A Semi-Supervised Machine Learning-Based Detection of Deceptive Opinions in Digital Social Networks

Nidhi A. Patel, Nirali R. Nanavati

Abstract— **Online reviews play an important part in guiding consumers' decisions in today's e-commerce and social networking environment. As the influence of Online Social Networks (OSNs) grows, so does the prevalence of deceptive or fraudulent reviews crafted to mislead customers, enhance product reputation, or undermine competitors. These fabricated opinions distort genuine user experiences and significantly impact sales and brand credibility. This work aims to accurately distinguish between factual and deceptive reviews using Positive and Unlabeled (PU) machine learning techniques within a semi-supervised framework. By integrating linguistic features and representation methods such as word2vec, trigram, bigram and unigram models, the proposed approach effectively identifies fraudulent content even with limited labelled data. Experimental results demonstrate improved accuracy and performance compared to traditional PU-based approach, highlighting the potential of advanced machine learning strategies for mitigating review spam and enhancing the reliability of online platforms.**

Index Terms— **Deceptive Review, Machine Learning, Online Social Networks, Semi-supervised Learning, Spam Review, User-Generated Content**

I. INTRODUCTION

The exponential growth of Online Social Networks (OSNs) and digital platforms has facilitated the large-scale creation and dissemination of user-generated content (UGC). Although this participatory environment enables rich information exchange, it operates largely without robust verification mechanisms, creating substantial challenges regarding the authenticity and reliability of online content [1]. In the absence of systematic content validation, misinformation and deceptive information have proliferated, posing significant risks to both users and commercial entities [2]. A major subset of online misinformation is deceptive opinion spam, which includes artificially generated or falsified reviews, misleading comments, fabricated posts, and promotional distortions, and has emerged as a particularly critical issue [3]. Detecting such deceptive reviews is inherently complex due to their contextual dependence, linguistic resemblance to genuine reviews, and deliberate construction to influence user perception [4].

Opinion reviews have become indispensable in shaping consumer attitudes and guiding purchase decisions in contemporary e-commerce ecosystems. However, the vast volume of reviews generated across OSNs makes it increasingly difficult to determine whether a review reflects an authentic user experience or originates from fraudulent manipulation [5]. Malicious actors exploit this ambiguity by generating misleading or biased reviews to artificially elevate product ratings, discredit competitors, or manipulate market dynamics. These practices distort consumer decision-making, erode trust, damage brand reputation, and compromise the overall integrity of digital marketplaces. Consequently, developing reliable mechanisms for distinguishing between genuine and deceptive reviews has become a critical research priority in social media analytics, information retrieval, and digital marketing [6], [7].

Conventional supervised and unsupervised learning techniques exhibit limitations in deceptive review detection due to their dependence on either fully labeled or completely unlabeled data. To overcome these constraints, recent research highlights the potential of semi-supervised approaches. In alignment with these advancements, the present study proposes a hybrid semi-supervised machine learning framework that integrates classification and clustering methodologies enhanced with n-gram representations and Word2Vec embeddings. The proposed model aims to improve detection accuracy and provide deeper insights into deceptive content patterns. Comparative evaluations against existing frameworks demonstrate the model's effectiveness in identifying deceptive reviews and mitigating the spread of misleading information across social platforms.

We review and analyze related work in the next area. The proposed approach is discussed in Section III. In Section IV, we present and discuss the experimental results, and the paper concludes with Section V.

II. RELATED WORK

Detection of deceptive reviews has been addressed through various methodologies, including supervised techniques on labeled data, unsupervised approaches on unlabeled data, and semi-supervised techniques on partially labeled datasets. Key studies employing these approaches are discussed in the following section.

Jha et al. [8] proposed an algorithm for supervised learning for deceptive review detection using LR, SVM, DT, KNN, RF, and MNB on a labeled review dataset. Their experimental evaluation demonstrated that the SVM classifier achieved the best performance among all tested models. Elmogy et al. [9] presented a supervised machine learning framework for detecting deceptive reviews by combining textual features with n-gram models (bi-gram

Nidhi A. Patel, Research Scholar, Gujarat Technological University, Ahmedabad, Gujarat, India.

Nirali Nanavati, Associate Professor, Department of Computer Engineering, Sarvajanik College of Engineering and Technology, Surat, Gujarat, India.

and tri-gram) along with reviewer behavioral features. The authors evaluated NB, KNN, RF, SVM, and LR on a real Yelp dataset and found that including behavioral features improved performance, with KNN (K=7) yielding the highest F-score. Banerjee et al. [10] applied various supervised learning algorithms to identify authentic and fraudulent reviews according to four linguistic cues: quantity of detail, ease of comprehension, perception indicators and writing style. A major drawback of this work is that it considers only labeled data. Rout et al. [11] utilized content similarity and sentiment polarity features to identify fake and genuine reviews. The authors applied three algorithms: NB, DT, and SVM; however, the approach is limited by the small number of features used. Etaiwi et al. [12] used supervised learning methods to detect deceptive reviews using preprocessing techniques and linguistic features such as POS and Bag-of-Words (BoW). They applied gradient boosted trees, NB, RF, DT, and SVM. As an outcome, SVM and NB perform greater. Silpa et al. [13] used a supervised learning approach based on textual data in reviews, as well as sentiment classification, to determine whether reviews were deceptive or genuine. The authors evaluated various classifiers, including LR, SVM, NB, and DT. Badresiya et al. [14] applied supervised algorithms to detect spam review in the domain of review mining on a 1600-item dataset. They utilized the text of reviews to identify deceptive reviews. The results show that the SVM outperforms all other supervised methods for identify review spam. However, utilize only the review's content to detect deceptive reviews. Barbado et al. [15] describe a framework for detecting false reviews by a Yelp review dataset of product. The authors applied supervised learning methods to the dataset, combining both review and reviewer centric approaches. As an outcome, they used the AdaBoost method for their testing. Wang et al. [16] introduced new strategy for identifying deceptive reviews that uses rolling collaborative training and the merging multiple attributes. Their method uses a multi-feature initial index system that incorporates sentiment, textual, and behavior characteristics from opinion. To represent text, use the related approach (Doc2vec), and then train an initial data set of seven classifiers. Asaad et al. [17] applied four preprocessing steps were utilized with normalization, tokenization, stemming and stop word removal. TF-IDF approaches were used to extract features. Three machine learning techniques used by the authors for the classification: stochastic gradient descent, support vector classifier, and Xgboost. Hassan et al. [18] developed a supervised learning system using LR, NB, and SVM to identify fake reviews. The authors used a dataset from the hotel review that included attribute such as sentiment polarity, Empath, and TF-IDF.

Unsupervised methods, such as Independent Component Analysis, Principal Component Analysis, and clustering techniques, operate on unlabeled data to learn meaningful patterns. Dong et al. [19] developed the Unsupervised Topic-Sentiment Joint (UTSJ), which uses the (LDA) Latent Dirichlet Allocation model and incorporates four levels: subject, document, sentiment, and word. Mothukuri et al. [20] constructed clusters from the extracted features using unsupervised techniques such as K-means clustering, GMM Diagonal covariance, and GMM Full covariance clustering, to identify deceptive reviews in Yelp's café dataset. The authors determined that K-means has the highest accuracy among the three. Li et al. [21] developed an unsupervised pattern-driven method that identifies compositional and linguistic irregularities in reviews using clustering and rule-extraction techniques. Their approach effectively distinguishes sincere from deceptive reviews in real-world scenarios where labeled datasets are scarce. A technique for identifying a set of deceptive reviews based on nominated topics has been proposed by Li et al. [22]. The three stages of the proposed model are as follows: first, identifying similar groups and target topics; second, clustering reviews using the K-means method; third, labeling suspicious group as deceptive using time burstiness and content duplication. Mukherjee et al. [23] introduced an unsupervised approach for identifying opinion deceptive using a novel generative model called the Latent Spam Model (LSM), which exploits both spammers' behavioral footprints and linguistic. As a result, the proposed model outperforms existing algorithms on real-world data.

Ligthart et al. [24] evaluated several semi-supervised learning techniques for deceptive review detection across hotel review datasets. Their results show that self-training combined with NB classifier. The authors used four semi-supervised methods, including co-training, self-training, label propagation plus spreading, and Transductive SVM. Tian et al. [25] introduced a non-convex semi-supervised Ramp-One Class SVM for detecting opinion spam using positive and unlabeled data. The method effectively handles outliers and lack of negative labels, achieving strong generalization on standard deceptive review datasets such as Yelp and Ott. Yılmaz et al. [26] proposed SPR2EP, a semi-supervised deceptive review detection method that iteratively labels unlabeled reviews using a bootstrapped classifier. The model integrates text features with semi-supervised refinement, improving detection performance over traditional supervised-only approaches.

A. Research Gap

As per the literature, the problems are slow convergence [8], inaccurate predictions [9], time-consuming and resource-intensive [10], computationally expensive [12][13], and computational complexity [20] [21] for correctly identifying deceptive review detection. To overcome this, we have proposed a detection model with a good learning paradigm.

III. Proposed Work

A. Dataset Description

The dataset that we have utilized includes 1600 reviews, 800 of which are deceptive reviews and 800 genuine reviews. Of these, 400 reviews have negative sentiment polarities, and 400 show positive sentiment polarities. Positive opinion reviews are a combination of deceptive and truthful reviews. We have collected the dataset from Ott et al. [27], Narayan et al. [28] for the review spam detection.

B. Data Preprocessing

The main standard preprocessing steps are considered in this paper including: tokenization and punctuation marks removal. The tokenization is separating the text into a small number of words or sentences. The crucial preprocessing step is punctuation mark removal, which divides the text into paragraphs, sentences, and phrases. Word2Vec is a method for creating word embeddings. The purpose of Word2Vec is to group the vectors of related words together in vector space [29].

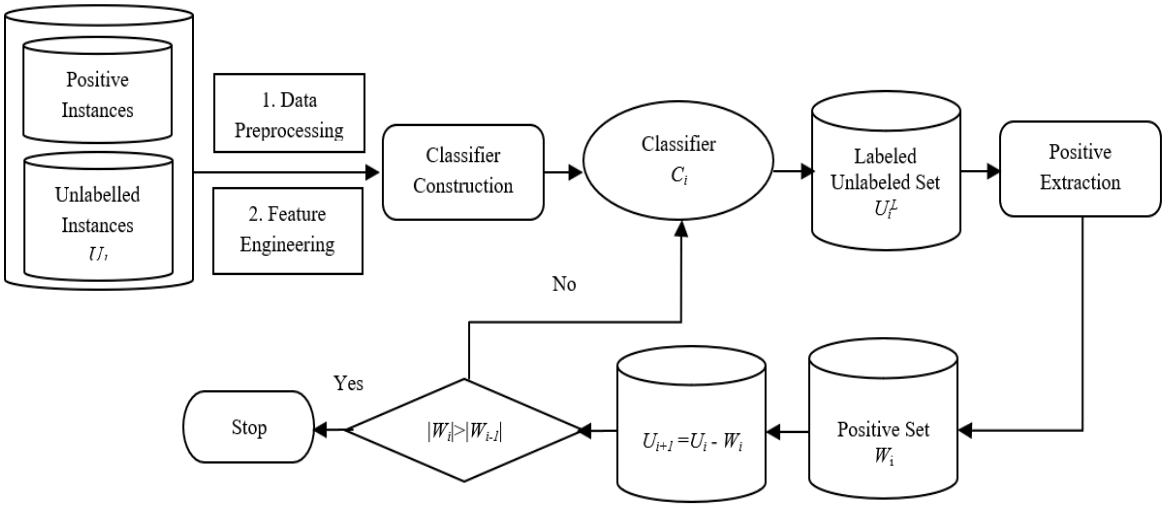

Fig. 1. Flowchart of the Proposed M ethodology.

Table I. The results of various classifiers using 40 deceptive reviews as training and 520 unlabeled reviews with feature methods of N-gram with Uni-gram, Bi-gram Tri-gram, Word2Vec and Tri-gram + Word2Vec

Classifier	N-gram						Word2Vec				Tri-gram+ Word2Vec			
	Uni-gram	Bi-gram	Tri-gram											
	A (%)	A (%)	A (%)	P	R	F	A (%)	P	R	F	A (%)	P	R	F
DT	54.38	58.36	61.35	0.68	0.63	0.65	63.72	0.69	0.65	0.67	67.28	0.71	0.64	0.67
NB	36.29	38.53	40.13	0.45	0.36	0.40	44.36	0.48	0.39	0.43	48.82	0.52	0.46	0.49
SVM	53.24	55.23	55.26	0.54	0.51	0.52	56.92	0.54	0.52	0.53	61.36	0.66	0.59	0.62
KNN	63.18	65.32	65.14	0.69	0.71	0.70	69.25	0.72	0.70	0.71	74.27	0.78	0.72	0.75
RF	53.28	56.27	57.32	0.62	0.56	0.59	60.01	0.64	0.61	0.62	65.23	0.69	0.63	0.66
LR	60.18	62.36	62.84	0.72	0.74	0.73	64.25	0.74	0.73	0.73	70.23	0.74	0.73	0.74

Table II. The results of various classifiers using 80 deceptive reviews as training and 520 unlabeled reviews with feature methods of N-gram with Uni-gram, Bi-gram Tri-gram, Word2Vec and Tri-gram + Word2Vec

Classifier	N-gram						Word2Vec				Tri-gram+ Word2Vec			
	Uni-gram	Bi-gram	Tri-gram											
	A (%)	A (%)	A (%)	P	R	F	A (%)	P	R	F	A (%)	P	R	F
DT	69.87	71.36	72.34	0.74	0.71	0.72	74.63	0.78	0.71	0.74	78.36	0.79	0.76	0.78
NB	53.42	55.51	56.24	0.54	0.51	0.52	56.93	0.61	0.52	0.56	59.28	0.61	0.58	0.60
SVM	71.18	73.46	72.35	0.76	0.78	0.77	75.36	0.79	0.80	0.79	79.19	0.80	0.76	0.78
KNN	73.30	76.25	76.89	0.84	0.79	0.81	79.25	0.81	0.78	0.79	82.47	0.84	0.81	0.83
RF	62.94	65.24	67.27	0.65	0.64	0.64	69.25	0.67	0.62	0.64	73.24	0.76	0.72	0.74
LR	76.23	77.25	77.26	0.73	0.75	0.74	81.53	0.83	0.76	0.79	84.37	0.87	0.83	0.85

Table III. The results of various classifiers using 120 deceptive reviews as training and 520 unlabeled reviews with feature methods of N-gram with Uni-gram, Bi-gram Tri-gram, Word2Vec and Tri-gram + Word2Vec

Classifier	N-gram						Word2Vec				Tri-gram+ Word2Vec			
	Uni-gram	Bi-gram	Tri-gram											
	A (%)	A (%)	A (%)	P	R	F	A (%)	P	R	F	A (%)	P	R	F
DT	46.30	48.25	48.82	0.53	0.49	0.51	51.36	0.54	0.51	0.52	55.23	0.57	0.54	0.55
NB	48.02	55.27	55.73	0.53	0.52	0.52	55.91	0.58	0.54	0.56	58.91	0.61	0.57	0.59
SVM	60.96	61.43	63.43	0.63	0.61	0.62	64.25	0.64	0.61	0.62	68.03	0.69	0.65	0.67
KNN	61.56	62.37	73.25	0.71	0.75	0.73	76.92	0.73	0.71	0.72	79.42	0.81	0.76	0.78
RF	48.41	48.62	55.91	0.52	0.54	0.53	58.04	0.59	0.56	0.57	62.28	0.64	0.61	0.63
LR	73.39	75.26	79.91	0.91	0.76	0.83	82.81	0.92	0.81	0.86	**86.37**	0.88	0.84	0.86

Table IV. The results of various classifiers using 120 deceptive reviews as training and 520 unlabeled reviews with existing and proposed with Tri-gram + Word2Vec feature methods

Classifier	Existing [28]				Proposed (N-gram (Tri-gram) + Word2Vec)			
	A (%)	P	R	F	A (%)	P	R	F
DT	45.31	50.00	45.71	47.76	55.23	0.57	0.54	0.55
NB	54.68	34.37	57.89	43.13	58.91	0.61	0.57	0.59
SVM	60.93	90.92	56.86	69.87	68.03	0.69	0.65	0.67
KNN	60.93	71.87	58.97	64.78	79.42	0.81	0.76	0.78
RF	46.87	56.25	47.36	51.42	62.28	0.64	0.61	0.63
LR	73.43	68.75	75.86	72.13	**86.37**	0.88	0.84	0.86

C. Feature Engineering

Feature engineering is the process of creating or extracting features from data. Our proposed approach used a "Bag-of Words" (BoW) strategy. In this approach, individual word groups are found in the text. These features, known as n-grams, are created by choosing a continuous word from a specific sequence. In the proposed approach, we have used

unigram, bigram and trigram (n = 1, 2 and 3), word2vec and combined n-gram with word2vec features and compared the results with the existing approach [28]. The results are shown in Section IV.

D. Proposed Algorithm

The below sequential steps show the pseudocode of the proposed PU Learning approach.

PU-Learning for Spam Review Detection

1 Preprocessing: Tokenization and Punctuation Marks Removal
2 Feature Engineering: N-gram and Word2Vec
3 $i \leftarrow 1$;
4 $|W_0| \leftarrow |U_1|$;
5 $|W_1| \leftarrow |U_1|$;
6 while $|W_i| \leq |W_i{-}1|$ do
7 $Ci \leftarrow$ Generate Classifier(P, U_i);
8 $U_i^L \leftarrow Ci\,(U_i)$;
9 $W_i \leftarrow$ Extract Positives(U_i^L);
10 $U_{i+1} \leftarrow U_i - W_i$;
11 $i \leftarrow i + 1$;
12 Return Classifier Ci

The proposed approach is based on the PU learning method [30]. It is an iterative procedure where unlabeled datasets are treated as negative classes in this approach. Next, we trained various classifiers using positive cases. Here, six classifiers have been used. These include DT, NB, SVM, KNN, RF, and LR classifiers. After, these classifiers to classify unlabeled datasets. All positive examples have been eliminated from instances of unlabeled data, and the remaining instances considered to be negative instances for the subsequent iteration. This process will continue until the stop condition is fulfilled. Figure 1 presents the flowchart for the proposed approach.

IV. Experimental Results and Discussion

Our results with the semi-supervised learning method, we used in our experiments yielded the following results: As mentioned in section III for the dataset, we have implemented our model in Python. Tables I, II, and III display the results for different training sets. For building test data, we randomly selected 160 opinion reviews with a combination of deceptive and truthful reviews. The 640 opinion reviews have been applied to training sets of various sizes. We consist of 40, 80, and 120 deceptive opinion instances, respectively. We have used 520 unlabeled instances in all the cases as per existing [28]. We utilized the following six classifiers: 1) Decision Tree (DT), 2) Naive Bayes (NB), 3) Support Vector Machine (SVM), 4) K-Nearest Neighbor (KNN), 5) Random Forest (RF), and 6) Logistic Regression (LR). We considered the accuracy (A), precision (P), recall (R), and f-score (F) parameters for evaluation and compared the results.

Tables I, II, and III compare the proposed with uni-gram, bi-gram, tri-gram of n-gram, word2vec, tri-gram + word2vec features. Out of all the results, tri-gram + word2vec got better results. Table IV shows the result of 120 deceptive reviews as training and 520 unlabeled reviews with the existing and proposed with tri-gram + word2vec feature methods.

A. Discussion

The highest level of accuracy we have achieved is **86.37 %** when using 120 deceptive opinion reviews as training and 520 unlabeled opinion reviews using logistic regression. The logistic regression works on containing maximum likelihood estimation and using a SoftMax classifier that divides multiple classes of data and works well with the textual dataset.

V. Conclusion and Future Scope

In this work, a PU based machine learning algorithm was applied using preprocessing and feature engineering for better prediction accuracy. For preprocessing, we simply applied tokenization and removed punctuation marks including white spaces. For feature engineering, we evaluated our approach using n-gram (namely unigram, bigram and trigram), word2vec, combined trigram with word2vec methods and observed that tri-gram with word2vec gives comparatively good results. We experimented our approach with different supervised machine learning algorithms namely decision tree, naive bayes, support vector machine, k-nearest neighbor, random forest, and logistic regression. From the results, we found that logistic regression-based approach outperforms existing PU based approach. In the future, the same work can be extended with more features with other machine learning algorithms.

Appendix

DT: Decision Tree
KNN: K-Nearest Neighbors
LR: Logistic Regression
MNB: Multinomial Naïve Bayes
NB: Naïve Bayes
OSNs: Online Social Networks
RF: Random Forest
SVM: Support vector Machine
UGC: User Generated Content

Acknowledgment

I, Nidhi Patel, would like to express my sincere gratitude to all those who have supported and contributed to this research. I extend my heartfelt thanks to my PhD supervisor, Dr. Nirali Nanavati, for her unwavering guidance, invaluable insights, and continuous encouragement throughout this work.

References

[1] D. Kavitha, G. Srujankumar, C. Akhil, and P. Sumanth, "Uncovering the truth: A machine learning approach to detect fake product reviews and analyze sentiment. Information Systems Engineering and Management", 2025, pp. 309–324. https://doi.org/10.1007/978-3-031-74885-1_20.

[2] P. T. Pandi and N. S. Kumar, "Fake review detection in e-commerce using machine learning and NLP technique", in Proc. 3rd Int. Conf. Inventive Computing and Informatics (ICICI), 2025, pp. 692–698. https://doi.org/10.1109/icici65870.2025.11069636.

[3] K. Mane, S. Dongre, and M. Madankar, "Fake review detection using random forest classifier," in Proc. IEEE Int. Students' Conf. Electrical, Electronics and Computer Science (SCEECS), 2025, pp. 1–6. https://doi.org/10.1109/sceecs64059.2025.10940605.

[4] Y. Zhang, H. Wang, and A. Stavrou, "A multiview clustering framework for detecting deceptive reviews,", Journal of Computer Security, pp. 1–22, 2023, https://doi.org/10.3233/jcs-220001.
[5] A. H. Alshehri, "An online fake review detection approach using famous machine learning algorithms," Computers, Materials & Continua, vol. 78, no. 2, 2024, pp. 2767–2786, https://doi.org/10.32604/cmc.2023.046838.
[6] H. Alawadh, A. Alabrah, T. Meraj, and H. Rauf, "Semantic features-based discourse analysis using deceptive and real text reviews," Information, vol. 14, no. 1, 2023, https://doi.org/10.3390/info14010034.
[7] M. Petrescu, H. Ajjan, and D. L. Harrison, "Man vs machine – detecting deception in online reviews", Journal of Business Research, vol. 154, 2023, https://doi.org/10.1016/j.jbusres.2022.113346.
[8] M. Jha, D. Maru, P. Sharma, and R. Khalkar, "Fake reviews detection using supervised machine learning algorithm", International Journal of Advances in Engineering and Management (IJAEM), pp. 647-651, 2022, doi: 10.35629/5252-0407647651.
[9] A. M. Elmogy, U. Tariq, A. Mohammed, and A. Ibrahim, "Fake reviews detection using supervised machine learning", International Journal of Advanced Computer Science and Applications, vol. 12, no. 1, 2021.https://doi.org/10.14569/ijacsa.2021.0120169.
[10] S. Banerjee, A. Y. K. Chua, and J. J. Kim, "Using supervised learning to classify authentic and fake online reviews", in Proc. ACM IMCOM, 2015, doi: 10.1145/2701126.2701130.
[11] J. K. Rout, S. Singh, S. K. Jena, and S. Bakshi, "Deceptive review detection using labeled and unlabeled data", Multimedia Tools and Applications, vol. 76, no. 3, pp. 3187–3211, 2016, https://doi.org/10.1007/s11042-016-3819-y.
[12] W. Etaiwi and G. Naymat, "The impact of applying different preprocessing steps on review spam detection", Procedia computer science, vol. 113, pp. 273-279. 2017.
[13] C. Silpa, P. Prasanth, S. Sowmya, Y. Bhumika, C. S. Pavan and M. Naveed, "Detection of Fake Online Reviews by using Machine Learning", In 2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA), 2023, pp. 71-77.
[14] A. Badresiya, S. Vohra, and J. Teraiya, "Performance analysis of supervised techniques for review spam detection", International Journal of Advanced Networking Applications (IJANA), pp. 21-24, 2014.
[15] R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers", Information Processing & Management, vol. 56, no. 4, pp. 1234–1244, 2019.
[16] J. Wang, H. Kan, F. Meng, Q. Mu, G. Shi, and X. Xiao, "Fake Review Detection Based on Multiple Feature Fusion and Rolling Collaborative Training," vol. 8, 2020, https://doi.org/ 10.1109/ACCESS.2020.3028588.
[17] W. H. Asaad, R. Allami, and Y. H. Ali, "Fake review detection using machine learning", Revue d'Intelligence Artificielle, vol. 37, no. 5, 2023, https://doi.org/10.18280/ria.370507.
[18] R. Hassan and M. R. Islam, "A supervised machine learning approach to detect fake online reviews", In proceedings international conference on computer and information technology (ICCIT), 2020, pp. 1-6.
[19] L. Y. Dong, S. J. Ji, C. J. Zhang, Q. Zhang, D. W. Chiu, L. Q. Qiu, and D. Li, "An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews", Expert Systems with Applications, vol. 114, pp. 210-223, 2018.
[20] R. Mothukuri, A. Aasritha, K. C. Maramella, K. N. Pokala, and G. K. Perumalla, "Fake review detection using unsupervised learning", In Proceedings International Conference on Sustainable Computing and Data Communication Systems (ICSCDS), 2022.
[21] J. Li, N. Li, L. Yang, and P. Zhang, "Identifying review spam with an unsupervised approach based on topic abuse", Proceedings of the 8th International Conference on Computing and Artificial Intelligence, 2022, pp. 350–356. https://doi.org/10.1145/3532213.3532265.
[22] J. Li, P. Lv, W. Xiao, L. Yang, and P. Zhang, "Exploring groups of opinion spam using sentiment analysis guided by nominated topics", Expert Systems with Applications, vol. 171, 2021, https://doi.org/10.1016/j.eswa.2021.114585.
[23] A. Mukherjee and V. Venkataraman, "Opinion spam detection: An unsupervised approach using generative models", Technical Report, University of Houston, 2014.
[24] A. Ligthart, C. Catal, and B. Tekinerdogan, "Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification", Applied Soft Computing, vol. 101. https://doi.org/10.1016/j.asoc.2020.107023.
[25] Y. Tian, M. Mirzabagheri, P. Tirandazi, B. Hosseini, and M. Seyed, "A non-convex semi-supervised approach to opinion spam detection by ramp-one class SVM", Information Processing & Management, vol. 57, 2020, doi: 10.1016/j.ipm.2020.102381.
[26] C. M. Yilmaz and A. O. Durahim, "SPR2EP: A semi-supervised spam review detection framework", in Proceedings IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 2018, pp. 306–313. https://doi.org/10.1109/asonam.2018.8508314.
[27] M. Ott, C. Yejin, C. Claire, and H. Jeffrey, "Finding deceptive opinion spam by any stretch of the imagination", In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies, Association for Computational Linguistics, vol. 1, 2011, https://doi.org/10.48550/arXiv.1107.4557.
[28] R. Narayan, J. K. Rout, and S. K. Jena, "Review spam detection using semi-supervised technique", in Advances in Intelligent Systems and Computing, 2017, pp. 281–286. https://doi.org/10.1007/978-981-10-3376-6_31.
[29] N. Patel, and N. Nanavati, "Deceptive Review Detection in Online Social Networks", Industrial Engineering Journal.
[30] D. Hernández, G. Rafael, and M. Manuel, "Using PU-learning to detect deceptive opinion spam", in Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2013, pp. 38-45.

Nidhi Patel, is a Ph.D. Research Scholar at Gujarat Technological University in Ahmedabad, Gujarat, India. She is working as an Assistant Professor with the Department of Computer Engineering at Institute of Shree Swami Atmanand Saraswati Institute of Technology, Surat, India. She holds Bachelor of Engineering in Computer Engineering, Master of Engineering in Computer Engineering with a specialized in software engineering. Her research focuses on machine learning, artificial intelligence, deep learning, data mining, and big data. Her academic output includes a strong portfolio of journal articles, review articles, and conference papers.

Dr. Nirali Nanavati, is an accomplished Associate Professor in Computer Engineering at Sarvajanik College of Engineering & Technology (SCET), Surat. She earned her doctoral degree from SVNIT, Surat, following an M.S. in Computer Science from NJIT, Newark, USA. Prior to academia, Dr. Nanavati gained industry experience as a Technical Consultant with IBM France and Infosys Technologies Ltd. Her research focuses on privacy□preserving data mining, database systems, artificial intelligence and machine learning. She contributed to patents and has published extensively—journal articles, book chapters, and conference papers. Beyond research, Dr. Nanavati is a celebrated mentor. Under her guidance, student teams won first prizes in international competitions including the CSI–InApp 2021 "Medical Image Translation" and 2020's "Generative AI based project", along with awards at the Smart India Hackathon and Innovations 2019. She regularly delivers expert workshops on AI, machine learning, privacy, and data analytics.