Measuring Misinformation Trends on Social Media in South Africa using Machine Learning

Jabu Mtsweni
Information and Cyber Security Centre
Council for Scientific and Industrial Research/Stellenbosch University
Pretoria, South Africa
mtswenij@gmail.com

Lungisani Ndlovu
Information and Cyber Security Centre
Council for Scientific and Industrial Research
Pretoria, South Africa
lndlovu3@csir.co.za

Sthembile Mthethwa
Information and Cyber Security Centre
Council for Scientific and Industrial Research
Pretoria, South Africa
smthethwa@csir.co.za

Nenekazi Mkuzangwe
Information and Cyber Security Centre
Council for Scientific and Industrial Research
Pretoria, South Africa
nmkuzangwe@csir.co.za

Abstract*—Misinformation, disinformation, malinformation, and/or fake news have gained attention for good and bad in South Africa, especially since the COVID-19 pandemic. The research-based and non-research-based interventions to tackle misinformation have also been slowly gaining traction, particularly through fact checkers, fake news reporting systems such as those by real411, research on automated systems to detect fake news online using machine learning, sentiment analysis of fake news, tagging of fake news data, and so on. Nevertheless, the spread of misinformation and/or fake news still presents a serious threat and challenge to social media platform owners, citizens, lawmakers, governments, and businesses alike. The awareness, engagement, influence, and impact levels of misinformation on citizens, politicians, journalists, and lawmakers are hypothesized to be relatively low, especially in South Africa. However, no sufficient research has been done in this area to understand engagements, awareness, and reporting of fake news online. This research, therefore, uses open-source intelligence and selected machine learning techniques to analyses publicly collected social media data to monitor and measure the awareness and engagements of fake news in South Africa over 30 days. The research further identifies key drivers of spreading or reporting misinformation online. The research concludes that misinformation engagements on social media in South Africa are active, but only in affluent regions and influenced by mobile device users, who are mostly male. The study recommends further research that may support raising misinformation awareness and positive engagements on social media.

Keywords*—*misinformation, disinformation, social media, natural language processing, machine learning

I. Introduction

It is accepted that fake news is a threat and challenge across the globe for governments, businesses, and citizens. It is also acknowledged that misinformation is an old-age trick used in politics, business, media, and social engagements. The term "fake news" gained its popularity after the United States (US) 2016 elections [1]. This challenge is complicated by the increasing usage of the internet, social media, and penetration of mobile phones across the world. Thus, increasing the potential reach when sharing unverified information [1].

On social networks, there is a large amount of misinformation perpetuated by various global events such as pandemics (e.g., COVID-19), disasters, elections, and various national discourses [2]. However, it is also evident that the understanding of fake news and/or misinformation is still not solid online. This emphasized the urgent need for effective strategies to combat misinformation in today's digital age. In this research paper, therefore, we provide our understanding and characteristics of the differences between misinformation, disinformation, and malinformation.

In our view, *misinformation* can be defined as misleading information that has a basis in false context, omissions, and/or inaccuracies. Generally, misinformation would be spread without any harm intended. *Disinformation*, on the other hand, has similar traits as misinformation but includes deliberately manipulated or fabricated content to cause harm and/or gain an unfair advantage. We also define another category referred to as *malinformation* that is based on some reality used in a false context to malign an individual, social group, or brand including perpetuating hate speech, harassment, and deliberate attacks. The term *propaganda* is also used often in the context of misinformation. It is defined by Britannica[1] as the "systematic dissemination of information such as facts, arguments, rumors, half-truths, or lies to influence public opinion".

These definitions are in line with the work of [3] [4]. There is still a debate about the term "fake news", and mostly driven by media because of the wrong connotation that the term has towards real news. However, fake news, in the context of this paper can be classified as any of the three categories of false information as defined above. In this paper, we use these terms intangibly, but our focus is generally on the spread of misinformation.

As misinformation evolves, it is even difficult to isolate those who spread it and those who try to contain it. The reason is that misinformation can generally be spread (advertently and/or mistakenly) by automated bots, trusted sources (e.g., news websites), politicians, celebrities, and ordinary citizens. The rise of generative artificial intelligence using enhanced learning language models makes it even easier for fake information to be generated and propagated online. Misinformation/fake news generated using these models proves to be realistic, and persuasive, thus increasing the difficulty of identifying fake news [5].

[1] https://www.britannica.com/topic/propaganda

The main objective of this paper is to investigate the engagement, spread, awareness, reporting, and impact of misinformation on social media. This research is conducted by adopting general machine learning techniques, such as topic modeling, sentiment analysis, and k-means clustering, to analyze social media data collected over 30 days.

The remainder of this paper is structured as follows: Section II deals with related work whilst Section III details the research methodology adopted for this research. In Section IV, the data analysis is conducted as well as the presentation of the results. In the same section, the implications of the results and recommendations are discussed. The paper is concluded in Section V.

II. Related Works

In South Africa, research on fake news and/or misinformation is emerging, particularly post-COVID-19 pandemic. The understanding of the appeal for fake news was studied in [2] by conducting sentiment analysis using South African Disinformation [Fake News] Website Data from 2020. They found that fake news contains strong emotive words and biases, and these findings could be of importance when developing tools to automatically flag or detect fake news on social media or online news websites. In addition, the impact of fake news on social cohesion is highlighted in a study by [6]. In their research, they argue that fake news in South Africa is spread by the proliferation of social media and provide a narrative on the xenophobic (anti-immigrant sentiment) attacks influenced by fake news.

In [3], a deep learning method was adopted to perform content and sentiment analysis on misinformation and true information datasets to understand the network diffusion characteristics. Their results suggest that the spread of misinformation on social media is influenced by content and emotions, aligning with the study by [2]. In Kenya, related studies have also been conducted from the qualitative perspective. In this regard, [7] conducted a focus group study by exploring how Kenyans experience and deal with misinformation. The study findings, interestingly, suggest that citizens react to misinformation based on personal interest in the topic, perceived resonance within their social networks, and perceived importance. [4] studied fake news and media trust in Kenya, Nigeria, and South Africa, and found that the perceived exposure to disinformation is high with the trust in social and national media being low, with South Africa showing a high correlation between exposure to disinformation and lower levels of media trust.

Authors [8] conducted research on misinformation in sub-Saharan Africa and provided more in-depth empirical data. Their findings revealed several key insights about the prevalence and sharing practices of misinformation in the region. Some of the findings of the study were that misinformation was common and a significant concern in sub-Saharan Africa, younger generations tend to blame older media users for sharing misinformation, young media consumers in sub-Saharan Africa were discerning users who employed various cues to evaluate information and some participants shared political misinformation intending to advance their political motives or ridicule those in power. The latter aligned with the prevalent use of satire and humor for political purposes in African media and political communication.

Furthermore, the study highlighted that due to the long-standing role of such practices in informal networks of media use, users in sub-Saharan Africa might have been less resistant to sharing information they knew was untrue. This indicates that certain cultural factors and social dynamics contribute to the spread of misinformation in the region.

In [9], they explored the relationship between age, social media political information seeking, perceived exposure to misinformation, and online political engagement, and examined how age moderates the relationship between social media political information seeking, perceived exposure to misinformation, and online political engagement in Kenya, Nigeria, and South Africa. Their investigation suggested that perceived exposure to misinformation and information-seeking have a positive relationship with online political engagement. Moreover, it was found that the effect of perceived exposure to misinformation on political engagement varies across age groups.

They collected data through an online survey, asking participants about their online political participation, information search behavior, and perception of exposure to misinformation on social media platforms. They used statistical analysis to study the relationship between these variables and the differences between age groups. However, online surveys can lead to bias towards urban, highly educated, and young respondents, further restricting the generalization of the findings to the wider population of these sub-Saharan African countries. Consequently, they proposed conducting face-to-face or mobile surveys to obtain a more representative sample of the population and further investigating more sophisticated measures of online political participation to gain a deeper understanding of the reasons for the observed relationships between information seeking, information exposure, and political participation.

Nistor and Zadobrischi [10] conducted a study focusing on the impact of social networks on generation and spreading fake news, highlighting the dangers to society relative to traditional media. The authors adjusted the analysis and machine learning classifiers to identify and expose false news at a success rate of 90%, limiting user interaction with misinformation. They developed browser extensions to prevent users from interacting with distorted information. The researchers proposed future developments, including graphical user interface (GUI)-type applications and completion of extension synchronization with social network application programming interfaces (APIs). The purpose of the study was to highlight the need for more effective models to combat negative aspects of fake news that can be avoided by current algorithms.

The effect of fake news awareness was tested in [11] as an intervention strategy for motivating the verification of news before sharing among Nigerian social media users. The study conducted a quasi-experiment with 470 participants where half of the participants were a control group, and the other half of the participants were a treatment group (exposed to the intervention). The study found that fake news awareness was an effective intervention strategy to motivate the verification of news before sharing.

Based on related research, it is apparent that the issue of misinformation is topical, quite broad, and diverse with impacts at various levels in society. It is also evident that this issue is tackled from different angles, from journalism, and political studies, and to technical research using varying techniques such as machine learning, quantitative and qualitative analysis. However, none of these studies concentrated on determining the extent of misinformation awareness in South African society.

III. Methodology

A. Data Collection

The data for this research was collected using data-scraping aided by the Talkwalker[2] platform over a month (26 June – 27 July 2023) from Twitter. Before the preprocessing of the data, over *19,440* records using "*South Africa*" and "*English*" as main filters were collected using non-case-sensitive keywords: *fake news*, *misinformation*, *disinformation*, *propaganda*, *and conspiracy theory*. The 30-day dataset was chosen to avoid collecting data over a longer period, which would require more resources while ensuring that we work with recent data. The data contained tweets and associated metadata such as date, author names and descriptions, followers, engagement score, source, regions, gender, and others. The hit rate of keywords used for searching was analyzed and Fig. 1 below shows how the results were distributed.

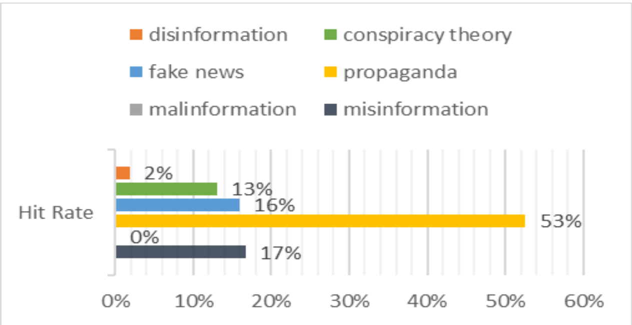

Fig. 1. Search Keywords Hit Rate.

B. Data Cleaning and Preprocessing

The raw data was cleaned and pre-processed for misplaced data (e.g., data outside South Africa), noise control using natural language handling stop-words [12] and removing text that was seen as having special characters, links, common words not part of stop-words that were observed in the social media data and punctuation. In addition, text that was observed as *"spam", "marketing", and "offensive"* was discarded from the analysis. On the other hand, the text was converted into lowercase, removed text in square brackets, removed punctuation, and removed words containing numbers.

In exploring the data, descriptive statistics were done on columns of interest including the analysis of the occurrences of words. These are explained in the next section.

IV. Data Analysis and Results

A. Exploratory data analysis

Exploratory data analysis refers to an approach used to analyze data sets, summarizing their primary characteristics by frequently employing statistical graphics and various data visualization techniques [13]. This method helps identify evident errors, gain a deeper understanding of patterns within the data, detect outliers or anomalous events, and reveal intriguing relationships among the variables [14]. In this study, we performed interactive visualization where tools like word cloud, plotly, etc. were used to enable the creation of interactive visualizations that allowed us to explore data dynamically.

The exploration of data analysis includes determining the most common or top words within the data, and as shown in Fig. 2, "*propaganda*" and "*misinformation*" are seen as the most common words, which somehow does give us insight that many social media (Twitter) users are aware of misinformation.

Fig. 2. Occurrences of Keywords.

B. Comprehensive Data Analysis

In analyzing the misinformation engagement trends in South Africa, the focus of our observations was on the *trend analysis of misinformation engagements, topic modeling, sentiment analysis, emotions analysis, gender distribution, source of content, region distribution, and the relationship between engagements and followership* using clustering.

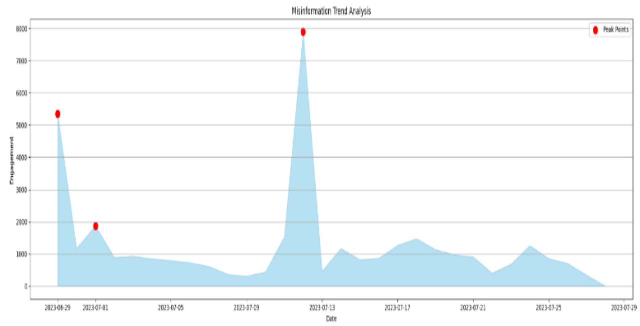

Fig. 3. Misinformation Engagement Trends in South Africa.

From the data collected for 30 days, it is evident in Fig. 3 that misinformation engagements hover around 1000 and as may be noted, peaked misinformation engagements tend to be triggered by certain events such as political activities, historical debates (which was the case for the 2023-07-20), and topical national discourse. As depicted in Fig. 4, the sentiment analysis mostly reflects a negative trend, primarily driven by users disseminating misleading information or not taking kind to the spread of misinformation, consequently leading to an abundance of negative sentiments. We compared Text Blob and Valence Aware Dictionary and Sentiment Reasoner (VADER) and decided to go with the VADER results due to its flexibility and popularity [15].

[2] https://app.talkwalker.com/

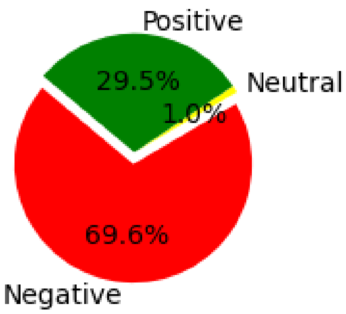

Fig. 4. Sentiment Analysis.

Based on the analysis, males dominate the engagements on misinformation and associated on social media by over 60%. This confirms the results of the study [16] and is supported by Twitter stats [17] that males are more inclined to adopt Twitter than females who are more frequent users of Facebook. This also explains the exaggerated emotional tones we observed in some of the data as well as strong debates that females would generally not be involved in male-dominated societies such as South Africa.

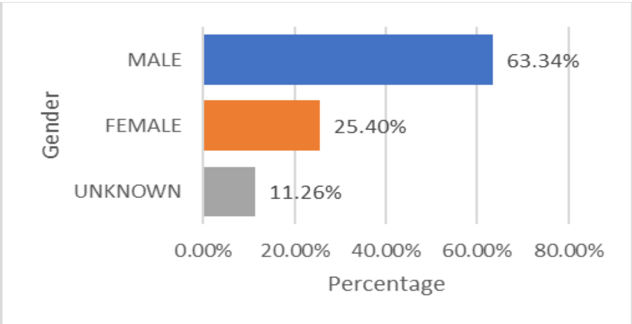

Fig. 5. Distribution of Engagement by Gender.

Other descriptive statistics of interest include that most users engaging in misinformation, at least as observed during the research period, use mobile devices with mobile Android covering over 55% followed by mobile iOS (26%). Automated bots played a minimal role (0.25%) in the spreading or alerting of misinformation. Lastly, it was analyzed that the Western Cape (47%) and Gauteng (34%) regions dominate the social media engagements on misinformation.

We applied unsupervised machine learning techniques in topic modeling to determine the key topics that dominate the misinformation engagement in South Africa. In [18], topic modeling is defined as a popular technique in natural language processing used to discover hidden topics or themes within a collection of texts. The research in [19] indicates that it is one of the most powerful techniques for text mining, data mining, finding hidden data, and finding relationships between data and text documents.

Some of the common algorithms for topic modeling are Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) [20]. LDA is a technique that classifies the data based on the content and by giving labels to each category [20]. This method automatically generates hidden topics using a discrete probability distribution on words. On the other hand, [20] also confirms that NMF is another technique widely used in various fields, especially the field of artificial intelligence, signal processing, and bioinformatics. The advantage is that it imposes non-negativity on factorized matrices and well matches the idea that text data topics are words mixed and should not have negative values. This method works on TF-IDF transformed data by breaking down a matrix into two lower-ranking matrices [21].

In this study, a combination of LDA and NMF was applied to derive the topics and themes of interest from the misinformation engagements data collected for South Africa. The modeling techniques were applied separately on the cleaned content by creating a dictionary and a corpus necessary for LDA and NMF.

In the word clouds below, we depict the generated topics using the LDA. The topics were vectorized based on *min_df <10 and max_df of 80%*. We discarded high-appearing words (80%) as we found them to be too common to be meaningful in topics. We also discarded low-appearing words (10 or less occurrences) for noise control.

Fig. 6. Results of the LDA Topic Modelling.

Thereafter, the two results from LDA and NMF were compared and analyzed further to derive concrete topics that form part of the misinformation engagements in South Africa.

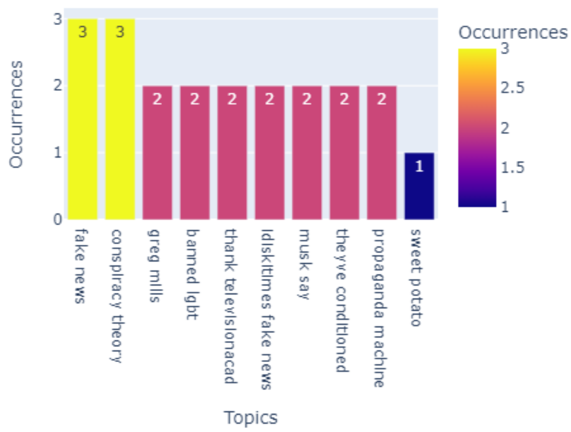

Fig. 7. LDA and NMF Combined Occurrence of Topics.

As shown in Fig. 7, common topics such as *conspiracy theory and fake news* clearly show that most users are aware of misinformation. In addition, topic modeling also points out that *@idiskitimes*, an online sports news publisher is alleged to be disseminating fake news or misinformation based on engagements and complaints by the users on social media during the research period.

Using k-means clustering, the relationship between the engagement score *(the sum of impressions such as likes, comments, and retweets on the article)* and the number of followers for the users' sharing information was analyzed and plotted as depicted in Fig. 8. The data between followers and engagements was normalized to a standard scale, rows with zero engagements and/or followers dropped. Outliers were also handled using the Z-score. The clustering is shown in a log scale.

Cluster 0 demonstrates aspects of a power-law relationship where straight dotted lines are observed at the bottom, suggesting that with certain engagements of misinformation, the relationship holds across different sizes of followers. Within the same cluster, there is also a strong relationship between engagement and followers. This also indicates a strong positive relationship between the number of followers and engagements in this cluster. Cluster 1 also indicates a positive relationship with increasing engagements whilst followers remain within similar ranges. Cluster two suggests a strong negative correlation with engagements decreasing as the number of followers increases. This may be viewed as an outlier, but as indicated above, for this dataset outliers were taken care of.

C. Discussions

Social media engagements happen more frequently in this digital age with over 450 million monthly active users as of July 2023 [17]. This provides opportunities for researchers and decision-makers to study the trend of misinformation continuously. The dataset relied upon for our analysis was large (with under 20k records), however, these come with limitations such as duplications due to retweets, which can cloud the topics as well as other analyses if not well handled.

Based on this research, it can be deduced that the level of engagement relating to misinformation in South Africa is active and polarized with negative sentiments, supporting the conclusion of [2]. However, it may be premature for us to conclude whether the engagements are high or low without comparing it with other nations.

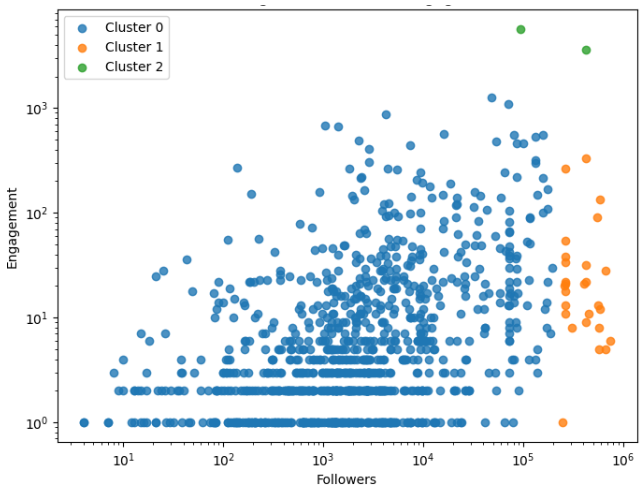

Fig. 8. Clustering of Misinformation Engagements vs. Followers.

Nevertheless, what is obvious is that the awareness of misinformation is evident in the spreading or alerting of such information on social media. The engagements are driven by the male gender, and this is not surprising in a male-dominated society such as South Africa where female-headed households are mostly disadvantaged due to the imbalances of the past. The misinformation engagements are also driven by users with mobile devices, and this is not surprising, considering that South Africa has a mobile usage penetration of over 100%, and smartphone usage is above 50% [22]. This is further supported by the analysis that most of the people engaged in misinformation in South Africa are in affluent regions such as Western Cape and Gauteng, meaning they have access to smartphones. KwaZulu Natal, which has the second highest population in South Africa featured less than 10% in the misinformation engagement on social media during the 30 days of the study. This may be pointing to a lack of awareness of misinformation (spreading or alerting) in that region. This conclusion is supported by the data from Statistics South Africa [23] which indicated that KwaZulu Natal is amongst the top 3 provinces with the highest levels of poverty. In the same report, Gauteng and Western Cape had the lowest proportion of adults living in poverty. Nevertheless, one also needs to be cautious since only a limited number of tweets are geolocated, and different social media platforms attract different users and genders [16].

Using a combination of NMF and LDA topic modeling techniques, we observed that national and international political and government events tend to drive engagements on this subject. For example, the Russia-Ukraine war features in several misinformation engagements in South Africa. Former President Zuma's name also features now and again on engagements related to misinformation and propaganda. The national crisis such as loadshedding does not feature that much, but reference to power stations is observed in our topic analysis. Misinformation and conspiracy theories around national disasters tend to also feature when using the NMF topic modeling.

Contrary to general expectations, the spread or alert of misinformation is not highly driven by users with a large following, but a mixture of users (low and high). This may require further analysis to understand the relationship deeper as our analysis indicates that there is a weak correlation between the number of followers and engagements.

Based on the results, further research is recommended to understand the awareness of misinformation in South Africa compared with other countries, particularly with similar contexts. In addition, the relationship between the users posting about misinformation and followers should be analyzed to get social graph patterns. This study did not focus on whether the information analyzed was fake or not but used keywords to measure the engagements on the topic. It is therefore recommended that the collected data be annotated for training a machine learning model that could automatically label and detect misinformation on social media.

V. Conclusion

This research study employed open-source intelligence and selected machine learning techniques to analyze publicly collected social media data to monitor and measure the awareness and engagements of fake news in South Africa over 30 days. This analysis included the identification of key drivers of spreading or reporting misinformation online.

Based on the research, the engagements trend on misinformation is influenced by specific events or conversations, mostly political and historical that tend to have a high negative sentiment and exaggerated emotions. The misinformation engagements in South Africa are found to be highly driven by males using mobile devices and geolocated in the Western Cape and Gauteng regions. Using a combination of topic modeling techniques, we determined that *conspiracy theory, propaganda, and fake news* are the topics that dominate misinformation engagements. Lastly, contrary to expectations, users with a large number of followers do not necessarily drive engagements but could be a vehicle for the propagation or awareness of misinformation. However, we did not analyze the relationship between users posting about misinformation and followers. Furthermore, we did not focus on whether the information analyzed was fake or not, instead, we used keywords to measure the engagements on the topic.

The study proposes further research that may support raising misinformation awareness and positive engagements on social media.

Acknowledgments

We would like to thank the Department of Science and Innovation for their support in the National Policy Data Observatory. We thank our collaborators for making the data available for this research work.

References

[1] C. S. Leon, M. Bonilla, L. I. Brusco, C. Forcato and F. U. Benitez, 2023.

[2] M. Gagiano and V. Marivate, “Emotionally driven fake news in South Africa,” in *Proceedings of Society 5.0 Conference 2023,* 2023.

[3] H. Luo, M. Cai and Y. Cui, “Spread of Misinformation in Social Networks: Analysis Based on Weibo Tweets,” *Security and Communication Networks,* vol. 21, pp. 2-23, 2021.

[4] H. Wasserman and D. Madrid-Morales, “An Exploratory Study of “Fake News” and Media Trust in Kenya, Nigeria and South Africa,” *African Journalism Studies,* vol. 40, no. 1, pp. 107-123, 2019.

[5] J. Zhou, Y. Zhang, Q. Luo, A. Parker and M. Choudhury, “Synthetic Lies: Understanding AI-Generated Misinformation and Evaluating Algorithmic and Human Solutions,” in *2023 CHI Conference on Human Factors in Computing Systems (CHI '23),* New York, USA, 2023.

[6] V. Chenzi, “Fake news, social media and xenophobia in South Africa,” *African Indentities,* vol. 19, no. 4, pp. 502-521, 2021.

[7] M. Tully, “Everday news use and misinformation in Kenya,” *Digital Journalism,* vol. 10, no. 1, 2022.

[8] D. Madrid-Morales, H. Wasserman, G. Gondwe, K. Ndlovu, E. Sikanku, M. Tully, E. Umejei and C. Uzuegbunam, “Motivations for sharing misinformation: A comparative study in six Sub-Saharan African countries,” *International Journal of Communication,* vol. 15, no. 2021, pp. 1200-1219, 2021.

[9] S. Ahmed, D. Madrid-Morales and M. Tully, “Social media, misinformation, and age inequality in online political engagement,” *Journal of Information Technology & Politics,* vol. 20, no. 3, pp. 269-285, 2023.

[10] A. Nistor and E. Zadobrischi, “The influence of fake news on social media: analysis and verification of web content during the COVID-19 pandemic by advanced machine learning methods and natural language processing,” *Sustainability,* vol. 14, p. 10466, 2022.

[11] O. D. Apuke, B. Omar and E. A. Tunca, “Effect of fake news awareness as an intervention strategy for motivating news verification behaviour among social media users in Nigeria: a quasi-experimental research,” *Journal of Asian and African Studies,* 2022.

[12] M. Anandarajan, C. Hill and T. Nolan, “Text Preprocessing,” *Practical Text Analytics. Advances in Analytics and Data Science,* vol. 2, pp. 45-59, 2019.

[13] K. Shoo, A. K. Samal, J. Pramanik and S. K. Pani, “Exploratory data analysis using Python,” *International Journal of Innovative Technology and Exploring Engineering,* pp. 4727-4735, 2019.

[14] R. W. Cooksey, “Descriptive statistics for summarising data,” in *Illustrating Statistical Procedures: Finding Meaning in Quantitative Data*, Singapore, Springer, 2020, pp. 61-139.

[15] E. Shihab and J. Yang, “Twitter sentiment analysis using natural language toolkit and VADER sentiment,” in *Proceedings of the international multiconference of engineers and computer scientists,* 2019.

[16] A. Budree, K. Fietkiewicz and E. Lins, “Investigating usage of social media platforms in South Africa,” *The African Journal of Information Systems,* vol. 11, no. 4, 2019.

[17] D. Ruby, “Twitter “X” Statistics In 2023 — (Users, Revenue & Trends),” 11 July 2023. [Online]. Available: https://www.demandsage.com/twitter-statistics/. [Accessed 30 July 2023].

[18] K. Taghandiki and M. Mohammadi, “Topic Modeling: Exploring the Processes, Tools, Challenges and Applications,” *TechRxiv,* 2023.

[19] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li and L. Zhao, “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” *Multimedia Tools and Applications,* vol. 15211, no. 78, p. 15169, 2019.

[20] d. P. F. Thiago and A. d. A. Lopes, “On the equivalence between algorithms for Non-negative Matrix Factorization and Latent Dirichlet Allocation,” *ESANN,* p. 6, 2016.

[21] TF-IDF, “Encyclopedia of Machine Learning,” *Encyclopedia of Machine Learning,* pp. 986-987, 2011.

[22] C. Memela and E. Chindega, “Emerging FinTech Models in Mobile Payments for Digital Finance in Rural Areas,” in *2023 IST-Africa Conference (IST-Africa)*, Pretoria, 2023.

[23] StatsSA, “Five facts about poverty in South Africa,” https://www.statssa.gov.za/?p=12075, Pretoria, 2019.