Do the Explanations Make Sense? Explainable Fake Review Identification and Users' Perspectives on Explanations

Md Shajalal[1,2], Md Atabuzzaman[3], Alexander Boden[2,4], Gunnar Stevens[1,4] and Delong Du[2]

[1]University of Siegen, Siegen, Germany
[2]Fraunhofer Institute for Applied Information Technology - FIT, Sankt Augustin, Germany
[3]Virginia Tech, USA
[4]Bonn-Rhein-Sieg University of Applied Sciences, Sankt Augustin, Germany

Abstract

Customer reviews and feedback play a crucial role in shaping purchase decisions on e-commerce platforms like Amazon, Zalando, and eBay. However, a major concern is the prevalence of fake or spam reviews, often posted by sellers to deceive potential customers and manipulate product perceptions. Machine learning (ML) models are widely used to detect fraudulent reviews, but their decisions can be difficult to interpret due to their complexity—often functioning as *black boxes*. In this paper, we propose an explainable framework for fake review detection that not only achieves high precision in identifying fraudulent content but also provides interpretable explanations. To assess the effectiveness of these explanations, we conduct an empirical user evaluation to determine which information is most valuable in understanding model decisions. Initially, we develop fake review detection models using deep learning (DL) and transformer-based architectures, including XLNet and DistilBERT. We then apply Layer-wise Relevance Propagation (LRP) to generate explanations by mapping word contributions to the predicted class. Experimental results on two benchmark fake review detection datasets demonstrate that our models achieve state-of-the-art performance, outperforming several existing methods. Furthermore, we conduct a user study with 12 participants to evaluate the comprehensibility and usefulness of LRP-generated explanations. The findings from this study provide key insights into how explanations can be improved to enhance transparency and user trust in fake review detection systems.

Keywords

Fake Review, Explainability, LRP, Transformers, BERT, DistilBERT, XLNet, Empirical Evaluation of XAI

1. Introduction

The rapid growth of e-commerce platforms for ordering various products online makes consumers' lives easier, saving potential time and cost for both ends. Trust and transparency issues are always critical, as they are directly associated with customer satisfaction and the revenue of companies or retailers [1, 2]. Generally, customers or buyers in e-commerce or service providers tend to check the ratings and reviews of previous customers who have already purchased the products to get an idea of the quality of the targeted products. Users usually prefer to buy products with higher ratings and better customer reviews.

However, identifying fake reviews can benefit customers, retailers, or companies by providing a trusted and transparent e-Commerce platform. In the last decade, significant attention has been paid to identifying fake reviews using automated methods with ML and DL-based classifiers. Notable ML methods such as SVM, NB, XGBoost, etc., generally use the TF-IDF or bag-of-words representation of textual reviews. However, these methods are traditional ways of representing text. The semantic representation of text using word embeddings has been employed in almost every natural language processing (NLP) task. Word embeddings can represent text's semantic and contextual information in a high-dimensional space. With these representations, multiple methods have been proposed using DL-based classifiers, including recurrent neural networks (RNN) and its variants such as LSTM, BiLSTM, GRU, etc [3, 4, 5, 6, 7]. After the invention of transformer-based text representations, NLP methods

Late-breaking work, Demos and Doctoral Consortium, colocated with The 3rd World Conference on eXplainable Artificial Intelligence: July 09–11, 2025, Istanbul, Turkey
md.shajalal@uni-siegen.de (M. Shajalal)

achieved high performance in almost every section. Transformer-based approaches, including BERT and its variants like DistilBERT, mBERT, and RoBERTa, have been used in many text classification tasks [8]. Recently, Electra, XLNet [9], GPT, and other large language models (LLMs) have also garnered significant attention in text classification, achieving high performance in numerous NLP tasks.

Generally, DL- and transformer-based approaches have complex architectures and involve a complicated decision-making process in predicting the original class. In the fake review detection task context, users are typically laypeople with minimal knowledge about predictive models. The decisions might surprise them when they see a particular review detected as fake, but they cannot determine why it is predicted as fake. Recently, explainable artificial intelligence (XAI) has gained significant attention in different fields, including business [10], bioinformatics [11], NLP [3, 8, 12, 13], and more. In the use case mentioned above, XAI comes into play to explain and validate the predictions made by the model. In this decade, XAI techniques have gained considerable attention in explaining model decisions, allowing AI practitioners and users to understand the reasons behind predictions and improve model performance and decision understanding. Several renowned XAI methods, including SHAP [14], LIME [15], LRP [13], Bert-interprete [16], can be applied to explain decisions related to NLP tasks.

In this research paper, we adopt a user-centric approach to introduce an explainable fake review detection system. We aim to make the complex ML models, which often function as “black boxes," more understandable and less surprising for general users. We examine the generated explanations with an empirical evaluation based on the verdicts of human subjects, ensuring that the explanations are meaningful and useful to the end users. With such empirical evaluation, we investigated how much the explanations can make sense to the general users in understanding the prediction. Initially, we developed fake review detection models using cutting-edge transformer models such as XLNet and DistilBERT. We also applied different DL models, including BiLSTM, CNN, CNN-LSTM, and CNN-GRU models, to detect fake reviews. We then introduce the LRP [12, 13] technique in the fake review detection task to interpret the decisions from DL models and present explanations for individual predictions, highlighting the contributed words for the predicted class.

Our research has practical implications for the field of fake review detection. We conducted experiments in multiple settings, and the experimental results on two benchmark fake review detection datasets demonstrate that our predictive models achieve state-of-the-art performance and outperform several existing methods. Furthermore, our generated explanations can interpret specific decisions, enabling users to understand why a particular review is classified as fake or genuine. The empirical evaluation with 12 human subjects was conducted to examine the effectiveness of the explanations and elicit further requirements in generating explanations in the context of fake review identification. The significant contributions in this research are threefold: i) We introduced two transformer-based fake review detection models applying DistilBERT and XLNet that demonstrated significantly better performance than DL methods and existing related works. ii) Our method can explain specific predictions with explanations introducing LRP technique. The explanations might enable users to understand why particular reviews have been predicted as fake. iii) Our conducted empirical evaluation of the generated explanations with human subjects and the results indicate further requirements in generating explanations for fake content identification tasks.

2. Literature Review

Various classical and deep learning-based text classification models have been applied to detect fake online reviews, including SVM, KNN, LR, LGBM, LSTM, CNN, RNN, and transformers like BERT [17, 18, 4, 19, 5, 20, 6]. Recent approaches combine CNN, PSO, and NLP techniques for credibility analysis [6] or use hybrid models integrating latent text features and aspect ratings [4]. Others propose ensemble-based learning [21] or explainability-driven models like SHAP [22]. LLMs are increasingly used to generate artificial reviews, requiring robust detection frameworks [1]. Additionally, PU learning [23] and voting-based techniques [24] have been explored. CNN-based models leveraging web-scraped content [25] and RNN variants extracting multiple review aspects [7] further advance fake review

detection methodologies.

From the recent literature review analyzing published research from the last five years, it is evident that most fake review detection methods lack explainability, except for one study [22], which only used SHAP values for global interpretability. Moreover, existing approaches fall behind modern state-of-the-art transformer-based methods. To address these limitations, we employ efficient transformer models such as XLNet and DistilBERT for fake review detection. We compare their performance against traditional deep learning models, including LSTM, BiLSTM, CNN, CNN-BiLSTM, and GRU, utilizing FastText word embeddings. Additionally, we apply layer-wise relevance propagation (LRP) to enhance model interpretability and explain predictions.

Unlike explainable text classification tasks such as hate speech detection [8] or sentiment analysis [13], explaining fake review identification is more challenging. In sentiment classification, negative or positive words typically indicate corresponding sentiments, while hate speech detection relies on specific offensive terms. However, fake and genuine reviews often use similar wording, making conventional interpretability methods less effective. To address this issue, we conducted a user study with twelve participants to evaluate the explanations generated by our models and determine what factors help users understand predictions in the fake review detection context.

3. Experiments

3.1. Dataset

Fake Review Dataset: This fake review dataset contains 40000 reviews in total. Among them, 50% reviews were originally written by humans (i.e., reviews collected from Amazon). The rest of the reviews are fake, generated by two different language models including ULMFit (Universal Language model Fine-tuning) and GPT-2 [1].

Yelp Review Dataset: We conducted experiments with another fake review dataset named *Yelp Review Dataset*. Compared to the previous one, this dataset is quite big and consists of more than 682K reviews and the distribution is quite imbalanced. The dataset is accessible at Kaggle[1].

3.2. Experimental Setting and Results

We first applied an ensemble machine learning approach with majority voting on SVM, Decision Tree, Random Forest, and XGBoost. Next, we explored deep learning models, including BiLSTM, CNN, CNN-LSTM, and CNN-GRU. BiLSTM used an embedding layer, Spatial dropout, bidirectional LSTMs, and fully connected layers. CNN included convolutional and dropout layers, while CNN-LSTM and CNN-GRU combined CNN with LSTM or GRU. For transformers, we used DistilBERT and XLNet, employing FastText embeddings to address vocabulary mismatches. Datasets were split into 70% train, 15% test, and 15% validation.

Performance on Fake Review Dataset. The performance of different fake review detection models on Fake Review dataset [1] is presented in Table 1 in terms of multiple evaluation metrics. Among four different deep learning models, BiLSTM performs better in terms of accuracy (0.9556) and F1-Score (0.9466). We can also see that CNN-GRU performs equally compared to BiLSTM in terms of F1-Score and Precision which is almost the same. However, the other two DL models CNN and CNN-LSTM also achieved consistent and effective performance. In terms of all evaluation metrics, our proposed two transformer-based fake review detection models achieved significantly higher accuracy (0.9592), precision (0.9906), and F1-Score (0.9821) among all employed models. The performance difference between XLNet and DistilBERT is not significant and it is only a 1% difference in terms of precision. DistilBERT achieved more than 4% performance gain in terms of F1-Score.

Performance on Yelp Review Dataset. We also present the performance for the Yelp dataset in table 1. The table summarized that transformers-based classification models here also performed better

[1]https://www.kaggle.com/datasets/abidmeeraj/yelp-labelled-dataset/data

than the deep learning models and ensemble ML model. Unlike the performance in the previous dataset, XLNet achieved higher accuracy, F1-Score and AUC compared to the DistilBERT-based classifier. But for the other measure, in terms of precision , DistilBERT performed better. However, the performance difference is not that big but compared to the deep learning-based methods, both DistilBERT and XLNet outperformed significantly with a way higher AUC. AUC is considered one of the best evaluation metrics to measure the performance when the dataset is imbalanced.

Overall, the performance on this dataset is lower than on the previous dataset. There are several probable reasons. One is the size of the dataset, the Yelp dataset is significantly larger than the fake review dataset and reviews are written by human. However, in the Fake review dataset, the fake reviews are generated by the large language models (LLM). Additionally, the Yelp data is considerably imbalanced. Since the reviews are generated by LLM, the transformer-based classification models might recognize the review patterns better than the reviews written by humans. However, considering the performance of a wide range of experiments on two different datasets, we can conclude that DistilBERT and XLNet achieved new state-of-the-art results in identifying fake reviews, both for human and machine-generated fake reviews.

Table 1
The performance of different methods compared to baselines on Fake Review and Yelp Dataset.

Type	Model	Fake Review Dataset			Yelp Dataset		
		Accuracy	Precision	F1Score	Accuracy	Precision	F1Score
Baseline	EnsembleML	0.8425	0.9147	0.9014	0.7848	0.7795	0.8156
Deep Learning	BiLSTM	0.9556	0.9750	0.9466	0.8947	0.8985	0.9444
	CNN	0.9252	0.9268	0.9259	0.8961	0.8966	0.9451
	CNN-LSTM	0.9486	0.9454	0.9493	0.8842	0.9007	0.9380
	CNN-GRU	0.9476	0.9751	0.9466	0.8964	0.8978	0.9452
Transformers	DistilBert	**0.9592**	**0.9906**	**0.9821**	0.9235	**0.9326**	0.9595
	XLNet	0.9580	0.9887	0.9779	**0.9349**	0.9278	**0.9654**

3.3. Explainability of the predictions

We implemented LRP technique for generating explanations with the same setting as detailed in [3]. The color intensity in highlighted text and size of the words *WordCloud* represent the degree of relevancy towards the class. The explanation is shown in Fig. 1. We can see that the highlighted words are related to the predicted class. The highlighted text and word cloud also show that words such as *read, chance, enjoy, liked, loved stars* are some most relevant for the prediction. We have a closer look at the review text, it is a review that exaggeratedly praises the book. The highlighted words are used for exaggerated praise.

3.4. Empirical User Evaluation

To evaluate the effectiveness of the LRP-generated explanations highlighting the important relevant words to the predicted class, we conducted an empirical user study with 12 human subjects. The subjects are studying master's in business informatics. We first give them overview of how our transformer-based model predicts the authenticity of the review. Then we provide them with a simple demo about the explanations and what those highlighted words mean.

We provided them three reviews (Review 1, Review 2, and Review 3) and asked them to score how authentic the reviews were. All three reviews were fake but we have not told them. Because we wanted to observe how they identify and what are the logic behind. We also provided the details about the products for which the reviews were posted. They were asked to put score for each review, and the score ranges from one star (*) to five star (*****). The highest value 5 (*****) indicates that the corresponding review is original, while the lowest value 1 (*) indicates the review is fake. The participants first score each review after carefully reading the reviews without the generated explanations.

We then provided them with the LRP-generated explanations for each review. We then instructed the participant to look at the explanations and re-score the reviews whether their assumptions changed

first let me say i'm an avid reader and this is a book that i read as a child i had to read it before i could have a chance to take it to colle ge i still enjoy reading it as a kid this book is still one of my favorit e books i have read the book over and over again and it is a must re ad i just can't put it down the only reason i gave it five stars is beca use i want to read more about the characters i liked the way they in teracted with the kids i loved their reactions to

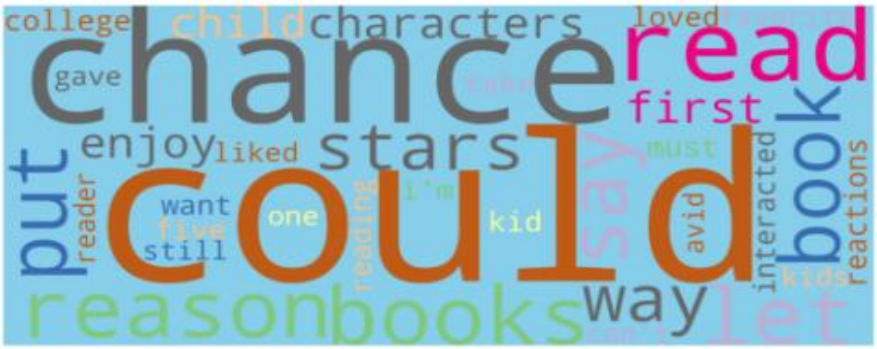

Figure 1: Explanation with highlighting relevant words for a predicted fake review.

after perceiving the explanations. We denote two scores before and after the explanations as *score 1* and *score 2*, respectively. We then discuss with each participant why they think that a particular review is original or fake. What are the reasons and rationale behind their scores? We also asked them about the efficiency or meaningfulness of the generated explanations and how they help the participants decide on the authenticity of the reviews.

Table 2
The user evaluations whether the reviews are fake or real, with and without explanations.

Subject	Review 1		Review 2		Review 3	
	Score1	Score2	Score1	Score2	Score1	Score2
1	**	**	****	***	**	**
2	**	**	****	****	****	****
3	*	*	*	*	**	**
4	***	**	*	*	****	***
5	*	*	*	*	****	****
6	**	**	**	***	*****	*****
7	**	**	*	**	****	****
8	*	*	*	**	*****	**
9	*	*	*	**	***	****
10	*	*	***	***	**	**
11	*	*	*	**	***	***
12	***	*	**	***	*****	*****

Table 2 represents the evaluation of the participants on whether those three reviews are original or fake. We can see that all participants thought that review 1 was fake except subjects 4 and 12. They provide three stars out of five, which concludes it is somewhat original. However, they changed their decision after having the explanations by putting two and one star, respectively. For review 2, except for participants 1 and 2, everyone considered the review to be fake. Interestingly, review 3 were considered solely as original by the majority of the participants. However, after considering the explanations generated by the LRP-enabled explainability technique, two participants (subject 4 and 8) changed their decision by decreasing the mark.

Discussion on participants' opinion: We had a detailed discussion with each participant on how they came up with the decision whether a particular review was fake or original. For example, we asked

the subject about the review 3. He said the following:

Subject 2: "The third review, because it was for me it was the most realistic. There was the name inside. So he seems to know the guy who's doing it and it's pretty, and it's really short."

He thought it was short and he believed in the text because it has a name. However, we asked him, what matters in predicting the review whether it is real or fake. He replied, its more about *linguistic form* (i.e., meaning grammatical structure and tone), not individual words.

Subject 2: "Yeah, and the second it's, uh, more about the words. And in the first, first, it's more about the linguistic form to me."

Similarly, Subject 3 also thought that highlighting relevant words as an explanation might not make sense in explaining review identification whether it is fake or real. It's about the whole text. He added, for the second review, based on the repetitive texts he identified review 2 as fake.

Subject 3: "Uh, the second one is, I also think it's completely made up by AI, um, because it's very repetitive and, uh, uh, some sentences you just read and you think no human would write like this. Um, and then the third one to me also was the most realistic one because it is kind of short. It's very, it kind of seems authentic in terms of like the excitement."

Subject 3: "No, to me, it's not the singular words. To me, it really is the structure and the whole like, the thing as a whole that, um, makes it seem like it's AI generated"

Subjects 4 and 5 provided their insight about whether our generated explanations make sense to understand the decision. They both thought that the current form of explanation might help to some degree to comprehend the decision, 2 out of 10.

subject 4: : "I think it might, it might make sense to some degree. But as, uh, my colleague just said, it's more about the, the overall. Two out of 10"

Subject 5: "No, no. Not from my part. I have to reread that, like out of 10. Uh, like two or three."

Subject 6 has found something very interesting in review 2, for example, information like age, and jobs are not relevant and these are not commonly used in review. He also identified that this is a very long review, generally, people are too lazy to write.

Subject 6: "Because no matter, um, his age or his, um, job and something like this or for buying gloves, um, and also it's, um, too long. And I guess, um, people are, most people are lazy to write this kind of message. Yeah."

Grammatical information is identified as important to understand whether the review is fake or real by subject 7. He considered the more the number of adjectives that exist in the review, the more realistic the review is.

Subject 7: "No, just any adjective. So for example, the ones that I have rated the, the, the realest, have more objectives than, than the other ones. So it could be just, um, your personal opinion, it's not about."

He also thought the individual word might have some importance towards certain classes, but it should be the whole context of the review.

Subject 7: "For me, for me, they didn't really help me to find out if they are or not real. Uh, I think it's more like a context thing. Only, I mean for me the word has, has to, um, it's okay. It was the, the same."

Interestingly Subject 8 found our generated explanations are effective. Before accessing the explanations, subject 8 provided review 3 as 5 starts, meaning a fully real review. But after he went through the generated explanations, he thought this was also a fake review. Though it has several good adjectives, but he thinks these are the reasons to be fake, contradictory to subject 7.

Subject 8: "So, the third one actually was from me, at the first, um, I gave them five stars. So that it's likely to happen because it's short. I think in real life everyone just give short recommendations and not long recommendations. But then after reading the AI, um, explanation, yes. Um, reading the words excellent, amazing, great. I also think that it's, it's not a real review."

In summary, the explanations generated by LRP technique to highlighting important relevant words in context of fake review identification can make general users sense in minimum scale. On contrary, for some application areas, for example, sentiment classification [12] and hate speech recognition [8], where LRP-based explanations are quite good to understand the reason behind the prediction. One of the main reasons identified through the empirical user study why explanations highlighting relevant words is the use of similar words in both fake and original review. For both positive and negative

reviews, we observed similar adjectives or other praising or criticizing words are used in both fake and original reviews. For example, in sentiment analysis task, there are some terms or negation elements that are used for specific positive and negative class [12]. For another example, patent classification task [3], terms related to specific scientific fields are used in the patent text. Rather, in the context of fake review identification task, grammatical structure of the sentence, tone and overall context matters most in explaining the decisions.

4. Conclusion and Future Direction

In this paper, we proposed transparent and interpretable fake review detection framework applying transformer models including DistilBERT and XLNet. We also apply multiple deep learning models including BiLSTM, CNN, CNN-LSTM, and CNN-BiLSTM for modeling the fake review detection task. Then, we adopted LRP technique to open the the black-box deep learning model. The LRP can explain why a particular prediction has been made. We conducted experiments in multiple settings and applied our models to two different benchmark datasets. Based on the experimental results, we demonstrated that our proposed DistilBERT- and XLNet-based fake review detection models significantly outperformed other ensemble ML and DL models. Compared to the previously known related methods, our method also outperformed NBSVM, OpenAI, and fakeRoBERTa methods on the same dataset. We conducted an user study with 12 participants and investigate how useful the generated explanations to understand the prediction, whether the review is fake or original. In the end, we demonstrated explanations provided by our adopted LRP technique for multiple example reviews for different categories. The empirical user evaluation with human subjects indicates further requirements to generate and present the explanations for any specific decision. In the future, we are planning to have an empirical study to measure the quality of the generated explanations. Further, it would be interesting to consider the elicited requirements and findings from user evaluation for explanation generation.

Acknowledgement

This research has been funded by the AntiScam Project (Defense against communication fraud), funded by BMBF Germany, Grant reference 16KIS2214

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

[1] J. Salminen, C. Kandpal, A. M. Kamel, S.-g. Jung, B. J. Jansen, Creating and detecting fake reviews of online products, Journal of Retailing and Consumer Services 64 (2022) 102771.

[2] M. Ngueajio, S. Aryal, M. Atemkeng, G. Washington, D. Rawat, Decoding fake news and hate speech: A survey of explainable ai techniques: A survey of explainable ai techniques., ACM Computing Surveys (2025).

[3] M. Shajalal, S. Denef, M. R. Karim, A. Boden, G. Stevens, Unveiling black-boxes: Explainable deep learning models for patent classification, in: World Conference on Explainable Artificial Intelligence, Springer, 2023, pp. 457–474.

[4] R. A. Duma, Z. Niu, A. S. Nyamawe, J. Tchaye-Kondi, A. A. Yusuf, A deep hybrid model for fake review detection by jointly leveraging review text, overall ratings, and aspect ratings, Soft Computing 27 (2023) 6281–6296.

[5] H. Paul, A. Nikolaev, Fake review detection on online e-commerce platforms: a systematic literature review, Data Mining and Knowledge Discovery 35 (2021) 1830–1881.

[6] N. Deshai, B. Bhaskara Rao, Unmasking deception: a cnn and adaptive pso approach to detecting fake online reviews, Soft Computing (2023) 1–22.
[7] G. Bathla, P. Singh, R. K. Singh, E. Cambria, R. Tiwari, Intelligent fake reviews detection based on aspect extraction and analysis using deep learning, Neural Computing and Applications 34 (2022) 20213–20229.
[8] M. R. Karim, S. K. Dey, T. Islam, S. Sarker, M. H. Menon, K. Hossain, M. A. Hossain, S. Decker, Deephateexplainer: Explainable hate speech detection in under-resourced bengali language, in: 2021 IEEE 8th International Conference on Data Science and Advanced Analytics (DSAA), IEEE, 2021, pp. 1–10.
[9] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, Q. V. Le, Xlnet: Generalized autoregressive pretraining for language understanding, Advances in neural information processing systems 32 (2019).
[10] M. Shajalal, A. Boden, G. Stevens, Explainable product backorder prediction exploiting cnn: Introducing explainable models in businesses, Electronic Markets 32 (2022) 2107–2122.
[11] M. R. Karim, T. Islam, M. Shajalal, O. Beyan, C. Lange, M. Cochez, D. Rebholz-Schuhmann, S. Decker, Explainable ai for bioinformatics: Methods, tools and applications, Briefings in bioinformatics 24 (2023) bbad236.
[12] L. Arras, F. Horn, G. Montavon, K.-R. Müller, W. Samek, " what is relevant in a text document?": An interpretable machine learning approach, PloS one 12 (2017) e0181142.
[13] L. Arras, G. Montavon, K.-R. Müller, W. Samek, Explaining recurrent neural network predictions in sentiment analysis, arXiv preprint arXiv:1706.07206 (2017).
[14] S. M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in neural information processing systems 30 (2017).
[15] M. T. Ribeiro, S. Singh, C. Guestrin, " why should i trust you?" explaining the predictions of any classifier, in: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining, 2016, pp. 1135–1144.
[16] S. Ramnath, P. Nema, D. Sahni, M. M. Khapra, Towards interpreting BERT for reading comprehension based QA, in: B. Webber, T. Cohn, Y. He, Y. Liu (Eds.), Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Association for Computational Linguistics, Online, 2020, pp. 3236–3242. doi:10.18653/v1/2020.emnlp-main.261.
[17] W. Choi, K. Nam, M. Park, S. Yang, S. Hwang, H. Oh, Fake review identification and utility evaluation model using machine learning, Frontiers in artificial intelligence 5 (2023) 1064371.
[18] A. M. Elmogy, U. Tariq, M. Ammar, A. Ibrahim, Fake reviews detection using supervised machine learning, International Journal of Advanced Computer Science and Applications 12 (2021).
[19] S. Yu, J. Ren, S. Li, M. Naseriparsa, F. Xia, Graph learning for fake review detection, Frontiers in Artificial Intelligence 5 (2022) 922589.
[20] N. A. Patel, R. Patel, A survey on fake review detection using machine learning techniques, in: 2018 4th international Conference on computing Communication and automation (ICCCA), IEEE, 2018, pp. 1–6.
[21] R. Singhal, R. Kashef, A weighted stacking ensemble model with sampling for fake reviews detection, IEEE Transactions on Computational Social Systems (2023).
[22] R. Mohawesh, S. Xu, M. Springer, Y. Jararweh, M. Al-Hawawreh, S. Maqsood, An explainable ensemble of multi-view deep learning model for fake review detection, Journal of King Saud University-Computer and Information Sciences (2023) 101644.
[23] Z. Shunxiang, Z. Aoqiang, Z. Guangli, W. Zhongliang, L. KuanChing, Building fake review detection model based on sentiment intensity and pu learning, IEEE Transactions on Neural Networks and Learning Systems (2023).
[24] Z. Wang, H. Li, H. Wang, Vote-based integration of review spam detection algorithms, Applied Intelligence 53 (2023) 5048–5059.
[25] D. K. Vishwakarma, P. Meel, A. Yadav, K. Singh, A framework of fake news detection on web platform using convnet, Social Network Analysis and Mining 13 (2023) 24.