Survey and Evaluation of Extreme Learning Machine on TF-IDF Feature for Sentiment Analysis

Manpreet Kaur, Dibyasundar Das and Smita Prava Mishra

December 5, 2022

Survey and Evaluation of Extreme Learning Machine on TF-IDF Feature for Sentiment Analysis

Manpreet Kaur
Department of CSE,
Faculty of Engineering (ITER), SOA University, BBSR Odisha, *India.*
mkaur1771@gmail.com

Dibyasundar Das
Department of CSE,
Faculty of Engineering (ITER), SOA University, BBSR Odisha, *India.*
dibyasundardas@soa.ac.in

Smita Prava Mishra
Department of CSE,
Faculty of Engineering (ITER), SOA University, BBSR Odisha, *India*
smitamishra@soa.ac.in

***Abstract*— Sentiment analysis is a tool to understand the emotion of the statement given by customers. This understanding helps the service provider to in improving the quality of service. Machine learning models are one of the popular choices for designing sentiment analysis systems. However, hyper-parameter tuning is one of the important concerns in most of these models. Moreover, the gradient-based training models are prone to the local-minima problem. In such a case one-pass learning model like Extreme Learning Machine (ELM), gives generalization performance with minimal hyper-parameter tunning. This work studies in depth the ability of ELM to learn a generalization model for the sentiment analysis problem. Here the study uses the airline Twitter review dataset to empirically analyze the ELM model and the required hyper-parameter setting.**

Keywords—TF-IDF, ELM, ML, SA

I. Introduction

Opinion Mining is the method involved in examining online work of writing to decide the profound tone they convey, whether they're positive, neutral, or negative. It takes a gander at the emotion communicated in a text. It is generally used to break down client feedback, overview reactions, and item audits. Further developing deals and holding clients are the center business objectives. It includes predicting or examining the hidden data present in the message. This hidden data is extremely valuable to get bits of knowledge of clients' preferences [32]. These bits of knowledge could uncover how to acquire a strategic advantage. It could likewise be applied to advertise reports and business diaries to pinpoint new open doors.

Twitter is an internet-based data and interpersonal relations website where individuals impart short messages also called a micro-blogging site [33]. Twitter is a mix of texting, contributing to a blog, and messaging, yet with compact substance and an expansive crowd. Likewise, Twitter Sentiment Analysis is used to analyze whether the text message is positive, neutral, or negative. Twitter Sentiment Analysis additionally is utilized for checking and examining social peculiarities, anticipating possibly hazardous circumstances, and deciding the overall state of mind of the people [1].

With the improvement in Deep Learning and Machine learning, the capacity to anticipate the sentiment inside a text has increased. It can be utilized to predict the sentiment as well as other abstract data inside a piece of text and because of these algorithms, it is possible to get the accurate sentiment of the text. Some traditional models like Naïve Bayes, Decision trees, Random Forest, Linear Regression, *k*-nearest neighbor, Logistic Regression, and Support Vector Machine (SVM) are used for prediction. But they have some limitations in the learning process.

Unlike these traditional algorithms, Extreme Learning Machine (ELM) is a one-pass learning algorithm for single hidden layer feedforward neural networks (SLFNs) [2]. ELM was submitted by Qin-Yu and Guang-Bin [16], which was concentrated to prepare single-hidden layer feedforward networks. ELM joins a lot quicker than conventional techniques since it learns without a cycle. ELM will in general flex better classification performance with few optimization constrain [3]. It considers a greater capacity to achieve global optimum. It works with different hidden nodes, including kernels [4]. Throughout the last 10 years, broad exploration of ELM has been done for three reasons: higher grouping precision, less manual intervention, and minimum training time. Liu and Wang [23] found that an Extreme learning machine might experience the agonize of overfitting as it is prepared by restricting shortcomings on the training set. Thus, they proposed septet ELM by presenting ensemble and cross-validation learning. Lin and Cao [24] begin voting mechanisms to strengthen the robustness of Extreme machine learning. They trained many independent ELMs with similar construction and the last characterization result was gotten by deciding on the consequences of these ELMs. Zhang and Shen [25] come up with their plan for contrast binary classification. They utilized a few tests as keys to training an Extreme learning machine autoencoder. Yao and Wu [28] employed dolphin swarm optimization to train the variables in ELM. Huang and Liu [26] utilized ELM to lessen features of multi-scale data by spectral regression where the proposed two-stage ELM merges quicker in case of high feature information.

II. Related Work

In this segment, a brief outline of the correlated projects on ELM with other traditional algorithms is presented. Ding and Huang [5] introduced a specific examination and correlation of the SVM and ELM. ELM and SVM are correspondents, in classification problems but ELM has fewer optimization constraints that are used to get better execution. Zhou and Huang [6] compared ELM with proximal Support vector machine (PSVM) and least square SVM (LS-SVM) on wide kinds of data sets. They saw that the ELM can be viewed as a bound structure of LS-SVM and PSVM. The execution of ELM in case of binary class problems was homogenous to PSVM and LS_SVM but, ELM outperformed. Zhang and Zhang 2016 [7] evaluate to get the foremost classifier for deep features which was taken out by CNN. They asserted that ELM indicates clear prevalence over Support vector machine in cross space acknowledgment undertakings and kernel elm accomplished state-of-the-art.

Merchant and Bhat [8] proposed their methodology for speculating the melting points of organic compounds based on ELM. They contrasted ELM based approach and other ML implementations and observed that ELM was better concurring than their investigations. Cheng and Liang [9] presented their improved model based on Extreme machine learning and stated it was better than some other popular neural networks. Hassan [10] utilized different classification strategies in the field of automated sleep apnea identification. As indicated by the end, ELM was better than different techniques in building a mechanized sleep apnea detection system. Dong et al. [11] to anticipate the tentative hourly electrical energy yield in Combined Cycle Power Plant proposed the Bagging-ELM network and compared it with other models. Their result shows that Bagging-ELM obtains higher accuracy and has a quicker learning speed with stability. Zhang and Zheng [12] took the Chinese dataset and compared it with SVM and ELM with the kernel model. Seeing the experiment result it is clear that ELM with kernel has high accuracy as compared to SVM, also the training and testing time is far less than SVM. Sonmez and Tuncer [13] perform ELM-based classification considering 30 different features including Phishing Websites facts in the UC Irvine ML data-set. They concluded the ELM has higher accomplishment contrasted with other classifier (SVM and NB) methods.

Kusrini et al. [14] purpose were to foresee understudies exiting the Education Management Doctoral Program by utilizing the factors of gender, working status, semester GPA 3, family status, and age. Because of the assessment utilizing the confusion matrix, Support Vector Machine has a precision pace of 63% while, elm has an exactness of 72%. Yusefi et al. [15] purpose are to decrease the mortality and early determination required for significant treatment for breast cancer. For that used ELM method and the discoveries are contrasted with those of other ML techniques. In this review, the extreme learning machine technique accomplished a training exactness of 96.403% and a test exactness of 99.248%. It was found that this analysis performed better compared to other ML models on the WBCD data-set while utilizing the ELM technique. MENG et al. [30] suggested a momentary wind speed foresee method centered on the extreme learning machine, that accomplished a quicker execution as compared to the orthodox practices. Olatunji [31] try researching how ELM and SVM analyzed the unique and significant issue of Email spam recognition, which is a classification problem. Guo-Jian Cheng, Hua-Xian Pan and Lei Cai [17] purpose are to access the feasibility and higher level of ELM for reservoir permeability prediction collated to SVM. For testing the standard of SVM they used correlation coefficient and RMSE (Root Mean Squared Error). The result shows that the ELM is more predominant than Support vector machine in learning speed for hidden samples. ELM can reserve a lot of training time. Hongming et al. [18] show that both PSVM and LS_SVM can be simplified further to unified with ELM can be built and ELM will in general have better versatility and accomplish comparable or much better speculation execution at a lot quicker training speed than conventional SVM. By comparing with the simulation results, ELM carry out better generalization presentation for classification.

Rajasekar Venkatesan and Meng Joo Er [19] proposed ELM based methods for multi-label classification problems. They differentiate with others existing state-of-the-art procedures. At last, they concluded that the learning speed of ELM is a few folds more noteworthy than the conventional neural networks. Hence the proposed ELM-based strategy will be an improved answer for the multi-label problems. Rajendra Kumar Roul and Pranav Rai [20] propose an approach called CRSC for choosing the top features of information and features the significance of Multi-layer ELM feature space in the arena of text classification. The proposed feature selection is called Commonality-Rarity Score Computation. The point of text classification is to characterize the text archives into a set of pre-characterized groups. All together reinforce the classification strategy, choosing prime features, and therefore eliminate the irrelevant ones. The experiment result shows that ML-ELM dominates all other state-of-the-art classifiers. Jiao and Zhao [29] proposed a sparse deep tensor extreme learning machine (SDT-ELM) merging the tensor effect with ELM which accomplished great classification execution on 3 open data-sets. IAENG at el. [21] solve to overcome the problem of high dimensions and sparse features produced by the vector space model (VSM) text representation which increases the burden of ELM. XIE Yongfang ,MIN Mengcan, and CHEN Xiaofang* XIE Yongfang [22] propose a method, named constrained voting extreme learning machine (CV-ELM) which is contrasted and the conventional ELM. The CV-ELM decide the bias and input weight based upon the variations in between-class samples. For SD states the analysis results indicate that, the CV-ELM exactness rate is better. Ding and Zhang [27] bring in wavelet examination to ELM for unsupervised learning and semi-supervised learning.

Feature extraction is a procedure used to diminish an enormous information informational collection into important features. Choice of right arrangement of features plays a vital part in polarity classification. Tamilarasi, et. al. [36] proposed a detain to find the contrariety of words from twitter using dictionary-based and feature extraction technique. In this they trained and tested both dictionary-based and feature engineering techniques. And the experiment result shows that feature engineering approaches like Count Vectorizer, TF-IDF and Word2Vec perform better compared to the dictionary based practices. They used Logistic Regression, the Count Vectorizer and TF-IDF both get 81% while Word2Vec, SentiWordNet and VADAR get 75%, 65% and 68% of accuracy respectively. There is a resembles of feature extraction methods between TF-IDF and BM25 [40], evaluated in terms of weight on Twitter. Experimental result shows that the TF-IDF outperform the BM25. The TF-IDF method is applied in two different tools such as the local formula and the global formula heading to effective results. Kissi et. al. [39] represent a pragmatic study on five distinguish classification models by using two Arabic datasets with three feature extraction techniques such as word2vec, TF-IDF and word count. It is seen that the LR classifier produced the finest result using TF-IDF. As it not just focuses on the recurrence of words present in the corpus yet in addition gives the significance of the words which help to analyze the data better. In Implementation of SVM for Shopee Customer Sentiment Analysis [37] with TF-IDF show a way to resolve the classification problem. This is demonstrated by the utilization of word weighting output using the term frequency-inverse document frequency method. Here it shows an average accuracy of 97.3%. In SA and Topic Modeling of Twitter Data: A Text Mining Approach to the US-Afghan War Crisis [38] shows an analysis using to feature extraction; Word2vec and TF-IDF were used to classify tweets using different machine learning classifiers. In experiment result shows that the TF-IDF performs better with classifiers but in any case, their performance altogether decayed utilizing the Word2vec highlights. This could be credited to the trouble of machine learning classifiers to make significance out of the compacted features of the Word2vec model, so TF-IDF is preferred. Kaixu Liu [41] proposed spam detection using in opinion mining. They examined small unconstrained terms which are trying to detect the irrelevant spam word in the social media. They used TF-IDF of text n-gram to determine the learning. This technique manifests the encoding rule to gain proficiency with the procedure and great outcomes are acquired by surveying the content. Abdelaal, et. al. [45] introduced a supervised learning framework that accurately predicts both sarcasm and polarity of the Arabic text with six supervised machine classifiers, along with different Ngram and TF-IDF. DT accomplishes the best accuracy results (F1-score of 64.4%) when predicting sarcasm while SVM shows the best exactness results (F1-score of 58.5%). The analyzed feature extraction techniques incorporate a plain Term Frequency Inverse Document Frequency (TF-IDF) approach and its two changes by non-identical dimensionality depletion approach [42] such as Linear Discriminate Analysis and Latent Semantic Analysis. It was seen, that while taking care of to the classifier the information where features were extracted utilizing TF-IDF upgraded with LDA technique, the classifier gradually decreased the validation loss. The outcome shows that TF-IDF outperform from others. Kamalanathan, et. al. [43] proposed model considered two features TF-IDF and NGrams (Bigram) on the Amazon Alexa and IMDB film to review dataset for sentiment analysis. From result it is seen that from those two element extractions, a huge increase in feature extraction with TF-IDF includes instead of based on N-Gram. Badjatiya, et. al. [44] considers 3 representations: Bag of Words Vector (BoWV), TF-IDF and Char n-grams for the assignment of hate speech detection. Among the benchmark strategies, they observed that the word TFIDF proved to be improve than the character n-gram method. Hakim, et. al. used TF-IDF calculation was utilized to group news stories in Bahasa Indonesia [46]. This calculation counts the heaviness of every word concerning its redundancy in the data furthermore, the quantity of documents where it exists. At the point when a word is rehashed too often in every one of the texts, it implies that that word isn't significant, and that a high accuracy has been accomplished in characterization. Most recent Research demonstrates that computer-based news arrangement is more proficient than grouping by human. So, news should be arranged into different groups, in order to recognize the most well-known news group by the clients in this ideal country at any time [47]. Elicited from tf-idf and SVM, author proposed a new classification technique. They compared with five groups of 20 Newsgroup datasets and two BBC dataset with 94.93% and 97.48% respectively. Through this they found that BBC dataset group gets more accuracy. Ishizuka and Bun [48] proposed an alternate term weighting plan TF-Proportional Document Frequency (TF-PDF), which assign more prominent weights to terms that show up often in many records from many channels and lower loads to those that are only sometimes referenced.

III. METHODOLOGY

A. *Extreme Learning Machine*

This segment of the paper illustrates the hypothetical study on basic ELM. Huang et al. [34] proposed ELM to train SLFNs. It is a feedforward network for clustering, compression, regression, feature learning, classification, and sparse approximation with the multiple layer or single layers of hidden nodes. Single hidden layer feedforward network consists of 3 layers: input layer, output layer and hidden layer [35].

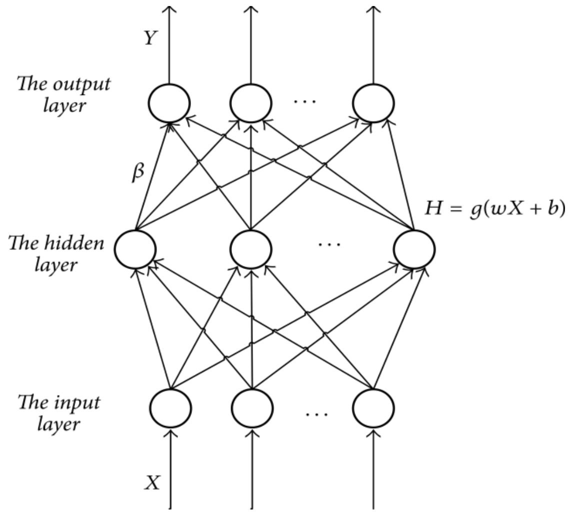

Fig. 1. Block-diagram of ELM

The layout of ELM model is shown in Fig. 1. Y and X represents the output and input vector. Here, b and W denotes the bias taken at the hidden layer and the assigned weight from input to hidden layer repectively, whereas β represents the output weight. The training of the model aims to limit the fault among the target and the result.

1) *ELM Training*

Here we will momentarily present the training problem for SLFN. The structure contains n input layer neurons, m output layer neurons and l hidden layer neurons. Let's consider the training sample $\{X, Y\} = \{x_i, y_i\}$ with an input feature and a desired matrix including the training trials with m and n dimension of output and input matrix respectively.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1Q} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nQ} \end{bmatrix}$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1Q} \\ \vdots & & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mq} \end{bmatrix} \qquad (1)$$

Next, the ELM arbitrarily sets the weights between the hidden layer and input layer where, wij addresses the weight between the ith hidden layer neuron and jth input layer neuron. Likewise, the ELM addresses βjk, weight between the jth hidden layer and the kth result layer. Then bias (b) of hidden layer neurons is set:

$$B = [b_1\ b_2\ \ldots.\ b_n]^T \qquad (2)$$

After setting the bias extreme learning machine choose the activation function g(x) with the output of the hidden layer (H):

$$H = g\,(wX + b) \tag{3}$$

Now output matrix T with each column vector as follows:

$$T = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ \vdots \\ t_Q \end{bmatrix} \tag{4}$$

The target output can be shown as

$$\sum_{i=1}^{N} \beta_i g(w_i x_j + b_i) = t_j \tag{5}$$

The formula from [4] and [5] can be abbreviated as

$$H\beta = T \tag{6}$$

By using Moore Penrose inverse we calculate the weight matrix to get unique solution with minimal error where H+ is the Moore-Penrose generalized inverse of matrix H.

$$\beta = H^{+}T \tag{7}$$

To boost the generalization of the network, regularization term is added to β. When total of training sample is more than hidden layer then [39]:

$$\beta = (\tfrac{I}{\lambda} + H^{T}H)^{-1} H^{T}T' \tag{8}$$

Likewise, when number of training sample is not exactly hidden layer then:

$$\beta = H^{T}(\tfrac{I}{\lambda} + H^{T}H)^{-1} T' \tag{9}$$

Algorithm 1: Algorithm of ELM

Training set: $\{x_i , y_i\}$ (i = 1,2,3,…,Q), activation function g(x) and each hidden layer with *l* hidden neurons

initialization:

1. Set arbitrary the input bias b_i and weight w_i
2. Then calculate the output matrix H of hidden layer.
3. Obtain the output weight β.

The fundamental training of ELM can be viewed as two stages: linear parameter solution and random initialization [35]. Without iterative tuned hidden framework ELM can give better generalization.

B. *Term Frequency – Inverse Document Frequency*

The factual measure gives an assessment about how relevant a word is to a corpus in a collection of records.

Evaluated on the basis of two metrics: -

- Term Frequency: - Words in a document. The weight of a label that occurs in a document is simply proportional to the tf.

  tf(t,d) = count of t in d / number of words in d

  Where, t = term

  d = document

  N = count of corpus

- Document Frequency: - It is the number of documents wherein the word is available.

  df(t) = occurrence of t in N documents

- Inverse Document Frequency: - It is the measure of how much data the word provides, i.e., assuming it is common or rare across all documents.

  idf(t) = N/df

If we have large number of N then IDF get explode, so to reduce the effect log is being used.

Sometime a fixed vocabulary is being missed out then df will be 0 but we can't divide it by 0, so to smooth it 1 is being added in the denominator.

idf(t) = log (N/(df + 1))

Then tf–idf is calculated as

tf-idf(t, d) = tf(t, d) * log (N/(df + 1))

Tf-idf is a system which set a probapility to every term in a corpus based on its frequency in corpus and the reciprocal document frequency. Fig. 2 depicts overall block diagram of TF-IDF.

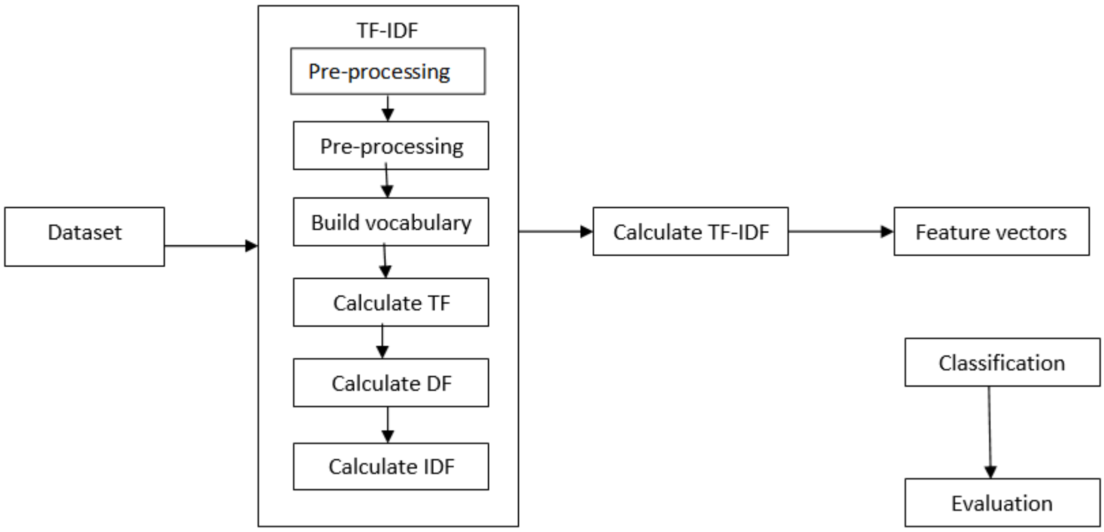

Fig. 2. Block-diagram for TF-IDF

IV. Results and Discussion

The model is carrying out in python and evaluate on a Linux device. The in-depth study is carried out on twitter airline revie dataset, which is publicly available. The training and testing ratio is set to 80:20 and the results obtained are tabulate as follows.

Fig. 3. Display the confusion matrix obtained for the testing set. The result shows that the model shows better performance for +ve classes. To further analyze the model various performance metrices are evaluated for each class. The obtained results are shown in TABLE I.

		Predicted		
		+ve	-ve	neutral
Actual	+ve	1709	93	33
	-ve	112	1499	225
	neutral	54	291	1491

Fig. 3. Confusion Matrix of test-set

TABLE I. Evalution of class wise performance metrices

label	tp	fn	fp	tn	acc	pre	rec
+ve	1709	126	166	3506	0.947	0.911	0.931
-ve	1499	337	384	3287	0.869	0.796	0.816
neutral	1491	345	258	3413	0.891	0.852	0.812

The results in TABLE I. shows that the ELM model with TF-IDF can recognize +Ve statements with hig accuracy than -ve or neutral. To further evaluate the execution of ELM the model accuracy is contrast to the well-known classifiers. The obtained outcomes are given in TABLE II.

TABLE II. Comparison of ELM with othe classifiers for TF-IDF features

Classifier	**Accuracy**
Random Forest	86%
SVM	85%
Naïve Bayes	76%
Decision Tree	84%
KNN	75%
ELM	86%

The results in TABLE II. shows that the ELM out performs may of the traditional models without any parameter tuning. However, the hidden node required for ELM is relatively high which pose as one of the limitations of the model.

V. Conclusion

In this work we made an in-depth survey of use-case of ELM and TF-IDF. Further the study evaluates ELM with respect to other well-known classifier models. From the results it is evident that ELM can outclass the traditional models without any parameter tuning and in one-pass learning. Although, the need of high number of hidden nodes in extreme learning machine is one of the drawbacks which needs to be addressed. Hereafter heuristic optimization can be integrated to input weight of ELM to obtain higher performance with a smaller number of hidden nodes.

References

[1] M. Wongkar and A. Angdresey, "Sentiment Analysis Using Naive Bayes Algorithm of the Data Crawler: Twitter," 2019 Fourth International Conference on Informatics and Computing (ICIC), 2019, pp. 1-5, doi: 10.1109/ICIC47613.2019.8985884.

[2] Li, M., Xiao, P. and Zhang, J., 2018. Text classification based on ensemble extreme learning machine. arXiv preprint arXiv:1805.06525

[3] Huang G-B, Ding X, Zhou H (2010) Optimization method based extreme learning machine for classification. Neurocomputing 74(1-3):155–163

[4] Pande, Shubhangi & Rathore, Neeraj & Purohit, Anuradha. (2021). A Survey and Analysis of Extreme Machine Learning Models and Its Techniques. 10.21203/rs.3.rs-599856/v1.

[5] Huang G-B, Zhou H, Ding X, Zhang R (2012) Extreme Learning Machine for Regression and Multiclass Classification. IEEE Trans Syst Man Cybernet—Part B: Cybernet 42(2):513–529.

[6] (Zhang L, Zhang D, Tia n F (2016) SVM and ELM: Who Wins? Object Recognition with Deep Convolutional Features from ImageNet. Springer International Publishing, Cham, pp 249–26 3)

[7] (Bhat AU, Merchant SS, Bhagwat SS (2008) Prediction of melting points of organic compounds using extreme learning machines. Ind Eng Chem Res 47(3):920–925).

[8] H. Liang et al., "Memristive Neural Networks: A Neuromorphic Paradigm for Extreme Learning Machine," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 3, no. 1, pp. 15-23, Feb. 2019, doi: 10.1109/TETCI.2018.2849721.

[9] A. R. Hassan, "A comparative study of various classifiers for automated sleep apnea screening based on single-lead electrocardiogram," 2015 International Conference on Electrical & Electronic Engineering (ICEEE), 2015, pp. 45-48, doi: 10.1109/CEEE.2015.7428288.

[10] K. Shang, X. Zeng and X. Dong, "Bagging-ELM Model for Heating Furnace Thermal Efficiency Prediction," 2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS), 2020, pp. 1329-1333, doi: 10.1109/DDCLS49620.2020.9275210.

[11] X. Zhang and X. Zheng, "Comparison of Text Sentiment Analysis Based on Machine Learning," 2016 15th International Symposium on Parallel and Distributed Computing (ISPDC), 2016, pp. 230-233, doi: 10.1109/ISPDC.2016.39.

[12] Y. Sönmez, T. Tuncer, H. Gökal and E. Avcı, "Phishing web sites features classification based on extreme learning machine," 2018 6th International Symposium on Digital Forensic and Security (ISDFS), 2018, pp. 1-5, doi: 10.1109/ISDFS.2018.8355342.

[13] M. I. Sa'ad, Kusrini and M. S. Mustafa, "Student Prediction of Drop out Using Extreme Learning Machine (ELM) Algorithm," 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), 2020, pp. 1-6, doi: 10.1109/ICORIS50180.2020.9320831.

[14] Ö. S. Keskin, A. Durdu, M. F. Aslan and A. Yusefi, "Performance comparison of Extreme Learning Machines and other machine learning methods on WBCD data set," 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021, pp. 1-4, doi: 10.1109/SIU53274.2021.9477984.

[15] Guang-Bin H, Qin-Yu Z, Chee-Kheong S (2006) Extreme learning machine: Theory and applications. Neurocomputing 70(1-3):489–50

[16] G. Cheng, L. Cai and H. Pan, "Comparison of Extreme Learning Machine with Support Vector Regression for Reservoir Permeability Prediction," 2009 International Conference on Computational Intelligence and Security, 2009, pp. 173-176, doi: 10.1109/CIS.2009.124.

[17] G. -B. Huang, H. Zhou, X. Ding and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," in IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), vol. 42, no. 2, pp. 513-529, April 2012, doi: 10.1109/TSMCB.2011.2168604.

[18] R. Venkatesan and M. J. Er, "Multi-label classification method based on extreme learning machines," 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV), 2014, pp. 619-624, doi: 10.1109/ICARCV.2014.7064375.

[19] Rajendra Kumar Roul and Pranav Rai. 2016. A New Feature Selection Technique Combined with ELM Feature Space for Text Classification. In Proceedings of the 13th International Conference on Natural Language Processing, pages 285–292, Varanasi, India. NLP Association of India.

[20] Li, M., Xiao, P. and Zhang, J., 2018. Text classification based on ensemble extreme learning machine. arXiv preprint arXiv:1805.06525.

[21] M. Mengcan, C. Xiaofang and X. Yongfang, "Constrained voting extreme learning machine and its application," in Journal of Systems Engineering and Electronics, vol. 32, no. 1, pp. 209-219, Feb. 2021, doi: 10.23919/JSEE.2021.000018.

[22] N. Liu and H. Wang, "Ensemble Based Extreme Learning Machine," in IEEE Signal Processing Letters, vol. 17, no. 8, pp. 754-757, Aug. 2010, doi: 10.1109/LSP.2010.2053356.

[23] Cao J, Lin Z, Huang G-B, Liu N (2012) Voting based extreme learning machine. Inf Sci 185(1):66–77

[24] C. Shen, S. -F. Zhang, J. -H. Zhai, D. -S. Luo and J. -F. Chen, "Imbalanced Data Classification Based on Extreme Learning Machine Autoencoder," 2018 International Conference on Machine Learning and Cybernetics (ICMLC), 2018, pp. 399-404, doi: 10.1109/ICMLC.2018.8526934.

[25] Liu P, Huang Y, Meng L, Gong S, Zhang G (2014) Two-stage extreme learning machine for highdimensional data. Int J Mach Learn Cybern 7(5):765–772

[26] Zhang, Nan & Ding, Shifei. (2017). Unsupervised and semi-supervised extreme learning machine with wavelet kernel for high dimensional data. Memetic Computing. 9.10.1007/s12293-016-0198-x.

[27] Wu T, Yao M, Yang J (2017) Dolphin Swarm Extreme Learning Machine. Cogn Comput 9(2):275–284

[28] J. Zhao and L. Jiao, "Sparse Deep Tensor Extreme Learning Machine for Pattern Classification," in IEEE Access, vol. 7, pp. 119181-119191, 2019, doi: 10.1109/ACCESS.2019 .2924647.

[29] ZHU K, YANG H M, MENG K. Short-term wind power forecast using extreme learning machine. Journal of Electric Power Science and Technology, 2019, 34(2): 106–111.

[30] S. O. Olatunji, "Extreme Learning machines and Support Vector Machines models for email spam detection," 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017, pp. 1-6, doi: 10.1109/CCECE.2017.7946806.

[31] H. Kaur, V. Mangat and Nidhi, "A survey of sentiment analysis techniques," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2017, pp. 921-925, doi: 10.1109/I-SMAC.2017.8058315.

[32] R. Wagh and P. Punde, "Survey on Sentiment Analysis using Twitter Dataset," 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), 2018, pp. 208-211, doi: 10.1109/ICECA.2018.8474783.

[33] Guang-Bin Huang, Qin-Yu Zhu and Chee-Kheong Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No.04CH37541), 2004, pp. 985-990 vol.2, doi: 10.1109/IJCNN.2004.1380068.

[34] Wang, J., Lu, S., Wang, SH. et al. A review on extreme learning machine. Multimed Tools Appl (2021). https://doi.org/10.1007/s11042-021-11007-7

[35] Dong Xiao, Beijing Li, Yachun Mao, "A Multiple Hidden Layers Extreme Learning Machine Method and Its Application", Mathematical Problems in Engineering, vol. 2017, Article ID 4670187, 10 pages, 2017. https://doi.org/ 10.1155/2017/4670187.

[36] D. Deepa, Raaji and A. Tamilarasi, "Sentiment Analysis using Feature Extraction and Dictionary-Based Approaches," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2019, pp. 786-790, doi: 10.1109/I-SMAC47947.2019.9032456.

[37] M. B. . Sitepu, I. R. . Munthe, and S. Z. . Harahap, "Implementation of Support Vector Machine Algorithm for Shopee Customer Sentiment Anlysis", SinkrOn, vol. 7, no. 2, pp. 619-627, May 2022..

[38] Akpatsa, Samuel & Addo, Prince & Lei, Hang & Li, Xiaoyu & Jr, Maxwell & Fiawoo, Duncan & Nartey, Juliana & Dagadu, Joshua. (2022). Sentiment Analysis and Topic Modeling of Twitter Data: A Text Mining Approach to the US-Afghan War Crisis. SSRN Electronic Journal. 10.2139/ssrn.4064560. Sentiment analysis and spam detection in short informal text using learning classifier systems

[39] Tarik Sabri, Omar El Beggar, Mohamed Kissi,Comparative study of Arabic text classification using feature vectorization methods, Procedia Computer Science,Volume 198,2022,Pages 269-275,ISSN 18770509,https:// doi.org /10.1016/j.procs.2021.12.239.

[40] A. I. Kadhim, "Term Weighting for Feature Extraction on Twitter: A Comparison between BM25 and TF-IDF," 2019 International Conference on Advanced Science and Engineering (ICOASE), 2019, pp. 124-128, doi: 10.1109/ICOASE.2019.8723825.

[41] Arif, M.H., Li, J., Iqbal, M. et al. Sentiment analysis and spam detection in short informal text using learning classifier systems. Soft Comput 22, 7281–7291 (2018).https://doi.org/10.1007/s00500-017-2729-x

[42] R. Dzisevič and D. Šešok, "Text Classification using Different Feature Extraction Approaches," 2019 Open Conference of Electrical,

Electronic and Information Sciences (eStream), 2019, pp. 1-4, doi: 10.1109/eStream.2019.8732167.

[43] Das, Mamata, Selvakumar Kamalanathan and Pja Alphonse. “A Comparative Study on TF-IDF Feature Weighting Method and Its Analysis Using Unstructured Dataset.” COLINS (2021)..

[44] Badjatiya, Pinkesh & Gupta, Shashank & Gupta, Manish & Varma, Vasudeva. (2017). Deep Learning for Hate Speech Detection in Tweets. 10.1145/3041021.3054223.

[45] Predicting sarcasm and polarity in arabic text automatically: Supervised machine learning approach, http://www.jatit.org/volumes/Vol100No8/16Vol100No8.pdf.

[46] Hakim, Ari Aulia, Alva Erwin, Kho I. Eng, Maulahikmah Galinium and Wahyu Muliady. “Automated document classification for news article in Bahasa Indonesia based on term frequency inverse document frequency (TF-IDF) approach.” 2014 6th International Conference on Information Technology and Electrical Engineering (ICITEE) (2014): 1-4.

[47] S. M. H. Dadgar, M. S. Araghi and M. M. Farahani, "A novel text mining approach based on TF-IDF and Support Vector Machine for news classification , " 2016 IEEE International Conference on Engineering and Technology (ICETECH), 2016, pp. 112-116, doi: 10.1109/ ICETECH. 2016. 7569223.

[48] Bun, KK & Ishizuka, M 2002, Topic extraction from news archive using TF∗PDF algorithm. in WISE 2002 - Proceedings of the 3rd International Conference on Web Information Systems Engineering., 1181645, Institute of Electrical and Electronics Engineers Inc., pp. 73-82, 3rd International Conference on Web Information Systems Engineering, WISE 2002, Singapore, Singapore, 02/12/12.https://doi.org/10.1109/ WISE.2002.1181645.