Soft Computing (2023) 27:11357–11378
https://doi.org/10.1007/s00500-023-08507-z

OPTIMIZATION

Unmasking deception: a CNN and adaptive PSO approach to detecting fake online reviews

N. Deshai[1] · B. Bhaskara Rao[1]

Accepted: 10 March 2023 / Published online: 3 June 2023

Abstract
Online reviews play a critical role in modern word-of-mouth communication, influencing consumers' shopping preferences and purchase decisions, and directly affecting a company's reputation and profitability. However, the credibility and authenticity of these reviews are often questioned due to the prevalence of fake online reviews that can mislead customers and harm e-commerce's credibility. These fake reviews are often difficult to identify and can lead to erroneous conclusions in user feedback analysis. This paper proposes a new approach to detect fake online reviews by combining convolutional neural network (CNN) and adaptive particle swarm optimization with natural language processing techniques. The approach uses datasets from popular online review platforms like Ott, Amazon, Yelp, TripAdvisor, and IMDb and applies feature selection techniques to select the most informative features. The paper suggests using attention mechanisms like bidirectional encoder representations from transformers and generative pre-trained transformer, as well as other techniques like Deep contextualized word representation, word2vec, GloVe, and fast Text, for feature extraction from online review datasets. The proposed method uses a multimodal approach based on a CNN architecture that combines text data to achieve a high accuracy rate of 99.4%. This outperforms traditional machine learning classifiers in terms of accuracy, recall, and F measure. The proposed approach has practical implications for consumers, manufacturers, and sellers in making informed product choices and decision-making processes, helping maintain the credibility of online consumer reviews. The proposed model shows excellent generalization abilities and outperforms conventional discrete and existing neural network benchmark models across multiple datasets. Moreover, it reduces the time complexity for both training and testing.

Keywords Fake reviews · Adaptive particle swarm optimization · Natural language processing · Convolutional neural network · Deep contextualized word representation

1 Introduction

In today's digital age, customers hold immense power to express their opinions and views through various websites. As a result, online reviews have become a crucial aspect for both customers and businesses in their decision-making processes. Reviews posted online have the potential to significantly impact a business's reputation and serve as a crucial parameter for evaluating the quality of products and services. However, the popularity of online reviews has also given rise to the unethical practice of fake review writing, where paid human writers produce deceptive reviews to manipulate readers' opinions. Such practices can have severe repercussions and adversely affect the credibility of the entire review system. Therefore, there is an urgent need for advanced research work to detect and combat fake online reviews, ensuring that customers can make informed decisions, and businesses can maintain their credibility (Zhaoa and Sunb 2022).

The Covid-19 pandemic that swept across the world in early 2020 has had a transformative effect on many aspects of life, including the world of global e-commerce and online shopping. However, one of the most significant impacts has been the proliferation of fraudulent reviews posted by customers regarding products or services offered by a brand or organization. This problem, commonly referred to as "Opinion Spamming," has become

✉ N. Deshai
desaij4@gmail.com

[1] Department of C.S.E.Gitam IT, GITAM University, Visakhapatnam, Andhra Pradesh 530045, India

increasingly complex and organized, as individuals and groups seek to profit from these activities (Salminen et al. 2022).

In today's digital age, online consumer reviews have become a crucial aspect of decision-making for new consumers looking to try out a business or product (Zhang et al. 2023a). The abundance and accessibility of online reviews empower consumers to evaluate a product or service's quality and value based on the experiences of others. As a result, online reviews have a significant impact on a business's reputation and success, making it critical for companies to maintain a positive online presence and monitor and address any negative feedback.

Online reviews have become an essential component of the modern consumer landscape, providing valuable insights into the quality and performance of products and services (Tufail et al. 2022). With the proliferation of e-commerce platforms, social media, and online communities, consumers have access to a wealth of information and feedback from other users, making it easier than ever to research products and make informed purchase decisions. Online reviews have gained significant importance as a primary source of consumption information in modern society, playing a crucial role in shaping consumer attitudes, preferences, and behaviors. They provide a platform for users to share their experiences and opinions, helping to build trust and credibility among consumers (Chatterjee et al. 2023). Moreover, online reviews are not only beneficial for consumers, but also for businesses, as they can help to increase visibility, sales, and brand loyalty. Positive reviews can serve as powerful endorsements, while negative reviews can provide valuable feedback for improvement and quality assurance. As such, online reviews have transformed the way we shop and consume, creating a more informed and empowered consumer base, while also driving innovation and competition in the marketplace (Zhang et al. 2023b).

Social media usage has increased to 4.48 billion users, with one-third using it regularly. A recent purge by Amazon involved the deletion of around 20,000 fake reviews (Hassan and Islam 2019). Consumers rely heavily on online reviews with 87% reading them for local businesses, and 79% trusting them as much as personal recommendations (Rout et al. 2017a). This emphasizes the need for businesses to manage their online reputation and combat fraudulent reviews to ensure trust in online reviews (Zhang et al. 2023c).

The article discusses the evolution of Natural Language Processing (NLP) and how it has come a long way from its inception to the present day, with 0.958 on the Amazon dataset and an F1-score of 0.955 on the Yelp dataset (Nasir et al. 2021). Additionally, a deep-learning approach using BERT and XLNet was proposed by the author, achieving a 95.15%. The BERT model has achieved outstanding results in detecting fake online reviews, surpassing other approaches with a 91% accuracy rate on a well-balanced dataset of reviews from various domains (Valliappan and Ramya 2023). Even more impressively, it achieved a 73% accuracy rate on an imbalanced third-party Yelp dataset of restaurant reviews, marking a significant stride in combating fake online reviews and offering promising prospects for creating more trustworthy and accurate review systems (Rout et al. 2017b). Our findings also demonstrated that achieving good results with real-world datasets, such as Yelp, using the bag-of-words (BOW) method requires at least 500,000 features. To evaluate the accuracy of fake review detection by BERT and GPT-2 models, we used the Yelp dataset and achieved an accuracy rate of 96.12% (Hajek and Sahut 2022). We also proposed using BERT for fake review detection, which resulted in an F1-score of 0.958 on the Amazon dataset and an F1-score of 0.955 on the Yelp dataset. Furthermore, we introduced a deep-learning approach using BERT and XLNet, achieving a 95.15% accuracy rate on the Yelp dataset (Budhi et al. 2021). Another method we explored used BERT and Siamese Networks, resulting in an F1-score of 0.951 on the Amazon dataset and an F1-score of 0.935 on the Yelp dataset (Paul and Nikolaev 2021). However, we discovered that the accuracy levels for the fake and genuine review classes were significantly imbalanced in the two large datasets (YelpNYC and YelpZIP). Numerous research studies have utilized CNN (Convolutional neural network) for detecting fake online reviews, which can extract features from textual data and classify them as genuine or fake. One study achieved 92.7% and 97.3% accuracy in detecting fake reviews on Yelp and Amazon, respectively, using CNN (Vidanagama et al. 2020). Another study combined CNN with other machine learning algorithms to improve the accuracy of fake review detection on TripAdvisor, achieving an accuracy rate of 87.0%. Combining CNN with other machine learning algorithms has demonstrated potential in detecting fake online reviews, such as the approach taken by Martínez Otero (2021). Birim et al. (2022) who achieved high accuracy rates on various datasets. In our approach to detecting fake reviews, we utilize APSO and CNN techniques, where APSO is an optimization algorithm used to identify critical features in text data, and CNN applies convolutional filters to detect patterns that indicate deception. By combining these techniques, we can create a powerful and adaptive system that achieves high accuracy in detecting fake reviews, even when fraudsters use sophisticated techniques. Our approach involves using machine and deep learning classifiers, including single and ensemble models, and developing specialized feature extraction methods that primarily analyze the linguistic characteristics of the text review and

reviewer behavior. We also address imbalanced data and implement a parallel version of n-fold cross-validation to expedite the investigation process. This approach has the potential to improve the accuracy and reliability of consumer feedback and maintain the integrity of online marketplaces.

In this paper, the following structure is adopted: Sect. 2 offers a summary of related studies pertaining to the research. Section 3 outlines the materials and methods employed in the study. A detailed description of the proposed methodology is presented in Sect. 4. The findings of the research are presented and analyzed in Sect. 5. Lastly, Sect. 6 provides a conclusion to the research findings. A list of abbreviations used in the paper is included at the end for reference.

2 Related work

The rapid advancement of technology in today's world has brought both advantages and disadvantages in the digital realm. One major drawback is the pervasive issue of fake and spam reviews, which significantly impact consumers' purchasing decisions. Customers rely heavily on online social networks such as Amazon, Yelp, and TripAdvisor to evaluate the quality of products and services. Online consumer reviews have become a critical source of information, covering a wide range of categories such as restaurants, products, hotels, and more (Narciso 2022). However, the credibility of these reviews is often undermined by businesses that create fake reviews, posing a significant challenge for consumers who rely on online reviews to make informed purchase decisions. The ubiquity of social media has provided an avenue for individuals to express their opinions and disseminate information, leading to the proliferation of fake reviews (Duma et al. 2023). To address this issue, various studies have proposed different techniques for detecting fake reviews, including the use of keywords, punctuation marks, entity recognition, and sentiment analysis. Furthermore, deep learning models based on feed-forward neural networks and LSTM have shown promise in accurately detecting fake reviews. Recent research has proposed a multitude of approaches to identify and mitigate the impact of fake reviews on social media platforms (Ren et al. 2017). Online reviews have gained immense popularity among consumers, who can conveniently access product or service-related feedback online. The trustworthiness of online reviews plays a crucial role in influencing customers' purchase intentions. However, the authenticity of online reviews can be compromised by fake reviews, which do not reflect genuine product experiences. Hence, detecting and preventing fake reviews has become increasingly crucial to maintain the credibility of online reviews and building trust among customers (Kurtcan and Kaya 2022). To detect fake reviews on a Chinese e-commerce platform using an unsupervised matrix iteration algorithm. The algorithm calculates fake degree values at the individual, group, and merchant levels. The test data set consists of 97,804 reviews from 93 online stores and 9,558 reviewers selected randomly. The proposed method achieves high accuracy with F-measure values of 82.62%, 59.26%, and 95.12% in detecting fake reviewers, online merchants, and groups with reputation manipulation, respectively (Arif et al. 2018). Fake reviews are intentionally written to deceive potential buyers and are often authored by individuals who have no experience with the products or services in question. Word of mouth, which is the personal communication between individuals regarding their perceptions of goods and services, is a crucial factor that influences consumer behavior. In recent years, there has been a significant shift towards online shopping, with over 60% of respondents in the Asia, Africa/Middle East, and Latin America regions expressing a willingness to shop online in the future (Kaghazgaran et al. 2017). The growth of e-commerce is reflected in the increasing online sales figures, with total U.S. retail e-commerce sales reaching US$105.7 billion in the first quarter of 2017, representing a 4.1% increase from the fourth quarter of 2016 (Elmurngi and Gherbi 2018). Experts predict that online sales will continue to grow, with an annual rate of 9.3% expected by 2020. Additionally, online grocery sales are projected to reach nearly US$100 billion by 2019 in the United States (Mohawesh et al. 2021). Research has presented a sophisticated two-phase approach to identify fraudulent online reviews across Amazon, Yelp, and TripAdvisor. In the first phase, the unstructured data was transformed into structured data, and in the second phase, the dataset underwent rigorous scrutiny by employing twenty-three supervised AI models. Moreover, researchers have proposed a cutting-edge deep learning model to detect fake online reviews by leveraging a Kaggle dataset (Rajamohana et al. 2017). To prepare the data for analysis, word embedding techniques (GloVe) were employed to construct a vector space of words and establish linguistic relationships. BERT and GPT are two significant natural language processing (NLP) models developed by Google's research team in recent years (Catal and Guldan 2017). BERT is known for its ability to capture the context of words in a sentence, while GPT can generate high-quality language representations using a decoder-only Transformer model. Both models have been used to improve the accuracy of NLP tasks, including fake online review detection. The attention mechanisms in BERT and GPT enable them to identify important words and phrases, assign higher weights to them, and focus on relevant parts of the input data (Goswami et al. 2017). Their feature extraction and

attention mechanisms make them useful for identifying the most important parts of text and distinguishing between genuine and fake reviews. The classification model suggested in the study utilized convolutional and recurrent neural network architectures, further enhancing the robustness of the proposed methodology. This paper delves into the intricate world of detecting malicious rumors, fake online reviews, and misinformation on social media platforms. To this end, researchers have proposed a plethora of innovative techniques including deep learning models, graph-based approaches, sentiment analysis, and entity recognition. These techniques have been applied to large-scale datasets and have yielded impressive results in terms of accuracy and performance. Notably, some studies have concentrated on identifying fake reviews on social media platforms using cutting-edge machine and deep learning techniques, sentiment analysis, and fact-checking websites (Brar and Sharma 2018). The range of methodologies and datasets explored in these studies demonstrates the complexity and importance of effectively detecting and combatting fake information on social media platforms. The primary objective of this study is to uncover the salient features of online reviews on social media and online platforms by harnessing the power of evolutionary-based techniques. In pursuit of this aim, four cutting-edge evolutionary classification techniques, namely LSTM, RNN, ANN, CNN.

2.1 Problem statement

One of the primary challenges in online review analysis is the context-dependent nature of language. The same word or phrase may have different connotations in different contexts, leading to incorrect sentiment analysis. For instance, the term "long" may describe a laptop's battery life positively, but the same term may have negative connotations when describing its start time (Dhingra and Yadav 2017). Therefore, opinion mining algorithms require assistance in identifying the context in which words are used to determine their sentiment accurately. Another challenge is the diversity of expression in people's opinions, making it difficult for traditional text processing techniques to determine the underlying sentiment correctly. Even slight differences in phrasing or word choice can significantly affect the meaning of a statement, particularly in opinion mining. Furthermore, people often express mixed or contradictory views about a product or service, making their opinions challenging to interpret accurately (Krishna 2019). For instance, a reviewer may have positive and negative comments about a product, which can be challenging for opinion-mining algorithms to comprehend. Finally, identifying fake reviews or fraudulent reviewers can be challenging. E-commerce sites and service providers must detect opinion spamming or fraudulent reviews to maintain the trust of consumers. However, fraudsters can use techniques such as emoticons or other punctuation marks, which may be removed during the data cleaning process, making it difficult to identify fake reviews accurately. Despite these challenges, it is essential to identify fraudulent reviews and reviewers to ensure that consumers can make informed decisions when purchasing products or services online.

2.2 Motivation and research goal

The problem of detecting fake online reviews has garnered significant interest from researchers worldwide, given the prevalence of social media platforms as a popular medium for accessing such reviews. Existing detection methods have predominantly relied on analyzing content or social context-based information extracted from news articles. However, there is still a pressing need for an efficient detection model capable of handling both content and community-level features with a tensor factorization approach. This research seeks to fill this gap by proposing an effective deep-learning model that can accurately detect and classify fake and real reviews.

3 Methodology

3.1 Dataset collection

In our research, we used the Ott dataset, a labeled dataset that includes truthful and deceptive hotel reviews of 20 hotels in Chicago (Wang et al. 2020). This dataset has 1600 reviews, equally divided into 800 truthful and 800 deceptive reviews. It was challenging to find a tagged dataset, and we could only find one labeled public dataset. The Yelp dataset, an unlabeled dataset, was used in our study. We collected the first 2000 review instances from the Yelp dataset, preprocessed them, and labeled them through an active learning process. 350 of the 2000 instances were labeled as "Spam," and 1650 were labeled as "Ham." Additionally, we compiled uncategorized reviews of other Amazon goods, Yelp, Tripadvisor, and IMDb to gather more data. We collected these reviews from Amazon's website, and they included product reviews, ratings, and metadata (Sa et al. 2017).

3.2 Data pre-processing

This study employs Data Acquisition and Data Pre-processing, Active Learning Algorithm, Feature Selection, and spam detection using traditional machine learning and deep learning classifiers. Pre-processing includes natural

language processing techniques, while feature selection utilizes TF-IDF, n-grams, and Word2Vec (Zhang et al. 2016). The spam detection phase involves SVMs, KNNs, NB, MLP, LSTMs, RNN, ANN and CNNs, and. In the pre-processing phase, text filtering removes less valuable elements like punctuation symbols to improve classification accuracy.

NLTK tokenizes the sentence and tags each word with its part of speech. Feature selection considers parameters like length count, bigram type, and sentiment word count. The dataset is split into training and testing samples using an 80–20 ratio to ensure model effectiveness and accuracy in classifying fake and genuine reviews. During the data pre-processing phase, cleaning steps were taken to remove undesired parts of the data. The first step was feature extraction, where the extracted tweets were tokenized to transform the text into machine-consumable forms, such as words, phrases, and sentences (Han et al. 2023). Three models, BoW, TF, and TF-IDF, were used to extract talk features and improve classification output. Next, stop words, such as "the", "and", "but", "or", etc., were removed to reduce data dimensionality and improve the efficiency of the classification model. This step involved eliminating both common and exclusive words. After tokenization and stop word removal, different stemmers, such as Snowball, Lovins, Porter, Dawson, Lancaster, and WordNet, were applied to reduce the data dimensionality and identify similar words in different forms (Yu et al. 2019). Table 1 displays the average processing time in milliseconds for the pre-processing step of various datasets, which includes stop word removal, word correction, and lemmatization. Stop word removal eliminates commonly used words that lack significant meaning, while word correction corrects spelling errors, and lemmatization reduces words to their base form. The table reveals significant variations in the average pre-processing time across different datasets. For example, the OTT dataset requires the least time, with an average of 0.79 ms, while Amazon data demands the most time, averaging 86.44 ms. This indicates that Amazon data requires considerably more pre-processing time than other datasets. The outcomes of the pre-processing stage are crucial for detecting fake online reviews. By removing irrelevant information and noise from the text, pre-processing makes it simpler for the model to identify patterns and classify reviews accurately. The findings from Table 1 can be utilized to optimize the pre-processing step for different datasets, depending on their average processing time, which can enhance the model's overall performance and lead to more precise detection of fake online reviews.

Detecting fake online reviews is a complex and challenging task, requiring several hyperparameters to be fine-tuned. Table 2 Batch size, optimizer, number of epochs, dropout rate, and Softmax function were explored in this study. Using a batch size of 128 achieved the best results, suggesting that larger batches are more effective in reducing variance and obtaining a more stable optimization trajectory. Among several optimizers, Adam performed the best with a learning rate of 0.001, and longer training times were found to produce optimal results at 50 epochs. A dropout rate of 0.8 was selected to prevent overfitting. The output logits across all labels were generated, and the probability of a specific sample belonging to one label was optimized when estimated by the Softmax function. Using these hyperparameters, a deep neural network was trained on labeled datasets to distinguish between genuine and fake online reviews across platforms such as OTT, Amazon, Yelp, TripAdvisor, and IMDb. The model's performance was evaluated using precision, recall, and F1-score metrics, leading to a highly effective and accurate system for detecting fake reviews. This model is an important tool for online review platforms to maintain the authenticity and reliability of reviews, leading to increased consumer trust and confidence. There are various stemming algorithms available for natural language processing, such as Snowball, Lovins, Dawson, Porter, and Lancaster. Snowball is an improved version of the Porter algorithm and can handle multiple languages, while Lovins uses substitution rules for English text but has some limitations. Dawson is a more complex algorithm that considers the context and part of speech of a word and is effective for non-English languages (Chua and Chen 2022). Porter is widely used and easy to implement, but it may over-stem or under-stem certain words. Lancaster is aggressive and produces shorter stems but may generate non-words. WordNet lemmatization considers part of speech and meaning to reduce words to their base form. To enhance the performance of the evolutionary classifier, unnecessary characters, such as extra spaces, punctuation marks, and symbols, were removed, and stop words were eliminated. In Table 3, the combination of hybrid techniques, including Tokenization, Lemmatization, and Stemming, enables the capture of both syntactic and semantic aspects of the text. These techniques are crucial for pre-processing the data from to effectively analyze and detect fake online reviews. It is recommended Ott, Amazon, Yelp, TripAdvisor, and IMDb datasets to experiment with different combinations and fine-tune these techniques based on the specific datasets and detection objectives to achieve optimal results. By carefully selecting and refining the hybrid techniques, you can enhance the accuracy and effectiveness of your fake review detection approach. The average processing time for this pre-processing step is shown in Table 2.

Table 1 Average processing time of pre-processing step

Dataset name	Average processing time of sample (ms)			
	Stop word removal	Word correction	Lemmatization	Total pre-processing
Ott	0.04	0.77	0.04	0.79
Yelp	0.32	9.55	0.36	9.63
Amazon	1.53	86.43	2.68	86.44
TripAdvisor	1.56	52.42	1.69	52.44
IMDb	0.23	0.89	0.08	0.92

Table 2 Parameters of deep learning classifiers

References	Model
Word vector dimension	300
Barch size	128
Layer filters	64
Kernel size	7
Padding	Valid
Pool size	4
Pooling activation function	ReLU, Maxout, Softmax, Sigmoid
Hidden units in the CNN layer	1024
Hidden layer	128
Output layer	4
Dropout rate	0.7, 0.8
Loss	Categorical_cross entropy, regression, object detection
Optimizers	Adam, RMSProp, Stochastic Gradient
Classifier	Softmax

3.3 Feature extraction

In this study, feature selection and feature extraction are used to create an informative feature subset from the original feature space, which may contain irrelevant, redundant, and noisy features (Moqueem et al. 2022). Feature selection methods score and choose the feature subset, while feature extraction methods transform the feature space into a lower dimension. Information gain, an entropy-based measure, was used to score each feature in this study. Three different feature selection methods were used to select a subset of 25 features. A mathematical formulation of the information gain is mentioned below:

$$\text{Gain}(Y, X) = H(Y)H(Y|X) \quad (1)$$

where H(Y) denotes the entropy of dataset. H(Y|X) denotes conditional entropy. Mathematical formulation of H(Y) and H(Y |X) is given below:

$$H(Y) = \sum p(y)\log p(y) \quad (2)$$

$$H(Y|X) = \sum_{x \in X} p(x)H(Y|X = x) \quad (3)$$

The paper used six popular feature extraction techniques, including bag-of-words (BoW), term frequency (TF), term frequency-inverse document frequency (TF-IDF), Word2Vec, GloVe, and FastText, to detect the most essential features of social media platform pandemic information (Jiang et al. 2020).

BoW is a method that treats a group of words as a collection of words, without considering syntactical or semantic dependencies. Term frequency counts how often a specific term appears in a text by dividing the number of occurrences by the total number of keywords in the document (Krishnan et al. 2022). To address the problem of unimportant common words, TF-IDF assigns more weight to rare words than to common ones in all documents. TF-IDF calculates the frequency of a word in the current document (TF) and assesses how rare the word is across all documents (IDF). The paper proposed a methodology consisting of five stages for developing and evaluating a model: data collection, data pre-processing and feature extraction, model development, and evaluation and assessment. The details of each stage can be seen in Fig. 1.

To achieve high accuracy in detecting fake online reviews, it is important to consider the specific characteristics of the task and Ott, Amazon, Yelp, TripAdvisor, and IMDb datasets.

Bag-of-Words (BoW) + Word Frequency: This technique is a good starting point as it provides a simple

representation of the text. It can capture the frequency of words, which may be indicative of fake reviews. However, it may not consider semantic relationships or the order of words.

Document-Term Matrix + TF-IDF: This technique builds on the BoW representation by incorporating TF-IDF weighting. It considers the importance of words in the document and the entire dataset. This can help to emphasize significant terms and potentially improve the accuracy of fake review detection.

Sentence Embeddings + Word2Vec (CBOW or Skip-gram): Sentence embeddings capture the overall meaning of sentences, while Word2Vec models learn word embeddings that encode semantic relationships. This combination allows for understanding both the context of sentences and the meaning of individual words. It can be effective in capturing the semantics of reviews and identifying suspicious patterns.

Named Entity Recognition (NER) Features + Part-of-Speech (POS) Features: NER features identify named entities, such as product names or brand mentions, which can be relevant in detecting fake reviews. POS features provide grammatical information that can help identify patterns associated with fake reviews. Combining these features can enhance the accuracy of the detection process.

Sentiment Lexicons + Document Embeddings: Sentiment lexicons provide information about the sentiment polarity of words, while document embeddings capture the overall sentiment of the entire review. By combining these features, you can consider both the sentiment of individual words and the overall sentiment expressed in the review. This can be useful for distinguishing between genuine and fake reviews.

3.3.1 BERT

BERT is a powerful NLP model used for fake online review detection. It uses masked language modeling during pre-training to understand the context and meaning of words in a sentence (Lo Presti and Maggiore 2021). BERT is fine-tuned on a dataset of labeled reviews to learn how to distinguish between genuine and fake reviews by assigning higher weights to important words and phrases in the text. BERT's deep contextualized word representations are also useful for feature extraction in fake review detection, as they capture the meaning of a word in its context, allowing BERT to identify the most relevant parts of the text and to predict randomly hidden words for classification (Vidanagama et al. 2021). This is especially helpful in detecting fake reviews, which often contain misleading or irrelevant information.

3.3.2 GPT

GPT is a pre-training technique for deep neural networks that generate high-quality language representations using unsupervised learning on large-scale datasets (Alsubari et al. 2023). It has a decoder-only Transformer model that generates text sequences and a powerful attention mechanism that can identify important words and phrases in the text and assign higher weights to them. This attention mechanism is particularly useful for detecting fake online reviews by identifying subtle differences between genuine and fake reviews that may be difficult for humans to spot (Ahmed et al. 2018). GPT generates a language representation for the review and then feeds it into a classification model, typically a deep neural network fine-tuned on labeled reviews, to distinguish between genuine and fake reviews. The attention mechanism in GPT can identify specific details about the product or service in a genuine review, which are not present in a fake review, and assign higher weights to them, making them more influential in the classification decision.

3.3.3 DCWR

The deep contextualized word representation (DCWR) approach is a deep learning-based approach used to compute deep features for natural language processing tasks. This approach was first introduced by Peters et al. in their 2018 paper "Deep contextualized word representations" (Asghar et al. 2019). The DCWR approach uses a deep bidirectional language model (biLM) to generate context-sensitive word representations. The biLM is trained on a large corpus of text data to learn a mapping from words to their corresponding contextualized embeddings. The biLM considers the entire sentence context, including both the left and right sides of a word, to generate its representation. This allows the model to capture complex linguistic phenomena such as polysemy, synonymy, and word sense disambiguation (Deshai and Bhaskara Rao 2023). The DCWR approach has been shown to outperform traditional word embedding techniques such as Word2Vec and GloVe in many of these tasks. It has also been incorporated into popular deep learning frameworks such as PyTorch and TensorFlow, making it more accessible to researchers and practitioners.

3.3.4 Word2Vec

Word2Vec is a natural language processing technique that learns word representations and was introduced by Google researchers in 2013 (Deshai and Bhaskara Rao 2022). It has two models: continuous bag-of-words (CBOW) and skip-gram models, and is popular for its ability to learn high-quality word embeddings that can capture semantic and

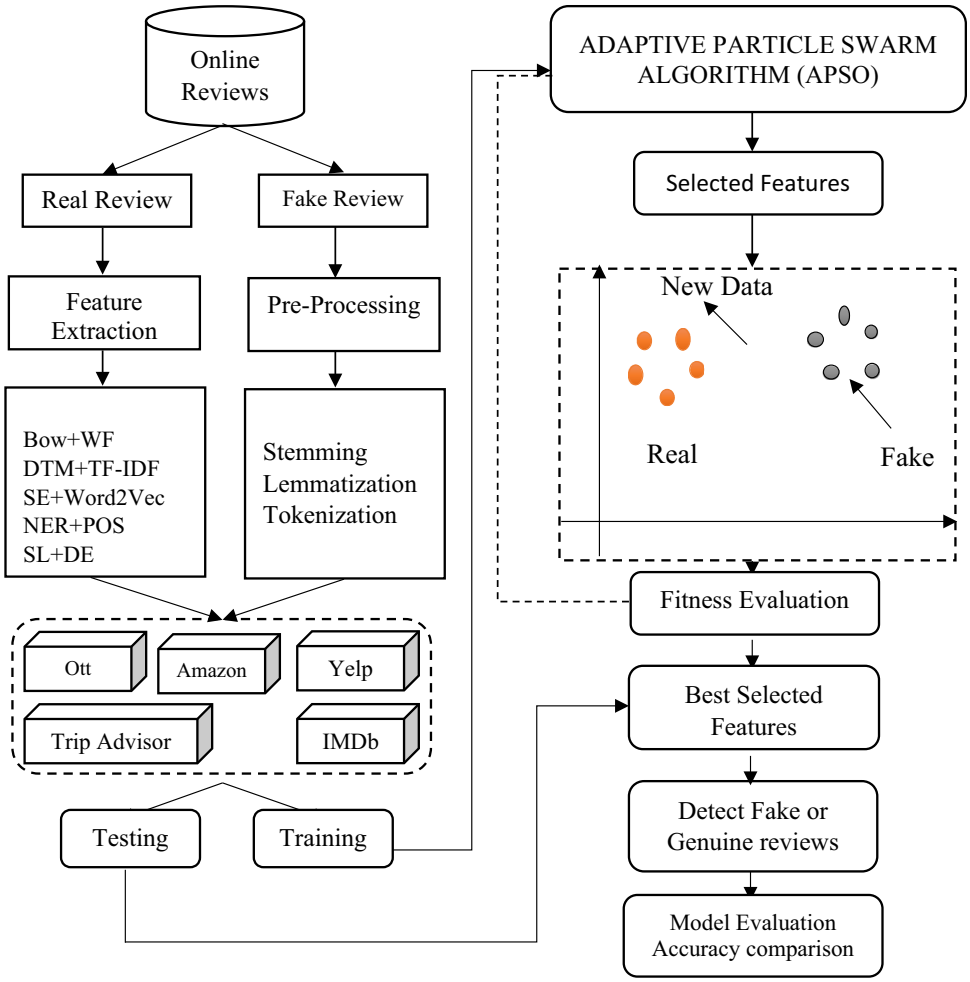

Fig. 1 Proposed Architecture

syntactic relationships between words (Barbado et al. 2019). The embeddings have numerous applications in NLP, and although variations like GloVe and FastText exist, Word2Vec remains popular and widely used.

3.3.5 GloVe

GloVe (Global Vectors for Word Representation) is a NLP technique that learns word embeddings, vector representations of words that capture their relationships (Barushka and Hajek 2018a). Introduced in 2014 by Stanford researchers, GloVe improves on previous techniques like Word2Vec by incorporating global word co-occurrence statistics into the learning process. It uses a weighted least-squares regression model to learn embeddings that maximize word co-occurrences in a corpus. GloVe embeddings have desirable properties, including the ability to capture semantic relationships and perform well on tasks like word similarity and analogy detection (Barushka and Hajek 2018b). GloVe is used in NLP applications and to create pre-trained embeddings. It is a popular technique due to its robustness and effectiveness.

3.3.6 FastText

FastText is a natural language processing library developed by Facebook AI Research in 2016, based on the Word2Vec model. It introduces a character-level n-gram embedding method that allows the model to handle out-of-vocabulary words and captures subword information (Barushka and Hajek 2019a). FastText also introduces a hierarchical softmax and subword information in the training process, resulting in faster training and better performance on rare words. It has several applications in NLP, particularly for handling tasks involving morphologically rich languages. FastText's embeddings have been shown to outperform those of Word2Vec and GloVe on several benchmarks. It is easy to use, fast, effective, and open source, with pre-trained models and APIs for quick integration into various NLP applications (Barushka and Hajek 2019b). Overall, FastText is a powerful NLP tool that has improved upon previous models and shown promising results in various NLP applications.

The table describes the datasets used in a research study on detecting fake online reviews. The datasets include Ott,

Yelp, Amazon, TripAdvisor, and IMDb. For each dataset, different tokenization techniques (Word2Vec, GloVe, FastText, DCWR) and stemming techniques (Snowball, Lovins, Dawson, Porter, Lancaster, WordNet) were used to preprocess the data, resulting in a total of 1230–1247 features per dataset. These preprocessed datasets were used to train and test the models for detecting fake online reviews on different online review platforms.

4 Models development

4.1 Feature Selection

Feature selection (FS) reduces dataset dimensionality by eliminating noisy and irrelevant features, improving learning performance by reducing overfitting and decreasing time and hardware resources (Krishna 2019). FS involves two processes: searching and evaluation. Searching finds the optimal feature subset using search algorithms with different time complexities. Evaluation can be done based on dataset characteristics or using a learning algorithm. Filters are fast but less accurate, while wrappers generate more accurate results using a learning algorithm. The statement discusses how metaheuristic algorithms can enhance the feature selection process by incorporating various methods such as new operators, encoding schemes, fitness functions, multi-objective optimization, and parallel algorithms.

4.2 Nature-inspired algorithms

Nature-inspired algorithms (NIAs) are computational methodologies that optimize complex problems by taking inspiration from natural processes (BrightLocal 2018; Chandy and Gu 2012). NIAs are categorized into evolutionary algorithms (EAs) and swarm-based algorithms (SI). EAs simulate biological evolution, while SIs model the behavior of social swarms. Examples of EAs include genetic algorithms, differential evolution, and the biogeography-based optimization algorithm (Chen et al. 2017). The detection process involves using the trained model to predict the output of the test dataset. Our model was trained using the CNN, LSTM, RNN, ANN, and adaptive particle swarm optimization (PSO) algorithms. To determine the most effective algorithm for accurately detecting and classifying fake reviews, we compared the performance of these algorithms using a comparison table. This enabled us to identify the algorithm that outperformed the others for the selected process.

4.3 Adaptive particle swarm optimization

Moreover, we are adding the Adaptive Particle Swarm Optimization and recursive feature elimination techniques to increase the detection accuracy rate of the fake profiles (Elmurngi and Gherbi 2017). PSO is an SI algorithm that uses particles to optimize problems. The position and velocity of particles are updated based on pbest and gbest in each iteration. To use PSO for feature selection, an Adaptive version is required. Transfer functions can be used to convert continuous variables to Adaptive variables, and two examples are given: a transfer function that defines the probability of updating a component of a solution and the sigmoid function used to convert velocity values to probability values in the range [0, 1] (Garcia 2018).

$$\mathrm{T}(\mathrm{vdi}(\mathrm{t})) = 1/(1 + \mathrm{e} - \mathrm{vdi}(\mathrm{t})) \tag{4}$$

In Adaptive optimization, Eq. (2) displays the S-shaped transfer functions utilized. Here, Xdi(t + 1) denotes the ith component in the X solution for dimension d in the (t + 1)th iteration. The function employs a random probability distribution (rand) to update the position vector's components based on each of their defined probabilities. The V-shaped transfer function is used in the next iteration to update the component based on the probability values acquired from Eq. (3). Researchers applied this equation to convert GSA into an Adaptive version.

$$X_i^d(t+1) = \begin{cases} 0, & \text{if } rand < S\_TF(v_i^d(t+1)) \\ 1, & \text{if } rand \geq S\_TF(v_i^d(t+1)) \end{cases} \tag{5}$$

$$T(X_i^d(t)) = |\tanh(X_i^d(t))| \tag{6}$$

$$X_{t+1} = \begin{cases} X_t, & \text{if } rand < V\_TF(\Delta X_{t+1}) \\ -X_t, & \text{if } rand \geq V\_TF(\Delta X_{t+1}) \end{cases} \tag{7}$$

To evaluate the effectiveness of the feature selection process, one major wrapper feature selection algorithm was utilized in this research, namely Adaptive Particle Swarm Optimization (APSO).

The adaptive particle swarm optimization (APSO) technique is more efficient in search than traditional particle swarm optimization techniques. This is due to its ability to converge more quickly and conduct a global search throughout the entire search space. APSO operates in two phases. The first phase involves real-time estimation of the evolutionary state by analyzing population distribution and particle fitness to identify one of the four designated evolutionary stages: exploration, exploitation, convergence, and leaping out (Ghai et al. 2019). This approach enables real-time adaptive tuning of search efficiency and convergence rate by adjusting algorithmic parameters such as inertia weight and acceleration coefficients.

The proposed algorithm segments the entire search space to enhance its search capability and strengthens the swarms' ability to share information. Each swarm expresses interest in or gathers information from other swarms that are fitter than it is, thereby improving the overall performance of the algorithm.

The following steps are involved in the algorithm:

1. Set up the PSO parameters, including the maximum number of iterations, population size, and initial particle positions and velocities
2. Define an objective function that calculates the fitness of each particle based on its feature subset
3. Evaluate the fitness of each particle using the objective function
4. Identify the particle with the best fitness as the global best particle
5. Begin the main PSO loop
6. Update the velocity and position of each particle using the PSO update equations
7. Evaluate the fitness of each particle based on its new position and update its personal best and the global best particle if necessary
8. Calculate the swarm diversity using a diversity measure
9. If the diversity is below a certain threshold, increase the exploration probability and decrease the exploitation probability to encourage exploration of the search space
10. Update the PSO parameters based on the exploration and exploitation probabilities
11. Continue the loop until the maximum number of iterations is reached or a stopping criterion is met
12. Select the best feature subset found by the PSO algorithm based on the global best particle's feature subset with the highest fitness

The velocity update equation of the algorithm is as follows:

Vi = Vi*wi + 1 / rank(i)rand()(pbest[ì] – Xi) + AdaptivePSO(i);

Xi = Xi + Vi;

Where AdaptivePSO(i) is defined as follows:

```
AdaptivePSO(i) {
Posx ← 0.0
For each individual k of the population
if pFitness[k] is better than fitness[i]
posx ← posx + 1 / rank(k) * rand()*(pbest[k]-Xi)
if(posx > Vmax)
return Vmax
else
return posx
}
```

Recursive feature elimination (RFE) is a feature selection technique that iteratively eliminates the features with the lowest predictive value, based on the ranking of features obtained from a model's attributes (Hajek 2018). RFE helps to eliminate dependencies and collinearity in the model. Cross-validation is used to score several feature subsets and determine the optimal number of features.

5 Experimental analysis

The following section presents an experiment that evaluated the performance of various machine learning, deep learning, and transformer models. Tables and graphs are provided to facilitate a comparison of the models' outcomes.

5.1 Experimental setup

The experiments are implemented using Python 3.6.9 on Google Colab, leveraging the powerful computing capabilities of GPU. The data preparation and tokenization are performed using Numpy 1.18.5 and Hugging face 3.5.1 libraries, while the pre-trained transformers are implemented using Hugging face 3.5.1. For implementing Machine learning models, Scikit-learn 0.23.2 is used. The deep learning models are created using either Pytorch 1.7.0 or Tensorflow 2.3.0. To visualize the experimental results, Matplotlib 3.2.2 is utilized.

5.2 Classification

As we delve into the world of machine learning and deep learning, we encounter a powerful tool that allows us to categorize and label the unknown with incredible accuracy. This tool, known as classification, is a true game-changer in the field of data analysis (https://doi.org/10.1145/1014052.1014073). In the research at hand, we seek to uncover the true nature of online profiles—are they genuine, or are they fake. To achieve this, we have enlisted the help of some of the most powerful classification techniques available to us today. Support Vector Machine, K-Nearest Neighbor, Random Forest, Logistic Regression, and Extra Tree are all lending their expertise to this critical task. With a value of 1 representing a fraudulent profile and a value of 0 indicating authenticity, we aim to sift through the sea of online profiles with unparalleled accuracy and efficiency.

5.3 Convolution Neural Network (CNN)

To evaluate the selected features, a CNN was chosen as the main classifier. The output features were represented as an adaptive vector of length n, where each element in the vector represented a feature in the dataset. If a feature was selected, the corresponding element was set to 1, otherwise, it was set to 0 (Hussain et al. 2019). The performance of the feature subset was evaluated based on the accuracy of the classification model and the number of selected features.

$$\textbf{Fitness} = \alpha \times (1 - \textbf{accuracy}) + \left(1 - \alpha * \frac{|S|}{|W|}\right) \quad (8)$$

Sigmoid: The sigmoid function is a type of activation function commonly used in neural networks. It takes in real numbers as input and restricts the output to a range between zero and one, making it useful for binary classification tasks. The sigmoid function is characterized by an S-shaped curve and can be represented mathematically by the Eq. (8)

$$\mathrm{y} = \frac{1}{(1 + \mathrm{e}\text{^} - \mathrm{x})} \quad (9)$$

The Tanh function is another popular activation function used in neural networks. Similar to the sigmoid function, it takes in real numbers as input, but its output is restricted to the range of − 1 to 1. The mathematical representation of the Tanh function can be shown using Eq. 3.

$$\mathrm{f(x)tanh} = 1\frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (10)$$

ReLU (Rectified Linear Unit) is the most widely used activation function in the context of convolutional neural networks. It works by converting all negative values in the input to zero, while keeping positive values unchanged (Jain et al. 2018). This simple and efficient function has become popular due to its computational advantages, as it reduces the load on the network and speeds up the training process. Its mathematical representation is in Eq. 4.

$$f(x)ReLU = max(0, x) \quad (11)$$

The loss function used in CNNs employs two parameters: the predicted output (also known as the CNN estimated output) and the actual output (referred to as the label), and is used to calculate the error.

The Softmax Loss Function, also known as the Cross-Entropy or log loss function, is frequently used to evaluate the performance of CNN models. It is used in multi-class classification problems instead of the square error loss function and produces a probability value ranging from 0 to 1. The function uses the softmax activation in the output layer to produce the output in the form of a probability distribution.

$$pi = \frac{e^{a_i}}{\sum_{k=1}^{N} e_k^a} \quad (12)$$

In the fitness equation, α is a weight parameter that controls the balance between classification accuracy and the number of selected features. The term (1-accuracy) represents the classification error, while the term (|S|/|W|) represents the ratio of selected features to the total number of features in the dataset. The fitness value is a measure of the quality of the feature subset, with higher fitness values indicating better feature subsets that have higher classification accuracy and fewer selected features (Kennedy et al. 2019). The wrapper feature selection algorithms search for the feature subset with the highest fitness value, and this feature subset is then used for classification with the CNN classifier. The pooling layer is an essential component in deep learning-based text classification for detecting fake online reviews. Its primary purpose is to sub-sample the feature maps generated by convolutional operations, shrinking large-size maps to create smaller feature maps while retaining the dominant information or features (Li et al. 2017a). The pooling layer is characterized by both the stride and kernel size. The fivefold cross-validation technique involves dividing the dataset into 5 equal-sized folds, using one fold as the testing set and the remaining 4 folds as the training set in each iteration. This process is repeated 5 times, with each fold used once as the testing set. The results are averaged to estimate the classifier's generalization ability and reduce the risk of overfitting.

5.4 Recurrent neural networks (RNNs)

RNN is a deep learning algorithm used for processing sequential data, particularly useful in natural language processing tasks (Li et al. 2017b). RNNs can remember previous inputs to inform current predictions. For detecting fake online reviews on social media, RNNs can be trained on a dataset of labeled genuine and fake reviews to predict the authenticity of new reviews. LSTMs, a popular type of RNN for natural language processing, address the vanishing gradients problem in RNNs by including specialized memory cells to retain information over longer periods.

Table 3 The Hybrid Pre-processing Techniques

Datasets	Hybrid Pre-processing Techniques			No of features
	Tokenization	Stemming	Lemmatization	
Ott	1.Transformer-basedTokenization.	1. Hybrid Stemming with Sentiment Analysis.	1. Machine Learning-based Lemmatization.	1230
Yelp	2.Byte Pair Encoding (BPE). Contextual Word Embeddings.	2. Hybrid Stemming with Named Entity Recognition (NER).	2. Hybrid Lemmatization with Word Embeddings.	1238
Amazon	3.Subword Tokenization.	3. Hybrid Stemming with Contextual Information.	3. Hybrid Lemmatization with Named Entity Recognition (NER).	1234
Trip Advisor				1247
IMDb				1225

RNNs' ability to capture the sequential nature of text data and learn from patterns in large datasets makes them a promising approach for detecting fake online reviews.

5.5 Long short-term memory (LSTM)

LSTM is a neural network used for natural language processing and sequence modeling tasks (Liu et al. 2019). In detecting fake online reviews on social media, LSTM models are effective in capturing long-term dependencies in sequential data, identifying the relationships between words, and distinguishing real from fake reviews (Jain et al. 2019). The model is trained on a dataset of labeled reviews and predicts the label based on the sequence of words seen so far. The trained LSTM model can be used to classify new reviews as real or fake and identify specific features or patterns in the text indicative of fake reviews.

5.6 Artificial Neural Networks (ANNs)

Artificial Neural Networks (ANNs) can be used for detecting fake online reviews on social media platforms through supervised learning on a dataset of genuine and fraudulent reviews (Madisetty and Desarkar 2018). ANNs consist of interconnected nodes that process and transmit information, allowing them to identify patterns and features indicative of fake reviews that may be difficult for humans or traditional machine learning algorithms to detect. ANNs can adapt and improve over time, making them a powerful tool for ensuring the integrity of online review systems.

6 Evaluation and assessment

Our research utilized four evaluation measures: accuracy, precision, recall, and g-mean. To calculate these measures, we constructed a confusion matrix in which true positives (TPs) represented fake news correctly predicted as fake, true negatives (TNs) represented non-fake news correctly predicted as non-fake, false positives (FPs) represented non-fake news incorrectly predicted as fake, and false negatives (FNs) represented fake news incorrectly predicted as non-fake (Malik and Hussain 2017). Accuracy was defined as the ratio of correctly classified instances, both fake and non-fake, overall, correctly, and incorrectly classified instances.

Equation (6) represents the calculation for accuracy, which is a measure of the overall performance of the classification model (Pandey and Rajpoot 2019). Accuracy is calculated as the ratio of the correctly classified fake and non-fake news instances over all the classified instances, both correct and incorrect. The equation for accuracy is:

$$\textbf{Accuracy} = \frac{\textbf{TP} + \textbf{TNTP}}{\textbf{TP} + \textbf{TN} + \textbf{FP} + \textbf{FN}} \tag{13}$$

where TP represents true positives, which are the number of fake reviews instances correctly predicted as fake; TN represents true negatives, which are the number of non-fake review instances correctly predicted as non-fake (Patel and Patel 2018). FP represents false positives, which are the number of non-fake reviews instances predicted as fake; and FN represents false negatives, which are the number of fake news instances predicted as non-fake.

Precision, which is the ratio of news correctly identified as fake over all fake (positive) news instances, it is represented as in Eq. (7).

$$\textbf{Precision} = \frac{\textbf{TP}}{\textbf{FP} + \textbf{TP}} \tag{14}$$

precision is a measure of how many of the news articles identified as fake reviews are fake, out of all the articles predicted to be fake. The precision score ranges from 0 to 1, where a score of 1 means that all articles identified as fake reviews are fake, and a score of 0 means that none of the articles identified as fake are actually fake (Ren and Ji 2017).

It measures the sensitivity of the model or how well the model can identify the positive examples (fake) of reviews. The recall is calculated as the ratio of true positive (TP) reviews correctly identified as fake overall actual positive (fake) news instances, as shown in Eq. (8):

$$\text{Recall} = \frac{\textbf{TP}}{\textbf{FN} + \textbf{TP}} \tag{15}$$

The F-measure is a weighted average of the precision and recall measures, and it is commonly used to evaluate classification models. It can be calculated using Eq. (9), where Precision is the ratio of true positives to the sum of true positives and false positives, and Recall is the ratio of true positives to the sum of true positives and false negatives (Rout et al. 2017b).

$$\text{F-Measure} = \frac{(2 \times \textbf{Precision} \times \textbf{Recall})}{(\textbf{Precision} + \textbf{Recall})} \tag{16}$$

The F-measure provides a more balanced measure of model performance than accuracy alone, as it considers both false positives and false negatives.

7 Results and discussion

The section reports on three experimental phases that were conducted on five different datasets, namely Ott, Yelp, Amazon, Trip Advisor, and IMDb. The first phase involved describing the performance of four versions of the CNN classification model, including CNN-APSO and the

Table 4 Parameters of PSO used with the PSO-based feature weighting

Parameter	Value
No. of practices	20
Maximum iterations	50
Local best weight	2
Global best weight	2
Intertia weight	1
Stop condition	Max no. of iteration

standard CNN, with various metaheuristic algorithms used for feature selection. In the second phase, the best model was chosen and compared with other classification algorithms. The five-fold cross-validation criteria were used for splitting the data, and the metaheuristic algorithms' settings are presented in Table 2. The best parameters for the proposed APSO-based feature weighting were selected through a trial-and-error process and are listed in Table 4.

Table 4 and Fig. 2 show the parameters used for PSO (Particle Swarm Optimization) in the PSO-based feature weighting approach to detect fake online reviews on platforms such as Ott, Amazon, Yelp, TripAdvisor, and IMDb. PSO is a computational method that optimizes a problem by iteratively trying to improve a candidate solution (Rout et al. 2018). In this approach, PSO is used to assign weights to the different features that are used to classify reviews as genuine or fake. The table provides information on the specific parameters used for this process, which are as follows:

No. of practices: This refers to the number of iterations or cycles that the PSO algorithm run to optimize the feature weights. In this case, the algorithm was run for 20 practices.

Maximum iterations: This parameter sets the maximum number of iterations that the PSO algorithm will perform in each practice. In this case, the maximum iterations were set to 50.

Local best weight: This parameter determines the influence of the best solution found by each particle in the swarm (i.e., a group of particles that work together to optimize the solution). In this approach, the local best weight was set to 2.

Global best weight: This parameter determines the influence of the best solution found by the entire swarm. In this approach, the global best weight was also set to 2.

Inertia weight: This parameter determines the trade-off between the particle's current velocity and its historical velocity. A higher inertia weight places more emphasis on the particle's historical velocity. In this approach, the inertia weight was set to 1.

Stop condition: This parameter sets the stopping criteria for the PSO algorithm. In this approach, the stopping condition was set to the maximum number of iterations (50).

By setting these parameters, the PSO algorithm can iteratively optimize the weights assigned to the different features used to classify reviews, to improve the accuracy of the classification process s (Tang et al. 2019). The results of this optimization process can be used to detect fake online reviews on platforms such as Ott, Amazon, Yelp, TripAdvisor, and IMDb.

The Table 5 shows the performance of traditional classifiers on different datasets to detect fake online reviews using various performance metrics such as accuracy, precision, recall, and F score. The classifiers used include

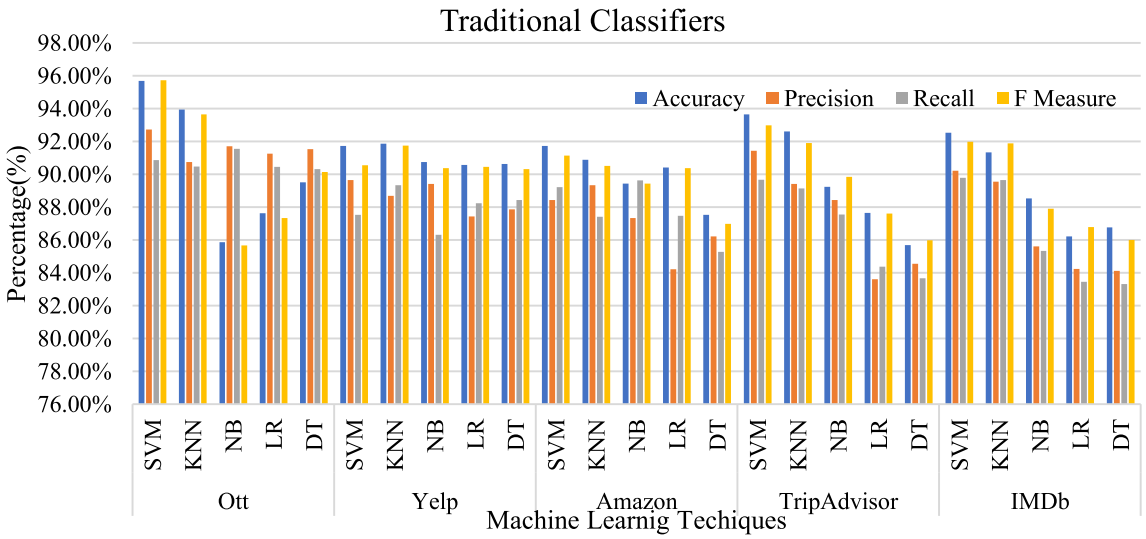

Fig. 2 Comparing performance of traditional classifiers on online review dataset using APSO optimization

Table 5 Performance of traditional classifiers on online review dataset with APSO

Dataset	Cross validation	Classifier	Accuracy (%)	Precision (%)	Recall (%)	F score (%)
Ott	Tenfold	SVM	95.68	92.73	90.87	95.72
	Fivefold	KNN	93.95	90.74	90.47	93.65
	Tenfold	NB	85.87	91.71	91.56	85.66
	Tenfold	LR	87.64	91.25	90.45	87.34
	Tenfold	DT	89.52	91.54	90.32	90.15
Yelp	Tenfold	SVM	91.73	89.65	87.54	90.56
	Fivefold	KNN	91.86	88.68	89.34	91.75
	Tenfold	NB	90.75	89.42	86.32	90.37
	Tenfold	LR	90.58	87.43	88.23	90.46
	Tenfold	DT	90.63	87.87	88.43	90.32
Amazon	Tenfold	SVM	91.73	88.44	89.22	91.15
	Fivefold	KNN	90.89	89.33	87.41	90.52
	Tenfold	NB	89.43	87.34	89.63	89.44
	Tenfold	LR	90.41	84.22	87.47	90.37
	Tenfold	DT	87.53	86.21	85.28	86.98
TripAdvisor	Tenfold	SVM	93.65	91.43	89.67	92.99
	Fivefold	KNN	92.61	89.42	89.14	91.90
	Tenfold	NB	89.23	88.43	87.56	89.84
	Tenfold	LR	87.65	83.62	84.38	87.61
	Tenfold	DT	85.68	84.55	83.67	85.98
IMDb	Tenfold	SVM	92.53	90.21	89.78	91.99
	Fivefold	KNN	91.34	89.56	89.65	91.88
	Tenfold	NB	88.54	85.62	85.34	87.91
	Tenfold	LR	86.22	84.24	83.45	86.79
	Tenfold	DT	86.76	84.13	83.31	86.01

Table 6 Performance comparison of LSTM Models using APSO optimization

Dataset	Train test ratio	Embedding and hidden dimension	Accuracy (%)	Precision (%)	Recall (%)	F score (%)
Ott	90:10	100, 200	95.66	96.54	96.45	95.60
	80:20	100, 200	95.33	96.65	95.23	95.44
	70:30	100, 50	95.56	95.34	96.12	95.52
Yelp	90:10	100, 200	95.55	96.32	96.24	95.21
	80:20	100, 200	95.12	96.45	96.56	96.26
	70:30	100, 50	95.01	96.77	96.44	95.78
Amazon	90:10	100, 200	95.66	95.66	95.66	95.66
	80:20	100, 200	96.44	95.30	95.67	96.25
	70:30	100, 50	96.56	95.11	95.15	96.78
TripAdvisor	90:10	100, 200	96.57	96.77	96.55	96.47
	80:20	100, 200	95.65	96.25	96.76	95.01
	70:30	100, 50	95.89	96.74	96.81	95.83
IMDb	90:10	100, 200	96.41	96.52	96.62	96.21
	80:20	100, 200	96.35	96.24	96.54	96.11
	70:30	100, 50	96.27	96.23	96.68	96.22

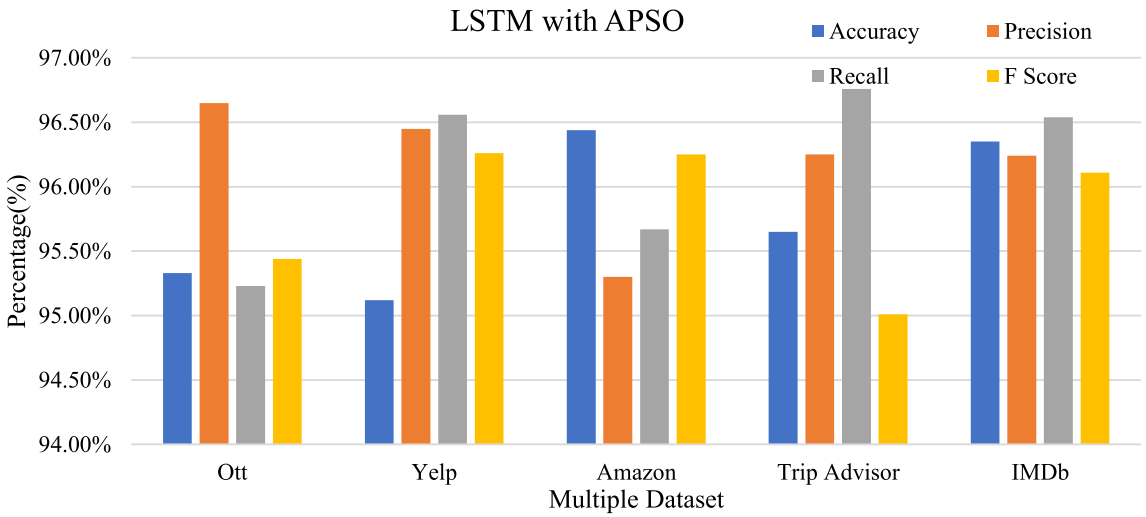

Fig. 3 Performance comparison of LSTM models using APSO optimization

Table 7 Result of RNN with APSO

Dataset	Train test ratio	Dimension	Accuracy (%)	Precision (%)	Recall (%)	F score (%)
Ott	90:10	50	97.78	98.49	98.45	97.60
	80:20	200	97.33	98.65	97.23	97.44
	70:30	200	97.56	97.34	98.12	97.52
Yelp	90:10	50	97.55	98.32	98.24	97.21
	80:20	200	97.23	98.45	98.56	98.26
	70:30	200	97.11	98.77	98.44	97.78
Amazon	90:10	50	97.78	97.78	97.78	97.78
	80:20	200	98.44	97.30	97.67	98.25
	70:30	200	98.56	97.11	97.15	98.78
TripAdvisor	90:10	50	98.57	98.77	98.55	98.47
	80:20	200	97.65	98.25	98.76	97.01
	70:30	200	97.89	98.74	98.81	97.83
IMDb	90:10	50	98.41	98.52	98.62	98.21
	80:20	200	98.35	98.24	98.49	98.11
	70:30	200	98.27	98.23	98.68	98.22

SVM, KNN, NB, LR, and DT. In the case of the "Ott" dataset, the SVM classifier had the highest accuracy of 95.68%, while NB had the lowest accuracy of 85.87%. For the "Yelp" dataset, the SVM classifier had the highest accuracy of 91.73%, and NB had the lowest accuracy of 90.75%. For the "Amazon" dataset, the SVM classifier had the highest accuracy of 91.73%, and DT had the lowest accuracy of 87.53%. For the "TripAdvisor" dataset, the SVM classifier had the highest accuracy of 93.65%, and DT had the lowest accuracy of 85.68%. Finally, for the "IMDb" dataset, the SVM classifier had the highest accuracy of 92.53%, and DT had the lowest accuracy of 86.76%. Overall, the SVM classifier performed the best on most datasets, while the DT classifier had the lowest performance.

Table 6 and Fig. 3 shows the results of using Long Short-Term Memory (LSTM) with Adaptive Particle Swarm Optimization (APSO) to detect fake online reviews on different datasets, including Ott, Yelp, Amazon, TripAdvisor, and IMDb. The table presents the accuracy, precision, recall, and F1 score for each dataset, calculated for three different experiments. The accuracy represents the proportion of correctly classified reviews, while precision indicates the proportion of true positives (correctly classified fake reviews) to the total number of reviews classified as fake. Recall represents the proportion of true positives to the total number of actual fake reviews, and the F1 score is a weighted average of precision and recall. Overall, the results show that LSTM with APSO performs well in detecting fake online reviews on all datasets, with accuracy ranging from 95.01 to 96.57%. The precision, recall, and

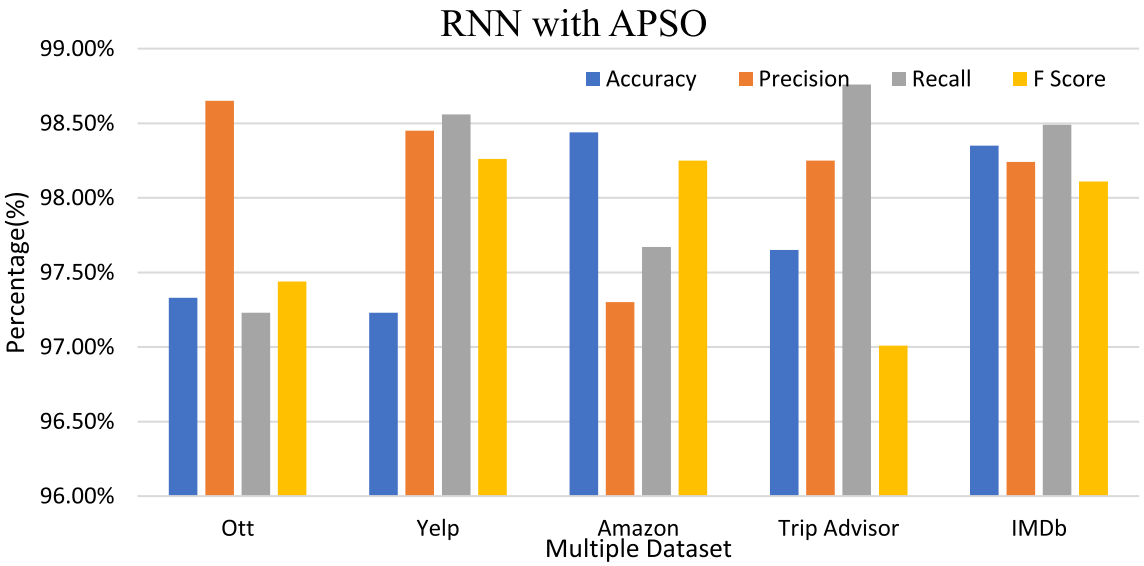

Fig. 4 Performance comparison of RNN using APSO on an online review dataset

Table 8 Result of ANN with APSO

Dataset	Train test ratio	Dimension	Accuracy (%)	Precision (%)	Recall (%)	F score (%)
Ott	90:10	100	96.88	95.24	97.37	96.23
	80:20	100	96.42	95.54	96.68	96.52
	70:30	100	96.65	96.72	97.24	96.67
Yelp	90:10	100	96.24	95.41	97.56	96.62
	80:20	100	96.25	95.53	97.76	97.32
	70:30	100	96.32	95.64	97.85	96.81
Amazon	90:10	100	96.78	96.71	96.64	96.73
	80:20	100	97.56	96.44	96.86	97.34
	70:30	100	97.89	96.25	96.73	97.89
TripAdvisor	90:10	100	97.41	95.68	97.69	97.52
	80:20	100	96.88	95.31	97.39	96.32
	70:30	100	96.91	95.74	97.64	96.76
IMDb	90:10	100	97.63	95.88	97.75	97.34
	80:20	100	97.84	95.36	97.63	97.26
	70:30	100	97.36	95.28	97.79	97.34

F1 score also indicate high performance on all datasets, with values ranging from 95.11 to 96.77%.

Table 7 and Fig. 4 shows the performance of a Recurrent Neural Network (RNN) model with APSO (Adaptive Particle Swarm Optimization) in detecting fake online reviews on different datasets, including Ott, Yelp, Amazon, TripAdvisor, and IMDb. For Ott, the RNN with APSO model achieved an average accuracy of 97.56%, with Precision, Recall, and F Score ranging from 97.34 to 98.65%. For Yelp, the model achieved an average accuracy of 97.3%, with Precision, Recall, and F Score ranging from 98.32 to 98.77%. For Amazon, the model achieved an average accuracy of 97.78%, with Precision, Recall, and F Score all at 97.78%. For TripAdvisor, the model achieved an average accuracy of 98.37%, with Precision, Recall, and F Score ranging from 98.25 to 98.77%. Finally, for IMDb, the model achieved an average accuracy of 98.34%, with Precision, Recall, and F Score ranging from 98.23 to 98.68%. The high accuracy, precision, recall, and F score values indicate that the model can effectively distinguish between real and fake reviews, which can help in improving the overall quality and reliability of online review platforms.

Table 8 and Fig. 5 show the performance of the Artificial Neural Network (ANN) with Adaptive Particle Swarm Optimization (APSO) in detecting fake online reviews on different platforms including Ott, Yelp, Amazon, TripAdvisor, and IMDb. For Ott, the ANN with APSO

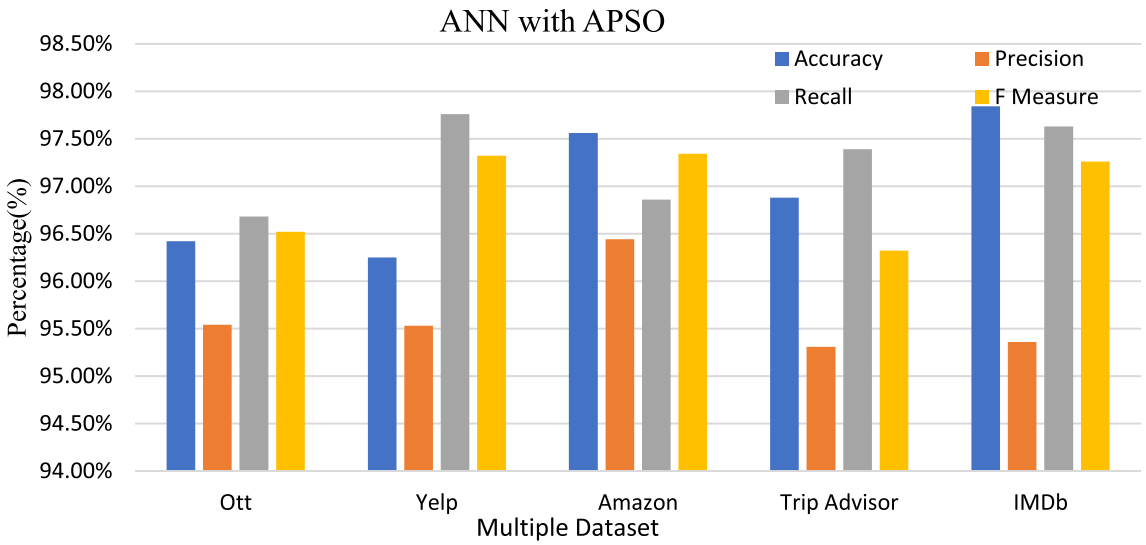

Fig. 5 Performance analysis of ANN on an online review dataset using APSO

Table 9 Result of CNN with APSO

Dataset	Train test ratio	Embedding and hidden dimension	Accuracy (%)	Precision (%)	Recall (%)	F score (%)
Ott	90:10	200,200	98.88	98.32	97.12	98.91
	80:20	50,100	99.00	98.53	98.01	99.02
	70:30	50,100	98.33	97.26	97.52	98.93
Yelp	90:10	200,200	98.33	97.32	96.24	98.21
	80:20	50,100	98.12	96.45	97.56	97.26
	70:30	50,100	98.01	97.77	97.44	98.78
Amazon	90:10	200,200	98.66	98.66	98.66	98.66
	80:20	50,100	99.44	98.30	98.67	99.25
	70:30	50,100	99.56	98.11	98.15	99.78
TripAdvisor	90:10	200,200	97.57	96.77	97.55	97.47
	80:20	50,100	98.65	97.25	97.76	98.01
	70:30	50,100	98.89	97.74	97.81	98.83
IMDb	90:10	200,200	97.41	96.52	96.62	97.21
	80:20	50,100	97.35	96.24	97.54	97.11
	70:30	50,100	97.27	96.23	96.68	97.22

achieved an accuracy of 96.88%, precision of 95.24%, recall of 97.37%, and F-score of 96.23%. Similar high performance was also observed for Yelp and TripAdvisor, with accuracy ranging from 96.24 to 97.41%, and F-score ranging from 96.32 to 97.52%. For Amazon and IMDb, the ANN with APSO achieved the highest performance, with accuracy ranging from 96.78 to 97.89%, precision ranging from 96.25 to 96.71%, recall ranging from 96.64 to 97.79%, and F-score ranging from 96.73 to 97.34%. The ANN with APSO can effectively detect fake online reviews on different platforms with high accuracy and precision.

Table 9 and Fig. 6 show the results of using convolutional neural networks (CNN) with APSO to detect fake online reviews in different datasets. For the Ott dataset, the CNN with APSO achieved an accuracy of 98.88%, with precision ranging from 97.26 to 98.53%, recall ranging from 97.12 to 98.01%, and F score ranging from 98.91 to 99.02%. For the Yelp dataset, the accuracy achieved was 98.33%, with precision ranging from 96.45 to 97.77%, recall ranging from 96.24 to 97.56%, and F score ranging from 97.26 to 98.78%. For the Amazon dataset, the CNN with APSO achieved a high accuracy of 98.66% and consistent precision, recall, and F score metrics across

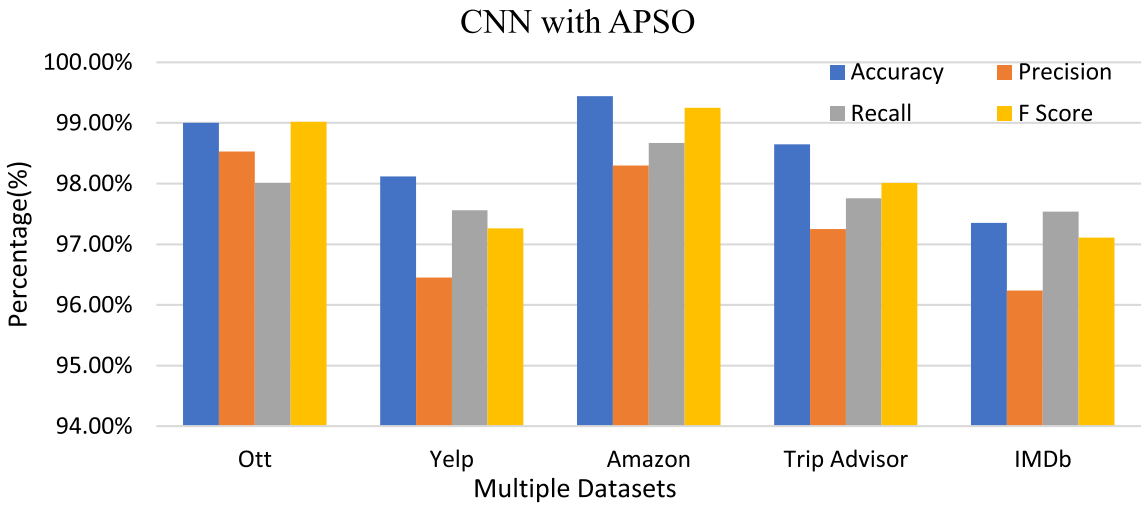

Fig. 6 Analysing the performance of CNN with APSO in detecting fake online reviews

different experiments. For the TripAdvisor dataset, the accuracy ranged from 97.57 to 98.89%, with precision ranging from 96.77 to 97.74%, recall ranging from 97.55 to 97.81%, and F score ranging from 97.11 to 98.83%. Finally, for the IMDb dataset, the CNN with APSO achieved an accuracy ranging from 97.27 to 97.41%, with precision ranging from 96.23 to 96.52%, recall ranging from 96.62 to 97.54%, and F score ranging from 97.11 to 97.22%. Finaaly, the CNN with APSO showed high performance in detecting fake online reviews in different datasets, achieving high accuracy and consistent precision, recall, and F score metrics.

To further highlight the superiority of CNN-APSO, the authors analyzed the convergence curves of all algorithms and datasets, which can be found in Fig. 4. The figure shows that CNN-APSO has a faster convergence rate compared to the other algorithms, confirming its better performance. It is important to note that the convergence curves were calculated using Eq. (7), which measures the fitness value of the best solution found by each algorithm at each iteration. These results support the authors' claim that CNN-APSO is a more effective method for feature selection and classification on text datasets, particularly on Yelp Dataset with the TF-IDF and Snowball, Lovins, Dawson, Porter, Lancaster, WordNet stemming representation techniques.

Due to its outstanding performance in the previous experiments, an additional examination and analysis were conducted on five datasets to further validate the quality of both the dataset and the proposed classification model. In this experiment, three other popular classification algorithms were applied to five datasets, namely Ott, Amazon,

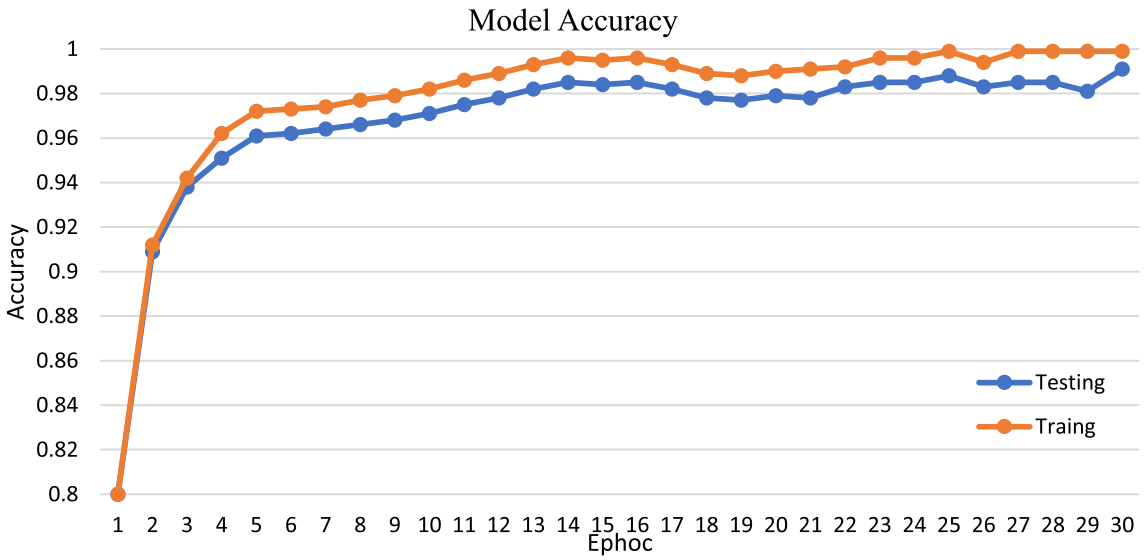

Fig. 7 Model accuracy during training and testing

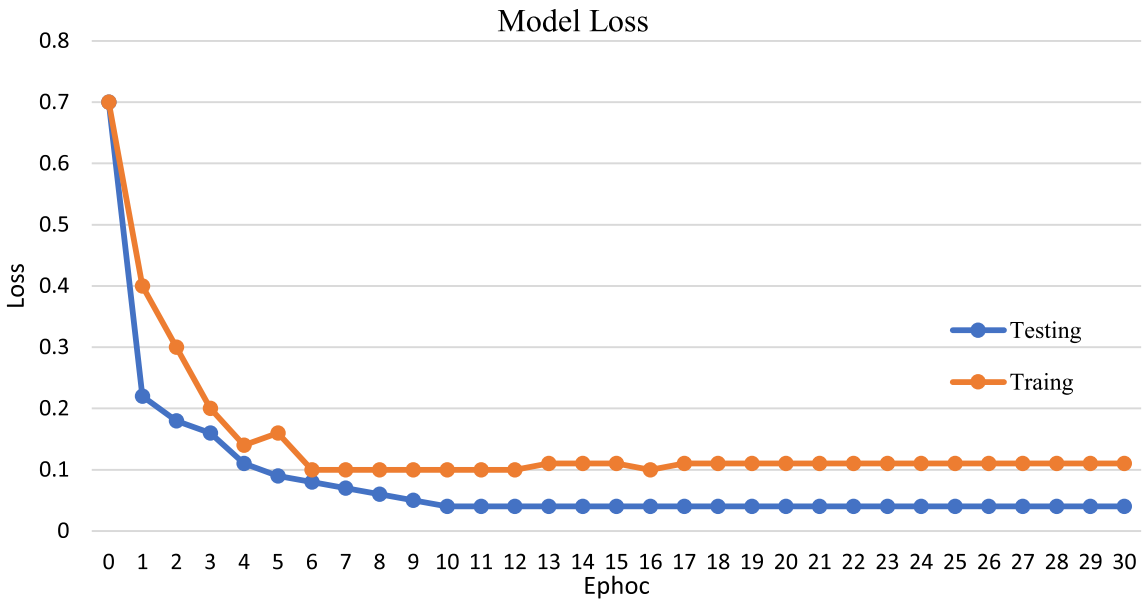

Fig. 8 Model loss during training and testing

Yelp, TripAdvisor, and IMDb. The results of the 4 classification models (LSTM, RNN, ANN, CNN) are presented in Table 9. CNN-APSO achieved the highest accuracy with 99.4%, followed by LSTM, RNN, ANN, SVM, KNN, NB, LR, DT and RF, respectively.

The proposed CNN model exhibited remarkable accuracy in identifying fake online reviews across all datasets during both training and testing phases, as shown in Fig. 7. The Ott and Amazon datasets had the highest accuracy rates at 99.44%, followed by IMDb at 97.35%. However, the testing accuracy was slightly lower than the training accuracy, implying that the model may have overfit to the training data. The model was trained for 30 epochs, and the detection of fake reviews was performed on five datasets, namely Ott, Amazon, Yelp, TripAdvisor, and IMDb. These findings suggest that the proposed deep learning-based text classification approach can be a promising solution for the detection of fake online reviews.

The CNN model was trained and tested on five datasets OTT, Yelp, Amazon, TripAdvisor, and IMDb) to detect fake online reviews. The model was trained for 30 epochs, with the training and testing being done with a split ratio of 80:20, as shown in Fig. 8. The model achieved high accuracy on all datasets during both the training and testing phases. The highest accuracy of 99.44% was achieved on the OTT dataset, while the lowest accuracy was 96.24% on the IMDb dataset. The accuracy scores for the other datasets were between 97.11% and 98.67%. Overall, the model achieved promising results in detecting fake reviews across different platforms, demonstrating the effectiveness of the proposed methodology that utilized deep learning-based text classification techniques.

8 Conclusions

In In conclusion, fake online reviews continue to be a major issue in the e-commerce industry, and our proposed deep learning-based text classification methodology offers a promising solution. We utilized advanced techniques such as BERT, GPT, DCWR, word2vec, GloVe, and fast Text for feature extraction in NLP tasks, along with a range of stemming algorithms. Our experiment demonstrated that the CNN-APSO model achieved the highest accuracy of 99.4% and outperformed other popular classification algorithms on social media platforms such as Ott, Amazon, Yelp, TripAdvisor, and IMDb. However, further research is required to refine the methodology and enhance its accuracy and efficiency in detecting fake reviews. Such techniques are critical to maintaining the authenticity of online reviews and empowering customers to make informed purchasing decisions. Our results also showed that CNN-APSO outperformed other classification models, including LSTM, RNN, ANN, SVM, KNN, NB, LR, DT and RF on the Ott, Amazon, Yelp, TripAdvisor, and IMDb dataset in terms of accuracy, precision, and F-measure.

Author contributions All authors are contributed equally.

Funding The authors have not disclosed any funding.

Data availability I taken the datasets which is available in public datasets (Amazon review data (2018) and kaggle websites

OTT Platforms	The OTT Platforms Review that support the findings of this study are available in ["OTT Platforms Review"] with the identifier(s) https://www.kaggle.com/code/anishwarammayappan/data-analysis-of-movies-on-ott-platforms
Data available in a public repository (Amazon review data (2018))	The Amazon review data that support the findings of this study are available in ["Amazon review data (2018)"] with the identifier(s) https://cseweb.ucsd.edu/~jmcauley/datasets/amazon_v2
Yelp Dataset	The Yelp review data that support the findings of this study are available in ["Yelp_academic_dataset_review "] with the identifier(s) https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset
Trip Advisor Hotel Reviews	The Trip Advisor review data that support the findings of this study are available in ["Trip Advisor Hotel Reviews "] with the identifier(s) https://www.kaggle.com/code/qusaybtoush1990/trip-advisor-hotel-reviews
IMDB Dataset of 50K Movie Reviews	The IMDB Dataset of 50K Movie Reviews that support the findings of this study are available in https://www.kaggle.com/datasets/lakshmi25npathi/imdb-dataset-of-50k-movie-reviews

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Ahmed H, Traore I, Saad S (2018) Detecting opinion spams and fake news using text classification. Secur Priv 1(1):e9. https://doi.org/10.1002/spy2.9

Alsubari SN, Deshmukh SN, Aldhyani THH, Al Nefaie AH, Alrasheedi M (2023) Rule-based classifiers for identifying fake reviews in e-commerce: a deep learning system. In: Som T (ed) Interdisciplinary mathematics. Springer, Singapore. https://doi.org/10.1007/978-981-19-8566-9_14

Arif MH, Li J, Iqbal M, Liu K (2018) Sentiment analysis and fake detection in short informal text using learning classier systems. Soft Comput 22(21):72817291

Asghar MZ, Ullah A, Ahmad S, Khan A (2019) Opinion spam detection framework using hybrid classification scheme. Soft Comput. https://doi.org/10.1007/s00500-019-04107-y

Barbado R, Araque O, Iglesias CA (2019) A framework for fake review detection in online consumer electronics retailers. Inf Process Manag 56(4):1234–1244. https://doi.org/10.1016/j.indmarman.2019.08.003

Barushka A, Hajek P (2018a) Spam filtering in social networks using regularized deep neural networks with ensemble learning. In: Iliadis L, Maglogiannis I, Plagianakos V (eds) Artificial intelligence applications and innovations. AIAI 2018, vol 519. IFIP advances in information and communication technology. Springer, Cham, pp 38–49. https://doi.org/10.1007/978-3-319-92007-8_4

Barushka A, Hajek P (2018b) Spam filtering using integrated distribution-based balancing approach and regularized deep neural networks. Appl Intell 48(10):3538–3556. https://doi.org/10.1007/s10489-018-1161-y

Barushka A, Hajek P (2019a) Spam detection on social networks using cost-sensitive feature selection and ensemble-based regularized deep neural networks. Neural Comput Appl. https://doi.org/10.1007/s00521-019-04331-5

Barushka A, Hajek P (2019b) Review spam detection using word embeddings and deep neural networks. In: MacIntyre J, Maglogiannis I, Iliadis L, Pimenidis E (eds) Artificial intelligence applications and innovations. AIAI 2019, vol 559. IFIP advances in information and communication technology. Springer, Cham, pp 340–350. https://doi.org/10.1007/978-3-030-19823-7_28

Birim ŞÖ, Kazancoglu I, Mangla SK, Kahraman A, Kumar S, Kazancoglu Y (2022) Detecting fake reviews through topic modelling. J Bus Res 149:884–900. https://doi.org/10.1016/j.jbusres.2022.05.081

Brar GS, Sharma A (2018) Sentiment analysis of IMDb review using supervised machine learning techniques. Int J Appl Eng Res 13(16):1278812791

BrightLocal (2018) Local consumer review survey 2018. https://www.brightlocal.com/research/local-consumer-review-survey/. Accessed 8 Nov 2019

Budhi GS, Chiong R, Wang Z, Dhakal S (2021) Using a hybrid content-based and behaviour-based featuring approach in a parallel environment to detect fake reviews. Electron Commer Res Appl 47:101048. https://doi.org/10.1016/j.elerap.2021.101048

Catal C, Guldan S (2017) Product review management software based on multiple classiers. IET Softw 11(3):8992

Chandy R, Gu H (2012) Identifying spam in the iOS app store. In: Proceedings of the 2nd joint WICOW/AIRWeb workshop on web quality. ACM, pp 56–59. https://doi.org/10.1145/2184305.2184317

Chatterjee S, Chaudhuri R, Kumar A, Wang CL, Gupta S (2023) Impacts of consumer cognitive process to ascertain online fake review: a cognitive dissonance theory approach. J Bus Res 154:113370. https://doi.org/10.1016/j.jbusres.2022.113370

Chen W, Yeo CK, Lau CT, Lee BS (2017) A study on real-time low-quality content detection on Twitter from the users' perspective. PLoS ONE 12(8):e0182487. https://doi.org/10.1371/journal.pone.0182487

Chua AYK, Chen X (2022) Online "helpful" lies: an empirical study of helpfulness in fake and authentic online reviews. In: Smits M (ed) Information for a better world: shaping the global future. iConference 2022. Lecture notes in computer science(), vol 13192. Springer, Cham. https://doi.org/10.1007/978-3-030-96957-8_10

Deshai N, Bhaskara Rao B (2022) A detection of unfairness online reviews using deep learning. J Theor Appl Inf Technol 100(13):4738–4779

Deshai N, Bhaskara Rao B (2023) Deep learning hybrid approaches to detect fake reviews and ratings. J Sci Ind Res 82:120–127. https://doi.org/10.56042/jsir.v82i1.69937

Dhingra K, Yadav SK (2017) Fake analysis of big reviews dataset using fuzzy ranking evaluation algorithm and Hadoop. Int J Mach Learn Cybern 10(8):21432162

Duma RA, Niu Z, Nyamawe AS et al (2023) A Deep Hybrid Model for fake review detection by jointly leveraging review text, overall ratings, and aspect ratings. Soft Comput 27:6281–6296. https://doi.org/10.1007/s00500-023-07897-4

Elmurngi E, Gherbi A (2017) An empirical study on detecting fake reviews using machine learning techniques. In: 7th international conference on innovative computing technology (INTECH). IEEE, pp 107–114. https://doi.org/10.1109/intech.2017.8102442

Elmurngi EI, Gherbi A (2018) Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques. J Comput Sci 14(5):714726

Garcia L (2018) Deception on Amazon—an NLP exploration. https://medium.com/@lievgarcia/deception-on-amazonc1e30d977cfd. Accessed 01 Sept 2019

Ghai R, Kumar S, Pandey AC (2019) Spam detection using rating and review processing method. In: Panigrahi B, Trivedi M, Mishra K, Tiwari S, Singh P (eds) Smart innovations in communication and computational sciences. Springer, Singapore, pp 189–198. https://doi.org/10.1007/978-981-10-8971-8_18

Goswami K, Park Y, Song C (2017) Impact of reviewer social interaction on online consumer review fraud detection. J Big Data 4(1):119

Hajek P (2018) Combining bag-of-words and sentiment features of annual reports to predict abnormal stock returns. Neural Comput Appl 29(7):343–358. https://doi.org/10.1007/s00521-017-3194-2

Hajek P, Sahut J-M (2022) Mining behavioural and sentiment-dependent linguistic patterns from restaurant reviews for fake review detection. Technol Forecast Soc Change 177:121532. https://doi.org/10.1016/j.techfore.2022.121532

Han S, Wang H, Li W et al (2023) Explainable knowledge integrated sequence model for detecting fake online reviews. Appl Intell 53:6953–6965. https://doi.org/10.1007/s10489-022-03822-8

Hassan R, Islam MR (2019) Detection of fake online reviews using semi-supervised and supervised learning. In: 2019 international conference on electrical, computer and communication engineering (ECCE), Cox'sBazar, Bangladesh, pp 1–5. https://doi.org/10.1109/ECACE.2019.8679186

Hussain N, Turab Mirza H, Rasool G, Hussain I, Kaleem M (2019) Spam review detection techniques: a systematic literature review. Appl Sci 9(5):987. https://doi.org/10.3390/app9050987

Jain G, Sharma M, Agarwal B (2018) Spam detection on social media using semantic convolutional neural network. Int J Knowl Discov Bioinform (IJKDB) 8(1):12–26. https://doi.org/10.4018/IJKDB.2018010102

Jain G, Sharma M, Agarwal B (2019) Spam detection in social media using convolutional and long short term memory neural network. Ann Math Artif Intell 85(1):21–44. https://doi.org/10.1007/s10472-018-9612-z

Jiang C, Zhang X, Jin A (2020) Detecting online fake reviews via hierarchical neural networks and multivariate features. In: Yang H, Pasupa K, Leung ACS, Kwok JT, Chan JH, King I (eds) Neural information processing. ICONIP 2020. Lecture notes in computer science(), vol 12532. Springer, Cham. https://doi.org/10.1007/978-3-030-63830-6_61

Kaghazgaran P, Caverlee J, Al M (2017) Behavioral analysis of review fraud: linking malicious crowdsourcing to Amazon and beyond. In: Proceedings of international AAAI conference web social media, vol 11

Kennedy S, Walsh N, Sloka K, McCarren A, Foster J (2019) Fact or factitious? Contextualized opinion spam detection. In: Proceedings of the 57th annual meeting of the association for computational linguistics: student research workshop. ACL, pp 344–350. https://doi.org/10.18653/v1/p19-2048

Krishna A et al (2019) Sentiment analysis of restaurant reviews using machine learning techniques. In: Sridhar V, Padma MC, Radhakrishna Rao KA (eds) Emerging research in electronics, computer science and technology. Springer, Singapore, p 687696

Krishnan HM, Preetha J, Shona SP, Sivakami A (2022) Detection of fake reviews on online products using machine learning algorithms. In: Abraham A, Haqiq A, Muda AK, Gandhi N (eds) Innovations in bio-inspired computing and applications. IBICA 2021. Lecture notes in networks and systems, vol 419. Springer, Cham. https://doi.org/10.1007/978-3-030-96299-9_31

Kurtcan BD, Kaya T (2022) Classification of authentic and fake online reviews with supervised machine learning techniques. In: Xu J, Altiparmak F, Hassan MHA, García Márquez FP, Hajiyev A (eds) Proceedings of the sixteenth ICMSEM 2022, vol 144. Springer, Cham. https://doi.org/10.1007/978-3-031-10388-9_22

Li H, Fei G, Wang S, Liu B, Shao W, Mukherjee A, Shao J (2017a) Bimodal distribution and co-bursting in review spam detection. In: 26th international conference on World Wide Web. ACM, pp 1063–1072. https://doi.org/10.1145/3038912.3052582

Li L, Qin B, Ren W, Liu T (2017b) Document representation and feature combination for deceptive spam review detection. Neurocomputing 254:33–41. https://doi.org/10.1016/j.neucom.2016.10.080

Liu Y, Pang B, Wang X (2019) Opinion spam detection by incorporating multimodal embedded representation into a probabilistic review graph. Neurocomputing 366:276–283. https://doi.org/10.1016/j.neucom.2019.08.013

Lo Presti L, Maggiore G (2021) Vulnerability on collaborative networks and customer engagement: defending the online customer experience from fake reviews. Qual Quant. https://doi.org/10.1007/s11135-021-01249-w

Madisetty S, Desarkar MS (2018) A neural network-based ensemble approach for spam detection in Twitter. IEEE Trans Comput Soc Syst 5(4):973–984. https://doi.org/10.1109/TCSS.2018.2878852

Malik MSI, Hussain A (2017) Helpfulness of product reviews as a function of discrete positive and negative emotions. Comput Hum Behav 73:290–302. https://doi.org/10.1016/j.chb.2017.03.053

Martínez Otero JM (2021) Fake reviews on online platforms: perspectives from the US, UK and EU legislations. SN Soc Sci 1:181. https://doi.org/10.1007/s43545-021-00193-8

Mohawesh R, Xu S, Tran SN, Ollington R, Springer M, Jararweh Y, Maqsood S (2021) Fake reviews detection: a survey. IEEE Access 9:6577165802

Moqueem A, Moqueem F, Reddy CV, Jayanth D, Brahma B (2022) Online shopping fake reviews detection using machine learning. In: Guru DS, Sharath-Kumar YH, Balakrishna K, Agrawal RK, Ichino M (eds) Cognition and recognition. ICCR 2021. Communications in computer and information science, vol 1697. Springer, Cham. https://doi.org/10.1007/978-3-031-22405-8_24

Narciso M (2022) The unreliability of online review mechanisms. J Consum Policy 45:349–368. https://doi.org/10.1007/s10603-022-09514-7

Nasir JA, Khan OS, Varlamis I (2021) Fake news detection: a hybrid CNN-RNN based deep learning approach. Int J Inf Manag Data Insights 1(1):100007. https://doi.org/10.1016/j.jjimei.2020.100007

Pandey AC, Rajpoot DS (2019) Spam review detection using spiral cuckoo search clustering method. Evol Intell 12(2):147–164. https://doi.org/10.1007/s12065-019-00204-x

Patel NA, Patel R (2018) A survey on fake review detection using machine learning techniques. In: 2018 4th international conference on computing communication and automation (ICCCA). IEEE, pp 1–6. https://doi.org/10.1109/ccaa.2018.8777594

Paul H, Nikolaev A (2021) Fake review detection on online E-commerce platforms: a systematic literature review. Data Min Knowl Discov 35:1830–1881. https://doi.org/10.1007/s10618-021-00772-6

Rajamohana SP, Umamaheswari K, Keerthana SV (2017) An effective hybrid cuckoo search with harmony search for review fake detection. In: Proceedings of 3rd international conference advances electrical electronics, information, communication and bio-informatics (AEEICB), p 524527

Ren Y, Ji D (2017) Neural networks for deceptive opinion spam detection: an empirical study. Inf Sci 385:213–224. https://doi.org/10.1016/j.ins.2017.01.015

Ren J, Ozturk P, Luo S (2017) Examining customer responses to fake online reviews: the role of suspicion and product knowledge. In: Fan M, Heikkilä J, Li H, Shaw M, Zhang H (eds) Internetworked world. WEB 2016. Lecture notes in business information processing, vol 296. Springer, Cham. https://doi.org/10.1007/978-3-319-69644-7_18

Rout JK, Singh S, Jena SK, Bakshi S (2017a) Deceptive review-detection using labeled and unlabeled data. Multimed Tools Appl 76(3):3187–3211

Rout JK, Dalmia A, Choo K-KR, Bakshi S, Jena SK (2017b) Revisiting semi-supervised learning for online deceptive review detection. IEEE Access 5:1319–1327

Rout JK, Dash AK, Ray NK (2018) A framework for fake review detection: issues and challenges. In: 2018 international conference on information technology (ICIT). IEEE, pp 7–10. https://doi.org/10.1109/icit.2018.00014

Sa PK, Sahoo MN, Murugappan M, Wu Y, Majhi B (2017) Progress in intelligent computing techniques: theory, practice, and applications. In: Proceedings of ICACNI, vol 2. Springer, Singapore, p 265271

Salminen J, Kandpal C, Kamel AM, Jung S, Jansen BJ (2022) Creating and detecting fake reviews of online products. J Retail Consum Serv 64:102771. https://doi.org/10.1016/j.jretconser.2021.102771

Tang X, Qian T, You Z (2019) Generating behavior features for cold-start spam review detection. In: International conference on database systems for advanced applications. Springer, Cham, pp 324–328. https://doi.org/10.1007/978-3-030-18590-9_38

Tufail H, Ashraf MU, Alsubhi K, Aljahdali HM (2022) The effect of fake reviews on e-commerce during and after Covid-19 pandemic: SKL-based fake reviews detection. IEEE Access. https://doi.org/10.1109/ACCESS.2022.3152806

Valliappan SA, Ramya GR (2023) Identifying fake reviews in relation with property and political data using deep learning. Procedia Comput Sci 218:1742–1751. https://doi.org/10.1016/j.procs.2023.01.152

Vidanagama DU, Silva TP, Karunananda AS (2020) Deceptive consumer review detection: a survey. Artif Intell Rev 53:1323–1352. https://doi.org/10.1007/s10462-019-09697-5

Vidanagama DU, Silva T, Karunananda A (2021) Content related feature analysis for fake online consumer review detection. In: Pandian A, Fernando X, Islam SMS (eds) Computer networks, big data and IoT. Lecture notes on data engineering and communications technologies, vol 66. Springer, Singapore. https://doi.org/10.1007/978-981-16-0965-7_35

Wang J, Kan H, Meng F, Mu Q, Shi G, Xiao X (2020) Fake review detection based on multiple feature fusion and rolling collaborative training. IEEE Access 8:182625182639

Yu C, Zuo Y, Feng B et al (2019) An individual-group-merchant relation model for identifying fake online reviews: an empirical study on a Chinese e-commerce platform. Inf Technol Manag 20:123–138. https://doi.org/10.1007/s10799-018-0288-1

Zhang D, Zhou L, Kehoe JL, Kilic IY (2016) What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. J Manag Inf Syst 33(2):456481

Zhang D, Li W, Niu B, Wu C (2023a) A deep learning approach for detecting fake reviewers: exploiting reviewing behavior and textual information. Decis Support Syst 166:113911. https://doi.org/10.1016/j.dss.2022.113911

Zhang D, Li W, Niu B, Wu C (2023b) A deep learning approach for detecting fake reviewers: Exploiting reviewing behavior and textual information. Decision Support Systems 166:113911. https://doi.org/10.1016/j.dss.2022.113911

Zhang D, Li W, Niu B, Wu C (2023c) A deep learning approach for detecting fake reviewers: exploiting reviewing behavior and textual information. Decis Support Syst 166:113911. https://doi.org/10.1016/j.dss.2022.113911

Zhaoa X, Sunb Y (2022) Amazon fine food reviews with BERT model, 7th international conference on intelligent, interactive systems and applications. Procedia Comput Sci 208:401–406