EPiC Series in Computing

Volume 93, 2023, Pages 56–67

Proceedings of Society 5.0 Conference 2023

Emotionally driven fake news in South Africa

Marc Gagiano and Vukosi Marivate

The University of Pretoria, Pretoria, Gauteng, South Africa
marcgagiano@gmail.com, vukosi.marivate@cs.up.ac.za

Abstract

To build a more inclusive society that better integrates cyberspace and physical space, we must understand the appeal behind misinformation. Misformation focuses on maintaining an exclusive society rather than integrating society. One of the challenges following the era of information explosion is the rapid spread of misinformation in the form of fake news. People's choices based on misinformation can have dire consequences, especially in smaller developing communities. Therefore, this paper focuses on the emotional tone of fake news in South Africa to better understand its appeal. Introducing an expected emotional score shows that fake news articles contain more overall emotions than non-fake news. Fake news articles are also written with different biases in mind. These biases were detected and separated using clustering algorithms. Introducing a transformer model allowed us to further classify different biases by creating a profile of the emotions each bias contains. It is found that fake news in South Africa contains a roller-coaster of strong emotive words combining feelings of anger, joy, sadness and fear. The ratio of how these words are combined depends on a particular bias. These findings can help build better detectors of fake news in the future and create a feedback loop to help write more captivating news articles to foster a more inclusive society.

Keywords— natural language processing, transformers, machine learning, fake news

1 Introduction

Since the rise of social media and digital transformations over the last decade, the amount of available text data has exploded exponentially. Books started being made available as both digital and hard copies. Newspapers transformed their business model into apps and websites rather than relying on physical documents. Optical character recognition (OCR) techniques became better at transcribing information; thus, the transformation of print books to digital books increased significantly. However, the influence of social media probably had the most significant impact on the amount of text information generated and the amount of text being shared daily [2, 10]. Unfortunately, with more data, more challenges arose. A few decades ago, it required effort from individuals and small companies to get their ideas out to the public. Any written statement had to go through a publishing house and editors. If the publishing house accepts it, their thoughts could be published, reaching the intended audience only if the intended audience buys and reads it. Therefore, there were multiple filters before ideas would come to the public space. Social media changed this. Initially, all the filters and gatekeepers that formed the middleman were removed from the process. This transformation led to the rise of the rapid spread of misinformation [7]. Thus, it is important to understand if a sentiment analysis of South

African fake news reveals an exaggerated emotional tone. Also, does fake news have inherent biases and does the bias differ in terms of one another on an emotional level?

Fortunately, machine learning, specifically Natural Language Processing (NLP), also evolved over the same period. Techniques for processing and extracting information from text data have improved drastically over the last few years. The methods for classifying unstructured text, part of speech tagging, topic modelling, sentiment analysis, Named entity recognition, and text generation improved as more text data became available [16]. The field of NLP was further advanced with the release of the paper "Attention is all you need" [17]. The transformer model proposed by this paper revolutionised the world of NLP, allowing better models to be trained more quickly and effectively. One applicable use case for Transformer models is classifying and understanding misinformation.

Over the past few years, fake news has become increasingly common worldwide. The main problem with this pandemic is that many people believe these fake news articles and need help distinguishing these articles from actual news articles. Many of these fake news articles target people on emotional levels rather than on facts. This evocation of strong emotions creates a strong bias for these readers to believe these fake news articles [11]. The intention behind many of these articles is not innocent. The best example would be the repeated application of fake news articles during the 2016 US elections to sway voter opinions in favour of Donald Trump. Therefore many of these articles are generally biased towards a specific plan or idea, typically a political agenda. As people start to believe these articles and take action, it has dire consequences for communities and countries. Overall, fake news leads to more polarized communities [18]. Therefore, from a security point of view and to ensure that we are building towards an inclusive society, understanding the sentiment behind fake news is of utmost importance, especially in developing countries like South Africa.

2 Literature Review

In the literature, sentiment analysis is often used to help identify fake news articles. In cases where sentiment is used to classify a fake news article, it typically focuses on a ternary positive, negative and neutral classification [3, 5, 4]. However, to understand the appeal behind fake news, the overall classification of fake news is less important than how emotions fluctuate throughout the given news article. Therefore context is essential. The context-capturing ability of transformers might be helpful for this purpose [17], but it requires supervised data. The challenge in developing communities is that most data captured are unsupervised and can be challenging to label. However, if working with supervised data, research has shown that transformer models successfully identified bias and emotions in fake news [15, 14, 12]. Alternatively to using transformers, there is the option to ignore the impact of the context within the fake news articles and to focus on the emotional response of individual words. [9] and [6] showed great promise using the Valence Aware Dictionary and sEntiment Reasoner (VADER) lexicon for the sentiment analysis of individual words. Once individual words have been scored, they can be aggregated into a single metric providing an overall emotional score for a given article. This score can be used to contrast fake news with non-fake news.

Another critical analysis to consider is the detection of bias. The idea is to try and extract protected characteristics and attack vectors similar to what was done with the Hateful Memes 2021 competition [1]. However, the considered dataset was supervised in that competition. Although some ideas implemented in the Hateful Memes 2021 competition might be useful, many techniques must be adapted to work in this context. Considering a few alternative sources, it is clear that bias detection using NLP and sentiment analysis has already been successful within the research [19, 13]. As highlighted earlier, in this context, the data is unsupervised and more than just a simple positive and negative of a news article is required. Some work has been done in South African news regarding detecting fake news. [8] explored using a Long Short-Term Memory (LSTM) model to detect whether a given South-African news article is fake. Beyond this, work remains scant and further research is needed. Very little information can be found on Fake News's bias and sentiment analysis in South Africa.

3 Methodology

Various NLP techniques are applied to analyse the bias and sentiment of South African fake news. A combination of predictive modelling, clustering and transformer models is used. Fortunately, these techniques can be applied independently, and the input dataset for the various methods only needs to be pre-processed once.

3.1 Data

The data set analysed is the South African Disinformation [Fake News] Website Data - 2020. The dataset contains South African fake news articles from various online platforms. The following is a comprehensive list of platforms from which the data was sourced: search67.com, hinnews.com, sa-news.com, whatsappgroup.co.za and mzanzistories.com. Each sourced dataset contains the publication date, title, article text, link to the article and platform name. There are more than 800 observations across all five data sources [8].

3.2 Approach

3.2.1 Pre-processing

Common stopwords and pronouns are stripped from the text. Where the standard libraries in Python fail, regular expressions (regex) are used to customise the pre-processing of the text data. Once the pre-processed text data is in a suitable format, the next step is to add the data to a tabular form where it can be accessed and analysed more easily.

3.2.2 Sentiment analysis

The sentiment analysis varies slightly from the resources freely available online. As mentioned, many resources focus on the ternary classification of an article. However, the ternary classification does not help to provide insights into fake news articles' appeal. The emotions of individual words within each article need to be analysed. Data is passed to the VADER lexicon from the **nltk** library to examine the individual words. Using the VADER lexicon, one can obtain the emotional score for each word within a given text. The only challenge is determining how strongly a given term is associated with a fake news article versus a non-fake news article. For this, a predictive model is trained, and the predictors' weights measure how strongly a given word is associated with fake news.

3.2.3 Clustering

Sentiment analysis forms one part of the analysis. Another challenge is detecting different biases. Abstract topic modelling establishes and identifies biases within fake news articles. The various free texts are clustered, and individual clusters are analysed to determine if they contain a particular bias. The assumption is that articles written with a specific bias have similar underlying semantic structures and use the same vocabulary. Clustering helps to separate various biases from one another.

3.2.4 Transformers

One must create an emotional profile for the different biases detected with the topic modelling. Transformer models help to extend beyond the ternary sentiment classification to emotional classification. Using one of the models available in the Huggingface library, the articles are classified on an extended set of emotions like joy, anger, fear etc. Once each article is classified, emotional profiles for each cluster can be established by tallying up the emotional scores of the articles within that cluster.

3.3 Evaluation

Working with unlabeled data has challenges, but the total records are small enough to evaluate by probing a small sample of results. Probing is used throughout the experiments to assess whether or not the sentiment analysis, clustering, and transformer results make sense. Many of the outputs can further be evaluated by studying the various graphs. A combination of word clouds, t-SNE plots and histograms illustrate the results. The data is also appended with non-fake news. The non-fake information contrasts with the fake-news articles regarding the observed results.

4 Experiments and Results

4.1 Preprocessing

Due to the small data size, no advanced software is required, and Python is used throughout for both data pre-processing and the actual analyses.

The data sheets were loaded into Python using **Pandas**. Each of the fake news articles was appended with two additional columns. The first extra column is called *description* and has the constant value *fake*, and the second column is called *category* and has the constant value of 5. Before consolidating the data, a bit of data cleaning is required. The different sources have different time-series formats. Some articles have a short date format of **dd/mm/yyyy** format, while others have a long date format of **dd MM yyyy**. The variety of date formats limits the ability of **Pandas** to parse the date column as time-series data correctly. A custom parser was used to solve the problem of formatting the date correctly. It uses a regex pattern to detect whether or not a date is in a long or short format and parses it accordingly.

Two levels of text pre-processing are required. The first level forms part of the general pre-processing applicable to all three independent analysis tasks. The second pre-processing involves the vectoriser process, where the free text is transformed into a vector representation for sentiment analysis. To ensure that all three independent analysis tasks use the same pre-processed data set as a starting point, the text pre-processing in this step is kept to a minimum. Text processing is also iterative, so a Python function that is easily readable, maintainable and updatable is used to pre-process the free text. This function converts the text column to lowercase and then uses text substitution to replace sub-strings found within the text with new sub-strings of text. Regex is used throughout this function to do the pattern matching. Some pattern matching includes transforming values like 50 000 to 50000, writing phrases like ”I'm” as ”i am”, removing unnecessary punctuation, trimming unnecessary whitespace and separating punctuation into stand-alone characters. While more could be done in pre-processing, it is a delicate balancing act before one removes helpful information based on human bias. An algorithm can easily pick up text that seems useless, as applicable.

It helps to add objective, non-fake news articles to the dataset to help evaluate the analysis. To supplement the dataset, the AG's corpus of news articles is added to the entire dataset. Before the AG news data set (AGNews) is added, it undergoes precisely the same pre-processing the fake news was subjected to. This includes the parsing of the dates and the pre-processing of the free text. The AGNews come with a *category* column, which contains the numbers 1 to 4, indicating whether a news article is related to world, sport, business and science/technology categories. This *category* is transformed into a *description* column indicating the four types of news articles. Combining this dataset with the fake news dataset, the *description* column contains the descriptions: 'world', 'sport', 'business', 'sci_tech' and 'fake' corresponding to the numbers 1 to 5 in the associated *category* column. See Figure 1 for the number of articles within each news category.

Notice that in Figure 1, it is clear that the number of fake news articles is much less than the number of articles in the other categories. This challenge is addressed by downsampling the non-fake news categories. See Figure 2 for the output. One way to improve this pre-processing is to include more than just the AGNews dataset. Enriching the base dataset with examples of UK fake news, USA fake news, UK non-fake news, and USA non-fake news is beneficial. This is a step in the right direction for generalising the results found within this paper to apply to a broader audience.

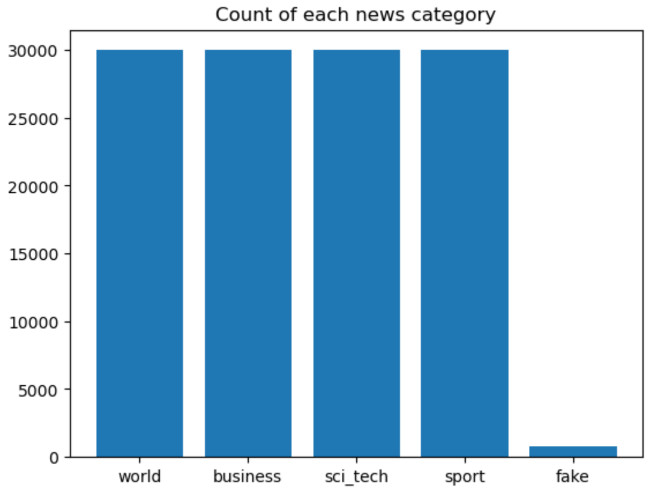

Figure 1: The number of articles within each news category.

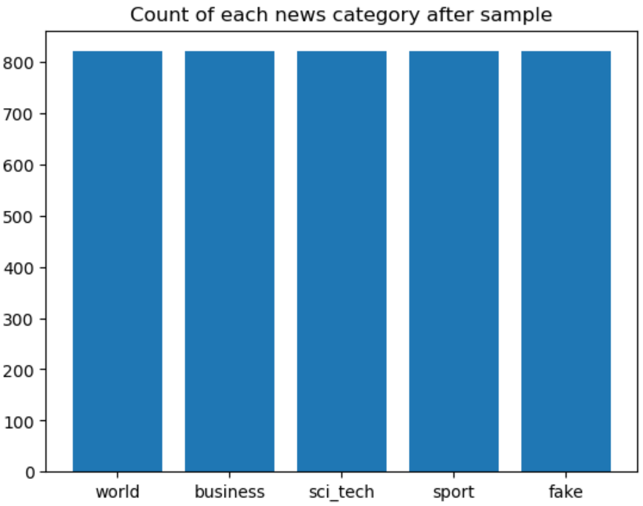

Figure 2: The number of articles within each news category after downsampling was applied.

4.2 Sentiment Analysis

The VADER model is used extensively to assist with sentiment analysis. VADER is a lexicon-style rule system pretrained on a large corpus of social media articles. It provides both the binary direction (positive and negative) of a word and a compound score indicating the magnitude of that given word in its current direction. In other words, it shows how negative or positive a given word is. The VADER model, therefore, provides a positive, neutral, and negative indicator and a compound score for each word. We omitted the neutral indicator and the neutral score (0). It might be helpful in future analysis to look at the impact and effect of neutral words in the context of fake news. The downside to VADER is that it does not consider the context of the words. However, it is still helpful in this context as the terms it was pretrained on had emotional loadings associated with them. While the exact emotional

	positive	negative	score
lack	0	1	-0.3182
challenged	0	1	-0.1027
win	1	0	0.5859
won	1	0	0.5719
hard	0	1	-0.1027
honest	1	0	0.5106
well	1	0	0.2732
value	1	0	0.34
relaxed	1	0	0.4939
played	1	0	0.34
glamour	1	0	0.5267
winners	1	0	0.4767
good	1	0	0.4404
great	1	0	0.6249

Table 1: The output of the VADER model applied to a fake news sports article.

	positive	negative	score
successful	1	0	0.5859
prize	1	0	0.5106
success	1	0	0.5719
creation	1	0	0.2732

Table 2: The output of the VADER model applied to a non-fake sports news article.

loading should vary based on the context, it does help to establish a proxy of the emotional loading of a fake news article. The differences can easily be highlighted by applying the VADER model to a sample fake news article and a sample non-fake news article. Table 1 shows the output of the VADER model applied to a single fake news article related to sports. Notice that the overall sentiment for this fake news article is positive. It contains many emotional words, each with solid emotional scores associated. To contrast, see Table 2.

While the VADER model applied to the non-fake news sports article is also positive overall, notice it contains fewer emotionally loaded words than the fake news article. Also, the fake news article switches between negative and positive polarities. This indicates the presence of possible inflexion points that might be a common feature in fake news articles. It might be worthwhile to study these inflexion points in the future. Having a way to quantify the emotional loadings of the words is helpful. Still, it is also essential to measure the likelihood of a word being associated with a fake news article compared to non-fake news. Considering the first word in Table 1 is **lack**, is it possible to tell if that word is more likely to be associated with a fake news article than a non-fake news article?

A model will be required to address the challenge of quantifying the association of a particular word with a fake-news article. The model chosen in this project is a simple logistic classifier. There are better models than this, SVM will probably work better for this type of problem, but it does help to illustrate the concept the easiest. Before the model is trained, the free text of the news articles needs to be vectorised. Various techniques were experimented with, but the count vectoriser performed the best in this scenario. If the base dataset were larger, TF-IDF would serve better. The classifier that was constructed is a simple binary classifier. The goal of the classifier was to predict if a news article is fake or not fake. See Figure 3 for the confusion matrix output on the test set.

The model did well in accurately classifying fake news from non-fake news. However, it must be noted that the AGNews data set is international compared to the fake news dataset, which is local to

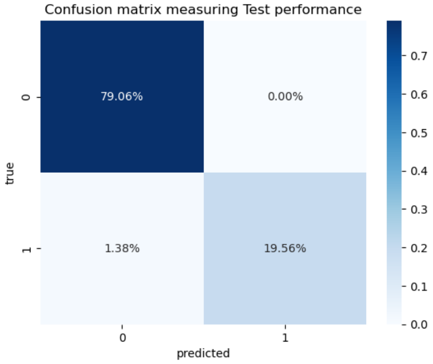

Figure 3: The confusion matrix showing the performance of the Logistic regression.

South Africa. The most important part of the model is not its predictive power but rather the weights of the predictors. These weights give the predictive strength of each word to how much it contributes to an article being fake. This predictive strength, combined with the emotional loading from the VADER score identified earlier, allows us to establish a new score, the **expected emotion**. The new score is analogues to the expected value, but instead of probability times weight, it represents probability times the emotional weight. A more rigorous exercise is required to define this value better and to ensure it confirms all the required rigorous statistical and mathematical tests. Still, with the experiments done here, we show that it is a valuable measure for analysing the sentiment of articles. Currently, two measures can be applied across all news articles. Each word in an article is evaluated in terms of the newly defined expected emotion score and the absolute value of the VADER score. The absolute value is used to establish the total emotional loading of the word. Direction is of less importance in this context. Once all the words in an article are evaluated, scores are tallied for that particular article. This process is iterated across all articles.

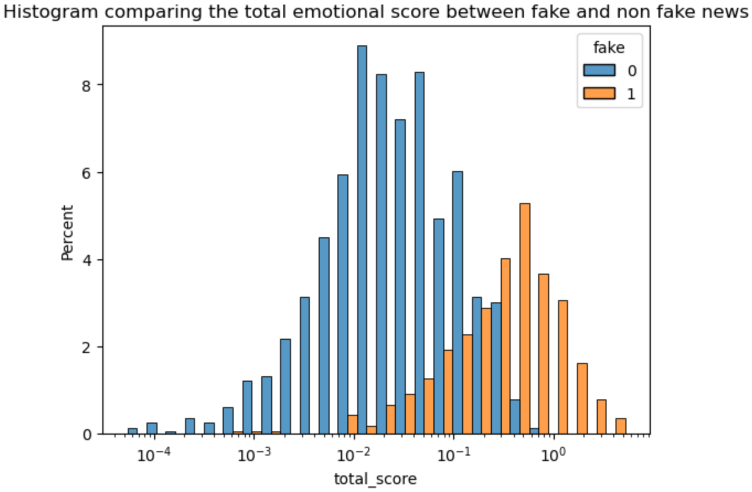

Figure 4: The histogram compares the total expected emotion per article between fake and non-fake news.

Figure 4 and Figure 5 show the final results after evaluating each article. The histograms use a log

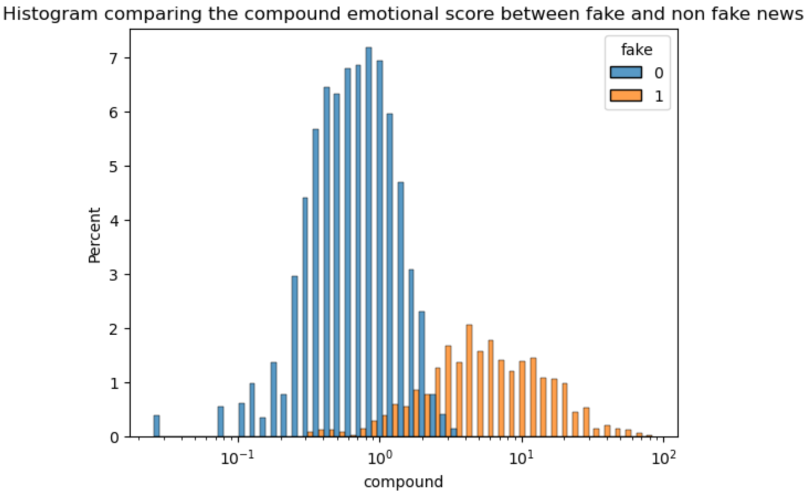

Figure 5: The histogram compares the total emotional loading per article for fake and non-fake news.

scale for the x-axis. Therefore the mean value for the fake news score in each graph is an order of ten higher than the non-fake news. It should be clear that there exists a clear separation between the two classes based on the scores highlighted in this section. From this analysis, it is clear that South African fake news contains an exaggerated emotional tone.

4.3 Clustering

The next step is establishing clusters that capture certain biases within the South African context. The assumption is that articles written with a particular bias in mind will have the same underlying semantic structure and contain similar words. A few topic analysis techniques were tried, but it was found that K-means clustering gave the best results. This is likely due to the small dataset size. To determine the optimal cluster size, the number of clusters varied from 0 to 25 clusters. Considering that there are roughly 800 South African Fake news articles if the cluster size is too large many clusters end up with only a few articles. It was found that eight clusters gave the most consistent and sensible results.

To help visualise the clustering exercise, a t-SNE plot is used to project the high-dimensional data down onto a two-dimensional graph. See Figure 6 for the t-SNE graph. The chart clearly shows that cluster 2 contains most of the articles. Manual inspection of this cluster reveals that it is the catch-all cluster—fake news articles that do not necessarily contain a particular bias. The more exciting clusters are the ones which do not form part of cluster 2. These are the clusters where the articles show a clear bias. See Figure 7 for the word clouds of cluster 1 and cluster 4.

Inspection of the word clouds and some of these articles' texts reveal a clear bias. The articles within the first cluster are all related to a bias that only certain races are victims of crime. The articles related to the fourth cluster are all about how Western medicine is wrong and that traditional medicine should be used instead. Naturally, this set of articles also contains several unnecessary fear spreading related to the Covid pandemic. With the fake news articles separated into clusters with a particular bias, the final step is to identify the emotions of the articles within these clusters.

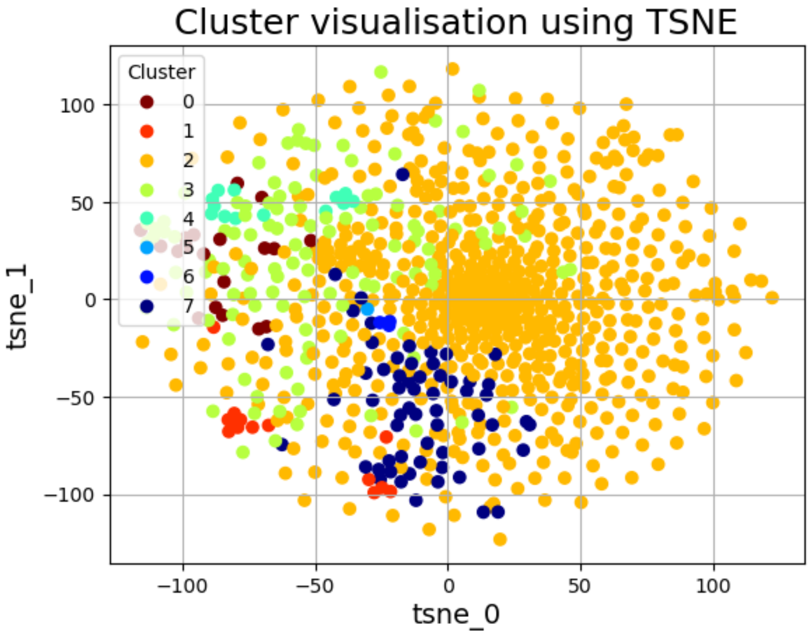

Figure 6: The t-SNE plot visualises the different clusters on a 2D graph.

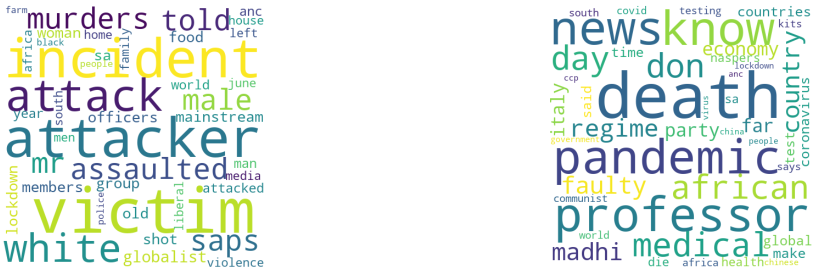

Figure 7: The word cloud for cluster 1(left) and cluster 4(right).

4.4 Transformers

Establishing an emotional profile for each cluster representing a particular bias is necessary. It needs more than a basic positive and negative sentiment in this context. Fortunately, the problem of classifying emotions has been studied before, and a pretrained transformer model is available in the Huggingface library that does this. Google's T5 transformer has already been trained on recognising emotions in free text. It has been pre-trained to recognise anger, fear, joy, love, sadness and surprise. This transformer takes free text as input and predicts the free-text article's overall emotion. Therefore, each article in each cluster can be passed through the transformer model. The total emotional response for each of the feelings in each cluster can be summed. Each cluster representing a particular bias will now have a total score for each emotion the transformer model predicts. These total emotional scores can be used to profile a specific bias. See Figure 8 and Figure 9 for the emotional profiles of the different clusters.

Figure 8 and Figure 9 represent the same data; it visualises it differently. What is important to notice from these graphs is that each cluster has a distinct emotional profile. Considering cluster 1,

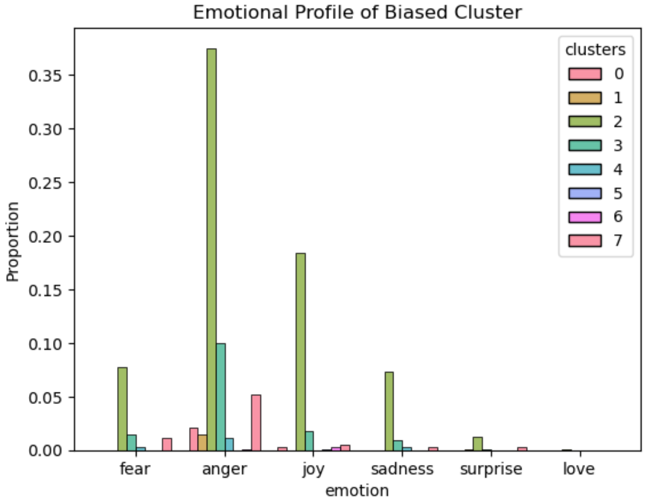

Figure 8: The emotional profiles of the biased clusters.

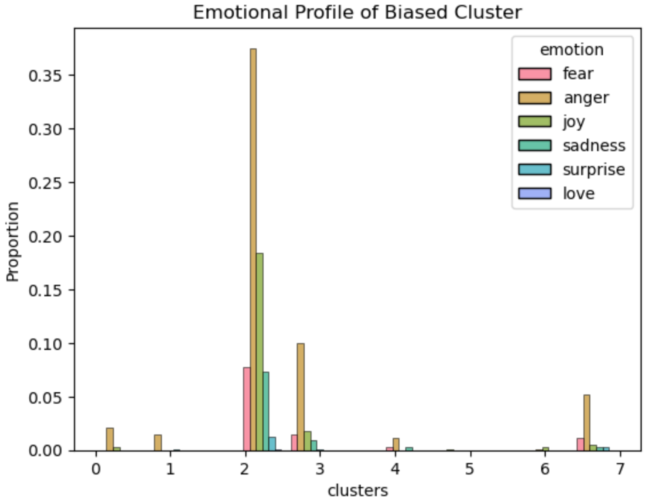

Figure 9: The biased clusters with different emotional scores.

highlighted earlier with the word cloud, it can be seen the emotional profile of this cluster mainly contains anger. In contrast, cluster 4 consists of fear, anger and sadness. Widespread anger is the most dominant emotion in South African fake news.

5 Conclusion

South African fake news was explored using the latest NLP techniques. The goal was to establish whether or not a sentiment analysis of South African news reveals an exaggerated emotional tone and if this emotional tone can be used to detect a bias. A dataset of South African fake news articles and AGNews articles was constructed to help facilitate the experiment. The first step of the analysis involved using the lexicon VADER model and logistic regression model to establish a new expected emotion score. Combined with the VADER score, this score was aggregated and shown to be much higher for fake news articles than non-fake news articles. This implies that South African fake news has a higher exaggerated emotional tone than non-fake news. The clustering of the data separated different biased articles into the same clusters. However, the biased topics are abstract in this context. Once the biases were divided, transformer models were used to detect the overall emotion of each article related to a specific bias. It was found that many South African fake news articles have a strong anger emotion associated with them. Each cluster represented different biases and had different and distinct emotional profiles.

This paper opens up the research to further deep-dive into why anger has such a strong emotive connection in South African fake news. The other emotions worth deep-diving include joy, fear and sadness. The research can also be expanded to include more global data sets that will help to generalise the results better. Finally, if the local fake news data is expanded and labelled, a better fit-for-purpose transformer model can be constructed to provide more predictive power and insights.

Combining the various analyses helps us better understand the appeal behind fake news articles in South Africa. These insights can be used as inputs into future fake news detectors. Alternatively, the insights obtained can also help inform how to write more appealing non-fake news articles that appeal to more people. These applications help us pave a better way to create a more inclusive society.

6 Acknowledgements

We would like to acknowledge funding from the ABSA Chair of Data Science, the TensorFlow Award for Machine Learning Grant. Many thanks to all the anonymous reviewers for their valuable feedback. We also want to acknowledge the assistance of different members of the Data Science for Social Impact research group.

References

[1] Workshop on online abuse and harms - 2021 Shared Task on Hateful Memes, September 2022. [Online; accessed 30. Sep. 2022]. URL: https://www.workshoponlineabuse.com/past-workshops/woah-2021-website/2021-shared-task-on-hateful-memes.

[2] Charu C. Aggarwal. *Mining Text Data*, pages 429–455. Springer International Publishing, Cham, 2015. doi:10.1007/978-3-319-14142-8_13.

[3] Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. Sentiment aware fake news detection on online social networks. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2507–2511, 2019. doi:10.1109/ICASSP.2019.8683170.

[4] Miguel A. Alonso, David Vilares, Carlos Gómez-Rodríguez, and Jesús Vilares. Sentiment analysis for fake news detection. *Electronics*, 10(11), 2021. URL: https://www.mdpi.com/2079-9292/10/11/1348.

[5] Bhavika Bhutani, Neha Rastogi, Priyanshu Sehgal, and Archana Purwar. Fake news detection using sentiment analysis. In *2019 Twelfth International Conference on Contemporary Computing (IC3)*, pages 1–5, 2019. doi:10.1109/IC3.2019.8844880.

[6] Anton Borg and Martin Boldt. Using vader sentiment and svm for predicting customer response sentiment. *Expert Systems with Applications*, 162:113746, 2020. doi:10.1016/j.eswa.2020.113746.

[7] Sijing Chen, Lu Xiao, and Akit Kumar. Spread of misinformation on social media: What contributes to it and how to combat it. *Computers in Human Behavior*, 141:107643, 2023. doi:10.1016/j.chb.2022.107643.

[8] Harm de Wet and Vukosi Marivate. Is it fake? news disinformation detection on south african news websites, 2021. doi:10.48550/ARXIV.2108.02941.

[9] Shihab Elbagir and Jing Yang. Twitter sentiment analysis using natural language toolkit and vader sentiment. In *Proceedings of the international multiconference of engineers and computer scientists*, volume 122, page 16, 2019. URL: https://www.iaeng.org/publication/IMECS2019/IMECS2019_pp12-16.pdf.

[10] Xia Hu and Huan Liu. *Text Analytics in Social Media*, pages 385–414. Springer US, Boston, MA, 2012. doi:10.1007/978-1-4614-3223-4_12.

[11] Y. Linlin Huang, Kate Starbird, Mania Orand, Stephanie A. Stanek, and Heather T. Pedersen. Connected through crisis: Emotional proximity and the spread of misinformation online. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, page 969–980, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2675133.2675202.

[12] Navakanth Reddy Naredla and Festus Fatai Adedoyin. Detection of hyperpartisan news articles using natural language processing technique. *International Journal of Information Management Data Insights*, 2(1):100064, 2022. doi:10.1016/j.jjimei.2022.100064.

[13] Kartikey Pant, Tanvi Dadu, and Radhika Mamidi. Towards detection of subjective bias using contextualized word embeddings. In *Companion Proceedings of the Web Conference 2020*, WWW '20, page 75–76, New York, NY, USA, 2020. Association for Computing Machinery. doi:10.1145/3366424.3382704.

[14] Shivangi Singhal, Rajiv Ratn Shah, Tanmoy Chakraborty, Ponnurangam Kumaraguru, and Shin'ichi Satoh. Spotfake: A multi-modal framework for fake news detection. In *2019 IEEE Fifth International Conference on Multimedia Big Data (BigMM)*, pages 39–47, 2019. doi:10.1109/BigMM.2019.00-44.

[15] Timo Spinde. An interdisciplinary approach for the automated detection and visualization of media bias in news articles. In *2021 International Conference on Data Mining Workshops (ICDMW)*, pages 1096–1103, 2021. doi:10.1109/ICDMW53433.2021.00144.

[16] Amirsina Torfi, Rouzbeh A. Shirvani, Yaser Keneshloo, Nader Tavaf, and Edward A. Fox. Natural language processing advancements by deep learning: A survey. *CoRR*, abs/2003.01200, 2020. URL: https://arxiv.org/abs/2003.01200.

[17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *CoRR*, abs/1706.03762, 2017. arXiv:1706.03762.

[18] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3):554–559, 2016. doi:10.1073/pnas.1517441113.

[19] Jianwei Zhang, Yukiko Kawai, Shinsuke Nakajima, Yoshifumi Matsumoto, and Katsumi Tanaka. Sentiment bias detection in support of news credibility judgment. In *2011 44th Hawaii International Conference on System Sciences*, pages 1–10, 2011. doi:10.1109/HICSS.2011.369.