Bulletin of Electrical Engineering and Informatics
Vol. 14, No. 1, February 2025, pp. 587∼595
ISSN: 2302-9285, DOI: 10.11591/eei.v14i1.8288

A scoring approach for detecting fake reviews using MRCS similarity metric enhanced by personalized k-means

Mohammed Ennaouri, Ahmed Zellou
Team of Software Project Management, ENSIAS, Mohammed V University, Rabat, Morocco

Article Info

Article history:

Received Feb 8, 2024
Revised Aug 26, 2024
Accepted Sep 4, 2024

Keywords:

Fake reviews detection
Linguistic exaggeration
Mean review cosine similarity
Opinion spam
Personalized k-means

ABSTRACT

Online commerce has grown in the digital age, and as a result, consumers now depend more than ever on other consumer feedback to make informed purchasing decisions. However, as the importance of reviews has increased, so has the prevalence of fake ones, which now infiltrate platforms and manipulate users' perceptions. This presents a significant challenge to preserving confidence and integrity in online marketplaces. This study addresses the difficulty of identifying fake reviews by introducing a distinctive methodology that incorporates advanced natural language processing (NLP) tools. By including a new metric, mean review cosine similarity (MRCS), which enhances textual similarity assessment for more accurate detection, we improve the identification process. Additionally, an exaggeration detection technique is included, enhancing the model's capacity to identify deceptive variations in review content. Furthermore, an adaptive clustering method differs from traditional k-means classification through modifying clusters to adjust to the constant evolution of misleading linguistic patterns. Empirical validation on the Yelp labeled dataset demonstrates the approach's accuracy (90%), with high precision (89%), recall (95%), and F1 score (92%), indicating its effectiveness and highlighting areas for further refinement.

Corresponding Author:

Mohammed Ennaouri
Team of Software Project Management, ENSIAS, Mohammed V University
Rabat, Morocco
Email: mohammed_ennaouri@um5.ac.ma

1. INTRODUCTION

Opinions expressed on the internet are increasingly influencing consumer decisions in the wide digital world [1], [2]. Genuine consumer reviews serve an important part in guiding consumers through the wide range of products and services available, impacting their attitudes and purchases, and eventually determining the profitability or ruin of a business. However, deceptive methods harm the honesty of content created by users. The most dangerous element is the meticulously planned technique for fake reviews, which are intended to deceive naïve clients. As platforms and businesses fight against the widespread propagation of these misinformation campaigns, there is an increasing demand for efficient techniques to distinguish between the real and the fake.

In this dynamic digital environment, the issue of fake review detection becomes crucial. Our investigation explores earlier studies, revealing methods for identifying the veracity of misleading opinions. While previous studies have investigated various approaches to fake review detection, there remain gaps in considering the impact of linguistic complexity in a review, the manner in which it was managed, and the effectiveness of existing detection systems in adapting machine learning approaches. Thus, we overcome these traditional limitations covered by supervised, unsupervised, and semi-supervised learning algorithms, existing data analysis, examining feature sets, and describing methods for evaluating reviews, reviewers, and products.

Our main goals are to develop a new approach and establish a robust foundation for identifying fake reviews. We introduce Word2Vec, a linguistic model that goes beyond conventional bag-of-words models to indicate semantic richness [3]. This model, when paired with cosine similarity measures, enables precise exploration of the semantic environment, enhancing the ability to detect nuances in language. We identify exaggeration, a technique commonly employed in fake reviews, and develop methods to detect it effectively. Our system is designed to discriminate between misleading reviews and genuine ones, adapting to the evolving linguistic strategies employed by dishonest entities [4]. We evaluate the performance of our approach using metrics such as accuracy, precision, recall, and F1 score, leveraging the Yelp labeled dataset [5]. The goal is continuous improvement, aiming to minimize false positives and negatives while maintaining high accuracy.

Jindal and Liu's foundational study in 2008 [6], [7] marks a significant starting point for understanding the impact of customer reviews on consumer decisions and the emergence of fake reviews as a challenge to the authenticity of online information. This research highlights the crucial role that user-generated content plays in shaping consumer attitudes and purchasing decisions, underscoring the need for reliable methods to distinguish between genuine and deceptive reviews. Indeed, the literature reveals a diverse array of approaches. In the context of supervised learning, Taneja and Kaur [8] proposed an ensemble model that combines various classification models to improve fake review detection, while Wang *et al.* [9] introduced a rolling collaborative training model that adjusts classifier parameters to account for data changes. Aiyar and Shetty [10] used specific heuristics, such as character n-grams, alongside classic machine learning approaches like support vector machine (SVM), random forest, and naïve Bayes (NB) to identify fraudulent reviews. The MGSD model, developed by Noekhah *et al.* [11], accurately detects opinion spam by calculating spamicity scores based on entity features and sample correlations.

Semi-supervised learning methods, on the other hand, provide a bridge between labeled and unlabeled data. Tian *et al.* [12] addressed labeled data limitations using the one-class SVM algorithm with ramp loss function. Ligthart *et al.* [13] explores six popular semi-supervised learning algorithms against typical supervised classification techniques, offering a thorough review of various approaches. Hassan and Islam [14] developed semi-supervised (expectation maximization) and supervised (SVM and NB algorithms) text mining models to detect false online reviews, highlighting the potential of partially labeled data for better classification. Zhang *et al.* [15] employed a semi-supervised method to find spammer groups and extract candidates using fequent items mining (FIM) in order to detect fake reviews.

Moreover, unsupervised learning algorithms provide useful options in situations where labeled data is rare. Huang and Liang [16] used unsupervised algorithms like k-means clustering to detect fake reviews based on semantic similarities. Gao *et al.* [17] suggested an unsupervised attention-based competitive model for detecting movie review spam, which outperformed existing approaches. Xu *et al.* [18] presented GSCPM, a Clique Percolation method-based approach for group spamer detection in online product reviews. These unsupervised learning algorithms provide insights into the underlying structure of reviews without the need for labeled data, making them a viable alternative to supervised methods. Table 1 summarizes machine learning approaches for detecting fake reviews.

Additionally, preprocessing methods like n-gram analysis, stop-word removal, and POS tagging [19] are essential for improving fake review detection. Additionally, methods such as text normalization with POS tagging [20] and using skip-gram models with SIF [21] enhance accuracy. Siagian and Aritsugi *et al.* [22] utilized various preprocessing steps to prepare data for ML models. These techniques, combined with modern ML approaches, are crucial for reliable fake review identification in competitive markets.

The next sections describe the proposed approach: in section 2 describe Word2Vec, similarity metric, exaggeration detection, cluster-based review assignment, and empirical results. Section 3 covers the findings and the discussion approach, which includes information, challenges, and perspectives. Section 4 summarizes the findings and emphasizes contributions.

Table 1. An overview of machine learning techniques for the detection of fake reviews

Approach	References	Features	Techniques	Dataset
Supervised learning	Taneja and Kaur [8]	Ensemble model	Bag-of-words and N-grams	Cloud armor dataset
	Wang *et al.* [9]	Multi-feature fusion	Sentiment analysis, review text, behavioral features	Custom dataset focused on customer electronics
	Aiyar and Shetty [10]	Custom heuristics	Character n-grams	Random forest, SVM, and NB
	Noekhah *et al.* [11]	Graph-based approach	Relationships between entities	data 4
Semi-supervised learning	Tian *et al.* [12]	K-means clustering	Semantic similarity	Cloud armor dataset
	Ligthart *et al.* [13]	Attention mechanism	Statistical components	Movie reviews
	Hassan and Islam [14]	Behavioral data	Clique percolation method	Random forest, SVM, and NB
	Zhang *et al.* [15]	Context and FIM	Extract spammer group candidates	
Unsupervised learning	Huang and Liang [16]	K-means clustering	Semantic similarity	Cloud armor dataset
	Gao *et al.* [17]	Attention mechanism	Statistical components	Movie reviews
	Xu *et al.* [18]	Behavioral data	Clique percolation method	Random forest, SVM, and NB

2. PROPOSED APPROACH

Our Python-based solution to detecting fake reviews has made a substantial contribution to the area. This section presents a rigorous process for distinguishing authentic reviews from fake ones. To address the numerous difficulties of language use, our multidimensional method employs modern natural language processing (NLP) tools, clustering algorithms, and rigorous linguistic evaluations.

2.1. Word2Vec: resolving semantic structure in review

To build carryout review vectors, Word2Vec, a powerful instrument for natural language comprehension by Mikolov *et al.* [23], Pennington *et al.* [24], is essential. It encodes semantic associations into vectors by learning from a review corpus. A preprocessing pipeline prepares reviews for Word2Vec's analysis, including text cleaning and stop word elimination [25].

By converting every word into a high-dimensional vector, Word2Vec goes beyond conventional bag-of-words models and captures contextual relationships. These word vectors can be used to create a single review vector that shows a range of human expression [26], [27]. The multidimensional representation improves the quality of the review by reflecting the complexity of the language ecosystem. Word2Vec can create dynamic representations that adapt to new linguistic patterns and are sensitive to the constantly changing language environment. Its application in reviewing exceeds computing; it sets out on a linguistic expedition, liberating semantic complexity and understanding the complex flow of meanings inside reviews.

2.2. The MRCS approach

Vectorization improves semantic understanding, and our approach uses cosine similarity to refine semantic navigation. This metric, ranging from -1 to 1, gives a normalized measure of textual proximity, addressing the limitations of traditional distance metrics such as Euclidean distance. Working with Word2Vec, it enables precise navigation through semantic context, recognizing obvious and hidden connections to context that indicate deceitful intention.

The recent metric similarity between two reviews, denoted as the mean review of Cosinus similarity $\mathrm{MRCS}\left(R_1, R_2\right)$, is formulated in (1):

$$\mathrm{MRCS}\left(R_1, R_2\right)=\left|\frac{\sum_{i=1}^{n} \sum_{j=1}^{m} \mathrm{CS}\left(W_i, W_j\right)}{n * m}\right|, i \neq j \tag{1}$$

where n present the amount of words in review R_1, m is the amount of words in review R_2, and $\mathrm{CS}\left(W_i, W_j\right)$ denotes the cosine similarity of the vectors of words W_i and W_j. This formula, which computes the cosine similarity between each word grouping in the two reviews, adds up the findings, and normalizes by the product of the word counts in each, captures the essence of our method.

We apply this formula not just to determine the textual closeness of reviews, but also to tailor our metric to identify misleading patterns that more basic similarity metrics may overlook. This comprehensive cosine similarity equation is included as a testament to our dedication to understanding language's finer points. Important aspects of the proposed MRCS metric increase its utility in review similarity analysis. The triangle inequality is preserved by MRCS, which is significant since it guarantees that the shortest path between two reviews always remains the shortest. This is crucial in situations when high processing efficiency is required. Mathematically, this is expressed as (2):

$$\mathrm{MRCS}\left(R_1, R_2\right) \leq \mathrm{MRCS}\left(R_1, R_3\right) + \mathrm{MRCS}\left(R_2, R_3\right) \tag{2}$$

The metric's symmetry, denoted as (3):

$$\mathrm{MRCS}\left(R_1, R_2\right) = \mathrm{MRCS}\left(R_2, R_1\right) \tag{3}$$

leads to reliable pairwise comparisons that produce equal results regardless of review order. The non-negativity property guarantees that the estimated dissimilarity is never negative (4):

$$\mathrm{MRCS}\left(R_1, R_2\right) \geq 0 \tag{4}$$

Moreover, a restricted range of $[-1, 1]$ is guaranteed by MRCS's dependence on cosine similarity. The cosine similarity formula (5) expresses how easily this range can be interpreted:

$$CS\left(W_i, W_j\right) = \frac{\sum_{k=1}^{d} W_{i,k} \cdot W_{j,k}}{\sqrt{\sum_{k=1}^{d} \left(W_{i,k}\right)^2} \cdot \sqrt{\sum_{k=1}^{d} \left(W_{j,k}\right)^2}} \tag{5}$$

The MRCS metric ranges from 0 to 1, with 1 indicating absolute similarity and 0 suggesting orthogonality between reviews. This scale is crucial for distinguishing structurally similar but semantically varied evaluations. A graphical representation (Figure 1) illustrates the MRCS calculation stages, guiding through vectorization, cosine angle computation, and MRCS derivation. This visual aids in understanding how this metric contributes to deep review analysis, facilitating fake behavior detection.

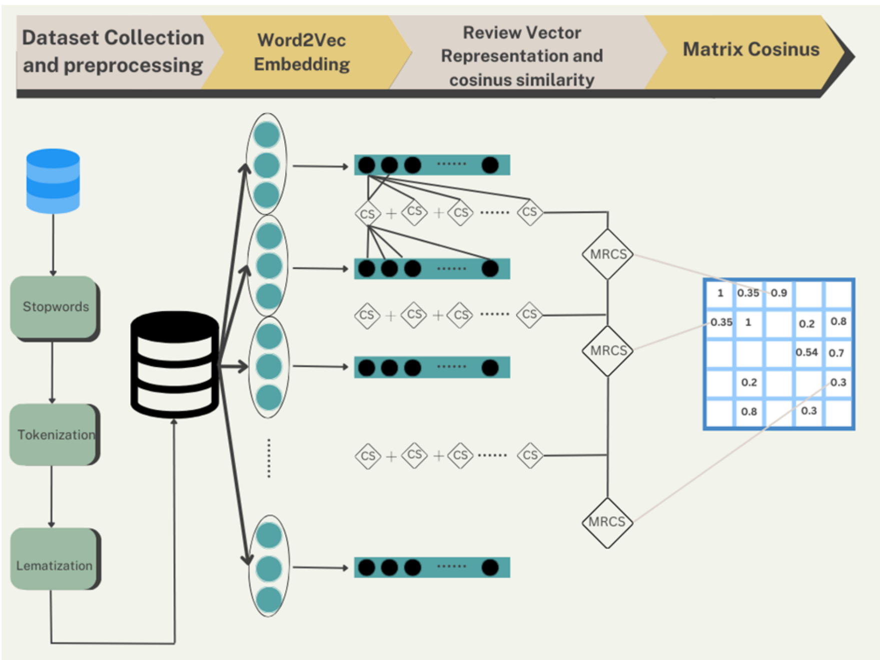

Figure 1. A graphical representation of the MRCS technique for detecting deceptive reviews

The innovative MRCS metric, leveraging cosine similarity and interacting with Word2Vec, offers a new method to review analysis and has the ability to uncover fake reviews. It enhances the understanding of semantic orientation patterns. MRCS is a key clustering direction, accurately matching genuine reviews and identifying deceptive patterns. The innovative analysis of MRCS transforms review assessments.

2.3. Exaggeration detection

Our solution employs an innovative exaggeration detection methodology based on Word2Vec vectorized forms. It captures extravagant words, giving a stylistic dimension to our research and revealing language intricacies suggestive of deceit. Extracting extravagant words aids in the detection of fake reviews and the understanding of evolving linguistic methods. Deceptive techniques purposefully use words to exaggerate, which are typically hidden in specific sentences. Our linguistic filter separates dishonesty from the reviews' rich lexicon.

This method makes use of a scoring system to identify real interest from misleading strategies by assigning scores to extravagant words, thus measuring exaggeration in reviews. We integrate multiple domains, such as an effective scoring system that takes into account the MRCS and the excessive word overlap within reviews, to create a comprehensive deception detection model. The (6) provide an equation for combining these scores:

$$\mathrm{ComS}\left(R_1, R_2\right) = \left(1 - \sigma_{EW}\right) \times MRCS\left(R_1, R_2\right) + \sigma_{EW}^2 \tag{6}$$

and

$$\sigma_{\mathrm{EW}} = \frac{EWO\left(R_1, R_2\right)}{\mathrm{n + m}} \tag{7}$$

where:
- ComS stands for combined score.
- The weight σ_{EW} is determined by the number of extravagant discrepancies between two reviews.
- EW stands for extravagant weight.
- The extravagant overlap (EWO) refers to the number of extravagant words that overlap between two evaluations.
- n represents the number of words in review R_1, m represents the number of words in review R_2.

This formula illustrates the delicate relationship between semantic similarity and linguistic extravagance by reflecting the combined effect of the MRCS score and extravagant word occurrence. The resulting composite score enhances our approach's capacity to detect specific dishonest behavior by incorporating many components of review analysis.

2.4. Cluster-based review assignment

In review clustering, our method differs from ordinary k-means by rethinking the process for adaptive precision. Unlike traditional k-means, which uses static centroid allocations, our adaptive technique iteratively refines cluster assignments to capture the dynamic nature of deceptive patterns. Traditional k-means may struggle with the variety of deceptive language use, however our dynamic clustering enables ongoing adaptation based on changing data, which is crucial in the case of fake reviews. This choice is based on recognizing the dynamic character of misleading operations and prohibiting the simplification or merging of deceptive and legitimate judgments. In cluster refining, computing new centroids is critical for responding to the linguistic complexities of each cluster. The mathematical expression for obtaining the new centroid, called NewCentroid (C), is given by (8):

$$\mathrm{NewCentroid}(C) = \max\left(\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\mathrm{ComS}\left(R_i, R_j\right)}{n-1}\right), i \neq j \tag{8}$$

where, n is the total amount of reviews within the cluster C, and $\mathrm{ComS}\left(R_i, R_j\right)$ is the combined score among the review vectors. R_i, R_j.

This formula's incorporation into our clustering technique enhances our approach's flexibility by enabling it to detect and adjust to the fine-grained linguistic changes seen in every cluster. Hence, adaptive accuracy is used to dynamically update cluster centroids in addition to detecting fraudulent language, improving our model for the constantly evolving online review environment.

Our adapted k-means approach, described in Algorithm 1, iteratively refines cluster assignments based on the highest similarity between reviews. Our approach improves the accuracy of detecting fake reviews by dynamically updating centroids and highlighting language signals indicating fraudulent intent. This adaptive technique ensures that our model tends to effectively represent the dataset's primary structure, capturing changing patterns of deceptive language over time.

Algorithm 1 Personalised k-means algorithm

```
Require: cluster1, cluster2, indexcentroid1, indexcentroid2, CosinesMatrix
Ensure: convergedclusters
Counter ⇐ 0
while true do
    for review in CosinusMatrix do
        Cluster1 ← review(index1)
        Cluster2 ← review(index2)
        if Cluster1 > Cluster2 then
            if review in Cluster2 then
                Remove review from Cluster2
                Append review to Cluster1
                Counter++
            end if
        else
            if review in Cluster2 then
                Remove review from Cluster1
                Append review to Cluster2
                Counter++
            end if
        end if
    end for
    if Counter = 0 then
        break
    end if
    indexCentroid1 ← Calculate New Centroid(Cluster1)
    indexCentroid2 ← Calculate New Centroid(Cluster2)
    centroid1 ← indexCentroid1
    centroid2 ← indexCentroid2
    Counter ← 0
end while
return Cluster1, Cluster2
```

Our updated approach enhances k-means adaptability to varying language aspects in fake review identification by prioritizing language signals indicating fraudulent intent. Adaptive weighting improves clustering precision, revealing deceptive patterns hidden in traditional k-means. Dynamic centroid updates ensure an accurate representation of evolving deceptive language patterns. This purposeful divergence elevates our technique to the forefront of adaptive precision in detecting fake reviews.

3. RESULTS AND DISCUSSION

The following section provides an extensive analysis of how well our technology distinguishes between genuine and fraudulent reviews.

3.1. Dataset description

The proposed model leverages the Yelp dataset, which consists of 1,600 reviews specifically curated to facilitate the detection of fake reviews. This dataset is a combination of genuine and deceptive reviews, making it ideal for training and evaluating our model. Table 2 illustrates the important properties of the Yelp labeled dataset, including the distribution of authentic and fake reviews.

Each review in the dataset is labeled, providing a clear distinction between genuine and fake reviews. TripAdvisor.com provides actual reviews of 20 popular hotels in the Chicago region, ensuring a realistic and diverse set of opinions. The fake reviews are generated by Amazon mechanical turk (AMT) workers, who were hired to create deceptive reviews for the same hotels. This dual nature of the dataset, with a balanced number of genuine and fake reviews, provides a robust basis for developing and testing the proposed model. By utilizing this dataset, our approach aims to capture the nuanced differences between genuine and deceptive reviews, thereby enhancing the accuracy and reliability of fake review detection.

Table 2. Overview of Yelp labeled dataset

Number of rows	Type	Source
400	Deceptive positive	Amazon
400	Truthful positive	TripAdvisor
400	Deceptive negative	Amazon
400	Truthful negative	TripAdvizor and yelp

3.2. Confusion matrix analysis

To gain a better understanding of the model's effectiveness, a confusion matrix and key assessment metrics are analyzed. The confusion matrix (Figure 2) summarizes 414 true positives, 51 false positives, and 21 false negatives, emphasizing the significance of memory and clarity in recognizing misleading linguistic difficulties.

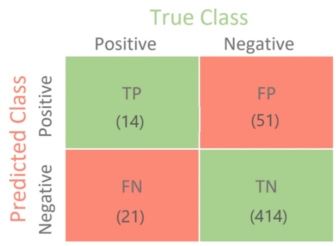

Figure 2. Confusion matrix

3.3. Model performance overview

The results give an intriguing overview of the model's predictions. While the 414 true positives (correctly detected authentic reviews) indicate the model's accuracy, the 51 false positives are instances in which false evaluations were misclassified as genuine. Similarly, the 21 false negatives show how honest evaluations were incorrectly identified as deceitful. These variances underline the importance of clarity and recall in understanding the complexity of deceptive language. Figure 3 shows the important metrics used to assess the model's performance.

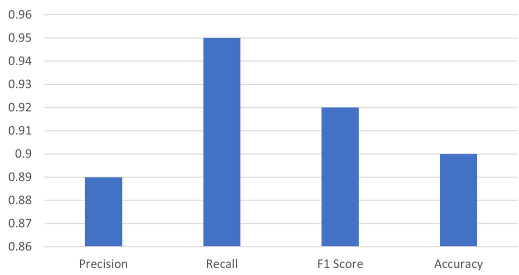

Figure 3. Key metrics diagram for model evaluation

The model's emphasis on false negatives and erroneous positives highlights the need for development even though its accuracy is outstanding. With an accuracy of roughly 89%, the model indicates that most expected genuine evaluations are, in fact, genuine. With a recall of approximately 95%, the model is deemed adequate for capturing real reviews; however, further refining may be necessary to minimize misclassifications.

This assessment influences our model enhancement efforts, demonstrating dedication to an optimized, adaptive approach. The ongoing pursuit of a more competent model aligns with the objective of accuracy in online review clustering. This inquiry into the identification about fake reviews handles the complex system of online interactions by evaluating existing approaches and presenting a multidimensional paradigm.

3.4. Contextual analysis

The emphasis is on developing a comprehensive strategy, as demonstrated by Word2Vec's ability to extract contextual information. The precision of the technique improves when the cosine similarity metric is used to navigate the semantic environment. In contrast to conventional distance measurements, the cosine similarity metric provides a normalized assessment of textual proximity that accounts for vector amplitude and direction. This measure serves as a discriminating lens through which the technique navigates the intricate semantic context of online reviews. It is almost exactly the same as Word2Vec.

The exaggeration identification method, which goes beyond conventional sentiment analysis, is one of the system's main contributions. The model may identify linguistic patterns that point to dishonest intent by eliminating extravagant word and employing a scoring system centred about cosine similarity and extravagant term correspondence. This method helps identify fraudulent reviews and encourages adaptive learning by keeping an eye out for changing language patterns.

A strategic choice was made to replace the classic k-means classification method with a flexible precision method which detects the fluidity of deceptive practices. The approach's adaptability is enhanced by iterative centroid creation, continuous cluster assignment optimization, and consideration of dynamic centroid modifications. The introduction of dynamic centroid updates and an improved approach to feature importance increase the precision of the clustering technique, enabling the model to detect deceptive patterns that were previously obscured by fewer significant linguistic features.

The Yelp labeled dataset was utilized to empirically evaluate the strategy and collect performance data. Despite the model's high accuracy, the argument shows that there is still potential for improvement, notably in dealing with false positives and erroneous negatives. The deliberate modification of the clustering process, as well as the use of linguistic precision instruments, illustrate the dedication to an adaptive strategy and the iterative nature of the search for a more nuanced effective model.

4. CONCLUSION

Recent findings in the field of fake review detection highlight the complexity of the problem and the need for innovative solutions. Our work has looked into a variety of approaches and methodologies, providing perspectives on the dynamic nature of fake review detection. We have demonstrated the efficacy of our comprehensive strategy, which combines advanced preprocessing approaches with powerful machine learning techniques. Our findings offer definitive proof that the proposed system, which includes Word2Vec and the MRCS metric, improves semantic detail and accuracy in detecting fake reviews. The inclusion of An exaggeration detection system adds a stylistic dimension to adaptive learning, allowing it to adjust to shifting verbal patterns and false intentions. Additionally, switching from classic k-means to an adaptive precision technique demonstrates a commitment to capturing dynamic, deceptive behaviors.

Empirical analysis using the Yelp labeled dataset highlights model accuracy and the ongoing need for improvement, serving as a benchmark for future work. This study lays the groundwork for future developments, emphasizing the need for more research, adaption to diverse domains, the incorporation of advanced linguistic factors, and collaborative initiatives to standardize datasets and evaluation criteria. The ever-changing digital ecosystem necessitates interdisciplinary partnerships to strengthen fake review detection systems toward future deceitful approaches.

REFERENCES

[1] P. Kalchgruber, W. Klas, N. Jnoub, and E. Momeni, "Factcheck - identify and fix conflicting data on the web," *Lecture Notes in Computer Science*, pp. 312–320, 2018, doi: 10.1007/978-3-319-91662-0_25.

[2] P. Yuruk-Kayapinar, "Digital Consumer Behavior in an Omnichannel World," *Managing Customer Experiences in an Omnichannel World: Melody of Online and Offline Environments in the Customer Journey*, pp. 55–73, Nov. 2020, doi: 10.1108/978-1-80043-388-520201007.

[3] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Distributed Representations of Words and Phrases and their Compositionality," *Neural information processing systems*, vol. 1, pp. 1–9, 2006.

[4] S. Banerjee, "Exaggeration in fake vs. authentic online reviews for luxury and budget hotels," *International Journal of Information Management*, vol. 62, p. 102416, Feb. 2022, doi: 10.1016/j.ijinfomgt.2021.102416.

[5] "yelp." https://www.yelp.com/

[6] N. Jindal and B. Liu, "Analyzing and detecting review spam," *in Proceedings - IEEE International Conference on Data Mining, ICDM, IEEE*, Oct. 2007, pp. 547–552, doi: 10.1109/ICDM.2007.68.

[7] N. Jindal and B. Liu, "Opinion spam and analysis," *WSDM'08 - Proceedings of the 2008 International Conference on Web Search and Data Mining*, pp. 219–229, 2008, doi: 10.1145/1341531.1341560.

[8] H. Taneja and S. Kaur, "An ensemble classification model for fake feedback detection using proposed labeled CloudArmor dataset," *Computers and Electrical Engineering*, vol. 93, 2021, doi: 10.1016/j.compeleceng.2021.107217.

[9] 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," *IEEE Access*, vol. 8, pp. 182625–182639, 2020, doi: 10.1109/ACCESS.2020.3028588.
[10] S. Aiyar and N. P. Shetty, "N-Gram Assisted Youtube Spam Comment Detection," *Procedia Computer Science*, vol. 132, pp. 174–182, 2018, doi: 10.1016/j.procs.2018.05.181.
[11] S. Noekhah, N. Salim, and N. H. Zakaria, "Opinion spam detection: Using multi-iterative graph-based model," *Information Processing and Management*, vol. 57, no. 1, 2020, doi: 10.1016/j.ipm.2019.102140.
[12] Y. Tian, M. Mirzabagheri, P. Tirandazi, and S. M. H. Bamakan, "A non-convex semi-supervised approach to opinion spam detection by ramp-one class SVM," *Information Processing and Management*, vol. 57, no. 6, 2020, doi: 10.1016/j.ipm.2020.102381.
[13] A. Ligthart, C. Catal, and B. Tekinerdogan, "Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification," *Applied Soft Computing*, vol. 101, 2021, doi: 10.1016/j.asoc.2020.107023.
[14] R. Hassan and M. R. Islam, "Detection of fake online reviews using semi-supervised and supervised learning," *in 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), IEEE*, Feb. 2019, pp. 1–5, doi: 10.1109/ECACE.2019.8679186.
[15] L. Zhang, Z. Wu, and J. Cao, "Detecting Spammer Groups from Product Reviews: A Partially Supervised Learning Model," *IEEE Access*, vol. 6, pp. 2559–2568, 2017, doi: 10.1109/ACCESS.2017.2784370.
[16] G. Huang and H. Liang, "Uncovering the effects of textual features on trustworthiness of online consumer reviews: A computational-experimental approach," *Journal of Business Research*, vol. 126, pp. 1–11, 2021, doi: 10.1016/j.jbusres.2020.12.052.
[17] Y. Gao, M. Gong, Y. Xie, and A. K. Qin, "An Attention-Based Unsupervised Adversarial Model for Movie Review Spam Detection," *IEEE Transactions on Multimedia*, vol. 23, pp. 784–796, 2021, doi: 10.1109/TMM.2020.2990085.
[18] G. Xu, M. Hu, C. Ma, and M. Daneshmand, "GSCPM: CPM-Based Group Spamming Detection in Online Product Reviews," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC), IEEE*, May 2019, pp. 1–6, doi: 10.1109/ICC.2019.8761650.
[19] N. Jnoub and W. Klas, "Declarative Programming Approach for Fake Review Detection," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization, IEEE*, Oct. 2020, pp. 1–7, doi: 10.1109/SMAP49528.2020.9248468.
[20] S. J. Nejad, F. Ahmadi-Abkenari, and P. Bayat, "Opinion Spam Detection based on Supervised Sentiment Analysis Approach," *2020 10h International Conference on Computer and Knowledge Engineering, ICCKE 2020*, pp. 209–214, 2020, doi: 10.1109/ICCKE50421.2020.9303677.
[21] Y. Liu, B. Pang, and X. Wang, "Opinion spam detection by incorporating multimodal embedded representation into a probabilistic review graph," *Neurocomputing*, vol. 366, pp. 276–283, 2019, doi: 10.1016/j.neucom.2019.08.013.
[22] A. H. A. M. Siagian and M. Aritsugi, "Robustness of Word and Character N-gram Combinations in Detecting Deceptive and Truthful Opinions," *Journal of Data and Information Quality*, vol. 12, no. 1, 2020, doi: 10.1145/3349536.
[23] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceeding*s, 2013.
[24] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, 2014, pp. 1532–1543, doi: 10.3115/v1/d14-1162.
[25] Q. Le and T. Mikolov, "Distributed representations of sentences and documents," *31st International Conference on Machine Learning, ICML 2014*, vol. 4, pp. 2931–2939, 2014.
[26] Y. Goldberg and O. Levy, "word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method," *arXiv preprint arXiv:1402.3722*, 2014.
[27] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching Word Vectors with Subword Information," *Transactions of the Association for Computational Linguistics*, vol. 5, pp. 135–146, Dec. 2017, doi: 10.1162/tacl_a_00051.

BIOGRAPHIES OF AUTHORS

Mohammed Ennaouri is a senior researcher in the High National School for Computer Science and Systems Analysis (ENSIAS), Rabat, Morocco. He received the Master's degree in Internet of Things: Software and Analytics in 2021. He is currently working as a teacher in secondary school. His research interest includes fake news, machine learning algorithms, and recommender systems. He can be contacted at email: mohammed_ennaouri@um5.ac.ma.

Ahmed Zellou received his Ph.D. in Applied Sciences at the Mohammedia School of Engineers, Mohammed V University, Rabat, Morocco 2008, his habilitation to supervise research work in 2014. He becomes full professor in 2021. His research interests include interoperability, mediation systems, distributed computing, data quality, and semantic web where he is the author/co-author of over 100 research publications. He can be contacted at email: ahmed.zellou@um5.ac.ma.