The Role of Social Media in Xenophobic Attack in South Africa

*Raborife, Mpho**
†*University of Johannesburg, South Africa.*
mraborife@uj.ac.za

*Ogbuokiri, Blessing**,†
York University, Toronto, Canada.
blessogb@yorku.ca

Aruleba, Kehinde†
University of Leicester, United Kingdom
ka388@leicester.ac.uk

* *Corresspondence*

Abstract

Xenophobia is a pressing issue in South Africa, with frequent instances of violence against immigrants. With the rise of social media, platforms like Twitter reflect public sentiment on this matter. This study examines tweets from 2017 to 2022 about xenophobia in South Africa, using NLP, sentiment analysis, and machine learning to understand public feelings and predict potential xenophobic incidents. The findings aim to help policymakers devise strategies to enhance social cohesion and promote a more inclusive society.

Keywords: Xenophobia, Xenophobic Attack, Twitter, Machine Learning, Sentiment Analysis.

1 Introduction

In Mogekwu (2005), xenophobia is defined as the fear or hatred of foreigners, expressed through discriminatory attitudes and behaviours, often culminating in violence and other forms of hatred. In South Africa, xenophobic violence against foreign nationals has increased since 1994, with factors like inadequate service delivery and resource competition exacerbating the issue (Gumede 2015, Harris 2001). Notably, this phenomenon predominantly affects immigrants of African descent, prompting a shift in terminology from xenophobia to Afrophobia (Dube 2018), as reflected in traditional media (Tarisayi & Manik 2020*a*).

Our study focuses on attitudes of South Africans on Twitter and their connection to xenophobia and xenophobic attacks. While traditional media remains a primary news source, social media's growing role in news consumption is evident (Chingwete et al. 2018). Twitter particularly serves as a stimulus for xenophobic narratives and has been used to analyse sentiments on societal issues (Tarisayi & Manik 2020*a*). Research has examined Twitter discourse on xenophobic attacks (van der Vyver 2019, Pontiki et al. 2020). For instance, van der Vyver (2019) analysed 3,784 tweets during the September 2019 Xenophobic attack, highlighting Twitter's potential to reveal insights, including widespread condemnation. In this study we generated a dataset containing 19,700 tweets spanning from January 2017 to December 2022, discussing xenophobia and xenophobic attacks in South Africa.

We manually labelled these tweets as positive, negative, or neutral and employed traditional machine learning classifiers for their classification, while also identifying urban/rural regions and attack counts. Notably, we utilised the Latent Dirichlet Allocation (LDA) model to extract interesting topics from these tweets. These labelled tweets play a pivotal role in enabling sentiment analysis and prediction, which are essential for the management of attacks and the formulation of policies. The remainder of the manuscript is structured as follows: Section 2 presents the background and related works, followed by Section 3 which details our research methodology. Section 4 presents the research findings, and Section 5 delves into the discussion. Finally, the conclusion summarizes our key findings and contributions in Section 6.

2 Background and Related Works

Social media have grown popular in Africa, with over 384 million users on the continent and about 28 million users in South Africa (statista 2023). This is a significant number when considering the digital divide between rural and urban Africa and the uneven technological distribution *on the continent (Aruleba & Jere 2022, Okolo et al. 2023). Over the years, social media applications have been used as a platform to share information and news about xenophobic attacks in South Africa. Users of these applications can easily share images, videos, and news of violent incidents, bringing local and international communities' attention to the issue and raising their awareness. This increased awareness can assist authorities in addressing the issue. The complex interaction of numerous factors, such as economic, political and social issues, influences social media applications' influence on xenophobia in South Africa.

In South Africa, movements and protests such as #FeesMustFall, #OperationDudula and #RhodesMustFall were quick to spread around the country through modern communication technologies such as Twitter, WhatsApp, and Facebook. The #ENdSARS movement against police brutality in 2020 witnessed widespread participation in protests due to the dissemination of information via various social media platforms. The emergence of a fifth estate has been discussed by researchers (Newman et al. 2013) considering the effects brought about by new communication technologies. This phenomenon suggests that the public is becoming more empowered by creating and disseminating user-generated content, thereby reducing their reliance on traditional media outlets, commonly referred to as the fourth estate. Similarly, the emergence of new communication technologies has prompted the internet, and consequently, social media form an alternative public sphere and, in some instances, a counter-public sphere. This is because users engage in discussions on a less regulated platform and express divergent opinions or sentiments that challenge established norms or authorities.

Recent research has shown that online hate speech has contributed significantly to the increasing horrific real-world hate crime (Mathew et al. 2019). The use of social media has enabled and created a conducive environment for the proliferation of misinformation, hate speech and extremist conspiracy theories. Social media platforms have been used to facilitate rapid and extensive information dissemination and offer a wide-ranging discursive framework within which potential perpetrators of violent actions can acquire visibility, resonance, and legitimacy (Williams et al. 2020). The literature also provides evidence that malicious content has a greater reach and spreads more extensively and rapidly than other types of content (Mathew et al. 2019). According to Mathew et al. (2019), posts generated by individuals who exhibit hateful behaviour exhibit a significantly greater dissemination rate due to their dense network connections and substantial contribution to content creation, despite their relatively small population size. The dissemination of false information and the promotion of hateful language on the internet play a crucial role in fostering scepticism and bewilderment within societies while also exacerbating pre-existing divisions based on nationalistic, ethnic, racial, and religious factors (Wardle & Derakhshan 2017, Fokou et al. 2022).

Anti-immigrant violence has been reoccurring in South Africa (Fokou et al. 2022). Since the end of apartheid, threats and acts of xenophobic violence have become common in the country. The tragic events of 2008 and 2015 made international headlines, and some heads of state have expressed concern and protest in a lacklustre manner. However, the international response to the September 2019 outbreak of violence was more pronounced, particularly in countries with a large number of South African citizens. African leaders have been more vocal in condemning the attacks and in their appeal for the South African government to restore peace and security. In some African nations, there were calls for a boycott of South African companies, the cancellation of an international football friendly, and retaliatory attacks (Chenzi 2021). Multiple nations' demonstrations and actions have demonstrated the continent's growing contempt for South Africa.

In South Africa, the production and dissemination of information have significantly influenced people's attitudes towards violent crime and anti-immigrant violence. The coverage of international migration by the South African press has been primarily anti-immigrant (Fokou et al. 2022). By creating and reinforcing ideologies, discourses, and policies concerning cross-border migration and the lives of migrants, the media have contributed to the high levels of xenophobia in South Africa (Danso & McDonald 2001). Twenty years later, this negative function of the media in exacerbating anti-immigrant sentiment has not changed significantly. Even today, the media are accused of misrepresentation, partiality, and unbalanced reporting, primarily regarding xenophobic violence (Mgogo & Osunkunle 2021). In most media coverage of the September 2019 assaults, foreign nationals have been portrayed as drug-related criminals (Tarisayi & Manik 2020*b*).

Hlatshwayo Hlatshwayo (2023) presented the findings from verified Internet news report and YouTube videos and argued that South Africa has entered a second wave of xenophobic violence aimed primarily at black immigrants from other African nations. What distinguishes this wave is the use of various strategies and tactics, such as social media, protest actions and marches to identify or position immigrants as the cause of unemployment, poverty and crime in South Africa by organised formations that are relentless in their campaigns against immigrants. Chenzi (2021) aims to highlight the impact of false news on the xenophobic discourse in South Africa. Despite scholarly neglect, the paper contends that fake news disseminated by social media platforms is becoming an increasingly important aspect of South Africa's contemporary xenophobia challenge. The paper argues further that fake news in South Africa has been primarily fueled by the proliferation of social media platforms, which have recently replaced traditional news sources for a growing number of South Africans despite their obvious flaws.

3 Methodology

3.1 Search Keyword Selection

The selection of search keywords for the data collection was based on the popular words and phrases that consistently formed the discussions during Xenophobic attacks. These words and phrases have become very popular because the locals regularly use them to form native slogans and chants against foreign nationals. Of course, they also form the significant topics that dominate social media, such as Twitter. The topics are represented in the form of hashtags which are tweeted, retweeted and liked by users on Twitter during Xenophobia. These keywords include foreigners, immigrant, operationDudula, and nhlanhla lux. Others are take back sa, take back south africa, dudula, and jellof.

3.2 Data Collection

Geo-tagged historical tweets dated between January 2017 to July 2022, with South Africa being the location of interest, were collected from the Twitter database (Udanor et al. 2022, Mgboh et al. 2021). We created an access token from Python version 3.6 script to authenticate and establish a connection to the Twitter database. The Python script was used to collect tweets that contain keywords relating to Xenophobia, such as the ones identified in Section 3.1. The preferred language of the tweet is all eleven (11) official South African languages, including tweets written in one or a combination of the following languages: Afrikaans, English, Xhosa, Zulu, Southern Sotho, Northern Sotho, Tsonga, Tswana, Venda, Swati, and Ndebele. A total of 15,912 tweets were collected. Each Tweet contains most of the following features in Table 1.

3.3 Data Annotation

The tweets were manually annotated as positive, negative, or neutral. A Tweet is annotated positive if it does not support Xenophobia or Xenophobic attack and negative if it supports Xenophobia

Table 1: Xenophobia Dataset Features

SN	Column name	Description	Type
1	Datetime	The date and time when the tweet was posted	Date
2	Location	The place the tweet originated from	Text
3	Text	A processed post that represents users' opinions	Text
4	Sentiment	A view of or attitude toward a situation or event; an opinion (positive, negative, or neutral)	Text
5	Likes	Number of users interested or agreed to the tweet	Integer
6	Retweets	Number of re-post a tweet gets	Integer
7	Coordinates	The geographic coordinate of the tweet, represented by latitude and longitude	Float
8	Province	The administrative division of the country of the tweet	Text
9	District	The locality of the tweet (urban or rural)	Text
10	Attacks	Number of places affected as a result of xenophobia	Integer

or Xenophobic attack. A tweet that is neither in support nor against Xenophobia is annotated neutral. The districts or regions where the tweets originated from were identified and manually annotated in two classes, namely urban and rural. We also manually identified the number of attacks from the location of the tweets. This was achieved with the services of ten locally trained human data annotators. We used human annotators to avoid the issue of subjectivity of text, the context of the sentiment, and its tone. Additionally, the presence of a sarcastic or ironic statements, emojis, idiomatic expressions or local-based colloquialisms and negations which are difficult for automatic pretrained annotators like VADER (Alsharhan et al. 2022) and Textblob (Praveen & Prasanna 2021, Laksono et al. 2019) to label appropriately were handled correctly by the human annotators.

3.4 Tweet Classification

We used five machine learning classifiers including Naive Bayes (NB) (Kewsuwun & Kajornkasirat 2022), Logistic Regression (LR) (Gulati et al. 2022, Jaya Hidayat et al. 2022), Support Vector Machines (SVMs) (Vasista 2018, Jaya Hidayat et al. 2022), Decision Tree (DT) (Ritanshi et al. 2022, Obaido et al. 2023), and XGBoost (Bachchu et al. 2022, Obaido et al. 2022). We also used a deep learning algorithm known as Long Short-Term Memory (LSTM) (Luan & Lin 2019, Bai 2018). The reason we chose these classifiers is because they have been successfully used to classify tweets in similar studies, such as (Ogbuokiri, Ahmadi, Bragazzi, Movahedi Nia, Mellado, Wu, Orbinski, Asgary & Kong 2022, Ogbuokiri, Ahmadi, Nia, Mellado, Wu, Orbinski, Asgary & Kong 2023, Qi 2020) and (Ogbuokiri, Ahmadi, Tripathi, Movahedi, Mellado, Wu, Orbinski, Asgary & Kong 2022).

The performance evaluation metrics were used as a measure of how well the classifiers classified the hand-labelled tweets according to their sentiments. They include accuracy, accuracy, Receiver Operating Characteristic (ROC), and Area Under the Curve (AUC).

3.5 Data Preprocessing

Raw tweets are very unstructured and contain a lot of redundant information about the data they represent. This redundant information may affect the process of data analysis, which may in turn affect the final output. To achieve an effective data analysis, there is a need to clean up the user tweets. We collected user tweets, date created, time created, and provinces from the dataset into a dataframe using Pandas version 1.2.4 (Li et al. n.d.). The tweets were prepared for Natural Language Processing (NLP) by first removing the URLs, duplicate tweets, tweets with incomplete information, punctuations, special and non-alphabetical characters using the tweets-preprocessor toolkit version 0.6.0 (Ogbuokiri, Ahmadi, Bragazzi, Movahedi Nia, Mellado, Wu, Orbinski, Asgary & Kong 2022), Natural Language Toolkit (NLTK version 3.6.2) (Ogbuokiri, Ahmadi, Nia, Mellado, Wu, Orbinski, Asgary & Kong 2023), and Spacy2 toolkit (version 3.2) (Honnibal 2017, Ogbuokiri, Ahmadi, Mellado, Wu, Orbinski, Asgary & Kong 2023). This process reduced the tweets in the dataset to 15,912 clean tweets.

3.6 Topic Modelling

We used the Latent Dirichlet Allocation (LDA) for topic modelling. LDA is a probabilistic topic modeling technique used to uncover underlying topics in a collection of documents. LDA assumes that each document is a mixture of various topics, and each topic is characterized by a distribution of words. The algorithm helps identify these topics and their corresponding word distributions, allowing for topic-based analysis of text data (Yu & Xiang 2023).

3.7 Test statistics

The Pearson correlation was calculated using the Python package. We utilised the "pearsonr" function from the "scipy.stats" module to analyse the correlation between tweets, likes, and retweets in relation to the occurrences of attacks (Ogbuokiri, Ahmadi, Bragazzi, Movahedi Nia, Mellado, Wu, Orbinski, Asgary & Kong 2022).

4 Results

In this part, we show the outcomes of our analysis. We begin by sharing the key findings from the summary statistics of the tweets based on time and location. Then, we display the results of classifying tweets based on their sentiments (positive, negative, and neutral). Lastly, we reveal the outcomes of our topic modeling analysis.

4.1 Summary Statistics

Table 2 provides a summary of the trends in posts about xenophobia and related attacks. This information is then visualised in Figure 1. We also explored the connection between the rise in tweet counts and the overall number of attacks by provinces. The results are shown in Figure 2. Specifically, we focused on two provinces, Gauteng (Corr=0.73, P=0.03) and Kwazulu-Natal (Corr=0.81, p=0.02), as their outcomes exhibited stronger and statistically significant patterns compared to other provinces, especially from 2021.

Table 2: Yearly xenophobia posts with numbers of attacks

SN	Year	Tweets	Likes	Retweets	Attacks
1	2017	641	762	467.0	328
2	2018	858	5078	2207.0	13
3	2019	2576	14232	5682.0	63
4	2020	3536	33457	9639.0	31
5	2021	2391	45750	10363.0	58
6	2022	5902	78687	19747.0	315

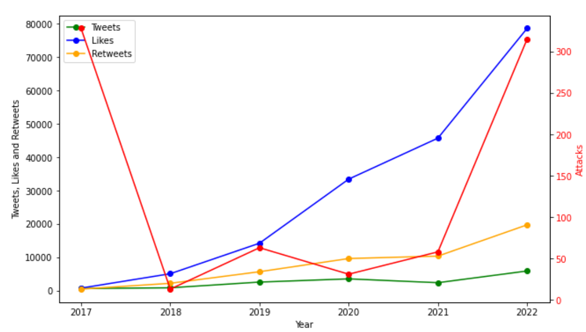


Figure 1: Xenophobia yearly tweets, likes, retweets, and attacks

Moreover, we categorised xenophobic attacks into two types: those occurring in rural and urban districts, and those happening in different provinces. According to our chosen year, four provinces reported xenophobic attacks. The results of this analysis are presented in Table 3.

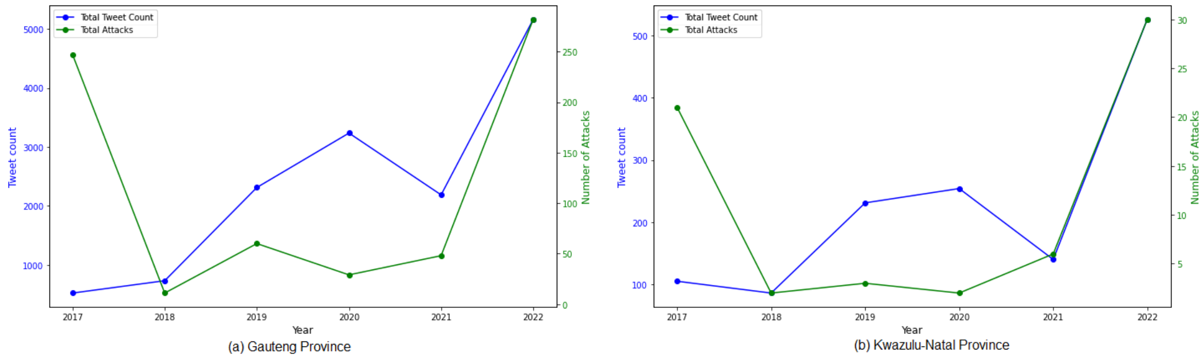


Figure 2: Comparison between the increase in tweets and number of attacks

Table 3: Number of attacks by province

SN	Province	District	Attacks
1	Gauteng	Rural	210
		Urban	467
2	KwaZulu-Natal	Rural	18
		Urban	48
3	Limpopo	Rural	61
		Urban	3
4	Mpumalanga	Rural	3
		Urban	0

4.2 Xenophobia tweet sentiment classification outputs

Here, we fitted the human-annotated data into machine learning classification models and a deep learning model. The purpose was to develop a machine learning model capable of classifying xenophobic tweets into three sentiment classes. The outcomes of the training are summarised in Table 4. From Table 4 it is observed that SVM and RF models outperformed other models. More details of the dominant word according to sentiment and other model performance matrices can be found in Appendices A and B.

The classification of data into positive, negative, and neutral by the models is summarised in Appendix C. A tweet is classified as positive if it is against xenophobia attacks, negative if the tweet is in favour of xenophobic attacks, and neutral if the tweet is neither in favour nor against xenophobic attacks. We summarise the classified tweets in districts (rural or urban) with respect to the year, as shown in Figure 3.

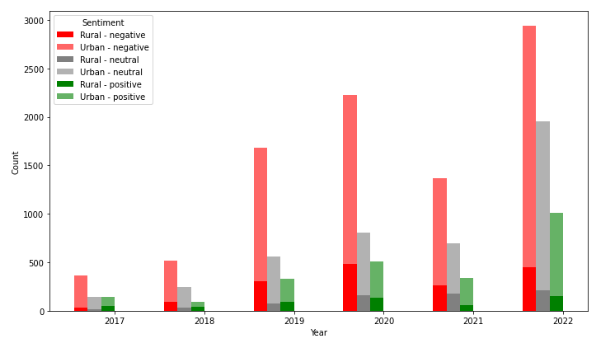


Figure 3: Sentiment of Rural and Urban District by Year.

To validate the performance of our models, we employed the visualisation of the ROC metric to assess the quality of the multi-classification output, in conjunction with the Area Under the Curve (AUC) measurement, as illustrated in Appendix D. The

Table 4: Model performance

SN	Model	Accuracy	Average AUC
1	NB	0.53	0.71
2	SVM	0.57	0.77
3	LR	0.51	0.69
4	RF	0.57	0.77
5	XGBoost	0.51	0.70
6	LSTM	0.50	0.50

ROC curve showcases the relationship between the true positive rate and the false positive rate. Consequently, a curve that aligns towards the upper left corner of the plot signifies the classifier's superior ability to accurately categorise tweets into different sentiment classes. Meanwhile, the AUC serves as an indicator of the extent to which the plot is situated under the curve, reflecting the overall performance of the classifier.

We conducted a comparison of the negative, neutral, and positive sentiment classes to gauge the effectiveness of our model's classification. In this context, the numerical values 0, 1, and 2 in some of the plots represent the negative, neutral, and positive sentiment classes, respectively. Similarly, the percentage demographics of the sentiments according to the province with respect to the year are displayed in Appendix D.

4.3 Topic generation and analysis of Xenophobia tweets

We utilised LDA to extract 10 distinct topics from the xenophobia dataset. These topics encompass a range of discussions: Topic 0 explores foreigner perception of Xenophobia, Topic 1 delves into the discourse surrounding South African identity, Topic 2 centers on conversations about urban development, and Topic 3 explores the societal Impact of Xenophobia. Additionally, Topic 4 investigates the legal framework, Topic 5 is centered around discussions about Operation Dudula, Topic 6 captures public opinion about Xenophobia, Topic 7 examines Racial Dynamics, Topic 8 focuses on immigration policy, and Topic 9 addresses migration issues. The primary aim of this analysis is to grasp user sentiments during xenophobia events and to comprehend the role of social media in the context of South African xenophobia unrest. The summarised Table in Appendix E presents the key topics, their associated keywords, and the respective comment counts for each topic. Similarly, the summary of the keywords and their corresponding probabilities for all the topics is shown in Appendix F.

In Figure 4, we identified the specific comments that contribute to each topic, along with their corresponding sentiments. These topics were subsequently categorised according to their prevailing sentiments. Interestingly, our observation revealed that all of the topics predominantly contain negative comments.

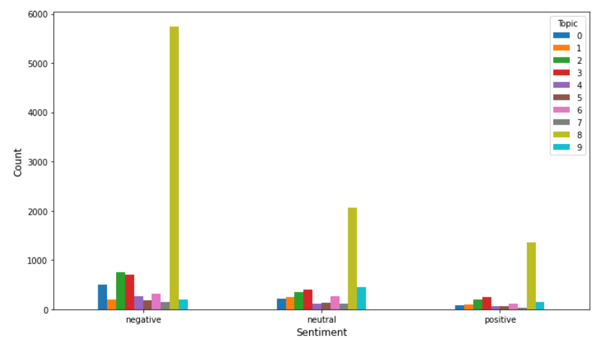


Figure 4: Grouped topic counts by sentiment.

5 Discussion

In this section, we have provided a comprehensive presentation and thorough discussion of the results obtained from our analysis. To begin with, we delve into the core findings of our study.

5.1 Findings

Our findings reveal a notable trend: the year 2022 witnessed a higher volume of tweets dedicated to discussing Xenophobia compared to other years. Additionally, there was a marked increase in the number of likes and retweets for Xenophobia-related content in 2022 compared to previous years. Concurrently, the year 2017 exhibited the highest frequency of attacks, followed closely by 2022, surpassing other years. This pattern indicates a noteworthy correlation: As the years progress, there is a visible rise in the number of tweets, likes, and retweets centered around Xenophobia, see Table 2. This suggests that users are progressively becoming more attuned to this issue and are utilising social media as a platform to express their thoughts on Xenophobia, comfortably from their own homes.

Our analysis revealed a robust correlation between the escalation in tweets and incidents of attacks. Specifically, the province of Gauteng exhibited a particularly pronounced and statistically significant association (Corr=0.73, P=0.03) between the increase in tweets and the occurrences of attacks, with a noticeable emphasis from 2021 to 2022. Similarly, Kwazulu-Natal province exhibited a compelling and statistically significant association (Corr=0.81, p=0.02) between the increase in tweets and the instances of attacks, see Figure 2. This implies that the narratives and discussions propagated on social media platforms could potentially exert an influence on real-life events, particularly within these provinces.

Furthermore, among the nine provinces in South Africa, only four registered instances of attacks based on our data. These provinces are Gauteng, KwaZulu-Natal, Limpopo, and Mpumalanga. Additionally, the majority of these attacks occurred in urban areas within Gauteng and KwaZulu-Natal. This is evident in the extent of unrest and damage that transpired in Johannesburg and Durban during the Xenophobia outbreaks. Conversely, the Limpopo and Mpumalanga provinces recorded a higher frequency of attacks in rural areas. Furthermore, our analysis brought to light a notable trend in the sentiment of Xenophobia-related tweets originating from rural areas over time. Notably, during the year 2022, a predominance of all sentiment categories shifted towards urban regions, see Table 3. Additionally, our findings unveiled an increased prevalence of negative sentiments within discussions surrounding Xenophobia across all provinces. In particular, the provinces of Gauteng, Limpopo, KwaZulu-Natal, and Mpumalanga exhibited a markedly higher occurrence of negative sentiment compared to other provinces, as depicted in the visual representation in the accompanying Appendix C.

We successfully generated a set of 10 distinct topics using LDA model. Among these, Topics 0, 1, 6, and 9 garnered a higher volume of comments dedicated to discussing Xenophobia in comparison to the other topics, as highlighted in Appendix E. Intriguingly, across all topics, there was a prevalent prevalence of negative sentiments compared to other sentiment categories, see Figure 4. Notably, provinces such as Gauteng and KwaZulu-Natal exhibited an increased engagement in urban areas discussing Topics 0, 3, 5, 6, 8, and 9. On the other hand, Topics 5, 6, and 9 saw a surge in comments addressing Xenophobia within provinces like Limpopo, Mpumalanga, and Eastern Cape, particularly within the context of rural areas. Our findings showed when there is an increase in negative sentiment of a topic, there is a Xenophobic attack. This goes to show the power of social media. It is very much possible that narrations or agendas being pushed on social media could influence or contribute to heightened unrest in real life.

Our findings vividly demonstrate a significant correlation: an upsurge in negative sentiment within a particular topic corresponds to the occurrence of a Xenophobic attack. This unequivocally underscores the potent impact of social media. It becomes evident that narratives or agendas propagated through these platforms hold the potential to exert influence and even contribute to escalating unrest in the real world.

5.2 Policy Implication

By analysing social media data, authorities can gain timely insights into public sentiment and discussions related to Xenophobia. This information can be used to develop early warning systems that detect emerging tensions and potential outbreaks of violence. Governments and law enforcement agencies can proactively prepare and allocate resources to prevent or mitigate Xenophobic attacks. The analysis of sentiment trends and topics can assist policymakers in identifying regions and communities with heightened levels of negative sentiment. Targeted interventions, awareness campaigns, and community engagement programs can be tailored to these areas to address misconceptions, promote understanding, and reduce the likelihood of Xenophobic incidents.

The findings derived from social media analysis can inform the formulation and implementation of policies aimed at addressing Xenophobia. Governments can design policies that counteract the negative narratives propagated online, promote inclusivity, and foster social cohesion. These policies can also be tailored based on geographical variations in sentiment and the nature of discussions. The insights from social media data can be used to develop educational initiatives that raise awareness about the harmful impacts of Xenophobia and debunk myths that contribute to its perpetuation. Online campaigns, workshops, and educational resources can play a pivotal role in changing public perceptions and attitudes. In the event of a Xenophobic attack, real-time sentiment analysis can provide rapid insights into public reactions, facilitating swift and appropriate crisis management responses. Authorities can use these insights to communicate effectively, dispel rumours, and prevent the spread of misinformation.

Additionally, social media analysis can reveal key influencers and opinion leaders who shape online discussions. Engaging with these individuals can help promote positive narratives, encourage dialogue, and counteract divisive messages. Collaborative efforts between government, civil society, and social media platforms can foster environments that prioritise inclusivity and understanding. Continuously monitoring social media sentiment and the correlation between online discussions and Xenophobic incidents allows policymakers to assess the effectiveness of their interventions. Data-driven insights can guide the refinement of strategies and policies over time.

Finally, social media data analysis can provide insights into cross-border discussions about Xenophobia. Collaboration with neighbouring countries and international organizations can lead to coordinated efforts to address the issue holistically and promote regional stability. Therefore, leveraging social media data, machine learning, topic modelling, and correlation analysis offers policymakers a valuable toolset to tackle Xenophobic attacks in South Africa. These approaches empower governments to proactively address negative sentiments, foster understanding, and implement targeted measures that contribute to a more inclusive and harmonious society.

5.3 Limitations

The Twitter data used for this research only reflects the opinions of Twitter users located in South Africa from January 2017 to December 2022. It is important to acknowledge that South Africa has a population of about 60 million people, and Twitter users may not represent the broader population accurately. With only an estimated 15% of online adults in the age range of 18–34 years using Twitter (Statista 2022), this research may not fully capture the opinions of all South Africans regarding xenophobia-related discussions. Nonetheless, this research provides valuable insights and analysis from the Twitter data for the selected age group to support policymaking.

5.4 Ethical Considerations

The study followed ethical guidelines, obtaining approval from Twitter to access tweets through their academic API. Only publicly available tweets were used. Personal information was carefully removed to ensure user confidentiality and anonymity, in line with Twitter's academic research terms and conditions.

6 Conclusion

We analyzed over 15,000 tweets from 2017 to 2022 related to xenophobia in South Africa, categorising them by sentiment. Our findings reveal a link between online sentiment, discussions, and actual xenophobic events. The research emphasizes the power of social media in influencing perceptions and the need for proactive engagement to prevent potential unrest. The insights offer valuable guidance for policymakers to address xenophobia effectively in South Africa.

References

Alsharhan, A. M., Almansoori, H. R., Salloum, S. & Shaalan, K. (2022), Three mars missions from three countries: Multilingual sentiment analysis using vader, *in* A. E. Hassanien, R. Y. Rizk, V. Snášel & R. F. Abdel-Kader, eds, 'The 8th International Conference on Advanced Machine Learning and Technologies and Applications (AMLTA2022)', Springer International Publishing, Cham, pp. 371–387.

Aruleba, K. & Jere, N. (2022), 'Exploring digital transforming challenges in rural areas of south africa through a systematic review of empirical studies', *Scientific African* **16**, e01190.

Bachchu, P., Anchita, G., Tanushree, D., Debashri, D. A. & Somnath, B. (2022), 'A comparative study on sentiment analysis influencing word embedding using svm and knn', *Cyber Intelligence and Information Retrieval* **291**, 199–211. Lecture Notes in Networks and Systems book series (LNNS,volume 291).

Bai, X. (2018), Text classification based on lstm and attention, *in* '2018 Thirteenth International Conference on Digital Information Management (ICDIM)', pp. 29–32.

Chenzi, V. (2021), 'Fake news, social media and xenophobia in south africa', *African Identities* **19**(4), 502–521.

Chingwete, A., Dryding, D. & Dessein, V. (2018), 'Afrobarometer round 7 survey in south africa', `https://www.afrobarometer.org/wp-content/uploads/migrated/files/publications/Summary%20of%20results/saf_r7_sor_13112018.pdf`.

Danso, R. & McDonald, D. A. (2001), 'Writing xenophobia: Immigration and the print media in post-apartheid south africa', *Africa today* pp. 115–137.

Dube, G. (2018), 'Afrophobia in mzansi? evidence from the 2013 south african social attitudes survey', *Journal of Southern African Studies* **44**.

Fokou, G., Yamo, A., Kone, S., Koffi, A. J. d. & Davids, Y. D. (2022), 'Xenophobic violence in south africa, online disinformation and offline consequences', *African Identities* pp. 1–20.

Gulati, K., Saravana Kumar, S., Sarath Kumar Boddu, R., Sarvakar, K., Kumar Sharma, D. & Nomani, M. (2022), 'Comparative analysis of machine learning-based classification models using sentiment classification of tweets related to covid-19 pandemic', *Materials Today: Proceedings* **51**, 38–41. CMAE'21.

Gumede, W. (2015), 'South africa must confront the roots of its xenophobic violence', `https://tinyurl.com/2p9sj5t2`.

Harris, B. (2001), A foreign experience: Violence, crime and xenophobia during south africa's transition.

Hlatshwayo, M. (2023), 'South africa enters the second wave of xenophobic violence: the rise of anti-immigrant organisations in south africa', *Politikon* **50**(1), 1–17.

Honnibal, M. (2017), 'spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing.', *Sentometrics Research* **1**(1), 2586–2593.
URL: *https://sentometrics-research.com/publication/72/. [Accessed 2022-09-21]*

Jaya Hidayat, T. H., Ruldeviyani, Y., Aditama, A. R., Madya, G. R., Nugraha, A. W. & Adisaputra, M. W. (2022), 'Sentiment analysis of twitter data related to rinca island development using doc2vec and svm and logistic regression as classifier', *Procedia Computer Science* **197**, 660–667. Sixth Information Systems International Conference (ISICO 2021).

Kewsuwun, N. & Kajornkasirat, S. (2022), 'A sentiment analysis model of agritech startup on facebook comments using naive bayes classifier', *International Journal of Electrical & Computer Engineering* **12**(3), 2829–2838. Accessed 2022-06-17.

Laksono, R. A., Sungkono, K. R., Sarno, R. & Wahyuni, C. S. (2019), Sentiment analysis of restaurant customer reviews on tripadvisor using naïve bayes, *in* '2019 12th International Conference on Information Communication Technology and System (ICTS)', pp. 49–54.

Li, F., Van den Bossche, J., Zeitlin, M., Meeseeks, M., PD, T., Hawkins, S. & et al (n.d.), 'What's new in pandas 1.2.4', Available on `https://pandas.pydata.org/pandas-docs/stable/whatsnew/v1.2.4.html`. [Accessed: 2022-06-01].

Luan, Y. & Lin, S. (2019), Research on text classification based on cnn and lstm, *in* '2019 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)', pp. 352–355.

Mathew, B., Dutt, R., Goyal, P. & Mukherjee, A. (2019), Spread of hate speech in online social media, *in* 'Proceedings of the 10th ACM conference on web science', pp. 173–182.

Mgboh, U., Ogbuokiri, B., Obaido, G. & Aruleba, K. (2021), Visual data mining: A comparative analysis of selected datasets, *in* A. Abraham, V. Piuri, N. Gandhi, P. Siarry, A. Kaklauskas & A. Madureira, eds, 'Intelligent Systems Design and Applications', Springer International Publishing, Cham, pp. 377–391.

Mgogo, Q. & Osunkunle, O. (2021), 'Xenophobia in south africa: An insight into the media representation and textual analysis', *Global Media Journal* **19**(38), 1–8.

Mogekwu, M. (2005), 'African union: Xenophobia as poor intercultural information', *Ecquid Novi* **26**, 5–20.

Newman, N., Dutton, W. & Blank, G. (2013), 'Social media in the changing ecology of news: The fourth and fifth estates in britain', *International journal of internet science* **7**(1).

Obaido, G., Ogbuokiri, B., Mienye, I. D. & Kasongo, S. M. (2023), A voting classifier for mortality prediction post-thoracic surgery, *in* A. Abraham, S. Pllana, G. Casalino, K. Ma & A. Bajaj, eds, 'Intelligent Systems Design and Applications', Springer Nature Switzerland, Cham, pp. 263–272.

Obaido, G., Ogbuokiri, B., Swart, T. G., Ayawei, N., Kasongo, S. M., Aruleba, K., Mienye, I. D., Aruleba, I., Chukwu, W., Osaye, F., Egbelowo, O. F., Simphiwe, S. & Esenogho, E. (2022), 'An interpretable machine learning approach for hepatitis b diagnosis', *Applied Sciences* **12**(21), 11127.

Ogbuokiri, B., Ahmadi, A., Bragazzi, N. L., Movahedi Nia, Z., Mellado, B., Wu, J., Orbinski, J., Asgary, A. & Kong, J. (2022), 'Public sentiments toward covid-19 vaccines in south african cities: An analysis of twitter posts', *Frontiers in public health* **10**, 987376.

Ogbuokiri, B., Ahmadi, A., Mellado, B., Wu, J., Orbinski, J., Asgary, A. & Kong, J. (2023), Can post-vaccination sentiment affect the acceptance of booster jab?, *in* A. Abraham, S. Pllana, G. Casalino, K. Ma & A. Bajaj, eds, 'Intelligent Systems Design and Applications', Springer Nature Switzerland, Cham, pp. 200–211.

Ogbuokiri, B., Ahmadi, A., Nia, Z. M., Mellado, B., Wu, J., Orbinski, J., Asgary, A. & Kong, J. (2023), 'Vaccine hesitancy hotspots in africa: An insight from geotagged twitter posts', *IEEE Transactions on Computational Social Systems* pp. 1–14.

Ogbuokiri, B., Ahmadi, A., Tripathi, N., Movahedi, Z., Mellado, B., Wu, J., Orbinski, J., Asgary, A. & Kong, J. (2022), 'The impact of omicron variant in vaccine uptake in south africa', *Research Square* . Preprint.

Okolo, C. T., Aruleba, K. & Obaido, G. (2023), 'Responsible ai in africa—challenges and opportunities', *Responsible AI in Africa: Challenges and Opportunities* pp. 35–64.

Pontiki, M., Gavriilidou, M., Gkoumas, D. & Piperidis, S. (2020), Verbal aggression as an indicator of xenophobic attitudes in greek twitter during and after the financial crisis, *in* 'Language

Resources and Evaluation Conference (LREC 2020)', pp. 19 – 26.

Praveen, G. J. & Prasanna, K. H. R. (2021), 'Sentiment analysis:textblob for decision making', *International Journal of Scientific Research Engineering Trends* **7**(2), 2395–566X.

Qi, Z. (2020), The text classification of theft crime based on tf-idf and xgboost model, *in* '2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)', pp. 1241–1246.

Ritanshi, J., Seema, B. & Seemu, S. (2022), 'Sentiment analysis of covid-19 tweets by machine learning and deep learning classifiers', *Advances in Data and Information Sciences* **318**, 329–339. Lecture Notes in Networks and Systems.

Statista (2022), 'Largest cities in south africa in 2021 by number of inhabitants', Available on `https://www.statista.com/statistics/1127496/largest-cities-in-south-africa/`. [Accessed: 2022-06-01].

statista (2023), 'Social media in africa - statistics facts'.
URL: *https://www.statista.com/topics/9922/social-media-in-africa/topicOverview*

Tarisayi, K. & Manik, S. (2020*a*), 'An unabating challenge: Media portrayal of xenophobia in south africa', *Cogent Arts Humanities* **7**, 1859074.

Tarisayi, K. S. & Manik, S. (2020*b*), 'An unabating challenge: Media portrayal of xenophobia in south africa', *Cogent Arts & Humanities* **7**(1), 1859074.

Udanor, C., Ossai, N., Nweke, E., Ogbuokiri, B., Eneh, A., Ugwuishiwu, C., Aneke, S., Ezuwgu, A., Ugwoke, P. & Christiana, A. (2022), 'An internet of things labelled dataset for aquaponics fish pond water quality monitoring system', *Data in Brief* **43**, 108400.
URL: *https://tinyurl.com/ypb96d7p*

van der Vyver, A. (2019), An analyisis of twitter discourse on xenophobia in south africa, *in* '23rd PATTAYA International Conference on "Advances in Engineering and Technology"'.

Vasista, R. (2018), 'Sentiment analysis using svm'. [Accessed on June 11, 2022].
URL: *https://medium.com/@vasista/sentiment-analysis-using-svm-338d418e3ff1*

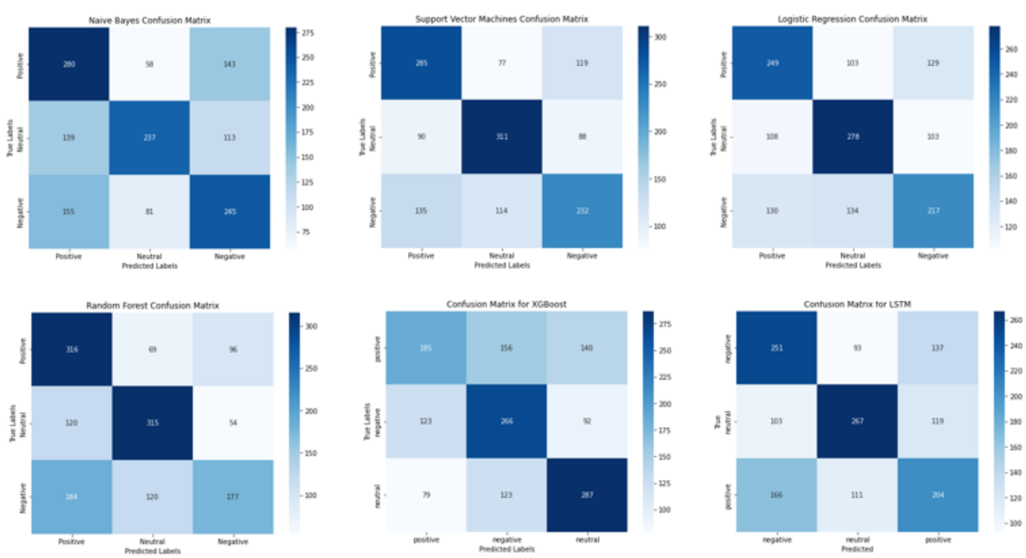
Wardle, C. & Derakhshan, H. (2017), *Information disorder: Toward an interdisciplinary framework for research and policymaking*, Vol. 27, Council of Europe Strasbourg.

Williams, M. L., Burnap, P., Javed, A., Liu, H. & Ozalp, S. (2020), 'Hate in the machine: Anti-black and anti-muslim social media posts as predictors of offline racially and religiously aggravated crime', *The British Journal of Criminology* **60**(1), 93–117.

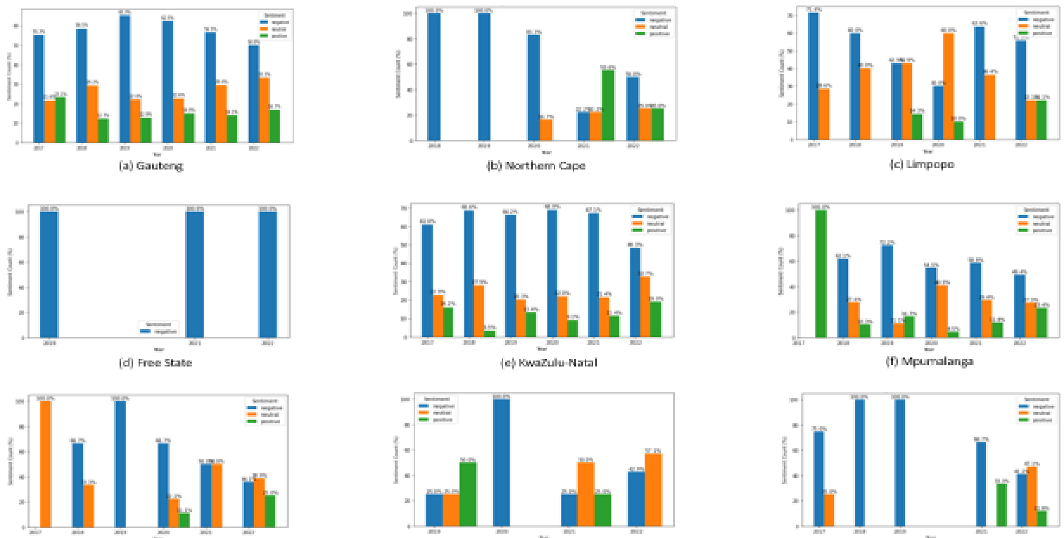
Yu, D. & Xiang, B. (2023), 'Discovering topics and trends in the field of artificial intelligence: Using lda topic modeling', *Expert Systems with Applications* **225**, 120114.
URL: *https://tinyurl.com/25prsh78*

A Wordcloud of sentiments

B Confusion matrix for different models.

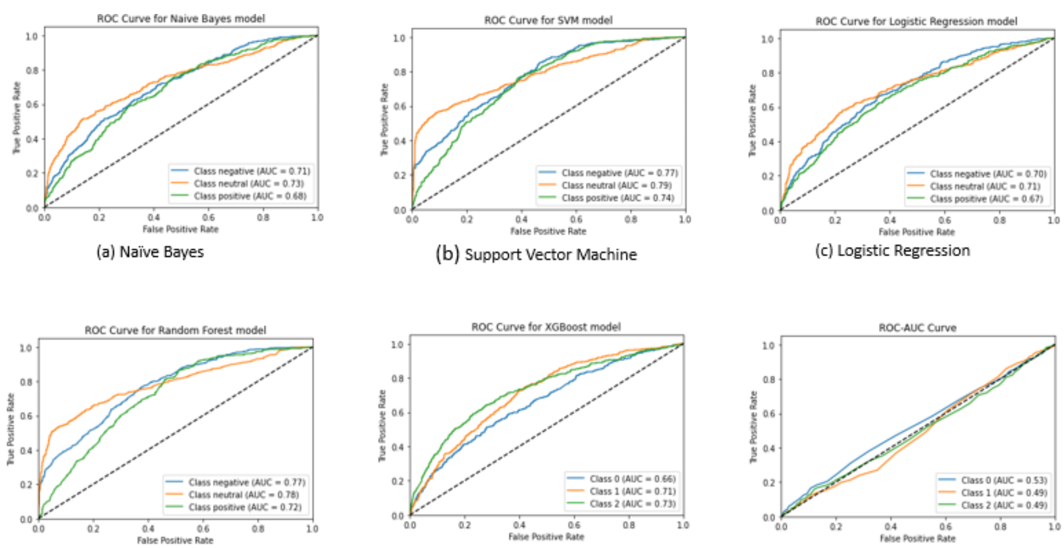


C South Africa provincial xenophobia sentiment count over time



(a) Gauteng (b) Northern Cape (c) Limpopo
(d) Free State (e) KwaZulu-Natal (f) Mpumalanga
(g) North West (h) Western Cape (i) Eastern Cape

D ROC-AUC Curve for different sentiment classifiers.



(a) Naïve Bayes (b) Support Vector Machine (c) Logistic Regression
(d) Random Forest (e) XGBoost (f) LSTM

E Top ten keywords of the 10 generated topic from xenophobia dataset.

Topic	Predicted Title	Representative Keyword	Comments
0	Foreigner Perception on Xenophobia	foreigners, sa, illegal, jobs, pay, n, immigrants, economy, drugs, country	3166
1	South African Identity	south, foreigners, africans, africa, african, operationdudula, people, country, must, never	1795
2	Urban Development	mashaba, city, spaza, create, european, abt, herman, press, ok, level	156
3	Societal Impact of Xenophobia	name, employers, lose, meant, rdp, playing, operationdudula, important, commsafety, houses	209
4	Legal Framework	foreigners, foreign, nationals, dudula, law, operation, amp, huge, drivers, drug	604
5	Operation Dudula	lux, nhlanhla, soweto, operationdudula, johannesburg, though, demand, amp, arrested, joburgcbd	861
6	Public Opinion about Xenophobia	foreigners, people, us, country, dont, like, u, amp, one, think	5404
7	Racial Dynamics	foreigners, whites, black, crime, blacks, white, joburg, like, commit, criminals	400
8	Immigration Policy	n, foreigners, safricans, sa, police, influx, would, ke, home, land	495
9	Migration Issues	immigrants, illegal, foreigners, country, must, anc, sa, go, borders, government	2698

F Topic Representative Keywords and their probabilities