Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 246 (2024) 1568–1579

Procedia Computer Science

www.elsevier.com/locate/procedia

28th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems (KES 2024)

SpamVis: A Visual Interactive System for Spam Review Detection

Nguyen Thanh Thao Lam[a], Nu Uyen Phuong Le[b], Md Rafiqul Islam[c], Md Kowsar Hossain Sakib[c], Shanjita Akter Prome[d], Cesar Sanin[c], Edward Szczerbicki[e], Jianlong Zhou[f]

[a]*College of Business & Management, VinUniversity, Hanoi, 12426, Vietnam*
[b]*Faculty of Engineering, Architecture, and Information Technology, University of Queensland, Brisbane, QLD 4072, Australia*
[c]*Business Information Systems, Australian Institute of Higher Education, Sydney, New South Wales, Australia*
[d]*School of Computer Science, Taylor's University, Malaysia*
[e]*Faculty of Management and Economics, Gdansk University of Technology, Gdansk, Poland*
[f]*School of Computer Science, University of Technology Sydney, Australia*

Abstract

In recent times, the number of spam reviews through various online platforms has emerged as a prime challenge, profoundly impacting businesses and consumers. These fake reviews not only distort clients' perceptions of products and services but also erode trust within the digital ecosystem. Despite the advent of machine learning (ML) techniques for identifying spam reviews, comparing text, and pinpointing groups of spammers, there remain notable gaps in both accuracy and the combination of interactive visualization for real-time decision-making. This paper presents *SpamVis*, a visual interactive system that leverages deep learning (DL) and ML blended with advanced visualization techniques for spam review detection, enabling analysts to conduct complicated analytical queries. The system allows users to input via click-on or touch to generate interactive charts and plots tailored for spam review analysis. The findings of the baseline test carried out on 67,395 review texts demonstrate that Bidirectional Encoder Representations from Transformers (BERT) carried out the best accuracy (86%) compared to other models. Our outcomes suggest that *SpamVis* can alleviate the gaps concerning accuracy and visualization needs in contemporary techniques, guiding analysts to make informed decisions for mitigating spam reviews and enhancing consumer trust. Furthermore, *SpamVis* empowers users to seamlessly discover the online reviews of various social media platforms in real-time, such as Facebook, Youtube, etc., giving them practical insights to navigate the online marketplace effectively.

Keywords: Deep Learning; Visual Interactive System; Spam Detection; Machine Learning; BERT; RoBERTa

* Corresponding author. Tel.: +61290208050
E-mail address: r.islam@aih.edu.au

1877-0509
10.1016/j.procs.2024.09.620

1. Introduction

In today's world, social proof is an essential aspect of the e-commerce domain because users often make their purchasing decisions about products or services based on what they read in online reviews. The sheer volume of user-generated content – with estimates suggesting 5 billion of reviews written annually across the globe [33]– has transformed the online shopping landscape. Additionally, according to another report, 87% of customers perceived relying on online reviews when purchasing products [21]. However, this emerging issue and the growing digital environment have given rise to spam reviews, which represent directed posts by highly motivated individuals who have significant economic incentives. These are otherwise referred to as fake reviews or untruthful opinions, designed to influence consumers' perceptions through exaggerated positive or negative feedback about a particular product on sale. According to existing studies, the percentage of such fake reviews differs and varies from 16% of reviews as mentioned by Luca and Zervas [18] to 33% of the total expressed by Salehi et al. [25], and even 50% fake reviews, according to Camilleri et al. [6]. Due to this, companies working to use algorithms to filter fake reviews still exist, and what Amazon found was that the proportion of fake reviews it deals with has gone up from 36% in the year 2019 to 42% in 2020 [4]. Such continuation affects the financial solvency and credibility of the organizations that encounter them, resulting in a projected annual fake investment on the international e-commerce expenditure of $152 billion [32]. For the UK, fake review content that targets products is claimed to have the potential of costing UK consumers up to £312 million annually [5]; this shows that the concern about spam review goes beyond a purely social issue of eroding consumer trust to have profound and quantifiable impacts on the economy, as well as the credibility of the market in the e-commerce system.

Therefore, to overcome this challenge, a large number of studies have been conducted aiming at employing ML and DL methods to classify content-based spam reviews. These methods identify fake content in reviews by analyzing the structural, sentimental, writing style, language, and format-oriented approaches, as identified in [20]. Popular ML classifiers like SVM, DT, and LR have also been used to address spam problems because of the various relationships and complicated data patterns that exist in spam review detection [19]. Similarly, DL models such as CNN, RNN, and LSTM have also been applied to multiple spam detection issues. These techniques help improve the efficiency of taxonomy-based content analysis for the identification of spam messages. For instance, Jacob et al. [13] put forward a system for identifying fake reviews on e-commerce websites, employing sentiment analysis by applying supervised ML such as logistic regression (LR), linear regression, CNN, and RNN on the cross-cutting products bought through Amazon websites, reviews of the hotels enlisted in TripAdvisor, and social media accounts. They detected that CNNs offered better accuracy in comparison to linear SVM, even though character-level SVM yielded good-performing paraphrasing tools with comparatively low complexities. More recently, Shang et al. [29] proposed a completely new approach for screening spam reviews by leveraging the concept of group intelligence and individual user sentiments through multi-dimensional representation. They also applied BERT to fine-grained sentiment analysis and text embedding together as a part of TNet and carried out many experiments on three datasets of Yelp (YelpChi, YelpNYC, YelpZIP), where their approach earned high recall rates and F1 scores and proved BERT's powerful modeling features for detecting spam reviews.

However, there are still unexplored scopes in this field, such as building a visual interactive system integrated with spam detection and real-time analysis of user reviews. Visual analytics has been used in numerous fields, such as recommendation systems [27, 12], mental health analysis, disease diagnosis [24], suicidal ideation detection [11], and lie detection [22, 23], but has not been delved into in previous studies of spam review detection domain, despite the enormous potential of visualized ML explanations on influencing human decisions [30]. Without a real-time analysis system of user reviews, it is particularly tough for all the involved stakeholders, including consumers, businesses, and e-commerce platforms, to take effective countermeasures in time. Notably, timely and accurate assessment of online reviews is of great significance for trust building, product authenticity control and user experience optimization. These gaps indicate the great demand for developing an innovative, efficient, and user-friendly solution in spam review detection with the integration of visualization. Hence, in this research, we focus on addressing these limitations and accordingly propose *SpamVis*, a visual interactive system to empower spam detection (SD) abilities. Whereas existing research focuses on ML and DL algorithms for SD but ignores efficient interaction and visualization of the results, *SpamVis* applies the state-of-the-art ML and DL algorithms with strong support for real-time review analysis,

interactive result exploration, and visualization. The dashboard is designed to enhance the efficiency of spam detection and make algorithmic spam detection more transparent and interpretable for end-users.

The contributions of this paper are listed as follows:

- **Application of ML and DL methods:** We utilized several ML and DL techniques, including BERT, RoBERTa, SVM, LR and DT, to categorise the reviews as spam or genuine. These multiple classification methods allow us to develop a more potent and robust spam review detection system, capable of detecting fake reviews with improved accuracy.
- **Development of *SpamVis*:** We introduced an interactive visualization system for the spam detection task. *SpamVis* is the first unified, user-friendly interactive visual analytics framework for visualization into the field of spam review detection.
- **Real-time review analysis:** Through this system, users can gather customer reviews from various online sources and analyze the genuine reviews as well as the spam reviews with corresponding sentiment analysis.

The remainder of this paper is organized as follows: Section 2 discusses the associated paintings; Section 3 details our method, which includes the pipeline of record series, processing, feature extraction, and model schooling. Section 4 describes the system, and Section 5 affords the results and compares them. Then discuss the key contributions in Section 6, and eventually Section 7 concludes the paper and shows future directions.

2. Literature Review

This section provides machine-learning approaches for spam detection, deep-learning approaches for spam detection, and spam detection through data visualization from recent work with their contribution and research flows.

2.1. Machine Learning approaches for spam detection

Nowadays, information can now be easily accessed in a variety of formats, including text, photos, sounds, films, and movies, from a wide range of sources. This data is important to have before making any kind of online purchase. On any internet site, users also offer their opinions and experiences; these comments are referred to as reviews. According to De et al. [8] online reviews are the most common type of electronic Word-Of-Mouth (eWOM) and are a way to transfer relevant knowledge about product usage, services, or seller behavior to consumers through internet-based reviews. Lo and Yao [17] describe reviews as user-generated content used by consumers for decision-making across purchases, travel, renting, and service sign-up. These views help others to determine purchasing decisions. Online reviews are also the richest informational source for retailers, manufacturers, and decision-makers. Nevertheless, we should point out that many statements of the kind posted by companies go where consumers are shown with hot profiles. The degree of deceptive positive feedback becomes a critical problem. Hence, several studies have been conducted to detect spam reviews through technological exploration.

Next, Saumya and Singh [26] presented a new semantic technique supporting sentiment valence assessment through RG, GB, and SVM classifiers. And 60% of the study participants disagreed with their evaluation, with 91% of them specifically examining the list of suspicious reviews. Recently, Barbado [2] and colleagues (2019) provided a novel idea to detect fake reviews while Khan et al. The detection of fake reviews using ML techniques was done by [15] In their result, RF, NB, DT, LR with SVM had better accuracy. Therefore, traditional ML approaches have been applied in this field, providing various insights. Whereas researchers applied ML techniques and got good results, there are still some limitations. Day by day, researchers are focusing on new technology that can handle large amounts of data, giving us more strategies to explore fake reviews.

2.2. Deep Learning approaches for spam detection

Most of the recent work has been carried out on DL models using neural networks and transfer learning-based approaches for classifying and predicting fake reviews. For instance, Wang et al. [34] employed an LSTM RNN model to detect spammers from fake reviews in Taiwan, which is a real scenario. Their work compared the results

with previous studies and identified the excellent performance of LSTM and SVM for fake review detection. In their model, an input layer, an LSTM layer, an output layer, and a hidden layer for dimensionality reduction were included. Jain et al. [14] proposed a detection system by multi-instance learning and a two-layer model that combines CNNs and GRNs in a hierarchical CNN-GRN. They used a 3-layer CNN for extracting raw n-gram features and added a GRN for modeling the semantic dependencies of the n-gram features. The study by Hajek et al. [10] presented two neural network models, which they intended to be used for improving the identification of fake reviews. These models combine traditional bag-of-words with word context and consumer emotions to learn document-level representation using three types of features: These may include n-grams, word embeddings, and indicators derived from a lexicon-based emotion model. This leads to a feature representation in a high-dimensional space for classifying fake review data into four categories. The efficiency of our approach is evidenced by the comparison with other methods of fake review detection for different datasets, with the sentiment polarity of the reviews and the types of products being diverse.

Then, Shan et al [28] proposed an auto-transformer model for distinguishing between fake news and Covid-related posts on social media. It also revealed that the domain-specific language models should be incorporated and the BCE-Dice Loss function should be utilized in order to achieve the enhanced classification rate. Meanwhile, Andresini et al. [1] proposed EUPHORIA, which is a classification framework for differentiating spammed contents from original ones, and this is done by utilizing multi-view learning fused with DL methods. This method contributes significantly to improving accuracy as a result of compiling as much information as possible regarding the contents of the reviews and the activity of the reviewers, as exemplified by the following real review datasets from Yelp.com. Gupta & Gandhi et al. [9] investigated transfer learning and other transformer-based pre-trained models for creating a fake review classifier where traditional models such as ML and neural networks lack proficiency. Models were trained on 10% and 50% of the Yelp dataset for the creation of a more generalized classifier, but not necessarily at the cost of high computational power.

2.3. Spam detection through data visualization

Data visualization is the idea of portraying information and data in a form that is quite easy to understand with the use of graphic images, including charts, graphs, and maps, among others. These masses of data are made easier to understand through the use of these other visual aids, which are useful in the effective interpretation of even complex figures. Specifically, data visualization aims to translate numeric data into visual language so that it becomes easier to detect trends, patterns, and anomalies. In the recent past, different parts of sectors have incorporated data visualization processes into creating an interactive dashboard to present the findings and conclusions in formats such as visual graphics, maps, and graphs. For instance, Chowdhury et al. [7] created a dashboard that integrates a frame-based dialog approach to identify the user's intent and slots of interactions by multiple input modalities, including voice, mouse, and keyboard. This integration increases navigation within multi-virtual interfaces, as seen in COVID-19-related artifacts. Su et al. [31] distinguished the following interaction requirements for the analysis of flow visualization and characteristics of human-computer interaction channels: multiple methods for the interaction were proposed. Their objective was to further enrich intuitive communication between people and computers by adopting gestures, head movement stability, eye comfort, and voice search for effective visualizations. Bjorkelund et al. [3] recommended a way of analyzing data retrieved from travel review websites. They explain how text mining offers a way of visualizing the analysis of textual review feedback, employing Google Maps as a way of identifying areas with highly rated hotels and other areas of interest for travelers. Moreover, their approach has extra value in interactions with instruments of faceted and filterable views and gives a more refined and targeted experience to users.

Generating insightful animations for spam detection with data visualization aims at creating engaging graphical interfaces for the detection and analysis of fake reviews online. It also increases the level of difficulty for consumers to distinguish between original and fake reviews since it provides a graphical view that shows them where trends and oddities might lie. Spam detection is one of the widely used and popular techniques that can be applied practically in all domains like emails. Lee et al. [16] proposed an improved spam SMS detection work that featured visualization functionality. This technique was divided into two main methods: a survey on the telephone with a random sample of companies and a survey on the Internet of companies that deal in computer hardware. The first method linked the Unicode value of each character with points in a two-dimensional space. The second method encoded each character in the string as an integer Unicode value and presented it in the form of an RGB triplet, where each value is a pixel on

the resulting image. Their proposed visualization method offered efficiency in classification since one was saved from tasks such as tokenization, stemming, and spell-checking, among others.

In summary, it can be stated that although ML and DL are commonly utilized for efficient spam review detection and that MDL techniques are applied to build emblematic models for this task, there is still a clear gap in the provision of an effective interactive visualization system in the current problem domain. Furthermore, as more attention is now paid to the effectiveness of data visualization, more research and practice are required to advance and apply these strategies in this field. Therefore, it is crucial to present a comprehensive interactive visual system for detecting spam reviews, where modeling is enriched by both ML and DL.

3. Methodologies

In this section, we describe the process by which we gather data, extract features, train the ML and DL classifiers, as well as create an interactive dashboard.

3.1. Data Collection

In this study, we utilized reviews sourced from YelpCHI. The primary datasets employed comprise reviews and metadata pertaining to hotels and restaurants. These review datasets are supplemented with annotations indicating whether a review is spam or not, facilitating training purposes. The metadata includes product IDs, reviewer IDs, and timestamps for the reviews. The hotel dataset encompasses approximately 5000 reviews, whereas the restaurant dataset comprises over 60000 reviews. Following data collection, we conducted preprocessing steps to prepare the data for subsequent analysis.

3.2. Feature Extraction and Data Splitting

To extract the key features, the datasets containing reviews were first combined with related metadata. After this initial step, the process of extracting features took place. It involved grouping the timestamps by year, figuring out how many words each review has, and discovering the total number of meaningful words in each review. This step aims to make the process more efficient by focusing on using only the data that gives us useful information for training. It helps cut down on the work needed to pull out information. Once the entire dataset was put together, the data for training and testing was divided into 4 training parts and a part for testing.

3.3. Training Methods

Both ML and DL methods have been used for the classification task. As for ML methods, support vector machines (SVM), logistic regression (LR), and decision trees (DT) have been used. BERT and RoBERTa were used for the DL methods. For DL methods, the input also covers textual content tokenization for evaluations and the feature extraction process.

3.4. Result and Evaluation

The predictions from each training method are evaluated using accuracy, precision, recall, and AUC metrics. The details of the result and the evaluation are described in Section 5.

3.5. Interactive Dashboard

Following the training results, we developed *SpamVis*, an interactive visualization system designed to thoroughly discover and examine spam reviews. *SpamVis* permits customers to select models and data, providing a complete and immersive evaluation throughout diverse domain names, which include restaurants, hotels, and social media platforms. The system incorporates two principal components: offline data evaluation using YelpCHI review data and online data analysis for real-time social media application review. The dashboard features a wide array of charts

and visualizations, along with bar graphs, horizontal bar charts, and result tables. This perspective helps analyze and interpret the results of spam detection analysis, providing customers with clear and actionable insights.

One of the key features of *SpamVis* is its online data analysis section. During this phase, customers can choose from several social media platforms, such as Facebook, Instagram, Messenger, WhatsApp, YouTube, and Facebook, to extract reviews. The system then offers real-time predictions, classifying the reviews as spam or real, and as positive or terrible. This functionality permits customers to respond without delay to emerging spam problems, thereby safeguarding the integrity and trustworthiness of user-generated content online.

Furthermore, the user interface of *SpamVis* is designed to be intuitive and user-friendly, permitting users to effortlessly navigate via special functionalities and customize their analysis in line with their unique needs. The integration of offline and online data ensures the availability of all technologies for spam detection, making *SpamVis* an effective system for agencies and investigators. *SpamVis*'s technical design and implementation details are provided in Section 4.

4. SpamVis System

Drawing inspiration from LieVis[22] and SidVis[11] dashboard, which serve as the basis for this dashboard, *SpamVis* is designed to offer a comprehensive and interactive platform for spam review detection and analysis. Its architecture supports two main functionalities: Offline Data Analysis and Online Data Analysis. This section provides an in-depth analysis of these functionalities, highlighting the system's design and capabilities.

4.1. Offline Data Analysis

In the offline data analysis component, *SpamVis* empowers users with extensive control over their spam detection tasks. Users can customize their analysis by selecting different types of input, including model choices, training datasets, and testing datasets. The available models range from advanced DL models like BERT and RoBERTa to traditional ML models such as SVM, Decision Trees, and LR, allowing for comprehensive analysis and comparison. The options for training and testing datasets include selecting between restaurant or hotel review data. Additionally, users have the option to upload their own CSV files containing text reviews, enabling personalized and specific testing scenarios.

Additionally, there is a horizontal plot that compares the average length of spam reviews versus genuine reviews, highlighting significant differences in review length that may indicate spam. Pie charts display the disparity in sentiment distribution between spam and authentic reviews, offering insights into how sentiments are manipulated in spam content. Lastly, the system provides interactive displays of key performance metrics from the training and testing phases of the chosen models. These metrics include accuracy, precision, recall, and the area under the curve (AUC), providing a transparent view of the model's effectiveness.

The offline data analysis component of *SpamVis* enables users to extract detailed insights from spam review datasets and compare the efficiency of different models. By providing comparative results between the training and testing phases, users can evaluate the performance and reliability of various models, facilitating a deeper understanding of model effectiveness and aiding in the selection of the most appropriate model for their specific needs.

4.2. Online Data Analysis

The online data analysis component of *SpamVis* provides users with real-time insights into reviews from online platforms. Our dashboard allows users to dynamically investigate reviews by fetching real-time data, and then a user can choose the number of reviews to analyze, ranging from 1 to 1000. When a user selects an app to investigate, the system can extract real-time text reviews from the chosen platform. We used Serpapi to fetch data from the Play Store. In order to provide users with flexibility and control over the scope of the analysis, users can define the exact number of reviews they wish to retrieve.

After the request is submitted, the system retrieves data through the API. This process entails initiating a request to SerpAPI to interrogate the chosen platform for the designated quantity of reviews. We cleaned up the data on the reviews we obtained from the Google Play Store and created a new data frame for analysis. After preprocessing, the

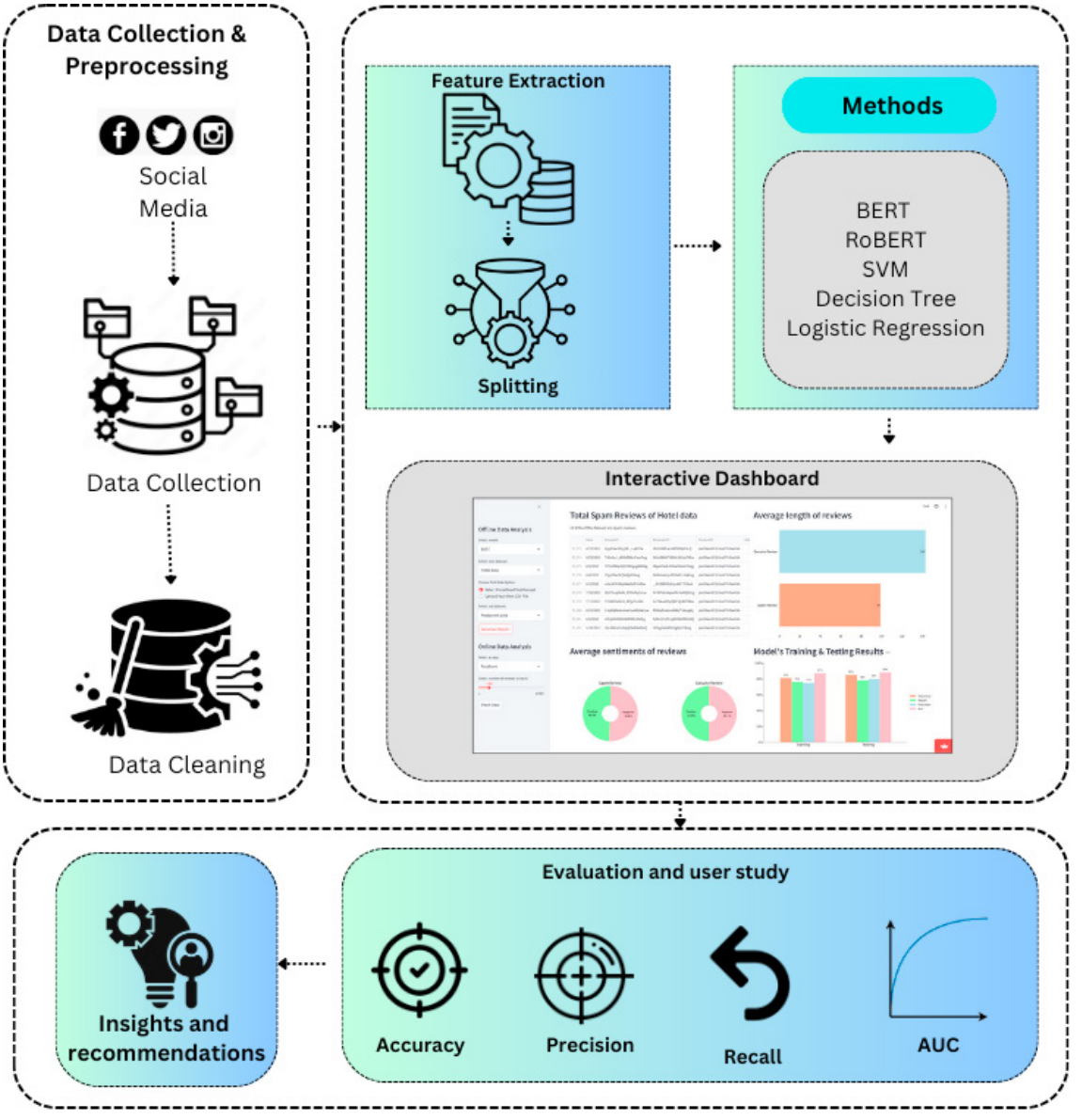

Fig. 1: *SpamVis* System Architecture

first dataset is displayed in the first plot. The system employs the trained BERT model to perform spam detection and sentiment analysis. Each review in the data frame is then analyzed and classified as either spam or not.

5. Experimental Analysis

For experiment analysis, we started with the experiment setup. Then, we analyzed the results and compared outcomes among models. We explained the individual model's performance by reviewing the experiment's validity. We ended the section with an evaluation of the models.

Dataset	Model	Accuracy	Precision	Recall	AUC
Restaurant Dataset Training	BERT	86.03	87.9	84.77	90.46
	RoBERTa	83.72	87.32	83.88	90.03
	SVM	80.43	83.26	79.65	87.9
	DT	82.19	85.33	77.67	89.6
	LR	81.45	78.43	80.87	88.59
Restaurant Dataset Testing	BERT	85.41	81.88	77.48	82.33
	RoBERTa	80.56	75.86	72.34	81.88
	SVM	77.61	83.66	74.55	81.35
	DT	79.07	78.9	77.89	80.76
	LR	80.2	76.59	75.49	87.41
Hotel Dataset Testing	BERT	83.57	79.00	77.02	80.90
	RoBERTa	76.33	71.78	71.90	80.77
	SVM	72.54	75.11	72.45	80.83
	DT	79.84	77.98	72.73	82.34
	LR	76.62	74.77	80.31	84.71

Table 1: Training performance using Restaurant Dataset

5.1. Experimental Setup

For DL, models are trained, tuned, tested, and deployed using Python, and checkpoints are saved in Huggingface. The ML methods are deployed inside the visualization dashboards.

5.2. Result analysis

Table 1 presents the performance of different ML and DL models for sentiment analysis on both datasets. On the DL side, BERT got the highest accuracy (86.03%) and AUC (90.46%) on the restaurant training dataset; on the other hand, RoBERTa achieved similar results on the testing dataset. From the ML side, we implemented SVM, DT, and LR. Among them, SVM performed well, especially on the restaurant training dataset, achieving an accuracy of 80.43%. DT and LR performed competently at a lower level than other DL and ML models. Overall, the findings indicate that DL models, especially BERT, are effective models for spam classification and sentiment analysis with complex text data.

Table 2 uses training results for the hotel dataset and testing results for both. Similar to Table 1, BERT performed consistently better in the training and testing phase with both datasets than other ML and DL models. We used the hotel dataset for training and both the restaurant and hotel datasets for testing. In the testing phase, with restaurant data, BERT performed exceptionally well in accuracy and other metrics such as precision, recall, and AUC. Among all the models and from the ML side, SVM showed poor performances in both training and testing.

In Tables 1 and 2, our model's accuracy is higher (85.41% when using the same training set and 84.04% when using a different training set). For RoBERTa, our model achieved comparable results in terms of accuracy, precision, recall, and AUC compared to the existing paper. Thus, it can be concluded that our model performs competitively with these current models.

5.3. Evaluation

We evaluated our system with different performance metrics—precision, accuracy, recall, and AUC and performed a user study to assess how the visuals could make our system more effective. We employed the user study to understand how real users interacted with the system across input formats and more complex queries on how the system reacted correspondingly. The study included ten participants: four females and six males. Each user was required to be familiar with the basic concepts of data visualization. After a brief introduction to the system, users were invited to play around with the visuals and provide their inputs, answering specific questions about our system. Most participants provided

Dataset	Model	Accuracy	Precision	Recall	AUC
Hotel Dataset Training	BERT	85.88	88.05	83.99	92.99
	RoBERTa	84.81	88.64	83.78	91.55
	SVM	79.88	81.24	80.79	85.73
	DT	81.41	83.97	78.45	90.83
	LR	81.99	80.31	81.49	90.60
Hotel Dataset Testing	BERT	84.67	80.74	77.31	83.28
	RoBERTa	81.92	78.48	76.65	82.39
	SVM	78.99	87.49	88.04	80.99
	DT	78.56	76.99	76.87	79.48
	LR	78.81	77.90	76.48	88.28
Restaurant Dataset Testing	BERT	84.04	76.99	76.08	78.95
	RoBERTa	79.09	74.88	75.01	82.79
	SVM	73.50	76.42	74.11	80.8
	DT	79.19	78.03	72.69	81.9
	LR	77.39	71.51	74.53	82.99

Table 2: Training performance using Hotel Dataset

positive feedback after engaging with our system. It should be noted that participants freely donated their time and were unpaid.

6. Discussion

This section describes the main findings and the spam detection result. While the literature focuses on the development of algorithmic ML and DL models for spam review detection, this work introduces a visual interactive system, as part of a hybrid approach, that enables the user to analyze reviews across different domains in both real-time and offline. By doing so, the study introduces an easy-to-use system that addresses a critical gap in current spam detection research. DL models, such as BERT remain notably effective. There are some following key observations:

- Firstly, *SpamVis* offers a number of insights in different forms such as charts, plots, and provides users with ample support to understand and analyze spam reviews across different domains. For offline data analysis, *SpamVis* will calculate the percentage of reviews labeled as spam, helping users get a sense of spam proportions within the data. By bringing up the table of spam reviews, users can clearly spot the features common to these fake reviews, such as phrases or keywords, which may help in identifying patterns and devising strategies against spam. The system also displays the average word length of spam reviews against non-spam ones. Lastly, sentiment analysis between spam and non-spam reviews can show whether spammers tend to manipulate the sentiment and the length of a review. By having a user-friendly interface, the system aims to accommodate a wide range of users with non-technical backgrounds in dealing with spam reviews effectively such as business professionals, researchers, and end-users.
- For online data analysis, *SpamVis* offers real-time insights, which is critical for investigations on social media platforms and so forth. This functionality is important as it helps to level immediacy and interactivity that is missing from traditional spam detection systems. Users could select certain online platforms to study in more detail, and this flexibility allows systems to capture online reviews to analyze. The system then processes these reviews in real-time, classifying them as spam or not spam and evaluating their sentiments. This real-time analysis provides immediate and actionable insights for human decision-making, empowering users to respond swiftly to emerging spam threats and sentiment shifts. Additionally, SpamVis can help companies identify between real customer concerns and manipulated spam when adjusting their products or services based on user feedback. This ensures that businesses prioritize addressing issues raised in authentic customer reviews only, aligning their offerings with customer expectations realistically. From the end-user's perspective, by analyzing spam reviews in real-time, they can seamlessly filter out fake reviews and focus on genuine customer experi-

ences. This leads to more informed buying decisions and prevents them from being affected by manipulated ratings.

- A key contribution of *SpamVis* is in providing cross-industry and cross-platform analysis: users can survey the spectrum of reviews across industries (hotels, restaurants, social media) and check if there are specific spam behaviors that are industry-specific or platform-specific. Likewise, using sentiment analysis from diverse kinds of reviews can provide a more general idea about the different ways reviews can vary from one industry to another, which can be relevant since various types of businesses are constantly on the lookout to improve customer satisfaction and engagement. From the end-user side, by consolidating review analysis into a single system, *SpamVis* allows for efficient comparison and information gathering between platforms, in which users will adjust their trust in reviews displayed on certain platforms with high volumes of spam.

Our initial version of SpamVis in this research work has several inevitable limitations that we would like to acknowledge and suggest some directions for further studies. Firstly, the live system currently focuses on three sectors only, which are: Hospitality, Food & Beverage, and Social Media. Expanding the system's capabilities to analyze data from an array of critical industries such as Health & Wellness, Luxury, and Automotive, could further increase its relevance, practicality, and usability. Secondly, SpamVis could benefit from integrating a multimodal interaction feature, where the system could take user inputs in the form of audio, text, click, touch, or even create a chatbot to enable human-like conversations with users. This feature could accommodate any additional follow-up queries from users such as providing analysis into the strengths and weaknesses of each DL or ML model, enabling users to choose the most suitable method for their review analysis problem. In short, we presented a web-based visual analytics system that combines visual analytics, real-time feedback, and a combination of ML and DL models to address a number of key gaps in previous studies for spam detection. Previous works were dominated by the development of individual algorithmic solutions that ignored the importance of user interaction and real-time feedback. *SpamVis* aims to fill this gap by offering an interactive, visual-driven approach to spam detection, in line with the growing trends in data science toward interpretability and user-centric design. The system's user interface and interactivity suggest it is an innovative system for a wide range of users, from researchers to industry practitioners.

7. Conclusion

Online reviews are becoming more and more prevalent every day, and they are quite helpful in assisting people in determining the quality of any platform or product. Before making online purchases, customers are accustomed to reading reviews, aiding their decision-making. However, spam reviews pose a significant risk to both consumers and sellers across the globe, eroding customer confidence in online purchases. As it remains one of the most important yet challenging problems of today's intense social activity online, this study explored it with advanced methods. Traditional classifiers such as SVM, DT, LR, and state-of-the-art DL models such as BERT and RoBERTa were used to detect spam reviews and showcase how well these models function. In order to enhance both usability and analytical value, we developed the *SpamVis* system as an interactive one. Endearing to users and tasked with analyzing real-time user interaction and reviews across various disciplines, *SpamVis* utilizes sentiment analysis as well as cross-discipline insights, making one ready to make certain decisions. Furthermore, *SpamVis* adds a significant contribution to the study of spam detection as it combines solid analytical models with effective visualization. It contributes to the effort of enhancing the scope and empowers the users to take swift action after analyzing the spam reviews. In addition, *SpamVis* lays the foundation for future studies and developments in the area of spam analysis and detection. One of the areas we expect to expand in the future is adding multilevel interactions. We also decided to expand the communication process by using chatbots to enhance visual content. These changes will improve the system and make it more efficient, customizable, and useful in the current world of spam.

References

[1] Andresini, G., Iovine, A., Gasbarro, R., Lomolino, M., de Gemmis, M., Appice, A., 2022. Euphoria: A neural multi-view approach to combine content and behavioral features in review spam detection. Journal of Computational Mathematics and Data Science 3, 100036.

[2] Barbado, R., Araque, O., Iglesias, C.A., 2019. A framework for fake review detection in online consumer electronics retailers. Information Processing & Management 56, 1234–1244.
[3] Bjørkelund, E., Burnett, T.H., Nørvåg, K., 2012. A study of opinion mining and visualization of hotel reviews, in: Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services, pp. 229–238.
[4] Bloomberg, . Amazon (amzn) cracks down on fake reviews, hitting chinese retailers. Bloomberg.com URL: https://www.bloomberg.com/news/articles/2021-08-18/amazon-amzn-cracks-down-on-fake-reviews-hitting-chinese-retailers.
[5] for Business, D., Trade, . Investigating the prevalence and impact of fake reviews. URL: https://www.gov.uk/government/publications/investigating-the-prevalence-and-impact-of-fake-reviews.
[6] Camilleri, A.R., 2019. How to spot a fake review: You're probably worse at it than you realise.
[7] Chowdhury, I., Moeid, A., Hoque, E., Kabir, M.A., Hossain, M.S., Islam, M.M., 2020. Miva: Multimodal interactions for facilitating visual analysis with multiple coordinated views, in: 2020 24th International Conference Information Visualisation (IV), IEEE. pp. 714–717.
[8] De Pelsmacker, P., Van Tilburg, S., Holthof, C., 2018. Digital marketing strategies, online reviews and hotel performance. International Journal of Hospitality Management 72, 47–55.
[9] Gupta, P., Gandhi, S., Chakravarthi, B.R., 2021. Leveraging transfer learning techniques-bert, roberta, albert and distilbert for fake review detection, in: Proceedings of the 13th Annual Meeting of the Forum for Information Retrieval Evaluation, pp. 75–82.
[10] Hajek, P., Barushka, A., Munk, M., 2020. Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. Neural Computing and Applications 32, 17259–17274.
[11] Islam, M.R., Sakib, M.K.H., Ulhaq, A., Akter, S., Zhou, J., Asirvathamt, D., 2023. Sidvis: Designing visual interactive system for analyzing suicide ideation detection, in: 2023 27th International Conference Information Visualisation (IV), IEEE. pp. 384–389.
[12] Islam, M.T., Islam, M.R., Akter, S., Kawser, M., 2020. Designing dashboard for exploring tourist hotspots in bangladesh, in: 2020 23rd international conference on computer and information technology (ICCIT), IEEE. pp. 1–6.
[13] Jacob, M.S., Rajendran, S., Michael Mario, V., Sai, K.T., Logesh, D., 2020. Fake product review detection and removal using opinion mining through machine learning, in: Proceedings of International Conference on Artificial Intelligence, Smart Grid and Smart City Applications: AISGSC 2019, Springer. pp. 587–601.
[14] Jain, N., Kumar, A., Singh, S., Singh, C., Tripathi, S., 2019. Deceptive reviews detection using deep learning techniques, in: Natural Language Processing and Information Systems: 24th International Conference on Applications of Natural Language to Information Systems, NLDB 2019, Salford, UK, June 26–28, 2019, Proceedings 24, Springer. pp. 79–91.
[15] Khan, H., Asghar, M.U., Asghar, M.Z., Srivastava, G., Maddikunta, P.K.R., Gadekallu, T.R., 2021. Fake review classification using supervised machine learning, in: Pattern Recognition. ICPR International Workshops and Challenges: Virtual Event, January 10–15, 2021, Proceedings, Part IV, Springer. pp. 269–288.
[16] Lee, H., Jeong, S., Cho, S., Choi, E., 2023. Visualization technology and deep-learning for multilingual spam message detection. Electronics 12, 582.
[17] Lo, A.S., Yao, S.S., 2019. What makes hotel online reviews credible? an investigation of the roles of reviewer expertise, review rating consistency and review valence. International Journal of Contemporary Hospitality Management 31, 41–60.
[18] Luca, M., Zervas, G., 2016. Fake it till you make it: Reputation, competition, and yelp review fraud. Management science 62, 3412–3427.
[19] Mewada, A., Dewang, R.K., 2023. A comprehensive survey of various methods in opinion spam detection. Multimedia Tools and Applications 82, 13199–13239.
[20] Michael, C., Khoshgoftaar, T.M., Prusa, J.D., Richter, A.N., Al Najada, H., 2015. Survey of review spam detection using machine learning techniques. Journal of Big Data 2, 1–24.
[21] Paget, S., . Local consumer review survey 2024: Trends, behaviors, and platforms explored. URL: https://www.brightlocal.com/research/local-consumer-review-survey/.
[22] Prome, S.A., Islam, M.R., Asirvatham, D., Sakib, M.K.H., Ragavan, N.A., 2023. Lievis: A visual interactive dashboard for lie detection using machine learning and deep learning techniques, in: 2023 26th International Conference on Computer and Information Technology (ICCIT), IEEE. pp. 1–6.
[23] Prome, S.A., Ragavan, N.A., Islam, M.R., Asirvatham, D., Jegathesan, A.J., 2024. Deception detection using ml and dl techniques: A systematic review. Natural Language Processing Journal , 100057.
[24] Rahman, M., Islam, M.R., Akter, S., Akter, S., Islam, L., Xu, G., 2021. Diavis: Exploration and analysis of diabetes through visual interactive system. Human-Centric Intelligent Systems 1, 75–85.
[25] Salehi-Esfahani, S., Ozturk, A.B., 2018. Negative reviews: Formation, spread, and halt of opportunistic behavior. International Journal of Hospitality Management 74, 138–146.
[26] Saumya, S., Singh, J.P., 2022. Spam review detection using lstm autoencoder: an unsupervised approach. Electronic Commerce Research 22, 113–133.
[27] Seeliger, A., Nolle, T., Mühlhäuser, M., 2018. Process explorer: an interactive visual recommendation system for process mining, in: KDD Workshop on Interactive Data Exploration and Analytics.
[28] Shan, G., Zhou, L., Zhang, D., 2021. From conflicts and confusion to doubts: Examining review inconsistency for fake review detection. Decision Support Systems 144, 113513.
[29] Shang, Y., Liu, M., Zhao, T., Zhou, J., 2021. T-bert: A spam review detection model combining group intelligence and personalized sentiment information, in: International Conference on Artificial Neural Networks, Springer. pp. 409–421.
[30] Stites, M.C., Nyre-Yu, M., Moss, B., Smutz, C., Smith, M.R., 2021. Sage advice? the impacts of explanations for machine learning models on human decision-making in spam detection, in: International Conference on Human-Computer Interaction, Springer. pp. 269–284.
[31] Su, C., Yang, C., Chen, Y., Wang, F., Wang, F., Wu, Y., Zhang, X., 2021. Natural multimodal interaction in immersive flow visualization. Visual Informatics 5, 56–66.

[32] Vaughn Schermerhorn, A., 2021. How amazon continues to improve the customer reviews experience with generative ai. URL: https://www.weforum.org/agenda/2021/08/fake-online-reviews-are-a-152-billion-problem-heres-how-to-silence-them/.
[33] Vaughn Schermerhorn, A., 2023. How amazon continues to improve the customer reviews experience with generative ai. URL: https://www.aboutamazon.com/news/amazon-ai/amazon-improves-customer-reviews-with-generative-ai.
[34] Wang, C.C., Day, M.Y., Chen, C.C., Liou, J.W., 2018. Detecting spamming reviews using long short-term memory recurrent neural network framework, in: Proceedings of the 2nd International Conference on E-commerce, E-Business and E-Government, pp. 16–20.