Fake review detection using enhanced ensemble support vector machine system on e-commerce platform

Seenia Joseph, S. Hemalatha
Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India

Article Info

Article history:

Received Jun 18, 2024
Revised Oct 21, 2024
Accepted Oct 30, 2024

Keywords:

Distance metric
Ensemble method
Fake reviews detection
Supervised machine learning
Feature selection
Support vector machine

ABSTRACT

Due to the quick growth of online marketing transactions, including buying and selling, fake reviews are created to promote the product market and mislead new customers. E-commerce customers can post reviews and comments on the goods or services they obtained. Before making a purchase, new customers frequently read the feedback and comments posted on the website. Nowadays customers find it very difficult to identify whether the reviews are fake or not, but doing so is essential. So, it's very crucial to develop an online spam detection system to help both consumers and producers in their decision-making. The reviewer's behaviour and important review characteristics can help you identify fake reviews. The importance of this study is to develop a fake review detection system on e-commerce platforms using an enhanced ensemble support vector machine system in which the Euclidean distance is replaced with the Mahalanobis distance metric. Review texts collected from Amazon and Yelp were given as input data sets into the constructed model and classified as fake or real.

Corresponding Author:

Seenia Joseph
Department of Computer Science, Karpagam Academy of Higher Education
Coimbatore, India
Email: seeniajoseph@gmail.com

1. INTRODUCTION

The internet has become a content creation platform where people express their opinions and experiences, significantly impacting customers and businesses. Potential customers often check reviews before making a purchase. Reviews help potential customers better understand other people's experiences, especially when choosing between purchasing a product or service. Chadchankar *et al.* [1] observed that 81% of individuals research products online before buying, and if verified buyers report bad customer service, over 58% will cease transactions. This underscores the significant influence of buyer feedback on purchasing decisions [2]. However, not all reviews of the product on the internet are genuine. Malicious users often post fake reviews to mislead customers into promoting or downgrading a target product or service. Fake reviews on e-commerce platforms mislead consumers, leading them to make poorly informed purchasing decisions and potentially receive subpar products. This erosion of trust can damage consumer confidence in the platform and the reputation of honest businesses. This article focuses on developing an efficient method to identify fake reviews on e-commerce platforms to help both consumers and producers in their decision-making.

To ensure the integrity of online reviews, it is crucial and necessary to create efficient tools to identify online reviewers. The type of review and features mentioned that are not directly related to the content play a role in identifying fake reviews. However, fake reviews may require the creation of other characteristics related to the reviewer himself, such as the time/date of the assessment or his writing style.

Therefore, the successful feature extraction of reviews leads to the successful recognition of fake reviews. This study aims to create a fake review detection system for e-commerce platforms by utilizing an advanced ensemble support vector machine (SVM) model, which replaces the Euclidean distance metric with the Mahalanobis distance metric. Euclidean distance is a common metric for measuring the distance between two points in a feature space. However, it assumes that the features are uncorrelated and have the same variance. On the other hand, Mahalanobis distance accounts for the correlation between features and the variance within the data. This makes it more suitable for cases with correlated features or differing scales. By replacing Euclidean distance with Mahalanobis distance in SVM, we aim to achieve a more nuanced and accurate distance measurement, improving the detection of fake reviews. This enhanced SVM is then ensembled with different classifiers for better results than the conventional method.

Greengrad [3] believes that implementing new novel algorithms and ideas can increase the performance of a spam detection system. Alternatively, Chavolla *et al.* [4] and Clune [5] argue that rather than looking for new ideas, it is more useful to improve the functioning of existing systems. The developed system involves enhancing the SVM classifier by optimizing the speed by removing irrelevant support vectors to reduce the number of computations involved and utilizing Mahalanobis distance matric to improve the construction of hyperplane in classification. Among online merchants, Amazon has been chosen for the study's application section. Because of Amazon's dominance in online retailing, we chose their dataset. Amazon is a sizable, well-established online retailer that offers a variety of datasets for machine learning applications. The Yelp dataset is provided as the studies' second dataset. Yelp.com is a user-generated website that reviews nearby establishments and resembles social networking sites in that it enables user communication.

Elmurngi and Gherbi [6] have taken movie reviews as a dataset. Text classification and sentiment analysis (SA) methods are used on the real dataset of movie reviews. They have applied two different approaches, with and without stop words in that they have compared Naive Bayes (NB), SVM, K-nearest neighbors (KNN-IBK), and decision tree (DT-J48) for sentiment classification of reviews. The measured outcomes show that the SVM algorithm outperforms rival algorithms for both techniques and achieves the highest level of accuracy in text classification and the detection of fake reviews. Abri *et al.* [7] thoroughly examined linguistic traits that differentiate fraudulent from reliable internet reviews. After examining fifteen characteristics, they discovered that fake reviews frequently employ more pauses, lengthier phrases, and duplicate terminology. Using these traits with various machine learning algorithms, they accurately distinguished fake from real reviews. Similarly, Wang *et al.* [8] employed supervised machine learning to propose two feature types subject features and readability characteristics for classifying Yelp reviews. Their results showed these features outperformed n-grams in identifying fraudulent reviews, and incorporating reviewers' behavioral traits significantly improved classification accuracy for actual Yelp opinion spam data.

Birim *et al.* [9] studied which feature combination emotion scores, topic distributions, cluster distributions, and a bag of words most effectively detect fraudulent reviews. The research addresses the significant issue of fake reviews influencing customer purchase decisions, using Amazon.com review data and various sentiment analysis features. Findings show that behavior-related features, particularly the verified purchase feature, significantly impact the classification of fraudulent reviews when combined with text-related features. Alsubari *et al.* [10] examined a Yelp dataset, applying methods like sentiment analysis, part-of-speech (POS) tagging, linguistic inquiry and word count (LIWC), and subjectivity for feature extraction. They extracted various attributes, including counts of adjectives, verbs, nouns, adverbs, polarity, objectivity, and subjectivity. Using information gain (IG), they selected the most valuable features. DTs, random forest, and adaptive boosting were employed to classify reviews as false or reliable, achieving accuracies of 96%, 94%, and 97%, respectively.

Gutierrez-Espinoza *et al.* [11] studied ensemble learning approaches for detecting false online content, specifically fake restaurant reviews. Their results showed that these methods outperform traditional machine learning algorithms. Stand-alone multilayer perceptron (MLP) classifiers achieved up to 68.2% accuracy, while an AdaBoost ensemble of MLPs reached 77.3%. Liu *et al.* [12], Rout *et al.* [13], and You *et al.* [14], used outlier detection techniques to classify reviews as spam or accurate. Outlier detection, a common data analysis topic, focuses on identifying anomalies in datasets [15] and is applied in fault detection, intrusion detection, and fraud detection. Current outlier detection methods fall into four categories: statistical distribution-based, distance-based, density-based, and subspace learning-based [16]. Additionally, studies using pre-trained language models like bidirectional encoder representations from transformers (BERT) and XLNet with latent dirichlet allocation (LDA) topic distributions found them effective for identifying fake COVID-19 news [17]. Muhammad and Ahmed [18] demonstrated good accuracy in identifying false reviews using sparse matrices of term frequency-inverse document frequency (TF-IDF), count vectorizer (CV), and n-gram features in a principal component analysis (PCA) feature set. Alsubari *et al.* [19] utilized a convolutional neural network-long short-term memory (CNN-LSTM) model on a multi-domain dataset, achieving in-domain accuracy rates of 87%, 86%, 85%, and 77%, with a cross-domain accuracy of 89%, surpassing previous studies.

2. PROPOSED METHOD

The main goal of this study is to create a method for detecting spam online reviews that is both effective and efficient. Several studies have been carried out to offer detecting methods that will solve the above-mentioned desirable features. Furthermore, existing solutions have a high false-positive rate, a long time to identify spam reviews, a large gap between the installation of spam detection methods and the guarantee of a positive result, and so on. The proposed system overcomes these difficulties and it first proposes enhanced algorithms to improve the working of each step, from which the best working method is combined to form the enhanced ensemble identification system. Features of the proposed system are

- To perform feature engineering construct a feature vector having only optimal features extracted from multiple entities, which helps to improve the performance of ham/spam detection systems.
- To design enhanced classification algorithms to improve the performance of the online review spam detection system.
- To design an enhanced ensemble classifier system to increase the accuracy of spam detection.

In this research method, the optimization of the SVM classifier is done in two manners. The first step is to remove irrelevant support vectors with no relevancy during classification. This lessens the quantity of computations and thus solves the high training time required. The second is to replace the conventionally used Euclidean distance with the Mahalanobis distance measure.

3. METHOD AND FINDINGS

The suggested spam online review detection is part of the online review security component since it protects users (or customers) against false details. The suggested system employs feature engineering, classification, and clustering methods to enhance online spam detection. The primary goal of the spam review detection system (SRD) is to identify all spam reviews using machine learning techniques. The algorithm starts by mapping all training features into the SVM vector space and computing the margins for each category. The smallest M margins are selected as relevant support vectors (SVs), with the rest discarded as noise. The identified SVs are then outlined in the prototypal vector space and remapped to the original space. Mahalanobis distance is used to calculate the average distance between new features and each SV set. The category with the closest SVs is assigned to the new feature.

The review datasets of smartphones collected from Amazon and Yelp are taken into consideration. The datasets received were pre-processed to meet the requirements of this study. The final pre-processed dataset therefore included details stored in a way that the algorithms could simply access. This final dataset includes 67,986 reviews from Amazon and 12876 from Yelp and is given as input to the system developed.

The SRD is a programme that uses a machine learning algorithm, A, to determine if a review R, is spam or not.

$$A\ (R, F) = \begin{cases} Spam\ if\ F\ has\ suspicious\ content\ Ham \\ \qquad\qquad\qquad otherwise \end{cases}$$

In the above equation, R is the online review that has to be categorized as spam or ham, F is a feature vector that represents the various characteristics of R. The detection algorithm, A, uses a learning algorithm (Q) to train the machine learning algorithm with the dataset that has features (F) pre-collected from reviews $q = Q(F, C)$, where $F\ =\ \{f1, \dots, fn\}$ and C is the set of target labels, which is {spam, ham} in this research. The detection algorithm, A, handles one review at a time and classifies them as ham or spam, depending on the result (q) obtained from (Q). Two actions are taken from the result.

$$\begin{cases} q = Spam\ \ Delete\ Review \\ q = Ham\quad Allow\ Review \end{cases}$$

Figure 1 provides a detailed methodology for the spam review detection model, with different phases. Phase I describes feature engineering, and phase II includes designing an enhanced ensemble classification system.

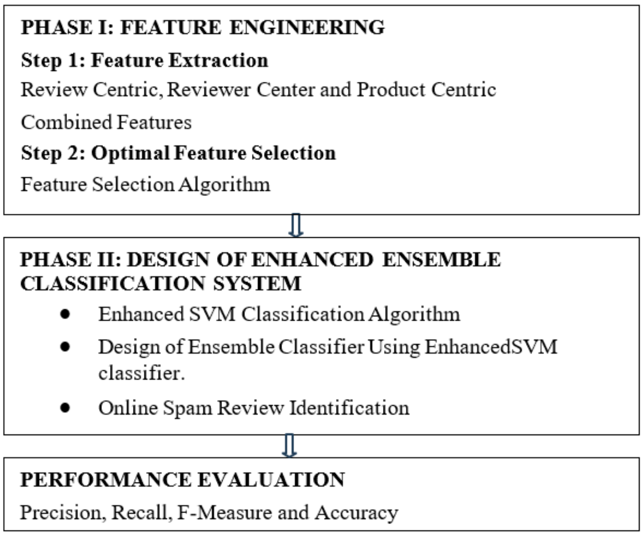

Figure 1. Different phases of methods

3.1. Phase I: feature engineering
3.1.1. Feature engineering

The approaches put forth in Phase I involve feature engineering to create a feature vector with only the best features and to improve review spam detection. Feature engineering is described as the process of creating or extracting features from data sources [20]. This is accomplished in Phase I through the two processes of feature extraction and feature selection.

3.1.2. Feature extraction

The feature extraction technique, which condenses raw data, is the most significant component of the system for detecting review spam. This reduced data is referred to as feature vectors or feature spaces. The three aspects to evaluate in an online review are content, reviewer, and product are also the topic of several characteristics obtained for this study. In the first stage of Phase I, a total of 53 sets of features are extracted. Out of 53 features, there are 38 review-centric features,13 reviewer-centric features and 2 product-centric features described in Table 1.

Table 1. Features extracted

Review centric features (38)	Textual features (9), metadata (8), content similarity using bag of words (5), POS tags (9), n-grams (4), rating (1), sentiment (1), burst patterns (1)
Reviewer centric features (13)	Reviewer activities, maximum number of reviews, percentage of positive reviews, review length, reviewer deviation, burst review ratio, ratio of verified purchases, reviewer burstiness, extreme ratings, reviewer average proliferation, reviewer spamicity, % of positive reviews, % of negative reviews
Product centric features (2)	Rank in sale, average rating

3.1.3. Feature selection

The features retrieved in step 1 may contain irrelevant, noisy, or duplicated properties. Overfitting and extra calculations are two classification difficulties that result in increased temporal complexity. A feature selection algorithm operates on the premise that not all extracted features are crucial for detecting review spam. Identifying and removing these irrelevant features can enhance the performance of the spam detection system [21]. In the first stage, an enhanced maximum relevance minimum redundancy (MRMR) feature selection approach is employed to locate distinguishing and significant characteristics.

According to the experimental results, using any feature selection technique might increase the performance of the spam review detection (SRD) system. Using the suggested feature selection using the modified genetic algorithm (FS_MGA) method, the SRD system achieved a precision of 79.25%, 76.94%, and 74.43% (Amazon dataset) and 78.04%, 75.04%, and 72.51% (Yelp dataset). Similarly, the SRD system with the MGA feature selection method had a high recall value of 80.25% (SVM), 77.88% (KNN), and

75.76% (NB) for the Amazon dataset and 78.53% (SVM), 76.12% (KNN), and 73.87% (NB) for the Yelp dataset. Codes used in the SRD system for different feature selections are described in Table 2.

Table 2. Coding scheme used – feature selection

Code	Algorithm
NFS	No feature selection
FS_	MRMR feature selection using MRMR algorithm
FS_MIMG	Feature selection using MRMR_IG and MRMR_MI algorithm
FS_ACO	Feature selection using ACO algorithm
FS_ACO+GA	Feature selection using ACO + GA algorithm
FS_MGA	Feature selection using FS_MIMG and ACO + GA algorithm

Performance matrices for different feature selections are given below. Figure 2 shows the percentage of precision when implemented in SVM, KNN and NB with Amazon and Yelp data sets. Similarly, Figure 3 shows performance of Recall.

The MGA method generated an F-measure of 79.75% (SVM), 77.41% (KNN), and 75.09% (NB) when utilizing the Amazon dataset shown in Figure 4. When evaluated using the Yelp dataset, the F-measure of the MGA method was 78.29%, 75.58%, and 73.18%, respectively, when tested with SVM, KNN, and NB classifiers. When evaluated using SVM, KNN, and NB classifiers, the proposed MGA method had a maximum accuracy of 80.08%, 78.40%, and 76.92% on the Amazon and 78.07%, 76.76%, and 74.94% on the Yelp datasets as shown in Figure 5.

Figure 2. Precision

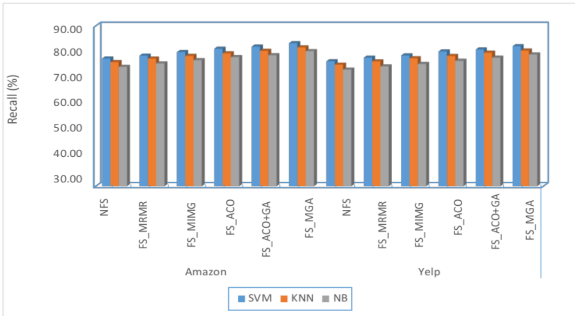

Figure 3. Recall

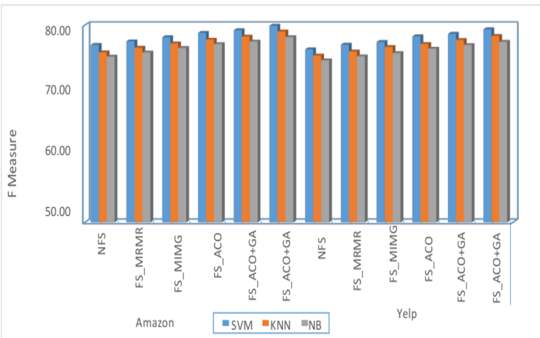

Figure 4. F-measure

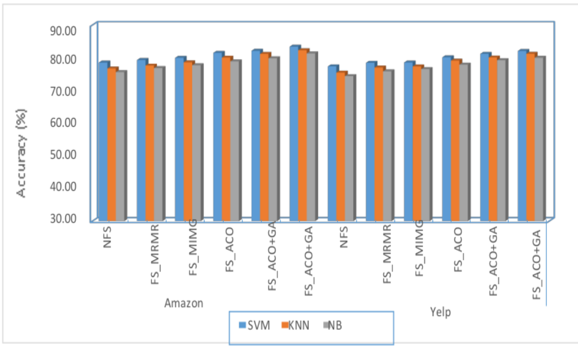

Figure 5. Accuracy

3.2. PHASE 2: design of enhanced ensemble classification system

The proposed system is designed in two steps. Step 1: enhance a classification algorithm and step 2: design ensemble systems using the classifier enhanced in step 1. SVM was selected as the classifier because it is the most widely used method for classification and prediction and also has an excellent track record of success in achieving high performance [22]-[24] when compared to various other classifiers. In this research study, the SVM classifier is optimized in two different ways. The first step is to eliminate support vectors that are not relevant for classification and are irrelevant. The second is to utilize the Mahalanobis distance metric in place of the commonly used Euclidean distance. This enhanced SVM is taken as the primary base classifier to design an ensemble classifier. According to Yildırım *et al.* [25], several methods for creating ensembles have been proposed like knowledge-based methods and randomization methods. Following the establishment of the ensemble system with base classifiers, the subsequent phase involves employing a technique to combine the outcomes of these base classifiers. This process employs two approaches: integration (fusion) methods and selection methods [26]. The second method is used in this research. The optimal feature vector produced by the MGA algorithm is used to train and test all the classifiers. Coding schemes used in enhanced classification systems for different algorithms are shown in Table 3.

The addition of a speed optimization strategy to the single SVM classifier increased its performance with the Amazon dataset by 4.5% (precision and recall), 4.47% (F-measure), and 3.66% (accuracy). The efficiency improvement gained when evaluated with the Yelp dataset was 3.71% (precision), 4.69% (recall), 4.20% (F-measure), and 3.34% (accuracy). When evaluated with the Amazon dataset, the optimization approach using the ensemble system improved performance by 0.79% (precision), 0.86% (recall), 0.83% (F-Measure), and 0.61% (accuracy). The efficiency improvement for the Yelp dataset was 1.34% (precision), 1.04% (recall), 1.19% (F-measure), and 1.14 seconds (accuracy). Precision evaluation when used with Amazon and Yelp datasets is given in Figure 6. Performance of Recall when used with different algorithms is given in Figure 7. F measure matrix when used with Amazon and Yelp datasets is given in Figure 8. The accuracy of models when used with different algorithms is presented in Figure 9.

Table 3. Coding scheme used–enhanced classification system

Code	Algorithm
S SVM	Classification system
ES	Ensemble SVM classification system
ES_SO	Enhanced ensemble SVM classification system with speed optimizers
ES_SO+ED	Enhanced ensemble SVM classification system with speed optimizers and Euclidean distance measure
ES_SO+MD	Enhanced ensemble SVM classification system with speed optimizers and Mahalanobis distance measure

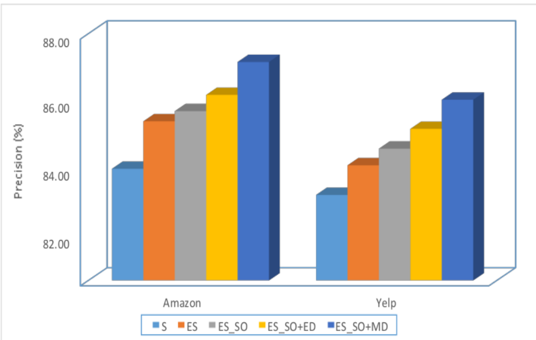

Figure 6. Precision

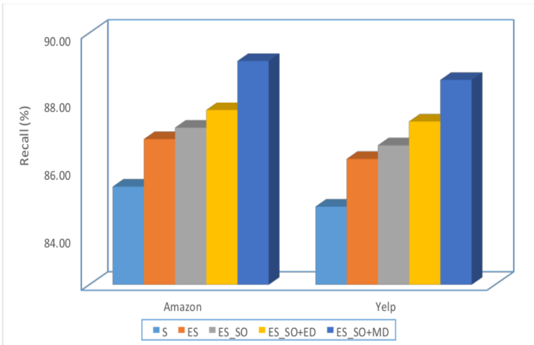

Figure 7. Recall

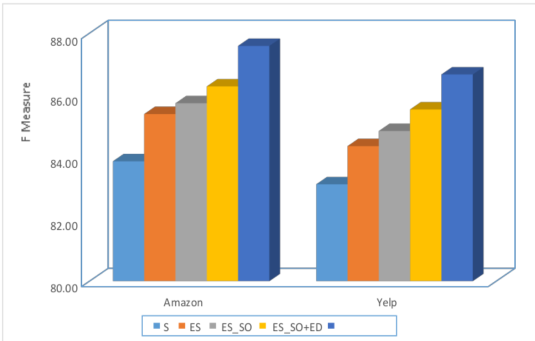

Figure 8. F-measure

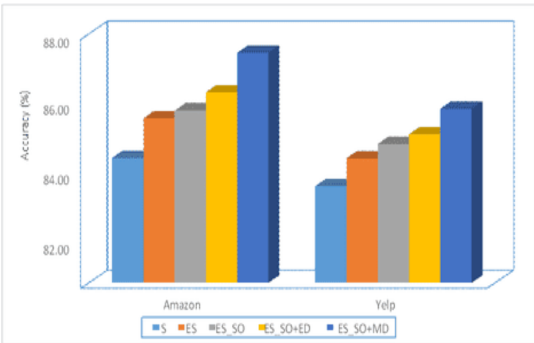

Figure 9. Accuracy

4. RESULTS AND DISCUSSION

In this research, we have developed a fraudulent review identification model using an online spam detection system in which enhanced ensemble SVM is used, which can help customers and marketing managers identify opinion spammers and their suspicious behavior when making decisions. The optimization accomplished via the usage of the distance measure demonstrated that the Mahalanobis distance measure outperformed the Euclidean distance in terms of classification performance. When compared to ES and utilizing the Amazon dataset, the system ES_SO+MD demonstrated an average efficiency improvement of 4.67% (precision), 5.95% (recall), 5.30% (F-Measure), and 5.09% (accuracy) in terms of precision, recall, F-measure, and accuracy. Using the Yelp dataset, the same approach demonstrated efficiency gains of 5.34%, 6.11%, 5.72%, and 3.97% in terms of precision, recall, F-measure, and accuracy, respectively. Phase I experimental results demonstrated that employing any feature selection algorithm positively impacted the performance of online spam review detection, with the proposed algorithm yielding the greatest improvement. Specifically, the combination of MRMR with mutual information (MI) and MRMR with IG, enhanced by ant colony optimization and genetic algorithms, achieved a 9.11% efficiency gain in accuracy for Amazon and a 9.08% gain for Yelp compared to classifiers without feature selection. In Phase II, it was confirmed that the optimization methods integrated into the SVM classifier were effective. The enhanced ensemble system, using the improved SVM as the base classifier, achieved high accuracies of 86.79% for Amazon and 83.20% for Yelp, and also reduced time complexity significantly. While the conventional SVM classifier took 22.03 seconds for Amazon and 17.37 seconds for Yelp, the optimized ensemble system cut these times to 18.04 seconds and 13.53 seconds, respectively.

5. CONCLUSION

The online spam detection system developed in this research enhanced the SVM system by replacing Euclidean distance with the Mahalanobis distance measure and then ensembled with a classifier giving a better result than a conventional method. The results obtained show that the combination of speed optimization and the hyperplane construction while using the Mahalanobis distance measure has a high impact on the performance of the SRD system both in terms of classification, accuracy and speed compared with classical support vector machine classifier. The proposed systems can be further improved by including an outlier detection algorithm, that can detect abnormal behaviors in reviews. Different linguistic constructs such as modifiers, negations, emojis, and ironic words are not taken into consideration, but they can all be used to improve the effectiveness of the proposed system.

REFERENCES

[1] A. S. Chadchankar, K. Borate, and P. Gombhi, "Opinion mining from online product review and online product's rating," *International Journal of Scientific and Technology Research*, vol. 9, no. 1, pp. 3317–3320, 2020.

[2] J. Zhang, W. Zheng, and S. Wang, "The study of the effect of online review on purchase behavior: comparing the two research methods," *International Journal of Crowd Science*, vol. 4, no. 1, pp. 73–86, 2020, doi: 10.1108/IJCS-10-2019-0027.

[3] Bar-Ilan University, "The brain inspires a new type of artificial intelligence," *Science News. Science Daily*, 2019.

[4] E. Chavolla, A. Valdivia, P. Diaz, D. Zaldivar, E. Cuevas, and M. A. Perez, "Improved unsupervised color segmentation using a modified HSV color model and a bagging procedure in K-means++ algorithm," *Mathematical Problems in Engineering*, vol. 2018, no. 1, p. 2786952, 2018, doi: 10.1155/2018/2786952.

[5] J. Clune, "AI-GAs: AI-generating algorithms, an alternate paradigm for producing general artificial intelligence," *arXiv preprint 1905.10985*, 2019, doi: 10.48550/arXiv.1905.10985.

[6] E. Elmurngi and A. Gherbi, "An empirical study on detecting fake reviews using machine learning techniques," in *2017 Seventh International Conference on Innovative Computing Technology (INTECH)*, Aug. 2017, pp. 107–114, doi: 10.1109/INTECH.2017.8102442.

[7] F. Abri, L. F. Gutierrez, A. S. Namin, K. S. Jones, and D. R. W. Sears, "Linguistic features for detecting fake reviews," in *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, Dec. 2020, pp. 352–359, doi: 10.1109/ICMLA51294.2020.00063.

[8] X. Wang, X. Zhang, C. Jiang, and H. Liu, "Identification of fake reviews using semantic and behavioral features," in *2018 4th International Conference on Information Management (ICIM)*, May 2018, pp. 92–97, doi: 10.1109/INFOMAN.2018.8392816.

[9] Ş. Ö. Birim, I. Kazancoglu, S. K. Mangla, A. Kahraman, S. Kumar, and Y. Kazancoglu, "Detecting fake reviews through topic modelling," *Journal of Business Research*, vol. 149, pp. 884–900, Oct. 2022, doi: 10.1016/j.jbusres.2022.05.081.

[10] S. N. Alsubari, M. B. Shelke, and S. N. Deshmukh, "Fake reviews identification based on deep computational linguistic features," *International Journal of Advanced Science and Technology*, vol. 29, no. 8s, pp. 3846–3856, 2020.

[11] L. Gutierrez-Espinoza, F. Abri, A. S. Namin, K. S. Jones, and D. R. W. Sears, "Fake reviews detection through ensemble learning," *arXiv preprint 2006.07912*, Jun. 2020, doi: 10.48550/arXiv.2006.07912.

[12] W. Liu, J. He, S. Han, F. Cai, Z. Yang, and N. Zhu, "A method for the detection of fake reviews based on temporal features of reviews and comments," *IEEE Engineering Management Review*, vol. 47, no. 4, pp. 67–79, Dec. 2019, doi: 10.1109/EMR.2019.2928964.

[13] J. K. Rout, A. Dalmia, K.-K. R. Choo, S. Bakshi, and S. K. Jena, "Revisiting semi-supervised learning for online deceptive review detection," *IEEE Access*, vol. 5, pp. 1319–1327, 2017, doi: 10.1109/ACCESS.2017.2655032.

[14] L. You, Q. Peng, Z. Xiong, D. He, M. Qiu, and X. Zhang, "Integrating aspect analysis and local outlier factor for intelligent review spam detection," *Future Generation Computer Systems*, vol. 102, pp. 163–172, Jan. 2020, doi: 10.1016/j.future.2019.07.044.

[15] I. Chatterjee, M. Zhou, A. Abusorrah, K. Sedraoui, and A. Alabdulwahab, "Statistics-based outlier detection and correction method for Amazon customer reviews," *Entropy*, vol. 23, no. 12, p. 1645, Dec. 2021, doi: 10.3390/e23121645.

[16] S. Afzal, A. Afzal, M. Amin, S. Saleem, N. Ali, and M. Sajid, "A novel approach for outlier detection in multivariate data," *Mathematical Problems in Engineering*, pp. 1–12, Oct. 2021, doi: 10.1155/2021/1899225.

[17] A. Gautam, V. Venktesh, and S. Masud, "Fake news detection system using XLNet model with topic distributions: CONSTRAINT@AAAI2021 shared task," in *Communications in Computer and Information Science*, Springer, 2021, pp. 189–200.

[18] F. Muhammad and S. Ahmed, "Fake review detection using principal component analysis and active learning," *International Journal of Computer Applications*, vol. 178, no. 48, pp. 42–48, 2019, doi: 10.5120/ijca2019919418.

[19] S. N. Alsubari, S. N. Deshmukh, M. H. Al-Adhaileh, F. W. Alsaade, and T. H. H. Aldhyani, "Development of integrated neural network model for identification of fake reviews in e-commerce using multidomain datasets," *Applied Bionics and Biomechanics*, pp. 1–11, Apr. 2021, doi: 10.1155/2021/5522574.

[20] R. M. Nabi, S. A. Saeed, and A. M. W. Abdi, "Feature engineering for stock price prediction," *International Journal of Advanced Science and Technology*, vol. 29, no. 12s, pp. 2486–2496, 2020, [Online]. Available: http://sersc.org/journals/index.php/IJAST/article/view/24722.

[21] K. S. Reddy and E. S. Reddy, "Using reduced set of features to detect spam in twitter data with decision tree and KNN classifier algorithms," *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 9, pp. 6–12, 2019, doi: 10.35940/ijitee.f3616.078919.

[22] T. V. Dinh, H. Nguyen, X.-L. Tran, and N.-D. Hoang, "Predicting rainfall-induced soil erosion based on a hybridization of adaptive differential evolution and support vector machine classification," *Mathematical Problems in Engineering*, pp. 1–20, Feb. 2021, doi: 10.1155/2021/6647829.

[23] B. Gaye, D. Zhang, and A. Wulamu, "Improvement of support vector machine algorithm in big data background," *Mathematical Problems in Engineering*, pp. 1–9, Jun. 2021, doi: 10.1155/2021/5594899.

[24] F. R. Lumbanraja, E. Fitri, Ardiansyah, A. Junaidi, and R. Prabowo, "Abstract classification using support vector machine algorithm (case study: abstract in a computer science journal)," *Journal of Physics: Conference Series*, vol. 1751, no. 1, 2021, doi: 10.1088/1742-6596/1751/1/012042.

[25] P. Yıldırım, U. K. Birant, and D. Birant, "EBOC: ensemble-based ordinal classification in transportation," *Journal of Advanced Transportation*, pp. 1–17, Mar. 2019, doi: 10.1155/2019/7482138.

[26] H. B. Ahmad, "Ensemble classifier based spectrum sensing in cognitive radio networks," *Wireless Communications and Mobile Computing*, pp. 1–16, Jan. 2019, doi: 10.1155/2019/9250562.

BIOGRAPHIES OF AUTHORS

Seenia Joseph is a research scholar at Karpagam Academy of Higher Education, Coimbatore, and an Assistant Professor at Bharata Mata College. She obtained her master's degree in Computer Applications from Calicut University in 2007. Her research focuses on data mining techniques applied across various domains. To date, she has authored 5 papers in national and international journals and contributed to 2 book chapters. She can be contacted at email: seeniajoseph@gmail.com.

Dr. S. Hemalatha is presently an Associate Professor and Research Guide at Karpagam Academy of Higher Education in Coimbatore. Her research interests encompass data mining, image processing, and networking. With a prolific scholarly output, she has authored over 10 papers, which have been published in Scopus-indexed journals, along with an additional 5 papers in peer-reviewed journals. She can be contacted at email: drhemashanmugam@gmail.com.