed through tokenization, lemmatization, and stemming operations. The Harris hawk optimization (HHO) algorithm selects two datasets' corresponding features. Finally, the sentiments from online reviews are classified into positive and negative categories regarding the proposed ERF-XGB approach. Hyperparameter tuning is used to find the optimal parameter values that improve the performance of the proposed ERF-XGB algorithm. The performance of the proposed ERF-XGB approach is analyzed using evaluation indicators, namely accuracy, recall, precision, and F1-score, for different existing approaches. Compared with the existing method, the proposed ERF-XGB approach effectively predicts sentiments of online product reviews with an accuracy rate of about 98.7% for the ChnSentiCorp dataset and 98.2% for the IMDB dataset.

Keywords: product reviews; sentiments; ensemble random forest; XG boost; Harris hawk optimization algorithm; prediction

1. Introduction

Nowadays, online shopping is a popular worldwide practice. Around 2 billion people use online shopping sites to purchase daily products. Customers usually buy products based on previous reviews in order to find better products [1]. The modern life of an individual is highly comfortable due to the advanced development of e-commerce. People can gather the needed products online without walking outdoors. Food delivery platforms have 70% more customers compared to other online delivery platforms. Uber Eats, Eat, and Grubhub are some of the fast-developing companies, with total transaction amounts of up to USD 94 billion. Choosing products by reading through many reviews can be tiresome. After the products are delivered to the clients, they usually share their opinion, comments, and positive and negative reviews on the Internet site [2]. Online shopping accounted for 14.1% of purchases in 2019 and is estimated to reach 22% in 2023. The products sold on online platforms need physical delivery, except for digital products. Compared to other methods, physical delivery causes more pollution, and the products

delivered can be highly damaged. The delivery of satisfactory products will improve the reviews of the customer. The reviews provided by the customer help both the company and another customer. The reviews are obtained as a platform to overcome all the issues and make it easy to evaluate positive and negative reviews from the customers and the company [3]. The development of e-commerce is mainly focused on two factors: transaction volume prediction and e-commerce index construction. Utilizing the machine learning model results in excellent nonlinear mapping ability, a simple iterative process, and strong generalization ability. Sentiment analysis is the natural language processing used to sort emotions from opinions, text, and tweets, similar to text mining. The three concepts that comprise sentiment analysis are popularity, opinion, and the subject [4]. Delaying the delivery of the ordered product will create a bad review from an individual customer, and in business, this affects the customer's retention rate. Trackable delivery products are delayed due to a lag in the shipment service. Popularity is defined by the negative and positive reviews. Opinion represents personal opinions. The subject represents the starting point of an opinion about an object [5]. An increase in the number of e-commerce platforms has generated significant financial transactions and increased fraud among users. E-commerce transaction volume prediction models, such as machine learning and statistical regression, have been conceived. The XG boost model is used to predict the transaction volume of e-commerce to protect the nonlinear part of online delivery. This contextual mining type helps extract and identify the subjective data to understand their social sentiment. This type of usage of e-commerce sites also impacts businesses greatly [6]. In addition, sentiment analysis contains lexical-based and machine learning approaches. The automated system also uses machine learning techniques to develop the hybrid model for sentiment analysis [7]. The machine learning approaches represent customer behavior using the plot's accurate visual representation [8]. Thus, the visual representation details are gathered from the overall consumer behavior on the e-commerce platform [9]. In addition, customer reviews may enhance the hit ratio, increase customer visits, and increase spending time on e-commerce sites [10]. They can also apply the voice of customers for customer services and target marketing [11]. Internet technology is the mainstream vehicle for the development of online shopping. The objective of online shopping is to buy and consume things satisfactorily. The satisfaction of consumers mainly depends on the sentimental analysis (SA) of a sufficient number of user reviews. Yet, there are challenges in accepting the reviews due to text length, unfaithful logic, and sequence length. Online shopping has been growing exponentially in recent years, resulting in increased environmental impacts from packaging, transportation, and production. Moreover, socially responsible consumers are increasingly demanding information on the sustainability credentials of the products they buy. Hence, sentiment analysis of e-commerce product reviews plays a crucial role in informing sustainable consumer decisions. The proposed ERF-XGB approach provides an accurate and efficient method for predicting the sentiment polarity of product reviews, allowing customers to make informed decisions about the service and product qualities of the purchased products. By promoting informed purchasing decisions, the proposed approach may indirectly contribute to sustainability efforts by guiding consumers toward more sustainable choices. Moreover, the proposed approach has potential applications in analyzing sustainability-related reviews, such as reviews of sustainable or eco-friendly products. By accurately classifying sentiment polarity in these reviews, the proposed approach helps to identify areas for improvement and guide sustainability efforts. In this paper, a novel sentiment analysis approach is designed to predict the sentiments of product reviews more precisely. The main contributions of this paper are discussed as follows:

- A new ensemble random forest-based XG boost (ERF-XGB) approach is proposed for the accurate and effective prediction and classification of sentiments of online reviews into two categories: positive and negative.
- Selection of more relevant feature information from preprocessed datasets using the Harris hawk optimization (HHO) algorithm.

- Analyzing the proposed ERF-XGB approach performances in terms of evaluation indicators: accuracy, recall, precision, and F1-score.

The remaining sections of this article are organized in the following ways: Section 2 deliberates on some recent literary works, Section 3 illustrates the proposed sentiment analysis methodology designed for predicting the sentiments of online product reviews, Section 4 portrays the experimental results, and, finally, Section 5 concludes the paper with its future scope.

2. Literature Survey

Zhao et al. [12] discussed the sentiment analysis of e-commerce products using machine learning techniques. A machine learning algorithm called LSIBA-ENN (Local Search Improvised Bat Algorithm-based Elman Neural Network) technique was obtained, which examined the differences obtained from online products. The presentation of the established technique was compared with some techniques such as NB, ENN, and SVM. As a result, the accuracy obtained for the established method was 93.91%. On the other hand, it provided less accuracy when applied to data from other domains, such as social media. Xu et al. [13] developed sentiment analysis in e-commerce using the naïve Bayes learning framework. The continuous naïve Bayes learning (CNBL) framework was deployed on more than two e-commerce products, which analyzed the sentiment categorization. The naïve Bayes enlarged the framework procedure, which resulted in a stable learning method and kept the increased efficiency. As a result, the products from Amazon as well as the appraisal obtained from the presented film improved the learning skill of the current domain and increased its capacity. However, the obtained results were wrong in some cases and therefore difficult to analyze.

Kumar et al. [14] illustrated that sentiment analysis and EEG identified customer gratification in product reviews. The recorded details collected from customers were recovered and the ratings were calculated by using the NLP technique. EEG signals display the product details in real time on the computer screen. Based on the values displayed on the screen, the local ratings were calculated easily and their report was obtained. The total presentation was improved by merging the local and global ratings based on the Artificial ABC algorithm. As a result, the established ABC method minimized the RMSE value compared to the unique model, and the rating was obtained as 0.29. On the other hand, the accuracy of the optimal value could not be achieved correctly in some cases. Parimala et al. [15] reviewed tweets for risk assessment in sentiment analysis using deep learning. The established RASA (Risk Assessment Sentiment Analysis) technique was applied and it identified clue words from the network, while the LSTM network categorized the tweet and sentiment analysis in each location. The developed technique was supported by different methods such as XG boost, naïve Bayes algorithm, SVM, multi-class, and dual class. As a result, the explained RASA technique obtained the dual scheme with a better accuracy rate of 1% when compared with XG boost, and in a multi-class scheme, it was obtained as 30% when related to other methods. However, the unique network performance was not processed with this method, and it was only operated for English. Ramshankar et al. [16] illustrated a novel system of fuzzy aiding using Black Hole-based Grey Wolf Optimization. A novel technique called BH-GWO was established, which obtained a coherent approval system. The mass of the product was efficiently obtained by the BH-GWO method. Finally, the sentiment analysis was determined using the diverse machine learning algorithm dataset. As a result, the BH-GWO system obtained 11.7% higher accuracy compared with fuzzy, 28.3% higher compared with KNN, 20.2% higher compared with SVM, and 18.75% higher compared with a neural network. On the other hand, the convergence speed was reduced. Gu et al. [17] discussed sentiment analysis in a deep neural network with a variational information bottleneck. The technique MBGCV was applied, which diminished the issues obtained in the network and achieved a satisfactory accuracy rate. The MBGCV technique utilized more than two channels and combined them with various methods. The developed model helped the traders examine the reviews from the customers and helped

to improve the products. As a result, the established technique attained a better accuracy of 94%, and the obtained sentiment analysis performance was better as well. Meanwhile, only limited reviews were collected in sentiment analysis.

Munna et al. [18] proposed two deep learning NLP models: one is sentiment analysis, and the other is product review classification aimed at improving the quality and services. They used several evaluation matrices such as accuracy, precision, recall, and F1-score. The experimental results demonstrate a high accuracy: 0.84 and 0.69 for sentiment analysis and product review classification, respectively. Xu et al. [19] proposed a continuous naïve Bayes learning framework for product review sentiment classification of largescale and multi-domain e-commerce platforms. The standard machine learning algorithms for sentiment classification are typically trained based on a single task or single domain basis. However, reviews in e-commerce platforms come from a large number of different domains. Experimental results on the Amazon product and movie review sentiment datasets show that our model can use knowledge learned from past domains to guide learning in new domains and can handle continuously updated reviews from different domains.

Alzahrani et al. [20] proposed a framework to use an opinion on consumers' reviews to help businesses and organizations continually improve their market strategies and obtain an in-depth analysis of the consumers' opinions regarding their products and brands. The long short-term memory (LSTM) and deep learning convolutional neural network integrated with LSTM (CNN-LSTM) models were used. The LSTM and CNN-LSTM algorithms achieved 94% and 91% accuracy, respectively. The test result shows the deep learning techniques used here to provide optimal results for classifying customers' sentiments toward the products. Huang et al. [21] developed a sentiment analysis model ERNIE-BiLSTM-Att (EBLA) to solve dimension mapping, disambiguation of sentiment words, and polysemy of Chinese words. The Attention Mechanism (Att) is used to optimize the weight of the hidden layer. Finally, softmax is used as the output layer for sentiment classification. The proposed model achieves an accuracy of more than 0.87 when compared to the existing one. Zhang et al. [22] propose a model to discover the helpfulness of online product reviews. Product reviews can be analyzed and ranked by our scoring system, and the reviews that may better help consumers than others will be found first. The experimental results confirm that our approach outperforms or performs the same as other machine learning methods.

3. Proposed Methodology

This paper proposes a novel ERF-based XG boost method for SA of online product reviews. The work of the proposed SA consists of three phases: data processing, feature selection, and sentiment classification, which are portrayed in Figure 1. The proposed model is mainly adopted for binary classification, and it classifies the e-commerce reviews into positive and negative opinions.
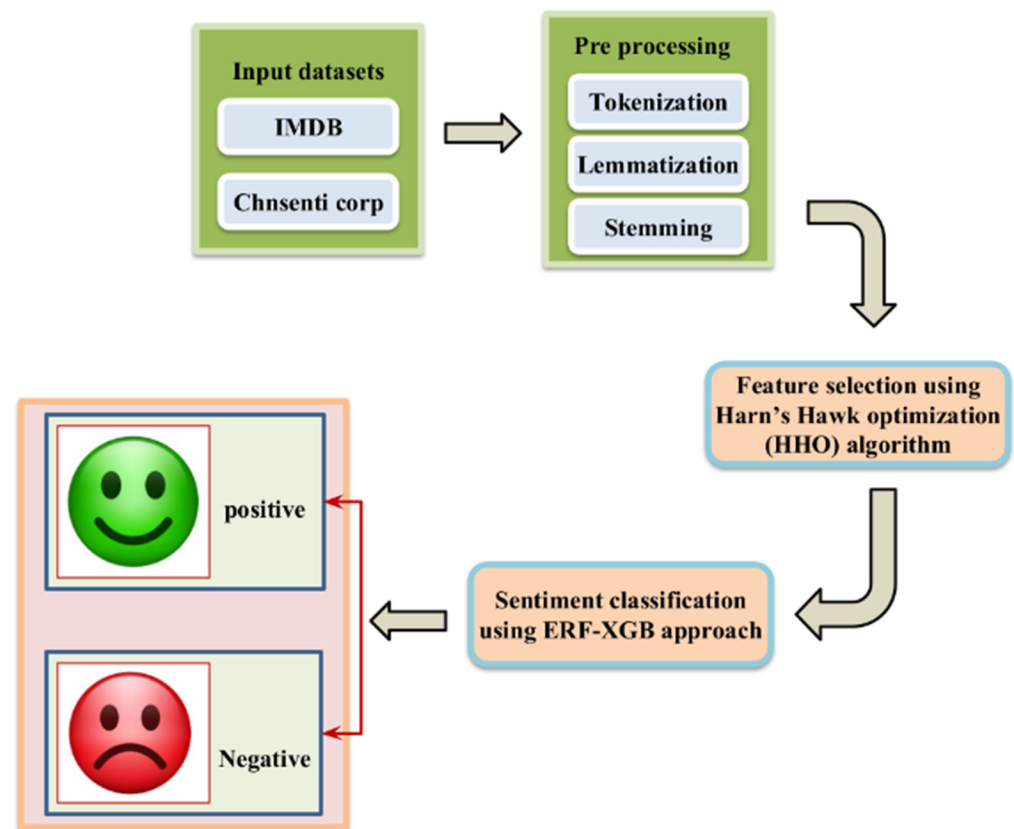
Figure 1. Proposed sentiment analysis architecture.

3.1. Data Preprocessing

The preprocessing approach is utilized for ignoring the undesirable elements from the set of databases. Three steps execute the preprocessing, and they are tokenization, Gensim lemmatization (GL), and snowball stemming (SBS).

3.1.1. Tokenization

This process encounters the whitespace character and breaks the input customer reviews into tokens or terms. The word sequences are analyzed to interpret their meaning [23].

3.1.2. Lemmatization

The groups of possible relations are determined by utilizing a morphological analysis process based on the multidimensional set. Then, it is utilized to solve problems in multidimension. Special characters, numeric removal, lower casing, and stand-alone punctuation are disparate operations in the Gensim package [24].

3.1.3. Stemming

During stemming, the smaller numbers of characters are neglected from the words by utilizing the stemming process. The lemmatization approach converts the words into a meaningful form without eliminating any characters [25].

3.2. Feature Selection

The insignificant features present in the data can decrease the accuracy of the design as well as make the model unable to studying the insignificant features. Then, the feature selection difficulties are developed as an optimization problem for detecting the informative features. The Harris hawk optimization (HHO) is the optimization technique for executing the feature selection, and it is described in the below section.

Harris Hawk optimization (HHO)

One of the population-based gradients-free optimizations is the HHO and it is utilized for solving optimization issues. The exploitation and exploration phases are inspired by the exploration of surprise pounce prey and various attacking methods of HH [26,27].

$$T^{\beta+1} = \begin{cases} T_\vartheta^\beta - d_1 \left| T_\vartheta^\beta - 2d_2 T^\beta \right| \\ (T_{RAB}^\beta - T_v^\beta) - d_3(P_r + d_4(u^B - l^B)) \end{cases} \tag{1}$$

From the above equation, $Z^{\beta+1}$ indicates the hawk's position, β denotes the next iteration value, the current position of hawks is represented by T_{RAB}^β, σ indicates the random assignment value, while u^B and l^B indicate the upper bound and lower bound, respectively. The Harris hawks detect and track the prey through their eyes, and the HH is the candidate solution.

$$X(q+1) = \begin{cases} X_\Re(q) - \gamma_1 |X_\Re(q) - 2\gamma_2 X(q)| & d \geq 0.5 \\ (X_{RAB}(q) - X_s(q) - \gamma_3(l^B + \gamma_4(u^B - l^B))) & d < 0.5 \end{cases} \tag{2}$$

From the above equation, the hawk's position vector is indicated by $X(q+1)$, the rabbit position is represented by $X_{RAB}(q)$, and the hawk's current position vector is represented by $X(q)$. The average positions of hawks are calculated, and this is expressed as

$$X_s(q) = \frac{1}{N} \sum_{c=1}^{N} X_c(q) \tag{3}$$

$$T^{\beta+1} = \Delta T^\beta - W \left| CT_{RAB}^\beta - T^\beta \right| \tag{4}$$

$$T^{\beta+1} = T_{RAB}^\beta - W \left| \Delta T^\beta \right| \tag{5}$$

$$T^{\beta+1} = \begin{cases} F & if \quad C(F) < C(T^\beta) \\ X & if \quad C(F) \geq C(T^\beta) \end{cases} \tag{6}$$

$$F = T_{RAB}^\beta - W \left| CT_{RAB}^\beta - T_N^\beta \right| \tag{7}$$

The values of s and v are assigned to 0 and 1. The above equation Γ_F represents the levy flight and the T_N^β value is determined in the exploration stage.

3.3. Sentiment Classification

This section illustrates the proposed ERF-based XGB approach for the efficient prediction and classification of online product reviews.

3.3.1. Ensemble Random Forest (ERF)

ERF works by generating multiple decision trees during regression and classification. The final results of the classification are intent on decision tree voting [28]. The random forest algorithm has no sensitivity for hyperparameter settings; therefore, with some small adjustments, it can be utilized for establishing an appropriate model. In the classification issues of the clustering and regressive analyses, the RF algorithm achieved better performance in solving the problem. In an ensemble random forest, every decision tree is generated individually with various bootstrap samples.

Steps for generating ERF

1. Bagging: The random forest algorithm selects random samples to extract $^2/_3$ from the initial training dataset $t = \{(X_1, Y_1), (X_2, Y_2)(X_m, Y_m)\}$ for establishing a training subset. For generating m decision trees, the bagging algorithm received m bootstrap sample sets. Unextracted data are called out-of-bad data (OOB). The calculation of

the OOB is more capable compared with cross-validation. OOB contains two-phase to extract the RF feature importance, such as the Gini index and OOB error rate. In the OOB error rate, the phase calculates the decision tree and the Gini index phase calculates the failed classification.

$$Gi_m = \sum_{C=1}^{C} P_C(1 - P_C) = 1 - \sum_{C=1}^{C} P_c^2 \tag{8}$$

The above equation m represents the total number of nodes, P_C indicates the class probability, and C represents the total number of classes.

The feature importance X_j scenario is represented as follows:

$$Vm_{JM}^{GINI} = Gi_M - Gi_L - Gi_R \tag{9}$$

From the above equation, after splitting, the right and left nodes of the Gini index are represented by Gi_L, Gi_R.

2. Decision tree construction: From m bootstrap sample sets, m classification trees are generated. The feature vector samples are represented by m. In m feature vectors, the original decision tree chooses the optimal feature vector. The results of the classification m can finally be received from m decision tree models.

$$y = \{Y_1(X), Y_2(X), \dots Y_m(X)\} \tag{10}$$

From the above equation, the classification model y with m the decision tree models are determined. The result of the final classification is determined by the multiple classification models.

$$t(X) = ArgMax \sum_{I-1}^{m} Z(Y_I(X)) = A \tag{11}$$

From the above equation, the fault class block is represented by A, the multiple classification models, and is denoted by $t(X)$, while the distinct decision tree model classification is represented by $Y_I(X)$. If the indicator function $Z(.)$ value is 1, then the two values are equal. Otherwise, the indicator function value is zero. To enhance the decision-making performance of the ensemble random forest (ERF) algorithm, the XG boost approach is integrated and hence the proposed ERF-XGB approach achieves the efficient detection of sentiments.

3.3.2. XG Boost (XGB) Algorithm

XG-boost is a widespread end-to-end and scalable tree-boosting model. It has been employed and optimized mostly in research, and it is the enhanced structure of the gradient boosting regression model (GBRT) [29–31]. GBRT contains a sequence of fundamental regression trees by way of the sequential technique and accommodates multiple trees to enlarge model ability. For this, the expression of the final prediction is formulated below.

$$\overset{\wedge^n}{x} = \overset{\wedge^{n-1}}{x} + \beta\, g_n(Y, \phi_n) = \beta \sum_{i=1}^{n} g_i(Y, \phi_i) \tag{12}$$

The above equation ϕ_i represents the parameter restraining tree structure, n represents the number of regression trees, $g_i(Y, \phi_i)$ represents i as the regression trees' output based on ϕ_ith the structure Y represents the predictor, and β represents the shrinkage factor, wherein Y predictors $x - \overset{\wedge^{n-1}}{x}$ are inputs. The main goal of the gradient boosting regression is

identifying the optimal φ_i constructing $g_i(Y, \phi_i)$ ith step to decrease the function of the objective as expressed below.

$$M = \sum_j m\left(\hat{x}_j, x_j\right) = \sum_j m\left[\hat{x}_j^{i-1} + \beta g_i(Y_j, \phi_i), x_j\right] \tag{13}$$

The above equation m represents a loss function that utilizes squared error between ground truth x and predictive value $\hat{x}$.

Figure 2 represents the flow diagram of the proposed ERF-XGB approach below. The input dataset is selected randomly and the ensemble RF and XGB parameters are initialized for bootstrap sampling. An RF- classifier is constructed and trained for each instance using majority voting. The capability of the model is improved using more instances via XG boost. If the minimum loss value is not obtained, the model is trained again until the loss value is minimized. The ERF-XGB model predicts the polarity of the input as negative or positive.
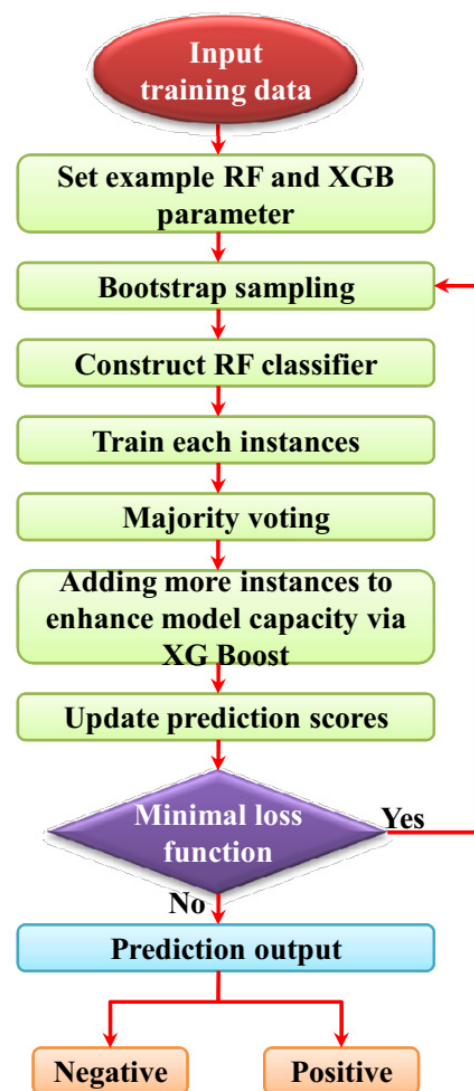

Figure 2. Flow diagram of proposed ERF-XGB approach.

Let us rewrite Equation (13) into the below equation.

$$M = \sum_j m\left(\hat{x}_j, x_j\right) + \sum_i \varphi\left(\phi_i\right) = \sum_j m\left[\hat{x}_j^{i-1} + \beta g_i(Y_j, \phi_i), x_j\right] + \sum_i \varphi\left(\phi_i\right) \tag{14}$$

The above equation $\varphi\left(\phi_i\right)$ represents the regularization item on the ith regression tree for the prevention of overfitting.

$$\varphi(\phi_i) = \delta\, S_i + \frac{1}{2}v\,\|u_l\| = \delta\, S_i + \frac{1}{2}v\sum_{l=1}^{S_i}\left[u_l^{(i)}\right]^2 \tag{15}$$

The above equation S_i represents the number of leaves in ith the tree, δ represents minimum loss reduction, v denotes regularization term on the leaves' weight, and $u_l^{(i)}$ represents the ith and lth regression tree. It is evident that δ penalizes S_i for decreasing the objective function. Figure 2 describes the flow diagram of the proposed ERF-XGB approach.

4. Experimental Results and Discussions

The ensemble random forest-based extreme gradient boosting (ERF-XGB) algorithm is proposed for the sentimental analysis of e-commerce product reviews. The experimental results are briefly explained in upcoming sections.

4.1. Experimental Setup

The experiments of this proposed ERF-XGB algorithm were conducted on the MATLAB 2019b platform along with the TensorFlow framework at PyCharm IDE and Python 3.5.

4.2. Dataset Description

Two types of datasets are used to effectively analyze the sentimental analysis of the proposed ERF-XGB algorithm. One of the datasets is the Internet Movie Database (IMDB) [32] dataset which has 25,000 tweets along with the polarity of 12,500 negative movie reviews and 12,500 positive movie reviews. The second dataset is the Chinese Emotional Corpus (ChnSentiCorp) dataset (https://github.com/hidadeng/cnsenti) which includes abundant Sentiment Corpus such as ChnSentiCorpMov and ChnSentiCorpHtl.The datasets are split into two types testing and training phases. Where the training phase contains 70% of the data, and the training phase contains 30% of the data. We provide a set of 25,000 highly polar movie reviews for training and 25,000 for testing. So, predict the number of positive and negative reviews using either classification or deep learning algorithms.

4.3. Performance Measures

The performance evaluation was performed by using performance metrics such as the accuracy ($C_{accuracy}$), precision ($C_{precision}$), recall (C_{recall}), and F1-score ($C_{F1-score}$), which are evaluated in terms of true positive ($C_{true\ positive}$), true negative ($C_{true\ negative}$), false positive ($C_{false\ positive}$), and false negative ($C_{false\ negative}$).

$$C_{accuracy} = \frac{C_{true\ positive} + C_{true\ negative}}{C_{true\ positive} + C_{true\ negative} + C_{false\ positive} + C_{false\ negative}} \tag{16}$$

$$C_{precision} = \frac{C_{true\ positive}}{C_{true\ positive} + C_{false\ positive}} \tag{17}$$

$$C_{recall} = \frac{C_{true\ positive}}{C_{true\ positive} + C_{false\ negative}} \tag{18}$$

$$C_{F1-score} = 2 * \frac{C_{precision} * C_{recall}}{C_{precision} + C_{recall}} \tag{19}$$

4.4. Hyperparameter Configuration

Hyperparameter tuning was used to find the optimal parameter values that improve the performance of the proposed ERF-XGB algorithm. Table 1 presents the optimized XG boost hyperparameter values.

Table 1. Optimized XGB parameters using the ERF algorithm.

Parameters	Value
Base Learner	Tree
Gamma	0
Learning rate	0.03
Number of pruning after control	0.2
Regularization	L2
Tree depth	4
Random sampling decision tree ratio	0.7
Minimum leaf node sample weight	2

4.5. Performance Analysis

Performance analysis of the ERF-XGB algorithm was conducted with respect to the metrics of accuracy, precision, recall, and F1-score. The overall performance analysis of the proposed ERF-XGB algorithm was compared with the standard XG boost, and the results are tabulated in Table 2.

Table 2. Overall performance analysis of the proposed ERF-XGB algorithm.

Sl. No	Performance Measures	Performance Ranges			
		Proposed ERF-XGB		XG Boost [24]	
		IMDB Dataset	ChnSenti Corp Dataset	IMDB Dataset	ChnSenti Corp Dataset
1	Accuracy	98.2%	98.7%	90.1%	90.7%
2	Precision	98.5%	98%	89.3%	88.6%
3	Recall	98.8%	98.3%	89.4%	89.3%
4	F1-score	98.1%	97.1%	89.0%	88.1%

4.6. Comparative Analysis

For comparative analysis, the methods namely Local Search Improvised Bat Algorithm based Elman Neural Network (LSIBA-ENN) model, Continuous Naïve Bayes Learning (CNBL) framework, Hybrid Black Hole based Grey Wolf Optimization (BH-GWO), Sentiment Lexicon Convolutional Neural Network with Attention-based Bidirectional Gated Recurrent Unit (SLCABG) model and proposed ERF-XGB algorithm are used. Figure 3a–d represents the Comparative analysis of accuracy, precision, recall, and F1-score for the ChnSentiCorp dataset. The comparative analysis is performed by using different methods such as LSIBA-ENN, CNBL, BH-GWO, SLCABG, and the proposed ERF-XGB algorithm. From this analysis, the proposed ERF-XGB algorithm achieved good performance for the sentimental analysis from e-commerce product reviews. Figure 3a denotes the accuracy analysis which depicts the better performance of the proposed ERF-XGB algorithm compared to other state-of-the-art methods. The accuracy rate of 78%, 86%, 83%, 90%, and 98.2% are achieved from the various methods like LSIBA-ENN, CNBL, BH-GWO, SLCABG, and proposed ERF-XGB algorithm respectively. Figure 3b shows the precision analysis that denotes the higher precision rate of 98.5% from the proposed ERF-XGB algorithm. The Methods such as LSIBA-ENN, CNBL, BH-GWO, and SLCABG attained a precision rate of 89%, 83%, 86%, and 90% respectively. Figure 3c portrays the recall rate analysis that

implies the high performance of the proposed ERF-XGB algorithm. The recall rate of 79%, 85%, 83%, 88%, and 98.8% are obtained from LSIBA-ENN, CNBL, BH-GWO, SLCABG, and the proposed ERF-XGB algorithm. Figure 3d shows the F1-score analysis and the methods such as LSIBA-ENN, CNBL, BH-GWO, SLCABG, and the proposed ERF-XGB algorithm provide the F1-score of 83%, 81%, 88%, and 92% and 98.1% respectively.
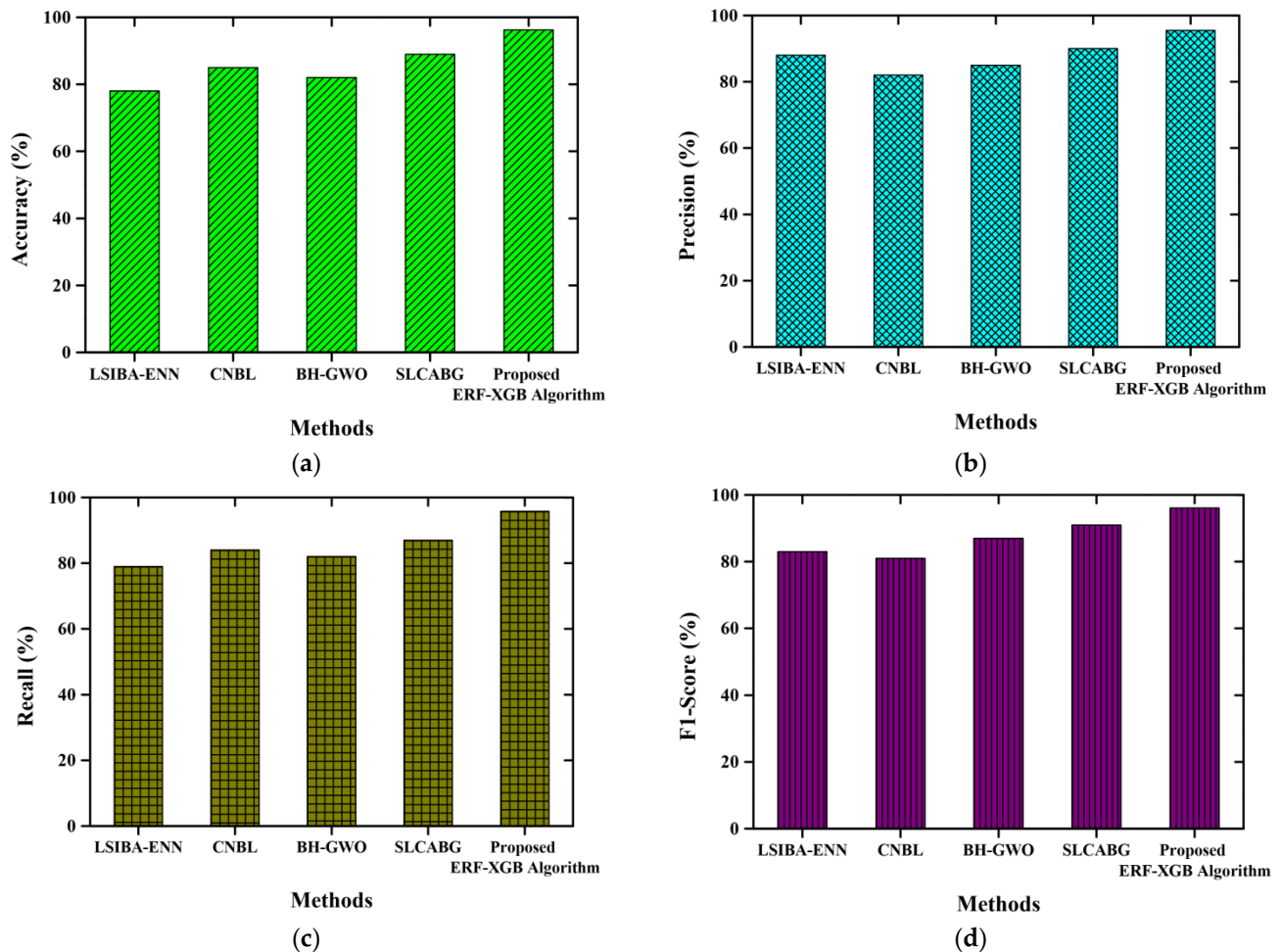

Figure 3. Comparative analysis using the ChnSentiCorp dataset (**a**) accuracy (**b**) precision (**c**) recall (**d**) F1-score.

Figure 4a–d depicts the comparative analysis of accuracy, precision, recall, and F1-score for the IMDB dataset. The comparative analysis is performed by using different methods such as LSIBA-ENN, CNBL, BH-GWO, SLCABG, and the proposed ERF-XGB algorithm. The accuracy analysis is performed in Figure 4a and the proposed ERF-XGB algorithm achieved a higher accuracy rate of 98.7%. Figure 4b represents the precision analysis of different methods such as LSIBA-ENN, CNBL, BH-GWO, SLCABG, and the proposed ERF-XGB algorithm which gives the precision rate of 79%, 87%, 83%, 90%, and 98% respectively. The recall rate analysis is depicted in Figure 4c which shows the best performance of the proposed ERF-XGB algorithm. The recall rate of 83%, 78%, 81%, 88%, and 98.3% is obtained from the LSIBA-ENN model, CNBL framework, BH-GWO, SLCABG model, and proposed ERF-XGB algorithm respectively. Figure 4d portrays the F1-score analysis and the proposed ERF-XGB algorithm achieved a higher performance of 98.1% compared to other state-of-the-art methods. Figure 5 shows the performance analysis of accuracy for testing and validation for different iterations. The smoothing accuracy is obtained by applying a smoothing algorithm and the training accuracy is obtained for each

mini-batch. The training is stopped when the network reaches a plateau and when there is no improvement noted in the accuracy.
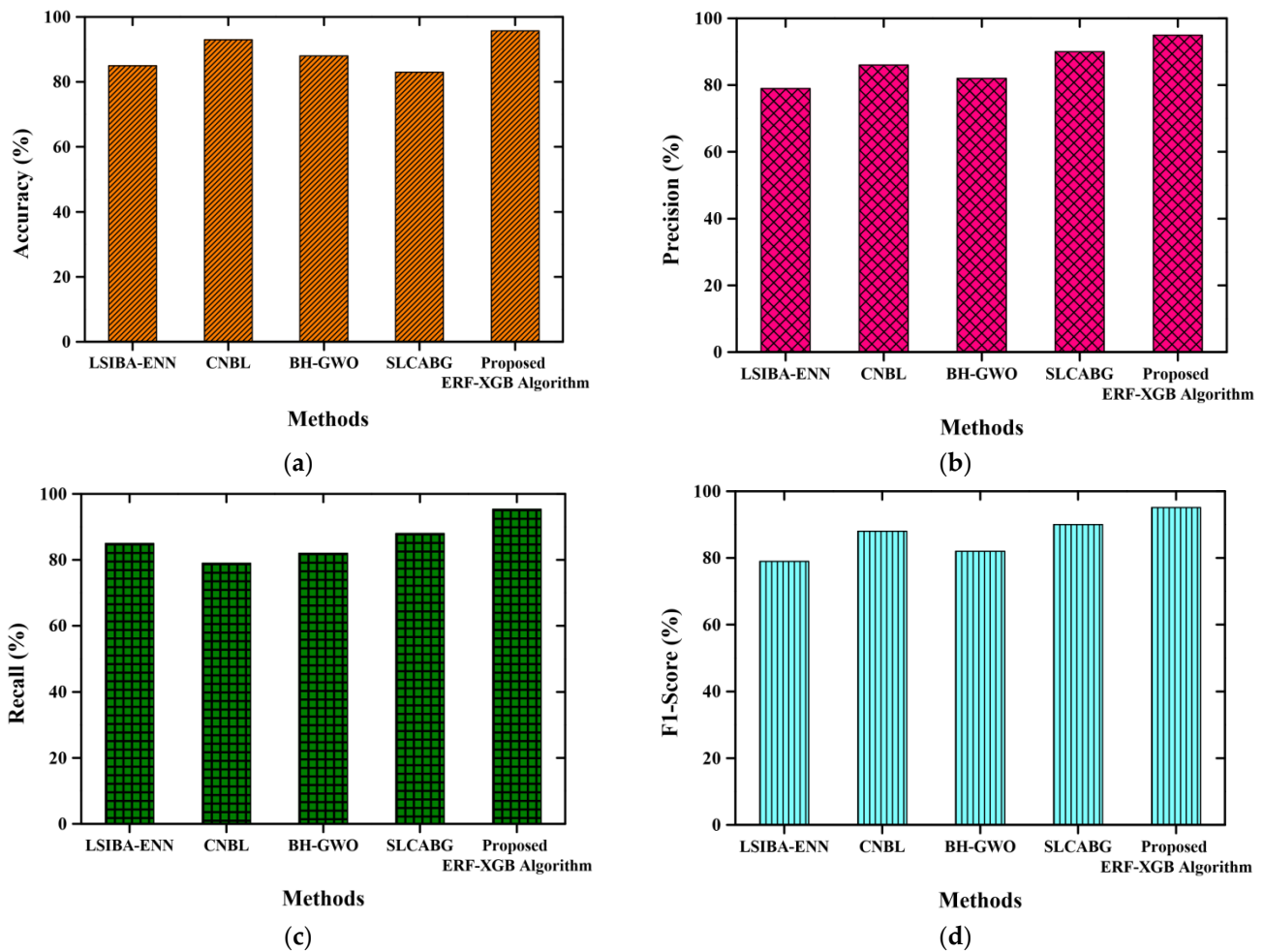

Figure 4. Comparative analysis of IMDB dataset. (**a**) accuracy (**b**) precision (**c**) recall (**d**) F1-score.
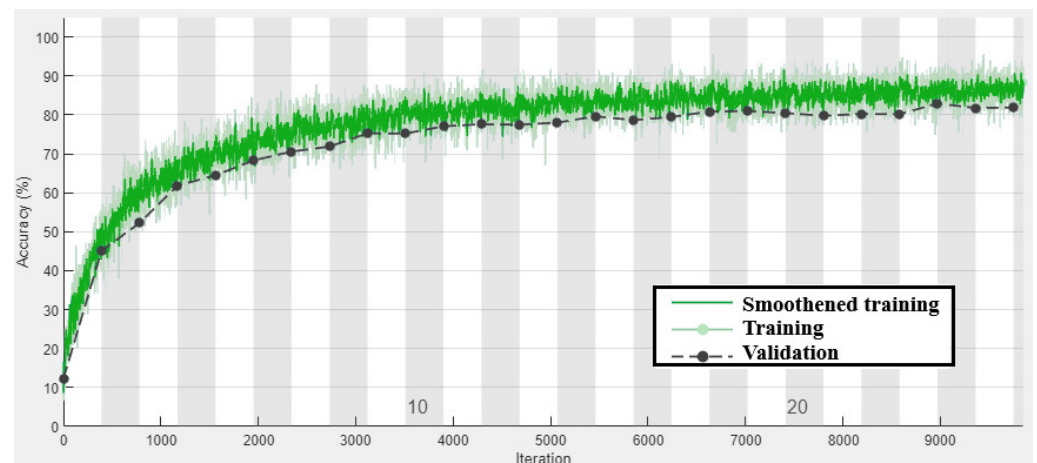

Figure 5. Performance analysis of accuracy for testing and validation for different iterations.

5. Conclusions

In this paper, the ERF-XGB algorithm was proposed for the sentimental analysis of e-commerce product reviews. Two types of datasets were used to effectively analyze the

sentimental analysis of the proposed ERF-XGB algorithm: the Internet Movie Database (IMDB) dataset and the Chinese Emotional Corpus (ChnSentiCorp) dataset. Hyperparameter tuning was used to find the optimal parameter values that improved the performance of the proposed ERF-XGB algorithm. For comparative analysis, the LSIBA-ENN model, CNBL framework, BH-GWO, SLCABG model, and proposed ERF-XGB algorithm were used. In the ChnSentiCorp dataset, the proposed ERF-XGB algorithm provided an accuracy of 98.2%, a precision of 98.5%, a recall of 98.8%, and an F1-score of 98.1%. In the IMDB dataset, the proposed ERF-XGB algorithm provided an accuracy of 98.7%, a precision of 98%, a recall of 98.3%, and an F1-score of 97.1%. The experimental results show that the proposed ERF-XGB algorithm attained better performance compared to the other state-of-the-art methods. The proposed ERF-XGB approach is effective in predicting the sentiments of online product reviews, with accuracy rates of about 98.7% for the ChnSentiCorp dataset and 96.2% for the IMDB dataset. In the future, the proposed ERF-XGB algorithm will be employed for processing and extracting complex semantic and syntactic features.

Author Contributions: Methodology, S.H.H.; Software, A.G.A.A.; Validation, S.K.B.; Resources, D.M.A. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

References

1. Munna, M.H.; Rifat, M.R.I.; Badrudduza, A.S.M. Sentiment analysis and product review classification in e-commerce platform. In Proceedings of the 2020 23rd International Conference on Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 19–21 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
2. Diwakar, D.; Kumar, R.; Gour, B.; Khan, A.U. Proposed machine learning classifier algorithm for sentiment analysis. In Proceedings of the 2019 Sixteenth International Conference on Wireless and Optical Communication Networks (WOCN), Bhopal, India, 19–21 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
3. Noor, A.; Islam, M. Sentiment Analysis for Women's E-commerce Reviews using Machine Learning Algorithms. In Proceedings of the 2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kanpur, India, 6–8 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6.
4. Singh, S.N.; Sarraf, T. Sentiment analysis of a product based on user reviews using random forests algorithm. In Proceedings of the 2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence), Noida, India, 29–31 January 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 112–116.
5. Yi, S.; Liu, X. Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review. *Complex Intell. Syst.* **2020**, *6*, 621–634. [CrossRef]
6. Hossain, M.S.; Rahman, M.F.; Uddin, M.K.; Hossain, M.K. Customer sentiment analysis and prediction of halal restaurants using machine learning approaches. *J. Islam. Mark.* 2022, *ahead-of-print*. [CrossRef]
7. Karn, A.L.; Karna, R.K.; Kondamudi, B.R.; Bagale, G.; Pustokhin, D.A.; Pustokhina, I.V.; Sengan, S. Customer centric hybrid recommendation system for E-Commerce applications by integrating hybrid sentiment analysis. *Electron. Commer. Res.* **2022**, *23*, 279–314. [CrossRef]
8. Shrirame, V.; Sabade, J.; Soneta, H.; Vijayalakshmi, M. Consumer Behavior Analytics using Machine Learning Algorithms. In Proceedings of the 2020 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2–4 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
9. Mehraliyev, F.; Chan, I.C.C.; Kirilenko, A.P. Sentiment analysis in hospitality and tourism: A thematic and methodological review. *Int. J. Contemp. Hosp. Manag.* **2022**, *34*, 46–77. [CrossRef]
10. Li, H.; Bruce, X.B.; Li, G.; Gao, H. Restaurant survival prediction using customer-generated content: An aspect-based sentiment analysis of online reviews. *Tour. Manag.* **2023**, *96*, 104707. [CrossRef]
11. Verma, P.; Dumka, A.; Bhardwaj, A.; Ashok, A. Product Review-Based Customer Sentiment Analysis Using an Ensemble of mRMR and Forest Optimization Algorithm (FOA). *Int. J. Appl. Metaheuristic Comput.* **2022**, *13*, 1–21. [CrossRef]
12. Zhao, H.; Liu, Z.; Yao, X.; Yang, Q. A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach. *Inf. Process. Manag.* **2021**, *58*, 102656. [CrossRef]

13.   Xu, F.; Pan, Z.; Xia, R. E-commerce product review sentiment classification based on a naïve Bayes continuous learning framework. *Inf. Process. Manag.* **2020**, *57*, 102221. [CrossRef]

14.   Kumar, S.; Yadava, M.; Roy, P.P. Fusion of EEG response and sentiment analysis of products review to predict customer satisfaction. *Inf. Fusion* **2019**, *52*, 41–52. [CrossRef]

15.   Parimala, M.; Swarna Priya, R.M.; Praveen Kumar Reddy, M.; Lal Chowdhary, C.; Kumar Poluru, R.; Khan, S. Spatiotemporal-based sentiment analysis on tweets for risk assessment of event using deep learning approach. *Softw. Pract. Exp.* **2021**, *51*, 550–570. [CrossRef]

16.   Ramshankar, N.; Joe Prathap, P.M. A novel recommendation system enabled by adaptive fuzzy aided sentiment classification for E-commerce sector using black hole-based grey wolf optimization. *Sādhanā* **2021**, *46*, 125. [CrossRef]

17.   Gu, T.; Xu, G.; Luo, J. Sentiment analysis via deep multichannel neural networks with variational information bottleneck. *IEEE Access* **2020**, *8*, 121014–121021. [CrossRef]

18.   Yang, L.; Li, Y.; Wang, J.; Sherratt, R.S. Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. *IEEE Access* **2020**, *8*, 23522–23530. [CrossRef]

19.   Zhao, W. Classification of Customer Reviews on E-commerce Platforms Based on Naive Bayesian Algorithm and Support Vector Machine. *J. Phys. Conf. Ser.* **2020**, *1678*, 012081. [CrossRef]

20.   Alzahrani, M.E.; Aldhyani, T.H.; Alsubari, S.N.; Althobaiti, M.M.; Fahad, A. Developing an intelligent system with deep learning algorithms for sentiment analysis of e-commerce product reviews. *Comput. Intell. Neurosci.* **2022**, *2022*, 3840071. [CrossRef]

21.   Huang, W.; Lin, M.; Wang, Y. Sentiment Analysis of Chinese E-Commerce Product Reviews Using ERNIE Word Embedding and Attention Mechanism. *Appl. Sci.* **2022**, *12*, 7182. [CrossRef]

22.   Zhang, R.; Tran, T.T. Helping e-commerce consumers make good purchase decisions: A user reviews-based approach. In *E-Technologies: Innovation in an Open World, Proceedings of the 4th International Conference, MCETECH 2009, Ottawa, Canada, 4–6 May 2009*; Springer: Berlin/Heidelberg, Germany, 2009; pp. 1–11.

23.   Garg, N.; Sharma, K. Text pre-processing of multilingual for sentiment analysis based on social network data. *Int. J. Electr. Comput. Eng.* **2022**, *12*, 776–784. [CrossRef]

24.   Kolajo, T.; Daramola, O.; Adebiyi, A.; Seth, A. A framework for pre-processing of social media feeds based on integrated local knowledge base. *Inf. Process. Manag.* **2020**, *57*, 102348. [CrossRef]

25.   Nafis, N.S.M.; Awang, S. The impact of pre-processing and feature selection on text classification. In *Advances in Electronics Engineering, Proceedings of the ICCEE 2019, Kuala Lumpur, Malaysia, 29–30 April 2019*; Springer: Singapore, 2020; pp. 269–280.

26.   Heidari, A.A.; Mirjalili, S.; Faris, H.; Aljarah, I.; Mafarja, M.; Chen, H. Harris hawks optimization: Algorithm and applications. *Future Gener. Comput. Syst.* **2019**, *97*, 849–872. [CrossRef]

27.   Alabool, H.M.; Alarabiat, D.; Abualigah, L.; Heidari, A.A. Harris hawks optimization: A comprehensive review of recent variants and applications. *Neural Comput. Appl.* **2021**, *33*, 8939–8980. [CrossRef]

28.   Zhou, X.; Xu, X.; Zhang, J.; Wang, L.; Wang, D.; Zhang, P. Fault diagnosis of silage harvester based on a modified random forest. In *Information Processing in Agriculture*; Elsevier: Amsterdam, The Netherlands, 2022.

29.   Zhu, X.; Chu, J.; Wang, K.; Wu, S.; Yan, W.; Chiam, K. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. *J. Rock Mech. Geotech. Eng.* **2021**, *13*, 1231–1245. [CrossRef]

30.   Luan, Y.; Lin, S. Research on text classification based on CNN and LSTM. In Proceedings of the 2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 29–31 March 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 352–355.

31.   Maas, A.; Daly, R.E.; Pham, P.T.; Huang, D.; Ng, A.Y.; Potts, C. Learning word vectors for sentiment analysis. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Portland, OR, USA, 19–24 June 2011; pp. 142–150.

32.   Maas, A. *Large Movie Review Dataset*. Sentiment Analysis. Available online: https://ai.stanford.edu/~amaas/data/sentiment (accessed on 6 March 2023).