Shedding Light on Loadshedding with Natural Language Processing: A social media case study on public perspectives of the South African electricity crisis in 2022

Avashlin Moodley[1], Privolin Naidoo[1],

[1]Natural Language Processing Research Group, Council for Scientific and Industrial Research (CSIR)

Email: amoodley1@csir.co.za pnaidooo4@csir.co.za

Abstract

In times of collective discomfort and dissatisfaction, people often find solace in shared adversity on social media platforms like X (formerly known as Twitter). These platforms offer a unique window into the public's emotions and viewpoints concerning common challenges. In 2022, South Africa experienced an electricity crisis, during which the country was subjected to rolling blackouts, commonly known as loadshedding, by Eskom, the country's primary electricity provider, to prevent a national electricity grid shutdown. This study conducted a data-driven exploration of the public discourse surrounding Eskom and loadshedding on X using natural language processing and data science techniques. The dataset utilised for this study comprised tweets containing keywords related to Eskom and loadshedding. The study delved into the topics of discussion by applying topic modelling techniques to uncover latent themes within the discourse. The topics were analysed through a multifaceted lens to unpack and highlight patterns within the sentiments, emotions and biases that underpin conversations related to loadshedding and Eskom. A notable inclusion in the analysis was the incorporation of sarcasm classifications, which enhanced the interpretation of the emotion and sentiment within the topics discussed. The findings uncovered from the analysis were contrasted with loadshedding-related events in 2022 to understand the public discourse as the electricity crisis escalated. The methodology of this study provides a framework for utilising natural language processing techniques to uncover and examine the perspectives of a collective within discourse related to events of shared interest.

Keywords: social media, text mining, natural language processing, topic modelling, loadshedding, Eskom

1 Introduction

In 2022, South Africa's energy landscape was under considerable strain, primarily attributed to challenges faced by its main electricity provider in coping with the electricity demand. As a result, the country was subjected to loadshedding, which is the controlled reduction of electricity supply to prevent a total grid collapse. In this time period, South Africa was subjected to recurring periods of stage 6 loadshedding (a 6000 megawatt reduction in electricity supply, each stage being equivalent to a 1000 megawatt reduction), which resulted in average citizens experiencing more than 25% of their day without electricity. This resulted in blackouts becoming a distressing norm, causing disruptions to daily life, economic activities, and national development.

Social media platforms like X (formerly known as Twitter) provided an outlet for citizens to voice their opinions and concerns about their personal experiences around the energy crisis and demand accountability from the state-owned enterprise. The populace experienced near-daily blackouts, with the severity becoming considerably worse as the year progressed. Discourse on X in 2022 serves as a microcosm of the national energy crisis experienced, offering a unique window to gauge public perception and opinion.

X discourse surrounding a specific topic, such as Eskom and Loadshedding in 2022, provides a rich source of text information. Navigating and comprehending large amounts of text-based tweets manually is a laborious and infeasible task. This is where Natural Language Processing (NLP) comes into play, leveraging NLP techniques such as topic modelling and sentiment analysis provides a mechanism for understanding high-level themes and the polarity of opinion in those themes. Topic models provide a categorisation that can be used for deeper analysis and as a feature in downstream classification tasks. Sentiment analysis provides a classification on a scale of positive, neutral and negative for a piece of text based on the language usage and context of the text. Enhancing sentiment with additional NLP classifications, such as sarcasm and emotion, allows for a deeper analysis of opinion.

In this study, we employed a combination of NLP techniques to dissect the discourse surrounding Eskom and loadshedding on Twitter. We classified sentiment, emotion and sarcasm within the tweets to understand the emotional undertones and biases of the discussions and identify prevalent topics to grasp the focal points of public concern. Through this multifaceted lens, we aimed to answer the following questions:What are the primary topics of conversation regarding Eskom and loadshedding? What observations can be made regarding the relationship between sarcasm and sentiment within key topics? Was the methodology taken a feasible framework to gauge emotion, opinion and bias on topics of public adversity?

2 Related Work

Sentiment analysis and topic modelling have proven to be useful tools for extracting valuable insights from social media data such as tweets.

Valence Aware Dictionary and sEntiment Reasoner (VADER) provides a robust rule-based model for sentiment analysis of social media text (Hutto and Gilbert, 2015). This model uses lexicons and grammar rules in order to capture complex sentiment expressions. The work by (Pak and Paroubek, 2010) underscores the complexities of sentiment classification in informal language, as seen in Twitter data.

(Albu and Spînu, 2022) introduced a novel approach to emotion detection by extending the WASSA dataset (Mohammad and Bravo-Marquez, 2017) to include emotions such as fear, sadness, joy, anger, and a neutral emotion. The crux of their approach lies in combining the BERTweet model with the SVM model, resulting in an ensemble model that outperformed other variants.

(Habib and Nithyanand, 2023) delved into narratives shared by COVID-19 victims on Twitter. Employing methodologies encompassing topic modelling and word embedding, the research aimed to unveil underlying themes and trends within these narratives.

(Tinsman et al., 2023) examined the potential link between social media usage during the COVID-19 pandemic and the emergence of trauma symptoms. Utilizing random forest regression and classification techniques on survey data, the study uncovers latent connections within the collected information.

(Aldous et al., 2019) scrutinised user engagement with news-related content across various social media platforms. Employing tools such as Twitter-LDA for topic modelling and VADER for sentiment analysis, the study offers insights into user interactions with diverse content.

(Pfeffer et al., 2023) presented a comprehensive analysis of a day's worth of Twitter data, shedding light on information dissemination, user behaviour, and linguistic patterns. While primarily focused on Twitter's societal impact, the study's methodologies encompass network analysis, sentiment analysis, and topic modelling.

(Moodley and Marivate, 2019) examined news articles from 2014 and 2019 using topic modelling techniques LDA and Non-negative Matrix Factorisation (NMF) to contrast the themes that were discussed relative to the South African general elections in both those years.

(Marivate et al., 2021) applied the semi-supervised Corex topic modelling technique from (Gallagher et al., 2017) to extract and categorise topics that gauge public response in relation to COVID-19 correspondence from official government communication.

(Butgereit, 2015) attempted to measure anger directed towards Eskom during loadshedding to gauge the extent of the public's emotional perception towards their electricity supply. (Butgereit, 2015) reports challenges with the data being used.

The studies discussed provide insights into methodologies, models and visualizations that inform the methodology employed in this study.

3 Methodology

This study aimed to gain an understanding of the latent themes within the Eskom Twitter corpus. Thereafter, analysing the themes through a multifaceted lens of different NLP classifications to draw high-level observations about the perceptions and opinions of the public concerning Eskom and loadshedding.

The methodology was broken down into 4 steps, which enabled the insights extracted in this study.

3.1 Data Collection & Preparation

The Eskom tweet corpus was collected using the Python *Snscrape* [1] library. The ingestion process consisted of collecting tweets that contained the keywords "Eskom" or "loadshedding". The collection process ran for 3 weeks and collected data that spanned 2008 to 2022.

The corpus was segmented to extract 2022 data for the context of this study. The data was subjected to a preprocessing pipeline to prepare the text for modelling and tagging at later stages. Website links, mentions, punctuation and emojis were removed from the data as part of this preprocessing step. The resulting text was lowercased and subjected to tokenisation and lemmatisation. This step reduces the vocabulary size of the text corpus by reducing tokens that are semantically the same but syntactically different. At this point, the data was in a suitable form to be tagged by the sentiment, emotion and sarcasm models.

For the topic modelling task, the data was enriched with the addition of bigrams and trigrams to provide tokens with richer context that aids in the creation of better-quality topics. The enriched data is then vectorised using a Term Frequency - Inverse Document Frequency (TF-IDF) vectoriser.

3.2 Sentiment & Emotion Tagging

Building classifiers that are capable of predicting sentiment and emotion that is localised to the language used by the South African Twitter population requires labelled data and numerous iterations of model training and evaluation. This was outside the scope of this study due to time and resource constraints; therefore, prebuilt models developed for sentiment and emotion predictions were used to tag the data with sentiment and emotion metadata to enrich the analysis.

The Valence Aware Dictionary and sEntiment Reasoner (VADER) has a nuanced understanding of emotions, making it well-suited for analysing emotions in short text, which is primarily found within Twitter data (Hutto and Gilbert, 2015). The VADER sentiment model was robust, easy to implement and able to achieve comparable results with recent transformer-based methods as showcased in the study (Saha et al., 2023). The VADER sentiment model was applied to the prepared data to assign sentiment tags that cater for 3 classes: positive, neutral and negative.

[1]https://github.com/JustAnotherArchivist/snscrape

For emotion recognition, the process included model inference on three preexisting models from the HuggingFace transformer repository trained for emotion recognition. The initial model was fine-tuned using Google's T5 (Raffel et al., 2019) on the emotion recognition dataset (Saravia et al., 2018). The model covers five emotions: joy, fear, sadness, anger, and surprise. The model was susceptible to misclassification errors in situations where sentences contained negating words. The second model employed a Roberta-based architecture on the Go Emotions dataset (Demszky et al., 2020), it handles multi-label classification with 28 Reddit-sourced labels. However, the model classified the majority of the tweets as neutral, which was undesirable. Utilising a pre-trained Roberta-based emotions model, the third model covers an extended range of emotions: anger, disgust, fear, joy, neutral, sadness, and surprise. The third model was selected for the recognition of emotions in tweets as it performed better on the shortcomings of its predecessors.

3.3 Sarcasm Tagging

Sarcasm is often defined as conveying the opposite of one's true intent through words, typically to mock, show irritation, or for humour (Merriam-Webster, 2023). The focus of this study is not to define and create a specialised sarcasm tagging model but to ascertain the impact understanding sarcastic tones within the corpus would have on sentiment and emotion analysis. This led us to utilise a fairly simple model. We employed a fine-tuned BERT model, specifically a 'bert-base-uncased' base that had been fine-tuned using an open-source Kaggle dataset using News Headlines for Sarcasm Detection. The rationale behind this selection was the parallels between the concise nature of news headlines and tweets. The data was cleaned as done previously for sentiment and emotion analysis before labelling the data.

3.4 Topic Modelling

Topic models provide a means to gain a high-level understanding of themes that exist within a text corpus. Twitter discourse contains short pieces of text with limited context, which can make it difficult for coherent topic models to be built using unsupervised techniques such as LDA and NMF. Semi-supervised topic models allow one to supply seed words that can guide the training of topic models. These models infer topics aligned to the keywords specified (Gallagher et al., 2017). Transformer-based models dominate the NLP domain at present; a popular semantic topic modelling technique that allows for supervision of topics is Bertopic (Grootendorst, 2022).

In this study, the data was prepared according to the process described in Section 3.1. Firstly, unsupervised Corex and Bertopic topic models were built. The unsupervised models produced topics that contained keywords related to highly used phrases in the discourse; however, they were not the most coherent to interpret. Keywords from the topics were extracted to inform the compilation of the seed words listed in Table 1.

Seed Words		
loadshedding,load,shed,shedding	solar	crisis, disaster
buy eskom	coal	useless, bad
andre, ruyter, ceo	stage	debt, owe eskom
corrupt, corruption	backup	privatize, privatization

Table 1: Seed tokens used to supervise topics

The seed words were used to build semi-supervised topic models using both techniques of interest. The Bertopic model took significantly longer than the Corex topic model to train and produced topics that were comparable to the Corex topics in both syntax and distributions of the topic clusters. This comparison led to the selection of the Corex topic modelling technique to enable simplicity in the framework being suggested.

4 Data Overview

The Twitter data collected on Eskom was segmented and cleaned to include only 2022 data. Table 2 provides a view of the characteristics of the tweets.

Statistic	
Number of tweets	136 597
Tweets mentioning Eskom	70 926
Number of token	2 709 777
Number of unique tokens	64 070
Average tokens per tweet	20
Average characters per tweet	111
Average word length	6

Table 2: Statistics describing the data used in this study

The tweets in the dataset average 20 tokens and contain slightly more than 100 characters on average. A peculiar insight from this table is the relationship between tokens and unique tokens. There are 2.7 million tokens, but only 64 thousand unique tokens. This indicates that much of the discourse contains common vocabulary, but is rich enough to contain diverse topics.

5 Exploring highlight themes for public opinion

5.1 Highlight Topics

Corex topic models were built using the seed words in Table 1. Two topic models were selected for further inspection. One was built on the entire corpus, the other was built on tweets labelled as sarcastic. Both models should exhibit topics aligned to the seed words, thus allowing for comparisons to be drawn. The sarcastic model is discussed later in Section 6.1; this section will focus on the full corpus.

Table 3 provides a view of the keywords and human-annotated labels for some of the highlight topics that represent public concerns. Expectedly, loadshedding-specific topics (1 & 3) surfaced quite prominently where Twitter users engaged in discourse around the outages and loadshedding stages, with a large topic containing tweets related to loadshedding in general. A topic arose with keywords indicative of the public urging for a state of disaster to be declared regarding the energy crisis. There was another topic that looked to provide reasons for the challenges, with keywords citing illegal electricity connections, looting, cable theft and informal settlements as possible causes of the energy crisis.

A topic arose regarding the ex-CEO of Eskom, which likely details public opinion related to the job he was doing as well as matters related to him that were newsworthy at the time. There was also a topic that included mentions of state-owned enterprises, likely to express dissatisfaction towards these entities.

Topic	Keywords	Topic Label
1	load,shed,load shed,shedding,load shedding, eskom load,stage load,load reduction,eskom load shed,reduction	Loadshedding
3	stage,hour,schedule,restore,update,power, outage,zone,morning,dobsonville	Stages and Outages
5	disaster,state disaster,crisis,declare,state,national state, national state disaster,utility eskom,declare state,sabcnews	State of Disaster
8	illegal,illegal connection,connection,cable, anc looting eskom,thank anc looting,anc looting, cable theft,illegal electricity,informal settlement	Outage Reasons/Complaints
9	ruyter,ceo,andre,andre ruyter,eskom ceo,pravin, andré ruyter,andré,ceo eskom,molefe	Andre de Ruyter (CEO)
10	saa,denel,transnet,prasa,post office,sabc, eskom saa,eskom transnet,eskom denel,saa denel	State Owned Enterprises
20	human right,basic human right,basic human,failed utility, pay failed utility,pay failed,electricity basic human,violation, right violation,human right violation	Basic Human Rights

Table 3: Highlight topics from 2022

Whilst inspection of the topic keywords provided a view into the discourse, topic timelines provided a window into when these topics were pressing issues. Figure 1 illustrates when spikes in volume occur in topics, which can be an indicator of an event occurring.

The two large spikes seen in topic 3 occur in February and September. These are notable because February 2022 was the first time in 2 months that the country experienced loadshedding, which likely resulted in uproar online due to the inconveniences that many thought were over, due to the 2-month hiatus. September, in contrast, was the start of loadshedding becoming a mainstay for South Africans; notable discussion points in this time were related to stage 4-6 loadshedding becoming a daily occurrence. The loadshedding topic received consistent discourse throughout the year, with the topic related to stages spiking with changes in the loadshedding schedule. The state of disaster topic also showed spikes in volume that correlated to the topic related to loadshedding stages increasing.

5.2 Sentiment & Emotion breakdowns

Analysis of the topics and timelines provides a view of the content that existed in the discourse and the temporal factors associated with it. The sentiment and emotion distributions in Figure 2 and Figure 3 indicate the feelings of the public in relation to Eskom and loadshedding.

Inspecting Figure 2 showcases the dominant negativity associated with each of the highlight topics. Topic 1 & 3 display some content which has been tagged positive; this is likely in times when the loadshedding stages are reduced. The large neutral constituent in these topics were related to informative posts that aimed to provide information and updates about loadshedding.

The distribution of emotions within the topics present an interesting observation. Figure 3 illustrates the emotions that are associated with each topic. The trend of negativity outweighing positivity is consistently seen in the emotion tags which indicates that the emotion modelling is at least

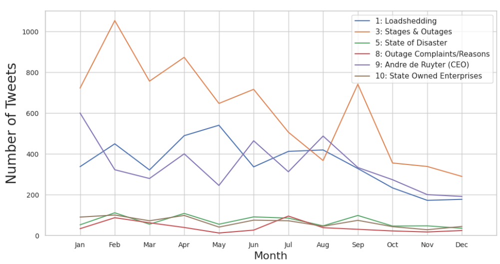

Figure 1: Topic timelines for highlight topics

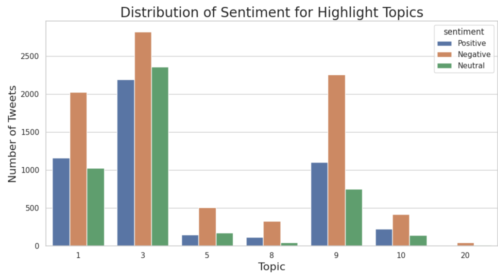

Figure 2: Sentiment distribution over highlight topics

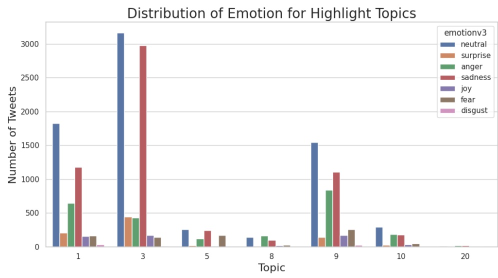

Figure 3: Emotion Distribution over highlight topics

working well enough to separate positive and negative related content. The large volumes of neutral content in topic 1 & 3 are likely indicative of informative messaging related to loadshedding. The similar volumes of anger and surprise on topic 3 correlated to announcements of higher stages of loadshedding. Sadness is a prominent emotion in topic 1,3 & 9. Topic 9 related to the ex-CEO of Eskom, this topic showcases the highest ratio of anger to sadness, which indicates that public discourse regarding the ex-CEO was highly negative, which aligns with the sentiment indicated in Figure 2.

6 Sarcasm Analysis

Due to the nature of tweets in combination with their replies, the pattern of rhetorical questions and sarcastic comments is prevalent. In our corpus, a notable 34% of the tweets were tagged as sarcastic, which is a significant subset of the data.

Recognising that our labelling model was not tailored specifically for Twitter data, we undertook a manual review of the categorised sarcastic tweets. By randomly selecting samples from the corpus and assessing the tweet content manually, we gained clarity on the model's precision and got a deeper understanding of the linguistic nuances involved. Among the sarcastic tweets that underwent manual analysis, a recurring theme was their dominant tone of complaint and distress conveyed through humour. These grievances were often articulated as questions or witty remarks, predominantly referencing problems related to loadshedding. This examination was not only enlightening but also instrumental in evaluating the efficacy of the sarcasm labelling approach.

Some tweets were fairly explicit in their sarcastic tone and did not need extra context to understand, for example, *"Great job again Eskom we really appreciate all your non service."*. However, this is not always the case, more often than not, more context is required to truly infer if a statement was sarcastic. *"Awesome I'll be sure to send any bills and costs for damages their way"* isn't overtly sarcastic until it is placed into context, being that the tweet is a reply to a post about damage caused by loadshedding. Additionally, many of the tweets were simply jokes and humour at the expense of Eskom, which often included popular culture tropes and memes *"Eskom can't turn off Chuck Norris's electricity instead Chuck Norris loadsheds Eskom"*.

The manual analysis revealed that most of the tweets were appropriately categorised as sarcastic. The remainder, although not sarcastic, displayed conflicting sentiments or exaggerated statements and questions, with a subset being generic spam tweets and updates regarding loadshedding. These tweets, though not fitting the sarcasm mould, were rich in emotional depth and expressive thought.

6.1 Topic Modelling

The process described in section 3.4 was used to extract topics from the subset of the corpus labelled as sarcastic. Both an unsupervised and supervised approach were taken, with the final topics that were settled on being the supervised topics. This was due to the supervised model producing more coherent topics. To keep the analysis consistent across the full corpus and this subset, we utilised the exact same vocabulary and vectoriser model. This ensured that the semantic information extracted from the full corpus stayed consistent across our evaluations.

Interestingly, the topics in 4 did not stray too far from the full corpus, indicating the topics discussed in a sarcastic tone do not deviate far from the full corpus. The common topics of Eskom, loadshedding, corruption and Eskom's CEO, Andre de Ruyter were expected as the corpus primarily revolved around these topics, as well as being included within the seed words. A unique view, however, was that Eskom was strong over multiple topics related to corruption within Eskom, the state of disaster, revamping the utility and lastly the uselessness of the utility.

Topic	**Keywords**	**Topic Label**
1	load, shed,shedding, load shed, load shedding, status, eskom status available, status available, eskom status, stage load	Loadshedding
4	coal, eskom coal, coal eskom, coal power, wet,export, supply coal, quality coal, wet coal, quality	Eskom Coal Supply
5	crisis, disaster,state disaster, eskom crisis, energy crisis,electricity crisis, disaster eskom, declare,crisis eskom, eskom disaster	State of Disaster
6	useless,bad, useless eskom, eskom useless, useless people, useless company, useless bunch, just useless,eskom bad, useless useless	Uselessness of Eskom
7	buy eskom, jobseekerssa, eskom dm, unit eskom dm,buy, dm,unit eskom, programme, revamp day, revamp day service	Revamp Eskom Service
9	ruyter, ceo,andre, eskom ceo, ceo eskom, andre ruyter, andre ruyter, war room,andre, ruyter eskom	Andre de Ruyter
10	corruption, corrupt, corruption eskom, eskom corruption, corrupt thug, denel,saa, voetsek corrupt, corrupt incompetent, corrupt thug useless	Corruption
12	privatize,privatization,privatize eskom,eskom privatize,ubva, privatization eskom,ndi,vha,mudagasi,uri	Privatization
19	make life difficult, life difficult, make life, shit hit fan, shit hit, hit fan, chuck norris, norris, patient eskom, past month	Life Difficulties

Table 4: Highlights of topics from tweets flagged as sarcastic

One of the topics that appeared had one major difference between the two sets of topics extracted. The sarcastic data extracted a topic related to the ANC which primarily focused on the term *"Voetsak ANC"* whereas the topic related to ANC across the full dataset mentioned *"Viva ANC"* primarily. Other notable differences lie within the distribution of keywords within similar topics, where the sarcastically labelled data showed less variation with regard to entities other than Eskom being mentioned.

6.2 Emotion and Sentiment Analysis

Both emotion and sentiment analysis were applied to the subset of data labelled as sarcastic. This helped understand if there was any significant deviations in the distribution caused by the sarcastic tweets. We see no major deviation in the distribution of the sentiment in Table 5 and emotion in Table 6. There was a slight increase in the negative sentiment and decrease in neutral, and a minor increase in sadness, anger and surprise.

	Full Corpus	**Sarcastic**	**Not Sarcastic**
Positive	0.28	0.29	0.28
Neutral	0.31	0.27	0.31
Negative	0.41	0.44	0.41

Table 5: Distribution of sentiment across data labelled sarcastic and not sarcastic

This did not remain consistent across topic level distributions when compared to the full corpus, with only topic 9 showing significant deviation with regards to emotion. Topic 9 relates primarily to the ex-CEO of Eskom and showed higher levels of anger with reduced levels of sad and neutral content.

Emotion	Full Corpus	Sarcastic	Not Sarcastic
Anger	15.4	16.1	15.1
Disgust	0.6	0.7	0.6
Fear	4.2	3.5	4.6
Joy	5.3	4.9	5.4
Sadness	23.6	25.7	22.5
Surprise	4.6	6.1	3.8
Neutral	46.3	43.0	48.0

Table 6: Distribution of sentiment across data labelled sarcastic and not sarcastic

Simply looking at the statistics showed that the sarcastic tweets did not have significant sway in the emotion and sentiment analysis. However, it has highlighted issues within the corpus that need to be addressed, for a cleaner representation of the emotions and sentiments within a corpus like this. Through the manual analysis of the sarcastically labelled data, we found that the majority of these tweets expressed more than one emotion and sentiment, often pointed towards different entities or topics. This trait was not unique to the sarcastic subset. This see-saw of emotions and sentiment, poses an issue that needs to be addressed further. Sarcasm tagging has assisted in identifying this in the corpus.

This could also help explain the level of neutrality picked up within the set as well as the major deviation between level of neutrality within sentiment and emotion. While analysis on the most likely sentiments does prove useful, fine-tuning the approach to account for the distribution that can occur within a single document can may yield more insightful results.

7 Conclusion

This study aimed to uncover latent themes and perspectives of South Africans on X in relation to Eskom and loadshedding in 2022. The analysis employed NLP techniques to assign topics, sentiment, emotion and sarcasm metadata onto the tweets for further analysis. The methodology employed in this study provides a simplistic framework to use NLP models to promote an understanding of a text-based corpus such as tweets.

The highlight topics extracted from the corpus related to loadshedding and the various stages and outages surrounding it. Topics arose relating to the ex-CEO, Andre de Ruyter, and calls from the public to declare a state of disaster. Spikes in volumes of tweets related to loadshedding stages were observed around announcement dates for prolonged bouts of blackouts. The emotion and sentiment distributions across topics emphasised the negative emotions that the public felt towards Eskom in 2022.

Sarcasm analysis outlined further work that needs to be performed to better understand and extract information from tweets. Determining the distribution of emotions on a single text document, especially shorter text, is a difficult task considering the limited context that exists. This will be considered for future work.

Future work on the Eskom tweet corpus will entail exploring temporal trends in a multi-year corpus and exploring the nuances of sentiment and sarcasm distributions that can co-exist within a text document.

Acknowledgements

This study was enabled by funding received from the Department of Science, Technology and Innovation (DSTI).

References

Ionuţ-Alexandru Albu and Stelian Spînu. 2022. Emotion detection from tweets using a bert and svm ensemble model.

Kholoud Khalil Aldous, Jisun An, and Bernard J. Jansen. 2019. View, like, comment, post: Analyzing user engagement by topic at 4 levels across 5 social media platforms for 53 news organizations. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01):47–57.

Laurie Butgereit. 2015. An algorithm for measuring relative anger at eskom during load-shedding using twitter. In *AFRICON 2015*, pages 1–5. IEEE.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. Goemotions: A dataset of fine-grained emotions. *Preprint*, arXiv:2005.00547.

Ryan J. Gallagher, Kyle Reing, David Kale, and Greg Ver Steeg. 2017. Anchored Correlation Explanation: Topic Modeling with Minimal Domain Knowledge. *Transactions of the Association for Computational Linguistics*, 5:529–542.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Hussam Habib and Rishab Nithyanand. 2023. The morbid realities of social media: An investigation into the narratives shared by the deceased victims of covid-19. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):303–314.

C.J. Hutto and Eric Gilbert. 2015. Vader: A parsimonious rule-based model for sentiment analysis of social media text.

Vukosi Marivate, Avashlin Moodley, and Athandiwe Saba. 2021. Extracting and categorising the reactions to covid-19 by the south african public-a social media study. In *2021 IEEE AFRICON*, pages 1–6. IEEE.

Merriam-Webster. 2023. Sarcasm.

Saif M. Mohammad and Felipe Bravo-Marquez. 2017. WASSA-2017 shared task on emotion intensity. In *Proceedings of the Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA)*, Copenhagen, Denmark.

Avashlin Moodley and Vukosi Marivate. 2019. Topic modelling of news articles for two consecutive elections in south africa. In *2019 6th International Conference on Soft Computing Machine Intelligence (ISCMI)*, pages 131–136.

Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Jürgen Pfeffer, Daniel Matter, Kokil Jaidka, Onur Varol, Afra Mashhadi, Jana Lasser, Dennis Assenmacher, Siqi Wu, Diyi Yang, Cornelia Brantner, Daniel M. Romero, Jahna Otterbacher, Carsten Schwemmer, Kenneth Joseph, David Garcia, and Fred Morstatter. 2023. Just another day on twitter: A complete 24 hours of twitter data. *Proceedings of the International AAAI Conference on Web and Social Media*, 17(1):1073–1081.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J. Liu. 2019. Exploring the limits of transfer learning with a unified text-to-text transformer. *arXiv preprint*.

Subrata Saha, Md. Imran Hossain Showrov, Md Rahman, and Md Majumder. 2023. Vader vs. bert: A comparative performance analysis for sentiment on coronavirus outbreak. pages 371–385.

Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. 2018. Carer: Contextualized affect representations for emotion recognition. In *Conference on Empirical Methods in Natural Language Processing*.

Claudine Tinsman, Siddharth Arora, and Max Van Kleek. 2023. News hurts: Exploring the link between pandemic-related social media use and trauma symptoms.