



Detecting fake reviews through topic modelling

Sule Öztürk Birim^a, Ipek Kazancoglu^b, Sachin Kumar Mangla^c, Aysun Kahraman^d, Satish Kumar^{e,g,*}, Yigit Kazancoglu^f

^a Manisa Celal Bayar University, Salihli Faculty of Economics and Administrative Sciences, Department of Business Administration, Manisa, Turkey

^b Ege University Faculty of Economics and Administrative Sciences, Department of Business Administration, Izmir, Turkey

^c Full Professor and Director, Research Centre on Digital Circular Economy for Sustainable Development Goals (DCE-SDG), Jindal Global Business School, O P Jindal Global University, Haryana, India

^d Manisa Celal Bayar University, Salihli Faculty of Economics and Administrative Sciences, Department of Business Administration, Manisa, Turkey

^e Department of Management Studies, Malaviya National Institute of Technology Jaipur, Jaipur, India

^f Yasar University, Department of Logistics Management, Yasar University 35100 Izmir, Turkey

^g Faculty of Business, Design and Arts, Swinburne University of Technology, Jalan Simpang Tiga, 93350 Kuching, Sarawak, Malaysia

ARTICLE INFO

Keywords:

Machine learning techniques
Fake online reviews
Natural language processing (NLP)
Online retailing
Purchasing decision

ABSTRACT

Against the uncertainty caused by the information overload in the online world, consumers can benefit greatly by reading online product reviews before making their online purchases. However, some of the reviews are written deceptively to manipulate purchasing decisions. The purpose of present study is to determine which feature combination is most effective in fake review detection among the features of sentiment scores, topic distributions, cluster distributions and bag of words. In this study, additional feature combinations to a sentiment analysis are searched to examine the critical problem of fake reviews made to influence the decision-making process using review from amazon.com dataset. Results of the study points that behavior-related features play an important role in fake review classifications when jointly used with text-related features. Verified purchase is the only behavior related feature used comparatively with other text-related features.

1. Introduction

Consumers seek information in order to make the right decisions in their purchasing processes. When planning to make a purchase, they benefit from the opinions and experiences of the people around them as well as their own experiences plus commercial recommendations. Internet and social media affect the buying behaviors of consumers in many ways (Pantano, 2021; Salehan & Kim, 2016). After purchasing the product, many consumers publish their own comments on e-commerce sites (Mo et al., 2015). Online consumer reviews are a type of product information based on users' personal usage experiences (Chatterjee, Goyal, Prakash, & Sharma, 2021; Zhang & Zhang, 2014).

Online reviews have become an influential factor on consumer attitude towards a brand. They help potential customers to make the right decisions by reducing uncertainty through the provision of information about quality of the product, strengths and weaknesses of different products, behavior of dealers, price and delivery time (Chakraborty & Bhat, 2018; Javed et al., 2021; Wang et al., 2015; Mohawesh et al.,

2021). User-oriented online consumer reviews and comments serve as the second most reliable source of product information after recommendations from family and friends (Salehan & Kim, 2016). In addition, 80% of consumers change their purchasing decisions after reading negative consumer reviews; 87% approve their purchasing decisions after reading positive consumer reviews (Zhang et al., 2016). These trends by consumers are an important input for the design of marketing strategies. By using consumer reviews, businesses identify product-related problems and gather market information about competitors; advertising can then be channeled appropriately to promote products and services (Javed et al., 2021). As a result of better monitoring of the market, there is an evolution of positive reflections on financial performance of enterprises such as increasing sales volumes and profitability (Chua & Banerjee, 2016; Agnihotri & Bhattacharya, 2016). These positive results happen when interpretations work properly. However, there are many fake reviews about businesses and their products as well as authentic ones. Fake consumer reviews are deceptive since they are written by people who have little or no experience with the products or

* Corresponding author.

E-mail addresses: sule.ozturk@cbu.edu.tr (S.Ö. Birim), ipek.savasci@ege.edu.tr (I. Kazancoglu), smangla@jgu.edu.in, sachinmangl@gmail.com (S. Kumar Mangla), aysun.kahraman@cbu.edu.tr (A. Kahraman), skumar.dms@mnit.ac.in (S. Kumar), yigit.kazancoglu@yasar.edu.tr (Y. Kazancoglu).

services under review (Zhang et al., 2016). They can have a negative impact on businesses, consumers, the market and society. Outcomes may involve disrupting customer-business relationships, damaging a brand, business image or reputation, hindering positioning strategies, affecting business performance negatively, disrupting information retrieval and processing, causing a decrease in consumer trust, leading consumers to wrong decisions, putting consumers in an ethical dilemma and in the long term, harming online purchasing itself (Chuang, 2020; Gentina, Chen, & Yang, 2020; Kauffmann et al., 2020; Daiv et al., 2020). Faced with these challenges, it is necessary to detect fake reviews and prevent them from appearing on platforms. However, fake reviews are not easily distinguished from authentic ones by only human interpretation (Banerjee et al., 2015). Therefore, for the detection of fake reviews, automated systems based on machine learning algorithms -such as text mining – have been established and developed (Mohawesh et al., 2021). Due to the negative impacts of fake reviews on businesses, it is necessary to use an effective tool which detects fake reviews with high success. Detecting whether a review is real or fake among huge amounts of data is very difficult for individuals or companies. Machine learning algorithms eliminate this difficulty through an ability to identify hidden patterns in the data that humans cannot identify (Mohawesh et al., 2021).

The motivation of this study is to reveal the importance of detecting fake comments due to their societal impact and to develop a method for detecting such comments. The main purpose of this study is to develop a method for the effective detection of fake reviews. The sentiment analysis, topic distributions, cluster distributions and bag of word were carried out in order to determine fake review. This will facilitate the decision-making process of marketing managers and consumers alike. Accordingly, the following research questions were set:

RQ1: How can fake reviews be detected within online purchasing?

RQ2: Which features are more suitable to detect fake reviews?

RQ3: How does the sentiment and topic of the review affect the detection of fake reviews?

The contribution of this study is to reveal which methods and features are effective for the detection of fake reviews; their negative societal impact is a real problem. Selected features for fake detection are sentiment scores, topic distributions, cluster distributions of reviews and bag of words extracted from reviews. The effect of topic distributions on fake detection is not clear according to existing literature. Some studies have found that topic distributions have strong discriminative power (Akoglu et al., 2013) while other studies show topic distributions are less important than behavioral features (Wang et al., 2018). To provide some clarity to the stated controversy, the effect of Latent Dirichlet Allocation (LDA) topic distributions based on fake detection is investigated in this study. Additionally, as an alternative to topic distributions, clusters extracted from Agglomerative Hierarchical Clustering (AHC) are used to predict fake detection. We compare the effect of topic distribution and cluster distribution of reviews with a traditional benchmark matrix, Count Vectorizer. Count Vectorizer (CV), which represents the content of reviews, is regarded as a base model to be compared with the proposed models. Since CV constructs a sparse matrix with a considerable amount of data, more efficient alternatives using a smaller amount of data will be valuable. In effect, we wish to examine the performance of topic and cluster distributions when compared to a sparse matrix. Additionally, the role of purchasing behavior in fake detection is a matter of concern in this study. When users make a purchase and use the product, the practice of fake review writing may diminish (Muhammad & Ahmed, 2019). Therefore, investigating the effect of the verified purchase on the performance of proposed models is another contribution to current literature. Examining fake reviews based on the results of this study will provide valuable information on Interpersonal Deception Theory (IDT), Information Processing Theory, Warranting Theory and Rhetorical Structure Theory (RST).

The study has seven sections. The first section covers the introduction and includes research motivation and research questions. Section 2

includes a literature review and theoretical background. In Section 2, firstly, an overview of the research topic and some previous related studies are discussed. Then, theories used in previous studies about detecting fake reviews and effects of fake reviews on consumer behavior are explored. In Section 3, the methodological structure is explained while in Section 4, results of the research are analyzed. In Section 5, research results are discussed with theoretical contribution, theoretical and managerial implications. Finally, the conclusion, future research and limitations are given in the last two sections.

2. Fake review detection

Consumers attach more importance to the experiences of other consumers, especially when they do not have enough information before making a purchase (Devika et al., 2020). It is possible to gather a range of ideas about products by accessing an increasingly wide user area via the internet. In this respect, the virtual environment offers platforms where consumers can come together. Consumers acquire information through social interaction by sharing their knowledge and experiences with other consumers about products or businesses by using online platforms that provide consumer generated reviews (Banerjee & Chua, 2014). These platforms can be the personal web site of the consumer or independent sites such as Yelp, eBay and Booking.com (Barbado et al., 2019; Plotkina et al., 2020). As they include reviews of consumers who have already bought or used a certain product, customers put faith in the reviews generated by their peers (Agnihotri & Bhattacharya, 2016; Kauffmann et al., 2020). This reduces the time needed, decreases the purchasing risk and avoids making a wrong purchasing decision (Park & Lee, 2008). Many online retailers have started to create systems such as cash, coupons or member points to encourage their consumers to write high-quality reviews (Mo et al., 2015).

Online reviews have become an important source of information for consumers. Online reviews have a significant impact on consumers' purchasing behavior and decision-making processes (Manes & Tchetchik, 2018). However, consumers are the most vulnerable to these reviews due to the existence of fake reviews. Consumers find it very difficult to distinguish between a fake and real review. According to research conducted with more than 10,000 consumers in the USA, UK, France, Germany and Australia, 72% of consumers' state that there should be a new set of standards to combat fake reviews. On the website, 43% of consumers stated that there should be confirmatory information such as a confirmatory purchase after a purchase is made. In addition, consumers, when there has been violation of the determined standards, suggest that a fine of up to 16% of the profit made by the seller should be imposed. Fake reviews from a brand's employees (42%) and other customers (34%) cause them to lose trust in the brand. Trust is important. 54% of consumers say they wouldn't buy a product if they suspect that an accompanying review is fake (Digital Information World, 2020).

Online reviews are classified as positive, negative or neutral (Anusha & Prasad, 2020; Zhuang et al., 2018). If the percentage of positive and influential comments is large, the overall impression of a product also tends to be positive (Devika et al., 2021). Another classification of comments is authentic and fake (Banerjee et al., 2015). Fake refers to false content (Su et al., 2020). Fake reviews are one of the most popular unethical methods and some of these fake comments are carefully written to look realistic. This situation can lead to damaging outcomes to consumers, businesses and the market at the macro level (Barbado et al., 2019). Fake reviews can be written by different parties for different reasons. For example, some companies ask their employees to make positive comments about their own products to boost the company brand; they also ask for negative comments to be made for competitors' products to criticize their merchandise (Banerjee et al., 2015). Thus, while increasing their profits unethically with positive comments, companies are also responsible for a negative impact on the performance of their competitors (Anusha & Prasad, 2020). It is also accepted that people may write reviews even if they do not make a purchase, raising

the suspicion of online comments even more (Anusha & Prasad, 2020). Other parties writing fake reviews may be independent businesses. Some companies engage in writing positive or negative comments for associated businesses in return for financial rewards (Salehan & Kim, 2016). Contents are all human-generated. Some firms have developed computer-generated methods to create fake reviews by using text-generation algorithms (Salminen et al., 2022). Nowadays, fake reviews by digital applications have become one of the primary agenda items in academic research and business management.

Due to their negative effects, businesses are looking for ways to detect and control fake reviews. Some businesses use individuals to detect fake reviews. With this method, businesses first detect clues for fake content; employees who are able to detect fake comments are then specially trained in line with these clues (Busioc et al., 2020). Nowadays the credibility of information varies excessively, with the current situation affecting the accuracy and reliability of the data generated (Su et al., 2020). Thus, detecting fake comments manually may become both difficult and ineffective. To overcome this problem, companies are working to create automatic detection and control systems such as human-crafted rules, traditional machine learning models and neural networks (Brodia, 2018; Devika et al., 2021; Kauffmann et al., 2020; Su et al., 2020).

2.1. Theories on fake reviews

Different theoretical models have been used in literature to investigate the effect of fake reviews on consumer behavior. These include the self-selection theory (Li & Hitt, 2008); interpersonal deception theory (IDT) (Zhang et al., 2016; Plotkina et al., 2020); information processing theory (Wang et al., 2015); source credibility theory (Visetin, 2019); deception theory, signal theory and attribution theory (Munzel, 2015); commitment-trust theory (Ameen et al., 2021); knowledge acquisition theory (Hu et al., 2018); warranting theory (DeAndrea et al., 2018); persuasion knowledge model and attribution theory (Munzel, 2016); belief function theory (Khalifa et al., 2018).

The purpose of individuals reading online product reviews is to obtain information and use it in their purchasing processes. This is stated in Information Processing Theory which explains how people perceive, use, remember and manipulate information (Sucharitha et al., 2020). Consumers in general like to reduce the effort needed for information processing or problem solving by reading online reviews (Wang et al., 2015). Consumers can only process a limited amount of information, depending on their total cognitive abilities, limited short-term memory duration and limited attention span (Siddiqi et al., 2020). According to Information Processing Theory, the quality of information available is a fundamental element (Zhang et al., 2016).

If the subject being examined is online comments and some comments are fake, this negatively affects the quality of the information. Consumers can accept a significant amount of deception as real, without realizing that it is deception in online reviews. Regardless of who wrote the fake comments, the situation that should be emphasized here is that there is a deception between the parties involved in the communication. Deception is a “message knowingly transmitted by a sender to foster a false belief or conclusion by the receiver” (Buller & Burgoon, 1996, p. 205). In other words, the sender of the message portrays a different meaning from the truth. This issue is associated with Interpersonal Deception Theory. The theory explains how individuals consciously or unconsciously handle a real deception when communicating face-to-face (Buller & Burgoon, 1996). Although the theory was developed for interpersonal relationships, it has also been used to explain deceptive messages (such as manipulated, spam or fake reviews) in online interactions since misleading messages disrupt fundamental communication (Tham & Wang, 2017; Wise & Rodriguez, 2013). In online consumer reviews, deception is expressed as a message intentionally posted by the sender to encourage a false belief or behavior in the receiver (Yang & Yu, 2018). Detecting deception in online written

communication is important, because in such cases it is more difficult to expose falsehoods due to a lack of behavioral cues such as consumers fidgeting, eye contact or body movements. With the growing use of social media and e-commerce, text messaging and review writing are increasing and becoming more readily available (Li et al., 2020; Wise & Rodriguez, 2013). Individuals may notice manipulative elements through verbal or non-verbal cues in fake comments and may doubt the information from the source (Maimon et al., 2019).

Another theory that deals with manipulated information is Warranting Theory; this explains what data people should use to guide their decisions. According to this theory, information that is resistant to manipulation is more valuable. By detecting fake comments, manipulated comments can be eliminated, thus providing users with more valuable information (Walther & Parks, 2002). As well as the information processing, the issue of online fake comments is related to the characteristics of the text from which the information is received.

Rhetorical Structure Theory is a theory of text organization that describes the relationships between parts and characteristics of the input text; it explains the coherence and the hierarchical structure of the content (Mann & Thompson, 1988; Rubin & Vashchilko, 2012). Rhetorical Structure Theory (RST) can be used to capture of rhetorical relations among sentences (Zhou & Zarafani, 2020). The origins of Rhetorical Structure Theory, which is a theory of text organization, created date back to the 1980s. The theory is about functions of the text, and how the text involves linguistic entities such as words, phrases, grammatical structure. It explains coherence by postulating a hierarchical, connected structure of texts (Mann & Thompson, 1988). Although the main motivation is text generation, Rhetorical Structure Theory has been used in very different ways. The theory has been used in news in order to identify contents such as description of discourse, detection of fake news (Della Vedova et al., 2018; Shu et al., 2017). Some of these studies used machine learning as a research technic (Han & Metha, 2019; Kraus & Feuerriegel, 2019; Prasanna, 2019; Rubin & Lukoianova, 2014; Rădescu, 2020; Rubin et al., 2015). Rhetorical Structure Theory has been also used for understanding deception in customer complaints (Pisarevskaya et al., 2019), detection of fake online reviews (Popola, 2017).

In this respect, Information Processing Theory, Interpersonal Deception Theory and Warranting Theory address research questions one and two, while Rhetorical Structure Theory addresses research question three of this study.

2.2. Societal impact of fake reviews

Today, fake reviews have become one of the most popular unethical methods of deceiving people (Barbado et al., 2019). The presence of deceptive messages negatively impacts on social welfare, with fake online reviews falling under the category of digital message deception (Anderson & Simester, 2014; Choi et al., 2020). Consumers trust and believe more consumer-to-consumer (C2C) reviews with the effect of eWoM rather than advertising from companies (Costa et al., 2019). Reviews generated by consumers can have a huge impact on the reputation of products and brands. When evaluating the quality of products for their online purchases, consumers rely on reviews written by previous users; these consist of their experiences, feedback and product recommendations. On the other hand, there are fake comments as well as real comments in these reviews. For this reason, companies encourage some consumers to write positive deceptive comments about their products/brands and negative comments about their competitors (Miller, 2009). They pay consumers or individuals to write these reviews (Moon et al., 2021). For example, when examining review scores, readability and responsiveness on Amazon, it was found that approximately 10.3% of product reviews were manipulated. In this way, companies write fake reviews to increase or decrease their average score (Hu et al., 2011, 2018). A large number of positive reviews encourage a consumer to purchase a product and improve the financial profits of the

manufacturer. However, fake reviews written by consumers without making a purchase tend to be significantly more negative than other reviews (Anderson & Simester, 2014). Fake reviews are defined as deviant customer behavior that “violates illegal and generally accepted norms of behavior” (Sigala, 2017, p. 609). Such actions have been evaluated as leading to a situation that can cause physical or psychological harm to companies, employees and customers. Misleading remarks made by reviewers are illegal, infringing on individuals’ behavior by causing physical or psychological harm to consumers or companies (Sigala, 2017).

Fake online reviews disrupt information retrieval and processing and negatively affect the message processing of communication. This situation has caused information asymmetry and deterioration in information (Agnihotri & Bhattacharya, 2016). Incorrect information also causes consumers to make false inferences (Cao et al., 2011). This situation also increases the tendency of consumers to be manipulated due to fake/false interpretations. In addition, the fact that companies create positive or negative fake reviews while pretending to be consumers makes it difficult for potential customers to distinguish a good quality product from a poor, low-quality product. This situation misleads consumers about products and causes them to make wrong judgments about the quality of products or services. Potential customers are also prevented from making purchasing decisions based on correct information (Chakraborty & Bhat, 2018; Pantano, 2021). This may damage the consumer’s trust in the business and cause many products to be returned to the company (Javed et al., 2021; Mohawesh et al., 2021). Consumer skepticism about online reviews may negatively affect genuine online review systems and online shopping in the long run (Abedin et al., 2020; DeAndrea et al., 2018; Javed et al., 2021). In addition, fake reviews encourage consumers to look for substitute products, resulting in revenue losses (Rayana & Akoglu, 2015). Fake reviews have a huge impact on consumer satisfaction. Therefore, when consumers realise that they have been deceived or misled by a fake review, they will become biased against all online reviews and not buy from that e-commerce site again (Alsubari et al., 2021). In addition, this situation reduces consumer confidence in online reviews (van Duyn & Collier, 2019) while creating confusion and doubt about information related to products/brands (Rapp & Salovich, 2018; Di Domenico et al., 2021). Consumers’ trust and attitudes towards the brand are negatively affected (Visentin et al., 2019). Fake reviews can damage the brand reputation of companies and eventually lead to product boycotts (Di Domenico et al., 2021).

While providing an extreme advantage to the competitors of a company due to fake reviews, targeted businesses can lose customers (Ho-Dac et al., 2013). Therefore, identifying fake and genuine reviews has become more important than ever as fake reviews influence social behavior, such as promoting non-existent products or disparaging competing products (Javed et al., 2021).

Most companies that make fake reviews do so to generate revenue for e-commerce sites due to financial problems (Javed et al., 2021). In terms of branding, most of the sellers who buy fake reviews online are not well-known brands; they are usually small, independent firms. These reviews are more effective and more important for these small firms. On the other hand, major brands avoid buying fake reviews in order not to damage their reputation (Hollenbeck, 2018). Star ratings, size and number of rooms are the main factors that are targeted by fake reviews of hotels (Hee et al., 2021). Independent hotels generally publish more false positive reviews than branded hotel chains. High quality hotels will not post fake reviews because their brand equity is strong with the hotel’s established star rating indicate its level of provision. However, independent hotels with weak brand equity benefit more from fake reviews. Hotels with more rooms post more fake reviews, giving the fake reviewer a greater monetary incentive (Moon et al., 2021). For businesses, loss of consumer confidence, especially in e-commerce, is the biggest risk for sellers, retailers and shopping platforms (Jacobs, 2011). In addition, it can negatively affect the performance of the business, damaging their reputation as a result (Xiao & Benbasat, 2011).

2.3. Machine learning as a method for fake review detection

In general, as they are fundamentally deceptive messages, fake reviews reduce societal benefits by manipulating consumer decisions and market information. The volume of fake reviews is increasing very rapidly as e-commerce continues to grow. Therefore, detecting fake reviews becomes a big data problem as traditional statistical techniques become ineffective when dealing with data of this size (Mohawesh et al., 2021). Machine learning is a type of artificial intelligence which aims to form smart systems that can learn from previous data and perform actions based on its acquired knowledge. Big data is used in ML algorithms so that the system learns hidden patterns in the data more easily within the huge amounts available. Datasets for online reviews include very large numbers of customer reviews. While training with the data, ML algorithms aim to learn the whole story about the data; this cannot be performed by individuals. ML algorithms will detect fake reviews automatically - which humans or businesses cannot detect manually - once they have trained on the big data of online reviews (Ahmed et al., 2021). Therefore ML methods increase efficiency in operations and enrich the decision making process substantially by accessing and analyzing big data (Donepudi et al., 2020).

It is vital to point out that although ML techniques are more reliable than traditional statistical or manual detection, existing ML methods do have some areas that need to be improved. A significant problem in ML is to find a definitive clue for classification of reviews as real or fake (Crawford et al., 2015). To find a definitive clue, a common approach is to use a bag of words approach in which a single word or group of commonly used words are detected as features to classify the review as fake or real. However, this approach is not sufficient in accurately classifying fake reviews (Crawford et al., 2015). Therefore, additional methods for feature extraction should be examined to identify a more accurate feature set that will better classify reviews as fake or real. Therefore, it is necessary to design automated fake-review detection approaches designed to elaborate sentiment analysis, natural language processing, polarity classification and review summarization (Anusha & Prasad, 2020). In current literature, various relevant features are described, including review centric and reviewer centric, to detect fake reviews. Reviewer centric features are also called behavioral features that do not depend on the content of the review. Some studies not only use content related features such as readability, topic distribution and distribution related features, but also use behavioral features including percentage of positive reviews and review length to detect fake reviews (Wang et al., 2018). In Wang’s study, reviewer centric behavioral features have more discriminative power than other features. Topic features were less significant than behavioral features. Akoglu et al. (2013) used the positive or negative sentiment of reviews for fake detection. In a recent study, pre-trained language models like BERT or XLNet with Topic distributions obtained from LDA, were used to discover fake news about Covid-19 (Gautam et al., 2021). Authors found that topic distributions are powerful tools in fake news detection. In another study, review content related sparse matrices of TF-IDF, CV and n-gram features were constructed for use in a Principal Component Analysis (PCA) feature set with results showing high accuracy in fake detection (Muhammad & Ahmed, 2019). Sandulescu and Ester (2015) proposed a novel method in which the similarity of the underlying topic distributions of reviews based on LDA is used to classify reviews as fake or genuine. Their results showed that although topic distribution models gave lower precision than bag of words models, their efficiency in performance and ability to infer review semantics was very successful. Another study compared topic distribution performance with unigram and bigram features with several small datasets (Lee & Kyungah Han, 2016). Authors found that topic distributions showed superior performance in some of the datasets.

A recent study utilized Amazon’s fake review dataset to construct a fake detection system (Elmurngi & Gherbi, 2017). Authors used rating, verified purchase, review length, CV and TF-IDF with top 1400 words as

features for classification. They compared performance of Count Vectorizer with TF-IDF and found that CV outperformed TF-IDF by a slight margin. They also found that verified purchase has a significant impact on the classification performance of the proposed models.

As previous research has indicated, various applications of sentiment analysis and topic distributions are widely used for fake prediction. Literature mainly shows that although the discriminative power of topic distributions is not superior to other features in every case, their efficiency and their capability in understanding underlying semantics make topic distributions powerful. In addition to these observations, some studies also indicate that topic distributions discriminate fake reviews better than other feature matrices. There is controversy currently in the body of literature about fake detection performance of topic distributions. This study aims to investigate the effect of topic distributions based on LDA on fake review detection. Topic distribution based on LDA is a multi-level clustering method in which documents can be associated with multiple topics (Blei et al., 2003). An alternative clustering method in which a document can only be associated with one cluster is proposed for LDA in this study. AHC, using a bottom-up method proceeding with a similarity calculation between clusters (Murtagh & Legendre, 2014), is used for comparison with LDA. It is a topic of interest to see whether AHC is comparable to LDA in fake detection in terms of prediction accuracy. We want to compare the effect of topic distribution and cluster distribution of reviews with a traditional benchmark feature matrix, CV. CV is chosen as a benchmark feature matrix to represent review content and will be used as a base model in comparison with LDA based topic distribution and AHC based cluster distribution matrices.

3. Methodology

This section describes the methodology used in this paper based on the research framework given in Fig. 1.

3.1. Dataset description

The dataset used for identification of fake reviews in this paper is Amazon's standard fake product reviews dataset. This dataset is taken from Aayush Saxena's GitHub repository on <https://github.com/aayush210789/Deception-Detection-on-Amazon-reviews-dataset>.

This dataset consists of 21,000 reviews in which there are 10,500 fake and 10,500 real reviews. In this dataset, each review has features of product ID, product name, reviewer name, verified purchase, star rating and a label indicating whether the review is fake or real. Labelling as fake was conducted by the Amazon filtering algorithm embedded in the Amazon website (Alsubari et al., 2021).

The online retail industry was chosen for the application. Consumer reviews have started to become dominant in online reviews, thus constituting a new source of WOM information (Chen & Xie, 2008). Research shows that 63% of individuals use the internet to search for information and 46.4% of individuals use the internet to research products and brands (We are Social & Hootsuite (2021)). Many products are sold on the same platform on online retail sites such as Amazon, eBay, Walmart, Costco; many companies that produce these goods can be found on the platforms as sellers. For this reason, the number of comments and their content are more likely to increase rapidly on such

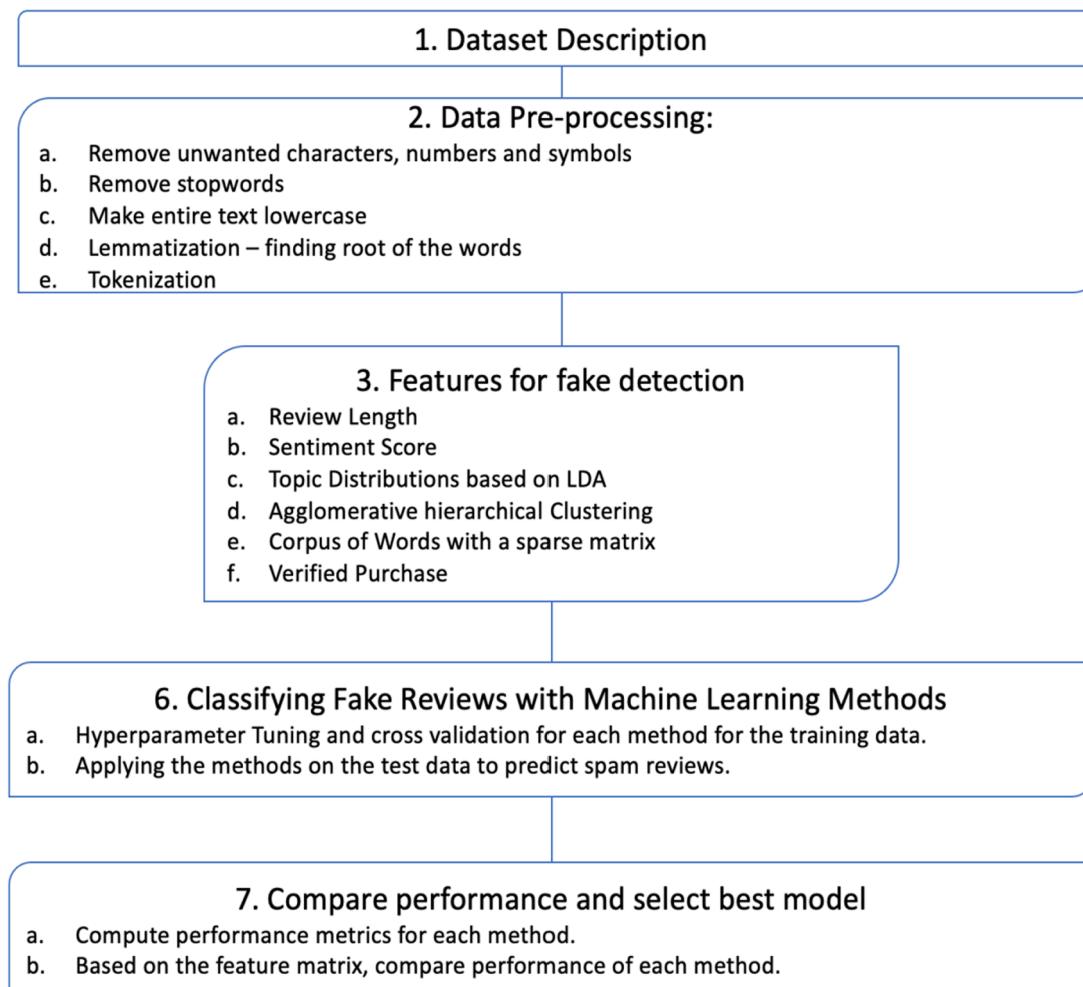


Fig. 1. Research Framework.

platforms. In addition, the online retail market has grown rapidly. In 2020, e-retail sales surpassed \$4.28 trillion worldwide with over two billion people purchasing products online (www.statista.com, 2021a). It has been projected that e-retail revenues will grow to \$5.4 trillion in 2022 (www.statista.com, 2021b). Worldwide, online retail sales accounted for 18% of all retail sales in 2021 and in 2024 it is expected to reach 21.8% (www.statista.com, 2021c). It can be safely concluded that fake reviews will increase with the growth of online retailing. The increase in fake comments and the failure to control them will cause more grievances for market actors. Among these actors, consumers are especially at risk because they can make wrong decisions based on misinformation (Mohawesh et al., 2021). Also, platforms with a large number of sellers, such as Amazon, can make it difficult to control fake reviews (Dacko et al., 2020; Salminen et al., 2022).

Amazon has been selected among online retailers for the application part of this study. The reason for choosing Amazon dataset is Amazon's prominence in online retailing. Amazon is a large, dominant online retailing brand (Chernonog, 2021) which provides many datasets for machine learning applications. More specifically, Amazon provides translucent information about reviewers and promotes huge diversity regarding product reviews (Pieper, 2016). Another reason for choosing Amazon's fake product review dataset is that it includes verified purchase data for each reviewer. From the dataset, we can observe whether the reviewer actually purchased a product or not. This leads us to investigate whether purchasing behavior has an influence on writing real or fake reviews. The last reason for choosing this dataset is it includes an equal amount of fake and real reviews, thus providing a balanced distribution. This balanced distribution enables ML algorithms to learn underlying patterns related to fake and real reviews with similar amounts of data.

3.2. Data Pre-processing

Data pre-processing is conducted to clean the data and make the text data ready for further analysis. With this aim in mind, the following procedures are applied to the customer review data:

Punctuation Removal: With this process, all punctuation marks such as comma, question mark, exclamation mark or colon are removed from the reviews. These marks do not have any meaning; therefore, they will not have any effect on textual analysis.

Stopwords Removal: Commonly used words in English such as "a, an, the, has, and, or, is, are, have etc." which carry no meaning individually are stopwords and they will not have any influence on textual analysis; hence they are removed from the review data.

Text to lowercase: All words in the dataset are converted to lowercase since the algorithms regard uppercase and lowercase versions of the texts as different formats.

Lemmatization: This technique is used to convert the word to its base or dictionary format. Examples of lemmatization are converting plural to singular, verb form to root form. For example, "stays" and "staying" are both converted to "stay" after lemmatization.

Tokenization: Tokenization is dividing text into individual words called tokens (Elmogy et al., 2021). This is one of the basic steps for further NLP techniques. For example, "I was looking for an inexpensive desk" is converted to ["I", "was", "looking", "for", "an", "inexpensive", "desk"] after tokenization.

3.3. Features for fake detection

The aim of fake or spam review detection is to develop a method that includes relevant information about the reviews or reviewers with the goal of accurately identifying fake reviews. In this study, fake detection is conducted by supervised learning using machine learning techniques. All the machine learning techniques in this study will return a label which will classify each review as real or fake. To conduct classification, features to be used in fake review prediction should be identified first.

In existing literature, various features are used to identify fake

reviews; these can be categorized as review centric or reviewer centric (Daiv et al., 2020). Review centric methods use features that are extracted based on bag of words from the reviews (Martinez-Torres & Toral, 2019). Reviewer centric methods aim to find the characteristic behaviors of the reviewers rather than the content of the reviews; these include helpfulness, number of reviews written, verified purchase, name or active membership duration (Pieper, 2016).

In this paper, both review centric and reviewer centric criteria are used in fake detection. The aim is to select the best feature combination which identifies fake reviews most accurately. Review centric features to be comparatively used for fake prediction in this study are sentiment scores, topic distribution, cluster distribution, corpus of words with a sparse matrix. More precisely, the performance of topic distribution, cluster distribution and corpus of words will be compared based on the accuracy of fake detection. Additionally, two reviewer centric behavioral features - review length and verified purchase - are used in this study. One non-verbal reviewer centric feature, verified purchase information of the reviewer, is a particular point of interest in this study. The effect of that non-verbal feature on fake detection performance will be analyzed for each proposed model. In the following section, an explanation of the features along with the rationale behind the selection of each feature is given in detail.

3.3.1. Review length

Studies show that legitimate reviews are lengthier than fake reviews since spammers may not want to devote too much time to writing (Z. Wang et al., 2018). Also, legitimate reviewers may have much more to share than a spam reviewer (Fayazi et al., 2015). There are a number of studies which show that review length can be used as an important predictor for fake detection (Daiv et al., 2020; Dong et al., 2018; Le, 2020). Given the popularity of review length as a predictor of fake reviews in this study, review length is selected as the feature to be used in every model for fake detection.

3.3.2. Sentiment score

Understanding the sentiment of user generated content has been a popular research area in recent years. The aim of sentiment analysis is to understand people's feelings about a specific product or a brand (Mahadevan & Arock, 2020). Sentiment analysis is conducted based on the assumption that people explicitly express their opinions about brands or products within the text that they write (Liu et al., 2017).

Sentiment analysis aims to find the degree of positivity or negativity of the written content about a specific issue such as blogs, news, product reviews or SNS (Social Network Service) (Kauffmann et al., 2020; Kim & Yoo, 2021). A popular type of sentiment analysis is using a lexicon of negative and positive words or phrases to identify the polarity of the document or review from this lexicon (Shivaprasad & Shetty, 2017). Polarity score of content reveals the degree of negativity or positivity of the written text. In this study VADER sentiment analysis is used to identify the polarity score of each review. VADER is a lexicon-based sentiment classifier used to label the texts as positive or negative. The reason for choosing VADER is that it has a gold-standard quality earned from its validation by experts in the field and it is more sensitive to sentiment expressions than its competitors in social media or other domains (Hutto & Gilbert, 2014).

In VADER, after analysis, a document is given a polarity score for each domain of negative, positive, neutral or compound. Among these polarity values, compound score is a useful metric to measure sentiment in a document (Elbagir & Yang, 2019). A threshold value is used to categorize reviews as positive or negative. The review is positive if the compound value ≥ 0.1 while the review is negative if the compound value < 0.1 . Based on the given logical comparison each review is identified as positive or negative.

3.3.3. Topic distribution based on latent Dirichlet allocation topic modelling

Topic modelling is an NLP technique which mines clusters related to

a specific topic based on the word frequency in a given text. Therefore a topic can be defined as a set of words which has a similar overall meaning (Kim & Yoo, 2021). From the range of topic modelling techniques, LDA is used in this study. LDA is a probabilistic model of a bag of words which assumes that documents are characterized as random blends of latent topics in which topics are denoted with a probability distribution of words; word distributions in topics share a common Dirichlet prior (Blei et al., 2003). Each review can also be shown as a probability distribution of identified topics. LDA adopts a certain process to identify topics in a corpus (Lucini et al., 2020):

1. Choose $N \sim \text{Poisson}(\xi)$
2. Choose $\theta \sim \text{Dirichlet}(\alpha)$
3. For each w_n in $w = \{w_1, w_2, \dots, w_n\}$
 - a. Choose $z_n \sim \text{Multinomial}(\theta)$
 - b. Choose w_n from $p(w_n|z_n, \beta)$

where N is the number of words in the document, ξ is Poisson distribution parameter showing the length of each review, θ is topic distribution of the document, α is the parameter of the Dirichlet prior, w_n is nth word used in the document w , z_n is the topic for the nth word, z_n is a set of topics and β is the parameter of the Dirichlet prior on the per topic word distribution.

Probability of the joint topic distribution of the document θ , with a z set of topics in the document w given the Dirichlet parameters of α and β is computed as (Blei et al., 2003):

$$p(\theta, z, w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta)p(w_n|z_n, \beta) \quad (1)$$

For a detailed explanation of the LDA process, Blei et al. (2003) should be consulted.

After the topics are identified as a probability distribution of words, each review is denoted as a probability distribution over topics. In this way, every review has a probability distribution represented as a vector. Probability of the review belonging to the specific topic is regarded as weight. For example, if there are 10 topics in the dataset, then one review will have the following vector given that each number in the vector represents the probability that the given review belongs to the corresponding topic; Review1 = [0.020, 0.020, 0.020, 0.020, 0.020, 0.820, 0.020, 0.020, 0.020, 0.020]. The given vector shows that Review1 belongs to the topic #6 with a probability of 0.820. Similar to the given vector, topic distributions are determined for each review. Determining the number of topics is a vital task in LDA analysis. An unnecessary number of topics could result in a very complex model, making interpretation difficult (Lucini et al., 2020). To determine the ideal number of topics and compare various topic models, perplexity based methods are used in associated literature (Lucini et al., 2020; Zhao et al., 2015). In information theory, perplexity is utilized to assess how well a statistical model describes a dataset (Zhao et al., 2015). In perplexity-using studies, different LDA models with different number of topics are compared based on perplexity metric. The lowest perplexity extracting method is regarded as the best. However, some studies have shown that perplexity based evaluations have some drawbacks such as perplexity lacking an absolute minimum; therefore, with increasing iterations it becomes asymptotic (Krasnov & Sen, 2019). Additionally, perplexity is unclear when used to evaluate the quality of topic modelling from the perspective of human judgement (Koltcov et al., 2014). When compared with human interpreted topics, perplexity is found not to be strongly correlated with human interpretation (Chang et al., 2009). Therefore, the necessity for another metric to compare various topic models has emerged. Topic coherence is another metric that is used to compare various topic models. A topic can be regarded as coherent if its most likely terms can be explained with a single concept (Newman et al., 2010). Topic coherence has several metrics and among these metrics, Normalized Pointwise Mutual Information (NPMI) has been found to

give the highest correlations with human interpretation scores (Röder et al., 2015; Vega-Carrasco et al., 2020). Therefore, in this study, the topic coherence measure NPMI is used to evaluate different LDA models based on a different number of topics. NPMI is calculated as follows:

$$\text{NPMI}(w_i, w_j) = \frac{\text{PMI}(w_i, w_j)}{-\log p(w_i, w_j)}, i \neq j \quad (2)$$

$$\text{where } \text{PMI}(w_i, w_j) = \log \left(\frac{p(w_i, w_j)}{p(w_i)p(w_j)} \right), i \neq j \quad (3)$$

$T = \{w_1, w_2, w_3, \dots, w_n\}$ is a topic generated from a topic model and w_n are the top n words; w_i and w_j are the members of T (Aletas & Stevenson, 2013).

In this study, different models having different number of topics are compared based on topic coherence score of NPMI. The model with the highest coherence score is selected as the best model and the number of topics in the best model is selected as the ideal number of topics for LDA analysis.

Topic modelling (TM) provides advantages over other techniques as it is a natural, scalable framework that can process millions of product reviews while maintaining the explanatory power of these reviews to discover, evaluate and understand customer opinions (Vega-Carrasco et al., 2020). The initial reason for choosing topic distribution for an NLP technique is based on its previous success for rating prediction (Mahadevan & Arock, 2020) and recommendation systems (Liang et al., 2017; Lucini et al., 2020). The second reason for topic distribution is related to the logic that the topics that users write in reviews can be an indicator of whether the review is fake or not. Previously, it has been found that topic distribution provides more discriminative power than other NLP techniques in classifying reviews as fake or real since it provides additional performance improving signals in the model (Gautam et al., 2021).

3.3.4. Agglomerative hierarchical clustering

AHC is a popular statistical method which works on the principle of grouping units according to their similarity. AHC is a bottom-up approach starting with every input as a cluster; as the process moves, clusters are compared based on distance metrics. Then the most similar clusters are grouped together. This process continues until one large cluster emerges in which all the data points are grouped (Vijaya et al., 2019). AHC does not require a fixed number of clusters in the beginning since the clustering dendrogram can be cut at an appropriate level from the user's point of view (Vega-Carrasco et al., 2020).

There are two important parameters in AHC; these are linkage method and distance metric. In this study, the linkage method 'Ward' is used since the most meaningful dendrogram is observed with this linkage. This method, proposed by Ward, 1963, is an AHC method which aims to minimize loss associated with each grouping based on sum of squares criteria (Murtagh & Legendre, 2014). Minimizing loss produces groups with the least inside divergence (Vijaya et al., 2019). In Ward's method, clusters are searched in multivariate Euclidean space (Murtagh & Legendre, 2014); this means that as the distance metric, Euclidean distance is used.

Previous studies have used review similarity as an indicator in fake review detection. In those studies, content similarity between reviews is computed and if the two reviews are above a predetermined threshold than they are regarded as fake (Kauffmann et al., 2020; Z. Wang et al., 2015). These studies have shown that the similarity between reviews is an indicator of whether the review is fake or real. This process is also applied by grouping reviews and observing group content similarity (Mukherjee et al., 2012). All these studies have used similarity measures directly in prediction of fake reviews and have obtained effective results. In this study, the reason for using AHC as an NLP technique is that clustering with AHC is based on item similarity; similarity between reviews can be used as an indicator in fake review detection as the

previous studies have shown. Similar reviews are combined in the same cluster in AHC, resembling group content similarity.

The reason for using AHC as a clustering algorithm is that clustering is based on item similarity and similarity between reviews can be used as an indicator in fake review detection. In previous studies, content similarity between reviews is computed and if the two reviews are above a predetermined similarity threshold then they are regarded as fake (Kauffmann et al., 2020; Wang, Hou, Li, & Song, 2015). This process is also applied by grouping reviews and observing group content similarity (Mukherjee et al., 2012). All these studies have used similarity measures directly in prediction of fake reviews. To the best of our knowledge, no study in current literature has used clusters obtained from AHC based on similarity in fake prediction. Using the obtained AHC cluster distribution matrix in fake detection, this study will make a valuable contribution to the body of literature.

AHC with Ward linkage is applied to the data and based on the dendrogram, the number of clusters is determined. After application of this method, each review's cluster is identified. Based on each review's cluster, one-hot encoding cluster distribution matrix is constructed for the review dataset. In AHC, each review can only be a member of one cluster as opposed to LDA in which reviews can belong to multiple topics. Therefore, the other reason for choosing AHC is that we want to see whether AHC can be an alternative to LDA in classification accuracy.

3.3.5. Corpus of words with a sparse matrix

Review content is also considered as one of the main predictors for fake classification in this study. To use review content in the model, a corpus of words is generated. A bag of words based on word frequency in the reviews is created using the CV method provided by Sklearn in python. To form this vectorizer, the most frequently used 500 words are chosen.

The Counter Vectorizer (CV) matrix, as a review content indicator, is popularly used in classification problems related to NLP. CV is found to bring more accurate predictions when compared to its counterpart Tf-IDF vectorizer. For example, one study used CV in detection of fake reviews on Amazon dataset and found that CV provided more accurate results when compared with Tf-IDF vectorizer (Daiv et al., 2020). Another study used CV and TF-IDF on Sentiment Predictions related to Covid-19 tweets (Raza et al., 2021) and found comparable results. Muhammad and Ahmed (2019) utilized PCA on TF-IDF and CV matrices for fake review detection.

In this study CV matrix is used as a base model to be compared with LDA and AHC models. The reason for choosing CV is to see how the proposed models perform compared to a traditional review content method.

3.3.6. Verified purchase

Verified purchase means Amazon has verified that the user who writes the review has actually purchased the product, also confirming that there was no huge discount (Daiv et al., 2020). By considering the idea that buying and using the product may have an impact on fake review writing behavior, some initial descriptive analysis is conducted to see the ratio of spammers in the verified purchase and non-verified purchase groups. The descriptive analysis result is shown in Table 1 as follows:

Table 1 show that the ratio of fake reviews for users who do not have verified purchase is high; the ratio of true reviews for the users who have verified purchase is even higher. This shows that Amazon verified purchase data may be a helpful predictor in fake review detection. Based on

Table 1

The ratio of spammers in the verified purchase and non-verified purchase group.

	Fake	True
Verified Purchase	Yes	0.137
	No	0.363

the above observation, the effect of verified purchase fake review detection performance is one of the main concerns in this study.

3.3.7. Models used in this study

For the sake of computation efficiency, we selected three main features for comparison; these are topic distribution, AHC distribution and corpus of words. All the models included review length and sentiment scores. We also want to explore the effect of verified purchase for each compared model. In total, there are six models as shown with abbreviations in Table 2:

3.4. Classifying fake reviews

After feature matrices are formed, machine learning classifiers are applied for fake detection analysis. For each machine learning classifier, hyperparameter tuning is applied with 5-fold cross validation. For every analysis, the training set is 80 percent of the data with testing set of 20 percent. Chosen machine learning classifiers and hyperparameters of each classifier are explained in the following section.

Decision Tree Classifier (DTC) is a machine learning classifier that works by building a tree based on instances of training data (Elmogy et al., 2021). In this study, tuned hyperparameters for DTC are maximum depth of decision tree, the minimum number of splits in an internal node and the minimum number of leaves at a terminal node.

Random Forest Classifier (RFC) is a special type of decision tree which eliminates overfitting problems to address traditional decision tree problems. Tuned hyperparameters for RFC are number of trees to construct maximum depth of each tree and the minimum number of leaves at a terminal node.

Support Vector Classifier (SVC) analyses data and determines the limits for decision by using hyperplanes. A hyperplane is used to separate one class from another in a binary classification where the space between the separated parts is kept as large as possible (Hussain et al., 2019). Generally, the recommendation is to use the RBF kernel (Adnane et al., 2019) while penalty parameter C and kernel coefficient gamma hyperparameters are tuned for this study.

Artificial Neural Network (ANN) includes connected processors called neurons which produce real valued activations. Input neurons obtain values from features of the models and establish weighted connections with the neurons in the next layers (Luo et al., 2017). A feed forward ANN uses a backpropagation learning algorithm in which data moves from the input layer to the output layer while weights at each layer are determined (Øyen, 2018). In the back propagation learning algorithm, the weights determined in the layers are updated backwards until the calculated differences between the target values and the actual values are minimized by the optimization function (Azzouni & Pujolle, 2017). Dependent on this procedure are two important hyperparameters in ANN - epoch size and batch size. The number of epochs resembles how many times one full forward and backward run occurs in the analysis (Hewamalage et al., 2021). Since the entire dataset is too large, it must be split into several smaller batches to pass through the network. Batch size is defined as the total number of observations in a single batch (Amirabadi et al., 2020). In this study epoch and batch size are chosen to be hypertuned in ANN classifier.

Table 2
Proposed Models.

Feature Matrices	Abb.
Length, Sentiment, Topic Distribution	TD
Length, Sentiment, Topic Distribution, Verified Purchase	TD-VP
Length, Sentiment, Agglomerative Hierarchical Cluster Distribution	AHC
Length, Sentiment, Agglomerative Hierarchical Cluster Distribution, Verified Purchase	AHC-VP
Length, Sentiment, Count Vectorizer Sparse Matrix	CV
Length, Sentiment, Sparse Matrix, Verified Purchase	CV-VP

3.5. Performance metrics

In this study, chosen performance metrics for the classification accuracy are Precision, Recall, F1-Score and Accuracy which are defined in the following equations. Performance of the classifier is better when these values are larger.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (5)$$

$$\text{F1Score} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (6)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (7)$$

where, TP (True Positive) is defined as the number of reviews classified correctly as fake by the classifier; TN (True Negative) refers to the number of reviews classified correctly as non-fake by the classifier; FP (False Positive) refers to the number of reviews incorrectly labelled as fake; FN (False Negative) refers to the number of reviews incorrectly labelled as non-fake (Dong et al., 2018).

4. Experimental results

Data pre-processing, Vader sentiment analysis, LDA and computations related to machine learning classifiers are conducted with Python 3.8 in the Jupyter-Lab environment. For the computations, several python libraries are implemented; these include nltk for data pre-processing and sentiment analysis, gensim for LDA, Scikit-learn for ML classifiers and Keras for ANN.

The first step of the analysis is to calculate review length for each review. Review length is calculated as the number of characters in each review and stored as a vector. In the second step, review data is pre-processed as explained in Section 2. After pre-processing, VADER sentiment analysis is applied to obtain a sentiment score for each review. Sentiment results are also stored as vectors in which each value represents sentiment of the corresponding review as positive or negative. Determining Number of Topics in LDA, hypertuning and implementation of ML classifiers and reporting the results are explained in detail in the following sections.

4.1. Determining number of topics in LDA

LDA analysis is run for the number of topics changing from 1 to 24. For each run, a coherence score is computed as shown in equation (2). The model which reveals the highest coherence score is selected as the best model. The number of topics used in the best model is selected as the ideal number of topics. Results of LDA runs for each compared number of topics are shown in Fig. 2. As can be seen from Fig. 2, the model with 4 and 10 topics brings the highest coherence value of 0.415 equally. To protect the diversity in the reviews, the ideal number of topics is selected as 10.

LDA analysis is run for 10 topics with topic distributions computed for each review. The topic distribution matrix has 10 columns representing topics. Each column represents the probability of belonging to the selected topic for the reviews.

4.2. Determining number of clusters and application of AHC

As explained in Section three, AHC with Ward linkage is applied to the data. A first dendrogram is constructed to determine the number of clusters in the data. To draw the dendrogram, review data is transformed into a TF-IDF (Term Frequency-Inverse Document Frequency) matrix. TF-IDF determines how important a certain word is in a collection; it is generated by multiplying the term frequency with the log of the ratio of the total number of reviews to the number of reviews in which the term appears (Ahsan et al., 2017). For the sake of computation efficiency, TF-IDF matrixes with the most frequently used 500 words are constructed. A dendrogram with Ward linkage is drawn based on the TD-IDF matrix. The resultant dendrogram can be seen in Fig. 3.

Although Fig. 3 shows most of the documents are grouped in one cluster, four clusters can be extracted from this data based on the given steps in identifying number of clusters based on a dendrogram (Maklin, 2019); the largest vertical distance that does not intersect any of the other clusters is determined first. Then, the optimal number of clusters is equal to the number of vertical lines crossing the horizontal line drawn on the vertical distance.

By applying the above steps, the number of clusters is determined as four. The AHC method is then applied to the pre-processed dataset with four clusters, Ward linkage and Euclidean distance. After AHC application, clusters are identified, meaning that each review belongs to only one of the four clusters. Based on this information, the one-hot encoding method is applied to form a cluster distribution matrix. This method makes sure that the review has a score of 1 for the cluster it belongs to and a score of 0 for the other clusters. Based on one-hot encoding, a

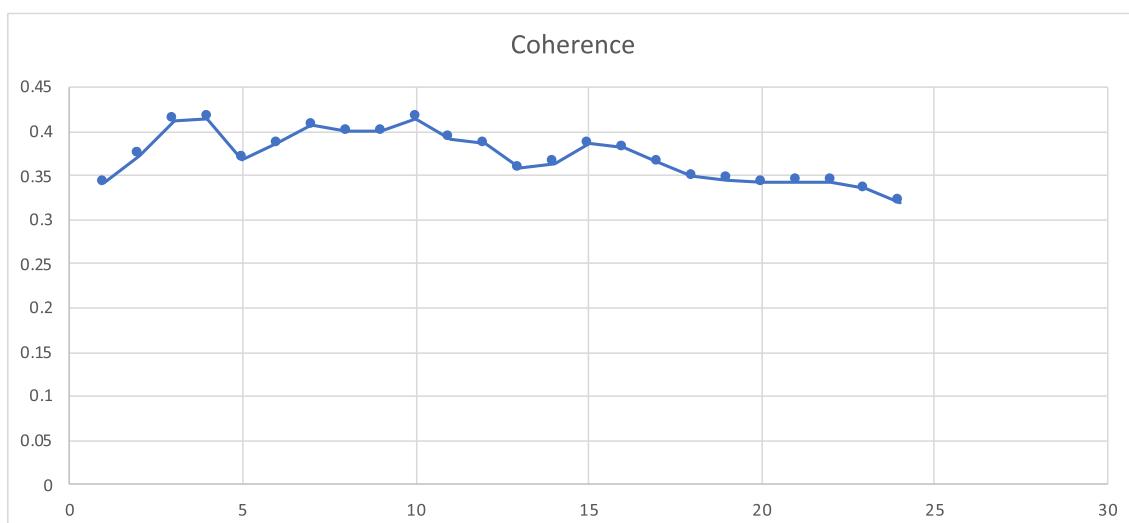


Fig. 2. Coherence values of the compared LDA models.

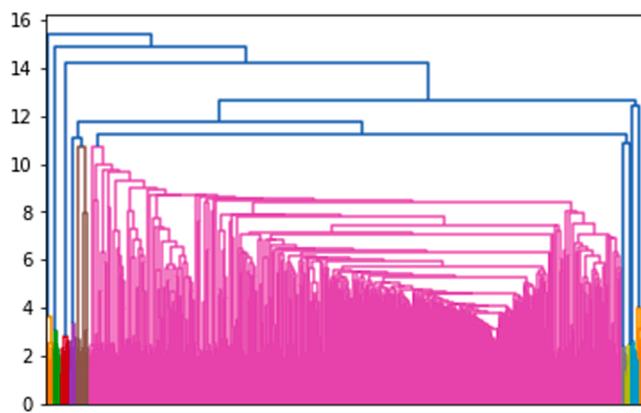


Fig. 3. Dendrogram for AHC.

cluster distribution matrix with four columns representing clusters is formed.

4.3. Constructing Count Vectorizer sparse matrix

In CV sparse matrix, frequency distributions of words are constructed for each review. To form the CV, the most frequently used 500 words are selected. For every review, a frequency distribution of the top 500 words is constructed to form the sparse matrix. This procedure results in a sparse matrix with 500 columns, each column representing a word.

4.4. Construction of feature matrices

Six feature matrices are constructed as explained in [Section 3](#). The aim is to compare the performances of Topic distribution, AHC cluster distribution and CV sparse matrix distribution with or without verified purchase data. Four selected ML classifiers are run for six models. This makes up a total of 24 models to be compared. Hypertuning and Implementation of Machine Learning Classifiers for each model are explained in the next section.

4.5. Hypertuning and implementation of machine learning classifiers

The selected possible values of the hyperparameters for the ML models can be seen in [Table 3](#). For each machine learning classifier, to find the best hyperparameter value, a grid search is applied along k-fold cross validation with k set to 5, where the training set is 80 percent of the data and the testing set is 20 percent. A robustness check is conducted by k-fold cross validation that makes models the least prone to volatile results in each new run.

The designed models for every hyperparameter combination are evaluated on the test dataset using accuracy parameter. Based on the accuracy scores, the best model for each ML technique is selected. The selected hyperparameters that reveal best performance can be seen in [Table 4](#).

Table 3
Candidate values for the hyperparameters.

Model	Hyperparameters	Values
DTC	Maximum depth	[40, 50, 60]
	Minimum splits in an internal node	[2, 4, 6]
	Minimum leaves at a terminal node	[4, 6, 8, 10]
RFC	Number of trees	[400, 800, 1200, 1400]
	Maximum depth	[40, 60, 80, 100]
	Minimum leaves at a terminal node	[6, 8, 10]
SVC	Gamma value	[1, 0.1, 0.01, 0.001]
	C Parameter	[0.1, 1, 10, 100]
ANN	Epoch size	[50, 100, 200]
	Batch Size	[25, 50, 100]

4.6. Results

All the analysis is conducted on MacBook Pro 2.6 GHz Dual-Core Intel Core i5 with 8 GB 1600 MHz DDR3 Ram. Each proposed model accuracy, recall, precision and F1 score is computed and the results are reported in [Table 5](#). At first glance, it appears that the best performance is from RFC in TD + VP model providing highest accuracy, precision and F1 score among the proposed models with values of 80.40%, 86.69% and 78.74 % respectively. The highest recall among the proposed models is observed from ANN in CV + VP model with a value of 78.38%. RF, with the features of review length, sentiment score, topic distribution and verified purchase, classifies fake reviews best when compared with the remaining proposed models.

The results of the proposed models can be seen in [Fig. 4](#). As [Fig. 4](#) shows, when the Verified Purchase (VP) feature is embedded in the models for every ML classifier, the performance of all the feature matrices increase dramatically. To quantify this performance increase, [Table 6](#) is constructed. [Table 6](#) reveals the average and individual performance increase for four ML classifiers based on accuracy, precision, recall and F1 score. The highest average increase is observed in AHC precision score with 28.20%, while the lowest average increase is observed in AHC recall score with 7.64%. The highest individual performance increase is observed in CV ANN recall score with 32.06%, while the lowest individual increase is observed in CV SVC recall score with 7.64%.

When topic and AHC cluster distribution models are compared with the benchmark CV model, it can be clearly seen that topic distribution has superior performance with the RF classifier over CV in most performance metrics. Only in recall, CV with ANN performed best. On the other hand, although AHC cluster distribution did not bring the highest performance scores, when compared to CV, AHC has acceptable results. When [Fig. 4](#) is examined, VP embedded AHC models have comparable results with CV models. Especially in precision and accuracy scores, AHC with VP has superior results over CV with VP in most of the performance metrics. In precision scores, except SVC, all ML classifiers of AHC + VP performed better than the ML classifiers of CV + VP. In accuracy, DTC and ANN of AHC + VP performed better than the corresponding classifiers of CV + VP. This result is a contribution to current literature, showing that AHC clusters can be used in fake prediction since it provides comparable results with CV with more efficiency. CV, with a 500-column matrix where AHC clusters use a four-column matrix denoting AHC, can bring similar results to CV with a smaller size matrix, thus showing greater efficiency.

As can be seen in [Table 5](#) and [Fig. 4](#), the results show that verified purchase, which is a behavior-related feature, plays an important role in fake review classifications when jointly used with text-related features. This finding is consistent with several other findings in literature, confirming the discriminative power of verified purchase over text related features. The behavioral feature set showed better results in fake detection when compared with semantic features ([Wang et al., 2018](#)). [Le, 2020](#), also showed that applying behavioral features is more successful than textual features in fake review classifying problems.

4.7. Robustness check for the results

In each of the proposed models, review length and sentiment score are used as predictors of review classification. To check the robustness of the results, classifiers for the chosen best models are run without the common variables of length and sentiment. The chosen best model is TD + VP for this study. As explained in the previous section, TD + VP yielded the three best performance scores among the four metrics. These three best scores are obtained from the RF classifier. To check the robustness for the RF classifier performance, length and sentiment are removed from the TD + VP model, thus separately conducting two new models. Then both variables are removed together and the three new models are constructed in total. The three new models for TD + VP are

Table 4
Selected Hyperparameters.

Model	Hyperparameters	TD	TD-VP	AHC	AHC-VP	CV	CV-VP
DTC	Maximum depth	40	40	40	40	40	40
	Minimum splits	2	2	2	2	2	2
	Minimum leaves	10	12	10	10	10	10
RFC	Number of trees	1200	400	1200	800	400	800
	Maximum depth	40	40	40	40	80	100
	Minimum leaves	10	8	8	10	6	6
SVC	Gamma value	0.01	0.001	0.001	0.01	0.001	0.001
	C Parameter	0.1	100	0.1	10	10	10
ANN	Epoch size	200	200	100	100	100	50
	Batch Size	50	50	100	100	100	50

Table 5
Performance scores of the proposed models.

Features	ML Technique	Accuracy	Precision	Recall	F1 Score
TD	DTC	59.02%	58.98%	57.83%	58.40%
	RFC	66.90%	65.94%	68.67%	67.28%
	SVC	59.52%	59.68%	60.13%	59.91%
	ANN	60.24%	62.63%	54.62%	58.35%
TD + VP	DTC	75.93%	76.38%	73.20%	74.76%
	RFC	80.40%	86.69%	72.12%	78.74%
	SVC	78.52%	84.94%	69.20%	76.26%
	ANN	80.10%	86.18%	71.28%	78.02%
AHC	DTC	56.40%	56.86%	57.48%	57.17%
	RFC	56.93%	57.16%	59.07%	58.10%
	SVC	59.93%	58.04%	63.14%	60.48%
	ANN	59.45%	57.53%	66.21%	61.57%
AHC + VP	DTC	78.21%	85.23%	68.08%	75.70%
	RFC	77.88%	85.60%	68.28%	75.96%
	SVC	78.43%	85.56%	69.06%	76.43%
	ANN	79.45%	86.01%	71.03%	77.81%
CV	DTC	58.52%	58.46%	60.95%	59.68%
	RFC	63.52%	62.96%	64.77%	63.85%
	SVC	61.10%	60.04%	66.89%	63.28%
	ANN	58.19%	60.61%	46.33%	52.51%
CV + VP	DTC	78.17%	80.75%	74.29%	77.39%
	RFC	80.33%	84.87%	73.14%	78.57%
	SVC	78.76%	86.26%	67.93%	76.01%
	ANN	76.48%	75.36%	78.38%	76.84%

run for each ML classifier to make predictions for fake detection. The results are examined to see whether the best performing RF method is still best overall. Table 7 shows the results for the modified TD + VP

models.

As Table 7 shows, for the three models, RFC shows best performance. For the model TD + VP without length, the model TD + VP without length and sentiment, RFC gives the best scores for accuracy, precision and F1 score. For the model TD + VP without sentiment, RFC shows the best scores for all performance measures. These results confirm that RF classifier continues to be the best method, even when the common variables are extracted from the model.

The aim of robustness check is to show that RF model still continues to yield the highest accuracy although there are some differences in the

Table 6
Performance Increase ratios with Verified Purchase.

	ML Classifier	Accuracy	Precision	Recall	F1 Score
CV	DTC	19.64%	22.29%	13.34%	17.71%
	RFC	16.81%	21.91%	8.37%	14.72%
	SVC	17.67%	26.22%	1.04%	12.73%
	ANN	18.29%	14.74%	32.06%	24.33%
	Totals	18.10%	21.29%	13.70%	17.37%
	AHC	21.81%	28.36%	10.61%	18.53%
AHC	RFC	20.95%	28.44%	9.21%	17.87%
	SVC	18.50%	27.51%	5.93%	15.95%
	ANN	20.00%	28.49%	4.82%	16.24%
	Totals	20.32%	28.20%	7.64%	17.15%
	TD	16.90%	17.39%	15.38%	16.36%
	DTC	13.50%	20.75%	3.46%	11.46%
TD	RFC	19.00%	25.26%	9.07%	16.36%
	SVC	19.86%	23.55%	16.66%	19.67%
	ANN	17.32%	21.73%	11.14%	15.96%
	Totals				

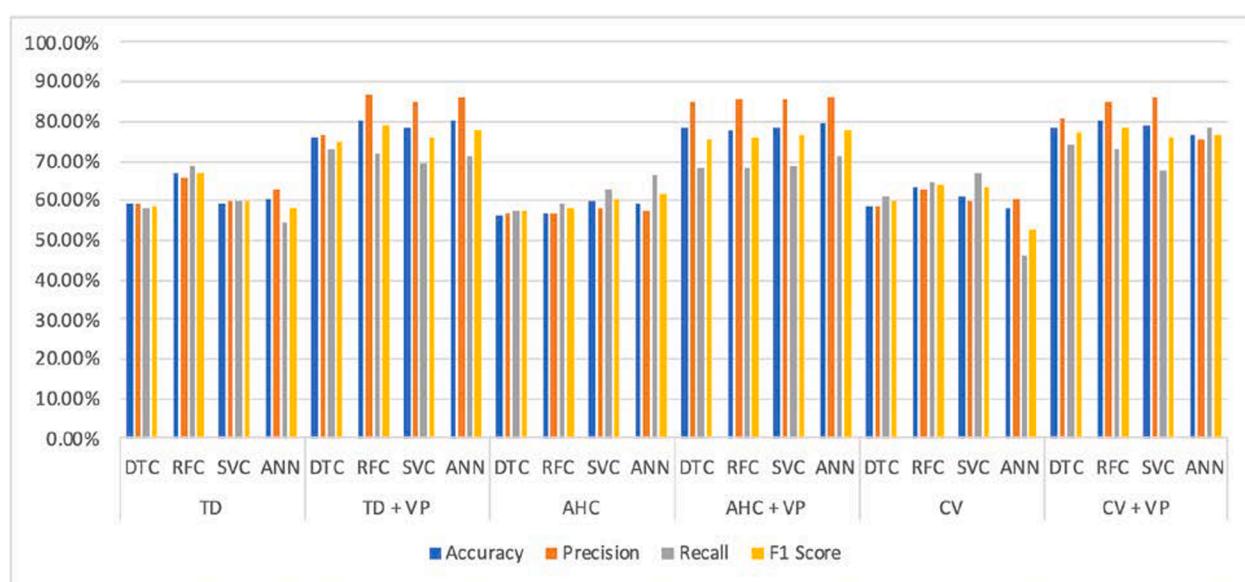


Fig. 4. Performance scores of the proposed models.

Table 7
Performance scores without the common variable/s.

Features	ML Classifier	Accuracy	Precision	Recall	F1 Score
TD + VP without length	DTC	76.57%	77.79%	74.69%	76.21%
	RFC	81.02%	86.45%	73.32%	79.35%
	SVC	77.98%	82.17%	72.44%	77.00%
	ANN	78.71%	82.53%	72.67%	77.29%
TD + VP without sentiment	DTC	75.02%	75.39%	74.06%	74.72%
	RFC	81.57%	86.78%	74.34%	80.08%
	SVC	78.88%	82.18%	73.58%	77.64%
	ANN	79.69%	84.75%	72.24%	78.00%
TD + VP without length and sentiment	DTC	75.93%	77.14%	73.88%	75.48%
	RFC	79.90%	85.50%	72.41%	78.41%
	SVC	78.43%	82.19%	72.27%	76.91%
	ANN	78.69%	84.15%	70.62%	76.80%

model. These results confirm that RF classifier continues to be the highest accuracy, even when the common variables are extracted from the model. When the sentiment score is dropped from the model the performance of RF increases slightly from 80.40% to 81.57% which is a very slight difference. This difference can change when the models run with different random numbers. Therefore, we regard this difference as negligible. In robustness check we are interested in the robustness of the RF model and our results show that RF model is robust.

4.8. Comparison with benchmark studies

Mohawesh et al. (2021) conducted a study on the performance metrics of fake detection methods on various datasets. The study used the Yelp consumer electronics dataset and deception dataset including hotel, doctor and restaurant reviews for comparison and analysis purposes. We should also compare our results with the studies that used Amazon dataset for fake review detection. Table 8 provides comparison of our study with the studies that used Amazon dataset for fake detection. Table 8 includes the results found in our robustness check. The performance metric used for comparison is the accuracy level.

The results show that our method, RF with features TD, VP and length, provide comparable accuracy levels with Daiv et al. (2020) and Hajek et al. (2020). On the other hand, Alsubari et al. (2021) provided a higher accuracy level using a deep learning method. The performance differences from other studies may depend on the method used and the random numbers and hyperparameters used inside the method.

Our study still provides a strong standpoint in classifying reviews as fake or real. RF method is an efficient method that can handle huge amounts of data with various features. RF can handle data with high efficiency (Hengl et al., 2018). Those organizations which prefer a fast working and high-performance tool can use RF technique for their purposes. Sentiment and topic of the review are utilized as the main predictors of fake review detection. Additionally, these predictors are found to be more beneficial when used with verified purchase information since the accuracy scores increased by integrating verified purchase in the data. This information is beneficial for the retailers since retailers use previous purchase information and the topics which users use most to communicate with their customers. Online retailers may prefer to use verified purchase data to understand customer behavior

depending on their reviews. Additionally, online retailers prefer a rapid data processing and high performing tool since high computational power will cost a lot (Martens & Maalej, 2019; Barbado et al., 2019; Salminen et al., 2022).

5. Discussion

There has been an upsurge in complex studies on detection of online fake reviews as online purchasing increases. Online reviews have become more important as information gathering gains influence on the consumer purchasing decision process. Fake reviews have a societal impact as they are deceptive messages, manipulating consumer decisions and harming firms' performance. As the influence of online consumer reviews increases, fake reviews written as if they are real and informative are becoming a worsening social problem, undermining consumers' trust in reviews (Moon et al., 2021). Fake reviews do not reflect the honest opinion of the consumer by giving a false impression at the point of purchase. This leads consumers to make inadequate or erroneous decisions. These fake reviews promote non-existent products or product features. Misleading information affects social behavior in a negative manner (Javed et al., 2021).

To prevent this situation and gain the trust of consumers, many website platforms moderate their own posts. Efforts are made to strengthen the validity and reliability of the website's posts (Luca & Zervas, 2016). However, detecting fake reviews has become an important area of study for researchers and practitioners (Mohawesh et al., 2021; Alsubari et al., 2021). In this respect, it is important to apply effective algorithms using machine learning techniques to make fake review detection automatically.

This study proposes a model to detect fake reviews. Fake reviews pose a threat to companies, consumers and all other actors in the economy. Detecting fake reviews is difficult, especially manually. Therefore, methods are being developed for automatic detection. Among these methods, it has been shown that machine learning techniques can detect fake reviews more effectively. In this respect, it is important to apply new algorithms using machine learning techniques to make fake review detection automatically. The NLP method has been used in previous studies on detecting fake reviews (Abri et al. (2020); Banerjee and Chua (2014); Bhopale et al. (2021); Daiv et al. (2020); Elmogy et al. (2021); Fornaciari et al. (2020); Hovy (2016); Kashti & Prasad (2019); Pieper's (2016); Wang et al. (2018) Wang et al., 2022). However, different data sets were used or the data set belonging to Amazon was analyzed based on different criteria. As Crawford et al. (2015), Hussain et al. (2019) and Pieper's (2016) concluded in their studies, there are too many criteria and there is no such finding as the best feature set. In these studies, criteria are analyzed in different combinations and effective working models are suggested. In this respect, each study contributes to the field because the same criterion set is not used, or the same criterion set is analyzed in different data sets. For instance, Daiv et al. (2020) has used the same data set, but Logistic Regression was used as the prediction method. In the study of Wang et al. (2022), Doc2vec was used for text vectorization by analyzing Yelp public data and Amazon's cell phone data set. However, this study used Count Vectorizer by analyzing a data set of Amazon's miscellaneous products. Similar to our study, RF has good robustness to noise data, so this study has used RF to identify fake reviews. In the study of Baishya et al (2021), deep learning was used, but the data set used is not the same data set used in our study. In addition, there is no used verified purchased criteria in Baishya et al. (2021)'s study. The similar aspect to our work is that it uses Count Vectorizer for bag of words. Deep learning was used as the prediction method. In our study, traditional machine learning methods were used. Verified purchasing was used in Pieper's (2016) study, but the criteria selected for analysis in that study are Reviewer Name, Review Date, Average Rating for Product, Reviewer Rating for Product, Amazon Verified Purchase (Yes/No), Review Text, ARI, Reviewskeptic Label (Nonspam/Spam). In our study, the following

Table 8
Comparing the results with the existing studies using Amazon Dataset.

Study	Method	Accuracy
Pieper (2016)	Reviewskeptic	53.6 %
Daiv et. Al. (2020)	LR with "verified purchase"	82.1%
Alsubari et al.(2021)	Convolutional Neural Network (CNN)	87%
Hajek et al. (2020)	Deep feed-forward neural network (DFFNN)	82.2%
This study	RF with features TD, VP, Sentiment, and length	80.40%
This study	RF with features TD, VP and length	81.57%
This study	RF with features TD, VP and sentiment	81.02%

criteria were used: review length, sentiment score, topic distribution, agglomerative hierarchical clustering, corpus of words and verified purchase. In Pieper's (2016) study, a semi-supervised method was used for spam review detection. In our study, a supervised method was used.

Genuine retailer data was used in this study. Sentiment scores, topic distributions, cluster distributions and bag of words were used to detect fake reviews and to determine which combination of features was most effective. The use of more than one feature in the prediction increases the originality of the research. Using different features improves the performance of the detection model (Zhang et al., 2017). Verbal and non-verbal features were used together in the model; this contributed positively to the performance of the prediction (Sauter, 2017; Chatterjee, Goyal, Prakash, & Sharma, 2021). The results of the study show that the best performance is achieved by RFC, providing the highest accuracy in the TD, VP model. According to the results, behavioral features play an important role in fake review classifications when used in conjunction with text-related features. VP is the only behavioral feature used in comparison to other text-related features.

5.1. Theoretical contributions

To detect fake reviews based on the results of this study will provide valuable information on Interpersonal Deception Theory (IDT), Information Processing Theory and Warranting Theory. The study shows that the theories have possibilities for expansions, modifications, and applications. In addition, the use of these theories can act as a catalyst for future studies built on machine learning to better detect fake reviews. It has also been revealed that these theories, which are generally used in face-to-face relations, can also form the basis of online communication-based studies. It is important to mention those theories which form the basis to the detection of fake reviews. In this context, the study contributes to Rhetorical Structure Theory which focuses on the organization and structure of the text as an input to the description of relations and features. The theory has been applied, compared to other approaches, and also criticized in a number of areas in discourse analysis, theoretical linguistics, psycholinguistics, and computational linguistics (Taboada and Mann, 2006a). Rhetorical Structure Theory is suitable for understanding the phenomenon which has been investigated in this study. It shows that Rhetorical Structure Theory is a good approach for deceptive vs. truthful stories clustering process in automatic solutions. It has been also discussed that Rhetorical Structure Theory can be used not only in manual discourse analysis, theoretical linguistics, psycholinguistics but also in computational linguistics (Taboada and Mann, 2006b). Taboada and Mann (2006a) emphasized that it would be worthwhile to explore how Rhetorical Structure Theory can help explain understanding processes in text, especially on-line (Taboada and Mann, 2006a). However, there is a lack of studies on this structure. Previous studies on computational linguistics have conducted generally to guide computational text generation, not for detecting structure of the text (Taboada and Mann, 2006b). However, this research has shown that Rhetorical Structure Theory is a good framework in computational linguistics. Therefore, it is supported Rhetorical Structure Theory could be applied beyond the original conditions. Prasanna (2019) has emphasized that linguistic techniques can be leveraged using deep learning techniques to achieve accuracy better compared with human. In addition to this, through our study, it has been demonstrated that linguistic techniques can be used in deep learning applications on false comment detection.

5.2. Theoretical and managerial implications

NLP techniques such as characteristics of sentiment scores, topic distributions, cluster distributions and bag of words were used in the proposed model to detect fake reviews. In this way, we have tried to classify fake reviews by using not only behavioral, but also text-related features, together. In addition, the effect on the performance of the

models was investigated by including the verified purchasing feature.

The study does not only reveal the potential negative effects of fake reviews on online retailers, but also emphasizes that fake reviews are a societal impact problem. This study highlights the importance of fake reviews; they are not only responsible for disruption to the fair-trade environment and damaged reputation of companies, meaning that greater consumer awareness against fake reviews must be targeted.

This study provides implications for consumers, companies as well as online shopping platforms. Fake reviews are becoming a particularly important social problem for e-commerce platforms.

Due to the presence of fake reviews, consumers have a problem with trust in both shopping platforms and sellers. Because of this, consumer awareness of fake reviews and warning of possible risks should be encouraged. When making a purchasing decision, consumers should be informed about reading and comparing review texts, rather than just making a decision based on total scores, stars and ratings. The consumer should be educated to prevent being victimized by fake reviews. In this regard, it would be useful for shopping platforms to support and guide the consumer. According to the results of this study, the widespread use of machine learning-based systems to detect fake reviews by shopping platforms or retailers will contribute to the protection of consumers. In this way, the trust of consumers will be gained and the consumer will show a greater tendency to use these platforms or retailers more frequently.

Companies selling in e-commerce on shopping platforms should avoid printing fake reviews. With fake reviews misleading consumers, an unfair competitive environment is created among honest sellers. By only allowing genuine recipients to post comments, non-recipients can be prevented from posting comments. If necessary, companies that promote fake accounts should be penalized (Salminen et al., 2022). Consumer confidence will be increased by eliminating information asymmetry. It is important that companies can benefit from their work in system development based on machine learning in order to eliminate information asymmetry and detect fake reviews. It is also important to point out how fake reviews affect consumers. It is necessary for the wellbeing of the consumer that platforms publish information on their websites to keep potential customers up to date with their processes and procedures.

The absence of any control over fake or spam reviews allows businesses or consumers to provide feedback and suggestions about the product. Reliable and effective detection of online shopping platforms and elimination of fake reviews should be a high priority to maintain the quality of product reviews (Budhi et al., 2021). Retailers should collect and regularly monitor consumer reviews posted on their online sites, intervene if necessary, and then publish review data sets. In this way, they can prevent low-quality comments created on e-commerce sites and monitor the reliability of digital content. This will provide a quality user experience for consumers. Too often, retailers on a platform contact buyers, ask them to write a glossy review and give out refunds and free products in return (Haider, 2021). It is important to ensure the reliability of a source. Requesting detailed information and expertise level of a reviewer should be pursued. Reviewer scoring systems can help to create a fake review identification framework. Consumer confidence can then be further increased by adding past review writing experiences to the comments section as features of users; this will encourage the creation of a user reputation score system. Detecting fake reviews is critical in protecting consumers from falsely advertised products and ensuring fair competition between vendors. Investment should be made in detection methods to reduce the proliferation of fake reviews as they undermine the credibility of the market. As in this study, retailers will be able to automatically detect fake reviews by using different statistical machine learning and artificial intelligence methods, considering security, privacy and terms of use. Business needs to proactively invest in fraudulent review controls. As a result of this study, the operation of a control system based on machine learning by companies will prove important in terms of gaining consumer trust.

This study should be suggested to online shopping platforms and retailers; benefits can result from using a detection model with machine learning-based methods. A filtering system, such as the verified purchase filter on Amazon, can be used to detect fake reviews on commercial websites. In this context, an easy way for platform users to take part in detecting fraudulent activities and to report fake reviews can be provided. Thus, the establishment of a system based on consumer trust can create an environment of fair competition for all companies selling on this platform. This will, prevent harm to the consumer and avoid company performance and reputation being adversely affected. Consumers will be more satisfied with the products they buy and have more faith in the market.

6. Conclusion

The major reasons for the increasing relevance of online fake reviews are due to customer purchasing choices, the damage done by online false reviews and the difficulty in manually detecting such reviews. These online fake reviews can seriously undermine the reliability of genuine customer reviews and may raise doubts about their potential value to businesses and individual consumers.

In this study, a modular system based on sentiment analysis was set up to tackle the critical problem of fake reviews in order to facilitate the decision-making process of marketing managers and consumers. The proposed model considers additional and comparative information from customer evaluations and analyses this data using NLP technology to give sentiment values, a new variable that clarifies the behavior of consumers. In this context, prior works on topics relevant to this framework, such as big data and marketing, sentiment analysis and fake reviews were scrutinized with the findings represented as a summary in the background of the study; this also lists our contributions to the benchmark model.

The findings of this study suggest that there are various scientific and practical implications to enhance the reliability and integrity of online review platforms and the identification of fake reviews online. The outcomes of this study will aid marketing managers and customers in their decision-making processes by detecting these fake reviews via various machine learning models.

7. Limitations and further research

This study has some limitations and future research opportunities. The consumer review data was collected using only Amazon.com's dataset. Therefore, in future studies, this can be done by collecting and comparing the findings of this study with consumer reviews on different product types and/or online shopping platforms. Brands aiming to launch new products could use these methods in the future to study the effects of fake reviews in field settings collaboratively through e-commerce sites such as Amazon.com. In this study, a limited number of features were added and evaluated. In future studies, new features, such as reviewer reputation and source reliability properties can be added to the estimation algorithm.

The results of this research can guide future studies on fake review detection, machine learning techniques, natural language processing in terms of marketing communication strategies and healthy development of e-commerce. This study was used to detect fake reviews through NLP techniques such as the features of sentiment scores, topic distributions, cluster distributions and bag of words. In further studies, a deep learning methodology can be used to classify reviews as fake or real. Other databases which include verified purchases features can also be used for further studies. In addition, further research can be conducted based on different industries such as personal care, pharmaceuticals, restaurants, theatres, supplements or health products. This is important for human health and welfare at a time when misleading consumers in their purchasing intention is prevalent. In this way, the social impact of these industries can be evaluated. Future studies in this area can then assess

the impact and variation between product categories.

In this study, fake online reviews were examined from both the perspective of the company and the consumer; why some consumers produce these fake accounts was not examined. In further studies, the behavioral features of the reviewer plus the psychological and motivational elements that lead consumers to this behavior can be investigated. In addition, the general effects of fake reviews, such as sales on the website, can be examined in future. Moreover, the study can be improved by separating the reviews according to positive and negative reviews and adding behavioral features about how often these reviews are published.

CRediT authorship contribution statement

Şule Birim: Writing – original draft, Validation, Investigation, Data curation, Conceptualization. **Ipek Kazancoglu:** Writing – original draft, Project Administration, Software, Methodology, Investigation, Data curation. **Sachin Kumar Mangla:** Writing – original draft, Project Administration, Software, Resources, Methodology, Investigation. **Aysun Kahraman:** Writing – original draft, Project Administration, Software, Resources, Methodology, Investigation. **Satish Kumar:** Visualization, Supervision, Resources, Project administration, Methodology. **Yigit Kazancoglu:** Writing – review & editing, Visualization, Supervision, Software, Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Abedin, E., Mendoza, A., & Karunasekera, S. (2020). Credible vs fake: A literature review on differentiating online reviews based on credibility. *ICIS 2020 Proceedings*. 2. https://aisel.aisnet.org/icis2020/social_media/social_media/2.
- Abri, F., Gutierrez, L. F., Namin, A. S., Jones, K. S., & Sears, D. R. W. (2020). Fake reviews detection through analysis of linguistic features. *IEEE international conference on machine learning applications (ICMLA '20)*.
- Adnane, M., El Haj Tirari, M., El Fkihi, S., & Oulad Haj Thami, R. (2019). Prediction demand for classified ads using machine learning: An experiment study. In *ACM international conference proceeding series, Part F1481*, pp. 1–6.
- Agnihotri, A., & Bhattacharya, S. (2016). Online review helpfulness: Role of qualitative factors. *Psychology & Marketing*, 33(11), 1006–1017.
- Ahmed, A. A. A., Aljarbouh, A., Donepudi, P. K., & Choi, M. S. (2021). Detecting fake news using machine learning: A systematic literature review. *ArXiv Preprint ArXiv: 2102.04458*. <https://doi.org/10.17762/pae.v58i1.1046>.
- Ahsan, M. N. I., Nahian, T., Kafi, A. A., Hossain, M. I., & Shah, F. M. (2017). An ensemble approach to detect review spam using hybrid machine learning technique. In: *19th international conference on computer and information technology, ICCIT 2016, December 2016*, pp. 388–394.
- Akoglu, L., Chandy, R., & Faloutsos, C. (2013). Opinion fraud detection in online reviews by network effects. In *Proceedings of the 7th international conference on weblogs and social media, ICWSM 2013*, pp. 2–11.
- Aletras, N., & Stevenson, M. (2013). Evaluating topic coherence using distributional semantics. In *Proceedings of the 10th international conference on computational semantics, IWCS 2013 - Long Papers, March 2013*.
- Alsubari, S. N., Deshmukh, S. N., Al-adhaileh, M. H., Alsaade, F. W., & Aldhyani, T. H. H. (2021). *Development of integrated neural network model for identification of fake reviews in E-commerce using multidomain datasets*. 2021.
- Ameen, N., Tarhini, A., Reppel, A., & Anand, A. (2021). Customer experiences in the age of artificial intelligence. *Computers in Human Behavior*, 114, 106548.
- Amirabadi, M. A., Kahaei, M. H., & Nezamalhosseini, S. A. (2020). Novel suboptimal approaches for hyperparameter tuning of deep neural network [under the shelf of optical communication]. *Physical Communication*, 41, 101057.
- Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research*, 51(3), 249–269.
- Anusha, P., & Prasad, K. L. (2020). Survey on fake online reviews using machine learning algorithms. *Journal of Critical Reviews*, 7(18), 2752–2758.
- Azzouni, A., & Pujolle, G. (2017). A long short-term memory recurrent neural network framework for network traffic matrix prediction. *Arxiv Preprint Arxiv:1705.05690v2*.

- Baishya, D., Deka, J. J., Dey, G., & Singh, P. K. (2021). Sentiment analysis-based fake review detection in E-commerce using deep learning. *SN Computer Science*, *2*, 479.
- Banerjee, S. & Chua, A. Y. (2014). Understanding the process of writing fake online reviews. In *Ninth international conference on digital information management (ICDIM)*, pp. 68–73.
- Banerjee, S., Chua A. Y. & Kim J. J., (2015). Using supervised learning to classify authentic and fake online reviews. In *9th international conference on ubiquitous information management and communication*, pp. 1–7.
- Barbado, R., Araque, O., & Iglesias, C. A. (2019). A framework for fake review detection in online consumer electronics retailers. *Information Processing and Management*, *56*, 1234–1244.
- Bhopale, J., Bhise, R., & Mane, A. (2021). A review-and-reviewer based approach for fake review detection. *4th international conference on electrical, Computer and Communication Technologies (ICECCT)*.
- Blei, D. M., Edu, B. B., Ng, A. Y., Edu, A. S., Jordan, M. I., & Edu, J. B. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022.
- Brodia, R. (2018). How to spot fake Amazon reviews. CNET. <https://www.cnet.com/how-to/spot-fake-amazon-reviews-with-fakespot/>.
- Budhi, G. S., Chiong, R., & Wang, Z. (2021). Resampling imbalanced data to detect fake reviews using machine learning classifiers and textual-based features. *Multimedia Tools and Applications*, *80*(9), 13079–13097.
- Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. *Communication Theory*, *6*(3), 203–242.
- Busioc, C., Ruseti, S., & Dascalu, M. (2020). A literature review of NLP approaches to fake news detection and their applicability to romanian language news analysis. *Revista Transilvania*, *10*, 65–71.
- Cao, Y., Li, W., Zhang, J., (2011). Real-time traffic information collecting and monitoring system based on the internet of things. In *6th international conference on pervasive computing and applications*, pp. 45–49. <https://doi.org/10.1109/ICPCA.2011.6106477>.
- Chakraborty, U., & Bhat, S. (2018). The effects of credible online reviews on brand equity dimensions and its consequence on consumer behavior. *Journal of Promotion Management*, *24*(1), 57–82.
- Chang, J., Boyd-Graber, J., Gerrish, S., Wang, C., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems 22 - Proceedings of the 2009 conference* (pp. 288–296).
- Chen, Y., & Xie, J. (2008). Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, *54*(3), 477–491.
- Chernonog, T. (2021). Strategic information sharing in online retailing under a consignment contract with revenue sharing. *Annals of Operations Research*, *300*(2), 621–641. <https://doi.org/10.1007/s10479-020-03807-1>
- Chatterjee, S., Goyal, D., Prakash, A., & Sharma, J. (2021). Exploring healthcare/health-product ecommerce satisfaction: A text mining and machine learning application. *Journal of Business Research* (in press).
- Choi, S. A. (2020). Amazon fake reviews. <https://arxiv.org/ftp/arxiv/papers/2009/2009.09102.pdf>.
- Chua, A. Y., & Banerjee, S. (2016). Helpfulness of user-generated reviews as a function of review sentiment, product type and information quality. *Computers in Human Behavior*, *54*, 547–554.
- Chuang, S. H. (2020). Co-creating social media agility to build strong customer-firm relationships. *Industrial Marketing Management*, *84*, 202–211.
- Costa, A., Guerreiro, J., Moro, S., & Henriques, R. (2019). Unfolding the characteristics of incentivized online reviews. *Journal of Retailing and Consumer Services*, *47*, 272–281.
- Crawford, M., Khoshgoftaar, T. M., Prusa, J. D., Richter, A. N., & Al Najada, H. (2015). Survey of review spam detection using machine learning techniques. *Journal of Big Data*, *2*(1). <https://doi.org/10.1186/s40537-015-0029-9>
- Dacko, S., Schmidt, R., Möhring, M., & Keller, B. (2020). Dealing with fake online reviews in retailing. In E. Pantano (Ed.), *Retail futures* (pp. 85–96). Bingley: Emerald Publishing Limited.
- Daiv, K., Lachake, M., Jagtap, P., Dhariwal, S., Gutte, P. V., Science, C., Science, C., Dhariwal, S., & Science, C. (2020). An approach to detect fake reviews based on logistic regression using review-centric features. June, 2107–2112.
- DeAndrea, D. C., Van Der Heide, B., Vendemia, M. A., & Vang, M. H. (2018). How people evaluate online reviews. *Communication Research*, *45*(5), 719–736.
- Della Vedova, M. L., Tacchini, E., Moret, S., Ballarin, G., DiPietro, M., and de Alfaro, L. (2018). Automatic online fake news detection combining content and social signals. In *2018 22nd conference of open innovations association (FRUCT)*. IEEE, pp. 272–279.
- Devika, P., Veena, A., Srilakshmi, E., Reddy, A. R., & Praveen, E. (2021, July). Detection of fake reviews using NLP & sentiment analysis. In *2021 6th international conference on communication and electronics systems (ICCES)*. IEEE, pp. 1534–1537.
- Di Domenico, G., Sit, J., Ishizaka, A., & Nunan, D. (2021). Fake news, social media and marketing: A systematic review. *Journal of Business Research*, *124*, 329–341.
- Digital Information World (2020). Buyers demand strict penalties against brands relying on fake reviews! (infographic). <https://www.digitalinformationworld.com/2020/03/what-fake-reviews-can-mean-for-your-business-infographic.html>.
- Donepudi, P. K., Ahmed, A. A. A., & Saha, S. (2020). Emerging market economy (EME) and artificial intelligence (AI): Consequences for the future of jobs. *PalArch's Journal of Archaeology of Egypt/Egyptology*, *17*(6), 5562–5574. <https://archives.palarch.nl/index.php/jae/article/view/1829>.
- Dong, L. Y., Ji, S. J., Zhang, C. J., Zhang, Q., Chiu, D. K. W., Qiu, L. Q., & Li, D. (2018). An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews. *Expert Systems with Applications*, *114*, 210–223.
- Elbagir, S., & Yang, J. (2019). Twitter sentiment analysis using natural language toolkit and Vader sentiment. *Lecture Notes in Engineering and Computer Science*, *2239*, 12–16.
- Elmogy, A. M., Tariq, U., Ibrahim, A., & Mohammed, A. (2021). Fake reviews detection using supervised machine learning. *International Journal of Advanced Computer Science and Applications*, *12*(1), 601–606.
- Elmurngi, E., & Gherbi, A. (2017). Detecting fake reviews through sentiment analysis using machine learning techniques. *The Sixth International Conference on Data Analytics Detecting*, 65–72.
- Fayazi, A., Lee, K., Caverlee, J., & Squicciarini, A. (2015). Uncovering crowdsourced manipulation of online reviews. In *SIGIR 2015 - Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval* (pp. 233–242).
- Fornaciari, T., Cagnina, L., Rosso, P., & Poesio, M. (2020). Fake opinion detection: how similar are crowdsourced datasets to real data? *Language Resources and Evaluation*, *54*(4), 1019–1058. <https://doi.org/10.1007/s10579-020-09486-5>
- Gautam, A., Venktesh, V., & Masud, S. (2021). Fake news detection system using XLNet model with topic distributions: CONSTRAINT@AAAI2021 shared task. In *Communications in computer and information science*. https://doi.org/10.1007/978-3-030-73696-5_18
- Gentina, E., Chen, R., & Yang, Z. (2020). Development of theory of mind on online social networks: Evidence from Facebook, Twitter, Instagram, and Snapchat. *Journal of Business Research*, *124*(4), 652–666.
- Han, W., & Metha, W. (2019). Fake news detection in social networks using machine learning and deep learning: Performance evaluation. In *Proceedings of IEEE International Conference on Industrial Internet (ICI)*, 375–380.
- Haider, K. (2021). Complete guide to identify fake reviews on amazon. <https://geekflare.com/identify-amazon-fake-reviews/>.
- Hajek, P., Barushka, A., & Munk, M. (2020). Fake consumer review detection using deep neural networks integrating word embeddings and emotion mining. *Neural Computing and Applications*, *32*(23), 17259–17274. <https://doi.org/10.1007/s00521-020-0475-2>
- Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., & Gräler, B. (2018). Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ*, *2018*(8). <https://doi.org/10.7717/peerj.5518>
- Hewamalage, H., Bergmeir, C., & Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, *37*(1), 388–427.
- Hlee, S., Lee, H., Koo, C., & Chung, N. (2021). Fake Reviews or Not: Exploring the relationship between time trend and online restaurant reviews. *Telematics and Informatics*, *59*, 101560.
- Ho-Dac, N. N., Carson, S. J., & Moore, W. L. (2013). The effects of positive and negative online customer reviews: Do brand strength and category maturity matter? *Journal of Marketing*, *77*(6), 37–53.
- Hollenbeck, B. (2018). Online reputation mechanisms and the decreasing value of chain affiliation. *Journal of Marketing Research*, *55*(5), 636–654.
- Hovy, D. (2016). The enemy in your own camp: how well can we detect statistically-generated fake reviews – An adversarial study. *Proceedings of the 54th annual meeting of the association for computational linguistics*.
- Hu, N., Bose, I., Gao, Y., & Liu, L. (2011). Manipulation in digital word-of-mouth: A reality check for book reviews. *Decision Support Systems*, *50*(3), 627–635.
- Hu, N., Dow, K. E., Chong, A. Y. L., & Liu, L. (2018). Double learning or double blinding: an investigation of vendor private information acquisition and consumer learning via online reviews. *Annals of Operations Research*, *270*(1), 213–234.
- Hussain, N., Mirza, H. T., Rasool, G., Hussain, I., & Kaleem, M. (2019). Spam review detection techniques: A systematic literature review. *Applied Sciences (Switzerland)*, *9*(5), 1–26. <https://doi.org/10.3390/app9050987>
- Hutto, C. J., & Gilbert, E. (2014). VADER: A parsimonious rule-based model for. In *Eighth international AAAI conference on weblogs and social media*, vol. 18. <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109>.
- Jacobs, J. (2011). Television, interrupted. In *Television as digital media* (pp. 255–280). Duke University Press.
- Javed, M. S., Majeed, H., Mujtaba, H., & Beg, M. O. (2021). Fake reviews classification using deep learning ensemble of shallow convolutions. *Journal of Computational Social Science*, *1*–20.
- Kashti, R. P., & Prasad, P. S. (2019). Enhancing NLP techniques for fake review detection. *International Research Journal of Engineering and Technology (IRJET)*, *6*(2), 241–245.
- Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. *Industrial Marketing Management*, *90*(July 2019), 523–537.
- Khalifa, M. B., Elouedi, Z., & Lefèvre, E. (2018). Fake reviews detection under belief function framework. In *International conference on advanced intelligent systems and informatics* (pp. 395–404). Cham: Springer.
- Kim, S. H., & Yoo, B. K. (2021). Topics and sentiment analysis based on reviews of omni-channel retailing. *Journal of Distribution Science*, *19*(4), 25–35.
- Koltcov, S., Koltsova, O., & Nikolenko, S. (2014). Latent dirichlet allocation: Stability and applications to studies of user-generated content. In *WebSci 2014 - Proceedings of the 2014 ACM web science conference* (pp. 161–165).
- Krasnov, F., & Sen, A. (2019). The number of topics optimization: Clustering approach. *CEUR Workshop Proceedings*, *2478*, 1–15.
- Kraus, M., & Feuerriegel, S. (2019). Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees. *Expert Systems with Applications*, *118*, 65–79. <https://doi.org/10.1016/j.eswa.2018.10.002>
- Le, H. (2020). Detection of fake reviews on social media using machine learning algorithms. *Issues in Information Systems*, *21*(1), 185–194.
- Lee, K. D., & Kyungah Han, S.-H. M. (2016). Capturing Word Choice Patterns with LDA for Fake Review Detection in Sentiment Analysis. In *WIMS '16: Proceedings of the 6th international conference on web intelligence, mining and semantics* (pp. 1–7).

- Li, L., Lee, K. Y., Lee, M., & Yang, S. B. (2020). Unveiling the cloak of deviance: Linguistic cues for psychological processes in fake online reviews. *International Journal of Hospitality Management*, 87, 102468.
- Li, X., & Hitt, L. M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4), 456–474.
- Liang, Q., Zheng, X., Wang, M., Chen, H., & Lu, P. (2017). Optimize recommendation system with topic modeling and clustering. In *Proceedings – 14th IEEE international conference on E-business engineering, ICEBE 2017 – Including 13th workshop on service-oriented applications, integration and collaboration, SOAIC 207, November 2017*, pp. 15–22.
- Liu, X., Burns, A. C., & Hou, Y. (2017). An investigation of brand-related user-generated content on twitter. *Journal of Advertising*, 46(2), 236–247.
- Luca, M., & Zervas, G. (2016). Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12), 3412–3427.
- Lucini, F. R., Tonetto, L. M., Fogliatto, F. S., & Anzanello, M. J. (2020). Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. *Journal of Air Transport Management*, 83(June 2019).
- Luo, N., Deng, H., Zhao, L., Liu, Y., Wang, X., & Tan, Z. (2017). Multi-aspect feature based neural network model in detecting fake reviews. In *Proceedings - 2017 4th international conference on information science and control engineering, ICISCE 2017*, pp. 475–479.
- Mahadevan, A., & Arock, M. (2020). Integrated topic modeling and sentiment analysis: A review rating prediction approach for recommender systems. *Turkish Journal of Electrical Engineering and Computer Sciences*, 28(1), 107–123. <https://doi.org/10.3906/elk-1905-114>
- Maimon, D., Santos, M., & Park, Y. (2019). Online deception and situations conducive to the progression of non-payment fraud. *Journal of Crime and Justice*, 42(5), 516–535.
- Maklin, C. (2019). Hierarchical agglomerative clustering algorithm example in python. *Towards Data Science*. <https://towardsdatascience.com/machine-learning-algorithm-s-part-12-hierarchical-agglomerative-clustering-example-in-python-1e18e0075019>.
- Manes, E., & Tchetchik, A. (2018). The role of electronic word of mouth in reducing information asymmetry: An empirical investigation of online hotel booking. *Journal of Business Research*, 85(C), 185–196.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text- Interdisciplinary Journal for the Study of Discourse*, 8 (3), 243–281.
- Martens, D., & Maalej, W. (2019). Towards understanding and detecting fake reviews in app stores. *Empirical Software Engineering*, 24(6), 3316–3355.
- Martinez-Torres, M. R., & Toral, S. L. (2019). A machine learning approach for the identification of the deceptive reviews in the hospitality sector using unique attributes and sentiment orientation. *Tourism Management*, 75(June), 393–403.
- Miller, C. (2009). Company settles case of reviews it faked. New York Times July 14. <http://www.nytimes.com/2009/07/15/technology/internet/15lift.html>.
- Mo, Z., Li, Y. F., & Fan, P. (2015). Effect of online reviews on consumer purchase behavior. *Journal of Service Science and Management*, 8(03), 419–427.
- Mohawesh, R., Xu, S., Tran, S. N., Ollington, R., Springer, M., Jararweh, Y., & Maqsood, S. (2021). Fake reviews detection: A survey. *IEEE Access*, 9, 65771–65802. <https://doi.org/10.1109/ACCESS.2021.3075573>
- Moon, S., Kim, M. Y., & Iacobucci, D. (2021). Content analysis of fake consumer reviews by survey-based text categorization. *International Journal of Research in Marketing*, 38 (2), 343–364.
- Muhammad, F., & Ahmed, S. (2019). Fake review detection using principal component analysis and active learning. *International Journal of Computer Applications*, 178(48), 42–48.
- Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. In *WWW'12 – Proceedings of the 21st annual conference on world wide web, May 2012*, pp. 191–200.
- Munzel, A. (2015). Malicious practice of fake reviews: Experimental insight into the potential of contextual indicators in assisting consumers to detect deceptive opinion spam. *Recherche et Applications en Marketing (English Edition)*, 30(4), 24–50.
- Munzel, A. (2016). Assisting consumers in detecting fake reviews: The role of identity information disclosure and consensus. *Journal of Retailing and Consumer Services*, 32, 96–108.
- Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Bacteriology*, 160 (1), 233–238.
- Newman, D., Lau, J. H., Grieser, K., & Baldwin, T. (2010). Automatic evaluation of topic coherence. In *NAACL HLT 2010 – Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics, proceedings of the main conference, June*, pp. 100–108.
- Oyen, S. (2018). Forecasting multivariate time series data using neural networks. In *Norwegian University of Science and Technology* (Issue June).
- Pantano, E. (2021). When a luxury brand bursts: Modelling the social media viral effects of negative stereotypes adoption leading to brand hate. *Journal of Business Research*, 123, 117–125.
- Park, D. H., & Lee, J. (2008). eWOM overload and its effect on consumer behavioral intention depending on consumer involvement. *Electronic Commerce Research and Applications*, 7(4), 386–398.
- Pieper, A.-T. (2016). Detecting review spam on amazon with review alarm. In *IBA bachelor thesis conference*, pp. 16.
- Pisarevskaya, D., Galitsky, B., Taylor, J., & Ozerov, A. (2019). An anatomy of a lie: Discourse patterns in customer complaints deception dataset. In *Proceedings of the web conference (WWW'19)*. <https://doi.org/10.1145/3308560.3316468>
- Plotkina, D., Munzel, A., & Pallud, J. (2020). Illusions of truth—Experimental insights into human and algorithmic detections of fake online reviews. *Journal of Business Research*, 109, 511–523.
- Popoola, O. (2017). Using rhetorical structure theory for detection of fake online reviews. In *Proceedings of the 6th workshop recent advances in RST and related formalisms* (pp. 58–63).
- Prasanna, D. G. (2019). Fake news detection – A survey of machine learning techniques. *International Journal of Information and Computing Science*, 6(5), 569–573.
- Rădescu, R. (2020). Automatic adaptive system for filtering fake news by methods based on artificial intelligence using machine learning. In *Proceedings of EDULEARN20 conference* (pp. 6659–6667).
- Rapp, D. N., & Salovich, N. A. (2018). Can't we just disregard fake news? The consequences of exposure to inaccurate information. *Policy Insights from the Behavioral and Brain Sciences*, 5(2), 232–239.
- Rayana, S., & Akoglu, L. (2015, August). Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th Acm Sigkdd international conference on knowledge discovery and data mining*, pp. 985–994.
- Raza, G. M., Butt, Z. S., Latif, S., & Wahid, A. (2021). *Sentiment analysis on COVID tweets: An experimental analysis on the impact of count vectorizer and TF-IDF on sentiment predictions using deep learning models*.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *WSDM 2015 – Proceedings of the 8th ACM international conference on web search and data mining* (pp. 399–408).
- Rubin, V. L., Conroy, N. J., & Chen, Y. (2015). Towards news verification: deception detection methods for news discourse. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS48) symposium on rapid screening technologies, deception detection and credibility assessment symposium*.
- Rubin, V. L., & Lukoianova, T. (2014). Truth and deception at the rhetorical structure level. *Journal of the Association for Information Science and Technology*, 66(5), 905–917.
- Rubin, V. L. & Vashchilko (2012). Identification of truth and deception in text: Application of vector space model to rhetorical structure theory proceedings of the EACL 2012 workshop on computational approaches to deception detection, pp. 97–106.
- Salehan, M., & Kim, D. J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30–40.
- Salminen, J., Kandpal, C., Kamel, A. M., Jung, S., & Jansen, B. J. (2022). Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services*, 64(2022), 102771.
- Sandulescu, V., & Ester, M. (2015). Detecting singleton review spammers using semantic similarity. In *WWW 2015 companion – Proceedings of the 24th international conference on world wide web* (pp. 971–976).
- Sauter, D. A. (2017). The nonverbal communication of positive emotions: An emotion family approach. *Emotion Review*, 9(3), 222–234.
- Shivaprasad, T. K., & Shetty, J. (2017). Sentiment analysis of product reviews: A review. In *Proceedings of the international conference on inventive communication and computational technologies*. <https://doi.org/10.1109/ICICCT.2017.7975207>
- Shu, K., Wang, S., & Liu, H. (2017b). Exploiting tri-relationship for fake news detection. arXiv preprint arXiv:1712.07709.
- Siddiqi, U. I., Sun, J., & Akhtar, N. (2020). Ulterior motives in peer and expert supplementary online reviews and consumers' perceived deception. *Asia Pacific Journal of Marketing and Logistics*, 33(1), 73–98.
- Sigala, M. (2017). How "bad" are you? Justification and normalisation of online deviant customer behaviour. In *Information and communication technologies in tourism 2017* (pp. 607–622). Springer, Cham.
- Statista.com (2021a). <https://www.statista.com/topics/871/online-shopping>.
- Statista.com (2021b). <https://www.statista.com/statistics/379046/worldwide-retail-e-commerce-sales>.
- Statista.com (2021c). <https://www.statista.com/statistics/534123/e-commerce-share-of-f-retail-sales-worldwide/>.
- Su, Q., Wan, M., Liu, X., & Huang, C. R. (2020). Motivations, methods and metrics of misinformation detection: an NLP perspective. *Natural Language Processing Research*, 1, 1–13.
- Sucharitha, G.; Matta, A.; Dwarakamai, K. & Tannmayee, B. (2020). In S. N. Mohanty (Ed.), *Theory and Implications of Information Processing* (Chapter 4), Emotion and information processing, Springer.
- Taboada, M., & Mann, W. C. (2006a). Rhetorical Structure Theory: looking back and moving ahead. *Discourse Studies*, 8(3), 423–459. <https://doi.org/10.1177/1461445606061881>
- Taboada, M., & Mann, W. C. (2006b). Applications of rhetorical structure theory. *Discourse Studies*, 8(4), 567–588. <https://doi.org/10.1177/1461445606064836>
- Tham, A., & Wang, M. (2017). There's a troll on the information bridge! An exploratory study of deviant online behaviour impacts on tourism cosmopolitanism. *Tourism Recreation Research*, 42(2), 258–272.
- van Duyn, E. V., & Collier, J. (2019). Priming and fake news: The effects of elite discourse on evaluations of news media. *Mass Communication and Society*, 22(1), 29–48. <https://doi.org/10.1080/15205436.2018.1511807>
- Vega-Carrasco, M., O'sullivan, J., Prior, R., Manolopoulos, I., & Musolesi, M. (2020). *Modelling grocery retail topic distributions: Evaluation, interpretability and stability*. <http://arxiv.org/abs/2005.10125>.
- Vijaya, V., Sharma, S., & Batra, N. (2019). Comparative study of single linkage, complete linkage, and ward method of agglomerative clustering. In *Proceedings of the international conference on machine learning, big data, cloud and parallel computing: Trends, perspectives and prospects, COMITCon 2019*, pp. 568–573.
- Visentin, M., Pizzi, G., & Pichierri, M. (2019). Fake news, real problems for brands: the impact of content truthfulness and source credibility on consumers' behavioral intentions toward the advertised brands. *Journal of Interactive Marketing*, 45, 99–112.

- Walther, J. B., & Parks, M. R. (2002). *Cues filtered out, cues filtered in: Computer-mediated communication and relationships*. In M. L. Knapp & J. A. Daly (Eds.), *Handbook of interpersonal communication* (3rd ed., pp. 529–563). Thousand Oaks, CA: Sage.
- Wang, N., Yang, J., Kong, X., & Gao, Y. (2022). A fake review identification framework considering the suspicion degree of reviews with time burst characteristics. *Expert Systems with Applications*, 190, 116207.
- Wang, X., Zhang, X., Jiang, C., & Liu, H. (2018). Identification of fake reviews using semantic and behavioral features. In *2018 4th international conference on information management, ICIM 2018*, pp. 92–97. <https://doi.org/10.1109/INFOMAN.2018.88392816>.
- Wang, Z., Hou, T., Li, Z., & Song, D. (2015). Spotting fake reviewers using product review graph. *Journal of Computational Information Systems*, 11(16), 5759–5767.
- Ward, J. (1963). *Hierarchical Grouping to Optimize an Objective Function*, 58(301), 236–244.
- We are Social & Hootsuite (2021). Digital 2021 global overview report. <https://wearesocial-cn.s3.cn-north-1.amazonaws.com.cn/common/digital2021/digital-2021-global.pdf>.
- Wise, M., & Rodriguez, D. (2013). Detecting deceptive communication through computer-mediated technology: Applying interpersonal deception theory to texting behavior. *Communication Research Reports*, 30(4), 342–346.
- Xiao, B., & Benbasat, I. (2011). Product-related deception in e-commerce: A theoretical perspective. *MIS Quarterly*, 35(1), 169–195.
- Yang G. H. & Y. Yu, (2018). Use of interpersonal deception theory in counter social engineering. In *CIKM workshops*.
- Zhang, D., Zhou, L., Kehoe, J. L., & Kilic, I. Y. (2016). What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews. *Journal of Management Information Systems*, 33(2), 456–481.
- Zhang, T., Li, G., Cheng, T. C. E., & Lai, K. K. (2017). Welfare economics of review information: Implications for the online selling platform owner. *International Journal of Production Economics*, 184, 69–79.
- Zhang, Y. & Zhang, D. (2014). Automatically predicting the helpfulness of online reviews. In *Proceedings of the 2014 IEEE 15th international conference on information reuse and integration*, August, pp. 662–668.
- Zhao, W., Chen, J. J., Perkins, R., Liu, Z., Ge, W., Ding, Y., & Zou, W. (2015). A heuristic approach to determine an appropriate number of topics in topic modeling. *BMC Bioinformatics*, 16(13), S8.
- Zhou, X., & Zarafani, R. (2020). A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 53(5), 1–40.
- Zhuang, M., Cui, G., & Peng, L. (2018). Manufactured opinions: The effect of manipulating online product reviews. *Journal of Business Research*, 87, 24–35.
- Sule Birim** Manisa Celal Bayar University, Salihli Faculty of Economics and Administrative Sciences, Department of Business Administration, Manisa, Turkey sule.ozturk@cbu.edu.tr ORCID: 0000-0001-7544-8588.
- Ipek Kazancoglu** Ege University Faculty of Economics and Administrative Sciences, Department of Business Administration, Izmir, Turkey ipek.savasci@ege.edu.tr ORCID: 0000-0001-8251-5451.
- Sachin Kumar Mangla** Full Professor and Director, Research centre on Digital Circular Economy for Sustainable Development Goals (DCE-SDG) Jindal Global Business School, O P Jindal Global University, Haryana, India sachinmangl@gmail.com; smangla@jgu.edu.in ORCID: 0000-0001-7166-5315.
- Aysun Kahraman** Manisa Celal Bayar University, Salihli Faculty of Economics and Administrative Sciences, Department of Business Administration, Manisa, Turkey aysun.kahraman@cbu.edu.tr ORCID: 0000-0003-4210-3924.
- Satish Kumar (Corresponding Author)** Department of Management Studies Malaviya National Institute of Technology, Jaipur, India skumar.dms@mnit.ac.in ORCID: 0000-0001-5200-1476.
- Yigit Kazancoglu** Yasar University, Department of Logistics Management Yasar University 35100 Izmir, Turkey yigit.kazancoglu@yasar.edu.tr ORCID: 0000-0001-9199-671X.