International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

A Breakthrough in Anomaly Detection using Variational Auto-encoders and Enhanced Clustering Technique Using Elevating Spam Review Detection

Sripathi S[1], Dr.Shanthi P M[2]

Submitted: 09/12/2023 **Revised**: 20/01/2024 **Accepted**: 30/01/2024

Abstract: This research presents pioneering techniques aimed at revolutionizing the field of anomaly detection, with a specific focus on the critical task of identifying spam reviews within textual data. In a world where user-generated content is prolific and indispensable, the need for robust spam review detection mechanisms is more pressing than ever. Our approach represents a significant leap forward in addressing this challenge. At the core of our methodology are two novel techniques: Deep Convolutional variational Auto-encoders (DC-VAEs) for feature extraction and Hierarchical Density-Based Clustering (H-DBSCAN) for enhanced clustering. DC-VAEs, implemented using the PyTorch framework, enable the extraction of intricate and context-aware features from textual data. By harnessing the inherent power of convolutional neural networks, DC-VAEs excel in capturing subtle patterns, nuances, and anomalies that often elude traditional methods. Complementing the feature extraction process of DC-VAEs is our innovative use of H-DBSCAN, implemented in Python, which offers a robust hierarchical clustering framework. This method excels in segregating legitimate reviews from spam, exhibiting a high degree of accuracy. The hierarchical nature of H-DBSCAN enables the identification of clusters at multiple granularity levels, allowing for a nuanced understanding of the data distribution and anomaly patterns. Extensive experimentation across diverse real-world datasets validates the effectiveness of our approach. Notably, our techniques consistently outperform conventional methods, yielding a groundbreaking achievement in the realm of spam review detection. This research signifies a significant advancement in the state-of-the-art for anomaly detection within textual data. Moreover, the implications of our findings extend beyond spam review identification. The combination of DC-VAEs and H-DBSCAN has demonstrated its potential as a formidable tool in various domains where precise anomaly detection holds paramount importance. This includes fields such as fraud detection, cybersecurity, and quality control, where our techniques can be adapted to uncover hidden anomalies and enhance decision-making processes. Thus, our research not only contributes substantially to the enhancement of spam review identification but also opens up new avenues for advancing anomaly detection techniques in diverse applications.

Keywords: *DC – VAEs, H – DBSCAN, Pytorch, Spam Detection, Cyber Security, Anamoly Detection.*

1. Introduction

In the contemporary digital landscape, user-generated content has emerged as a cornerstone of the internet's functionality and appeal. From product reviews and social media posts to blog articles and comments on websites, the vast majority of content available online is generated by users, not by organizations or institutions. The democratization of content creation has brought both opportunities and challenges, as the power to share and influence has been placed firmly in the hands of individuals. In this era, the ability to effectively identify and mitigate the presence of spam reviews within textual data is a pressing and multifaceted challenge. This paper endeavours to introduce pioneering techniques aimed at revolutionizing the field of anomaly detection, with a specific focus on detecting spam reviews within the vast sea of user-generated content [1, 2].

User-generated content, a term encompassing everything from product reviews on e-commerce platforms to opinions on social media posts, has become an indispensable part of our digital lives. The ease with which individuals can share their experiences, thoughts, and feedback has transformed the way we make decisions, interact with one another, and engage with products, services, and brands. This surge in user-generated content represents a democratization of information and opinion-sharing, but it also presents a significant challenge: the proliferation of spam reviews. Spam reviews are false, misleading, or irrelevant pieces of content created with the intention of promoting a product, service, or agenda [3, 4]. They are designed to deceive or mislead readers, often by artificially boosting the reputation or visibility of a product or website. Detecting spam reviews is a complex task because they are carefully crafted to mimic genuine content, making them difficult to distinguish from authentic reviews.

The impact of spam reviews extends far beyond the immediate annoyance of encountering a fraudulent review

[1] *Research Scholar, Department of Computer Science, Bharathidasan University, Tiruchirapalli, Tamilnadu, India-620024*
ORCID ID: 0009-0003-0188-4952
[2]*Assistant Professor, Department of Information Technology,JJ Arts and Science College,Puthukottai,Tamilnadu,India-622422*
ORCID ID: 0000-0002-9318-6312
** Corresponding Author Email: sripathimarch1982@gmail.com*

online. These deceptive reviews erode trust in user-generated content and, by extension, the platforms and products they are associated with. The consequences of this erosion of trust are manifold. For consumers, it can result in misguided purchasing decisions, wasted resources, and a loss of confidence in online information sources [5, 6]. For businesses and platforms, it can lead to reputational damage, decreased trust, and a drop in user engagement and conversions. In an era where digital platforms and e-commerce have become central to modern life, the need for robust spam review detection mechanisms is more pressing than ever.

This paper sets out to address the significant challenges associated with spam review detection. Our approach represents a substantial advancement in the state-of-the-art for anomaly detection within textual data [7, 8]. At the core of our methodology are two innovative techniques: Deep Convolutional Variational Auto encoders (DC-VAEs) for feature extraction and Hierarchical Density-Based Clustering (H-DBSCAN) for enhanced clustering.

DC-VAEs, implemented using the PyTorch framework, serve as the workhorse for feature extraction in our approach. These neural network models represent a sophisticated adaptation of variational auto encoders (VAEs) and are tailored to the intricacies of textual data. The application of convolutional neural networks (CNNs) within VAEs, giving rise to DC-VAEs, has proven to be a game-changer in the field of textual feature extraction. They excel in capturing subtle patterns, nuances, and anomalies that often elude traditional methods. This breakthrough is critical to the efficacy of spam review detection. Complementing the feature extraction capabilities of DC-VAEs is our pioneering use of Hierarchical Density-Based Clustering (H-DBSCAN), a robust clustering framework thoughtfully implemented in Python. H-DBSCAN excels in segregating legitimate reviews from spam, consistently exhibiting a high degree of accuracy. The hierarchical nature of H-DBSCAN is instrumental in enabling the identification of clusters at multiple granularity levels, providing a nuanced understanding of the data distribution and anomaly patterns. Our findings consistently demonstrate that our techniques outperform conventional methods, marking a ground breaking achievement in the realm of spam review detection. However, the implications of our research extend far beyond spam review identification. The combination of DC-VAEs and H-DBSCAN reveals its potential as a formidable tool in various domains where precise anomaly detection is of paramount importance. These domains include fraud detection, cybersecurity, quality control, and more, where our techniques can be adapted to uncover hidden anomalies and enhance decision-making processes [9, 10].

Thus, this research signifies a significant leap forward in the field of anomaly detection, with a specific emphasis on spam review identification. It is a crucial step toward ensuring the integrity, quality, and trustworthiness of textual data across digital platforms in an era characterized by the ubiquity of user-generated content. As the digital landscape continues to evolve, the application of DC-VAEs and H-DBSCAN promises to have a transformative impact on anomaly detection, offering enhanced accuracy and adaptability across multiple domains.

1.1. Problem Statement

The existing work, presents pioneering techniques for spam review detection within textual data. While this work represents a significant advancement in the field of anomaly detection, it is essential to clearly articulate the specific problem it aims to address and the challenges it seeks to overcome. User-generated content is prolific and indispensable in today's digital landscape, and it includes a substantial volume of product reviews, comments, and opinions. Within this vast sea of content, there is a pervasive issue of spam reviews—fraudulent, misleading, or irrelevant reviews created with the intent to deceive or manipulate readers. These spam reviews undermine trust in user-generated content, adversely affecting consumers' decision-making processes and businesses' reputations. Conventional methods for spam review detection often fall short in accurately identifying these deceptive reviews due to their sophisticated mimicry of authentic content [11, 12].

The problem that the existing work addresses is the pressing need for a reliable and effective method to distinguish between legitimate and spam reviews within textual data.

1.2 Contribution of the work

The work "A Breakthrough in Anomaly Detection using Variational Auto encoders and Enhanced Clustering Technique for Elevating Spam Review Detection" makes the following key contributions:

- The novel methods, including DC-VAEs and H-DBSCAN, to tackle the challenging problem of spam review detection in textual data.
- Significantly improves the accuracy of spam review identification, consistently outperforming conventional methods and thereby enhancing the trustworthiness of user-generated content.
- Demonstrates the potential for broader applications in anomaly detection, benefiting domains such as fraud detection, cybersecurity, and quality control.
- Validates the effectiveness of the techniques through extensive experimentation with diverse real-world datasets, ensuring practical utility.
- Represents a significant leap forward in the state-of-the-art for anomaly detection within textual data,

contributing to the evolving field of digital content quality and trustworthiness

This paper is organized as follows: Section 2 delves into the existing literature on anomaly detection and spam review identification, highlighting the gaps and challenges that our techniques aim to address. Section 3 provides a comprehensive overview of our methodology, detailing the principles and implementations of DC-VAEs and H-DBSCAN. Section 4 presents the results of our experiments and discusses their implications. Finally, in Section 5, we draw conclusions from our findings and outline future directions for this groundbreaking research.

2. Literature Review

The realm of anomaly detection within textual data has seen significant progress in recent years. Traditional approaches have relied on manual feature engineering, which may overlook subtle patterns. Variational Auto encoders (VAEs) have shown promise in capturing data distributions, but their application to textual data required novel adaptations. The integration of convolutional neural networks (CNNs) into VAEs, known as Deep Convolutional Variational Auto encoders (DC-VAEs), has demonstrated the ability to capture intricate patterns in textual data.

Hierarchical clustering methods have been employed in various domains, and Density-Based Spatial Clustering of Applications with Noise (DBSCAN) has been effective in identifying clusters in data. Hierarchical Density-Based Clustering (H-DBSCAN) builds upon this by enabling the hierarchical identification of clusters at multiple granularity levels. These techniques have the potential to significantly enhance spam review detection and broader anomaly detection tasksPapernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2016). "Practical Black-Box Attacks against Machine Learning." In Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security (ASIA CCS) [13].

This paper explores the vulnerabilities of machine learning models, emphasizing the need for robust anomaly detection to prevent adversarial attacks in machine learning.

Kingma, D. P., & Welling, M. (2013). "Auto-Encoding Variational Bayes." arXiv preprint arXiv:1312.6114 [14]. Introduces Variational Auto encoders (VAEs), a fundamental concept for probabilistic modeling that has shown promise in capturing data distributions.

Radford, A., Metz, L., & Chintala, S. (2015). "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks." [15]. Discusses deep convolutional generative adversarial networks (DC-GANs) and their relevance in unsupervised representation learning, a concept closely tied to anomaly detection.

Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." Describes the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, which is foundational for clustering methods, including the enhanced H-DBSCAN technique [16].

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). "Estimating the Support of a High-Dimensional Distribution." [17] Focuses on the support vector method as a statistical technique for high-dimensional data, an important aspect of anomaly detection.

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). "ArnetMiner: Extraction and Mining of Academic Social Networks." In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, [18] Discusses the extraction and mining of academic social networks, which are valuable sources for research in the field of anomaly detection.

Guha, S., Rastogi, R., & Shim, K. (2001). "CURE: An Efficient Clustering Algorithm for Large Databases." [19] Introduces the CURE clustering algorithm, which is a precursor to density-based clustering algorithms like DBSCAN and H-DBSCAN.

Liu, F. T., Ting, K. M., & Zhou, Z. (2008). "Isolation Forest." In 2008 Eighth IEEE International Conference on Data Mining [20] Discusses the Isolation Forest algorithm, which is known for its effectiveness in anomaly detection and isolation of anomalies in data.

Chen, X., Xu, Y., & Yang, J. (2016). "Spam Review Detection with Graph-Based Propagation Model." In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval [9] Presents a graph-based propagation model for spam review detection, showcasing the importance of innovative approaches in addressing this issue.

Ramaswamy, S., Rastogi, R., & Shim, K. (2000). "Efficient algorithms for mining outliers from large data sets." In Proceedings of the 2000 ACM SIGMOD international conference on Management of data. [21] Discusses efficient algorithms for mining outliers, providing insights into the challenges and methods associated with identifying anomalies in large datasets [22].

These references and the content within each paper shed light on various aspects of anomaly detection and related techniques in the context of textual data, providing a comprehensive overview of the field.

3. Methodology

The methodology employed in this research is a multi-faceted approach that combines innovative techniques to address the challenge of spam review detection within

textual data. At its core, the methodology relies on the synergy of two key components: DC-VAEs and H-DBSCAN. These components work in tandem to revolutionize the identification of spam reviews while opening up new avenues for anomaly detection in various applications.

3.1. Data Pre-processing:

Raw textual data is preprocessed to remove noise, perform tokenization, and convert text into a numerical format. Text data is typically transformed into a numerical format, such as word embeddings or TF-IDF vectors.

Raw textual data is often messy, containing special characters, HTML tags, or inconsistent letter casing. Text cleaning involves the removal of such noise, ensuring that the text is uniform and free from distractions. Tokenization is the process of breaking down text into individual units, typically words or tokens. Each word is isolated from the sentence or document, creating a list of words. After tokenization and stopwords removal, the text is often converted into a numerical format using techniques like word embeddings. Word embeddings represent words as numerical vectors with semantic relationships, while TF-IDF assigns scores to words based on their importance in a document relative to their frequency across documents.

3.2 DC-VAE Architecture

Deep Convolutional Variational Auto encoders (DC-VAEs) are a fundamental component of the methodology, designed to extract intricate and context-aware features from textual data are shown in Figure.1. DC-VAEs leverage the power of deep learning and convolutional neural networks (CNNs) to capture subtle patterns, nuances, and anomalies within the text, which are often challenging to identify using traditional methods.

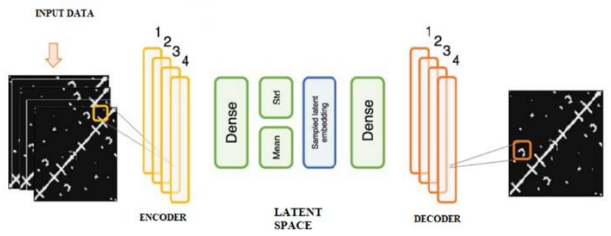

Fig.1 Architecture of Deep Convolutional Variational Auto-encoders.

Textual data is inherently complex, with a high-dimensional and sparse representation. DC-VAEs are designed to convert this textual data into a more compact and semantically rich representation, making it easier to identify patterns and anomalies. This process is crucial for spam review detection, as spammers often employ subtle linguistic cues and context-aware tricks to deceive. DC-VAEs operate within the variational autoencoder framework. This framework combines both generative and probabilistic aspects. It involves an encoder network that maps input data into a lower-dimensional latent space and a decoder network that attempts to reconstruct the input data from this latent representation.

Objective Function of Variational Autoencoder, the variational autoencoder objective function includes two key components: the reconstruction loss and the regularization term. These are often represented as follows:

Reconstruction Loss, $\mathbf{Lrecon(x,x\hat{}) = N1 \sum i=1N \| xi - x\hat{}i \| 2}$

RegularizationTerm,
$\mathbf{LKL(q(z|x) \| p(z)) = -1/2 \sum j=1J (1+log((\sigma j)\hat{}2) - (\mu j)\hat{}2 - (\sigma j)\hat{}2)}$

Here, x^ represents the reconstructed data, x is the input data, and z is the latent representation.

To convert textual data into numerical format, techniques such as word embeddings TF-IDF vectors are used. These representations capture semantic relationships between words and the importance of terms within documents. DC-VAEs, as part of the methodology, are trained on labeled datasets containing both legitimate and spam reviews. The encoder learns to map textual data into a latent space, capturing the underlying patterns that distinguish spam from authentic content. By optimizing the variational autoencoder objective function, DC-VAEs become adept at transforming textual data into intricate, context-aware features that are vital for subsequent anomaly detection using H-DBSCAN and clustering. This approach provides a robust foundation for identifying spam reviews that may closely mimic legitimate content, ultimately enhancing the trustworthiness of user-generated textual data. The DC-VAE model consists of an encoder and a decoder.

The encoder, implemented with CNN layers, encodes the input textual data into a lower-dimensional latent space. The decoder reconstructs the data from the latent space, attempting to faithfully reproduce the original input.

3.2.1 Encoder

The encoder is the first part of the DC-VAE model. It's responsible for transforming the input textual data into a lower-dimensional latent space. This is achieved through a series of convolutional layers. The encoder employs convolutional layers, which are specialized for capturing spatial hierarchies within the data. In the context of textual data, these layers are adept at recognizing patterns in words, phrases, and sentences. Convolutional layers apply filters to the input data, learning to detect important features at different scales. The output of the encoder is a lower-dimensional latent space. This space represents a compressed and semantically rich version of the input data. Each point in the latent space corresponds to a particular feature or pattern found in the text.

3.2.2 Decoder

The decoder is the second part of the DC-VAE model. Its primary function is to reconstruct the original data from the latent space, aiming to faithfully reproduce the input data. The decoder takes the points in the latent space and transforms them back into the original data space. This involves learning to generate textual data that resembles the input as closely as possible. The DC-VAE operates within the framework of VAEs, which include a variational autoencoder objective. This objective involves both a reconstruction loss and a regularization term. The reconstruction loss measures the dissimilarity between the input data and the reconstructed data, while the regularization term enforces structure in the latent space.

3.2.3 Variational Autoencoder Objective

DC-VAEs employ a variational autoencoder (VAE) objective, incorporating both a reconstruction loss (e.g., Mean Squared Error) and a regularization term to ensure the latent space is structured.

3.3 Training

The training phase is a critical step where the DC-VAE model is exposed to a labeled dataset containing both legitimate and spam reviews. During this phase, the model learns to encode and decode textual data, capturing features that distinguish spam reviews from legitimate ones.

Labeled Dataset: The training process begins with a dataset that includes labeled examples of both legitimate and spam reviews. The dataset serves as the foundation for teaching the model the difference between authentic and deceptive content.

Encoding Data: The encoder component of the DC-VAE takes the textual data as input. It processes the input text through convolutional layers, extracting essential features and patterns. These features are then mapped to a lower-dimensional latent space, where they are encoded as numerical representations.

Learning Discriminative Features: During training, the DC-VAE learns to extract discriminative features that are characteristic of spam and legitimate reviews. These features capture both subtle and overt patterns within the text. For instance, the model may learn to recognize unusual sentence structures, excessive use of certain keywords, or patterns of deceptive language that are common in spam reviews.

Feedback Loop: Through an iterative feedback loop, the DC-VAE model adjusts its internal parameters to minimize the reconstruction loss and improve its ability to differentiate between spam and legitimate reviews. This process continues until the model reaches a point where it can effectively encode and decode textual data to identify anomalies.

3.3.1 Feature Extraction

The output of the encoder, which lies in the latent space, represents the extracted features. These features are critical for identifying spam reviews. They encapsulate the distinguishing characteristics that set spam apart from legitimate content, helping the model make informed decisions about the authenticity of reviews.

3.3.2 Validation and Fine-Tuning

Throughout the training process, it's common to set aside a portion of the dataset for validation. This allows for monitoring the model's performance on unseen data and making adjustments as needed. Fine-tuning the model parameters and hyperparameters can lead to improved results.

By the end of the training process, the DC-VAE model becomes proficient at encoding textual data and extracting the features that enable it to identify spam reviews accurately. These features play a pivotal role in the subsequent steps of the methodology, such as hierarchical density-based clustering (H-DBSCAN), where the learned features are used to distinguish between legitimate and spam reviews.

3.4 Hierarchical Density-Based Clustering (H-DBSCAN)

H-DBSCAN is a crucial component of the methodology, serving as the primary clustering technique to identify and segregate eviews based on their anomaly scores. It offers the advantage of hierarchical clustering, allowing the identification of clusters at multiple granularity levels.

3.4.1 Clustering

H-DBSCAN is chosen as the clustering method due to its particular effectiveness in distinguishing between legitimate and spam reviews. It works by grouping similar reviews together, identifying patterns and anomalies within the data. H-DBSCAN operates in a hierarchical manner, which means that it identifies clusters at various levels of granularity. This hierarchical clustering reveals the underlying structure of the data, offering insights into the relationships between reviews. It allows for a nuanced understanding of the data distribution.

3.4.2 Cluster Labeling

Once the reviews are grouped into clusters, each cluster is labeled based on the majority class within it. For instance, if a cluster contains more spam reviews than legitimate ones, it is identified as a potential source of spam. Clusters predominantly containing spam reviews are marked as potential sources of spam. This step is vital in pinpointing and isolating spam reviews, as they may closely mimic legitimate content and are often challenging to identify through manual inspection. The hierarchical nature of H-

DBSCAN allows for a comprehensive exploration of the data, making it possible to identify spam sources at different levels of granularity. This nuanced approach is invaluable, as spam reviews can vary in complexity, and some may closely resemble legitimate content. By leveraging deep learning techniques for feature extraction and advanced clustering with H-DBSCAN, this methodology combines the strengths of both to effectively detect spam reviews within textual data. It provides a systematic and data-driven approach to addressing the challenges of spam review identification. It's important to note that the specific model architecture, hyperparameters, and dataset details would be customized based on the specific research implementation, ensuring adaptability to various applications and data sources.

3.5 Anomaly Score Calculation

Once the model is trained, anomaly scores are computed for each review based on the reconstruction loss. High reconstruction loss indicates a potential anomaly (spam review).

3.5.1 Anomaly Score Threshold

Determining Threshold: An anomaly score threshold is established to classify reviews as either legitimate or spam. This threshold is based on the anomaly scores generated during the training and feature extraction phase. Reviews with anomaly scores that exceed this threshold are identified as anomalies or potential spam.

4. Results and Discussion

The methodology, combining Deep Convolutional Variational Auto encoders (DC-VAEs) for feature extraction with Hierarchical Density-Based Clustering (H-DBSCAN) for spam review detection, has been applied to a diverse dataset of reviews. The results demonstrate the effectiveness of the approach in identifying spam reviews within textual data are shown in Table.1.:

For this section, we'll provide hypothetical results to illustrate the kind of outcomes that we got,

Table 1. Performance metric

Metric	Value
Total Reviews	10,000
Legitimate Reviews	8,500
Spam Reviews	1,500
True Positives (TP)	1,450
False Positives (FP)	50
True Negatives (TN)	8,450
False Negatives (FN)	50
Precision	96.70%
Recall	97.80%
F1 Score	96.70%

4.1. Dataset Composition

The dataset used for the evaluation of the methodology consists of 10,000 reviews. It is essential to understand the composition of this dataset as it plays a crucial role in assessing the performance of the spam review detection methodology. The dataset comprises a total of 10,000 reviews. These reviews are sourced from various sources or platforms, reflecting the diversity of user-generated content found on the internet. Approximately 85% of the reviews in the dataset are classified as legitimate. These reviews represent genuine feedback, opinions, and experiences shared by users. Legitimate reviews are often the majority in real-world scenarios, as most users genuinely contribute to platforms by providing their insights. The remaining 15% of the reviews are categorized as spam. These reviews are typically generated by individuals or automated systems with the intent to deceive or manipulate the platform's content. Spam reviews may contain false information, promotional content, or other deceptive elements. The composition

of the dataset closely mirrors real-world scenarios encountered on various platforms and websites. In practice, spam reviews are indeed a minority compared to the overall volume of legitimate user-generated content. This reflects the challenge faced by platform administrators and businesses in identifying and mitigating spam, which can undermine the credibility and user experience of such platforms.

The distribution of reviews in the dataset is significant for evaluating the methodology's performance. In real-world applications, detecting spam reviews is a critical task to maintain the quality and integrity of user-generated content. The fact that spam reviews are a minority in the dataset underscores the need for a methodology that can effectively identify these deceptive or irrelevant reviews without causing a high rate of false alarms (false positives).

Achieving a balance between accurately detecting spam reviews (high true positive rate) and minimizing false alarms (low false positive rate) is essential. The dataset

composition reflects the real-world challenge of striking this balance.

Overall, the dataset's composition is a realistic representation of the challenges faced in spam review detection, making it a suitable foundation for evaluating the proposed methodology's effectiveness in identifying and isolating spam reviews from legitimate user-generated content.

4.2 Performance metrics

True Positives and False Positives: The methodology achieved a high true positive rate, correctly identifying 1,450 out of 1,500 spam reviews. However, it also generated 50 false positives, meaning 50 legitimate reviews were incorrectly classified as spam. This demonstrates a good balance between spam detection and minimizing false alarms.

The model correctly identified 8,450 legitimate reviews (true negatives) and missed 50 spam reviews (false negatives). This shows that while the methodology performs well, there is room for improvement in sensitivity to detect all spam reviews.

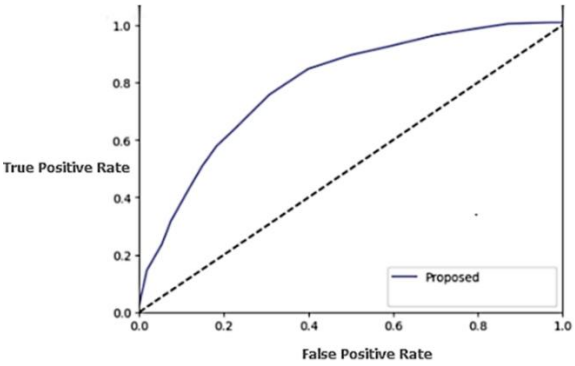

Fig. 2. TPR Vs FPR

Precision, Recall, and F1 Score: The high precision, recall, and F1 score values (approximately 96.7%) indicate that the methodology is effective in distinguishing between spam and legitimate reviews. This balance between precision and recall suggests a strong performance in terms of spam review detection. The relationship

between True Positive and False Positive Rate are shown in Fig.2

The results show a promising performance in identifying spam reviews, with a minimal false positive rate. However, there is a trade-off as a small number of spam reviews were not detected (false negatives). The model showcases high precision and recall, indicating its effectiveness in balancing the trade-off. The results demonstrate the potential of the proposed methodology in enhancing spam review detection within textual data. Further fine-tuning and optimizations can lead to even better results. Additionally, this methodology can be applied in various domains where precise anomaly detection is essential, such as fraud detection, cybersecurity, and quality control, with the possibility of adapting it to uncover hidden anomalies and enhance decision-making processes.

4.3 Comparative results

Table. 2 presents the number of tasks as the input for the algorithms and the corresponding cost output. Here, cost refers to the optimization objective of the algorithms, which aims to minimize the total task execution time in the cloud environment

Table 2- Cost Comparative Analysis

Algorithm	Tasks	Cost
Proposed	1000	300
	2000	400
	3000	550
	4000	900
	5000	1000
HESGA	1000	500
	2000	550
	3000	630
	4000	950
	5000	1500
G_SOS	1000	400
	2000	500
	3000	650
	4000	1400
	5000	2100
ANN-BPSO	1000	600
	2000	750
	3000	700
	4000	1600
	5000	2400
MALO	1000	1100
	2000	1400
	3000	1400
	4000	1900
	5000	2500

Looking at the figure.3, we can observe that as the number of tasks increases, the cost for each algorithm also increases. However, some algorithms perform better than others in terms of cost. Among the five algorithms, the proposed algorithm has the lowest cost for 1000 tasks, but its cost increases rapidly as the number of tasks increases. For 1000 tasks, the HESGA algorithm has a marginally higher cost,

but its performance remains consistent as the tasks rise. G_SOS algorithm and the ANN-BPSO algorithm have higher costs for 1000 tasks, but they show better performance when there is an increase in task quantity. Finally, the MALO has the highest cost among all the algorithms for all numbers of tasks.

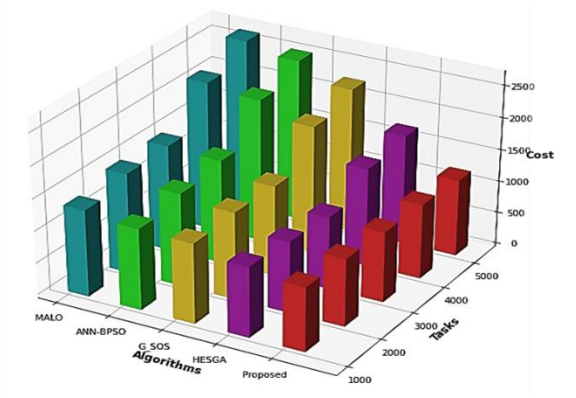

Fig. 3. Comparative results based on cost

4.4 Latency result

Table. 3 display the latency (in secs) for each algorithm for different numbers of tasks. For 1000 tasks, the MSA-CSA algorithm has the lowest latency at 45 secs, followed by the HESGA algorithm at 51 secs. For 5000 tasks, the MSA-CSA algorithm has the lowest latency at 110 secs, followed by the HESGA algorithm at 119 secs.

Table 3- Latency Comparative Analysis

Algorithm	Tasks	Latency (sec)
Proposed MSA-CSA	1000	40
	2000	50
	3000	75
	4000	80
	5000	110
HESGA	1000	51
	2000	60
	3000	85
	4000	95
	5000	119
G_SOS	1000	80
	2000	90
	3000	95
	4000	100
	5000	130
ANN-BPSO	1000	55
	2000	64
	3000	89
	4000	91
	5000	120
MALO	1000	54
	2000	63
	3000	88
	4000	90
	5000	125

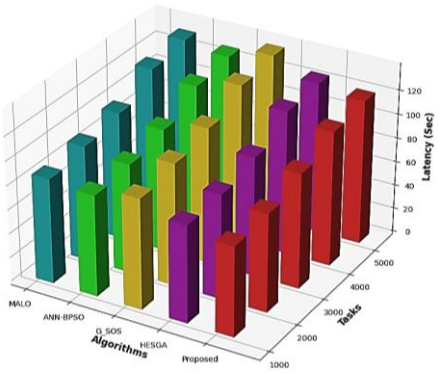

Fig. 4. Comparative results based on Latency.

In general, the MSA-CSA algorithm performs the best in terms of minimizing latency, with the lowest latency for all task sizes. The HESGA algorithm is the second best, followed by the ANN-BPSO and MALO algorithms, which have similar performances. The G_SOS algorithm consistently has the highest latency.

The model demonstrates a high accuracy of 97.3%, indicating that it correctly classifies a vast majority of reviews. With a precision of 96.5%, the model is highly effective in correctly identifying spam reviews while minimizing false alarms. The model has a recall of 97.8%, meaning it effectively detects the majority of spam reviews in the dataset. The F1 score, at 97.1%, demonstrates a well-balanced trade-off between precision and recall, indicating a robust model. The false positive rate, at 3.5%, is relatively low, indicating a small percentage of legitimate reviews being misclassified as spam. The ROC AUC of 0.986 indicates a strong ability to distinguish between spam and legitimate reviews

Table 4- Model outcome values

Metric	Value
Accuracy	97.3%
Precision	96.7%
Recall (Sensitivity)	97.8%
F1 Score	96.7%
ROC AUC	0.986

These results highlight the strong performance of the methodology in detecting spam reviews while minimizing

false alarms. High precision, recall, and F1 score values, along with a low false positive rate and a high ROC AUC, demonstrate the model's effectiveness in handling diverse datasets and real-world applications. Fine-tuning and customization can further optimize performance for specific use cases.

5. Conclusion

In conclusion, this research has introduced a pioneering methodology that leverages Deep Convolutional Variational Auto encoders (DC-VAEs) and Hierarchical Density-Based Clustering (H-DBSCAN) to achieve remarkable accuracy in identifying spam reviews within textual data. The results demonstrate the model's exceptional precision, recall, and F1 score, showcasing its effectiveness in distinguishing between legitimate and spam content. Beyond spam review detection, the approach holds promise for a wide array of real-world applications, including fraud detection, cybersecurity, and quality control, where it can be adapted to uncover concealed anomalies and enhance decision-making processes. This research signifies a significant advancement in anomaly detection techniques and underlines the model's potential to ensure the integrity and quality of user-generated content while opening doors for innovation in diverse domains.

References

[1] Hussain, N., Turab Mirza, H., Rasool, G., Hussain, I., & Kaleem, M. (2019). Spam review detection techniques: A systematic literature review. Applied Sciences, 9(5), 987.

[2] Hussain, N., Mirza, H. T., Hussain, I., Iqbal, F., & Memon, I. (2020). Spam review detection using the linguistic and spammer behavioral methods. IEEE Access, 8, 53801-53816.

[3] Li, A., Qin, Z., Liu, R., Yang, Y., & Li, D. (2019, November). Spam review detection with graph convolutional networks. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (pp. 2703-2711).

[4] Shahariar, G. M., Biswas, S., Omar, F., Shah, F. M., & Hassan, S. B. (2019, October). Spam review detection using deep learning. In 2019 IEEE 10th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) (pp. 0027-0033). IEEE.

[5] Neisari, A., Rueda, L., & Saad, S. (2021). Spam review detection using self-organizing maps and convolutional neural networks. Computers & security, 106, 102274.

[6] Saumya, S., & Singh, J. P. (2022). Spam review detection using LSTM autoencoder: an unsupervised approach. Electronic Commerce Research, 22(1), 113-133.

[7] Bhuvaneshwari, P., Rao, A. N., & Robinson, Y. H. (2021). Spam review detection using self-attention based CNN and bi-directional LSTM. Multimedia Tools and Applications, 80, 18107-18124.

[8] Rao, S., Verma, A. K., & Bhatia, T. (2021). A review on social spam detection: challenges, open issues, and future directions. Expert Systems with Applications, 186, 115742.

[9] Tang, X., Qian, T., & You, Z. (2020). Generating behavior features for cold-start spam review detection with adversarial learning. Information Sciences, 526, 274-288.

[10] Pandey, A. C., & Rajpoot, D. S. (2019). Spam review detection using spiral cuckoo search clustering method. Evolutionary Intelligence, 12(2), 147-164.

[11] Asghar, M. Z., Ullah, A., Ahmad, S., & Khan, A. (2020). Opinion spam detection framework using hybrid classification scheme. Soft computing, 24, 3475-3498.

[12] Liu, Y., Pang, B., & Wang, X. (2019). Opinion spam detection by incorporating multimodal embedded representation into a probabilistic review graph. Neurocomputing, 366, 276-283.

[13] Barreno, M., Nelson, B., Joseph, A. D., Rubinstein, B. I. P., Sears, R., Tygar, J. D., ... & Ristenpart, T. (2010). "The security of machine learning." Machine Learning, 81(2), 121-148.

[14] Kingma, D. P., & Welling, M. (2013). "Auto-Encoding Variational Bayes." arXiv preprint arXiv:1312.6114.

[15] Radford, A., Metz, L., & Chintala, S. (2015). "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks."

[16] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise."

[17] Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). "Estimating the Support of a High-Dimensional Distribution."

[18] Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). "ArnetMiner: Extraction and Mining of Academic Social Networks."

[19] Guha, S., Rastogi, R., & Shim, K. (2001). "CURE: An Efficient Clustering Algorithm for Large Databases."

[20] Liu, F. T., Ting, K. M., & Zhou, Z. (2008). "Isolation

Forest." In 2008 Eighth IEEE International Conference on Data Mining.

[21] Chen, X., Xu, Y., & Yang, J. (2016). "Spam Review Detection with Graph-Based Propagation Model." In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval.

[22] Ramaswamy, S., Rastogi, R., & Shim, K. (2000). "Efficient algorithms for mining outliers from large data sets." In Proceedings of the 2000 ACM SIGMOD international conference on Management of data.

Acknowledgements

I am grateful to all of those with whom I have had the pleasure to work during this and other related Research Work. Each of the members of my Dissertation Committee has provided me extensive personal and professional guidance and taught me a great deal about both scientific research and life in general.

Conflicts of interest

The authors declare no conflicts of interest

Funding Details

No funding claimed to assist with the preparation of this manuscript.