 doi:10.20944/preprints202011.0369.v1

Understanding the Strategies of Creating Fake News in Social Media

Dinusha Vatsalan
Data61-CSIRO
Sydney, NSW,, Australia

Jeyakumar Samantha Tharani
University of Jaffna, Sri Lanka
Jaffna, Sri Lanka

Nalin A.G. Arachchilage
La Trobe University
Melbourne, Australia

Abstract

Social media giants like Facebook are struggling to keep up with fake news, in the light of the fact that disinformation diffuses at lightning speed. For example, the COVID-19 (i.e. Coronavirus) pandemic is testing the citizens' ability to distinguish real news from falsifying facts (i.e. disinformation). Cyber-criminals take advantage of the inability to cope with fake news diffusion on social media platforms. Fake news, crafted as a means to manipulate readers to perform various malicious IT activities. However, no previous study has investigated the strategies used to create fake news on social media. Therefore, we have analysed five data-sets that contain online news articles (i.e. both fake and legitimate news) to investigate strategies of crafting fake news on social media platforms. Our study findings revealed a threat model understanding strategies of crafting fake news which may highly likely diffuse on social media platforms.

Keywords: sentiment analysis, machine learning, fake news

ACM Reference Format:
Dinusha Vatsalan, Jeyakumar Samantha Tharani, and Nalin A.G. Arachchilage. 2020. Understanding the Strategies of Creating Fake News in Social Media. In *Proceedings of xxxx xxxx*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Alice is on the train travelling to work and she suddenly receives a Facebook notification. Scrolling through her Facebook feed, she's stopped for a second by the news that her company is going through a redundancy process. She immediately thought to respond to or share the story with her colleagues at the office, so that they can discuss it later. However, Alice has a disturbing thought, concerning what if the story was not real (i.e. fake news or disinformation). Fake news diffusion has become a serious threat not only to individuals, but also to businesses as well as the government [13]. Fake news makes up stories that aren't true or there may be some truth in it, but aren't entirely accurate [25]. The most serious case is when the fake news escalates to a life-threatening situation.

During the COVID-19 pandemic [23], as the life-threatening disease continues to spread, so does disinformation via social media platforms. There was some exciting news spreading through social media about COVID-19, which seems to come from a legitimate source that contain valid advice from real medical professionals, but turned out to be erroneous and in some cases dangerously bad. Johns Hopkins University cautioned of such messages that were being shared on social media platforms [23]. Most of them claimed to have an "excellent summary" of COVID-19 pandemic though revealed as wrongly attributed.

It is easy to diffuse fake news at lightning speed, but hard to stop. Social media giants like Facebook and Twitter are struggling to cope with disinformation on their platforms [10]. On the other hand, fake news is a threat to national democracy [6]. Pew Research Centre survey [19] reports in June 2019, that almost 70% of US citizens believe fake news and disinformation have greatly influenced on their confidence in government institutions due to the 2016 US presidential election. The election was marked by a large number of fake news, which carried disinformation, shared on social media platforms [9]. Similarly, fake news is also said to have loomed over the Australian federal election in 2019 and influenced the UK European Union membership (Brexit) referendum [11]

Social media platforms have become a common place for the rapid prevalence of fake news [6, 24]. Almost anyone, who has a profile with them, can publish and share their opinions. When people ignore fact-checking (e.g. source) before sharing, it is hard to stop fake news "going viral" over the social media platforms [13]. Previous research has developed countermeasures to combat against the prevalence of disinformation on social media platforms [2, 17]. Most of the work focused on fake news (accuracy) detection using meta data [8, 12, 21, 25].

So far, there has been little work reported on the readers' interactions to fake news shared by other people (e.g. (unknown) friends or celebrities) [13, 17, 22, 25]. This requires studying users' emotion and behaviour when falling for fake news in social media. However, one cannot deny

xxxx xxxx, xxxx xxxx, xxxx, xx, xxx

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

 doi:10.20944/preprints202011.0369.v1

the fact that fake news is intentionally created as a means to psychologically manipulate users to perform malicious IT activities, such as clicking on fraudulent links associated with fake news/posts [17]. Once the fraudulent (i.e. phishing) link is clicked, they may even disclose their sensitive information (e.g. user-names/passwords or online banking details) to hackers [7].

Moreover, there are circumstances, where hackers will get malicious IT applications (i.e. "ransomware") installed on the victim's computer, which encrypts pretty much all the data. Therefore, it is imperative to investigate the strategies used by cyber-criminals to create fake news on social media that people can fall for. In this work, we develop a threat model, understanding the strategies used to create fake news that may highly likely diffuse on social media. We also discuss how those strategies entice people to perform various malicious IT activities. To achieve that, we have analysed five data-sets using Machine Learning (ML) that contain online news articles (i.e. both fake and legitimate news) to investigate strategies that are used by cyber-criminals to create fake news on social media platforms. The remainder of the paper is structured as follows: we discuss the related work, we present the methodology and then the experiment evaluation discussing the threat modelling and our findings, limitations and future work, and conclude the paper.

2 Related Work

Fake news often overlaps with missing information, such as false or misleading information (a.k.a misinformation) and disinformation (false information but purposely created to shared to deceive people) [17]. Other than the apparent fact-checking, the research supporting its efficacy is, at best, mixed.

Previous research employed a sentiment analysis technique in Machine Learning (ML) to detect fake news [5, 14, 26]. In addition, there has been a number of studies conducted in the direction of analysing emotions in news to detect fake news [3, 5, 14]. For example, a study focused on examining how sentiment polarity influences the fake news detection [14]. However, these studies are mostly limited to the negative and positive polarities of the keywords. However, they failed to consider numerous lexical features of the news/posts in their sentiment analysis (i.e. emotion), for example, emojis (e.g. ':-)'), sentiment related acronyms and initialism (e.g. 'LoL', 'WTF') and commonly used slang (e.g. 'nah', 'meh', 'giggly'). Furthermore, emotions expressed in different forms in social media, for example, through the title of the news/posts, images and videos used, or users' comments and reactions, are not considered in the previous studies [3, 5, 14]. Lee et al. have conducted a online experiment with 261 participants to investigate people's susceptibility to fake news on social media. However, their investigation focused on how fake news are associated with real news that individuals viewed previously, as well as their cognitive ability [18].

Moreover, there has no attention been paid on analysing how users' emotion influences on their behaviour when falling for fake news in social media. On the one hand, it is imperative to understand the fact that perpetrators use people's (i.e. readers') emotions and behaviours against them, while on the other hand it is also important to learn how emotional and behavioural features are used to create fake news that can manipulate people (i.e. readers) to fall for fake news in social media. Therefore, this study investigated strategies used by attackers based on emotional, behavioural and metadata features to create fake news that are highly like to diffuse in social media.

3 Methodology

We have analysed five data-sets using Machine Learning (ML) that contain online news articles (i.e. both fake and legitimate news) to investigate strategies that are used by cyber-criminals to create fake news on social media platforms. Therefore, in this section, we describe the proposed framework for our study (outlined in Figure 1). It consists of three main phases, which are analysing attackers' strategy, training the model, and validating the outcome.

3.1 Analysing Attackers' Strategy

To analyse the attackers' strategy, we categorise the features into: 1) Content and title of the news, 2) users' behavioural features and 3) article's metadata features.

1. **Sentiment based strategy in the content and title of news**: Emotions conveyed through the news can influence the users in reading the news [4, 15]. Attackers can take advantage of highly sensitive emotional aspects in the news to target victims to fall into the fake news and consequently disseminate the news in the social media platforms. Emotions in the text data of the news content, title of the news/post, images and/or videos included in the news, as well as in the comments and replies to the news trigger the users' behaviour towards reading and disseminating fake news. As our experimental results reveal, crafting fake news with highly sensitive emotions is a major strategy used by the attackers to diffuse fake news rapidly.
   To study the emotions associated within text data of the news, we use VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis tool [14], which is specifically developed to extract sentiments expressed through posts or news in social media. It considers several aspects of emotions, including the use of exclamation marks, capitalization, intensifying words (e.g. extremely), conjunctions (e.g. nevertheless), emojis, slangs, acronyms and emoticons. It returns the positive, negative, and neutral sentiment scores

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

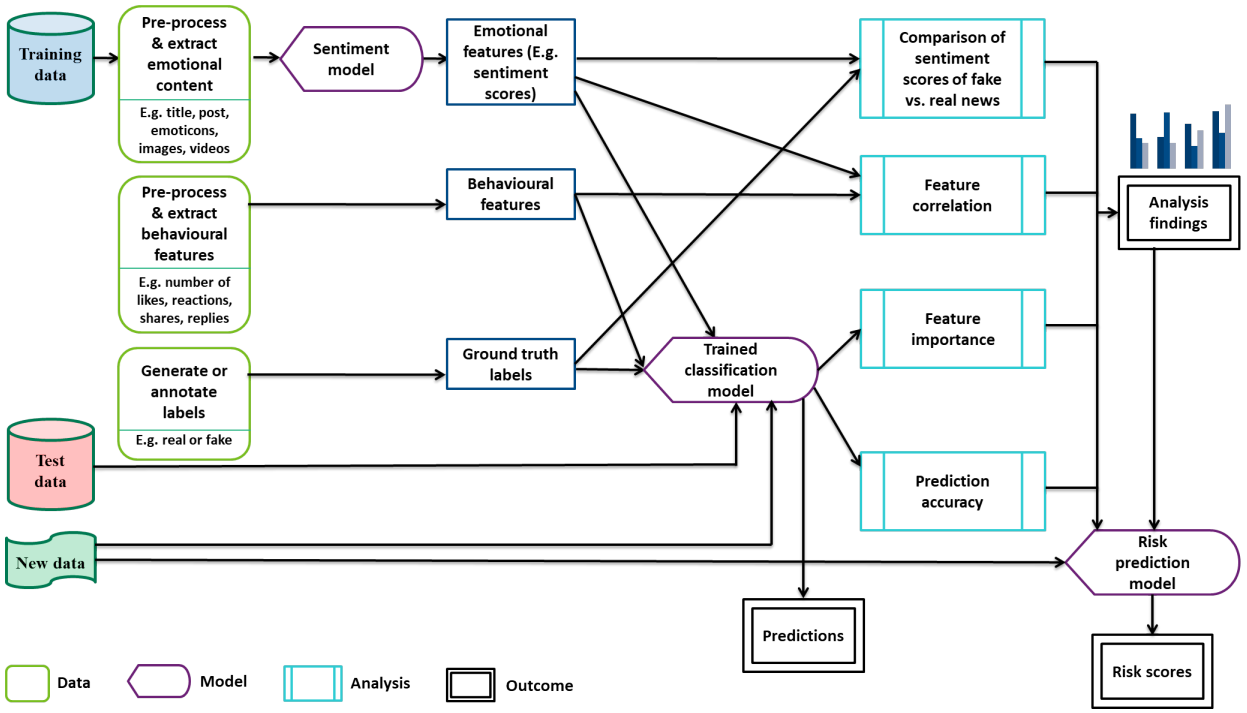

Figure 1. Proposed Model for Identifying Attackers' Strategies on Fake News

(that are summed to 1.0), as well as the compound score, which is the sum of lexicon ratings (normalized between -1 and +1).

We compare the sentiment scores of fake and legitimate news to study the differences in the sentiment scores. This allows us to study, which sentiment (negative, positive, or neutral) is heavily used by the attackers in fake news. For example, attackers might use a news title with highly negative sentiment to lure victims.

2. **Users' Behaviour Based Strategy**: Users' behaviours concerning to the news/posts are often influenced by the emotional aspects, and consequently the behavioural aspects can impact the diffusion of fake news widely to other users. Usually if people feel emotional when they read some articles, for example posts related to COVID-19 [10], they tend to share that information with their close ones. This behaviour of sharing the news leads to diffusion of the news in social media platforms.

   Behavioural data that corresponds to not only sharing, but also the number of likes, reactions, comments, and replies to the news/posts in social media [13]. These features can greatly influence the users in reading and sharing the news, as it is not uncommon in social media platforms that users get influenced by the behaviour of other users (e.g. known friends, celebrities, the users in the common circle of friends with their friends or even unknown users). This can be used as a strategy by the attackers to diffuse fake news in social media to trap users to perform malicious IT activities. Therefore, we use the behavioural features to identify the correlation amongst those features and the correlation with the emotional features. For example, negative sentiment in the news might have a positive correlation with the number of shares, i.e. users might tend to share the news with high likelihood if the news contains more negative emotions.

3. **Metadata Based Strategy**: Metadata has widely been used in the literature to identify fake news [8, 21]. Such metadata includes a top image in the post, source, hash tags in tweets, topic area of the article, etc. Attackers use strategies utilising article's metadata for fake news composition. For example, as validated by our experimental results, attackers use strategies to craft fake news that looks and feels of the legitimate news coming from sources like BBC, CNN, WHO, etc., or with eye-catching or popular images to make people fall for fake news.

   Similarly, using misleading hash tags in tweets or using titles based on the current trend or with a celebrity's name can increase the popularity of the news/posts. In our study, we analyse the correlation between the emotional features and the metadata, as well as the behavioural features and metadata to understand attackers' strategies of crafting news utilising these features.

Moreover, we identify the most influential metadata by ranking the features using feature importance scores. The feature importance scores can be calculated for decision tree or Random forests classifier based on the reduction in the criterion used to select split points in the trees, such as Gini or entropy [16]. High-scored features are more influential in fake news detection. Thus, the identified high-scored features are highly related to the strategies used by the attackers for crafting the fake news.

3.2 Threat model training

To analyse the attackers' strategies, we identify influential or more useful features for the fake news detection. With such feature selection method, we then extract those important features that are highly used in the attackers' strategies. A machine learning-based supervised method (e.g. random forest or support vector machine) is used to train the threat model with the selected features. The trained model is then used on the test data to validate the effectiveness of the threat model.

3.3 Risk score prediction

Based on the trained threat model, a risk score prediction model can be developed to predict the likelihood of news/posts in social media being fake based on the attackers' strategies using metadata, emotional and behavioural data available in the news. The feature importance scores can be used as weights for the different features and the risks in these features need to be quantified and normalized between 0.0 and 1.0, such that 1.0 indicates the highest risk and 0.0 indicates no risk. Sentiment scores of emotional features are already in the range $[0.0, 1.0]$, while behavioural features need to be normalized, e.g. using the min-max normalization [20]. However, quantifying the risk in metadata is challenging and requires further study (left as future work). New data can then be predicted using this model and the predicted score can be used to alert the users of the potential risk based on attackers' common strategies.

4 Experimental Evaluation

We have analysed the different types of attackers' strategies for fake news generation and diffusion using the following datasets:

1. **Covid-19 Public Media Dataset** is a recent dataset [1] containing 50795 news focusing on the non-medical aspects of COVID-19. The data is scraped from a range of more than 20 high-impact blogs and news websites with five topic areas: general, business, finance, tech and science.
2. **Buzz_feed** [2] contains 1932 fake and 2537 real online news articles with title, text, URL, top-image, authors, source, publish-date, movies, images, and canonical-link metadata features.
3. **FA-KES** [3] consists of 804 fake and legitimate news articles from several media outlets including mobilisation press, loyalist press, and diverse print media.
4. **SMNews** [4] contains around 13000 social media news with several attributes including title, text, and number of replies, likes, comments, participants, and shares.
5. **Liar** dataset [5] includes 12800 human labeled short statements sampled from various contexts/venues including news releases, campaign speeches, tweets, and Facebook posts.

In the following, we discuss attackers' strategies based on sentiment analysis, feature correlation analysis and feature importance analysis.

4.1 Strategy behind the content and title of the fake news:

We have performed the sentiment analysis on the content and the title of the fake news and compared them with the real news. Our results shown in Figures 2, 3, and 4 reveal that negative emotions are high in both real and fake news. However, the negative sentiment scores are generally higher with the fake news than real news. This shows that attackers often craft fake news associated with high negative emotions to psychologically manipulate (i.e. emotionally) victims to fall for fake news. The research study conducted in [1] shows that human brains evolved to react much more strongly to negative experiences than positive ones. Hence, attackers take advantage of this human nature in fake news composition and earn benefits by manipulating users to perform various malicious IT activities through fake news.

4.2 Strategy behind the feature correlation:

We have analysed feature correlation between emotional, behavioural, and metadata features available in the **Buzz_feed** and **SMNews** datasets to understand their impact on attackers' strategies.

1. **Sentiment score vs. metadata features**
   Figures 5 and 6 show the correlation between the sentiment scores and the metadata. These results reflect that composing fake news with the title that has highly negative emotion (almost 80%) directly influences the domain rank of the news. Further, attaching a (sensitive/popular) image in the news/posts, will make the news diffusing viral on social media platforms. Making

[1] https://anacode.de/download-covid-19-public-media-dataset/

[2] https://www.kaggle.com/mdepak/fakenewsnet

[3] https://www.kaggle.com/mohamadalhasan/a-fake-news-dataset-around-the-syrian-war

[4] https://www.kaggle.com/mrisdal/fake-news

[5] https://www.cs.ucsb.edu/~william/ data/liar_dataset.zip

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

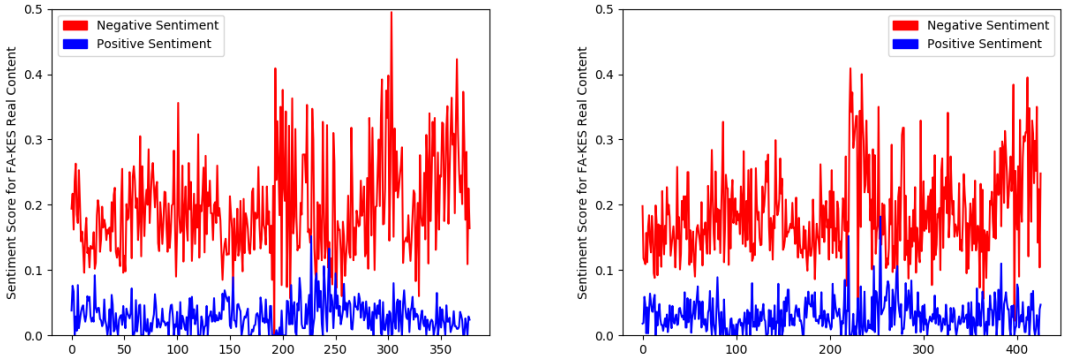

Figure 2. Sentiment score of news content for fake (left) and real (right) news on the FA-KES dataset.

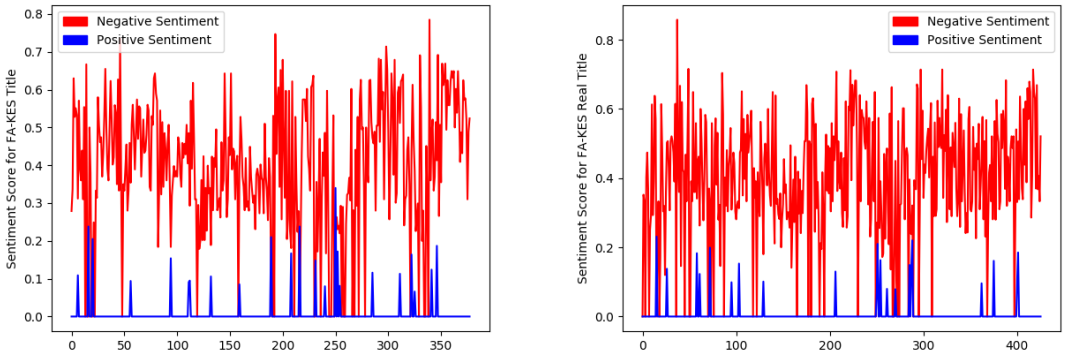

Figure 3. Sentiment score of news title for fake (left) and real (right) news on the FA-KES dataset.

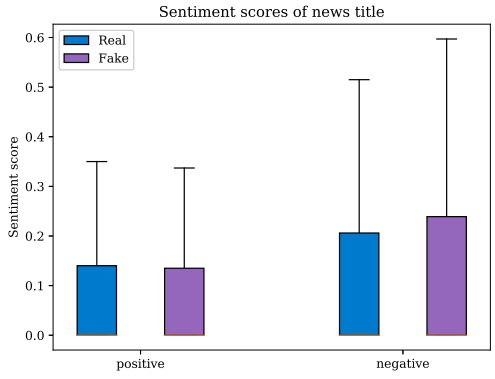

Figure 4. Comparison of sentiment scores for fake and real news title on the SMNews dataset.

the title of the fake news with negative emotions will improve the crawling of the website. These strategies show how attackers craft fake news to lure victims and how they enable their fake news becoming popular within a short time period and disseminating among many victims.

In addition, we have also analysed the correlation between the topic of the news and the sentiment score in the latest COVID-19 dataset. The results are presented in Figures 7 and 8 which show that all five categories of the topics have high sentiment scores for neutral emotion than the other emotions. From the results, one can argue that attackers' strategies have evolved with using sentiments in the news. Since Covid-19 dataset is very recent, it reveals that attackers have started to use neutral emotions commonly in the fake news generation to mutate their strategies of using negative emotions into neutral emotions to be successful.

2. **Sentiment score vs. user behavioural features**
As can be seen in Figure 9, our results reveal that the neutral sentiment score of the title have positive correlation with the replies count and participants count. Further, the number of likes and shares are 100% correlated with each other. These results show that fake

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

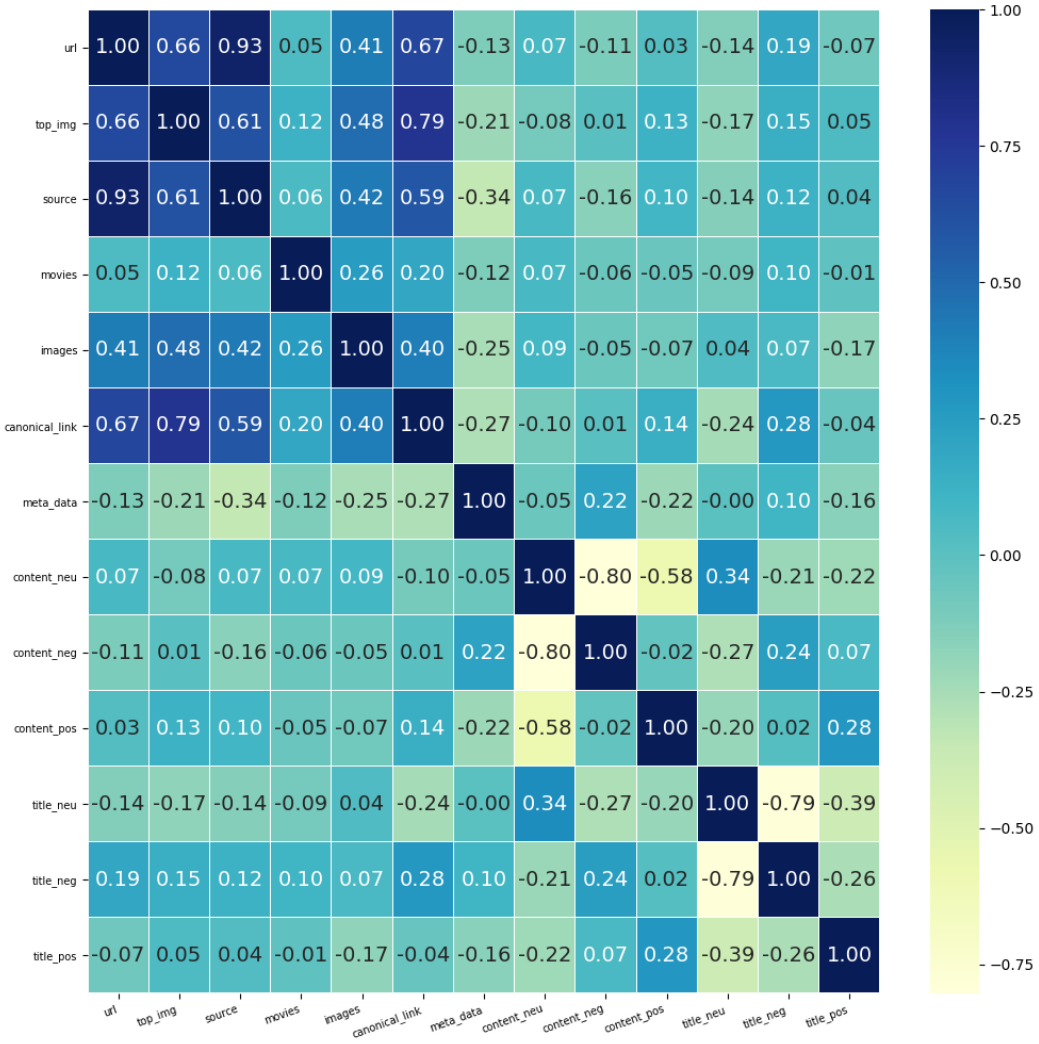

Figure 5. Correlation between the sentiment score and metadata in Buz_feed.

news crafted by attackers with the title having neutral sentiment will influence the behaviour of readers reacting to the fake news (e.g. number of replies) in social media. News with large number of likes or reactions are highly likely to be shared by users.

3. **User behavioural features vs. metadata**
Figure 10 shows the results for correlation amongst the metadata and user behavioural features in SMNews dataset, which reveals that the number of likes and shares are positively correlated with the main image URL. In addition, the news released by the country is also positively correlated with the number of likes, participants and replies. This finding shows how attackers improve the likes and shares by simply attaching a popular/sensitive image in the news/posts. Further, people have a tendency to immediately trust news if it is released by a popular, trust-worthy or well-established country. Hence, attackers also consider this as one of their strategies to manipulate the fake news. This will improve the number of participants, likes as well as

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

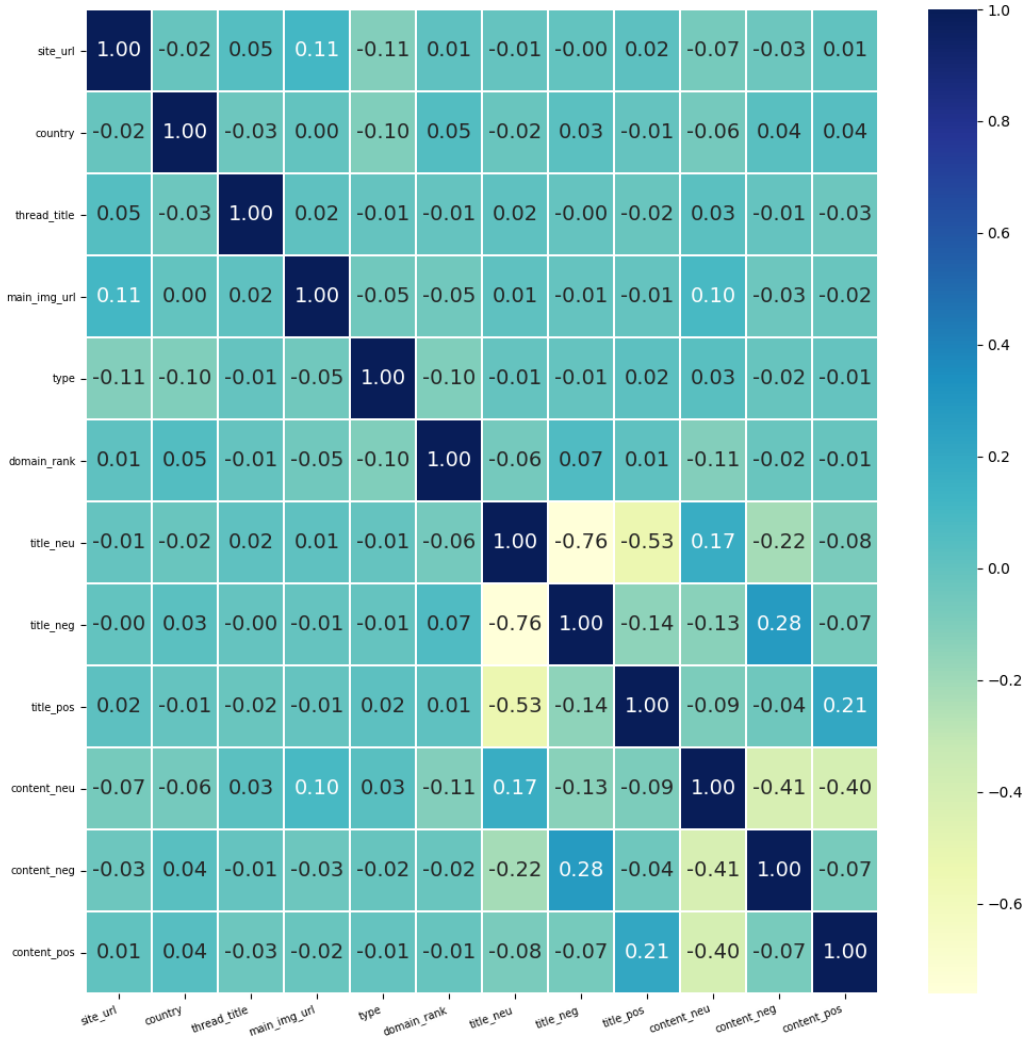

Figure 6. Correlation between the sentiment score and metadata in SMNews dataset.

shares, which will open a back door for increasing number of phishing attacks.

4.3 Strategy based on the feature importance:

We have identified the influential features of news used in strategies by attackers for fake news generation and diffusion based on the Gini index used by the random forest model to classify news as fake or real. Figures 11 and 12 present the results of this analysis, which shows that the metadata attributes like canonical link, source, images and top image achieve higher importance scores than the emotional features in the title and content of news. Instead of creating fake news with emotional content or title, attackers can simply lure victims by using the metadata features. For example, if the news comes with a trustful organisation's image (e.g. BBC, CNN, etc.) it gives more trust to the users on the news. Not only the image, but also other metadata features, such as the website URL, also help improving the trust of the users on the news. While such metadata features allow easier identification of fake news than emotional features, emotional

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

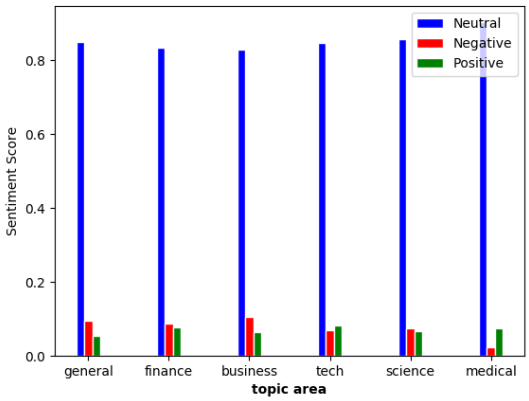

Figure 7. Comparison on title's sentiment score and topic area in Covid-19 dataset.

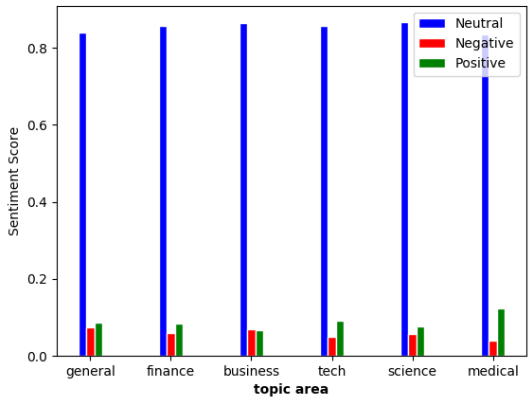

Figure 8. Comparison on content's sentiment score and topic area in Covid-19 dataset.

features tend to attract users more towards fake news (as the results of the correlation analysis between emotional and behavioural features reflect) and consequently can prevent them from looking into the metadata.

Feature importance analysis on **FA-KES** dataset in Figure 13 illustrates that the negative and neutral emotions in the news content and title are essential in distinguishing fake and real news than positive emotions. Interestingly, the negative and neutral sentiment scores in the title of the news have higher importance scores as similar to the sentiment scores in the news content.

Finally, we evaluate the accuracy of threat model built using the metadata, behavioural and emotional features of news in Figure 14. These results show that a high accuracy of classification is achieved when such features are used. The SMNews dataset contains more emotional, behavioural and metadata features compared to other datasets, Buzz_-feed dataset contains many metadata features, and FA-KES dataset does not have more behavioural or metadata features but it has title and content with higher range of sentiment scores (Figures 2 and 3). Liar dataset contains only the news content which are also in short texts, and therefore does not provide much information to the classification model. Similarly, the FA-KES and Kaggle News datasets do not contain any behavioural features, and thus the accuracy of the classification model on these datasets is lower. Higher accuracy achieved with datasets containing more emotional, behavioural and metadata features validates the significance of these features for fake news identification. On the other hand, these results reveal that attackers use combination of metadata, user behavioural, and emotional features to effectively craft fake news and successfully trap victims into the fake news.

4.4 Threat modelling based on our findings

Our key findings for the threat model are:

1. Attackers use the strategy of crafting fake news with more negative emotions than legitimate news.
2. Another strategy used by attackers is having the title of the fake news with high negative emotions.
3. Attackers evolve over time to be successful with their cyber-crime objectives. This is validated from the correlation analysis results on the recent Covid-19 dataset, which states that attackers use more neutral emotions in the news of different topics as a strategy to build trust with the users.
4. One of the interesting observations of the correlation analysis between emotional and behavioural features is that negative sentiment has a strong positive correlation with the number of likes or shares. Since news with negative emotions have more likelihood of being widely diffused in social media, attackers use such strategy of crafting fake news with negative emotions to create the trap for many victims.
5. We have found another strategy of using the metadata of news by the attackers. When fake news is crafted with popular top image or images from well known sources, it will improve the number of shares, likes, comments and participants. Thus, this is another simple technique used by the attackers to lure victims.
6. Article's metadata has high influence in fake news detection than the content based sentiment score. Thus, we can classify or detect the fake news by analysing the metadata features, for example, the accuracy of the image or source of the article. However, people tend to overlook metadata when they are highly emotional, and therefore attackers also use emotional features as a successful strategy for fake news diffusion. Unlike metadata features, emotional features are difficult to

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

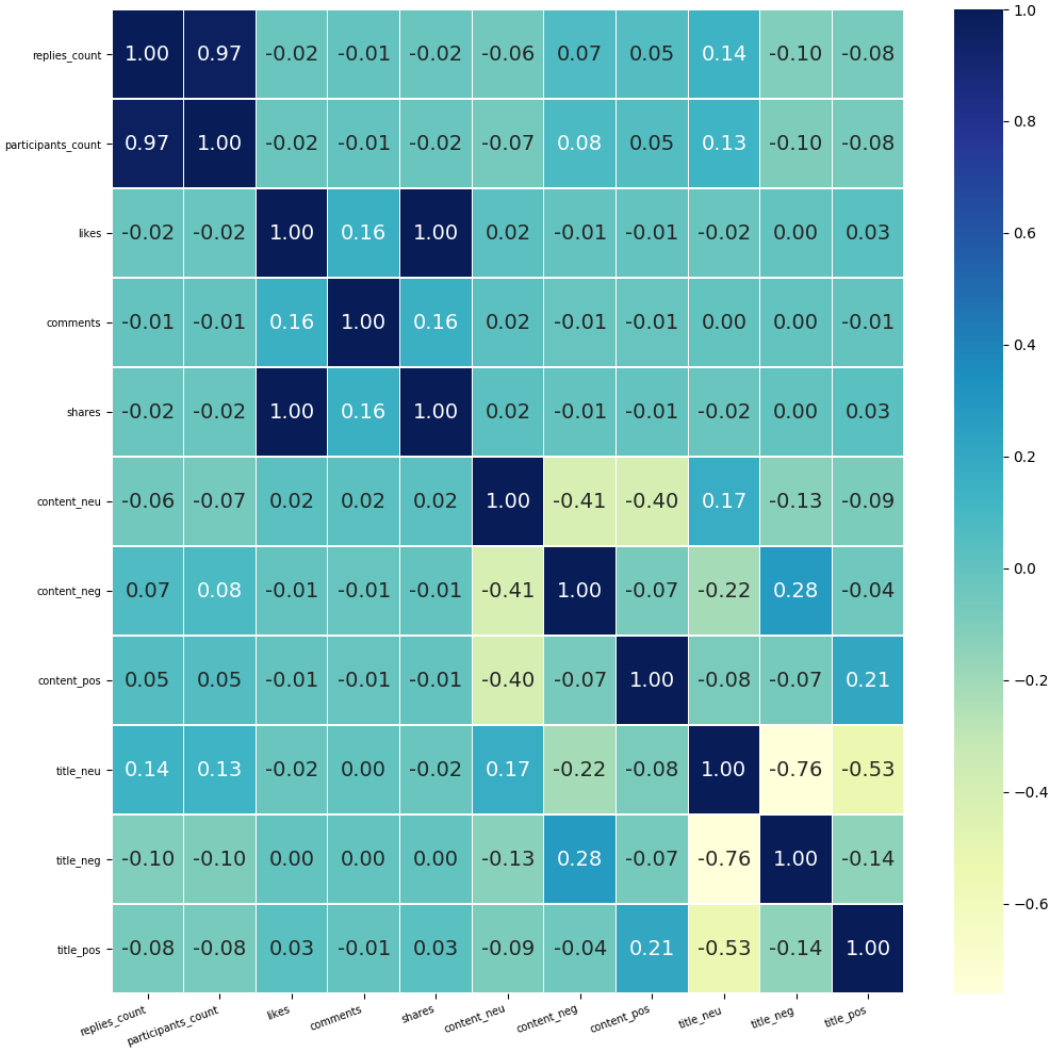

Figure 9. Correlation between the sentiment score and user behavioural features in SMNews dataset

be distinguished by people if the emotions conveyed by the news are fake or real. This benefits the attackers in building a successful strategy.

7. Finally, combining more emotional, behavioural and metadata features, significantly improves the accuracy of fake news classification, which means that another successful strategy used by attackers is manipulating the combined emotional, behavioural, and metadata features. Moreover, these features are abundantly available in social media, which the attackers utilize for successful generation and dissemination of fake news/posts in social media.

4.5 Limitations and future work

The initial results of our study showed some interesting observations and strategies used by cyber-criminals for conducting phishing attacks via fake news in social media. However, there are several limitations and open challenges that need to be solved in the future:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

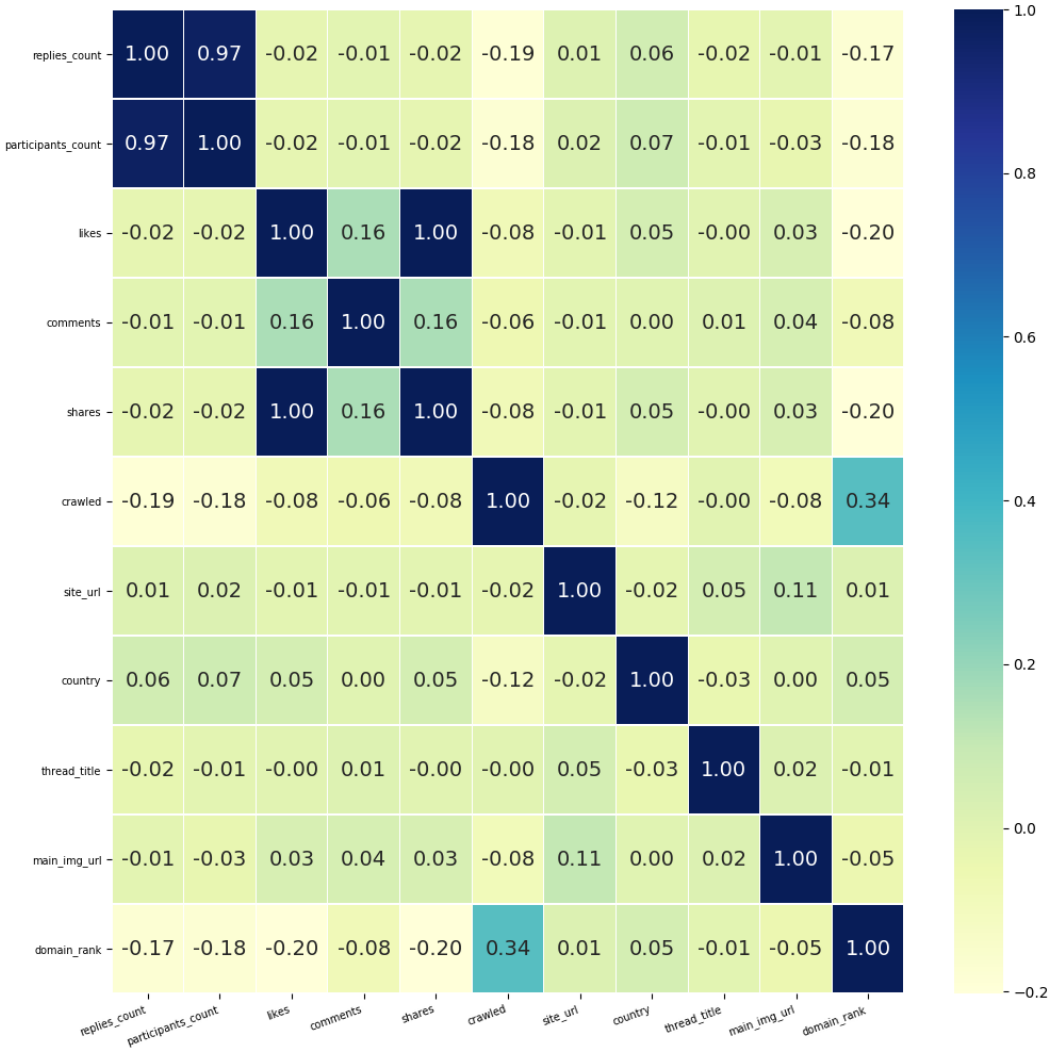

Figure 10. Correlation between the metadata and user behavioural features in SMNews dataset

1. Images and videos in the news often contain a lot of hidden emotions. The importance of emotions available in such images and videos is not validated in this paper. Moreover, some of the metadata features, such as top image, can contain emotional features. This study requires appropriate datasets containing images and videos with ground-truth labels, and using deep learning techniques [12].
2. Emotions related to religious beliefs, sexual interests, and racism aspects are not considered in this paper. These can be used by attackers to target specific victims (for example, people who are highly religious share religious posts with high likelihood). The main challenge of such a study is that it requires collecting (based on a user study) or crawling (from social media) relevant datasets, which might not be possible due to privacy concerns and restrictions.
3. Behavioural features in this study are also limited to the total number of shares, comments, likes, reactions,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

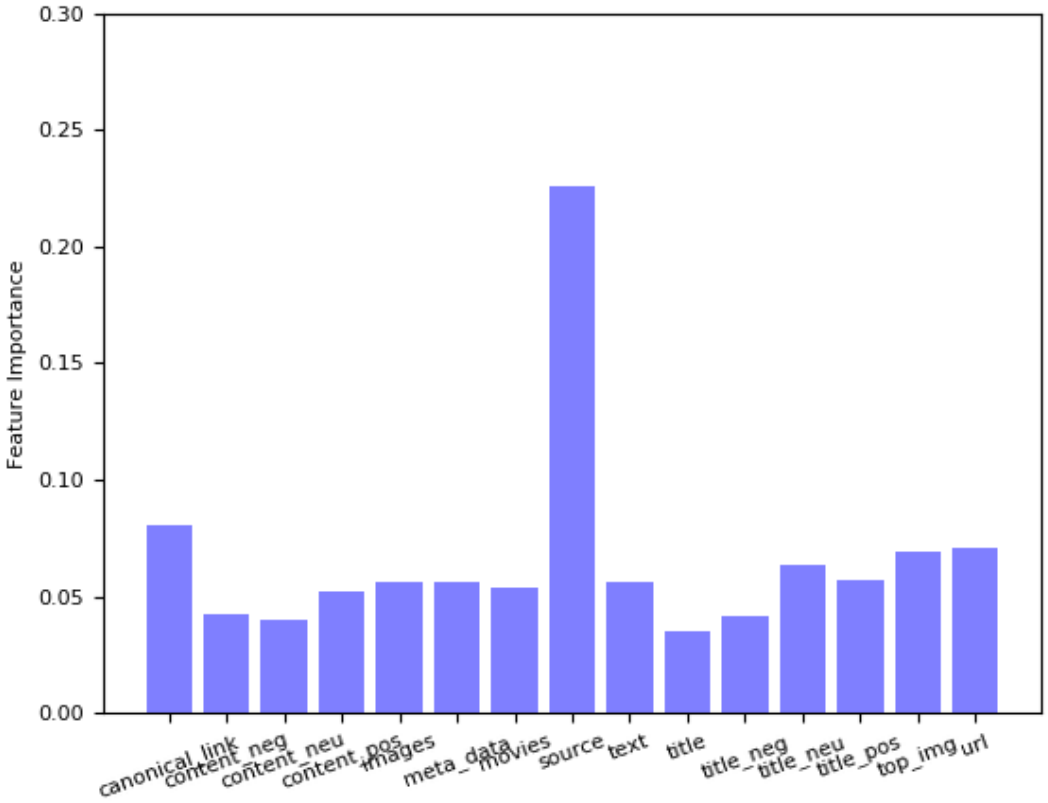

Figure 11. Feature importance for classifying/predicting using Random forest using behavioural features on the BuzzFeed dataset.

replies, and participants. However, there are other behavioural features that could play an important role in attackers' strategies for fake news diffusion, such as the number of mutual friends or common users who have liked, shared, or commented, as well as number of popular users and celebrities who have reacted, shared or replied to posts/news. Since users often get influenced by the behaviour of other users who are in the common circle of friends or celebrities, it would be interesting to study how attackers use such features with the aim of fake news dissemination in social media.

4. Finally, our study is limited to news/posts in the English language. More experiments are required to conduct an analysis of news/posts in other languages, specifically to understand if there exist common strategies used by attackers for crafting and diffusing fake news in social media.

5 CONCLUSION

With the wide-spread use of social media, diffusion of fake news is increasingly becoming common leading to several psychological, financial and economical damages to individuals, organisations and the government. In this paper, we have analysed strategies used by the attackers based on emotional, behavioural and metadata for fake news crafting and diffusion in social media. So far, there has been little work that focused on the emotional aspect of fake news detection. No previous study has investigated attackers' techniques nor strategies used to craft fake news on social media.

Our findings have revealed the key strategies commonly used by attackers. Attackers often employ emotional strategies of using negative emotions both in the content and title of the fake news. These emotions related strategies will have an impact on users' behaviour when dealing with fake news, such as the number of likes, reactions, shares, comments, re-tweets, and participants. In addition, the metadata related strategies have a great influence on the fake news than other features. Attackers may use a combination of strategies based on these features to leverage their attack (e.g. phishing) through crafting fake news.

Our study findings can be used in a risk prediction model to predict the likelihood of news/posts being fake based on the emotional, user behavioural, and metadata features. A

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 November 2020 doi:10.20944/preprints202011.0369.v1

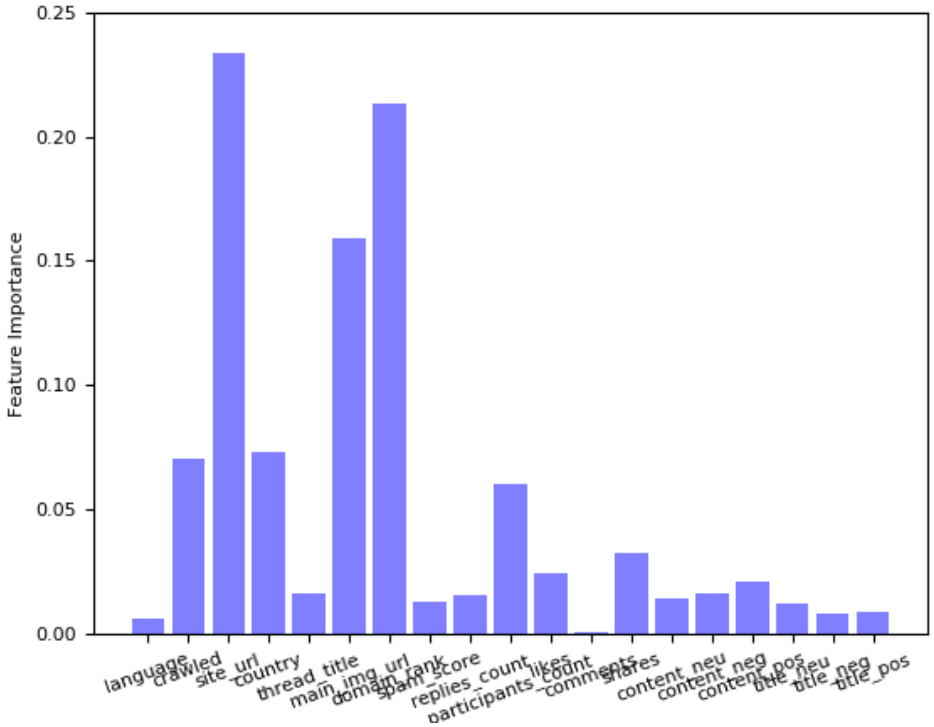

Figure 12. Feature importance for classifying/predicting using random forests using behavioural features on the SMNews dataset.

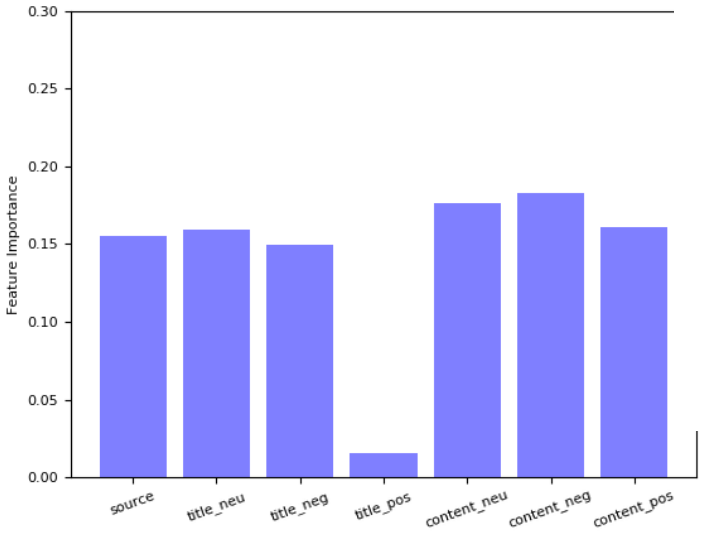

Figure 13. Feature importance for classifying/predicting using random forests using behavioural features on **FA-KES** dataset

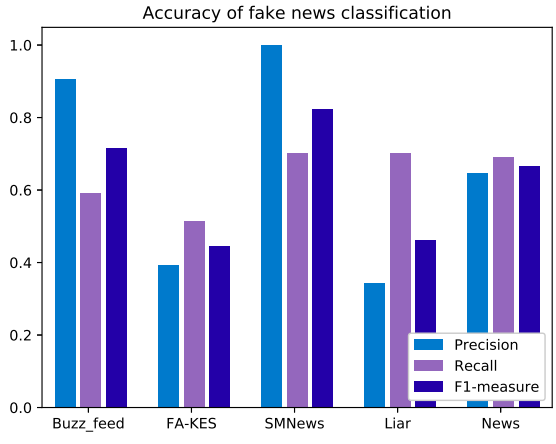

Figure 14. Accuracy of fake news classification.

browser plug-in can be implemented in social media platforms to alert users to differentiate fake news/post from legitimate ones.

References

[1] 2019. Negativity Bias. http://blog.idonethis.com/negativity-bias/, Accessed date on September 05, 2020.

[2] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2019. Sentiment aware fake news detection on online social networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2507–2511.

[3] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2019. Sentiment aware fake news detection on online social networks. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2507–2511.

[4] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2019. Sentiment aware fake news detection on online social networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2507–2511.

[5] Jonathan Albright. 2017. Welcome to the era of fake news. *Media and Communication* 5, 2 (2017), 87–89.

[6] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.

[7] Nalin Asanka Gamagedara Arachchilage, Steve Love, and Konstantin Beznosov. 2016. Phishing threat avoidance behaviour: An empirical investigation. *Computers in Human Behavior* 60 (2016), 185–197.

[8] Gaurav Bhatt, Aman Sharma, Shivam Sharma, Ankush Nagpal, Balasubramanian Raman, and Ankush Mittal. 2018. Combining neural, statistical and external features for fake news stance identification. In *Companion Proceedings of the The Web Conference 2018*. 1353–1357.

[9] Alexandre Bovet and Hernán A Makse. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature communications* 10, 1 (2019), 7.

[10] ABC News Breakfast. [n.d.]. As the coronavirus spread, an experiment showed Facebook was struggling to keep up with fake news. *ABC News Breakfast* ([n. d.]). https://www.abc.net.au/news/2020-04-24/facebook-approves-ads-with-covid-19-misinformation/12172168, Accessed date on May 05, 2020.

[11] Luke. Buckmaster and Tyson Wils. [n.d.]. Responding to fake news. *Parliament of Australia* ([n. d.]). https://www.aph.gov.au/About_Parliament/Parliamentary_Departments/Parliamentary_Library/pubs/BriefingBook46p/FakeNews/, Accessed date on September 20, 2019.

[12] Limeng Cui, Suhang Wang, and Dongwon Lee. 2019. SAME: sentiment-aware multi-modal embedding for detecting fake news. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 41–48.

[13] Christine Geeng, Savanna Yee, and Franziska Roesner. April 25–30, 2020. Fake News on Facebook and Twitter: Investigating How People (Don't) Investigate. (April 25–30, 2020), 1–14.

[14] Clayton J Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Eighth international AAAI conference on weblogs and social media*.

[15] Jozef Kapusta, L'ubomír Benko, and Michal Munk. 2019. Fake News Identification Based on Sentiment and Frequency Analysis. In *International Conference Europe Middle East & North Africa Information Systems and Technologies to Support Learning*. Springer, 400–409.

[16] Max Kuhn and Kjell Johnson. 2013. *Applied predictive modeling*. Vol. 26. Springer.

[17] David MJ Lazer, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[18] Sian Lee, Joshua P Forrest, Jessica Strait, Haeseung Seo, Dongwon Lee, and Aiping Xiong. 2020. Beyond Cognitive Ability: Susceptibility to Fake News Is Also Explained by Associative Inference. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[19] Amy. Mitchell, Geffrey. Gottfried, Galen. Stocking, Mason. Walker, and Sophia Fedeli. [n.d.]. Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed. *Pew Research Center* ([n. d.]). https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/, Accessed date on September 20, 2019.

[20] S Patro and Kishore Kumar Sahu. 2015. Normalization: A preprocessing stage. *Veer SurendraSai University of Technology (VSSUT), Burla, Odisha, India.* (2015).

[21] Karishma Sharma, Feng Qian, He Jiang, Natali Ruchansky, Ming Zhang, and Yan Liu. 2019. Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology (TIST)* 10, 3 (2019), 1–42.

[22] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.

[23] Peter Suciu. [n.d.]. During COVID-19 Pandemic It Isn't Just Fake News But Seriously Bad Misinformation That Is Spreading On Social Media. *Forbes* ([n. d.]). https://www.forbes.com/sites/petersuciu/2020/04/08/during-covid-19-pandemic-it-isnt-just-fake-news-but-seriously-bad-misinformation-that-is-spreading-on-social-media/3ca0280c7e55, Accessed date on April 10, 2020.

[24] Edson C Tandoc Jr, Zheng Wei Lim, and Richard Ling. 2018. Defining "fake news" A typology of scholarly definitions. *Digital journalism* 6, 2 (2018), 137–153.

[25] Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. In *Advances in Neural Information Processing Systems*. 9051–9062.

[26] Xinyi Zhou, Atishay Jain, Vir V Phoha, and Reza Zafarani. 2020. Fake news early detection: A theory-driven model. *Digital Threats: Research and Practice* 1, 2 (2020), 1–25.