The Role of Fake Review Detection in Managing Online Corporate Reputation

R. E. Loke[a] and Z. Kisoen
Centre for Market Insights, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands

Keywords: Online Reviews, Feature Extraction, Spam Detection, Supervised Machine Learning, Strong Corporate Reputation Management.

Abstract: In a recent official statement, Google highlighted the negative effects of fake reviews on review websites and specifically requested companies not to buy and users not to accept payments to provide fake reviews (Google, 2019). Also, governmental authorities started acting against organisations that show to have a high number of fake reviews on their apps (DigitalTrends, 2018; Gov UK, 2020; ACM, 2017). However, while the phenomenon of fake reviews is well-known in industries as online journalism and business and travel portals, it remains a difficult challenge in software engineering (Martens & Maalej, 2019). Fake reviews threaten the reputation of an organisation and lead to a disvalued source to determine the public opinion about brands. Negative fake reviews can lead to confusion for customers and a loss of sales. Positive fake reviews might also lead to wrong insights about real users' needs and requirements. Although fake reviews have been studied for a while now, there are only a limited number of spam detection models available for companies to protect their corporate reputation. Especially in times with the coronavirus, organisations need to put extra focus on online presence and limit the amount of negative input that affects their competitive position which can even lead to business loss. Given state-of-the-art derived features that can be engineered from review texts, a spam detector based on supervised machine learning is derived in an experiment that performs quite well on the well-known Amazon Mechanical Turk dataset.

1 INTRODUCTION

The last few months have changed the landscape of the world drastically (McGrath & Ross, 2020). The outbreak of COVID-19, or the coronavirus, is already stamped as a human tragedy and has a growing impact on the global economy. To sustain, especially the business industry is facing a huge number of challenges to cope with (Gerdeman, 2020). Iansiti et al. (2020) state that business leaders all over the world are struggling with a wide variety of problems from decreasing sales and stalling supply chains to keeping employees safe and ensure that the operational core can continue operating without too many obstacles from the coronavirus. Another recently published study from McKinsey (2020) shows that although the coronavirus has caused the biggest quarterly drops of shares since 1987, a record of unemployment claims and a crude drop of oil prices globally, it has turned more people to technology than ever. Governments around the world have urged people to work from home where possible, this together with the lockdown measures leads to a new way of using technologies in our daily lives. According to the Dutch Institute of International Relations (2020), "COVID-19 is a digital pandemic in terms of its origin, and it is also one in its effects". As workplaces instruct employees to work from home, universities shift fully to online teaching and the restaurant industry transitions faster than before to online ordering and delivering; one of the most rapid organizational transformations in the history of the modern firm is happening right now (Iansiti & Richards, 2020). In this huge digital transformation, organizations are forced to move to a fundamentally new operating architecture based on software, data, and digital networks. With more digitally at stake for organisations, the online corporate reputation has become more important than ever and can mean the deal breaker between surviving in times with the coronavirus or not.

[a] https://orcid.org/0000-0002-7168-090X

Loke, R. and Kisoen, Z.
The Role of Fake Review Detection in Managing Online Corporate Reputation.
DOI: 10.5220/0011144600003269
In *Proceedings of the 11th International Conference on Data Science, Technology and Applications (DATA 2022)*, pages 245-256
ISBN: 978-989-758-583-8; ISSN: 2184-285X

According to Chandler (2020), the coronavirus has driven a massive rise in the use of technology globally. In their recently published article it is stated that "the coronavirus boosted online spending and usage in Q1 of 2020 to the highest in history". It also shows that digital platforms are thriving as consumers seek more entertainment, shopping opportunities and new ways of connecting during the crisis. This increase of online behavior generates more data for organizations to work with improving their online corporate reputation. More organisations start to realize the importance of having an online strategy and strong digital visibility as part of corporate reputation. In times of the corona pandemic, organisations rely more than ever on strong online presence in terms of their websites and apps (Lincoln, 2020). Recent research from The New York Times (2020) stated that people are spending almost one hour a day extra on websites since the outbreak of COVID-19. This means that an important way to reach a broader audience is by having a multi-channel strategy including an app, social media pages and websites. However, with more organisations strengthening their digital strategies; the online market becomes more crowded in terms of competitors. Also, organisations that shift from a traditional marketing toolbox to multi-channel become more vulnerable in terms of corporate reputation. The rise of social media and reviewing websites has empowered consumers and weakened the position of organisations by exposing them to negative publicity, customer attacks and reputation damage (Horn, Taros & Dirkes, 2015). In order to provide a very actual and up-to-date research, this study will focus on the rising concern of fake reviews and its relationship with corporate reputation.

Fake reviews can quite easily be written by anyone on the Internet. Martens and Maalej (2019) state that reviews as a feedback form is often used by managers to prepare organisations for business decisions and to measure corporate reputation of organisations. Research shows that positive feedback improves app downloads, sales and the reputation of the company. However, as a side effect, a market for fake reviews has emerged which can turn into very negative consequences for organisations (Martens & Maalej, 2019). For several years now, there has been done extensive research on the effects of negative and fake reviews on online corporate reputation. Many researchers indicate that small insignificant comments or reviews can have a far-reaching impact on an organisation (DiMauro & Bulmer, 2014). According to Otar (2018), negative and fake reviews can damage corporate reputation online and business growth. Stats show that only four negative or fake reviews can cost an organisation 70% of potential customers (Otar, 2018). Especially fake reviews are recognized as a real challenge by both the research community and the e-commerce industry. As many giant app stores as Google and Apple try to combat against fake reviews, almost 15-30% of all reviews are estimated to be fake per product or service (Barbado et al., 2019). Therefore, fake reviews in app stores can be seen as an actual, critical business problem that affects all layers of businesses.

Fake review detection has been a hot topic in research and industry for many years now (Li, Lui & Qin, 2018). However, it remains interesting to analyse the background and effect of fake reviews in business and, because of a generally noted low accuracy of detecting fake reviews by people, how these can be detected using machine learning methods. With the rising market for apps, organisations have become more vulnerable to user feedback in form of app ratings and reviews. As research shows that even a single fake review can have a significant impact on business, it will be important to take this problem seriously and analyse it below in a survey in more detail. In addition, the outcomes of an experiment that was conducted are reported on below and made available for other companies in order to tackle the issue of fake app reviews.

The remainder of the paper below has been logically structured into sections on literature review (Section 2), research methodology (Section 3), results (Section 4), discusion (Section 5), and conclusions (Section 6).

2 LITERATURE REVIEW

In this section, we first give an in-depth description and background of corporate reputation and then explain and discuss on developments in the phenomenon of fake reviews that can be related to online corporate reputation. Thereafter, we stipulate a preliminary conceptual model and report on common spam review detection techniques.

The aim is to provide an overview of current state-of-the-art knowledge addressing relevant theories, methods, and unforeseen gaps in existing research.

2.1 Corporate Reputation

The definition of corporate reputation has been widely discussed over the years in the research industry and is in continual change. Although it is a hot topic, this concept is still vague and has many different definitions that sometimes even contradict each other. According to Giovanni (2010, p.74), "the reputation of a company can be considered one of the most valued organizational assets". Chun (2005) and Dowling (2016) both agree that corporate reputation has one aligning element; the term is often described as a reflection of the company to insiders and outsiders. Also, corporate reputation is often linked with terms as corporate identity, corporate image, and corporate goodwill (Wartick, 2002; Barnett et al., 2006).

For this study, it will be important to set one straight direction for corporate reputation; therefore, the definition from Fombrun and van Riel (1997) will be maintained throughout the paper. According to their early days research corporate reputation can be identified as "a perceptual representation of a company's past actions and future prospects that describes the firm's overall appeal to all of its key constituents when compared with other leading rivals". Another important finding that comes across in most academic papers on corporate reputation is that many researchers define corporate reputation as a collective concept; it is seen as the sum of the perception of external stakeholders (Barbado et al., 2019; Barnett et al., 2006; Horn et al., 2015). Chun (2005) states that corporate reputation can be seen as an umbrella construct for corporate image and corporate identity.

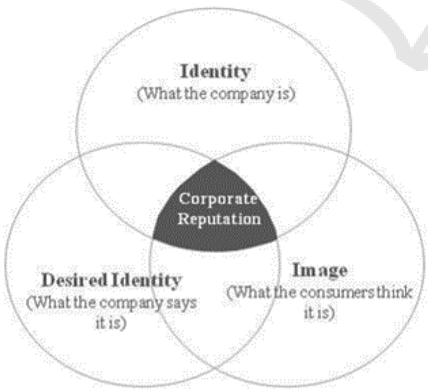

Figure 1: Key elements of corporate reputation (Chun, 2005).

Figure 1 shows the statement from Chun (2005) that identity, desired identity, and image are partly independent variables that form corporate reputation. Image can be described as the perception of others of a company or how it is formulated "how others see us" (Chun, 2005). Identity can be described as an internal view of the company what means how members of the organization perceive, feel, and think about the company ("how we see ourselves"). Desired identity describes how an organisation wants to be perceived which refers to the name, logo, symbol as well as strategic actions and philosophy ("how we want others to perceive ourselves"). The gap in the middle represents how an organisation is being perceived internally and externally, as well as how it wants to be perceived (Chun, 2005). A wide gap indicates inconsistencies in strategy or communication and can damage the corporate reputation of an organisation. Walker (2010) states that alignment between these variables can lead to strategic benefits, such as increasing profitability, lower costs, and a competitive advantage.

We will now discuss on the relevant concepts of electronic worth-of-mouth (EWOM), online corporate reputation and corporate reputation management.

Table 1: Touch points of EWOM (adapted from Mishra & Satish, 2016).

Stage	Touch points of EWOM
Problem or interest	External stimuli (ads on websites, social media personalization and recommendations)
Information search	Search engines, social media, product websites, e-retailers
Evaluation of alternatives	Websites with compare options, social media for feedback, online reviews, and rating websites
Purchase decision	Channels (e-commerce websites), discussion and feedback on social media
Post-purchase behaviour	Review sites, social media, online rating and reviews, feedback on social media or product sites

2.1.1 EWOM (Electronic Word-of-Mouth)

Internet and social media platforms have added a new element to the traditional word-of-mouth (WOM) term. Electronic word-of-mouth, or EWOM, refers to any positive or negative content made by potential, actual, or previous customers about a product or company, which is made available to an audience of people and institutions via the web 2.0 (Mishra & Satish, 2016). EWOM is expressed in different forms of communication such as opinions, online ratings, online feedback, reviews, comments, and experience sharing via online communication channels. According to a study from Mishra & Satish (2016) on EWOM, it plays a critical factor in marketing efforts and has an impact on different stages in the consumer purchase decision process. Table 1 shows how consumers are in touch with EWOM during the purchasing process (Mishra & Satish, 2016; Dewey, 1910).

Although there seems to be a clear link between EWOM and corporate reputation, there is little literature on this connection. Hoyer and Macinnis (2001) found out that WOM is the most credible and objective influence on corporate reputation. Other researchers agree that in meeting or exceeding customers' expectations, customer satisfaction is achieved, EWOM is uttered, and good reputations are built (Davies et al. 2010). However, the corporate reputation of companies is considered fragile; while it may take time to build, it can be easily destroyed.

2.1.2 Online Corporate Reputation

A concept that often simultaneously appears with corporate reputation is online reputation. According to Jones, Temperley and Lima (2010), "online corporate reputation is a reputation, which involves a corporate reputation created in the online environment". Online reputation is not only created on social media but is also created by groups of people sharing and collaborating online and through search engines as Google, Ask and Yahoo (Weber Shandwick & KRC Research, 2019). In this digital era, online corporate reputation is as important as offline reputation (Abimbola & Vallaster, 2009). The emergence of social media platforms and review websites allows people to have new tools to publically judge companies at a much greater and faster pace than before. On these platforms, consumers do not only discuss content from companies, but they also create it (Barnett et al., 2006). Fournier and Avery (2011) have defined social media as "a venue for open-source branding" in which consumers can co-create the nature of reputations of a brand. Companies try to influence this process of co-creation by creating solid online presence and strong online marketing strategies. The online presence, according to Waters et al. (2009), "offers various benefits to companies like the opportunity to communicate directly with customers, strengthen relationships, stimulate co-creation and to assess consumer's brand attitudes". Nowadays, companies experience more pressure from outside to take part in online conversations that influence corporate reputation. Therefore, the online corporate reputation is associated with increased loss of control and increased need for active monitoring (Gensler et al., 2015).

2.1.3 Corporate Reputation Management

Since the overall goal of this research is to contribute to a good online reputation management for companies (for example, by emphasizing genuine reviews in EWOM to consumers and eliminating fake ones), it is important to understand the meaning behind reputation management.

According to Hutton et al. (2001), reputation management, which is considered a business function, is based on the traditional term "public relations", or also known as "corporate affairs". Beal & Strauss (2009) state that online reputation management is placed between marketing communications, public relations, and search engine optimization (SEO). Jones et al. (2010) agree with this definition as they list: "online reputation management is the process of positioning, monitoring, measuring, talking and listening as the organization engages in a transparent and ethical dialogue with its various online stakeholders". What comes across from different literature is that to build and maintain corporate reputation, it is important for a company to understand who its stakeholders are and how they perceive the company (Beal & Strauss, 2009). This can be linked to the umbrella theory of Chun (2005) and is aligned with the perception that reputation is formed by a collective perception of different individuals. The more the perceptions of several individuals are aligned with each other, the stronger the corporate reputation of a company (Gensler et al., 2015).

When looking at how corporate reputation can best be maintained, research from Page and Fearn (2005) indicates that organisations should focus on aligning the perceptions of different stakeholders. To do so, organisations should focus on clear communication about leadership and successes of the organisation and the organisation's perspective on consumer fairness in advertisement, marketing, websites, reviews, and other forms of communication. To go more in-depth on this: the reputation of an organisation is reflected by the leadership style and its successes from the CEO. A clear example of this is Tesla, an automotive company that is mainly known for its famous CEO, Elon Musk. The reputation is also reflected by consumer fairness including the fair treatment of consumers regarding pricing, quality of products and services and transparency in advertisement which also includes reviews.

To conclude on reputation management, literature indicates that it is important for organisations to measure, monitor and co-ordinate the different stakeholder reputations with the overall goal to align these as much as possible. Page and Fearn (2005) and (Gensler et al., 2015) emphasize strongly that the more different stakeholder reputations are similar, the stronger the corporate reputation of an organisation is. To create alignment, organisations should focus on creating clear and transparent messages with regards to leadership style, successes of an organisation, advertisement and marketing communication. It is important for an organisation to be authentic and transparent towards all its stakeholders.

2.2 The Role of Fake Reviews in Corporate Reputation

In today's tech-savvy world, review websites, social media and mobile applications have become the most important source for consumers to express themselves. It is considered very easy for people to share their views about products and services using e-commerce websites as TripAdvisor and Trustpilot, forums and blogs (Hussain, Mirza, Rasool, Hussain & Kaleem, 2019). In app stores in particularly, users can rate downloaded apps on a scale from 1 to 5 stars and write a review message in which they can express satisfaction, report bugs, or make suggestions (Martens & Maalej, 2019). A recent study on online consumer buying behaviour confirms the statement that most people read these reviews about products and services before buying them (Xhema, 2019). In case of apps, consumers often read through the reviews before deciding to download the app. Harman, Jia, and Zhang (2012) identified in their research that there is a positive relationship between the number of positive ratings and reviews to sales and download ranks of apps. As is stated, "stable numerous ratings lead to higher downloads and sales numbers", which will have a positive effect on corporate reputation (Barnett et al., 2006).

As a result of the positive connection between reviews and sales, a new illegal market that is focused on producing fake reviews has emerged. The phenomenon of producing fake reviews on products and services with the goal to boost sales is also referred to in academic studies as "spam attack" (Hussain et al., 2019). In regular situations, real users are motivated by their satisfaction level to provide feedback on apps; however, fake reviewers get paid or similarly rewarded to submit reviews (Martens & Maalej, 2019). An important distinction between real users and fake reviewers is that fake reviewers might not even be real app users and thus their reviews might not be truly reflecting honest opinions. According to Martens and Maalej (2019), fake reviews can be defined as non-spontaneous, requested and rewarded. Another definition states that a fake review is a positive, neutral, or negative review that is not an actual consumer's honest and impartial opinion or that does not reflect a consumer's genuine experience of a product, service, or business (Fontanarava et al., 2017).

Many studies agree that fake reviews have a negative effect on the online corporate reputation of a company (Horn et al., 2015; Barbado et al., 2019; Xhema, 2019; Hussain et al., 2020). One of the main issues with opinion sharing websites and apps is that fake reviews can easily create hype about a particular product based on misleading information. These fake reviews can become the key factor for consumers in their buying decision and thus lead to negative financial consequences. Although it seems clear for people that not everything on the Internet is believable, research shows that almost 84% of consumers consider online reviews to be as trustworthy as personal recommendations. However, for organisations to make use of fake reviews or to have fake reviewers can harm the corporate reputation by creating false expectations. Also, true reviews can help organisations learn where to improve and can be beneficial in increasing success for business. Secondly, if an organisation gets caught buying fake reviews for its own products or for decreasing the value of those of its competitors, it will lead to much more reputation loss than it possibly would gain. An example from 2013 is Samsung which was fined for paying people to negatively review HTC products. Another example is a report from BBC that showed that fake online reviews get openly bought and sold and that shoppers often can get products for free in return for fake reviews.

We now first underline why it can be extremely important for a company to focus on *strong* corporate reputation management and then elaborate on the role of fake reviews in consumer buying behavior that can be related to corporate reputation.

2.2.1 Benefits of Strong Corporate Reputation Management

The above-mentioned examples indicate what can happen if organisations do not put effort in strong reputation management and alignment of stakeholder reputations as discussed by Page and Fearn (2005). Positive reputation can strengthen the overall performance of an organisation, while negative reputation is considered a competitive disadvantage (Aula, 2010).

According to Helm and Klode (2007), there are five major benefits that strong corporate reputation can bring to an organisation. These are as follows: (1) Increased financial performance; (2) Greater competiveness; (3) Higher satisfaction and loyalty among consumers; (4) Attract and retain employees; (5) Support in crisis. Some explaining notes on this: Firstly, the first benefit logically can result in an increased stock value. According to Helm and Klode (2007), a strong reputation limits risks for investors, who are more willing to spend money on the organisation. Secondly, the second benefit goes hand-in-hand with increased financial performance. Helm and Klode (2007) identify that organisations with strong corporate reputation can easily charge higher prices due to the fact that consumers perceive the quality of products and services as better. Thirdly, several studies indicate that a good corporate reputation can increase benefit number (3) (Helm and Klode, 2007; Chun, 2005; van Riel & Fountain; 2008). Fourthly, a positive company image attracts more highly skilled employees, hence, benefit number (4) (Helm & Klode, 2007). Lastly, according to Helm and Klode (2007), in times of crisis for an organisation, a positive reputation can help companies to overcome economic consequences. Organisations with a strong image experience less market decline compared to organisations with a weak reputation (van Riel & Fombrun, 2008).

To conclude, strong corporate reputation can bring several major benefits to an organisation. These benefits are linked to financial, strategic, and competitive advantages that all have a positive effect on the performance of an organisation. Therefore, it is highly advisable and important for an organisation to focus on strengthening its corporate reputation and on limiting threats as fake reviews.

2.2.2 Fake Reviews in Consumer Buying Behaviour

A study from Constantinides and Fountain (2008) describes relationships when consumers are exposed to information about organisations. There are four identified stimulating factors, A, B, C and D, see Figure 2, that each affect the purchasing decision. Although purchasing behaviour should be threatened separately from corporate reputation, it is important to describe the theory from Constantinides and Fountain (2008) in order to emphasize the role that fake reviews play in purchasing behaviour. Organisations that use fake reviews, attempt to make from stimuli D a controllable stimulating factor. Since Constantinides and Fountain (2008) postulate that all stimulating factors are equally distributed, this explains why organisations with bad reputations, as part of their sales strategies, focus on making the uncontrollable controllable (Grutzmacher, 2011).

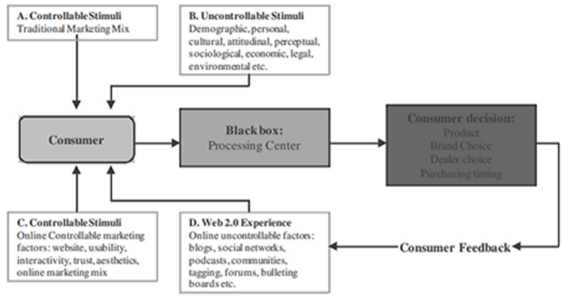

Figure 2: Four stimuli on consumer behaviour (Constantinides and Fountain, 2008).

2.3 Conceptual Model

The goal of the conceptual model that we postulate is to visualize the concepts in this study and indicate the modeling playground between fake reviews and corporate reputation. The model, inspired by Fombrun (1997), shows how several variables that have been identified in the above literature frame are related to each other and eventually create corporate reputation; see Figure 3.

According to Dowling (2016), firstly, corporate identity is, in short, how people recognize an organisation. Secondly, corporate image is defined as "a set of beliefs and feelings an audience has about an organization". This all leads to corporate reputation, that is formed by the judgement about the organisation's attributes as is indicated in the conceptual model.

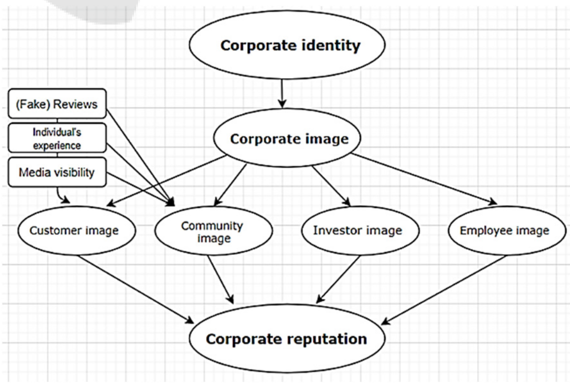

Figure 3: Fake reviews and corporate reputation: a conceptual framework that we propose in this paper that has been derived from scientific literature (see text in 2.3).

We stipulate that (fake) reviews play an important role on the perception of the customer and community image.

2.4 Spam Review Detection

As fake reviews are becoming much more of a problem with more review websites popping up and with consumers' ability to produce feedback at any time, demand for spam detection methods is rising. As we have discussed, such methods are needed for strong corporate reputation management. However, as much more research has recently appeared on the topic of spam detection, the practical implication seems to remain a challenge. Major review websites as Yelp and Amazon have already taken first steps in detection of fake reviews on their websites; however there seems to be a lot of room for improvement. For instance, Hussain et al. (2019) researched several spam detection techniques. According to their paper, spam detection consists out of the following steps: (1) Gather a review dataset; (2) Select feature engineering; (3) Apply, for example, machine learning techniques. Below, each of these three steps will be separately discussed in depth in order to generate useful findings for implementing a spam detection model in the experiment that we set up.

2.4.1 Gathering a Review Dataset

To be able to set up a machine learning model for review spam detection, it is important to have a dataset to work with. However, in terms of spam detection it is considered difficult to find an available, labelled dataset (Hussain et al., 2019). A prior inventory on spam detection models indicates that there is only one labelled hotel review dataset available that includes review text and has no other features available (Kaggle, 2020). Many of the studies that analyze spam detection methods do not publish used datasets publicly, which makes it difficult for new researchers to continue to optimize and improve on spam detection models. It can be stated that after researching multiple studies on spam detection, only a limited number of labelled datasets are available which is contradicting the high current urgency for spam detection methods in society.

2.4.2 Feature Engineering

According to Hussain et al. (2019), the linguistic approach is the most common approach for feature extraction from review datasets. As they explain in their research, this approach focuses on review text and includes data pre-processing, tokenization, transformation, and feature selection. In the next section, in Section 3, an experimental setup of how all these practical steps can be executed for spam detection will be given; we will now proceed with discussing step 4 which is the most crucial step because it has the most significant effect on the performance of spam detection models. Previous research on feature selection, according to Hussain et al. (2019), shows that the following spammer features are used to detect spam and non-spam reviews: (1) Maximum number of reviews: previous research indicates that spammers write often more than one review per day. (2) Percentage of positive reviews: most spammers write positive and favourable reviews; therefore, a high percentage of positive reviews could indicate spam reviews. (3) Review length: most spammers do not write very lengthy reviews with a lot of details. Therefore, short reviews can indicate spam reviews. (4) Reviewer deviation: spammers give often very high ratings, therefore this rating deviates from the average review rating. (5) Maximum content similarity: research shows that similar reviews are used for multiple products and services over different organisations.

After analyzing several sources in the study of Hussain et al. (2019) it shows that the linguistic approach holds the highest accuracy in terms of spam detection methods. However, it all depends on the feature selection process as features become the input for the actual spam review detection method that might be in place.

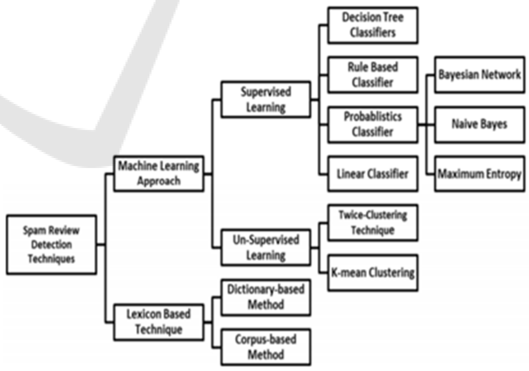

Figure 4: Taxonomy of spam review detection techniques (Hussain et al., 2019, p. 13).

2.4.3 Machine Learning Techniques

To be able to classify reviews in the two classes of spam and non-spam, it will be needed to choose the appropriate classification model. Hussain et al. (2019) published a taxonomy of spam detection techniques (Figure 4). It was created to enable other researchers "to classify existing approaches and to figure out the most appropriate technique to solve a spam detection problem". Spam detection models fall into two categories (see again, Figure 4): machine-learning based methods and lexicon-based approach. The first approach can be classified into supervised and unsupervised learning. Research shows that the accuracy of supervised learning in terms of Support Vector Machine and Naïve Bayes is best; for unsupervised learning, Aspect Based, and K-Nearest Neighbour is best. In this paper, the focus will lie on machine learning techniques, therefore, the Lexicon-based approach will not be further discussed. For an overview of accuracy rate per approach, please refer to Hussain et al. (2019).

3 RESEARCH METHODOLOGY

We adopted an exploratory research methodology, as we intended to generate general insights about the fake review problem that the business industry is currently dealing with in society. Of course, we held our main drive that is targeted towards the relevancy of online corporate reputation for e-commerce in the back of our minds.

Figure 5 graphically represents our research methodology that consisted of data collection, preparation, and analysis processes. Below, we give some more detail about the datasets that we employed as well as how we concretely implemented our data processing and machine learning.

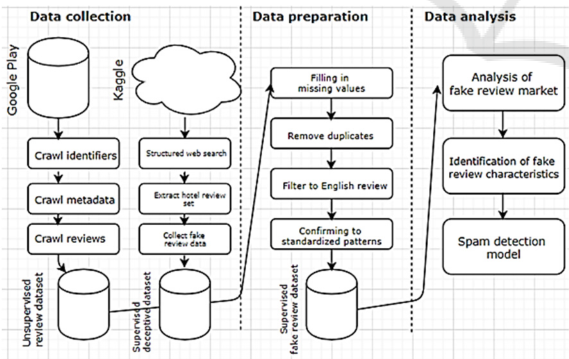

Figure 5: Research design process.

3.1 Datasets

Our main dataset was obtained from the open-source data platform Kaggle (2020). This is a well-known hotel reviews dataset from Amazon Mechanical Turk in combination with TripAdvisor. It has been created for researchers to provide new solutions on the fake review issue and to develop and test new spam detection models. During the production of this (supervised) dataset, a group of people were paid to write 400 fake positive and 400 fake negative reviews about hotel experiences. These deceptive reviews were added together with 800 genuine, thruthfull reviews (again, 400 positive and 400 negative). The total number of reviews is hence 1600. Positive and negative refer to the sentiment in a review text. The dataset was obtained in April 2020, via the Kaggle website.

As a good common practice, the dataset was descriptively analysed on any differences between the groups deceptive versus truthful, negative versus positive, and TripAdvisor versus non-TripAdvisor. We found no significant statistical difference (p-value = 0.2) for the length of words in deceptive and truthful reviews, but significant statistical difference (p-value = 0) for the length of words in negative and positive reviews as well as in TripAdvisor and non-TripAdvisor reviews.

In addition, with the goal to test and apply our algorithms to other datasets that are relevant for many businesses, a scraper was built to crawl about 200.000 product reviews of eight food and beverage suppliers from the review site Google Play. For this scraper, an algorithm was built in Python by using the package Google Play Scraper. However, it was too challenging and too costly to turn this dataset into a supervised dataset with spam identification labels that we needed to feed our supervised machine learning algorithms.

3.2 Data Processing

After understanding the dataset and the structure of the data, the next step was to process the data.

Non meaningful stop words were removed from reviews using a natural language toolkit library.

It is important to explain how relevant review text features were computed. First, we logically computed the length of words variable that we already have mentioned in the previous section. Second, we included the sentiment polarity (positive or negative). Third, since Ott et al. (2011) state that in review classification there is a large difference between informative and imaginative writing, namely that the former typically consists of more nouns, adjectives, prepositions, determiners, and coordinating conjunctions, while the latter includes more verbs, adverbs, pronouns, and pre-determiners, for each word in a review, Parts of Speech components were extracted to be able to feed this as a feature vector in the machine learning model. Fourth, we experimented with weighting meaningful words to form topics.

All review text data were vectorized using TfidfVectorizer.

3.3 Machine Learning

Once relevant features were extracted, it was time to split the data into a training and test set. We used the following, common split: 80% training and 20% testing. Only the training data was used to implement machine-learning models.

We implemented the following machine learning models: Logistic Regression, Decision Tree, Support Vector Machine (SVM), and Random Forrest.

We used GridSearchCv to finetune and find best hyperparameters for ML algorithms.

We also systematically tested our machine learning models using Random States.

4 RESULTS

The machine learning model that did best was Support Vector Machine; see Table 2 for several of its model performance scores on the test data set of 320 reviews. The accuracy rate was 89%. When we systematically tested this model with random states, a slightly higher accuracy rate could be obtained.

Table 2: Several performance scores of our SVM machine learning model.

	precision	recall	f1-score	support
deceptive	0.88	0.90	0.89	155
truth	0.90	0.88	0.89	165
avg / total	0.89	0.89	0.89	320

To check on any welcome generalization capabilities of the machine learning model, we also tested the model on several reviews from Yelp that are likely to be in the same application domain and the outcomes were promising; see Figure 6.

```
test_string("The hotel was good.The room had a 27-inch Samsung led tv, a microwave.The room had a double bed")
array(['truth'], dtype=object)
test_string("we stayed at the schicago hilton for 4 day and it was an amazing experience")
array(['deceptive'], dtype=object)
```

Figure 6: Generalization capabilities of our SVM machine learning model on some Yelp reviews.

Table 3: Comparison of several good-scoring spam detection models, i.e., different supervised learning techniques on different datasets (adapted from Hussain et al. (2019)).

Dataset	Learner	Accuracy
Amazon.com	Logistic Regression (LR)	78%
Epinions.com	Naive Bayes (NB)	63%
Hotel reviews through Amazon Mechanical Turk (AMT)	Support Vector Machine (SVM)	89.9%
Amazon.com	Support Vector Machine (SVM)	71%
Yelp's real-life data	Support Vector Machine (SVM)	86.1%
Hotel reviews through Amazon Mechanical Turk (AMT)	Support Vector Machine (SVM)	84%
Arabic reviews from Tripadvisor.com and Booking.com	Naive Bayes (NB)	99%
IMDb movie	Decision Tree (DT)	75%
Chinese Language micro-blog	Random Forest (RF)	65%
Yelp restaurant reviews	Logistic Regression (LR)	79%
	Naive Bayes (NB)	72%
	Random Forest (RF)	76%
	Support Vector Machine (SVM)	78%
Amazon product reviews	Random Forest (RF)	91%

5 DISCUSSION

When we compare the results of our study to other related work, see Table 3, we may note that our spam detection model yields a high performance when compared to other models. Our best performing machine learning model would score fourth place in terms of accuracy in this list, with an accuracy like that was obtained in the two studies that use the same dataset that would be on a third and fifth place in terms of accuracy, respectively. It clearly outperforms other studies in terms of accuracy and, since accuracy is one of the most important performance indicators, could therefore logically serve and practically be applied in any defence strategy that an organisation might want to define in order to be able to tackle fake reviews.

Our study contributes to the research community by providing another successful example of how fake reviews can be detected.

Although fake reviews are a rising concern and a hot topic in the machine learning domain, unfortunately, not many datasets in which fake reviews have been identified are accessible. Obviously, in supervised learning, a large, diverse dataset is needed for proper training of classifiers in different application domains.

The scarcity of labelled datasets forms a real challenge for further research in the field. It can be recommended to synthesize and produce a new labelled dataset to bring more variety into the domain of spam detection. Currently, many models are being built upon the same sort of data, and, therefore, it will be valuable for future research to have different or more ample datasets to analyse.

6 CONCLUSIONS

Many organisations struggle with defending their online corporate reputations against fake reviews. It can be argued that positive and neutral fake reviews have, similar as is the case for negative fake reviews, negative consequences on corporate reputation. To provide organisations with an asset for corporate reputation management, a state-of-the-art machine learning model has been built that separates fake reviews from regular ones. The model yields a high accuracy rate compared to others, and, therefore, it can be said that this model could be successfully implemented by organisations as part of their corporate reputation management strategy. In the future, the model should be further optimized and extended to incorporate new datasets that are relevant for organisations by finetuning the processing steps that we have inpictured in this paper.

ACKNOWLEDGEMENTS

This paper has been inspired on the MSc master project of Zoë Kisoen who was involved via the master Digital Driven Business at HvA. Thanks go to Frederik Situmeang as well as several anonymous reviewers for providing some useful suggestions to an initial version of this manuscript. Rob Loke is assistant professor data science at CMIHvA.

REFERENCES

Abdulhamid, S.M., Abd Latiff, M.S., Chiroma, H., Osho, O., Abdul-Salaam, G., Abubakar, A.I., Herawan, T., 2017. A Review on Mobile SMS Spam Filtering Techniques. IEEE Access 5, 15650–15666. https://doi.org/10.1109/ACCESS.2017.2666785

Abimbola, T., Vallaster, C., 2007. Brand, organisational identity and reputation in SMEs: An overview. Qualitative Market Research: An International Journal 10, 341–348. https://doi.org/10.1108/13522750710819685

ACM, 2017. Eindrapportage ACM-verkenning naar online reviews: ‘Reviews gereviewd’. https://www.acm.nl/nl/publicaties/publicatie/17217/Eindrapportage-ACM-verkenning-naar-online-reviews-Reviews-gereviewd

Aula, P., 2010. (PDF) Social media, reputation risk and ambient publicity management [WWW Document]. ResearchGate. http://dx.doi.org/10.1108/10878571011088069

Barbado, R., Araque, O., Iglesias, C.A., 2019. A framework for fake review detection in online consumer electronics retailers. Information Processing & Management 56, 1234–1244. https://doi.org/10.1016/j.ipm.2019.03.002

Barnett, M., Jermier, J., Lafferty, B.A., 2006. Corporate Reputation: The Definitional Landscape. Corporate Reputation Review, 9, 26-38.

Beal, A., Strauss, J., 2009. Radically Transparent: Monitoring and Managing Reputations Online. John Wiley & Sons.

Bryant, D., 2019. How Chinese Sellers are Manipulating Amazon in 2019 [WWW Document]. URL https://www.ecomcrew.com/chinese-sellers-manipulating-amazon/ (accessed 6.20.20).

Carlisle, K., 2015. Fake Online Reviews are Bad for Business. 910 West. URL https://910west.com/2015/02/fake-online-reviews-bad-business/ (accessed 6.20.20).

Chandler, S., 2020. Coronavirus Drives 72% Rise In Use Of Fintech Apps [WWW Document]. URL https://www.forbes.com/sites/simonchandler/2020/03/30/coronavirus-drives-72-rise-in-use-of-fintech-apps/#68435c9066ed (accessed 6.17.20).

Chun, R., 2005. Corporate reputation: Meaning and measurement. International Journal of Management Reviews 7, 91–109.

Constantinides, E., & Fountain, S. J. (2008). Web 2.0: Conceptual foundations and marketing issues. Journal of direct, data and digital marketing practice, 9(3), 231-244.

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, 23. https://doi.org/10.1186/s40537-015-0029-9

Davies, G., Chun, R., Kamins, MA., 2010. Reputation gaps and the performance of service organizations. Strategic Management Journal 31, 530–546.

Dewey, J. (1910). How we think. Boston: D.C. Health and Company.

DigitalTrends (2018) Can you really trust app store ratings? We asked the experts. https://www.digitaltrends.com/android/can-you-really-trust-app-store-ratings/

DiMauro, V., & Bulmer, D. (2014). The Social Consumer Study. The Society for new Communications Research.

Dowling, G. R. (2016). Defining and measuring corporate reputations. European Management Review, 13(3), 207-223.

Dutch Institute of International Relations, S., n.d. Economic impact of COVID-19 [WWW Document]. Statistics Netherlands. URL https://www.cbs.nl/en-gb/dossier/coronavirus-crisis-cbs-figures/economic-impact-of-covid-19 (accessed 6.20.20).

Fombrun, C., van Riel, C., 1997. The Reputational Landscape. Corporate Reputation Review.

Fontanarava, J., Pasi, G., Viviani, M., 2017. Feature Analysis for Fake Review Detection through

Supervised Classification, in: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 658–666. https://doi.org/10.1109/DSAA.2017.51

Fournier, S., Avery, J., 2011. The uninvited brand. Business Horizons, SPECIAL ISSUE: SOCIAL MEDIA 54, 193–207. https://doi.org/10.1016/j.bushor.2011.01.001

Gensler, S., Völckner, F., Egger, M., Fischbach, K., Schoder, D., 2015. Listen to Your Customers: Insights into Brand Image Using Online Consumer-Generated Product Reviews. International Journal of Electronic Commerce, 1 20.

Gerdeman, D., 2020. How the Coronavirus Is Already Rewriting the Future of Business [WWW Document]. HBS Working Knowledge. URL http://hbswk.hbs.edu/item/how-the-coronavirus-is-already-rewriting-the-future-of-business (accessed 6.17.20).

Giovanni, S., 2010. Managing Knowledge Assets and Business Value Creation in Organizations: Measures and Dynamics: Measures and Dynamics. IGI Global.

Google (2019) In reviews we trust - making Google Play ratings and reviews more trustworthy. https://android-developers.googleblog.com/2018/12/in-reviews-we-trust-making-google-play.html

Gov UK, n.d. Online reviews: letting your customers see the true picture [WWW Document]. GOV.UK. URL https://www.gov.uk/government/publications/online-reviews-and-endorsements-advice-for-businesses/online-reviews-giving-consumers-the-full-picture (accessed 6.20.20).

Grützmacher, A., n.d. Reputation 2.0: The role of social media in corporate reputation - Case Nokia 160.

Harman, M., Jia, Y., Zhang, Y., 2012. App Store Mining and Analysis: MSR for App Stores.

Helm, S., Klode, C., 2011. Challenges in Measuring Corporate Reputation. pp. 99–110. https://doi.org/10.1007/978-3-642-19266-1_11

Horn, I., Taros, T., Dirkes, S., 2015. (PDF) Business Reputation and Social Media: A Primer on Threats and Responses [WWW Document]. ResearchGate. http://dx.doi.org/10.1057/dddmp.2015.1

Hoyer, WD., Macinnis, DJ., 2001. Consumer Behavior. 2nd ed., Boston: Houghton Mifflin.

Hussain, N., Mirza, H., Rasool, G., Hussain, I., Kaleem, M., 2019. Spam review detection techniques: a systematic literature review.

Hutton, J., Goodman, M., Alexander, J., Genest, C., 2001. Reputation management: the new face of corporate public relations? Public Relations Review 27, 247–261. https://doi.org/10.1016/S0363-8111(01)00085-6

Iansiti, M., Richards, G., 2020. Coronavirus Is Widening the Corporate Digital Divide. Harvard Business Review.

Jamie, 2019. Fake Reviews Are a Real Problem: 8 Statistics That Show Why [WWW Document]. BrightLocal. URL https://www.brightlocal.com/learn/fake-reviews-are-a-real-problem-8-statistics-that-show-why/ (accessed 6.20.20).

Jankauskaite, D., Urboniene, A., 2016. Organization's Reputation Management Through Content Creation And Sharing In The Social Media 3, 35.

Jones, B., Temperley, J., Lima, A., 2010. Corporate Reputation in the Era of Web 2.0: The Case of Primark. Journal of Marketing Management November 2009, 927–939. https://doi.org/10.1362/026725709X479309

Kaggle, 2020. Find Open Datasets and Machine Learning Projects | Kaggle [WWW Document]. URL https://www.kaggle.com/datasets?sortBy=relevance&group=public&search=spam&page=1&pageSize=20&size=sizeAll&filetype=fileTypeAll&license=licenseAll (accessed 6.19.20).

Kaggle Reviews, 2020. Rome wasn't built in a day: spotting fake reviews [WWW Document]. URL https://kaggle.com/nicodds/rome-wasn-t-built-in-a-day-spotting-fake-reviews (accessed 6.17.20).

Kaplan, A., Haenlein, M., 2010. Users of the World, Unite! The Challenges and Opportunities of Social Media. Business Horizons 53, 59–68. https://doi.org/10.1016/j.bushor.2009.09.003

Li, L.-Y., Qin, B., Liu, T., 2018. Survey on Fake Review Detection Research. Jisuanji Xuebao/Chinese Journal of Computers 41, 946–968. https://doi.org/10.11897/SP.J.1016.2018.00946

Lincoln, J., n.d. As Coronavirus Spreads, Digital Marketing Becomes More Important Than Ever [WWW Document]. Business 2 Community. URL https://www.business2community.com/small-business/as-coronavirus-spreads-digital-marketing-becomes-more-important-than-ever-02296683 (accessed 6.17.20).

Martens, D., Maalej, W., 2019. Towards understanding and detecting fake reviews in app stores. Empir Software Eng 24, 3316–3355. https://doi.org/10.1007/s10664-019-09706-9

McGrath, J., Ross, T., 2020. Corporate reputation and the coronavirus. Ipsos.

McKinsey, 2020. Customer experience: new capabilities, new audiences, new opportunities. McKinsey & Company.

Mishra A, Satish SM., 2016. eWOM: Extant Research Review and Future Research Avenues. Vikalpa 41(3):222-233. doi:10.1177/0256090916650952

Otar, C., 2018. https://www.forbes.com/sites/forbesfinancecouncil/2018/10/05/how-review-sites-can-affect-your-business-and-what-you-can-do-about-it/

Page, G., & Fearn, H. (2005). Corporate reputation: what do consumers really care about? Journal of Advertising Research, 45(3), 305-313.

Siano, A., Vollero, A., Confetto, M., Siglioccolo, M., 2013. Corporate communication management: A framework based on decision-making with reference to communication resources. Journal of Marketing Communications 19. https://doi.org/10.1080/13527266.2011.581301

Statista, 2020. Amazon fake product review categories 2018 l Statistic [WWW Document]. Statista. URL https://www.statista.com/statistics/997026/amazon-sh

opping-categories-largest-share-fake-product-reviews/ (accessed 6.20.20).
The New York Times, 2020. https://www.nytimes.com/2020/02/28/business/economy/companies-coronavirus-economy.html
Walker, K. (2010). A systematic review of the corporate reputation literature: Definition, measurement, and theory. Corporate reputation review, 12(4), 357-387.
Wartick, S., 2002. Measuring Corporate Reputation Definition and Data. Business & Society 41, 371–392. https://doi.org/10.1177/0007650302238774
Waters, R.D., Burnett, E., Lamm, A., Lucas, J., 2009. Engaging stakeholders through social networking: How nonprofit organizations are using Facebook. Public Relations Review 35, 102–106. https://doi.org/10.1016/j.pubrev.2009.01.006
Weber Shandwick, KRC Research, 2019. The state of corporate reputation in 2020: everything matters now.
Xhema, J., 2019. Effect of Social Networks on Consumer Behaviour: Complex Buying. IFAC-PapersOnLine, 19th IFAC Conference on Technology, Culture and International Stability TECIS 2019 52, 504–508. https://doi.org/10.1016/j.ifacol.2019.12.594