data

MDPI

Article

Using Social Media to Detect Fake News Information Related to Product Marketing: The FakeAds Corpus

Noha Alnazzawi [1,*], Najlaa Alsaedi [2], Fahad Alharbi [3] and Najla Alaswad [4]

[1] Computer Science and Engineering Department, Yanbu University College, Royal Commission for Jubail and Yanbu, Yanbu Industrial City 41912, Saudi Arabia
[2] Computer Science Department, King Abdul Aziz University, Jeddah 21589, Saudi Arabia; nalsaedi0024@stu.kau.edu.sa
[3] Data Management Specialist, Ministry of Interior, Public Security, Riyadh 12732, Saudi Arabia; fahadaljedia@gmail.com
[4] Data Analyst Specialist, Princess Norah University, Riyadh 11671, Saudi Arabia; najlaAlaswad@my.com
* Correspondence: alnazzawin@rcyci.edu.sa

Abstract: Nowadays, an increasing portion of our lives is spent interacting online through social media platforms, thanks to the widespread adoption of the latest technology and the proliferation of smartphones. Obtaining news from social media platforms is fast, easy, and less expensive compared with other traditional media platforms, e.g., television and newspapers. Therefore, social media is now being exploited to disseminate fake news and false information. This research aims to build the FakeAds corpus, which consists of tweets for product advertisements. The aim of the FakeAds corpus is to study the impact of fake news and false information in advertising and marketing materials for specific products and which types of products (i.e., cosmetics, health, fashion, or electronics) are targeted most on Twitter to draw the attention of consumers. The corpus is unique and novel, in terms of the very specific topic (i.e., the role of Twitter in disseminating fake news related to production promotion and advertisement) and also in terms of its fine-grained annotations. The annotation guidelines were designed with guidance by a domain expert, and the annotation is performed by two domain experts, resulting in a high-quality annotation, with agreement rate F-scores as high as 0.815.

Keywords: social media; fake news; corpus construction; text mining

Citation: Alnazzawi, N.; Alsaedi, N.; Alharbi, F.; Alaswad, N. Using Social Media to Detect Fake News Information Related to Product Marketing: The FakeAds Corpus. *Data* **2022**, *7*, 44. https://doi.org/10.3390/data7040044

Academic Editors: Gianni Costa and Riccardo Ortale

Received: 25 February 2022
Accepted: 30 March 2022
Published: 7 April 2022

1. Introduction

Social media is a very fast and easy-to-access channel that disseminates news and, every second of the day, huge numbers of people are accessing and interacting with online news [1]. Over the last decade, social media channels, including Twitter, Facebook, YouTube, and Instagram, have become an integral part of our daily lives [2].

As an increasing amount of our time is spent interacting online through social media platforms, more and more people tend to seek out and consume news from social media sources, rather than traditional news organizations. Twitter is a very popular social media platform and its number of users has been growing rapidly since its creation in 2006. Today, it represents a very important and widely used source for news dissemination and also for marketing and promoting new products. For example, 62 percent of U.S. adults got their news from social media in 2016, while in 2012, only 49 percent reported reading the news on social media [3]. In recent years, Twitter has provided a panel where people can interact with each other and maintain social ties. People use Twitter to share their daily activities, happenings, thoughts and feelings with their contacts, which makes twitter both a valuable data source and a great target for various areas of research and practice. According to a report published in 2021 [4], Twitter has 340 million users and delivers 500 million tweets a day, 200 billion tweets a year [4,5].

It is fast and easy to use social media to obtain the latest news or to see advertisements for different products [6]. Any news spreads much faster via social media, no matter where in the world an event takes place [1]. However, despite the advantages provided by social media platforms, the credibility and quality of news on social media is lower than traditional news channels, including TV, newspapers, and other trusted news sources, due to the freedom afforded to social media channels in expressing (false) ideas and circulating (fake) news and (misleading) adverts [1]. Therefore, social media enables the wide and rapid dissemination of "fake news", i.e., low-quality news, which contains intentionally false information. Although the survey report [7] found that almost 60% of users expect news on social media to be inaccurate, it still leaves millions of people who will spread (retweet) fake news believing it to be true.

Twitter is widely used to spread false information promoting products and brands. For example, in the United States alone, 60 percent of adults who depend on social media for news consumption also share false information [5,8]. Individuals receive advertisements on social media based on their interests and consciousness about the facts and the content mentioned in the circulated advertisements. Around 54% of people around the globe have expressed their concerns about fake news [1]. Additionally, the younger generation is more heavily influenced by online-based news than older generations. This, in turn, results in the quick dissemination of news to millions and billions of people [9]. Additionally, online advertisements for products tend to target the younger generations and try to promote products relevant to their lifestyles, such as skincare products and technological gadgets, in eye-catching ways, to reach as many people as possible around the globe [1].

The widespread dissemination of false information has the potential to have an extremely negative impact on individuals and society [10]. For example, in 2008, a false report about the United Airlines parent company's bankruptcy caused the company's stock price to drop by 76%. Twitter [11,12] has been widely used to spread fake and biased news during the last two U.S. presidential election periods [13]. Following the last presidential election, it was estimated that over 1 million tweets were related to fake news "Pizzagate". Thus, the word "Fake news" was even named the word of the year by the Macquarie dictionary in 2016 [3].

Fake news is created for a variety of reasons, but mainly for financial and political gain [3]. For example, as most of the fake news is spread by propagandists, it usually conveys influential messages and persuades individuals, in different ways, to accept biased or false information [3,14,15]. From marketing perspectives, fake news also presents false information to promote a specific idea or product. If spread with malicious intent, fake news can be used by a competitor to damage the reputation of a specific brand or company.

According to Twitter's policy, a warning tag will be applied to any tweet containing disputed or misleading information related to COVID-19 that goes directly against guidance on COVID-19 from authoritative sources. However, Twitter is still working on the public conversation to make sure that credible and authentic information is available to the user [16].

Therefore, fake news detection on social media in general, and Twitter in particular, has recently become an emerging research topic that is attracting tremendous attention [3]. Although considerable effort has been made towards fake news detection on websites and news articles, very little effort has been put in to explore Twitter and, to the best of our knowledge, no prior work has focused on the influence of fake news and false information on marketing and promoting products, solely focused on Twitter, in order to tackle the rise and spread of fake news and to enhance the automatic detection of fake news and false information on this specific social media platform. To facilitate research into fake news detection on Twitter about misleading advertisements for differing products that target the consumer, and to help mitigate the negative effects caused by fake news—both to benefit consumers and the news ecosystem—it is critical that we develop methods to automatically detect fake news on social media. Machine Learning (ML)-based text mining (TM) tools have the potential to automatically detect fake news and false information related

to product marketing. However, developing TM tools is reliant upon textual corpora, in which pertinent information is marked up by expert annotators. Such annotated corpora serve both as training datasets for ML-based Named Entity Recognition methods and as a gold standard for the systematic evaluation of new methodologies.

This research aims to explore how Twitter is used to disseminate false marketing information through deliberately misleading/fake adverts; the contribution of this research is threefold:

1. To use Twitter as a social media resource to explore the use of fake and false news to promote products.
2. To build annotated datasets for fake and real advertisements related to cosmetics, fashion, health and technology products.
3. The corpus is freely available to stimulate the development of ML-based Text Mining (TM) systems for the automatic extraction and classifications of details, relating to fake news intended to mislead the consumer by promoting false products. The developed TM systems can ultimately be a useful data resource for the research community to further the study of social media credibility, in promoting products and circulating fake advertisements.

2. Related Work

Most of the previous work tackles the problem of detecting fake news using textual sources from news articles. Fake news detection on social media in general, and Twitter in particular, has recently become an emerging research topic that is attracting considerable attention [3]. However, fake news detection is a very challenging task, as the purpose of the distribution of such news and events is to deliberately mislead people [17].

The previous studies on fake news detection in social media focused on different topics, including: bot detection [5,18,19], predicting spammer behavior and detecting these spammers [20–25], tweets related to natural disasters that were fake, spam and legitimate [26] clickbait detection [27]. Although there are many studies focused on spam detection problems related to online reviews for products [28–31], to the best of our knowledge, there is no prior work devoted to the study of fake advertisement marketing products on social media (i.e., Twitter in particular).

Supervised ML methods are widely used to identify fake news and most of the existing work is based on supervised ML methods, by formulating fake news detection as a binary classification problem, the main concern of this approach is to find effective features for training and evaluating the ML models [32]. Supervised ML-based approaches require a reliable pre-annotated dataset to train a classification model on a set of features that help the model to recognize and correctly classify the information in the unseen dataset [3,32]. The features that have been used by fake news detection ML algorithms generally fall into two categories: news contents and social context [33]. The content-based approaches usually rely on using features, such as linguistics-based [34] and visual-based features [35,36], while social context approaches incorporate features from users' profiles [37], posts content and social networks [3,32]. However, supervised ML algorithms are strongly dependent on domain knowledge for designing features, which makes the method difficult to generalize to new tasks [38]. To the best of our knowledge, no prior work has focused on the influence of fake news on marketing products by giving false information and using false advertisements, which makes it difficult to adapt ML algorithms that are trained in different domains.

Fake news detection on social media presents unique challenges that make existing detection algorithms from traditional news media ineffective or not applicable [3]. The lack of manually labeled fake news datasets for text drawn from social media (i.e., Twitter) limits the advancement of ML-based approaches that could automatically detect fake news in social media. Examples of the publicly available dataset are LIAR [3], BuzzFeed News [34,39], CREDBANK [40]. All the mentioned datasets include text drawn from news websites, only the text in the CREDBANK dataset was drawn from Twitter. Recently,

most of the work on fact checking comes in the form of shared tasks, e.g., CheckThat 2021 Cross-Language Evaluation Forum (CLEF) [41,42]. The shared task consists of various tasks related to fact checking about tweets related to COVID-19 and predicting the veracity of a news article and its topics (i.e., health, election, crime, climate, economy and education) [43]. Table 1 shows the comparison of the characteristics of the popular datasets, in terms of size, text genre, topic and annotation level.

Table 1. Comparison of the characteristics of some of the popular fake news datasets.

Dataset	Size	Text Genre	Topic	Categories	Annotation Level
BuzzFeed [39]	2283	News articles	Presidential election, political biases: mainstream, left-leaning, and right-leaning.	Mostly true, mostly false, mixture of true and false, and no factual content	Sentence level
LIAR [3]	12.8 K	Sentences collected from politifact	politics	false, barely true, half true, mostly true, and true	Sentence level
CREDBANK [40]	60 M	Tweets	Real world events	-Certainly inaccurate -Probably inaccurate -Uncertain -Probably accurate -Certainly accurate	Sentence level
FaceBookHoax [44]	15,500	Facebook posts	Scientific news sources vs. conspiracy news sources	Hoax, no-hoax	Sentence level
CheckThat sub-task 1: check-worthiness estimation on Twitter CT–CWT–21 [41]	1312	Tweets	COVID-19 and politics	-Not worth fact checking -Worth fact checking	Sentence level
CheckThat sub-task 3: (CT-FAN-21) Multi-class fake news categorization of news articles [42]	1254	News articles	Health, climate, economy, crime, elections, and education	False, partially false, true, and other.	Sentence level

However, all the mentioned datasets are annotated at sentence level, to either classify the text as fake or real, or relate to the credibility level. Moreover, no prior dataset was dedicated to the impact of fake news and false advertisements on marketing and promoting products. Fake news is false information and facts disseminated with the main intention of deceiving the reader [45]. The term 'fake news' is often described in related literature using different terms, including 'misinformation', 'disinformation', 'hoax', and 'rumor', which are actually different variations of false information. Most of the previous works on fact checking and fake news detection [46] examine the problem from the angle of a veracity classification. However, there is no system that can automatically and completely stop the dissemination of fake news in social media and the consequent negative impact of the fake news on society without the involvement of humans [46]. Classical ML approaches can be applied to automatically extract fake news information, given that they have similar cases in the training dataset [9,17]. However, the development of text-mining tools depends on the availability of an annotated corpus.

In this research, we built a new corpus, named FakeAds. The corpus is collected from tweets for the topic of fake news in the marketing domain. The corpus is unique and novel, in terms of the very specific topic in the fake news domain (i.e., knowing how fake news

influences marketing) and the fine-grained annotation provided at word level to classify each product into one of the following classes: fashion, cosmetics, health, and electronics.

3. Results and Discussion

To ensure that the generated corpus was of high quality, the annotations provided in the corpus closely followed the guidelines set by the annotators, who were English native speakers and experts in the field of annotation. We calculated the Inter Annotator Agreement (IAA) between the two annotators and a high IAA score provided assurance that the corpus annotations were reliable and of high quality.

We followed a number of other related studies [47–49] by calculating the IAA in terms of F-score. The F-score is the same whichever set of annotations is used as the gold standard [49,50]. To carry out such calculations, the set of annotations produced by one of the annotators was considered the 'gold standard', i.e., the set of correct annotations and the total number of correct entities was the total number of entities annotated by this annotator.

In this study, the annotations produced by the first annotator were considered as the 'gold standard', i.e., the set of correct annotations, and the total number of correct entities was the total number of entities annotated by this annotator. Based on the gold standard, the Inter Annotator Agreements (IAA), by means of precision, recall and F-score, were calculated. Precision (P) refers to the percentage of the correct positive annotated entities annotated by the second annotator in comparison to the annotation produced by the first annotator, which was assumed to be the gold standard. The precision was calculated as the ratio between the true positive (TP) entities and the total number of entities annotated by the second annotator (the sum of true positives (TPs) and false positives (FPs)), according to the following formula:

$$P = TP/TP + FP.$$

Recall (R) is the percentage of positive annotated entities recognized by the second annotator. It is calculated as the ratio between the TP and the total number of annotations in the gold standard, according to the following formula:

$$R = TP/TP + FN.$$

The F-score is the harmonic mean of precision and recall and is calculated according to the following formula:

$$\text{F-score} = 2 * (\text{Precision} \times \text{Recall})/\text{Precision} + \text{Recall}.$$

Table 2 shows the statistics and the IAA for the annotation of product types in the FakeAds corpus. Overall, the annotators agreed most of the time on annotating cosmetics and health products and the F-scores for these two classes were the highest, at 0.94 and 0.86, respectively. The reason for this high score is because the mentions and examples of cosmetics and health were very straightforward, and the annotators could easily recognize and classify the mentions. On the other hand, the F-scores for fashion and electronic products were generally lower than those for cosmetics and health because the number of tweets for electronics and fashion products were the fewest in the corpus compared with the number of examples of cosmetics and health products. In addition, there were a greater number of disagreements between the annotators, with regard to which type of products belonged to these two classes. For example, the second annotator annotated general words, e.g., clothes, bags, jumpers, etc., and this contributed to the low precision, especially for the fashion class, where the second annotator annotated irrelevant products as fashion (i.e., annotating very general descriptions of a fragrance instead of mentioning specific products e.g., luxurious scents). It was noticed that the low recall for electronics products was because the second annotator did not annotate every mention of electronic products and did not annotate broad coverage of electronics devices. For example, he did

not correctly annotate electronic devices related to skincare and healthcare, such as skincare device and airbrush.

Table 2. Comparison of FakeAds corpus with other comparable corpora.

Dataset	Size	Text Genre	Topic	Categories/Labels	Annotation Level	# of Annotators	Agreement Measurement
FakeAds	5000	Tweets	Marketing and fake news	Binary classes: Fake Real Multi-classes: Health Cosmetics Fashion Electronic	– Tweet level – Mention level	3 annotators for the binary annotation 2 annotators for the multi-class annotation	F-score (0.815)
CREDBANK	60 M	Tweets	Real world vents	– Certainly inaccurate – Probably inaccurate – Uncertain – Probably accurate – Certainly accurate	Sentence level	1736 unique annotators from AMT	Intraclass correlations (ICC) (0.77)
CheckThat sub-task 1: check-worthiness estimation on Twitter	1312	Tweets	COVID-19 and politics	– Not worth fact checking – Worth fact checking	Sentence level	3 annotators	Majority voting Averaged (0.597)

To show the importance of our generated dataset, we compared the FakeAds corpus with other publicly available datasets in the fake news detection domain, which are reported in Table 1. In particular, we compared our dataset with CREDBANK [24] and CheckThat sub-Task 1: check worthiness [26] on twitter datasets, because they used Twitter as a textual source and they share some of the characteristics with the FakeAds corpus, e.g., they are annotated for similar classes related to reporting false or uncertain information. As shown in Table 2, the FakeAds corpus differs from the existing datasets in terms of the very specific domain, which is false advertisement to promote products, and also the rich annotations at two levels of annotation at tweet level, where the tweet is classified as real or fake, and at mention level, where the product mention is given one of the following classes: health, cosmetics, fashion or electronics. This makes it a valuable resource for training and evaluating ML-based techniques. The results of the annotation are satisfactory and are measured in terms of F-score at 0.815.

4. Materials and Methods

Corpus Construction

The FakeAds corpus consists of tweets that were collected from Twitter using the TweetScraper tool [51] for the period between 1 January 2015 and 30 December 2020. We targeted this particular five-year span as product marketing through social media was very common during this period. The following list of keywords was used to collect the relevant tweets: marketing, advertisement, digitalMarketing, socialmediaMarketing and onlinePromotion. We used the hashtagify tool [52] to find highly ranked, trending and popular hashtags, and also to find hashtags highly related to marketing and advertising. We found that the used search keywords represent hashtags ranked by hashtagify to be highly related to marketing hashtags. The tweets were further filtered by the annotators of this task who are English instructors, and only tweets that include information directly related to our task in question were retained, resulting in 5000 tweets. Manual inspection of the collected tweets revealed that the products that are discussed in the tweets generally belong to one of the following broad categories: cosmetics, health, fashion, and electronics. Thus, these categories were used as the classes for the products in the FakeAds corpus.

The tweets were annotated at two levels:

1. At tweet level so that tweets were annotated as fake or real.
2. At word level so that for each tweet, the product was classified into one of the following classes: cosmetics, health, fashion, and electronics.

In the tweet-level annotation task, the tweets were annotated as either fake or real. This annotation task is considered binary classification and we used the Amazon Mechanical Turk (AMT) tool to annotate the tweets. AMT is a crowdsourcing marketplace introduced by Amazon and which is becoming increasingly popular as an annotation tool for NLP research including: word sense disambiguation, word similarity, text entailment, and temporal ordering [53,54]. To ensure the quality of annotations produced by AMT we applied the country and high acceptance rate crowd filters so that only annotators with a 95% success rate on previous AMT Human Intelligence Tasks (HITs) and restricted to those who were located in the United States were accepted for the task. The reason to choose these two filters was because it lowered the pool of workers and it has been shown to be effective in reducing incidents of spamming and cheating found in previous studies [55,56]. The same set of annotation guidelines was shared/used by the annotators to ensure high-quality and reliable annotations. As per the guidelines, the annotators need to consider two factors before deciding if a tweet is fake or real: the account-related features (e.g., the profile information such as number of followers and following users) and the tweet's related features (e.g., lexical and syntactical features of the tweet) [57].

Each of the 5000 tweets in our corpus was annotated by three workers, resulting in $5000 \times 3 = 15{,}000$ annotations in total. For each tweet the majority given class was chosen and hence the tweet was given that label: fake or real. In total, we collected 5000 tweets, out of which 2914 (0.5828) were labeled as real news while 2086 (0.4172%) were labeled as fake news. Figure 1 shows the distribution of tweets that contained either fake or accurate content in the FakeAds corpus. It has been noted that while 41% of the tweets in FakeAds were annotated to be fake, distributing real information related to product promotion still represents a higher percentage of product advertisements. This was something we expected as the Twitter platform is used by many trustworthy organizations to disseminate real adverts and factual information.

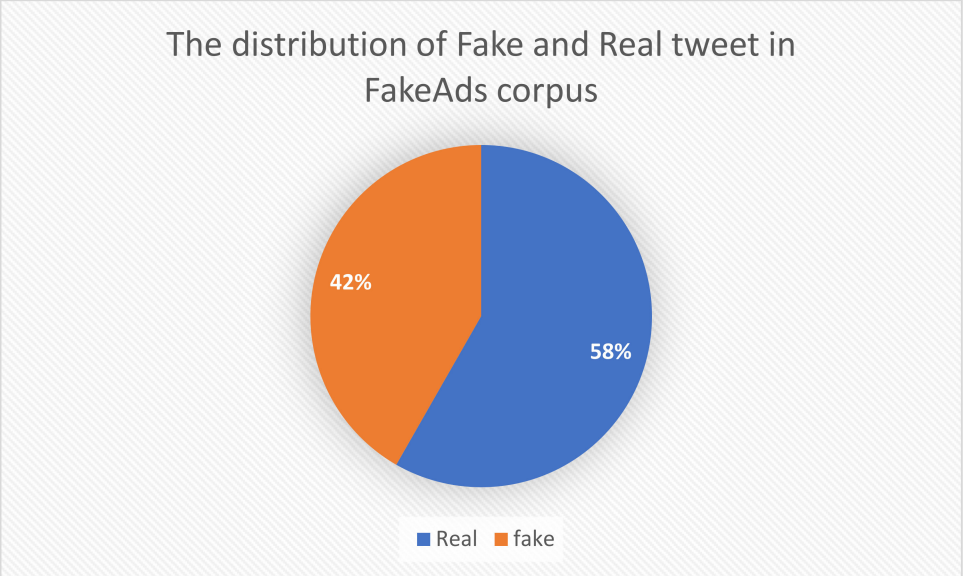

Figure 1. The distribution of fake and real tweets in the FakeAds corpus.

However, for the multi-class annotation task the tweets were annotated at word level, which denotes mentions for products including the following classes: cosmetic, fashion, health and electronics. The annotation was done through the COGITO service where

each tweet was annotated by two annotators for the mentions of the product type. Each tweet was annotated by two annotators for the entity types related to the product types by using the same set of annotation guidelines provided in Supplementary Materials File S1. The annotation included marking up all entity mentions in the corpus related to the four semantic types mentioned in Table 3.

Table 3. Annotated entity classes in the FakeAds corpus.

Entity Type	Description
Cosmetic	Is product mention related to skincare, body care or make-up, for example, lipsticks, creams etc.
Electronic	Is products that require electric currents or electromagnetic fields to work. Examples are electronic devices, phones, cameras, computers etc.
Health	Is product mention related to supplement(s) that promotes the wellbeing of individuals, e.g., vitamins, herbs, etc.
Fashion	Is product related to accessories such as clothing, shoes, bags jewelry, fragrances, etc.

Figure 2 describes the most common product types in the corpus and their distribution in the FakeAds corpus. As shown in Figure 2, there was considerable emphasis on Twitter in promoting cosmetic products e.g., skincare, makeup, etc. The cosmetic class represents 83% of the annotations in the FakeAds corpus, health-related products come next after the cosmetic products with 10% of the total annotation in the FakeAds corpus. The less-dominant and lesser-targeted products in advertisements on Twitter and in the FakeAds corpus in particular were electronics and fashion. It was also noted that people on Twitter tended to discuss fashion and electronics products less frequently in the context of advertising when compared with cosmetics. Table 4 summarizes the statistics of the corpus, and it shows the total number of fake and real annotations and the distribution for the different products among fame and real tweets. Figure 3 visualizes the distribution of products that are targeted most by fake news and false information, which are cosmetic and health products compared to real information for these two types of products. On the other hand, it is worth mentioning that the number of real news information related to electronic and fashion products is significantly higher than the number of fake news that targets these two types of products. This is because online advertisements for products in social media platforms tend to target the younger generations and try to promote products relevant to their lifestyles, such as skincare products and different supplements that match their lifestyles.

Table 4. Statistics of the FakeAds corpus annotations.

	Total Number of Annotations	Health	Cosematics	Fashion	Electronics
Real	6159	300	5691	135	33
Fake	4807	200	4527	66	14

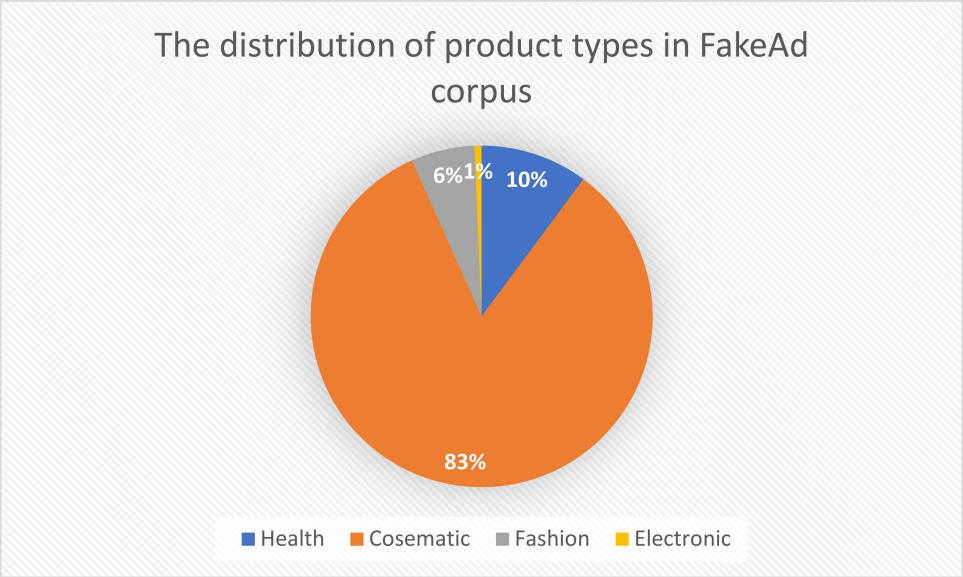

Figure 2. The distribution of product types in FakeAds corpus.

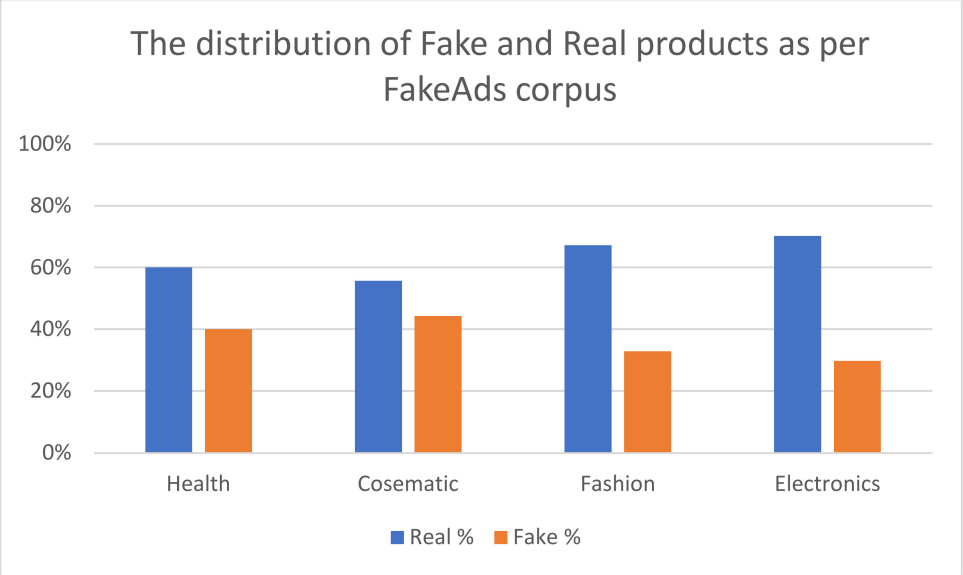

Figure 3. The distribution of fake and real products in FakeAds corpus.

5. Conclusions

Our central goal in this paper was to provide the research community with a dataset that could serve the study of fake news detection on Twitter that targets information that misleads the consumer by falsely promoting products. The corpus consists of 5000 tweets, annotated at two levels: (1) each tweet is annotated as fake or real, (2) each tweet is annotated at word level. This is to classify the product into one of the following classes: cosmetics, health, fashion, or electronics. We envision that this will be a useful data resource for the community to further the study of social media credibility in promoting products and circulating fake advertisements. The proposed research could also provide a broader view about fake news related to marketing and help to enhance Twitter's policy to provide more credible and authentic information related to promoting products. It will also help to give an idea about which types of products are targeted more by propagandists to distribute fake news in an attempt to attract more consumers. The generated corpus can serve as a gold standard for the development and evaluation of TM tools that can classify each

tweet as real or fake and extract product mentions, related to cosmetics, health, fashion and electronics. For example, in the future, we are planning to use classical ML-based NER and compare them with state-of-the-art contextual word embedding (e.g., BERT) on the FakeAds corpus to automatically classify tweets as fake or real, and also extract the product type discussed in the tweets.

6. Limitations and Future Works

Despite the contributions presented earlier, we acknowledge certain limitations. This work's main limitation lies in the product classes. While we tried to make sure that the product categories (i.e., classes) were broad enough to include all the products mentioned in the FakeAds corpus, the categories used may be not broad enough to include products that were not mentioned in the FakeAds corpus; for example, products related to sports equipment, furniture, cars, etc. Another potential limitation is the size of the corpus due to the cost of the manual annotation, in terms of time and money, so we were only able to annotate 5000 tweets. However, the size of the corpus is comparable to popular fake news datasets mentioned in Table 1 and exceeds the size of some datasets, e.g., the BuzzFeed and CheckThat datasets. ML-based text mining methods require a large dataset for accurate models to be trained and tested and, hence, increasing the size of the corpus would give more accurate results for training and evaluating ML models. In the future, we are planning to use the generated corpus as a gold standard for the development and evaluation of TM tools and we are also planning to broaden and increase the range of the corpus by including more product classes, and also by including text from other social media platforms, e.g., Facebook.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/data7040044/s1, File S1: The annotation guidelines.

Author Contributions: Conceptualization, N.A. (Noha Alnazzawi) and N.A. (Najlaa Alasaedi); methodology, N.A. (Noha Alnazzawi); validation, N.A. (Noha Alnazzawi), N.A. (Najlaa Alasaedi) and F.A.; formal analysis, N.A. (Noha Alnazzawi); investigation, N.A. (Noha Alnazzawi); resources, N.A. (Noha Alnazzawi); data curation, N.A. (Noha Alnazzawi); writing—original draft preparation, N.A. (Noha Alnazzawi); writing—review and editing, F.A. and N.A. (Najlaa Alaswad); supervision, N.A. (Noha Alnazzawi); project administration, N.A. (Noha Alnazzawi) All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is freely available on Kaggle website https://www.kaggle.com/datasets/nohaalnazzawi/the-fakeads-corpus, accessed on 24 February 2022.

Acknowledgments: The authors extend their appreciation to the Saudi Society for Data Science for supporting this work through Research Group no. RGP-13.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Meel, P.; Vishwakarma, D.K. Fake news, rumor, information pollution in social media and web: A contemporary survey of state-of-the-arts, challenges and opportunities. *Expert Syst. Appl.* **2020**, *153*, 112986. [CrossRef]
2. Wang, W.; Chen, L.; Thirunarayan, K.; Sheth, A.P. Cursing in English on Twitter. In Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing, Baltimore, MD, USA, 15–19 February 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 415–425.
3. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explor. Newsl.* **2017**, *19*, 22–36. [CrossRef]
4. Aslam, S. *Twitter by the Numbers: Stats, Demographics & Fun Facts*; Omnicore: San Francisco, CA, USA, 2018.
5. Khan, T.; Michalas, A.; Akhunzada, A. Fake news outbreak 2021: Can we stop the viral spread? *J. Netw. Comput. Appl.* **2021**, *190*, 103112. [CrossRef]

6. Aldwairi, M.; Alwahedi, A. Detecting fake news in social media networks. *Procedia Comput. Sci.* **2018**, *141*, 215–222. [CrossRef]
7. Martin, N. How Social Media Has Changed How We Consume News. Forbes. Available online: https://www.forbes.com/sites/nicolemartin1/2018/11/30/how-social-media-has-changed-how-we-consume-news/?sh=40c30d723c3c (accessed on 20 February 2022).
8. Wong, Q. Fake News Is Thriving Thanks to Social Media Users, Study Finds. CNET. Available online: https://www.cnet.com/tech/social-media/fake-news-more-likely-to-spread-on-social-media-study-finds/ (accessed on 20 February 2022).
9. Nasir, J.A.; Khan, O.S.; Varlamis, I. Fake news detection: A hybrid CNN-RNN based deep learning approach. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100007. [CrossRef]
10. Aslam, N.; Ullah Khan, I.; Alotaibi, F.S.; Aldaej, L.A.; Aldubaikil, A.K. Fake detect: A deep learning ensemble model for fake news detection. *Complexity* **2021**, *2021*, 5557784. [CrossRef]
11. Murayama, T.; Wakamiya, S.; Aramaki, E.; Kobayashi, R. Modeling the spread of fake news on Twitter. *PLoS ONE* **2021**, *16*, e0250419. [CrossRef]
12. Carvalho, C.; Klagge, N.; Moench, E. The persistent effects of a false news shock. *J. Empir. Financ.* **2011**, *18*, 597–615. [CrossRef]
13. Bovet, A.; Makse, H.A. Influence of fake news in Twitter during the 2016 US presidential election. *Nat. Commun.* **2019**, *10*, 7. [CrossRef]
14. Shu, K.; Wang, S.; Liu, H. Exploiting Tri-Relationship for Fake News Detection. In Proceedings of the 12th ACM International Conference on Web Search and Data Mining (WSDM 2019), Ithaca, NY, USA, 20 December 2017; Cornell University: Ithaca, NY, USA, 2017.
15. Klein, D.O.; Wueller, J.R. Fake news: A legal perspective. *Australas. Polic.* **2018**, *10*, 11.
16. Roth, Y.; Pickles, N. Updating Our Approach to Misleading Information. Twitter Blog. Available online: https://blog.twitter.com/en_us/topics/product/2020/updating-our-approach-to-misleading-information (accessed on 20 February 2022).
17. Kaliyar, R.K.; Goswami, A.; Narang, P. FakeBERT: Fake news detection in social media with a BERT-based deep learning approach. *Multimed. Tools Appl.* **2021**, *80*, 11765–11788. [CrossRef]
18. Cresci, S.; Di Pietro, R.; Petrocchi, M.; Spognardi, A.; Tesconi, M. The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race. In Proceedings of the 26th International Conference on World Wide Web Companion, Perth, Australia, 3–7 May 2017; International World Wide Web Conferences Steering Committee: Geneva, Switzerland, 2017; pp. 963–972.
19. Gibert, D.; Mateu, C.; Planes, J. The rise of machine learning for detection and classification of malware: Research developments, trends and challenges. *J. Netw. Comput. Appl.* **2020**, *153*, 102526. [CrossRef]
20. Bakhteev, O.; Ogaltsov, A.; Ostroukhov, P. Fake News Spreader Detection Using Neural Tweet Aggregation. In Proceedings of the CLEF 2020 Labs and Workshops, Notebook Papers, Thessaloniki, Greece, 22–25 September 2020; Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A., Eds.; CEUR: Uzhhorod, Ukraine, 2020.
21. Lee, K.; Caverlee, J.; Webb, S. Uncovering Social Spammers: Social Honeypots+ Machine Learning. In Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Geneva, Switzerland, 13–19 July 2010; ACM: New York, NY, USA, 2010; pp. 435–442.
22. Ghosh, S.; Korlam, G.; Ganguly, N. Spammers' Networks within Online Social Networks: A Case-Study on Twitter. In Proceedings of the 20th International Conference Companion on World Wide Web, Hyderabad, India, 28 March–1 April 2011; ACM: New York, NY, USA, 2011; pp. 41–42.
23. Wang, A.H. Don't Follow Me: Spam Detection in Twitter. In Proceedings of the 2010 International Conference on Security and Cryptography (SECRYPT), Athens, Greece, 26–28 July 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 1–10.
24. Stringhini, G.; Kruegel, C.; Vigna, G. Detecting spammers on social networks. In Proceedings of the 26th Annual Computer Security Applications Conference, Austin, TX, USA, 6–10 December 2010; ACM: New York, NY, USA, 2010; pp. 1–9.
25. Yardi, S.; Romero, D.; Schoenebeck, G. Detecting spam in a Twitter network. *First Monday* **2010**, *15*. [CrossRef]
26. Rajdev, M.; Lee, K. Fake and spam messages: Detecting misinformation during natural disasters on social media. In Proceedings of the 2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), Singapore, 6–9 December 2015; IEEE: Piscataway, NJ, USA, 2015; pp. 17–20.
27. Potthast, M.; Köpsel, S.; Stein, B.; Hagen, M. Clickbait detection. In *Advances in Information Retrieval*; Springer International Publishing: Cham, Switzerland, 2016; pp. 810–817.
28. Ott, M.; Cardie, C.; Hancock, J. Estimating the prevalence of deception in online review communities. In Proceedings of the 21st International Conference on World Wide Web, Lyon, France, 16–20 April 2012; ACM: New York, NY, USA, 2012; pp. 201–210.
29. Danescu-Niculescu-Mizil, C.; Kossinets, G.; Kleinberg, J.; Lee, L. How opinions are received by online communities: A case study on amazon.com helpfulness votes. In Proceedings of the 18th International Conference on World Wide Web, Geneva, Switzerland, 20–24 April 2009; ACM: New York, NY, USA, 2009; pp. 141–150.
30. Feng, S.; Xing, L.; Gogar, A.; Choi, Y. Distributional footprints of deceptive product reviews. In Proceedings of the International AAAI Conference on Web and Social Media, Dublin, Ireland, 4–7 June 2012; pp. 98–105.
31. Xie, S.; Wang, G.; Lin, S.; Yu, P.S. Review spam detection via temporal pattern discovery. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; ACM: New York, NY, USA, 2012; pp. 823–831.
32. Jin, Z.; Cao, J.; Zhang, Y.; Luo, J. News verification by exploiting conflicting social viewpoints in microblogs. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; ACM: New York, NY, USA, 2016.

33. Yang, S.; Shu, K.; Wang, S.; Gu, R.; Wu, F.; Liu, H. Unsupervised fake news detection on social media: A generative approach. In Proceedings of the AAAI Conference on Artificial Intelligence, Palo Alto, CA, USA, 8–12 October 2019; AAAI Press: Palo Alto, CA, USA, 2019; pp. 5644–5651.
34. Potthast, M.; Kiesel, J.; Reinartz, K.; Bevendorff, J.; Stein, B. A stylometric inquiry into hyperpartisan and fake news. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), Melbourne, Australia, 15–20 July 2018; Association for Computational Linguistics: Stroudsburg, PA, USA, 2017; pp. 231–240.
35. Jin, Z.; Cao, J.; Zhang, Y.; Zhou, J.; Tian, Q. Novel visual and statistical image features for microblogs news verification. *IEEE Trans. Multimed.* **2016**, *19*, 598–608. [CrossRef]
36. Gupta, A.; Lamba, H.; Kumaraguru, P.; Joshi, A. Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In Proceedings of the 22nd International Conference on World Wide Web, Rio de Janeiro, Rio de Janeiro, Brazil, 13–17 May 2013; ACM: New York, NY, USA, 2013; pp. 729–736.
37. Castillo, C.; Mendoza, M.; Poblete, B. Information credibility on Twitter. In Proceedings of the 20th International Conference on World Wide Web, Hyderabad, India, 28 March–1 April 2011; ACM: New York, NY, USA, 2011; pp. 675–684.
38. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep learning–based text classification: A comprehensive review. *ACM Comput. Surv.* **2021**, *54*, 1–40. [CrossRef]
39. Horne, B.; Adali, S. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. In Proceedings of the Eleventh International AAAI Conference on Web and Social Media, Montreal, QC, Canada, 15–18 May 2017; Association for the Advancement of ArtificialIntelligence: Palo Alto, CA, USA, 2017; pp. 759–766.
40. Mitra, T.; Gilbert, E. Credbank: A large-scale social media corpus with associated credibility annotations. In Proceedings of the Ninth International AAAI Conference on Web and Social Media, Oxford, UK, 26–29 May 2015; AAAI: Palo Alto, CA, USA, 2015; pp. 258–267.
41. Shaar, S.; Hasanain, M.; Hamdan, B.; Ali, Z.S.; Haouari, F.; Nikolov, A.; Kutlu, M.; Kartal, Y.S.; Alam, F.; Da San Martino, G. Overview of the CLEF-2021 CheckThat! Lab task 1 on check-worthiness estimation in tweets and political debates. In Proceedings of the CLEF 2021—Conference and Labs of the Evaluation Forum, Bucharest, Romania, 21–24 September 2021; CEUR: Uzhhorod, Ukraine, 2021; pp. 369–392.
42. Shahi, G.K.; Struß, J.M.; Mandl, T. Overview of the CLEF-2021 CheckThat! Lab: Task 3 on fake news detection. In Proceedings of the CLEF 2021—Conference and Labs of the Evaluation Forum, Bucharest, Romania, 21–24 September 2021; CEUR: Uzhhorod, Ukraine, 2021.
43. Nakov, P.; Da San Martino, G.; Elsayed, T.; Barrón-Cedeño, A.; Míguez, R.; Shaar, S.; Alam, F.; Haouari, F.; Hasanain, M.; Babulkov, N.; et al. The CLEF-2021 CheckThat! Lab on detecting check-worthy claims, previously fact-checked claims, and fake news. In Proceedings of the ECIR: European Conference on Information Retrieval, Lucca, Italy, 28 March–1 April 2021; Springer International Publishing: Cham, Switzerland, 2021; pp. 639–649.
44. Tacchini, E.; Ballarin, G.; Della Vedova, M.L.; Moret, S.; de Alfaro, L. *Some Like It Hoax: Automated Fake News Detection in Social Networks, Technical Report UCSC-SOE-17-05*; University of California: Santa Cruz, CA, USA, 2017.
45. Tandoc, E.C., Jr.; Lim, Z.W.; Ling, R. Defining "fake news": A typology of scholarly definitions. *Digit. J.* **2018**, *6*, 137–153. [CrossRef]
46. Zubiaga, A.; Aker, A.; Bontcheva, K.; Liakata, M.; Procter, R. Detection and resolution of rumours in social media: A survey. *ACM Comput. Surv.* **2018**, *51*, 1–36. [CrossRef]
47. Thompson, P.; Daikou, S.; Ueno, K.; Batista-Navarro, R.; Tsujii, J.; Ananiadou, S. Annotation and detection of drug effects in text for pharmacovigilance. *J. Cheminform.* **2018**, *10*, 37. [CrossRef]
48. Hripcsak, G.; Rothschild, A.S. Agreement, the F-measure, and reliability in information retrieval. *J. Am. Med. Inform. Assoc.* **2005**, *12*, 296–298. [CrossRef]
49. Thompson, P.; Iqbal, S.A.; McNaught, J.; Ananiadou, S. Construction of an annotated corpus to support biomedical information extraction. *BMC Bioinform.* **2009**, *10*, 349. [CrossRef]
50. Brants, T. Inter-annotator agreement for a German newspaper corpus. In Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00), Athens, Greece, 31 May 2000; European Language Resources Association (ELRA): Paris, France, 2000.
51. TweetScraper. Jonbakerfish/TweetScraper Is a Simple Crawler/Spider for Twitter Search without Using API. Available online: https://github.com/jonbakerfish/TweetScraper (accessed on 5 February 2022).
52. Hashtagify. Search and Find the Best Twitter Hashtags-Free. Available online: https://hashtagify.me/hashtag/thebookofbobafett (accessed on 5 February 2022).
53. Yetisgen-Yildiz, M.; Solti, I.; Xia, F.; Halgrim, S. Preliminary experiments with Amazon's mechanical turk for annotating medical named entities. In Proceedings of the NAACL HLT 2010 Workshop on Creating Speech and Language Data with Amazon's, Los Angeles, CA, USA, 6 June 2010; Association for Computational Linguistics: Stroudsburg, PA, USA, 2010; pp. 180–183.
54. Snow, R.; O'Connor, B.; Jurafsky, D.; Ng, A.Y. Cheap and fast–but is it good? Evaluating non-expert annotations for natural language tasks. In Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing, Honolulu, HI, USA, 25–27 October 2008; Association for Computational Linguistics: Stroudsburg, PA, USA, 2008; pp. 254–263.
55. Eickhoff, C.; De Vries, A.P. Increasing cheat robustness of crowdsourcing tasks. *Inf. Retr.* **2013**, *16*, 121–137. [CrossRef]

56. Gravano, A.; Levitan, R.; Willson, L.; Beòuš, Š.; Hirschberg, J.B.; Nenkova, A. Acoustic and prosodic correlates of social behavior. In Proceedings of the 12th Annual Conference of the International Speech Communication Association, Florence, Italy, 27–31 August 2011; Interspeech: Brno, Czech Republic, 2011; pp. 97–100.
57. Gurajala, S.; White, J.S.; Hudson, B.; Voter, B.R.; Matthews, J.N. Profile characteristics of fake Twitter accounts. *Big Data Soc.* **2016**, *3*, 2053951716674236. [CrossRef]