Social Media Discourse Analysis on the Tigray War Through Natural Language Processing

MSc Human and Social Data Science Dissertation

UNIVERSITY OF SUSSEX

Author: Jungyeon Lee (277183)

Supervisor: Dr Jack Pay

August 2024

Contents:

Figures and Tables

Abstract

This research aims to delve into the social media discourse surrounding the Tigray War, a representative case of armed conflict in Ethiopia. Given the ongoing conflicts in the region and the need for ceasefire planning and implementation, diverse sources of information are crucial. Social media discourse holds significant potential to provide new insights.

Therefore, this project addresses the following research questions to obtain additional information: the major research questions are "How did social media discourse during the Tigray War evolve across different stages, as analysed through methodologies of Natural Language Processing (NLP): topic modelling, sentiment analysis, and word frequency analysis?" and "Can the outputs of topic modelling, sentiment analysis, and word frequency analysis provide additional information to assist in the preparation and implementation of a ceasefire for the conflict?."

This research utilised natural language processing (NLP) techniques, including NLTK, BERTopic, and a RoBERTa-base sentiment analysis model, to analyse social media data, primarily focusing on YouTube comments. The BERTopic model was developed by optimising its hyperparameters to conduct topic modelling. The analysis showed that peace-related topics increased significantly as the conflict neared its end.

For sentiment analysis, we compared basic models like NLTK and TextBlob with pre-trained models from Hugging Face. The RoBERTa-base sentiment analysis model was chosen due to its high performance. Through this model, sentiment trends were analysed using bounds to calculate values at each phase. It was observed that negative discourses decreased as the Pretoria Agreement drew near.

Lastly, discourses were tokenised using NLTK, and word frequency was analysed to identify the major terms of the Tigray War and peace-related words. The results indicated that as the final peace agreement approached, words and topics related to peace became significantly more frequent.

In conclusion, this study identified the trends in topics, sentiments, and word frequencies during each phase of the Tigray War. By addressing the research questions, it demonstrated that using NLP methods to analyse social media discussions about an armed conflict can provide additional information. This information could serve as evidence to help plan and facilitate ceasefire negotiations.

Acknowledgement

This piece of work began with the effort of bringing data science into humanitarian actions. I would like to express my gratitude to my supervisor, Dr Jack Pay. Without his support, completing this study would have been challenging.

In addition, I hope that further data science research will aim more at humanitarian actions. It will contribute to achieving the UN Sustainable Development Goals (SDGs). I aspire that this work will provide ideas and assist in introducing data science into humanitarian actions.

1. Introduction

The Tigray War was a military confrontation between the Ethiopian federal government and the Tigray People's Liberation (TPLF). This war lasted for two years, from 3 November 2020 to 3 November 2022 (Addis Standard, 2022). According to York Geoffrey (2022) and UNICEF (2022), approximately between 378,000 and 600,000 people died due to the Tigray war and around 2,750,000 civilians were internally displaced during the war. Even if the Tigray War had been ended by the Pretoria agreement signed on 3 November 2022, Ethiopia is still suffering internal and external conflicts such as the War in Amhara (United Nations, 2023).

The persistence of conflicts in Ethiopia poses a threat of humanitarian crises and economic downturns, risking the nation's stability and development (Al Jazeera, 2023). The Ethiopian Ministry of Finance announced that around $20 billion was estimated to rebuild the damages caused by the war (Schipani, 2023). Besides, the Tigray War ruined the relationship with foreign creditors. As a result, the European Union (EU) terminated its financial assistance to Ethiopia in November 2020 on the grounds of human rights abuse (Mwakideu, 2023). Therefore, ceasefire negotiations for current armed conflicts are crucial to resolve the conflicts in Ethiopia. For this purpose, humanitarian actors such as the United Nations (UN) and the African Union (AU) with stakeholders are required to secure enough information and data to facilitate and implement a ceasefire negotiation.

Nowadays. social media is one of the primary spaces where enormous amounts of data exist because it facilitates real-time discussions and engagements among the public (Wafiq, 2023). Consequently, social media users can easily express their opinions and publish their ideas about social issues, politics, and more on the platform (Wallace, 2017). Hence, social media analysis has a significant potential to provide valuable information to humanitarian actors.

At the same time, advanced technical methodologies have emerged and enhanced, making it easier to analyse social media. Natural Language Processing (NLP) allows us to analyse language data such as social media discourse through tokenisation and vectorisation (Jurafsky and Martin, 2008). Moreover, Machine Learning (ML)-based NLP models have been developed to perform sentiment analysis and topic modelling. According to Manica et al. (2019), these tokenised and vectorised data points can be used as input data for ML-based NLP models, enabling computer-based social media discourse analysis.

Hence, this research aims to conduct a social media discourse analysis of the Tigray War, a typical case of an armed conflict in Ethiopia. To enhance the understanding of an armed conflict in Ethiopia, this project collects data from YouTube, one of the main social media platforms.

Following this, the data is examined through NLP and ML methods particularly, sentiment analysis, topic modelling, and word frequency analysis. Furthermore, this research investigates whether the social media discourse analysis can provide further information ① to understand an armed conflict and ② has the potential to prepare and drive a ceasefire to solve the conflict.

This project begins with the research aims and questions section. This section explains the project's purposes, research questions, and how each question will be addressed. The five literature review sections discuss the nature of armed conflict, focusing on the Tigray War, and explain why social media discourses can be utilised to analyse the war using NLP and ML. This provides the background necessary to understand the context of this dissertation.

Consequently, the six methodology sections detail how the data is collected and pre-processed using Python with NLTK and Pandas libraries. Furthermore, these explain how the hyperparameters for BERTopic, used for topic modelling, were determined and how the final model is structured. Additionally, these describe the derivation of the sentiment analysis model for analysing the sentiments of social media discourses in the dissertation. Moreover, these illustrate how the techniques were employed to conduct the analysis, interpret the results, and ultimately address the research questions.

On the other hand, the results section initially presents a descriptive analysis and qualitative analysis of comments through topic modelling, sentiment analysis, and word frequency analysis. Subsequently, it offers several analyses related to research question 2. Lastly, the discussion and conclusion section interprets all analyses to address the two research questions. It also provides insights and recommendations for future research, as well as a discussion of the critical analysis of this project.

The abbreviation for this research is as follows:

TPLF	Tigray People’s Liberation	**EU**	European Union
UN	United Nations	**AU**	African Union
NLP	Natural Language Processing	**ML**	Machine Learning
UCDP	Uppsala Conflict Data Program	**EPRDF**	Ethiopian People's Revolutionary Democratic Front
OPDO	Oromo Peoples' Democratic Organisation	**DL**	Deep Learning
BERT	Bidirectional Encoder Representations from Transformers	**LDA**	Latent Dirichlet Allocation
EM	Expectation-Maximisation	**NLTK**	Natural Language Toolkit
NMF	Non-negative Matrix Factorization	**HDBSCAN**	Hierarchical Density-Based Spatial Clustering of Applications with Noise
VADER	Valence Aware Dictionary and sEntiment Reasoner	**TDF**	Tigray Defense Forces

2. Research Aims and Questions

In this project, the two main aims are: ① to understand discourse topics, sentiments, and word frequencies related to the Tigray War using social media discourse analysis through NLP techniques including BERTopic, NLTK, and an ML-based sentiment model, and ② to investigate whether analysing social media discourse on the Tigray War using NLP methods can provide additional information that could serve as evidence to plan and facilitate ceasefire negotiations. This study defines additional information as data that has not been formally reported or released by the parties involved in the conflict, nor has it been officially announced by any stakeholders.

For the first and second research goals, the research questions and sub-research questions have been formulated as follows:

1) Research Question: How did social media discourse during the Tigray War evolve across different stages, as analysed through methodologies of NLP: topic modelling, sentiment analysis, and word frequency analysis?

1-1) What kind of topics emerged from social media discourse during the Tigray War through BERTopic, which is a topic modelling method, and what proportion of these topics was present in each stage of the war?

1-2) What is the overall trend of sentiment in discourses regarding the Tigray War as revealed by sentiment analysis using an ML-based sentiment model, and what proportion of these sentiments was expressed in each phase of the war?

1-3) What are the major terms within the entire social media discourse on the Tigray War and what are the trends of these terms in each conflict phase, as examined by the top 30-word frequency analysis through NLTK?

2) Research Question: Can the outputs of topic modelling, sentiment analysis, and word frequency analysis provide additional information to assist in preparing and implementing a ceasefire for the conflict?

2-1) Does the proportion of peace-related topics such as "Peace" and 'Ceasefire' increase as the conflict progresses towards the final ceasefire stage?

2-2) When comparing the peace talks phase with the preceding phase, is there a decrease in the proportion of negative discourses regarding the war?

2-3) Is there an observed increase in the proportion of the frequency of words related to peace and ceasefire, such as 'peace', 'stop', 'ceasefire', 'cessation', and similar terms, as the Pretoria Agreement approaches?

This project seeks to understand the Tigray War through the first research question, utilising social media discourse. Meanwhile, the second research question examines the potential of the outputs of social media discourse analysis that can be additive information for conflict resolution. However, it is challenging to assess its potential because it does not have standard criteria and enough relevant studies. Therefore, this project operates on three main assumptions: ① the peace-related topics and words mean that the public thinks the war has to be resolved ② negative discourses mostly imply that they contain negative content and use more negative words than peaceful discourses or situations ③ discourses desiring peace are typically composed of positive content and words, so negative discourses are contrary to those seeking peace.

3. Literature Review

3-1. Armed Conflict

An armed conflict is defined as a disagreement over government territory or related issues, where two parties use armed force resulting in at least 25 battle-related casualties within a year (UCDP, 2006; Dixon, 2009). Between these two parties, one must be the government of a state.

Armed conflict brings numerous adverse impacts on both internal and external stakeholders (Oetzel and Getz, 2011). In 2014, 11 conflicts occurred and around 101,400 people died because of the armed conflicts (Pettersson and Wallensteen, 2015). Not only this but also an armed conflict possesses serious collateral damage like economic loss and violation of human rights (Serneels and Verpoorten, 2013).

A ceasefire acts as a way to resolve conflicts by halting a war, with both sides agreeing to stop fighting, often with mediation by a third party such as the United Nations (Fortna, 2018). Ceasefires can involve countries or groups that are not officially recognised as states. Peace agreements also may be used to address short-term needs like providing aid to reduce the impact of a conflict or move towards a peaceful resolution (Clayton et al., 2022).

Various sources of information must be thoroughly referenced to prepare for peace talks, from the timing of the negotiations to the detailed content of the ceasefire, to improve the durability of the ceasefire (Fortna, 2018). These sources include official reports and documents published by conflict parties and stakeholders, comprehensive analysis results of the conflict parties and national and local stakeholders, public opinion, and so on (UN DPPA, 2022).

3-2. Tigray War

Since the cessation of the Ethiopian Civil War in 1991, the Ethiopian People's Revolutionary Democratic Front (EPRDF), led by the TPLF, had been in charge of governing the country (Burke, 2020). Nonetheless, the former Prime Minister, Hailemariam Desalegn, from EPRDF, resigned because of public discontent in 2018 (Al Jazeera, 2018). Subsequently, Abiy Ahmed, the chairman of the Oromo Peoples' Democratic Organisation (OPDO), was elected as Ethiopia’s Prime Minister in 2018 and has held the position since then. Abiy Ahmed was awarded the Nobel Peace Prize for his dedicated efforts in conducting peace talks with Eritrea, which led to the resolution of the longstanding armed conflict between Ethiopia and Eritrea (Burke and Henley, 2019). However, in 2019, Abiy Ahmed attempted to integrate most of the Ethiopian political parties into the Prosperity Party, but the TPLF disagreed with this proposition (Balehegn, 2021). This disagreement became the catalyst for the beginning of the Tigray War.

The Tigray War commenced in November 2020, primarily initiated by tensions between the Tigray regional government and the federal government about the Tigray region's local elections (Caruso and Akamo, 2024). Particularly, a key trigger was the Tigray regional government's opposition to the federal government's policy that postponed the general elections due to the COVID-19 pandemic (Kelecha, 2024). The Tigray regional government decided to hold separate local elections. This incident accelerated existing tensions between the Tigray region and the federal government, which had originated from disputes that happened by the Prosperity Party (Blanchard, 2021). These tensions escalated rapidly, culminating in armed conflict between the Ethiopian federal government forces and the TPLF, and the Tigray War began.

The Tigray War persisted for 2 years from 3 November 2020 to 3 November 2022, and it was resolved through the Pretoria Agreement (Mekonen, 2023). The peace agreement was mediated by AU (African Union, 2022). In the peace agreement, both the TPLF and the Ethiopian federal government agreed to a "permanent cessation of hostilities", the end of the Tigray War (Winning and Cocks, 2022).

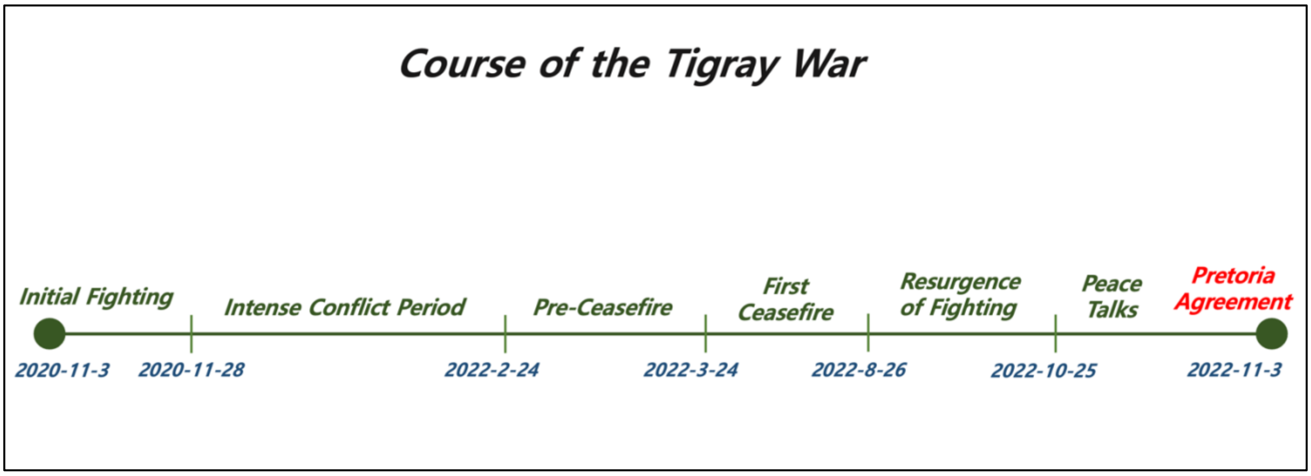


Figure 1. Course of the Tigray War

The Tigray War is a representative case of armed conflict in Ethiopia. In this project, the Tigray War is divided into 6 phases to comprehend the progression of the war over time. The first phase, termed the 'Initial Fighting', encompasses approximately the first month of the conflict. The second phase, from, 29th November 2020, after when the Ethiopia forces' attack on Mekelle and the 'Mekelle Offensive' occurred thereby the war intensified (Center for Preventive Action, 2022), to one month before the first ceasefire, will be called the 'Intense Conflict Period.' It spans around one year and three months.

Following this, the 'Pre-Ceasefire' and 'First Ceasefire' will be the third and fourth phases, spanned the first ceasefire period, which commenced on 24th March 2022, and one month before the first ceasefire. Subsequently, the period from the breakdown of the first ceasefire to the peace talks conducted by the AU, aimed at mediating the final cessation, is termed the 'Resurgence of War', constituting the fifth phase. Lastly, the 'Peace Talks', which steered toward the Pretoria Agreement, will be the last phase of the war.

The Tigray War already reached a ceasefire agreement on 2 November 2022, and the peace agreement has been maintained since then. Hence, analysing the social media discourse on the war and evaluating whether the results of this analysis provide various insights into understanding the war and reflect public reactions to the course of the war has the potential to answer whether social media discourse analysis can be utilised as a source of data to prepare for a ceasefire.

3-3. Social Media Discourse

The concept of 'social media' is relatively recent, but it has rapidly become a famous place for public discussions (Leppanen, Westinen and Kytola, 2016). According to Figure 2, in 2024, globally, 62.3% of people, equivalent to 5.04 billion out of a total population of 8.08 billion, have an account on at least one social media platform.

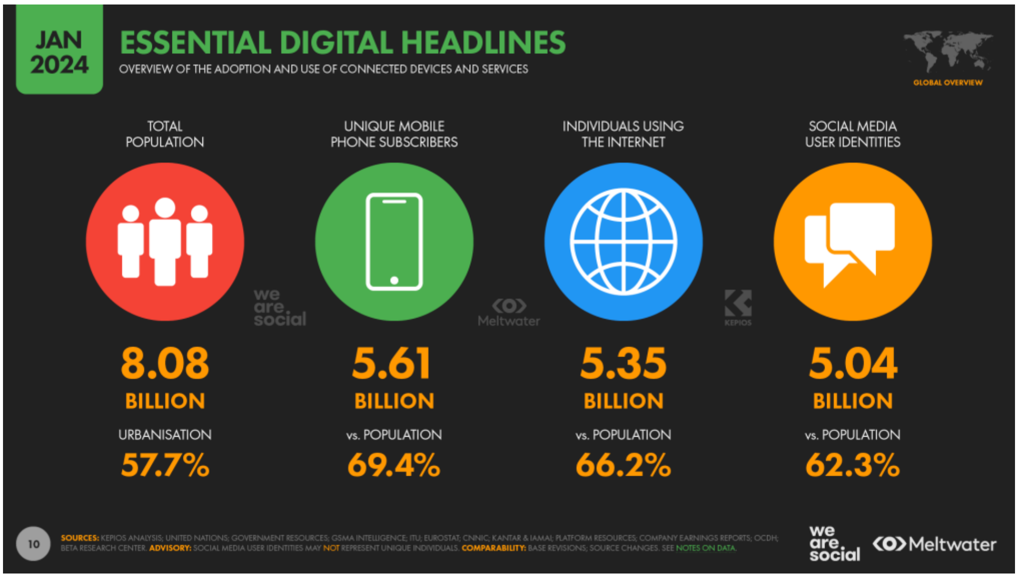


Figure 2. Essential Digital Headlines in January 2024

Source: https://wearesocial.com/uk/blog/2024/01/digital-2024-5-billion-social-media-users/

Besides, according to Figure 3 by Hughes (2019), the monthly active users on Facebook are 2.3 billion and YouTube has 1.9 billion. Furthermore, Domo's Data Never Sleeps 10.0 report (2022) highlights that 1.7 million pieces of Facebook content, 347,200 Twitter tweets, and 500 hours of YouTube video are generated every minute. This data, coupled with the statistic that globally 62.3%

of people have at least one social media account in 2024, underscores the significant role of social media as a platform for individuals to express their viewpoints on various issues of interest.

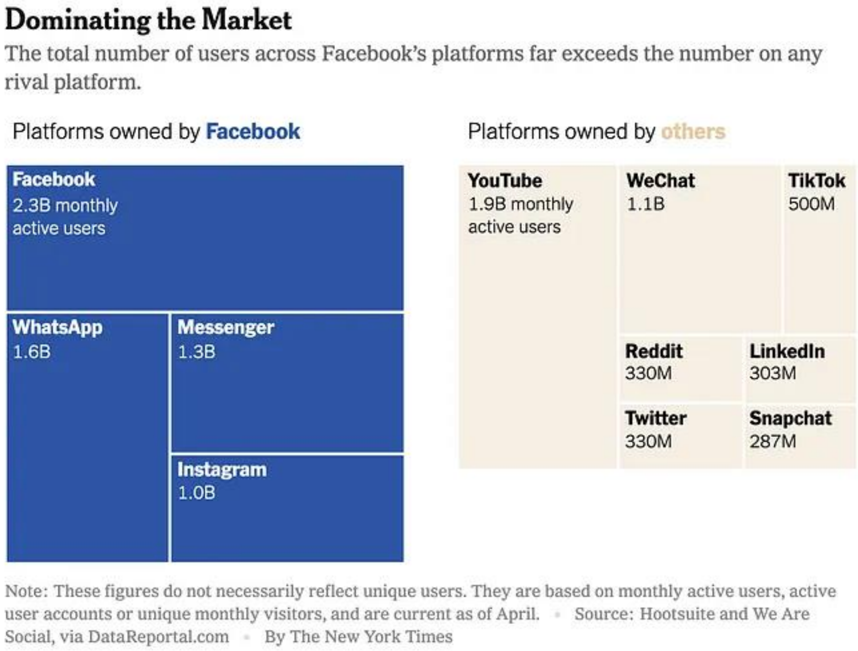


Figure 3. SNS Market Information

Source: https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html

Recently, social media have transformed how people communicate, interact, and access various kinds of data and information (Tucker et al., 2017). According to Tucker et al. (2017), social media allows everyone, including those excluded from mainstream politics, to express their views and opinions as a discourse. Thus, social media platforms have taken a major role in public discourse, and these are now regarded as important channels, much like traditional media such as news media and television (Chadwick, 2013). In a democracy, decision-makers are influenced by public opinion, and therefore public support is crucial for the continuation of conflict actions (Simmons, 2009). Hence, analysing public opinion through social media can offer valuable insights into attitudes and perspectives regarding the conflict.

Nevertheless, it is important to note that social media analysis may not comprehensively explain all aspects of a conflict due to the presence of propaganda and misinformation from fake news. Fake news causes distrust in governments and professionals, fosters anxiety, and can directly and adversely impact people's lives (Rocha et al., 2021). The research titled "The Spread of True and False News Online" by Vosoughi, Roy, and Aral (2018) demonstrated that politics was the largest sector in which fake news emerged from their dataset, with terrorism and war also ranking 4th.

However, it has strong potential to be utilised as a powerful method to obtain additional information beyond officially reported data regarding a conflict. Therefore, it is necessary to examine the potential value of social media discourse whether it provides information on understanding a conflict using the case of the Tigray War and whether it can be used as an additional resource to help prepare and implement a ceasefire.

3-4. Machine Learning and Natural Language Processing

Machine Learning is categorised as a component of artificial intelligence, encompassing the ability of a machine to imitate intelligent human behaviour (Brown, 2021). Machine Learning can be trained using data without predetermined equations or specific instructions (Tondak, 2020).

Within ML, there are two main approaches: 'Supervised Learning' and 'Unsupervised Learning'. Supervised Learning works with labelled training data, where pre-defined classifications or actual dependent values are provided, while unsupervised learning operates with unlabelled datasets (Sathya and Abraham, 2013).

Besides, Deep Learning (DL) is considered a subset of ML, especially utilising multi-layered neural networks (LeCun, Bengio and Hinton, 2015). In DL, semi-supervised learning is also employed alongside supervised and unsupervised learning methods.

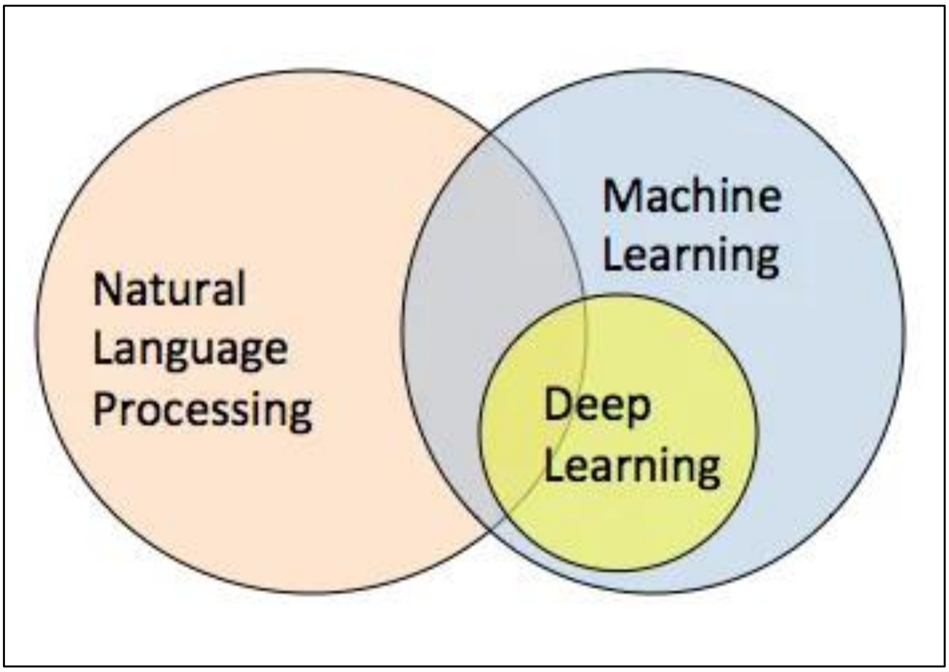


Figure 4. Machine Learning and NLP

Source: https://medium.com/syntax-and-semantics/natural-language-processing-vs-machine-learning-vs-deep-learning-ba072ed242ca

On the other hand, Natural Language Processing is a field at the intersection of computer science and information retrieval, focusing on enabling computers to understand human language and manipulate it (Liu, Lin and Sun, 2020). Its main goal is to process natural language data, such as written text or spoken words, using statistical and neural network-based approaches. According to Figure 4, NLP closely intertwines with ML, utilising ML to construct and refine its model such as sentiment analysis and topic modelling.

Through NLP, text data can be vectorised and used as input for Machine Learning (Jurafsky and Martin, 2008). This makes NLP a fundamental method for analysing social media discourse using ML-based methodologies (Kowsari et al., 2019).

These days, several ML-based NLP methods have been invented. Bidirectional Encoder Representations from Transformers (BERT) is a representative and famous model that employed ML specifically DL to develop this NLP model. BERT consists of an extensive model architecture, built upon self-attention layers, enabling the model to learn relationships between individual words and across entire sentences (Devlin et al., 2019). BERT stands out as the leading-edge technology in NLP, from text modelling to question answering and translation (Sharma et al., 2024).

Therefore, this dissertation employs BERTopic, which is based on the architecture of BERT, in order to perform topic modelling on social media discourse on the war. In addition, this research attempts to find an ML-based sentiment analysis model by comparing the outputs of open-source models to capture the sentiments of the social media discourse.

3-5. Previous Literature on Social Media Discourse

There is no relevant research on social media discourse that has specifically targeted an armed conflict case. However, Törnberg and Törnberg (2016) revealed the patterns of representation regarding the terms "Muslim" and "Islam" through topic modelling and critical discourse analysis. In this research, Latent Dirichlet Allocation (LDA) was employed for topic modelling. LDA is an ML-based topic modelling method that utilises both Bayesian methods and an Expectation-Maximisation (EM) algorithm during the training phase (Blei, Ng and Jordan, 2003). According to the study, Muslims are often portrayed as intertwined with conflict, particularly regarding topics such as terrorism, sexual abuse, and the perception of women in Islam in social media discourse. Their study demonstrated that social media discourse can contribute to understanding the public's perception of specific topics in social media.

Since no research papers exist that investigate whether social media discourse can be used as a source of information for ceasefire negotiations, this dissertation can play a pivotal role in exploring this possibility and providing research ideas.

4. Methodology

4-1. Technical Requirement, Experiment Set-Up and Ethical Considerations

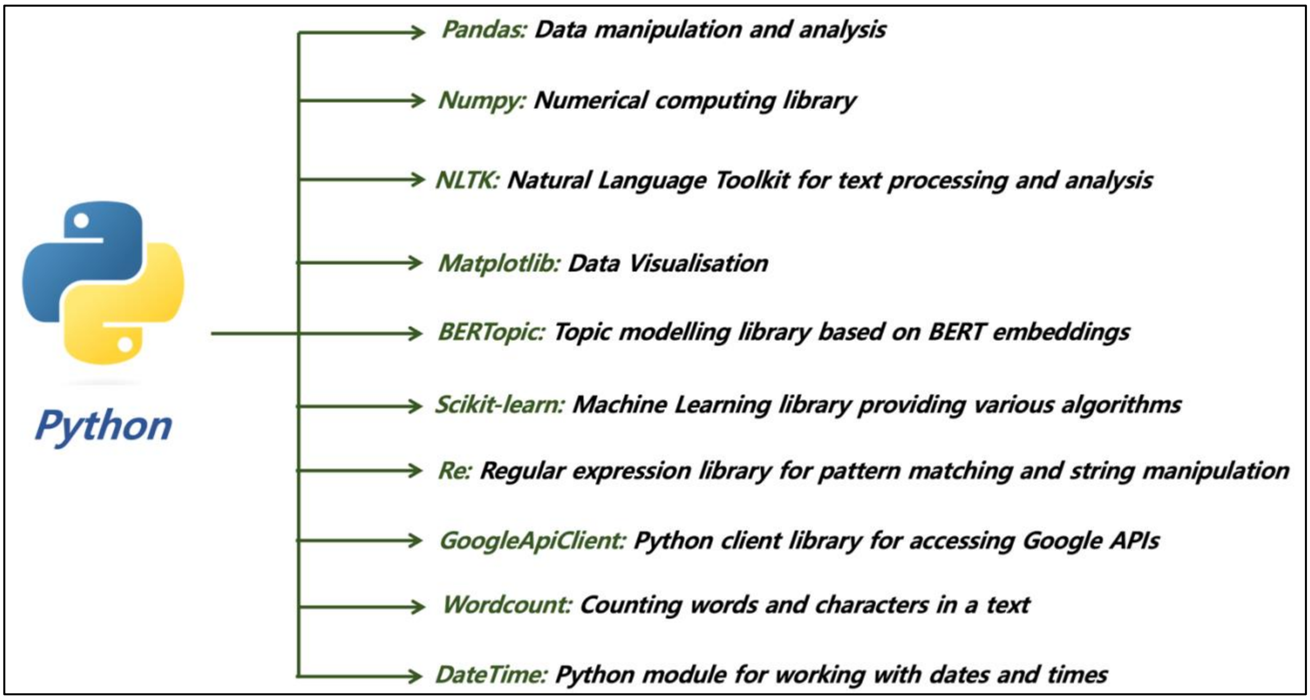


Figure 5. Technical Requirement: Python Libraries

In this project, Python is opted as a primary programming language. To conduct data collection, data preprocessing, word frequency analysis, topic modelling, and sentiment analysis, several Python libraries are employed: NLTK, Pandas, NumPy, Matplotlib, GoogleApiClient, Scikit-learn and so on. The detailed purpose of each library is demonstrated in Figure 5.

Likewise, it is vital to establish certain definitions to implement this research. Firstly, "data" refers to a discourse about the Tigray War found on YouTube and "dataset" is a collection of this data. Moreover, the Tigray War is split into 6 stages, purpose of this task was mentioned in the 3-2 section. In addition, the datasets undergo a preprocessing phase to refine and cleanse them to ensure their relevance and significance.

This project strictly considers and follows ethical considerations to avoid any ethical issues. The data collection was performed solely using YouTube API, which reduces the risk of violating data ethics. This method makes sure this project keeps the ethical standards by operating within the confines of the platform's designated access protocols. Thereby, it prevents unauthorised data acquisition or privacy violation. Furthermore, the project strictly follows all clauses from the ethical standards called 'the Code of Conduct' by BSC – The Chartered Institute for IT in Appendix 1. These standards guide this project to avoid the occurrence of any ethical issues. Overall, this research ensures that no ethical violations would arise by considering ethical standards throughout the entire stages of research: data collection, data preprocessing, modelling, and data analysis.

4-2. Data Collection

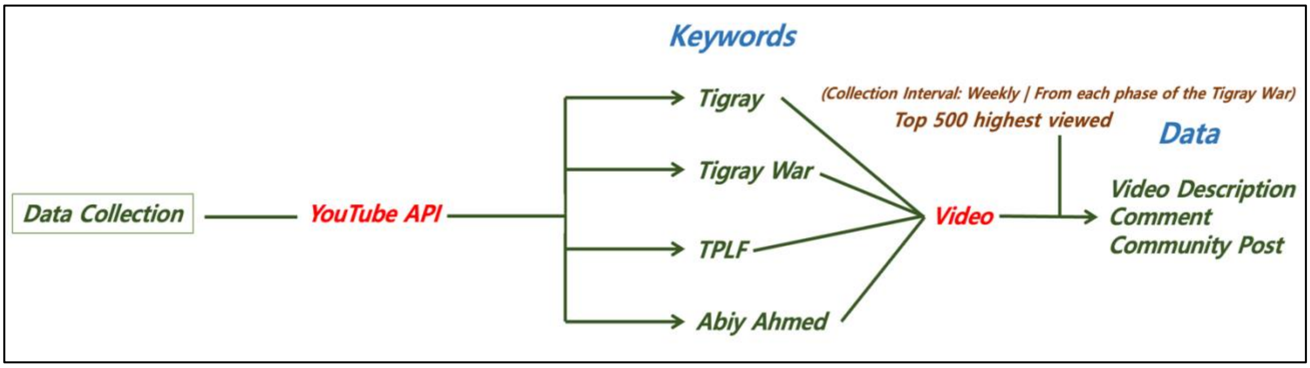


Figure 6. Data Collection Process

To compile political discussions from social media, YouTube was chosen as the primary source for data extraction. According to Belay et al. (2020), in their study on 'social media use practices in the Ethiopian socio-political landscape in Ethiopia in 2020', Facebook dominated as the most popular platform, with a usage rate of 53.99 percent. YouTube followed behind, accounting for 17.73 percent of usage. However, Meta is currently not accepting additional academics to access their data. Therefore, YouTube remains accessible for academics to gather data through its API, making it the next viable option.

Keywords related to the war, such as 'Tigray War', 'Tigray', 'TPLF', and 'Abiy Ahmed' were used for data collection. 'Tigray War' is the official title of the conflict, and 'Tigray' must be included as it is the region where the war takes place. Furthermore, 'TPLF' is a main party in the conflict, fighting against Ethiopia's current Prime Minister, 'Abiy Ahmed', making these keywords essential.

It is important to note that other keywords may not be directly relevant to the core aspects of the conflict. In the context of social media data collection, selecting the right keywords is essential for ensuring the quality and relevance of the data obtained (Lian et al, 2022). Hence, this research decided to focus on the major elements of the war, such as its official name, the region, and the major actors and stakeholders. Thus, this study concluded that using these four keywords can effectively gather data directly related to the Tigray War, data from the Tigray region and comprehensive data on two major stakeholders of the conflict.

The data collection process encompassed not only video descriptions and comments from videos related to these keywords but also community posts from the channels that uploaded such videos. Data were gathered weekly during each phase of the war, focusing on the top 500 highest-viewed videos.

The data collection was initiated on the 16th of July 2024 and finalised on the 20th of July 2024. Throughout this process, data from a total of 68,249 YouTube users was gathered. YouTube user IDs were replaced with pseudonymised IDs upon collection, and after calculating the total number of participants, these IDs were immediately dropped from the dataset.

The data collected from each phase were merged into a single dataset, resulting in a total of six datasets. The number of data points in each dataset is described in Table 1.

No	Name	Period	Number
1	Initial Fighting	2020.11.3. ~ 2020.11.28.	195,030
2	Intense Conflict Period	2020.11.29 ~ 2022.2.24.	1,274,844
3	Pre-Ceasefire	2022.2.25 ~ 2022.3.24.	78,852
4	First Ceasefire	2022.3.25. ~ 2022.8.26.	385,408
5	Resurgence of Fighting	2022.8.27. ~ 2022.10.25.	162,856
6	Pease Talks	2022.10.26. ~ 2022.11.3.	111,046

Table 1. Data Points of Each Dataset

These datasets were pre-processed using NLTK, specified in the pre-processing section, and subsequently integrated into a single dataset named the 'Tigray War dataset' for the project. This dataset was stored on OneDrive, authorised by the University of Sussex during the research process. It was then discarded upon the completion of the entire dataset analysis.

4-3. Natural Language Toolkit (NLTK)

The Natural Language Toolkit, often referred to as NLTK, offers a range of libraries and tools for statistical NLP in English written and some languages such as Portuguese, Arabic, Spanish and so on (Yao, 2019). It provides essential and powerful NLP functions such as classification, stemming, tagging, and tokenisation.

Based on tokenisation using NLTK, this study analyses word frequency in the entire discourse as well as within each stage of the war. In addition, NLTK is employed to build a basic sentiment analysis model to compare other models and implement tokenisation and stop-word handling in the data pre-processing stage.

4-4. Data Pre-processing

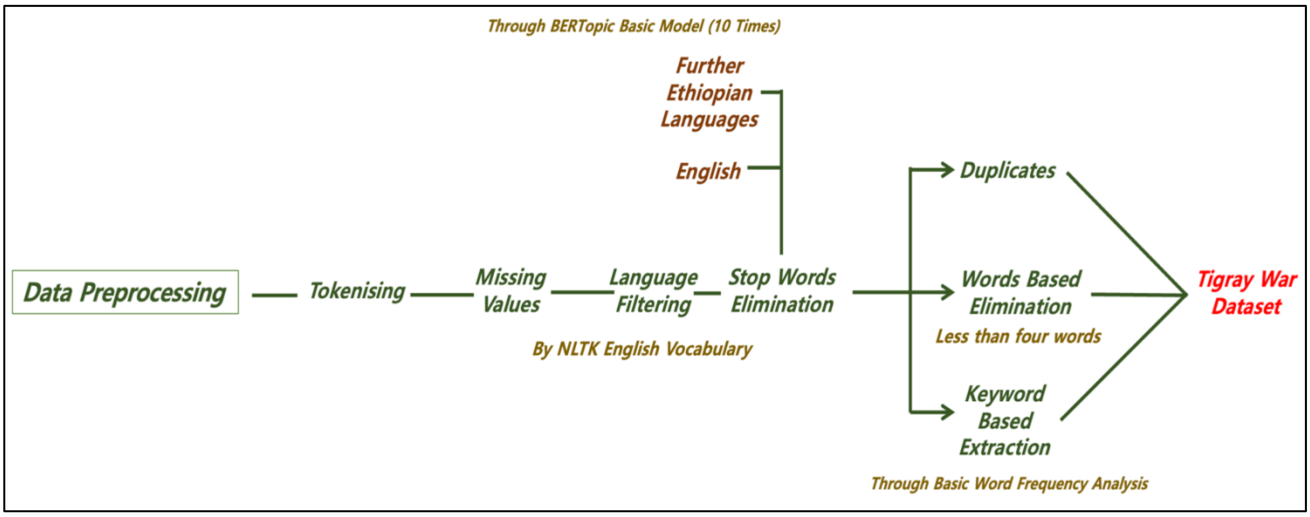


Figure 7. Flow of Data Pre-processing

This project handles text-based data. Thus, full-text data were tokenised to become input data. Afterwards, the missing values, 1,994,485 out of 2,208,036 data points from six datasets, were dropped. The reason for the existence of a large number of missing values is due to the process of merging comments, community posts, and video descriptions into a single data frame after data collection. Moreover, since more than 80 languages exist in Ethiopia, only English vocabulary was used (Ethnologue, 2013). Analysing various languages in text-based datasets is challenging due to the inherent complexity of human languages (Gudivada, Rao, and Raghavan, 2015). Hence, any words not in the English vocabulary were excluded based on the NLTK basic English dictionary. This NLTK basic English dictionary also contains the keywords in Table 3 below.

Furthermore, English stop-words were eliminated from the text in the dataset. In addition, due to the difficulty in distinguishing between English and some Ethiopian language-based words using English alphabets, certain words in Ethiopian languages were removed as stop-words. To implement this, the six datasets were initially merged after only removing missing values, prior to any other pre-processing. Subsequently, topic modelling was conducted using the BERTopic model with default parameters. During this process, words from Ethiopian languages that use the English alphabet were identified in some topics. This identification was repeated ten times to build a stop-words dictionary specifically for Ethiopian languages. Using this dictionary, words from Ethiopian languages that use the English alphabet were removed. The constructed dictionary contains 104 words and this is detailed in Table 2.

Stop-words Dictionary (English Alphabet-Based Ethiopian Languages)							
No	**Vocabulary**	**No**	**Vocabulary**	**No**	**Vocabulary**	**No**	**Vocabulary**
1	iyo	27	kala	53	ya	79	hin
2	ka	28	wax	54	ta	80	rabbi
3	ay	29	ama	55	ye	81	kan
4	ah	30	marka	56	ante	82	nu
5	la	31	che	57	gin	83	kana
6	ha	32	bara	58	hulu	84	ani
7	kale	33	de	59	ga	85	fi
8	wa	34	wey	60	aho	86	al
9	si	35	en	61	hula	87	alula
10	di	36	badam	62	shi	88	ly
11	dan	37	kea	63	bel	89	blo
12	sile	38	gena	64	baa	90	naga
13	badan	39	na	65	hala	91	ey
14	bur	40	dal	66	dib	92	ale
15	ilka	41	mise	67	bal	93	manta
16	ahey	42	lama	68	rag	94	haya
17	ayu	43	nay	69	yang	95	waar
18	anba	44	aa	70	io'	96	sie
19	sidi	45	harka	71	dha	97	san
20	gara	46	yar	72	kalo	98	jeer
21	marae	47	lo	73	dari	99	ba
22	mid	48	dhan	74	bay	100	aha
23	mar	49	amba	75	kula	101	el
24	es	50	se	76	con	102	un
25	las	51	para	77	pais	103	hay
26	das	52	ist	78	ich	104	den

Table 2. Stop-words Dictionary (English Alphabet-Based Ethiopian Languages)

Following this, duplicates were identified and deleted from the dataset. The total number of duplicates found across the six datasets was 98,227. This indicates a significant tendency for text data on social media to contain a considerable number of duplicates.

In addition, text data containing fewer than four words were excluded, as it is challenging to classify such brief texts as discourses. This threshold was set by the researcher due to the absence of an official standard. Subsequently, a word frequency analysis on the combined Tigray War Dataset was implemented to identify the top 100 most frequently occurring words. Then, fourteen keywords related to the Tigray War from this list were extracted, specified in Table 3. It is important to note that three keywords, 'Forces', ‘Battle’ and 'Ceasefire', were added as the researcher deemed them highly relevant to this research, despite not appearing in the top 100 keywords.

No	Keywords	Basis	No	Keywords	Basis
1	Tigray	Top 100	8	Eritrea	Top 100
2	Ethiopia	Top 100	9	Ahmed	Top 100
3	TPLF	Top 100	10	Peace	Top 100
4	War	Top 100	11	Military	Top 100
5	Abiy	Top 100	12	Forces	Researcher discretion
6	Amhara	Top 100	13	Battle	Researcher discretion
7	Government	Top 100	14	Ceasefire	Researcher discretion

Table 3. Tigray War Keywords

Afterwards, any discourse that does not include at least one of these keywords was excluded from this research. This approach is crucial for the project as it enables us to analyse social media discourses relevant to the Tigray War. It is essential to highlight that data points are not deleted during language filtering and stop-word handling, only the relevant words were removed. The data points subjected to these methods were removed in subsequent stages if they were deemed invalid.

Dataset	Initial	Missing Value	Language Filtering	Stop-words	Duplicates	Words Count	Keywords Filtering
Initial Fighting	195,030	25,047 (-169,983)	25,047 (-0)	25,047 (-0)	10,804 (-14,243)	6,451 (-4,353)	4,237 (-2,214)
Intense Conflict Period	1,274,844	143,906 (-1,130,938)	143,906 (-0)	143,906 (-0)	72,517 (-71,389)	48,262 (-24,255)	30,004 (-18,258)
Pre-Ceasefire	78,852	4,465 (-74,387)	4,465 (-0)	4,465 (-0)	2,513 (-1,952)	1,384 (-1,129)	806 (578)
First Ceasefire	385,408	19,996 (-365,412)	19,996 (-0)	19,996 (-0)	11,251 (-8,745)	6,875 (-4,376)	4,067 (-2,808)
Resurgence of Fighting	162,856	14,102 (-148,754)	14,102 (-0)	14,102 (-0)	9,104 (-4.998)	5,817 (-3,287)	3,877 (-1,940)
Pease Talks	111,046	6,035 (-105,011)	6,035 (-0)	6,035 (-0)	4,142 (-1,893)	2,676 (-1,466)	1,766 (-910)

Table 4. Data Pre-processing Result (Data Points)

In conclusion, after six steps of data pre-processing, six datasets were merged into a single dataset called the "Tigray War dataset." This dataset has been stored at OneDrive. Phases 3 and 6, known as 'Pre-ceasefire' and ‘Peace Talks’, span only one month, resulting in relatively fewer data points. In contrast, Phase 2, the 'Intense Conflict Period', covers 15 months, thereby yielding a greater quantity of gathered data compared to other stages of the war. The number of data points removed at each stage is demonstrated in Table 4. As a result, a total of 44,757 data points were secured for conducting topic modelling, sentiment analysis, and word frequency analysis.

4-5. Topic Modelling

4-5-1. BERTopic

Topic modelling is often utilised in text-mining tools to identify the hidden semantic structures within a body of text (Srivastava and Sahami, 2009). Topics generated by such models consist of clusters of related words. This is based on a mathematical framework particularly the statistics algorithms to discover the latent semantic architectures of the text body (Landauer et al., 2007).

A neural topic model, BERTopic, can represent words as multi-dimensional vectors, capturing contextual information effectively (Grootendorst, 2022), which has been noted as more efficient for topic modelling compared to prior methodologies. According to Egger and Yu (2022), BERTopic demonstrates superior performance over models such as LDA and NMF (Non-negative Matrix Factorization) in terms of topic coherence and interpretability when analysing social media data from Twitter. This suggests that BERTopic is well-suited for examining social media content and other intricate textual sources where context and nuance are crucial.

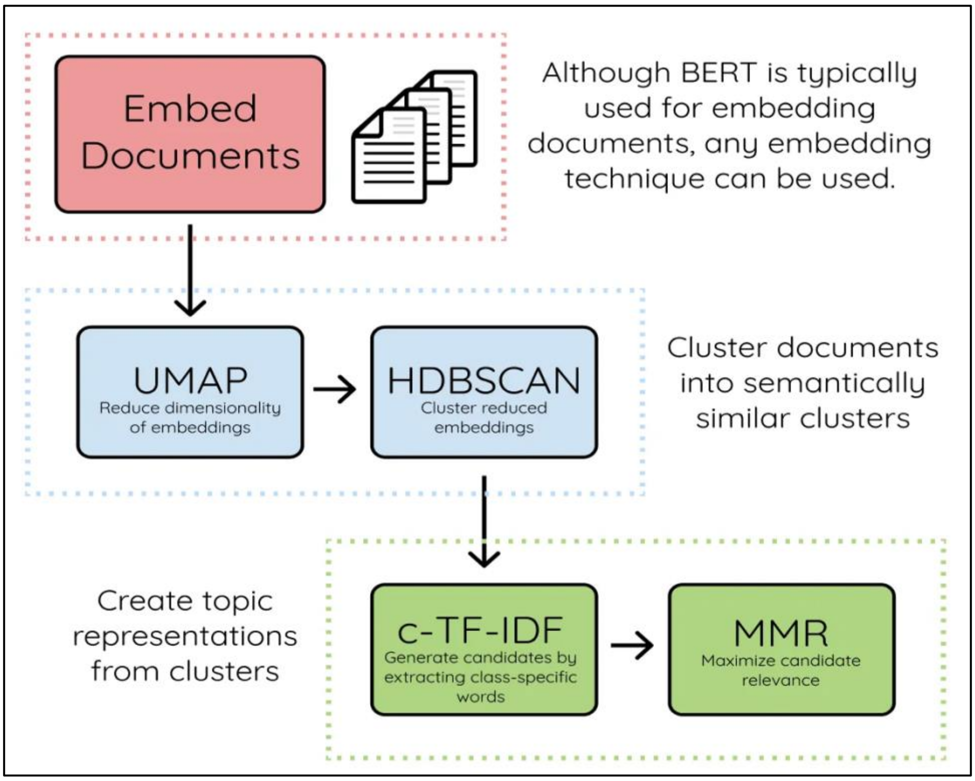


Figure 8. BERTopic Algorithm

Source: https://medium.com/data-reply-it-datatech/bertopic-topic-modeling-as-you-have-never-seen-it-before-abb48bbab2b2

Grootendorst (2022) explained that BERTopic includes three major steps: document embedding, document clustering and topic representation. Document embedding involves transforming documents into dense vectors in a lower-dimensional space, keeping their semantic meaning and structure. Based on the document embedding, document clustering groups similar topics together as a cluster.

Hence, BERTopic is utilised to conduct topic modelling on the Tigray War dataset. This approach helps identify the topics within social media discourse surrounding the Tigray War.

4-5-2. Hyperparameter Tuning

Hyperparameters	Value
umap_n_neighbors	10 | 15 | 20
umap_n_components	5 | 10 | 15
hdbscan_min_cluster_size	5 | 10 | 15
top_n_words	10 | 15 | 20

Table 5. Possible Hyperparameters

To find the best combination of hyperparameters for BERTopic, all possible combinations of hyperparameters were selected, described in Table 5, and compared using 5,000 randomly sampled data points from the dataset.

The all-MiniLM-L6-v2 model from the SentenceTransformer library was chosen to balance computational efficiency and semantic understanding. This model is known for producing high-quality sentence embeddings with relatively low computational overhead, making it suitable for large-scale text data (MLflow, 2023). As this research involves handling large-scale text data from social media, using the all-MiniLM-L6-v2 as the embedding model for BERTopic is considered appropriate.

For UMAP hyperparameters, setting the 'n_neighbors' parameter within the range of 10 to 20 is informed by the need to balance local and global structure preservation. Smaller values of this parameter may preserve local structures better, while larger values can capture global structures more effectively (McInnes et al., 2020). The selected range aims to explore this balance comprehensively. In addition, the 'n_components' parameter determines the number of dimensions in the reduced space. Values between 5 and 15 are selected to investigate how different levels of dimensionality reduction affect clustering performance. Lower dimensional spaces can simplify the clustering task, whereas higher dimensions might retain more nuanced information (Sanguansat, 2012).

Similarly, the 'min_cluster_size' parameter from HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) is utilised for clustering the reduced embeddings. Values from 5 to 15 are opted to examine the model's ability to identify both small and large clusters. According to Campello et al. (2013), smaller min_cluster_size values can detect fine-grained clusters, while larger values prevent the formation of overly fragmented clusters. This range allows for a thorough exploration of the clustering granularity.

Regarding the 'top_n_words' parameter from BERTopic, it is recommended to keep this value below 30 and preferably between 10 and 20 (Grootendorst, 2021). Therefore, 10, 15, and 20 values are used for grid search.

Lastly, according to Grootendorst (2021), if the data size exceeds 1,000,000, the 'min_topic_size' parameter for BERTopic should be adjusted to more than 100. However, since the data size in this research is relatively smaller, a total of 44,757 data points, the default value of 10 is deemed sufficient.

Name	C_V	C_{UMass}	C_{NPMI}
Segmentation	S^{one}_{set}	S^{one}_{pre}	S^{one}_{one}
P. Calculation	$P_{sw(110)}$	P_{bd}	$P_{sw(10)}$
C. Measure	$\tilde{m}_{\cos(nlr)}$	m_{lc}	m_{nlr}
Aggregation	σ_a	σ_a	σ_a

Figure 9. Topic Coherence Measures

Source: https://towardsdatascience.com/understanding-topic-coherence-measures-4aa41339634c

The model's performance was assessed using coherence scores. Topic coherence refers to the average or median of pairwise word similarities among the top words within a given topic (Lau et al., 2014). These word similarities rely on external data, which is not utilised during topic modelling. Several automatic coherence measures exist, including NPMI, LCP, C_v and UMass.

According to Röder et al. (2015), C_v is generally the most stable method for measuring coherence. Meanwhile, Lau et al. (2014) found that NPMI effectively replicates human performance and is deemed the best method for measuring coherence. Hence, this research employed both NPMI and C_v coherence scores to evaluate the topic model's performance, as they are considered the most robust methods for calculating coherence scores.

No	Neighbours	Components	Minimum Cluster Size	Number of Top Words	Score
1	10	10	5	10	0.581
2	10	15	5	10	0.572
3	20	10	5	10	0.571

Table 6. Models Performance (C_v)

No	Neighbours	Components	Minimum Cluster Size	Number of Top Words	Score
1	15	10	15	10	0.012
2	15	5	15	10	0.010
3	15	15	15	10	0.009

Table 7. Models Performance (NPMI)

Nevertheless, relying completely on coherence scores to determine hyperparameters is risky, as quantitative methods do not always align with all data and models and may introduce bias. It is preferable to assess the model's performance qualitatively together with quantitative methods. Consequently, this research visualised the output of two models constructed under the hyperparameters with the C_v coherence score and the highest NPMI coherence score. The coherence scores are described in Table 6 and Table 7, and the models' visualisation outputs are detailed in Figures 10 and 11.

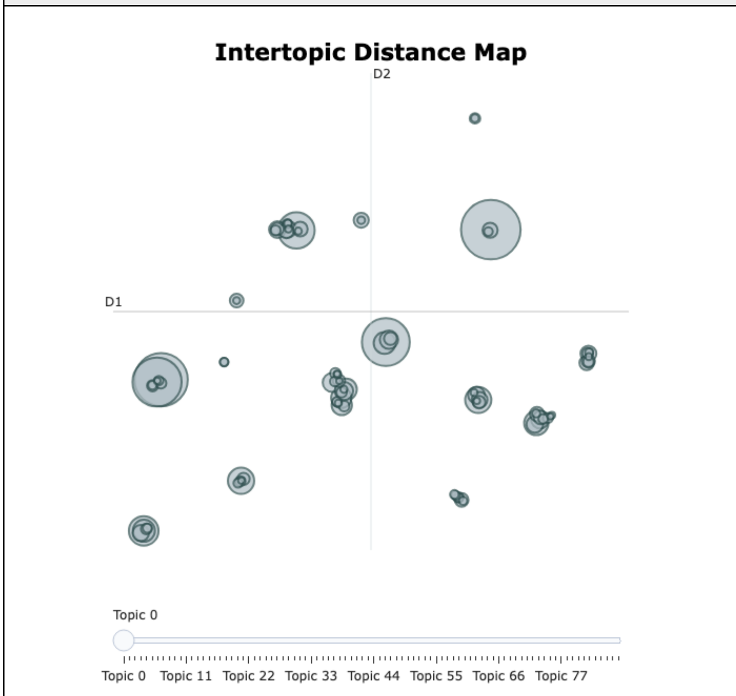
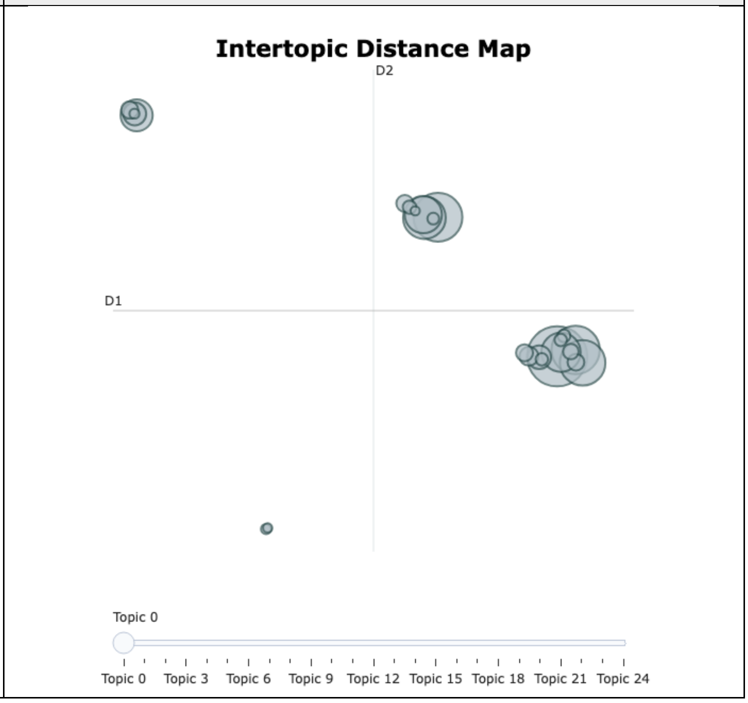
Highest C_v Coherence (Model 1)	Highest *NPMI* Coherence (Model 2)
	

Figure 10. Intertopic Distance Map

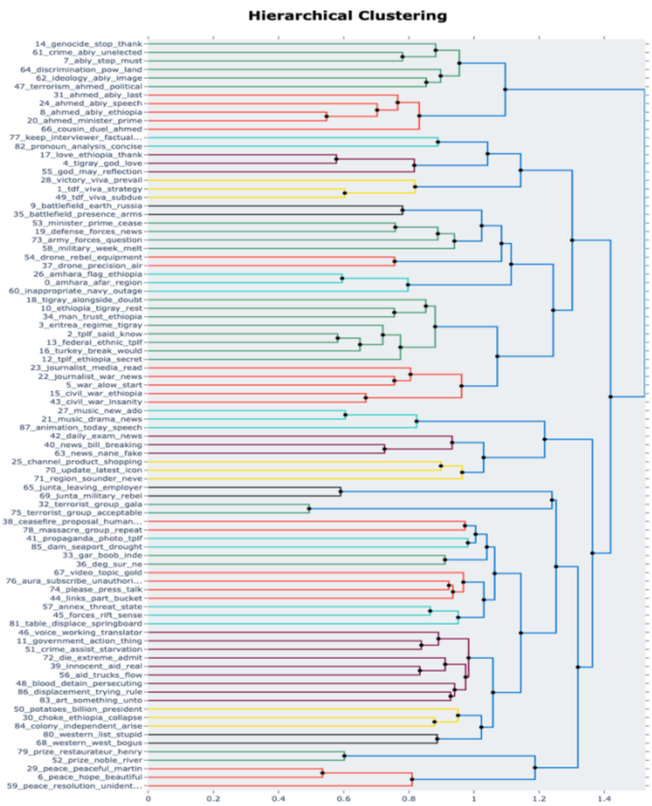
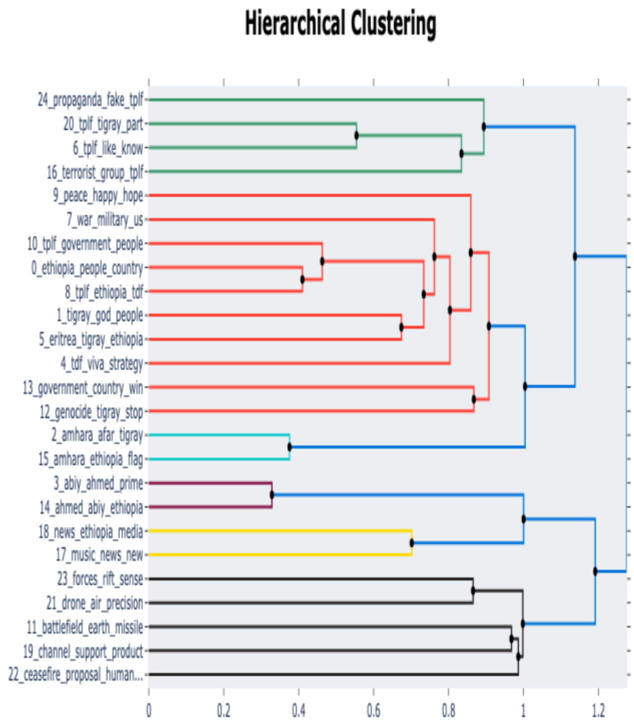
Highest C_v Coherence (Model 1)	Highest *NPMI* Coherence (Model 2)
	

Figure 11. Hierarchical Clustering

Based on the intertopic distance maps (Figure 10), the topic clusters in Model 2 are often very close and overlapping. In contrast, although Model 1 also has some overlapping clusters, it shows more distinctly separated clusters compared to Model 2. Therefore, from the perspective of the distances between clusters, it can be concluded that Model 1 demonstrates better performance.

Likewise, Model 1 seems to perform better in hierarchical clustering. Model 2 grouped Topic 9 ('peace_happy_hope') with Topic 7 ('war_military_us') and Topic 22 ('ceasefire_proposal_human') with Topic 21 ('drone_air_precision'), indicating bad performance. Moreover, it failed to identify a distinct topic related to 'Peace'. Conversely, Model 1 effectively categorised topics, such as Topics 7, 14, 57, 61, 62, and 64, which criticise Abiy Ahmed for initiating the war, into one category. It also distinguished topics like Topics 8, 20, 24, 31, and 66 related to Abiy Ahmed's speeches and general information. Furthermore, it identified topics related to 'Peace', such as Topics 6, 29, and 59, which would belong to a specific topic category.

As a result, the Highest C_v Coherence model demonstrated a superior C_v coherence score and clear topic differentiation in the visual analysis. Hence, combination A was selected as the optimal hyperparameter setting for BERTopic.

4-5-3. Model Construction

Initially, the topic model yielded 763 topics from 44,757 text data points. According to Figure 12, the model with 763 topics contained a considerable number of overlapping topics. At least two topics overlap, with numerous similarities identified in the representative documents and relevant keywords. Therefore, the topics have been consolidated and reduced by around half, and the model was re-visualised.

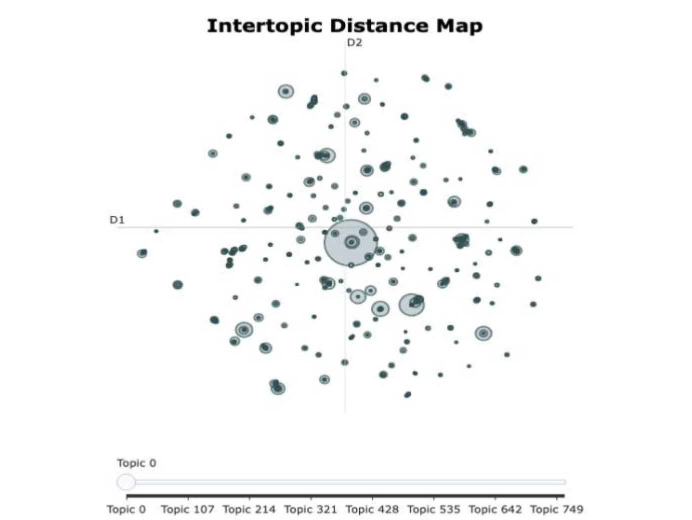
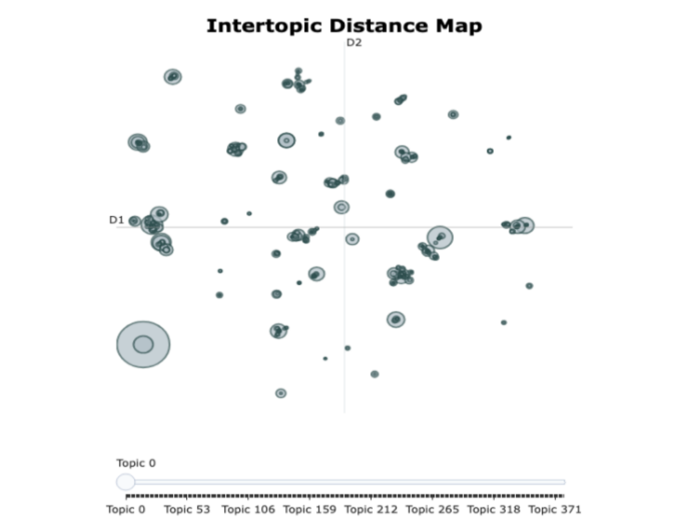
Model with 763 topics	Model with 380 topics
	

Figure 12. Intertopic Distance Map 1 – Final Model

Nevertheless, there were still many overlapping topics within 380 topics. Hence, the topics were further consolidated, reducing the number by half again, and a hierarchical analysis was conducted on the topics.

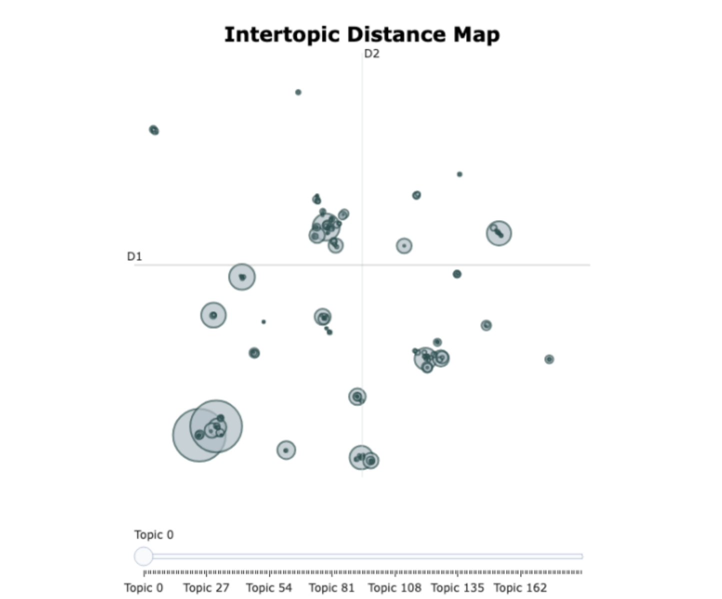
Model with 190 topics


Figure 13. Intertopic Distance Map 2 – Final Model

After consolidating the topics into 190, significant improvement in distances between topics was observed. Thus, it has been decided not to proceed with further consolidation to minimise the risk of over-reducing.

As a result, the initial 763 topics were merged into 190 topics. The information on the 190 topics with the total number is described in Appendix 2. Subsequently, the researcher has categorised each topic as 'Tigray War - Peace', 'International Politics', 'Humanitarian Crisis', and so on, for analysis purposes based on the representative keywords and documents.

4-5-4. Topic Labelling

Labels of the Tigray War Topics					
No	**Label**	**Group**	**No**	**Label**	**Group**
1	Abiy Ahmed	4, 19, 32, 86, 126, 188	2	Anti-Abiy Ahmed	10, 22, 68, 96, 103, 132
3	Ethiopian Government	8	4	Ethiopian News and Media	0, 23, 38, 39, 41, 46, 49, 50, 60, 65, 69, 76, 79, 81, 89, 102, 105, 129, 141, 138, 159, 151, 149, 146, 145, 178, 177, 176
5	Humanitarian Crisis	21, 33, 51, 56, 73, 106, 156, 155, 144, 185, 184, 175, 174	6	International Politics	9, 26, 34, 53, 75, 128, 158, 157
7	Politics	137, 152, 153, 169	8	Religion	72, 101
9	Tigray War: Genocide	16	10	Tigray War: Anti-TPLF	11, 15, 24, 55, 59, 61, 62, 63, 67, 80, 97, 98, 112, 133, 139, 131, 143, 163, 162, 183, 187
11	Tigray War: Battle and Conflict	1, 14, 17, 18, 28, 42, 48, 57, 74, 84, 91, 90, 88, 100, 104, 118, 120, 121, 119, 130, 127, 140, 135, 136, 167	**12**	**Tigray War: Ceasefire**	6, 40, 43, 111, 113, 115, 165
13	Tigray War: Eritrea	3, 181	**14**	**Tigray War: Peace**	12, 44, 92, 94, 116, 123, 172, 170, 186, 180
15	Tigray War: Pro-TPLF	2, 5, 30, 64, 77, 78, 93, 107, 142, 171, 168, 161, 179	16	Tigray War: Propaganda	20, 37, 58
17	Tigray Region	25, 31, 36, 45, 70, 82, 109, 110, 114, 122, 125, 134, 147, 182	18	Tigray War: TPLF	7, 13, 27, 52, 71, 83, 85, 87, 108, 160, 150, 148, 164, 166
19	np.nan	-1, 29, 35, 47, 54, 66, 95, 99, 117, 124, 154, 173			

Table 8. Labels of the Tigray War Topics

Firstly, topics related to Abiy Ahmed, the Prime Minister of Ethiopia, such as speeches and general information, were consolidated under Label 1. Meanwhile, topics regarding criticism and negative content about Abiy Ahmed were designated as Label 2. Additionally, topics concerning the Ethiopian Government were assigned Label 3, those related to Ethiopian politics were assigned Label 7, and topics concerning religions were designated Label 8.

Furthermore, topics directly involved in the Tigray War were assigned different labels according to their specific content. Topics about the humanitarian crisis resulting from the war were allocated to Label 5 while opposing views on the TPLF were integrated into Labels 10 and 15 respectively. The topic concerning general discourses about the TPLF was labelled as 18. Moreover, topics on battles and conflicts from the war were assigned to Label 11, and those concerning the propaganda issues within the war involved in the war were allocated to Label 16. Topics generally advocating for peace in Ethiopia were given Label 14, while those discussing ceasefire agreements and efforts to promote a ceasefire were assigned Label 12. This research considers discourse from Labels 12 and 14 as indicating a need for the cessation of the war or general peace in Ethiopia. Topics referring to the Tigray War as genocide were allocated Label 9.

In addition, topics about international politics were assigned Label 6. General discourses about the Tigray region were designated as Label 17, while discourses about Eritrea were assigned Label 13.

In the BERTopic model, discourses assigned to Topic -1 are considered outliers and are not associated with any specific topic (Grootendorst, 2021). This model categorised a total of 22,052 discourses into -1 topics. Including -1 topic, topics that are difficult to analyse or whose meanings are not clearly understood such as '117_jeff_thank_truth_deliverer' topic were given Label 19 and excluded from subsequent analyses. These labelling details are illustrated in Table 8.

To assess the labels, around 10 percent of the dataset, 4,440 data points, have been randomly extracted. Through these data points, this research first assigned topics using the topic model and labelled them. Based on these results, the distribution of labels from five iterations of topic modelling on the same data has been compared to evaluate the quality of the labels. The evaluation matrix is detailed in Equation 1.

Method	Equation
Accuracy	$\frac{TP + TN}{N}$
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
F1 Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$
Where N stands for 'Total number of examples', TP represents 'True Positive', TN denotes 'True Negative', FP refers to 'False Positive' and FN means 'False Negative'.	

Equation 1. Evaluation Matrix

No	Accuracy	Precision	Recall	Average F1
1	0.878	0.764	0.753	0.758
2	0.876	0.756	0.750	0.752
3	0.868	0.739	0.749	0.743
4	0.875	0.754	0.740	0.746
5	0.880	0.759	0.747	0.752
Mean	**0.875**	**0.754**	**0.748**	**0.750**

Table 9. Performance of Topic Labelling

The labelling results yielded an overall accuracy of 87.5 percent, a precision of 75.4 percent, a recall of 74.8 percent, and an average F1 score of 75 percent. Although the precision, recall, and F1 score are around 75 percent, the accuracy is notably high at 87.5 percent. Therefore, this study concluded that meaningful analysis can be conducted using these labels.

4-6. Sentiment Analysis Model

4-6-1. Sentiment Analysis

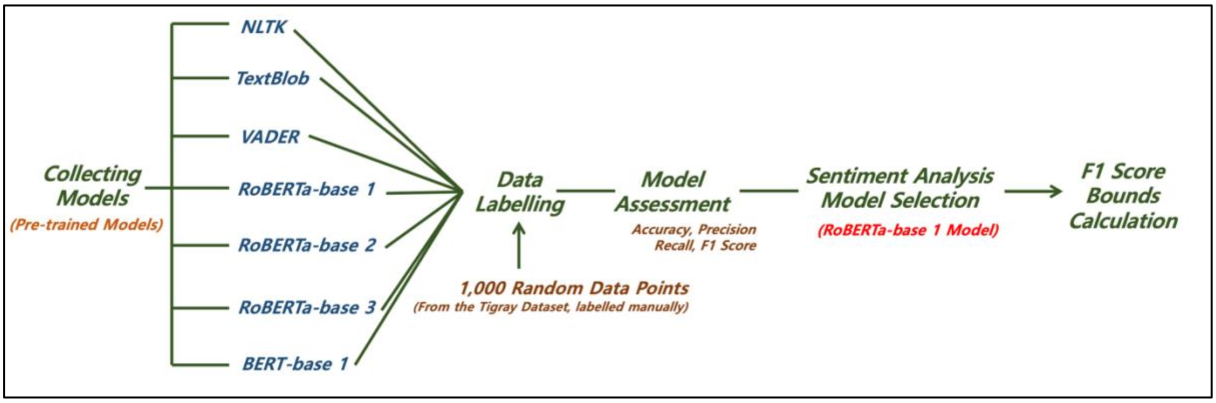


Figure 14. Sentiment Analysis Process

Zagibalov (2010) explained that sentiment analysis in NLP aims to automatically classify the sentiment expressed in a document or text. Unlike the dominant BERTopic in topic modelling, there is no clear frontrunner in sentiment analysis models. In addition, performance can vary significantly depending on the data applied to the model (Sharma et al., 2024). Therefore, this research opted to compare the performance of several sentiment analysis models to find the most suitable one for the sentiment analysis of the Tigray War dataset.

4-6-2. Model Comparison

Recently, several sentiment analysis models have been invented. This research collected seven pre-trained sentiment analysis models such as NLTK, TextBlob, and models from Hugging Face to decide on one model for this project. Hugging Face, an open-source machine-learning tool originally designed for NLP projects has increasingly found application not only in commercial and business purposes but also in academic research (Nicolas Azevedo, 2024).

To assess the models, this project used a random selection of 1,000 text samples including more than ten words from the Tigray War dataset in order to capture sentiments appropriately. Due to the unavailability of an officially labelled dataset for sentiment analysis of armed conflict discourses, the research determined to manually label the dataset. These text samples were assigned sentiments of positive, neutral, and negative. The data points were labelled as 'positive' when they expressed the need for peace of cessation or used primarily positive language. Discourses that were informative but difficult to gauge sentiment were labelled as 'neutral'. On the other hand, discourses that criticised the war or used mainly negative words were labelled as 'negative'.

Model	Feature
NLTK	- Build-In Sentiment Analyser from NLTK
TextBlob	- Sentiment Analyser from TextBlob Library
VADER	- VADER (Valence Aware Dictionary and sEntiment Reasoner) is a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media (Hutto, C.J. & Gilbert, E.E., 2014) - The model was constructed using default hyperparameters
RoBERTa-base 1	- Trained with SemEval 2017 corpus Dataset (Pérez et al., 2021) - The base model is BERTweet, a RoBERTa model trained on English tweets - Pretrained model from Hugging Face - Link: https://huggingface.co/finiteautomata/bertweet-base-sentiment-analysis
RoBERTa-base 2	- Trained on around 124 million tweets from January 2018 to December 2021 - Fintuned for sentiment analysis with the TweetEval benchmark - Pretrained model from Hugging Face - Link: https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest
RoBERTa-base 3	- Trained on around 198 million tweets and finetuned for sentiment analysis - Multilingual XLM-roBERTa-base model (Barbieri, Anke and Camacho-Collados, 2022) - Pretrained model from Hugging Face - Link: https://huggingface.co/cardiffnlp/twitter-xlm-roberta-base-sentiment
BERT-base 1	- Trained in English language using a masked language modelling (MLM) objective (Devlin et al., 2019) - The base model is BERT, and this model is uncased - Pretrained model from Hugging Face - Link: https://huggingface.co/sbcBI/sentiment_analysis_model

Table 10. Sentiment Analysis Models

The models specified in Table 10 were evaluated based on accuracy, precision, recall, and F1 scores.

Rank	Model	Accuracy	Precision	Recall	Average F1
1	RoBERTa-base 1	0.705	0.781	0.668	0.669
2	RoBERTa-base 2	0.657	0.711	0.618	0.622
3	RoBERTa-base 3	0.654	0.703	0.613	0.617
4	TextBlob	0.563	0.601	0.602	0.560
5	BERT Based 1	0.525	0.605	0.499	0.464
6	NLTK	0.490	0.564	0.509	0.444
7	Vader	0.490	0.564	0.509	0.444

Table 11. Sentiment Analysis Models Performance

As outlined in Table 11, the RoBERTa-base 1 model demonstrates the highest scores among these seven models across the evaluation metrics: accuracy, precision, recall, and average F1 score. Specifically, this model achieves approximately 70 percent accuracy. Although this accuracy does not reach the 80 to 90 percent range, this study considers that this level of accuracy can produce meaningful and valuable results. The confusion matrix of this model is detailed in Figure 15.

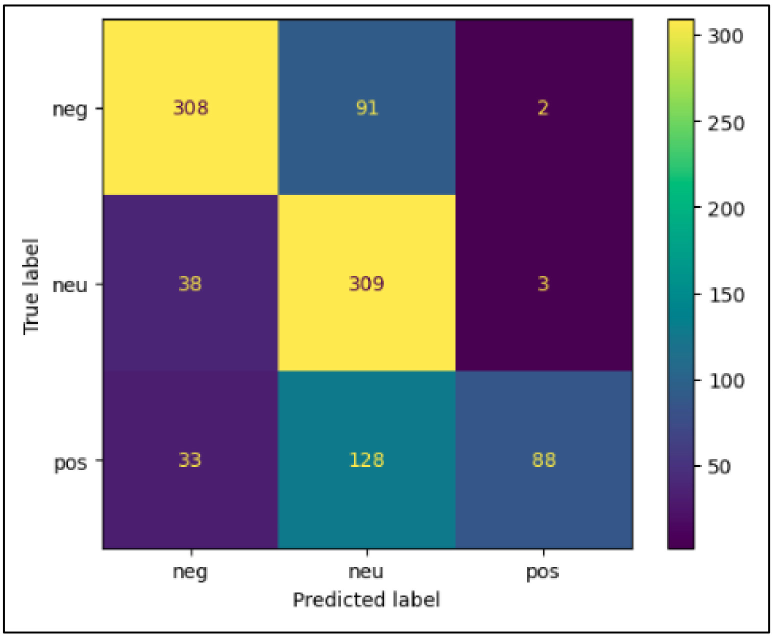


Figure 15. Confusion Matrix

However, the F1 scores for this model are 0.7897 for negative sentiment, 0.7038 for neutral sentiment, and 0.5146 for positive sentiment. On average, the F1 score is 0.669. Nonetheless, the relatively low F1 score for positive sentiment indicates that accurately determining the sentiment distribution is challenging.

Hence, this study has determined to analyse the sentiment distribution by calculating the upper and lower bounds of the F1 scores and examining the sentiment distribution within these bounds. Although fine-tuning the model could be a potential solution, conducting the analysis using the bounds is deemed more viable due to time constraints.

4-6-3. F1 Score Bounds Calculation

The bounds of social discourse sentiment can be estimated through the confidence intervals for the F1 Score. This can be achieved through bootstrapping, a method that employs random sampling with replacement. It is a type of resampling technique that provides estimates of accuracy measures such as bias, variance, confidence intervals, and prediction error for sample statistics (Efron and Tibshirani, 1998). By using random sampling methods, bootstrapping enables the estimation of the sampling distribution for nearly any statistic (Wehrens, Putter and Buydens, 2000).

In this study, bootstrapping has been performed 1,000 times using the manually labelled data. For each sample, the F1 score was computed and then estimate the distribution of these scores. The distribution was calculated based on a 95 percent confidence interval. The upper and lower bounds of the F1 Score for each sentiment are as follows:

Sentiment	F1 Score	Upper Bound	Lower Bound
Negative	0.7897	0.7578	0.8203
Neutral	0.7038	0.6681	0.7358
Positive	0.5146	0.4499	0.5815

Table 12. F1 Score Bounds

Based on the upper and lower bounds of the F1 score, this research approximates the results of sentiment analysis for social media discourse in each phase to address research questions 1-2 and 2-2, using the RoBERTa-base 1 model.

5. Results

5-1. The topics occurred during the Tigray War

Overall Trend of the Tigray War Topics							
No	Label	Count	Ratio	No	Label	Count	Ratio
1	Tigray War: Battle and Conflict	5,429	24.31%	2	Ethiopian News and Media	4,981	22.30%
3	Tigray War: Pro-TPLF	2,224	9.95%	4	Tigray War: TPLF	1,359	6.08%
5	Abiy Ahmed	1,220	5.46%	6	Tigray War: Anti-TPLF	1,190	5.32%
7	Tigray War: Eritrea	1,004	4.49%	8	**Tigray War: Ceasefire**	999	4.47%
9	International Politics	759	3.39%	10	Anti-Abiy Ahmed	617	2.76%
11	**Tigray War: Peace**	527	2.35%	12	Ethiopian Government	484	2.16%
13	Tigray Region	462	2.06%	14	Tigray War: Humanitarian Crisis	397	1.77%
15	Tigray War: Genocide	330	1.47%	16	Tigray War: Propaganda	284	1.27%
17	Religion	38	0.17%	18	Politics	28	0.12%

Table 13. Overall Trend of the Tigray War Topics

Overall, on social media discourses about the Tigray War on YouTube, topics about battles and conflicts from the war make up the largest portion at 24.31 percent. Topics related to 'Ethiopian News and Media' follow closely with 22.30 percent. These two topic labels are the most common throughout all stages from Phase 1 to Phase 6.

Topics under the label of 'Battle and Conflict', during Phase 1 to Phase 6, mostly focus on the events and situations related to various battles or conflicts that occurred during these periods. For instance, discussions often revolve around scenarios such as “tigray wish withdraw battlefield leave amhara forces fight battlefield analysis take less week tdf occupy amhara god bless” and “first tplf attack northern command army right track used transport forces amhara afar region.”

On the other hand, throughout all phases, general issues concerning Ethiopia and comments on media, such as “video excellent explanation ethiopia country world” and “upcoming election pervious five would say best election ethiopia among political electoral commission ability,” are consistently present. These topics, which are less related to conflict, appear across the entire period. From Phases 1 to 6, these two labels, 'Battle and Conflict' and 'Ethiopian News and Media' exhibit similar trends and content, therefore they are no longer discussed in further analysis.

The proportion of topics supporting the TPLF, a major party in the Tigray War, is 9.95 percent, considerably higher than the proportion of topics criticising the TPLF, which is 5.32 percent. Topics that are neutral or merely mention the TPLF make up 6.08 percent. Meanwhile, topics concerning Abiy Ahmed, a key figure directly associated with the Tigray War and opposed to the TPLF, appear in 5.46 percent of the discussions. At the same time, negative topics about Abiy Ahmed account for 2.76 percent.

In comparison, topics related to the Ethiopian Government are much less prevalent, comprising only 2.16 percent. Abiy Ahmed is often referenced more frequently than the Ethiopian government itself, given his strong representation as a prominent politician in Ethiopia.

Following this, topics providing information and general status about the Tigray region make up 2.06 percent. Despite being a neighbouring country, topics about Eritrea, indirectly involved in the Tigray War, appear in 4.49 percent of the discussions.

Discussions on ceasefire negotiations and peace, hoping for a cessation of the Tigray War, collectively accounted for 6.83 percent, with ceasefire negotiations at 4.47 percent and peace-related topics at 2.35 percent. Besides, topics criticising the Tigray War as genocide make up 1.47 percent.

Moreover, topics demonstrating the humanitarian crisis caused by the Tigray War appeared in about 1.77 percent of the discourse. In addition, the proportion of topics discussed propaganda from the war stands at 1.27 percent.

Furthermore, topics about international politics and national politics appear occasionally, accounting for 3.39 percent and 0.12 percent respectively. On the other hand, categories such as 'Religion,' which represents basic information and mentions any religion within Ethiopia, and topics not closely related to the Tigray War, are excluded from the analysis.

1) Phase 1

Phase 1 - Trend of the Tigray War Topics							
No	**Label**	**Count**	**Ratio**	**No**	**Label**	**Count**	**Ratio**
1	Ethiopian News and Media	329	20.66%	2	Tigray War: Battle and Conflict	315	19.78%
3	Abiy Ahmed	179	11.24%	4	Tigray War: Anti-TPLF	151	9.48%
5	**Tigray War: Ceasefire**	118	7.41%	6	Tigray War: TPLF	100	6.28%
7	Tigray War: Pro-TPLF	62	3.89%	8	International Politics	61	3.83%
9	Tigray War: Eritrea	57	3.58%	10	Anti-Abiy Ahmed	46	2.88%
11	Ethiopian Government	43	2.70%	12	Tigray War: Genocide	41	2.57%
13	**Tigray War: Peace**	32	2.01%	14	Tigray Region	21	1.31%
15	Tigray War: Propaganda	16	1.00%	16	Tigray War: Humanitarian Crisis	13	0.81%
17	Politics	5	0.31%	18	Religion	3	0.18%

Table 14. Phase 1 - Trend of the Tigray War Topics

According to Table 14, during the initial stage of the conflict, topics related to Abiy Ahmed, a key figure directly associated with the war, account for 11.24 percent, while topics associated with the TPLF make up 6.28 percent. Regarding Abiy Ahmed, the majority of discussions express belief and support for him, as seen in statements like "abiy smart leader world believe." In the case of the TPLF, the discourse generally centres around general information, such as "truly uninformed analysis tplf army stand fight even short."

Meanwhile, criticisms of Abiy Ahmed make up 2.88 percent, while criticisms of the TPLF constitute 9.48 percent. Additionally, discussions supporting the TPLF account for 3.89 percent. For instance, topics critical of Abiy Ahmed have emerged, such as "shame colonel abiy coz peace price going war." In contrast, topics supporting the TPLF often focus on wishing for victory in the conflict, as seen in "tigray people win real fight." However, topics opposing the TPLF predominantly include statements like "tplf must stop war" and "terrorist group tplf must take full responsibility mess", which call for the TPLF to cease fighting and refer to them as a terrorist group.

Topics concerning peace, such as 'Ceasefire' and 'Peace,' accounted for 7.41 percent and 2.01 percent respectively, totalling 9.45 percent. Moreover, topics referring to the Tigray War as a genocide perpetrated by the Ethiopian government comprised 2.57 percent. In particular, discourses expressing a desire for peace and prayers for its attainment frequently appear in 'Peace' related topics, such as "may god help bring peace please stop save one going." Within the 'Ceasefire' label, the focus is more direct on cessation of the war, with calls to "stop war stop killing innocent people must go."

Furthermore, topics about Eritrea appear at a rate of 3.58 percent, and topics related to international politics occur at 3.83 percent. In the case of international politics, the topics mainly concern Turkey's support for the Ethiopian government, as seen in "turkey give open support ethiopia." Regarding Eritrea, the discussions focus on its involvement in the conflict, as illustrated by "civil war involve eritrea."

Within Phase 1 of the Tigray War, the proportion of topics discussing the Ethiopian government is 2.70 percent, while those about the information and status of the Tigray region make up 1.31 percent. Meanwhile, topics such as 'Tigray War: Propaganda, Humanitarian Crisis', 'Politics', and 'Religion' are almost non-existent in Phase 1, each accounting for less than 1 percent.

2) Phase 2

Phase 2 - Trend of the Tigray War Topics							
No	**Label**	**Count**	**Ratio**	**No**	**Label**	**Count**	**Ratio**
1	Tigray War: Battle and Conflict	3,698	30.79%	2	Ethiopian News and Media	2,513	20.92%
3	Tigray War: Pro-TPLF	1,230	10.24%	4	Tigray War: TPLF	731	6.08%
5	Tigray War: Anti-TPLF	594	4.94%	6	Abiy Ahmed	587	4.88%
7	**Tigray War: Ceasefire**	398	3.31%	8	International Politics	387	3.22%
9	Tigray War: Eritrea	374	3.11%	10	Anti-Abiy Ahmed	277	2.30%
11	Tigray Region	238	1.98%	12	Tigray War: Humanitarian Crisis	216	1.79%
13	Ethiopian Government	202	1.68%	14	**Tigray War: Peace**	196	1.63%
15	Tigray War: Propaganda	173	1.44%	16	Tigray War: Genocide	169	1.40%
17	Religion	21	0.17%	18	Politics	6	0.05%

Table 15. Phase 2 - Trend of the Tigray War Topics

As described in Table 15, as the war escalated during Phase 2, the proportion of topics supporting the TPLF increased to 10.24 percent. In contrast, the proportion of topics critical of the TPLF drops sharply to 4.94 percent. General topics about the TPLF remain stable at 6.08 percent, similar to the previous phase. In the label supporting the TPLF, discourses wishing for continued victory for Tigray and the TDF in the conflict appear, such as "great keep victory people tigray" and "tdf amazing keep going." However, in topics opposing the TPLF, negative portrayals of the group are predominant, with statements like "criminal junta attack federal military base fall sleep mixed person wright" casting the TPLF as criminals, and "defense forces terrorist junta Tigray" accusing them of being terrorists.

On the other hand, topics concerning Abiy Ahmed decreased to 4.88 percent but still represent a significant portion of the discourse. Moreover, topics critical of Abiy Ahmed slightly declined to 2.30 percent. This suggests that in Phase 2, discussions are more focused on the TPLF rather than on Abiy Ahmed. Specifically, discourses about Abiy Ahmed often describe him as an excellent leader, as seen in comments like "abiy stunning leader love" and "always support ahmed real leader condemn tplf." In contrast, topics opposing him describe Abiy Ahmed as a criminal, with comments such as "abiy ahmed war criminal must arrest." Additionally, topics related to Eritrea, which began to play a role in the Tigray War, appear relatively frequently at 3.11 percent.

Topics associated with 'Ceasefire' and 'Peace' account for 3.31 percent and 1.63 percent respectively, making up 4.94 percent. As the war extended into a prolonged conflict, the frequency of peace-related topics declined. Within peace-related topics, there are many discussions advocating for peace and supporting innocent Tigrayan people, such as "peace love stop killing innocent tigray love tigray." Moreover, in the 'Ceasefire' label, topics include calls for the TPLF to negotiate with the government, as seen in "tigray part must negotiate defy government ceasefire", as well as arguments for a ceasefire to achieve peace.

The proportion of topics mentioning the humanitarian crisis caused by the Tigray War increased to 1.79 percent, appearing more frequently than in the previous phase. As seen in comments such as "dying lack medicine food medicine days famine inside tigray right" and "food medicine needs people tigray", the prolonged conflict has led to a humanitarian crisis, with shortages of medicine and food.

Besides, topics related to propaganda from the war constitute 1.44 percent. Meanwhile, the proportion of topics criticising the Tigray War as genocide decreased to 1.40 percent.

The proportion of topics about the Tigray region is 1.98 percent, while topics concerning the Ethiopian government similarly account for 1.68 percent.

3) Phase 3

Phase 3 - Trend of the Tigray War Topics							
No	**Label**	**Count**	**Ratio**	**No**	**Label**	**Count**	**Ratio**
1	Tigray War: Battle and Conflict	130	30.87%	2	Ethiopian News and Media	103	24.46%
3	Tigray War: Pro-TPLF	28	6.65%	4	International Politics	27	6.41%
5	Ethiopian Government	17	4.03%	6	Abiy Ahmed	16	3.80%
7	Tigray War: Eritrea	15	3.56%	8	Tigray War: Anti-TPLF	15	3.56%
9	Tigray War: TPLF	13	3.08%	10	**Tigray War: Ceasefire**	11	2.61%
11	Tigray Region	10	2.37%	12	Anti-Abiy Ahmed	9	2.13%
13	**Tigray War: Peace**	9	2.13%	14	Tigray War: Humanitarian Crisis	7	1.66%
15	Tigray War: Genocide	5	1.18%	16	Tigray War: Propaganda	3	0.71%
17	Religion	2	0.47%	18	Politics	1	0.23%

Table 16. Phase 3 - Trend of the Tigray War Topics

In the month preceding the first cessation, during Phase 3, topics related to the Ethiopian government and Abiy Ahmed, key figures in the first peace agreement, account for 4.03 percent and 3.80 percent respectively. Meanwhile, the proportion of topics supporting the TPLF, as well as those critical of the TPLF and general topics about the TPLF, all decreased, representing 6.65 percent, 3.56 percent, and 3.08 percent. The comments from topics opposing the TPLF often include strong criticisms, such as “entire horn stable n peaceful tplf terrorist terrorist disarm send war”, which argues that peace can only be achieved if the TPLF is disarmed. Conversely, within the ‘Pro-TPLF’ label, discourses frequently praise the TDF’s positive performance, as seen in comments like “tdf professional army well comes usual.”

Moreover, topics concerning the Tigray region make up 2.37 percent, while those linked to Eritrea account for 3.56 percent. Topics about peace, specifically labelled as 'Peace' and 'Ceasefire,' collectively constitute around 4.75 percent, with 'Peace' at 2.13 percent and 'Ceasefire' at 2.61 percent. Within peace-related topics, there are numerous discourses advocating for a peaceful resolution to the conflict, such as “god peaceful resolution tigray war better chance.” It is important to note that the proportion of peace-related topics declined, compared to Phase 2.

Besides, topics related to the humanitarian crisis resulting from the war amount to 1.66 percent. At the same time, topics criticising the war as genocide account for 1.18 percent, which is almost the same as in Phase 2. Additionally, topics talked about international politics also appear more frequently, accounting for 6.41 percent. The trend of topics is illustrated in Table 16.

4) Phase 4

Phase 4 - Trend of the Tigray War Topics							
No	Label	Count	Ratio	No	Label	Count	Ratio
1	Tigray War: Battle and Conflict	499	29.56%	2	Ethiopian News and Media	419	24.82%
3	Tigray War: Pro-TPLF	117	6.93%	4	Tigray War: Eritrea	86	5.09%
5	Tigray War: TPLF	83	4.91%	6	Tigray War: Anti-TPLF	60	3.55%
7	Abiy Ahmed	56	3.31%	8	**Tigray War: Ceasefire**	53	3.13%
9	**Tigray War: Peace**	47	2.78%	10	International Politics	46	2.72%
11	Tigray War: Humanitarian Crisis	45	2.66%	12	Tigray War: Genocide	40	2.36%
13	Ethiopian Government	38	2.25%	14	Tigray Region	37	2.19%
15	Anti-Abiy Ahmed	36	2.13%	16	Tigray War: Propaganda	21	1.24%
17	Religion	3	0.17%	18	Politics	2	0.11%

Table 17. Phase 4 - Trend of the Tigray War Topics

Within Phase 4, the first cessation of the Tigray War, topics related to battles and conflicts that arose from the war still account for a significant portion at 29.56 percent. In alignment with the peace agreement, the frequency of topics about the main stakeholders, the TPLF, increased to 4.91 percent. Topics supporting the TPLF rose to 6.93 percent, while those opposing the TPLF remained at 3.55 percent, similar to the previous phase. Within topics supporting the TPLF, numerous comments express hope for Tigray's independence in addition to achieving peace through a ceasefire. Examples include “tigray gain access huge amount become independent” and “tigray independent sometime future.”

However, topics concerning Abiy Ahmed and the Ethiopian government, other key stakeholders, decreased compared to Phase 3, constituting 3.31 percent and 2.25 percent respectively. In comments about Abiy Ahmed, there is a notable amount of praise for his role in creating the first ceasefire, with remarks such as “abiy ahmed change job good actor.”

On the other hand, discussions on 'Peace' and 'Ceasefire' represented 3.13 percent and 2.78 percent of the topics respectively, totalling 5.92 percent, which is a somewhat increase from Phase 3. In peace-related topics, similar to previous discussions, comments expressing a wish for peace are predominant, such as “peace tigray peace anywhere” and “praying peace care worried would hope country.” Within the ‘Ceasefire’ label, most comments praise the first ceasefire, but there are also concerns about the potential for the ceasefire to break down, as reflected in comments like “also suspected war might starting derail.”

In addition, topics characterising the Tigray War as genocide and those addressing the humanitarian crisis resulting from the war consistently appear, with their proportions slightly increasing to 2.36 percent and 2.66 percent respectively. Topics related to international politics decreased by more than half compared to the previous phase, accounting for 2.72 percent. The trend of topics within Phase 4 is demonstrated in Table 17.

5) Phase 5

Phase 5 - Trend of the Tigray War Topics							
No	**Label**	**Count**	**Ratio**	**No**	**Label**	**Count**	**Ratio**
1	Tigray War: Battle and Conflict	328	22.14%	2	Ethiopian News and Media	311	20.99%
3	Tigray War: Pro-TPLF	201	13.57%	4	Tigray War: Anti-TPLF	115	7.76%
5	Tigray War: TPLF	99	6.68%	6	**Tigray War: Ceasefire**	84	5.67%
7	Tigray War: Eritrea	69	4.65%	8	Anti-Abiy Ahmed	37	2.49%
9	Ethiopian Government	36	2.43%	10	Abiy Ahmed	34	2.29%
11	International Politics	33	2.22%	12	**Tigray War: Peace**	33	2.22%
13	Tigray Region	33	2.22%	14	Tigray War: Humanitarian Crisis	27	1.82%
15	Tigray War: Propaganda	21	1.41%	16	Tigray War: Genocide	16	1.08%
17	Religion	3	0.20%	18	Politics	1	0.06%

Table 18. Phase 5 - Trend of the Tigray War Topics

During Phase 5, following the failure of the first ceasefire and the resumption of the Tigray War, topics supporting the TPLF significantly increased to 13.57 percent. Simultaneously, discourses opposing the TPLF also more than doubled, rising to 7.76 percent. This indicates a heightened polarisation of opinions regarding the TPLF in the wake of the breakdown of the peace agreement. General topics about the TPLF are present at 6.68 percent. In discourses critical of the TPLF, there are intense criticisms such as “going understand tplf never get meaning peace blood stream forget never happen”, which accuse the TPLF of not understanding the meaning of peace, and “long live working eliminate terror group tplf”, which labels the TPLF as a terrorist group and supports for its elimination. Within the ‘Pro-TPLF’ label, comments are frequently in line with the TDF’s actions and supporting them as the conflict continues, such as “tdf fighting protect people empire.”

As the war resumed, the frequency of topics about Abiy Ahmed decreased to 2.29 percent compared to Phase 4. However, topics critical of Abiy Ahmed slightly increased, reaching 2.49 percent. Meanwhile, topics related to Eritrea, which has been consistently involved in the conflict, appear steadily at 4.65 percent. Within the discourse criticising Abiy Ahmed, there are numerous criticisms suggesting that he does not like peace in Ethiopia, as seen in discourse like “abiy must go peace works, never like peaceful”, which emerged following the breakdown of the ceasefire.

Importantly, peace-related topics appeared in about 7.90 percent of the discourses, with 'Peace' labelled topics making up 2.22 percent and 'Ceasefire' labelled topics comprising 5.67 percent. Among peace-related topics, there are discussions urging the search for peaceful solutions to the conflict, as mentioned by “please find peaceful solve problem, want see beautiful continent.” Meanwhile, within the ‘Ceasefire’ label, most comments focus on the necessity of cessation of the war, with statements like "solve war need invade stop war quick" predominating.

Topics demonstrating the Tigray War as genocide and those mentioning the humanitarian crisis caused by the war continued to be discussed, representing 1.82 percent and 1.08 percent of the discourse, respectively.

6) Phase 6

Phase 6 - Trend of the Tigray War Topics							
No	**Label**	**Count**	**Ratio**	**No**	**Label**	**Count**	**Ratio**
1	Ethiopian News and Media	145	22.72%	2	Tigray War: Battle and Conflict	108	16.92%
3	**Tigray War: Ceasefire**	68	10.65%	4	Tigray War: Anti-TPLF	54	8.46%
5	Tigray War: TPLF	53	8.30%	6	Tigray War: Pro-TPLF	46	7.21%
7	**Tigray War: Peace**	32	5.01%	8	Tigray War: Eritrea	31	4.85%
9	International Politics	18	2.82%	10	Abiy Ahmed	17	2.66%
11	Tigray Region	16	2.50%	12	Tigray War: Genocide	13	2.03%
13	Anti-Abiy Ahmed	13	2.03%	14	Tigray War: Humanitarian Crisis	10	1.56%
15	Ethiopian Government	9	1.41%	16	Tigray War: Propaganda	4	0.62%
17	Religion	1	0.15%	18	Politics	0	0.00%

Table 19. Phase 6 - Trend of the Tigray War Topics

In the final phase of the war, with negotiations for a ceasefire through peace talks underway, peace-related topics account for 15.67 percent of the overall discourses. This can be interpreted as a result of the ongoing peace talks, indicating a significant public focus on this agenda. Within peace-related topics, the majority of discourses express a desire for the long-lasting continuation of peace, as evidenced by comments like "hopefully peace remain long time", or seek a complete end to the conflict, as in "peace coming back, battlefield hope leave us alone."

As the peace talks were mediated by the AU, the proportion of topics mentioning Abiy Ahmed and the TPLF increased to 2.66 percent and 8.30 percent, respectively. In contrast, topics associated with the Ethiopian Government constitute only 1.41 percent of the discourses in Phase 6.

Furthermore, topics opposing the TPLF rose to 8.46 percent, while topics supporting the TPLF decreased by over 6 percent to 7.21 percent. Topics critical of Abiy Ahmed also capture a slight decrease, accounting for 2.03 percent. During the peace talks, strong criticism of the TPLF still exists, with comments such as "put tplf terrorism criminal group" and "investigate tplf continue terrorist." Conversely, within the discourse supporting the TPLF, there are also some comments, such as "tplf keep fighting for self-determination", advocating for the continuation of the fight for autonomy.

Moreover, topics referring to the Tigray War as genocide and those discussing the humanitarian crisis caused by the war continued to appear, at 2.03 percent and 1.56 percent, respectively. In topics under the 'Humanitarian Crisis' label, discourses often ask for immediate humanitarian actions in Tigray, as seen in comments like "tigray needs rapid humanitarian intervention." The trend of topics is described in Table 19.

5-2. The Sentiments of Social Media Discourses on the Tigray War

	Lower Bound	Median	Upper Bound
Negative	13,991 (31.3%)	14,580 (32.6%)	15,144 (33.8%)
Neutral	22,755 (50.8%)	23,974 (53.6%)	25.061 (56.0%)
Positive	5,423 (12.1%)	6,203 (13.9%)	7,009 (15.7%)

Table 20. Overall Sentiment Distribution (Tigray War)

Through the sentiment analysis model with F1 score ranges in Table 12, the distribution of each sentiment in social media discussions about the Tigray War is estimated including their upper and lower bounds.

Among 44,757 social media discourses on the Tigray War, 12.1 percent to 15.7 percent of discourses exhibit positive sentiment, while 31.3 percent to 33.8 percent show negative sentiment. This indicates that negative discourses dominated during the armed conflict. Meanwhile, neutral sentiment discourses account for 50.8 percent to 56.0 percent of the total discourses. It is essential to note that neutral sentiments are the most prevalent across all phases of the war, accounting for between 49 and 60 percent. This suggests that similar to the 'Ethiopian News and Media' topic label being almost frequent across all phases of the war, a significant amount of the discourses was focused on conveying neutral information or opinions.

1) Phase 1 to Phase 2

Phase 1	Lower Bound	Median	Upper Bound
Negative	1,588 (37.5%)	1,655 (39.1%)	1,719 (40.6%)
Neutral	2,078 (49.0%)	2,190 (51.7%)	2,289 (54.0%)
Positive	343 (8.1%)	392 (9.3%)	443 (10.5%)

Phase 2	Lower Bound	Median	Upper Bound
Negative	9,319 (31.1%)	9,712 (32.4%)	10,088 (33.6%)
Neutral	15,125 (50.4%)	15,935 (53.1%)	16.658 (55.5%)
Positive	3,809 (12.7%)	4,357 (14.5%)	4,923 (16.4%)

Table 21. Phase 1 to 2 - Sentiment Distribution

In Phase 1, according to Table 21, the proportion of negative sentiment discourses ranges from 37.5 percent to 40.6 percent. While positive sentiment discourses are much less frequent, occurring between 8.1 percent and 10.5 percent. Most negative discourses support the progress of the Tigray War, such as "golden opportunity turning point government stop tigray" and "majority support war people terrorist tigray." In contrast, positive discourses often oppose the conflict and advocate for peace, with phrases like "may bring peace understanding remove conflict" and "love Ethiopia, wish for peace."

During Phase 2, although negative sentiment discourses remain predominant, accounting for 31.1 percent to 33.6 percent, positive sentiment discourses have increased, ranging from 12.7 percent to 16.4 percent. This marks a rise of between 3.6 percent and 5.9 percent compared to Phase 1. Among negative discourses, there are many adverse statements about the TPLF, such as "need tplf to die together" and "finally ending the terrorist tplf so that the people are finally free from cancer destabilising ethiopia—last tplf ever." On the other hand, positive discourses often express hopes for peace, with comments like "holy land tigray peace loving people tigray grace bless" and "hope ethiopia peace good future."

2) Phase 3 to Phase 4

Phase 3	Lower Bound	Median	Upper Bound
Negative	228 (28.3%)	238 (29.5%)	247 (30.6%)
Neutral	415 (51.5%)	437 (54.2%)	457 (56.7%)
Positive	115 (14.3%)	131 (16.3%)	148 (18.4%)

Phase 4	Lower Bound	Median	Upper Bound
Negative	1,228 (30.2%)	1,279 (31.4%)	1,329 (32.7%)
Neutral	2,107 (51.8%)	2,219 (54.6%)	2.320 (57.0%)
Positive	498 (12.2%)	569 (14.0%)	643 (15.8%)

Table 22. Phase 3 to 4 - Sentiment Distribution

During Phase 3, the portion of negative sentiment discourses decreased compared to the previous phase, ranging from 28.3 percent to 30.6 percent. Conversely, positive discourses increased, rising to a range of 14.3 percent to 18.4 percent. Among the negative discourses, numerous comments mention both the TPLF and the Amhara region, such as "tplf also return stolen amhara region." This likely reflects the many battles that took place in the Amhara and Afar regions during Phase 3. In addition, some negative discourses include criticisms of Eritrea, with statements like "go eritrea eritrea must blame everything eritrea." Meanwhile, positive discourses, such as "god peaceful resolution tigray war better chance" and "may god bring peace horn", predominantly focus on seeking peace.

However, in Phase 4, the proportion of negative discourses increased once more, reaching between 30.2 percent and 32.7 percent. At the same time, the proportion of positive sentiment discourses declined again, falling to between 12.2 percent and 15.8 percent. These results indicate that despite the approaching and actual implementation of the first ceasefire, negative sentiment discourses remained more prevalent than positive discourses. In negative discourses, there are still many statements that consistently label the TPLF as a terrorist group, such as "tdf terrorist group working amhara tigray opposition party." Moreover, negative comments often include references to fake news, with discourse like "responsible fake news supporting ethiopia way." On the other hand, positive discourses frequently include supportive statements about the peace agreement, such as "appreciate united despite perception real issue respecting opposing becomes visible peace everyone needs to think live peacefully within future tigray." The sentiment distributions of Phases 3 and 4 are demonstrated in Table 22.

3) Phase 5 to Phase 6

Phase 5	Lower Bound	Median	Upper Bound
Negative	1,200 (31.0%)	1,251 (32.3%)	1,299 (33.5%)
Neutral	2,087 (53.8%)	2,199 (56.7%)	2,299 (59.3%)
Positive	373 (9.6%)	427 (11.0%)	1483 (12.5%)

Phase 6	Lower Bound	Median	Upper Bound
Negative	427 (24.2%)	445 (25.2%)	462 (26.2%)
Neutral	943 (53.4%)	994 (56.3%)	1.039 (58.8%)
Positive	286 (16.2%)	327 (18.5%)	369 (20.9%)

Table 23. Phase 5 to 6 - Sentiment Distribution

In Phase 5, with the resumption of war, negative discourses become more prevalent, ranging from 31 percent to 33.5 percent. Positive discourses significantly decreased, comprising only 9.6 percent to 12.5 percent. Many negative discourses consistently blame the TPLF and Tigray region, as seen in comments like "tigray always problem around year know making always like war" and "tplf responsible every damage happening ethiopia." There are also several discourses wishing for the TPLF to surrender, with comments like "one tplf must surrender." However, there are also numerous comments criticising Abiy Ahmed, highlighting how innocent people in the Tigray region are suffering due to the Tigray War, such as "remove abiy killing innocent regardless."

On the other hand, positive discourses often express a desire for peace and cessation of the war, with comments like "hope see resolve issue ethiopia common need peace" and "hope best country ethiopia hopefully find agreement get fast growing economy become power." Additionally, some discourses show support for both armies, such as "victory tdf god tigray" and "hero ethiopian military forces victory." Table 23 illustrates the sentiment distributions of Phases 5 and 6.

Meanwhile, in Phase 6, approaching the Pretoria Agreement, the proportion of negative discourses fell sharply to between 24.2 percent and 26.2 percent, while positive discourses significantly increased to a range of 16.2 percent to 20.9 percent. The period during which more positive discourses emerge may indicate that peace talks hold greater significance. With the peace agreement approaching, there is an increase in positive discussions for peace, as seen in comments like "happy people land decided put peace unity way peace" and "god want peace finally big day country god bless ethiopia thanks support one love." Many also highlight the need for humanitarian aid in the Tigray region, through the peace agreement, with statements like "hopefully succeed humanitarian tigray region also whole ethiopia get peace point view thanks."

Nevertheless, negative discussions frequently oppose humanitarian aid to the Tigray region, with comments such as "cut humanitarian aid use aid tigray", or they view the TPLF as terrorists rather than as part of the agreement, as shown in "investigate tplf continue terrorist." This implies that strong feelings remain on both sides as a result of the prolonged war.

5-3. The Word Frequency from Social Media Discourses on the Tigray War

During the Tigray War, 12,771 words were utilised to form the social media discourse. This research vectorised the entire social media discourse through NLTK and analysed word frequency using the Wordcount library in Python, covering overall trends as well as each phase of the war. The Wordcount library processes text data to count the frequency of each word (Navaro, 2022). Table 24 describes the Top 30 most frequently occurring words and their respective proportions among all words during the Tigray War on YouTube.

Tigray War Word Frequency								
No	**Vocabulary**	**Ratio**	**No**	**Vocabulary**	**Ratio**	**No**	**Vocabulary**	**Ratio**
1	Tigray	3.16%	11	Eritrea	0.69%	21	Know	0.45%
2	Ethiopia	2.56%	12	One	0.68%	22	New	0.44%
3	TPLF	2.33%	13	Ahmed	0.63%	23	Time	0.41%
4	People	1.61%	14	Peace	0.63%	24	Never	0.36%
5	War	1.41%	15	Like	0.61%	25	Support	0.35%
6	Abiy	1.29%	16	Forces	0.56%	26	God	0.35%
7	Amhara	1.02%	17	Us	0.55%	27	Also	0.35%
8	Government	0.90%	18	Military	0.45%	28	Would	0.35%
9	TDF	0.90%	19	Region	0.45%	29	Get	0.34%
10	News	0.78%	20	Country	0.45%	30	Please	0.33%

Table 24. Overall Word Frequency (Tigray War)

The word 'Tigray', representing the region affected by the war, is the most frequently used term, accounting for 3.16 percent of all words. Additionally, the 'Amhara' region, also related to this conflict, is mentioned in 1.02 percent of the text. Keywords directly connected to the war, such as 'War', 'TDF', (the Tigray Defense Forces), 'Forces', and 'Military', contribute 1.41 percent, 0.90 percent, 0.56 percent, and 0.45 percent, respectively. Terms associated with the parties and stakeholders involved in the war, like 'TPLF', 'Abiy', 'Government', and 'Ahmed', hold proportions of

2.33 percent, 1.29 percent, 0.90 percent, and 0.63 percent. On the other hand, the word ‘Peace’, representing the desire to finish the war, makes up 0.63 percent of the discourse. The word ‘One’, implying unity, appears with a relative frequency of 0.68 percent. While other words are frequently used, their meanings may not be clear without additional context. The details of each phase’s word frequency are demonstrated in Appendix 3.

Tigray War Major Terms				
Category	**Term 1**	**Represent**	**Term 2**	**Represent**
1	TPLF	Tigray region	Abiy	Ethiopian Government
2	TDF	Forces of TPLF	Military	Forces of the Ethiopian Government
3	Peace	Desire to end the war	One	Unity

Table 25. Tigray War Major Terms

To analyse the Tigray War, among thirty words, six words closely related to the conflict were selected as major terms. These terms represent the two opposing sides, 'TPLF' and 'Abiy'; their forces, 'TDF' and 'Military'; and symbols of the end of the war and peace, 'Peace' and 'One.' Subsequently, these terms were categorised into three groups, as illustrated in Table 25, and their frequency trends throughout the war were examined. This approach allows us to understand the dynamics of the conflict and how different aspects evolved.

1) TPLF and Abiy

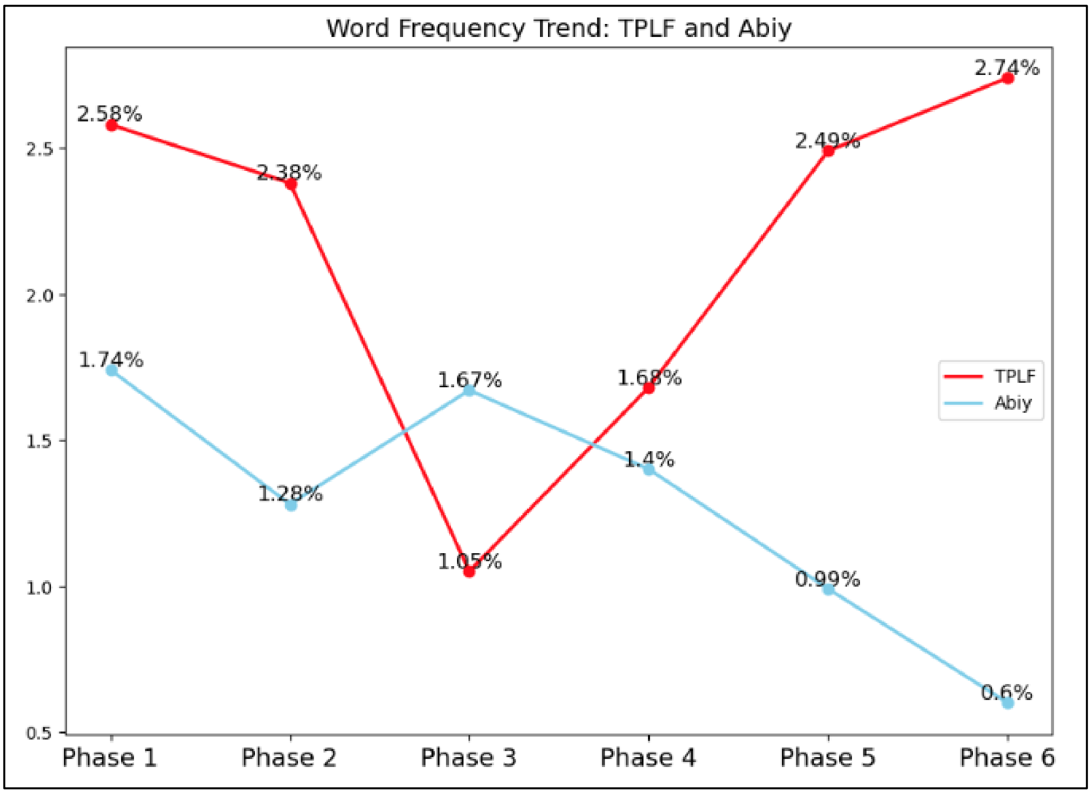


Figure 16. Word Frequency Trend: TPLF and Abiy

According to Figure 16, the frequency of appearances of TPLF and Abiy Ahmed, the representative of the Ethiopian government, in discourse related to the Tigray War, changes as the conflict progresses.

In Phase 1, at the start of the war, the term TPLF appears in 2.58 percent of the discourse, while Abiy Ahmed is mentioned in 1.74 percent. As the war continues into Phase 2, the frequency of both terms slightly decreases, with TPLF appearing in 2.38 percent and Abiy Ahmed in 1.28 percent of the discourse. However, during Phase 3, which is the month leading up to the first ceasefire, the frequency of Abiy Ahmed rose significantly to 1.67 percent, surpassing TPLF, which dropped sharply to 1.05 percent. This change is likely due to discussions surrounding, the Ethiopian government's declaration of an indefinite humanitarian truce on 24th March 2022 (Dahir and Marks, 2022), which aimed to allow the delivery of humanitarian aid into Tigray.

Consequently, during the first ceasefire, the frequency of mentions of Abiy Ahmed decreased to 1.4 percent, while mentions of the TPLF increased by 0.63 percent to 1.68 percent. As the conflict resumed in Phase 5, the frequency of TPLF mentions rose to 2.49 percent, whereas mentions of Abiy Ahmed fell to 0.99 percent. The war resumed with both sides blaming each other. According to the Ethiopian News Agency (2022), tensions escalated on the 24th of August 2022 when the Ethiopian Air Force shot down a plane, claiming it was transporting weapons for the TPLF. In the period leading up to the Pretoria Agreement, the TPLF continues to be mentioned more frequently, making up 2.74 percent of the discourse, compared to just 0.6 percent for Abiy Ahmed. This discrepancy might be attributed to the fact that the peace talks were mediated by the African Union (AU), which could have resulted in less emphasis on Abiy Ahmed. Furthermore, since the conflict predominantly unfolded in the Tigray region, there is likely greater interest in the TPLF.

It is important to note that this study does not consider the term 'Ahmed' since this is his last name, so his first name, 'Abiy', is used in the analysis.

2) TDF and Military

As described in Figure 17, during the initial period of the war, the TDF did not feature in the discourse, whereas the Ethiopian Military was mentioned with a frequency of 0.6 percent. In Phase 2, as the conflict intensified, the TDF began to appear in 1.04 percent of the discourse, while mentions of the military decreased to 0.47 percent. In Phase 3, one month before the first ceasefire, the frequency of TDF mentions fell by about 0.4 percent to 0.6 percent, and the mentions of the military completely ceased. This lack of frequency might be related to the short duration of Phase 3, which lasted only a month.

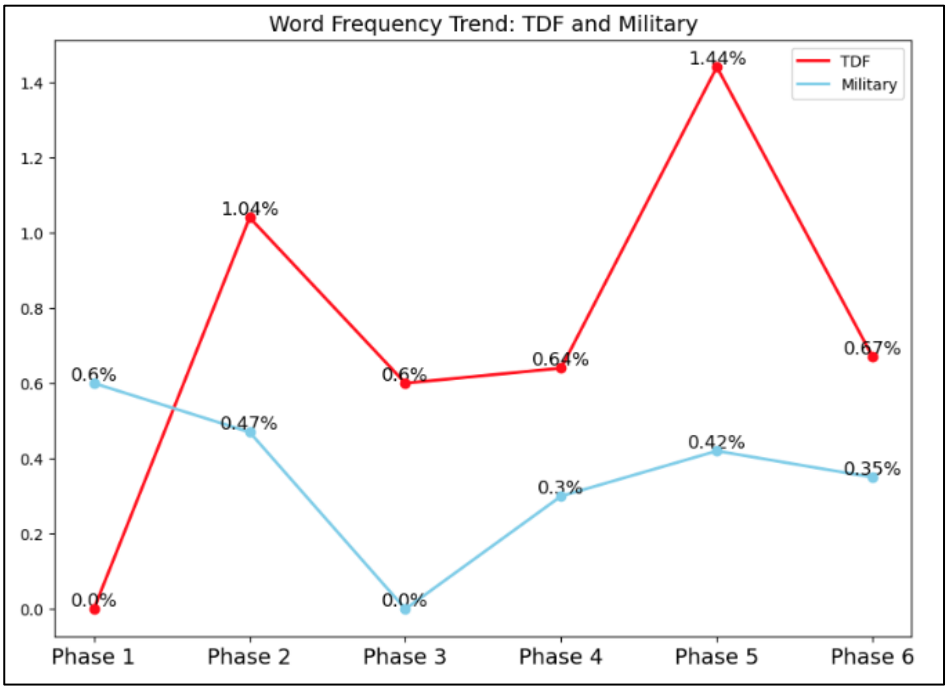


Figure 17. Word Frequency Trend: TDF and Military

During the first ceasefire, mentions of the TDF are at 0.64 percent, similar to the previous phase, while the military reappears in the discourses at a lower frequency of 0.3 percent. This lower frequency is likely due to the cessation of active battles during the ceasefire. With the reactivation of the war, the frequency of mentions of the TDF increased largely to 1.44 percent as battles recommenced, while mentions of the military also elevated slightly to 0.42 percent. As peace negotiations advanced, the frequencies of both TDF and military mentions declined, reaching 0.67 percent and 0.35 percent, respectively.

3) Peace and One

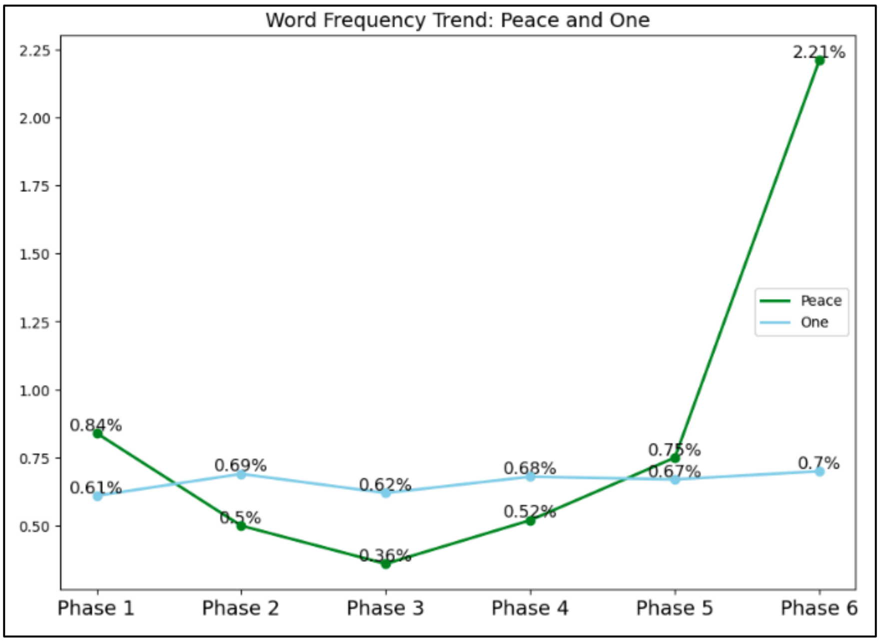


Figure 18. Word Frequency Trend: Peace and One

The word frequency trends of terms symbolising peace, such as 'Peace' and 'One,' are demonstrated in Figure 18. For 'One,' the frequency remains relatively stable, ranging from 0.60 to 0.70 percent across Phases 1 to 6. In contrast, 'Peace' appears with a frequency of 0.84 percent in Phase 1, but as the conflict intensifies, this frequency decreases to 0.5 percent and drops further to 0.36 percent in the month leading up to the first ceasefire.

Afterwards, during the first ceasefire, the frequency of 'Peace' slightly rises to 0.52 percent compared to the previous phase. This increase can be attributed to the halt of combat and the arrival of humanitarian aid in the Tigray region. In Phase 5, when the war recommences, 'Peace' is mentioned more frequently at 0.75 percent, indicating that despite the renewed conflict, the focus on peace grows stronger because of the first peace agreement. Subsequently, during the peace talks, the frequency of 'Peace' surged significantly to 2.21 percent, suggesting that public attention was strongly concentrated on achieving peace in Ethiopia.

6. Discussion and Conclusion

6-1. Discussion

Through topic modelling, sentiment analysis, and word frequency analysis, this research delved into social media discourse on the Tigray War from YouTube. For research question 1, the data analysis in the results section revealed the topics, sentiments, and major terms of the war and their word frequency trends related to the overall conflict and each phase of the Tigray War.

Firstly, as captured in Table 13, various topics emerged during the Tigray War. These topics include labels such as 'Ethiopian News and Media,' 'International Politics,' 'Ethiopian Government,' 'Religion,' and 'Politics,' as well as Tigray War-related labels like 'Battle and Conflict,' 'TPLF,' 'Abiy Ahmed,' 'Humanitarian Crisis,' 'Peace,' and 'Ceasefire.' The proportion of the labels 'Battle and Conflict' and 'Ethiopian News and Media' were consistently predominant throughout the war. In contrast, other topics such as 'TPLF,' 'Abiy Ahmed,' 'Peace,' and 'Ceasefire' fluctuated according to the dynamics of the conflict.

For instance, in Phase 1, topics under the 'Anti-TPLF' label appeared more frequently than those under 'Pro-TPLF.' But, in Phase 2, 'Pro-TPLF' related topics became more prevalent. Subsequently, in Phase 6, 'Anti-TPLF' topics once again dominated. Additionally, topics under the 'Ceasefire' label saw an increase in frequency during both the first ceasefire (Phase 4) and the final ceasefire (Phase 6), compared to previous stages.

Furthermore, Table 20 illustrates the overall trend of discourse sentiment during the Tigray War. In general, comments containing neutral opinions or information are most prevalent, with a lower bound of 50.8 percent and an upper bound of 56 percent. Negative discourse ranges from 31.3 percent to 33.8 percent, which is higher than the positive discourse bound of 12.1 percent to 15.7

percent. As the war progresses, the proportion of negative and positive discourse shifts, as described in Tables 21 to 23, but the bounds for neutral discourse remain relatively unchanged.

Comments with positive sentiment mainly expressed wishes for peace across all phases of the war. However, in Phase 5, there was a notable increase in comments supporting and complimenting TDF and the Ethiopian Military. Meanwhile, negative sentiment comments generally criticised TPLF. In Phase 4, there was also a significant amount of criticism directed at fake news. In Phase 6, the focus shifted to increased opposition to humanitarian aid for the Tigray region.

In addition, the top 30-word frequency, as demonstrated in Table 24, highlighted the 30 terms most frequently mentioned in the context of the Tigray War. Among these terms, this study identified 'TPLF,' 'Abiy,' 'TDF,' 'Military,' 'Peace,' and 'One' as major terms. These words were selected because they stand for the two opposing sides, their forces, and the hope for peace, rather than other words.

According to Figure 16, 'TPLF' appeared more frequently than 'Abiy' in all phases except Phase 3, with the gap widening particularly during the period following the breakdown of the first ceasefire and the subsequent peace talks. Meanwhile, 'TDF' was used more often than 'Military' overall, although both terms saw a slight increase in usage during the first ceasefire. However, their frequency decreased during the peace talks, as shown in Figure 17.

Words related to peace, such as 'Peace' and 'One,' saw an increase in frequency during the first ceasefire. In Phase 6, 'Peace' more than doubled in frequency, while 'One' also showed a slight increase, though only by 0.03 percent. This result indicates that terms related to the Tigray region were mentioned more frequently than those related to the Ethiopian Government. It also shows that the frequency of six terms was influenced by the first ceasefire (Phase 4) and the peace talks (Phase 6).

On the other hand, regarding research question 2, this research concludes that the outputs from topic modelling, sentiment analysis, and word frequency analysis can provide valuable insights that can aid in preparing and implementing a ceasefire for armed conflict.

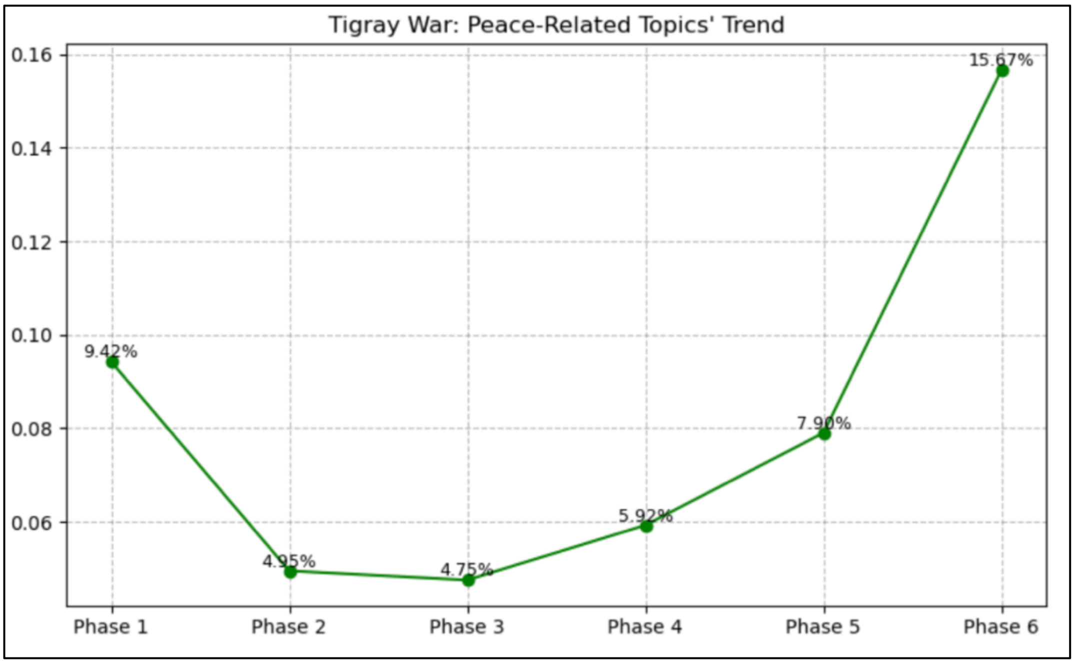


Figure 19. Trend of Peace-Related Topics (Tigray War)

More specifically, through topic modelling, this study discovered labels for peace-related topics, including 'Peace' and 'Ceasefire.' As illustrated in Figure 19, the proportion of peace-related topics significantly increased as the conflict approached its final phase (Phase 6). Although there is a noticeable drop from Phase 1 to Phase 3, the proportion of peace-related topics begins at 4.75 percent one month before the first ceasefire (Phase 3), and rises to 5.9 percent, 7.90 percent, and ultimately reaches 15.67 percent in the final phase.

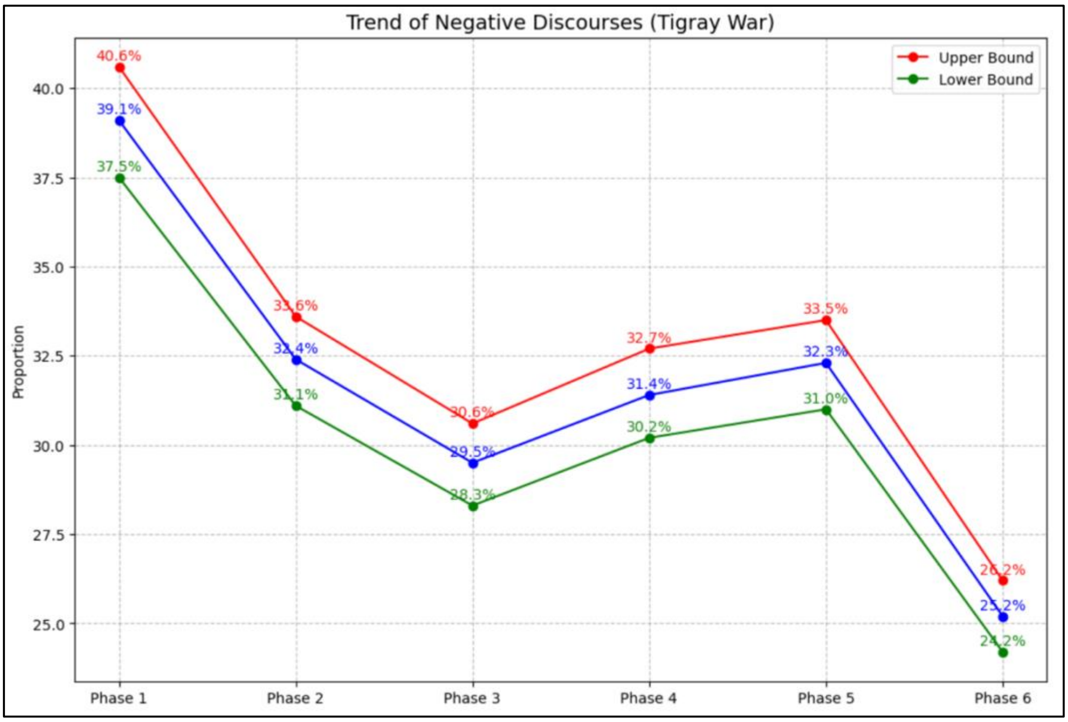


Figure 20. Trend of Negative Discourses (Tigray War)

Besides, as shown in Figure 20, this research reveals that the proportion of negative discourses decreased as the Pretoria Agreement approached. Although there is an increase in negative discourses during the first ceasefire, a significant decline is observed when transitioning from Phase 5 to Phase 6. Particularly, in Phase 5, negative discourses range from 31 percent to 33.5 percent, while in Phase 6, they drop to between 24.2 percent and 26.2 percent.

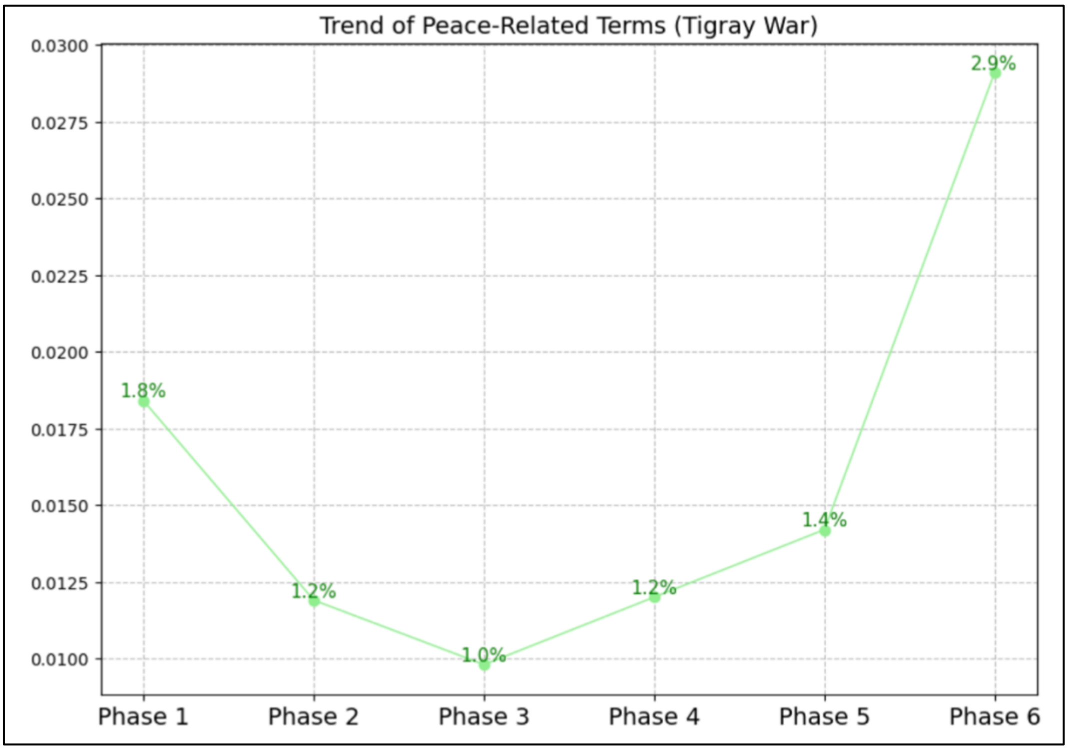


Figure 21. Trend of Peace-Related Terms (Tigray War)

Moreover, this study identified peace-related terms such as 'Peace' and 'One' using word frequency analysis. As detailed in Figure 21, the use of peace-related words, 'Peace' and 'One', in Phase 1 was 1.8 percent, but it gradually decreased to 1.2 percent in Phase 2 and 1.0 percent in Phase 3. However, during the first ceasefire period, the usage increased to 1.2 percent, and in Phase 5, when the war resumed, it rose further to 1.4 percent. Finally, during Phase 6, leading up to the final peace agreement, the usage of such terms reached approximately 2.9 percent, showing a notable increase. This suggests that the frequency of words related to peace and ceasefire, such as 'peace', 'stop', 'ceasefire', 'one', and similar terms, increased as the Pretoria Agreement drew nearer.

In conclusion, as the final peace agreement approaches, the frequency of topics and words related to peace increases significantly. Conversely, as the agreement draws nearer, the proportion of negative discourses tends to decrease. These results imply that by continuously analysing social media discourse during an armed conflict, it is possible to predict changes in public support for war and ceasefire. Since public support is a crucial factor in ceasefire negotiations and agreements

(Tellez, 2019), understanding the level of public support during the conflict can be used as additional information for preparing and implementing peace talks aimed at the cessation of the conflict. By leveraging these insights, mediators and stakeholders could better align their strategies with public support, ultimately improving the prospects for achieving a successful ceasefire.

6-2. Conclusion

This study conducted the social media discourse analysis on the Tigray War through NLP methodologies such as topic modelling, sentiment analysis, and word frequency analysis.

The data was collected from YouTube using four keywords: Tigray War, Tigray, TPLF, and Abiy Ahmed. A total of 2,208,035 data points were initially collected and pre-processed. As a result, a total of 44,757 data points were secured for conducting social media discourse analysis.

BERTopic was utilised to conduct topic modelling, while the RoBERTa-base 1 model was employed to implement sentiment analysis among seven sentiment analysis models, described in Table 10. For word frequency analysis, this research tokenised and vectorised the text data and analysed these using the Wordcount Python library.

Based on the analysis results of the research questions, this study concludes that analysing social media discourse on an armed conflict using NLP methods can ① enhance understanding of the conflict through insights into related topics, sentiments, and word frequency and ② offer supplementary information that could be used as evidence to plan and facilitate ceasefire negotiations.

6-3. Recommendation

This study utilised social media discourse from YouTube due to the constraints of accessing data from Facebook due to Meta's policies. However, as previously mentioned, Facebook is the most widely used social networking site in Ethiopia. Therefore, utilising discourses from Facebook could potentially enhance the quality of data.

Meanwhile, Ethiopia's state-owned telecommunications company, Ethio-Telecom, frequently blocked internet access from 2020 to 2022, preventing the use of social media platforms during these periods (Zelalem, 2022). Although there is no precise information on when these internet blockages occurred, they overlap with the Tigray War. Thus, there may have been periods when data generation was low due to internet blockages, potentially affecting the collection of data across the entire phase of the war.

Besides, as noted earlier, there is no relevant research on social media discourse that mainly targets an armed conflict case. As a result, this research manually labelled social media discourses, which may contain the possibility of researcher bias. Nonetheless, this study provides a valuable example of analysing social media discourse on armed conflict, and it is expected to serve as an impactful reference for future research in similar areas.

For future research, developing sentiment analysis models specifically tailored to social media discourses about armed conflicts could enable more accurate sentiment extraction. Additionally, creating models capable of distinguishing political discourses within social media content would facilitate effective data pre-processing and allow for efficient yet robust analysis in similar future studies.

7. References

- Addis Standard. (2022). "Agreement for Lasting Peace through a Permanent Cessation of Hostilities between the Government of the Federal Democratic Republic of Ethiopia and the Tigray People's Liberation Front.” https://addisstandard.com/wp-content/uploads/2022/11/AU-led-Ethiopia-Peace-Agreement.pdf
- African Union. (2022). Cessation of Hostilities Agreement between the Government of the Federal Democratic Republic of Ethiopia and the Tigray Peoples’ Liberation Front (TPLF). African Union - Peace and Security Department. https://www.peaceau.org/en/article/cessation-of-hostilities-agreement-between-the-government-of-the-federal-democratic-republic-of-ethiopia-and-the-tigray-peoples-liberation-front-tplf
- Al Jazeera. (2018). Ethiopia prime minister Hailemariam Desalegn resigns. Al Jazeera. https://aje.io/k4xp6
- Al Jazeera. (2023). What’s behind the crisis in Ethiopia’s Amhara region? A simple guide. Al Jazeera. https://www.aljazeera.com/news/2023/8/10/whats-behind-the-crisis-in-ethiopias-amhara-region-a-simple-guide
- Azevedo Nicolas. (2024). What is Hugging Face? The ML Platform For Building AI-Powered Apps. Scalable Path. https://www.scalablepath.com/machine-learning/hugging-face
- Balehegn, M. (2021). The politics and problems of Prosperity Party Gospel. Ethiopia Insight. https://www.ethiopia-insight.com/2021/04/04/the-politics-and-problems-of-prosperity-party-gospel/
- Barbieri, F., Anke, L.E. and Camacho-Collados, J. (2022). XLM-T: Multilingual Language Models in Twitter for Sentiment Analysis and Beyond. arXiv:2104.12250 [cs]. https://arxiv.org/abs/2104.12250
- Belay, E. G., Mengesha, G. H., & Asale, M. A. (2020). The Tributes and Perils of Social Media Use Practices in Ethiopian Socio-political Landscape. Lecture notes in computer science, pp.199–209. doi:https://doi.org/10.1007/978-3-030-60152-2_16
- Blanchard, L.P. (2021). Ethiopia’s Transition and the Tigray Conflict. Congressional Research Service. https://sgp.fas.org/crs/row/R46905.pdf
- Blei, D., Ng, A. and Jordan, M. (2003). Latent Dirichlet Allocation. Journal of Machine Learning Research, 3, pp.993–1022
- Brown, S. (2021). Machine learning, explained, https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained
- Burke, J. (2020). Rise and fall of Ethiopia’s TPLF – from rebels to rulers and back. The Guardian. https://www.theguardian.com/world/2020/nov/25/rise-and-fall-of-ethiopias-tplf-tigray-peoples-liberation-front
- Burke, J. and Henley, J. (2019). Abiy Ahmed, Ethiopia’s prime minister, wins 2019 Nobel peace prize. The Guardian. https://www.theguardian.com/world/2019/oct/11/abiy-ahmed-ethiopian-prime-minister-wins-2019-nobel-peace-prize?CMP=share_btn_url
- Campello, R.J.G.B., Moulavi, D., Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In: Pei, J., Tseng, V.S., Cao, L., Motoda, H., Xu, G. (eds) Advances in Knowledge Discovery and Data Mining. PAKDD 2013. Lecture Notes in Computer Science(), vol 7819. Springer,

Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-37456-2_14

- Caruso, F. and Akamo, J. O. (2024) 'EU Policy towards Ethiopia amidst the Tigray War: The Limits of Mitigating Fragmentation', The International Spectator, 59(1), pp. 120–139. doi: 10.1080/03932729.2024.2302473
- Center for Preventive Action. (2022). Conflict in Ethiopia. Global Conflict Tracker. https://www.cfr.org/global-conflict-tracker/conflict/conflict-ethiopia
- Chadwick, A. (2013). The hybrid media system: Politics and power. New York, Ny: Oxford University Press
- Clayton, Govinda; Nygård, Håvard Mokleiv; Rustad, Siri Aas; Strand, Håvard. (2022). Ceasefires in Civil Conflict: A Research Agenda. Journal of Conflict Resolution, 67 (7–8): 1279–1295. doi:https://doi.org/10.1177/00220027221128300
- Dahir, A.L. and Marks, S. (2022). Ethiopia Declares 'Humanitarian Truce' in War-Ravaged Tigray Region. The New York Times. https://www.nytimes.com/2022/03/24/world/africa/ethiopia-tigray-conflict-truce.html
- Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv.org. doi:https://doi.org/10.48550/arXiv.1810.04805
- Dixon, J. (2009). What Causes Civil Wars? Integrating Quantitative Research Findings. International Studies Review, 11(4), pp.707–735. https://www.jstor.org/stable/40389163
- Domo. (2022). Data Never Sleeps 10.0 | Domo. https://www.domo.com/data-never-sleeps#data
- Efron, B. and Tibshirani, R. (1998). An introduction to the bootstrap. Boca Raton, Fla Chapman & Hall/Crc.
- Egger, R. and Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. Frontiers in Sociology, 7. doi:https://doi.org/10.3389/fsoc.2022.886498
- Ethiopian News Agency. (2022). Ethiopia Downs Aircraft Transporting Weapons to Terrorist TPLF. Ethiopian News Agency. https://www.ena.et/web/eng/w/en_37974
- Ethnologue. (2013). Ethiopia | Ethnologue Free. https://www.ethnologue.com/country/ET/
- Fortna, V. P. (2018). Peace Time Cease-Fire Agreements and the Durability of Peace. Princeton University Press. ISBN 978-0-691-18795-2
- Grootendorst, M. (2021). BERTopic. GitHub. Available at: https://github.com/MaartenGr/BERTopic
- Grootendorst, M. (2021). Parameter tuning - BERTopic. Github.io. https://maartengr.github.io/BERTopic/getting_started/parameter%20tuning/parametertuning.html
- Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. arXiv pre- print arXiv:2203.05794
- Gudivada, V.N., Rao, D. and Raghavan, V.V. (2015). Big Data Driven Natural Language Processing Research and Applications. pp.203–238. doi:https://doi.org/10.1016/b978-0-444-63492-4.00009-5
- Hughes, C. (2019). Opinion | It's Time to Break Up Facebook. The New York Times. 9 May. https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html
- Hutto, C.J. & Gilbert, E.E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Ann Arbor, MI, June 2014

- Jurafsky, D. and Martin, J.H. (2008). Speech and language processing : an introduction to natural language processing, computational linguistics, and speech recognition. Upper Saddle River, N.J: Pearson Education, Harlow
- Kelecha, M. (2024). Understanding Ethiopia's Tigray War. International Affairs, 100(1), pp.445–446. doi:https://doi.org/10.1093/ia/iiad328
- Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L. and Brown, D. (2019). Text Classification Algorithms: A Survey. Information, 10(4), p.150. doi:https://doi.org/10.3390/info10040150
- Landauer, T.K., McNamara, D.S., Dennis, S., & Kintsch, W. (Eds.). (2007). Handbook of Latent Semantic Analysis (1st ed.). Psychology Press. doi:https://doi.org/10.4324/9780203936399
- Lau, J.H., Newman, D. and Baldwin, T. (2014). Machine Reading Tea Leaves: Automatically Evaluating Topic Coherence and Topic Model Quality. Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics. doi:https://doi.org/10.3115/v1/e14-1056
- LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep Learning. Nature, 521(7553), pp.436–444. doi:https://doi.org/10.1038/nature14539
- Leppanen, S., Westinen, E., & Kytola, S. (2016). Social Media Discourse, (Dis)identifications and Diversities (1st ed.). Routledge. doi:https://doi-org.sussex.idm.oclc.org/10.4324/9781315624822
- Lian, Y., Lin, X., Dong, X. and Hou, S. (2022). A Normalized Rich-Club Connectivity-Based Strategy for Keyword Selection in Social Media Analysis. Sustainability, 14(13), pp.7722–7722. doi:https://doi.org/10.3390/su14137722
- Liu, Zhiyuan., Lin, Yankai. and Sun, Maosong. (2020) Representation Learning for Natural Language Processing. 1st ed. 2020. Singapore: Springer Nature Singapore. doi:10.1007/978-981-15-5573-2
- Manica, M., Mathis, R., Cadow, J. and Rodríguez Martínez, M. (2019). Context-specific interaction networks from vector representation of words. Nature Machine Intelligence, 1(4), pp.181–190. doi:https://doi.org/10.1038/s42256-019-0036-1
- McInnes, L., Healy, J. and Melville, J. (2020). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv.org. doi:https://doi.org/10.48550/arXiv.1802.03426
- Mekonen, Fikir G. (2023). A year after the Pretoria agreement, hard work remains for Ethiopia. ISS Africa. https://issafrica.org/iss-today/a-year-after-the-pretoria-agreement-hard-work-remains-for-ethiopia
- MLflow. (2023). Introduction to Advanced Semantic Similarity Analysis with Sentence Transformers and MLflow. MLflow 2.15.1 documentation. https://mlflow.org/docs/latest/llms/sentence-transformers/tutorials/semantic-similarity/semantic-similarity-sentence-transformers.html
- Mwakideu, C. (2023) 'Can Berlin, Paris reset 'strained' EU-Ethiopia ties?', DW. https://www.dw.com/en/can-berlin-and-paris-reset-strained-eu-ethiopia-ties/a-64343852
- Navaro, P. (2022). Wordcount — Python tools for Big data. ENSAI. https://pnavaro.github.io/big-data/04-WordCount.html
- Neeraj Anand Sharma, Ali and Muhammad Ashad Kabir (2024). A review of sentiment analysis: tasks, applications, and deep learning techniques. International journal of data science and analytics. doi:https://doi.org/10.1007/s41060-024-00594-x

- Oetzel, J. and Getz, K. (2011). Why and how might firms respond strategically to violent conflict? Journal of International Business Studies, 43(2), pp.166–186. doi:https://doi.org/10.1057/jibs.2011.50
- Pérez, J.M., Rajngewerc, M., Giudici, J.C., Furman, D.A., Luque, F., Alemany, L.A. and Martínez, María Vanina (2021). pysentimiento: A Python Toolkit for Opinion Mining and Social NLP tasks. arXiv (Cornell University). doi:https://doi.org/10.48550/arxiv.2106.09462
- Pettersson, T. and Wallensteen, P. (2015). Armed conflicts, 1946–2014. Journal of Peace Research, 52(4), pp.536–550. https://www.jstor.org/stable/24557437
- Rajaraman, A. and Ullman, J.D. (2011) 'Data Mining', in Mining of Massive Datasets. Cambridge: Cambridge University Press, pp. 1–17. doi:https://doi.org/10.1017/cbo9781139058452.002
- Rocha, Y.M., de Moura, G.A., Desidério, G.A., de Oliveira, C.H., Lourenço, F.D. and de Figueiredo Nicolete, L.D. (2021). The impact of fake news on social media and its influence on health during the COVID-19 pandemic: a systematic review. Journal of Public Health, 1(10). doi:https://doi.org/10.1007/s10389-021-01658-z
- Röder, M., Both, A. and Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. Proceedings of the Eighth ACM International Conference on Web Search and Data Mining - WSDM '15, pp.399–408. doi:https://doi.org/10.1145/2684822.2685324
- Sanguansat, P. ed., (2012). Principal Component Analysis. InTech. doi:https://doi.org/10.5772/2340
- Sathya, R. and Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. International Journal of Advanced Research in Artificial Intelligence, 2(2), 34-35. doi:https://doi.org/10.14569/ijarai.2013.020206
- Schipani, A. (2023). Ethiopia seeks help to find $20bn for post-conflict reconstruction. Financial Times. https://www.ft.com/content/cdd73748-58f3-41d4-add1-cd563372c6de
- Serneels, P. and Verpoorten, M. (2013). The Impact of Armed Conflict on Economic Performance. Journal of Conflict Resolution, 59(4), pp.555–592. doi:https://doi.org/10.1177/0022002713515409
- Simmons, B. A. (2009), Mobilizing for Human Rights: International Law in Domestic Politics, Cambridge University Press, New York
- Srivastava, A.N., & Sahami, M. (Eds.). (2009). Text Mining: Classification, Clustering, and Applications (1st ed.). Chapman and Hall/CRC. doi:https://doi.org/10.1201/9781420059458
- Tellez, J.F. (2019). Peace agreement design and public support for peace: Evidence from Colombia. Journal of Peace Research, 56(6), pp.827–844. doi:https://doi.org/10.1177/0022343319853603
- Tondak, akshay. (2020). Deep Learning Vs Machine Learning | Know The Difference, https://k21academy.com/datascience-blog/deep-learning/dl-vs-ml/
- Törnberg, A. and Törnberg, P. (2016). Muslims in social media discourse: Combining topic modeling and critical discourse analysis. Discourse, Context & Media, 13, pp.132–142. doi:https://doi.org/10.1016/j.dcm.2016.04.003
- Tucker, J.A., Theocharis, Y., Roberts, M.E. and Barberá, P. (2017). From Liberation to Turmoil: Social Media And Democracy. Journal of Democracy, 28(4), pp.46–59. doi:https://doi.org/10.1353/jod.2017.0064

- UCDP. (2006). Definitions, sources and methods for Uppsala Conflict Data Program Battle-Death estimates. Uppsala Conflict Data Program (UCDP). https://ucdp.uu.se/downloads/old/brd/ucdp-brd-conf-41-2006.pdf
- United Nations Department of Political and Peacebuilding Affairs (UN DPPA). (2022). Guidance on the Mediation of Ceasefires. https://peacemaker.un.org/thematic-areas/ceasefires-security-arrangements
- UNICEF. (2022). UNICEF Ethiopia Humanitarian Situation Report No. 9 - September 2022 - Ethiopia | ReliefWeb. https://reliefweb.int/report/ethiopia/unicef-ethiopia-humanitarian-situation-report-no-9-september-2022
- United Nations (2023). Ethiopia: Mass killings continue, risk of further 'large-scale' atrocities | UN News. https://news.un.org/en/story/2023/09/1140872
- Wafiq, N.F. (2023). The Power of Social Media: Shaping Political Discourse in the Digital Age. Modern Diplomacy. https://moderndiplomacy.eu/2023/05/11/the-power-of-social-media-shaping-political-discourse-in-the-digital-age/
- Wallace, J. (2017). Modelling Contemporary Gatekeeping. The rise of individuals, algorithms and platforms in digital news dissemination. Digital Journalism, 6(3), pp.274–293. doi:https://doi.org/10.1080/21670811.2017.1343648
- Wehrens, R., Putter, H. and Buydens, L.M.C. (2000). The bootstrap: a tutorial. Chemometrics and Intelligent Laboratory Systems, 54(1), pp.35–52. doi:https://doi.org/10.1016/s0169-7439(00)00102-7
- Winning, A. and Cocks, T. (2022). Combatants in Ethiopia's Tigray war agree to stop fighting. Reuters. https://www.reuters.com/world/africa/african-union-parties-ethiopia-conflict-have-agreed-cease-hostilities-2022-11-02/
- Yao, J. (2019). Automated Sentiment Analysis of Text Data with NLTK. Journal of Physics: Conference Series, 1187(5), p.052020. doi:https://doi.org/10.1088/1742-6596/1187/5/052020
- York, Geoffrey. (2022). "Surge of dehumanizing hate speech points to mounting risk of mass atrocities in northern Ethiopia, experts say". The Globe and Mail. https://www.theglobeandmail.com, Archived on 22 October 2022
- Zagibalov, Taras (2010). Unsupervised and knowledge-poor approaches to sentiment analysis. University of Sussex. Thesis. https://hdl.handle.net/10779/uos.23314979.v1
- Zelalem, Z. (2022). Six million silenced: A two-year internet outage in Ethiopia | Context. www.context.news. https://www.context.news/digital-rights/six-million-people-in-the-dark-tigrays-two-year-internet-outage

8. Appendix

8-1. Appendix 1: Code of Conduct*

*Source: https://www.bcs.org/membership-and-registrations/become-a-member/bcs-code-of-conduct

BCS, The Chartered Institute for IT

CODE OF CONDUCT FOR BCS MEMBERS

Version 8 last reviewed and approved by Trustee Board 8 June 2022

Introduction

As a professional body, the British Computer Society (known as BCS, the Chartered Institute for IT), has a responsibility to set rules and professional standards to direct the behaviour of its members in professional matters. It is expected that these rules and professional standards will be higher than those established by the general law and that they will be enforced through disciplinary action which can result in expulsion from membership.

Members are expected to exercise their own judgement (which should be made in such a way as to be reasonably justified) to meet the requirements of the code and seek advice if in doubt.

Appendix 1 to the code sets examples of interpretation of the tenets of professional conduct and form part of this Code of Conduct.

Breaches of the Code of Conduct

If a member of BCS should know of, or become aware of, any breach of this Code of Conduct by another member they are under an obligation to notify BCS immediately.

Breaches of the Code of Conduct may also be brought to the attention of BCS by others who are not members of BCS.

Any breach of the Code of Conduct brought to the attention of BCS, or of which BCS becomes aware, will be considered under the Institute's disciplinary procedures.

Correspondence in connection with this Code of Conduct should be directed to:

Customer Service team
BCS, The Chartered Institute for IT,
3 Newbridge Square
Swindon
SN1 1BY

Email: custsupport@bcs.uk

BCS CODE OF CONDUCT

This Code of Conduct:

- sets out the professional standards required by BCS as a condition of membership.
- applies to all members, irrespective of their membership grade, the role they fulfil, or the jurisdiction where they are employed or discharge their contractual obligations.
- governs the conduct of the individual, not the nature of the business or ethics of any Relevant Authority*.

1. Public Interest

You shall:

a. have due regard for public health, privacy, security and wellbeing of others and the environment.

b. have due regard for the legitimate rights of Third Parties.

c. conduct your professional activities without discrimination on the grounds of sex, sexual orientation, marital status, nationality, colour, race, ethnic origin, religion, age or disability, or of any other condition or requirement.

d. promote equal access to the benefits of IT and seek to promote the inclusion of all sectors in society wherever opportunities arise.

2. Professional Competence and Integrity

You shall:

a. only undertake to do work or provide a service that is within your professional competence.

b. **NOT** claim any level of competence that you do not possess.

c. develop your professional knowledge, skills and competence on a continuing basis, maintaining awareness of technological developments, procedures, and standards that are relevant to your field.

d. ensure that you have the knowledge and understanding of Legislation* and that you comply with such Legislation, in carrying out your professional responsibilities.

e. respect and value alternative viewpoints and, seek, accept and offer honest criticisms of work.

f. avoid injuring others, their property, reputation, or employment by false or malicious or negligent action or inaction.

g. reject and will not make any offer of bribery or unethical inducement.

3. Duty to Relevant Authority

You shall:

a. carry out your professional responsibilities with due care and diligence in accordance with the Relevant Authority's requirements whilst exercising your professional judgement at all times.

b. seek to avoid any situation that may give rise to a conflict of interest between you and your Relevant Authority.

c. accept professional responsibility for your work and for the work of colleagues who are defined in a given context as working under your supervision.

d. **NOT** disclose or authorise to be disclosed, or use for personal gain, or to benefit a third party, confidential information except with the permission of your Relevant

Authority, or as required by Legislation.

e. **NOT** misrepresent or withhold information on the performance of products, systems or services (unless lawfully bound by a duty of confidentiality not to disclose such information), or take advantage of the lack of relevant knowledge or inexperience of others.

4. Duty to the Profession

You shall:

a. accept your personal duty to uphold the reputation of the profession and not take any action which could bring the profession into disrepute.

b. seek to improve professional standards through participation in their development, use and enforcement.

c. uphold the reputation and good standing of BCS, the Chartered Institute for IT.

d. act with integrity and respect in your professional relationships with all members of BCS and with members of other professions with whom you work in a professional capacity.

e. encourage and support fellow members in their professional development.

APPENDIX 1

Interpretation of the BCS Code of Conduct

The explanatory notes below are offered for guidance only. The examples are not, and not intended to be, exhaustive.

If you are in a leadership position and especially if you hold an executive position you are expected to:

- encourage and facilitate colleagues to develop as professionals.
- seek to ensure that no one is penalised for raising areas of concern or conflicts of interest.
- encourage colleagues to follow this code of conduct.

Definitions:

Legislation

The term "Legislation" means any applicable laws, statutes and regulations.

Third Parties

The term "Third Parties" includes any person or organisation that might be affected by your activities in your professional capacity, irrespective of whether they are directly aware or involved in those activities.

Relevant Authority

The term "Relevant Authority" in this document is used to identify the person(s) or organisation(s) which has/have authority over the activity of individuals in their professional capacity. For practising BCS members this is normally an employer or client. For student members, this is normally an academic institution.

Public Interest

- Acting in the public interest may be governed by Legislation.
- Legitimate rights of third parties include protecting personal identifiable data to prevent unlawful disclosure and identity theft, and also respect for copyright, patents and other intellectual property.
- Assess the requirement for, and where necessary, hold appropriate indemnity insurance.

Professional Competence and Integrity

- All members are required to undertake professional development activities as a condition of membership. Continuing professional development activities should broaden your knowledge of the IT profession and maintain your competence in your area of specialism.
- You should seek out and observe good practice exemplified by rules, standards, conventions or protocols that are relevant in your area of specialism.
- You should only claim current competence where you can demonstrate you have the required expertise e.g. through recognised competencies, qualifications or experience.
- Legislation that may apply in carrying out your professional responsibilities might include that applicable to:
  - your Relevant Authority.
  - the geographic area in which you are carrying out your professional

responsibilities.

- the geographic area in which your responsibilities will be discharged.

You may need to seek guidance from your Relevant Authority.

- Where you are leading a first of kind project, you will ensure that you make use of peer review and support where appropriate.

Duty to Relevant Authority

- Exercising of your professional judgement:
    - Where there is conflict between full and committed compliance with the Relevant Authority's instructions and the independent and considered exercise of your professional judgement, you will indicate the likely risks and consequences.
    - If any conflict is likely to occur or be seen by a third party as likely to occur you will make full and immediate disclosure to your Relevant Authority.
    - If for any reason you are unable to complete any assigned tasks in accordance with their requirements (e.g. on time or within budget) you will advise the Relevant Authority as soon as practicable.

Duty to the Profession

- As a member of BCS you have a responsibility to:
    - share knowledge and understanding of IT and support inclusion of every sector of society.
    - encourage and support fellow members in their professional development.
    - Support a colleague or any other person to whom they have a duty of care who in good faith raises any concern about a danger, risk, malpractice or wrongdoing which affects others (blows the whistle).

- In circumstances where a member is also a member of another professional body the clauses of any other applicable code of conduct cannot be employed to diminish or negate the clauses of the BCS Code of Conduct.
- You will not make any statement on behalf of BCS or purport to represent BCS through any public medium, including digital social media, unless authorised to do so by BCS.

8-2. Appendix 2: Topics – Tigray War Social Media Discourse

Tigray War Social Media Discourse's Topics					
No	Count	Name	No	Count	Name
0	4,195	0_ethiopia_channel_shopping_product	1	4.061	1_amhara_afar_land_region
2	1,108	2_tigray_god_prevail_bless	3	999	3_eritrea_war_tigray_forces
4	921	4_ahmed_prime_minister_abiy	5	885	5_tdf_viva_victory_fighting
6	823	6_war_end_loo_stop	7	724	7_tplf_justice_game_leadership
8	484	8_government_federal_federalism_constitution	9	476	9_zee_global_political_world
10	414	10_abiy_killer_go_king	11	400	11_tplf_surrender_peace_land
12	389	12_peace_pray_peaceful_hope	13	364	13_viva_tdf_tigray_tplf
14	364	14_military_defence_forces_army	15	331	15_junta_fascist_dictator_tplf
16	330	16_genocide_stop_massacre_tigray	17	311	17_defense_forces_town_near
18	226	18_afar_forces_tdf_special	19	177	19_prize_noble_winner_peace
20	166	20_propaganda_tplf_party_government	21	138	21_food_famine_hunger_feed
22	137	22_journalist_media_information_propaganda	23	133	23_song_music_love_listening
24	126	24_terrorist_group_terror_organization	25	108	25_regional_region_state_government
26	102	26_dam_water_renaissance_building	27	99	27_abiy_tplf_gala_flour
28	97	28_drone_air_strike_turkey	29	96	29_civil_war_brewing_stop
30	95	30_independent_independence_republic_country	31	95	31_voice_voiceless_thank_suffering
32	91	32_leader_hero_abiy_leadership	33	91	33_aid_humanitarian_guardian_access
34	87	34_turkey_china_standing_mediate	35	82	35_voice_talk_talking_joke
36	79	36_western_land_tigray_nelson	37	79	37_fake_news_breaking_reserved
38	77	38_music_drama_new_movie	39	73	39_video_war_footage_suggest
40	70	40_agreement_disarmament_agreed_south	41	68	41_news_reporter_reserved_television
42	65	42_forces_onslaught_force_control	43	65	43_ceasefire_unilateral_agreement_basic
44	63	44_brave_freedom_blessed_resilience	45	60	45_ethnic_federalism_livable_group
46	59	46_copyright_sport_disclaimer_news	47	58	47_sur_nous_premier_pour
48	57	48_withdraw_withdrawal_withdrawing_retreat	49	53	49_opinion_news_war_media
50	50	50_video_tdf_mixer_picture	51	49	51_trucks_missing_truck_aid
52	48	52_jeel_wali_tu_blah	53	48	53_russia_china_interfering_internal
54	45	54_forward_awareness_look_looking	55	44	55_horn_peace_gang_stuff
56	41	56_food_shortage_appetite_aid	57	40	57_forces_attack_shipment_magnetic
58	39	58_lie_truth_fool_lying	59	37	59_crime_criminal_rebellion_reasonable
60	36	60_video_media_house_orang	61	35	61_uniform_wear_wearing_clothes
62	35	62_pay_price_money_priceless	63	35	63_evil_thanks_dawn_thank
64	32	64_strong_long_stay_people	65	30	65_strictly_unauthorized_distribution_content
66	30	66_deg_video_ula_dag	67	30	67_ground_buried_grave_bury
68	29	68_court_aby_jail_criminal	69	27	69_bell_icon_informative_latest
70	26	70_born_father_name_fake	71	26	71_tdf_inhuman_peace_bolden
72	25	72_church_orthodox_irrespective_priest	73	25	73_amnesty_report_widespread_international
74	25	74_battle_field_victory_life	75	23	75_eu_aid_essentially_money
76	23	76_music_orthodox_media_film	77	23	77_endure_civilian_freedom_brave
78	22	78_song_gallant_saga_amazing	79	21	79_music_classical_copyright_collection

80	21	80_shire_defensive_town_front	81	20	81_warning_wary_warn_lees
82	19	82_federal_border_defense_forces	83	19	83_drama_story_mona_mascara
84	19	84_pay_price_heaver_lour	85	17	85_crying_cry_grief_cried
86	17	86_doctor_king_god_protect	87	17	87_militia_sanctity_infringe_mendacious
88	16	88_follow_accuracy_impartiality_finance	89	16	89_west_western_cyclical_middle
90	16	90_claim_evidence_inconsistency_gains	91	16	91_conflict_crisis_proportionally_sir
92	16	92_east_west_cat_peace	93	16	93_supporter_hear_stressful_bout
94	16	94_family_forever_love_hate	95	15	95_coward_brave_imagination_dedication
96	15	96_foreign_killing_ahmed_innocent	97	15	97_mouthpiece_mouth_chocolate_rude
98	14	98_chaw_gay_western_hange	99	14	99_black_white_simplistic_screen
100	14	100_northern_ambo_north_year	101	13	101_god_worship_ie_father
102	13	102_civil_gist_terrestrial_sky	103	13	103_price_pay_deserve_sell
104	13	104_cancer_burning_burn_figure	105	12	105_album_apple_singing_singer
106	12	106_million_siege_possible_artery	107	12	107_leadership_practiced_belligerently_galvanized
108	12	108_martin_ounce_corroborate_diehard	109	12	109_tiger_tigery_lion_mining
110	12	110_buried_cemetery_coffin_grave	111	12	111_mediation_negotiation_trump_scraped
112	12	112_lied_cleansing_enslave_fire	113	12	113_negotiate_negotiation_communicate_sequence
114	12	114_map_drown_misrepresent_regin	115	11	115_negotiation_gone_negotiate_narrower
116	11	116_upon_glance_greeting_praise	116	11	117_jeff_thank_truth_deliverer
118	11	118_link_section_member_given	119	10	119_rainy_western_westing_stall
120	10	120_city_appoint_mayor_tdf	121	10	121_sea_red_leader_job
122	10	122_new_amt_sew_min	123	10	123_justice_eternal_drumbeat_lasting
124	10	124_cynicism_creed_necessary_individually	125	9	125_million_population_stile_contradictory
126	9	126_election_board_electorial_strangely	127	9	127_voice_voiceless_mesh_fry
128	9	128_ambassador_tue_marshal_interview	129	9	129_price_anther_noble_peace
130	9	130_zone_altitude_kit_camel	131	9	131_die_bis_er_bin
132	9	132_trust_fitted_awkward_barred	133	9	133_rape_phenomena_allege_shedding
134	8	134_brother_sister_chill_thank	135	8	135_coup_insecurity_attempt_undersecretary
136	8	136_election_pompous_unmature_gusto	137	8	137_parliament_haughty_grandiose_unhelpful
138	8	138_false_formed_zip_information	139	8	139_arrest_pardoning_kidnap_administrator
140	8	140_battlefield_spinning_month_doubt	141	8	141_pilot_prisoner_second_bombardment
142	8	142_love_support_decent_precious	143	7	143_silly_ammunition_gun_weapon
144	7	144_trust_worthy_mast_ruthless	145	7	145_martin_reconciliation_sleepless_book
146	7	146_watching_video_reflect_stream	147	7	147_fuel_plowing_lend_transported
148	7	148_dialogue_national_tobe_presence	149	7	149_politics_video_subject_discover
150	7	150_contextual_speaking_bee_manipulation	151	7	151_choosing_produce_range_corner
152	7	152_minority_group_imaginary_unsuspecting	153	7	153_deg_prime_minister_rip
154	7	154_dance_circle_avid_embracement	155	7	155_massacre_executed_cold_replica
156	7	156_electricity_shutdown_transmission_electric	157	7	157_invite_participation_torrent_crazed
158	7	158_flag_color_blue_company	159	7	159_diamond_subscribe_comment_share
160	7	160_dress_wear_warmed_sweat	161	6	161_come_thank_good_brilliant
162	6	162_river_throw_body_transporting	163	6	163_combine_accumulate_accelerate_unjust
164	6	164_shire_freed_digit_correctly	165	6	165_loyalist_retake_retribution_otherwise
166	6	166_dad_sac_cade_ula	167	6	167_batch_optional_foundation_static

168	6	168_forever_independence_stay_made	169	6	169_investigation_objection_os_sue
170	6	170_soul_rest_legendary_may	171	6	171_canada_victory_unjust_association
172	6	172_wish_peacefully_critic_peaceful	173	5	173_advancing_leap_lean_towards
174	5	174_food_unpredicted_talking_desperately	175	5	175_economy_average_removal_deleterious
176	5	176_flint_disputable_report_simultaneously	177	5	177_migrate_home_pursuance_trace
178	5	178_informant_interviewer_interview_irksome	179	5	179_control_acknowledged_unfortunate_fail
180	5	180_chantry_musician_nostalgic_song	181	5	181_extend_seismological_unpunished_criminal
182	5	182_printing_bank_wholesale_froze	183	5	183_evidence_curious_store_honestly
184	5	184_water_added_drop_add	185	5	185_electric_service_trucks_main
186	5	186_blood_valueless_plead_shed	187	5	187_dog_pupate_lap_toothless
188	5	188_relevance_wat_power_mediate	-1	22,052	-1_tplf_ethiopia_people_tigray

8-3. Appendix 3: Word Frequencies – Tigray War Social Media Discourse

Tigray War Phase 1 Word Frequency								
No	Vocabulary	Ratio	No	Vocabulary	Ratio	No	Vocabulary	Ratio
1	Tigray	2.58%	11	Amhara	0.66%	21	Forces	0.45%
2	TPLF	2.58%	12	Country	0.63%	22	Must	0.45%
3	Ethiopia	2.51%	13	One	0.61%	23	Group	0.43%
4	People	2.17%	14	Military	0.60%	24	Power	0.41%
5	War	2.09%	15	Like	0.60%	25	Please	0.40%
6	Abiy	1.74%	16	News	0.58%	26	Stop	0.39%
7	Government	1.25%	17	Federal	0.50%	27	International	0.39%
8	Ahmed	0.92%	18	Us	0.50%	28	Civil	0.38%
9	Peace	0.84%	19	Eritrea	0.49%	29	Conflict	0.37%
10	Region	0.72%	20	Know	0.47%	30	World	0.37%

Tigray War Phase 2 Word Frequency								
No	Vocabulary	Ratio	No	Vocabulary	Ratio	No	Vocabulary	Ratio
1	Tigray	3.16%	11	One	0.69%	21	Region	0.44%
2	Ethiopia	2.57%	12	Ahmed	0.63%	22	New	0.42%
3	TPLF	2.38%	13	Forces	0.62%	23	Time	0.39%
4	People	1.65%	14	Like	0.61%	24	Never	0.39%
5	War	1.31%	15	Eritrea	0.60%	25	God	0.37%
6	Abiy	1.28%	16	Us	0.57%	26	Would	0.36%
7	Amhara	1.09%	17	Peace	0.50%	27	Also	0.35%
8	TDF	1.04%	18	Know	0.48%	28	Get	0.34%
9	Government	0.84%	19	Military	0.47%	29	Thank	0.34%
10	News	0.72%	20	Country	0.46%	30	Please	0.34%

Tigray War Phase 3 Word Frequency								
No	Vocabulary	Ratio	No	Vocabulary	Ratio	No	Vocabulary	Ratio
1	Ethiopia	3.33%	11	Eritrea	0.82%	21	Video	0.53%
2	Tigray	3.21%	12	Ahmed	0.78%	22	Us	0.47%
3	News	1.79%	13	Like	0.72%	23	Time	0.46%
4	Abiy	1.67%	14	Support	0.67%	24	Get	0.44%
5	Amhara	1.61%	15	Afar	0.66%	25	God	0.43%
6	People	1.20%	16	Channel	0.63%	26	Please	0.43%
7	TPLF	1.05%	17	One	0.62%	27	Forces	0.42%
8	New	1.02%	18	TDF	0.60%	28	Country	0.37%
9	War	0.93%	19	Music	0.60%	29	Region	0.36%
10	Government	0.87%	20	Today	0.59%	30	Peace	0.36%

Tigray War Phase 4 Word Frequency								
No	Vocabulary	Ratio	No	Vocabulary	Ratio	No	Vocabulary	Ratio
1	Tigray	3.33%	11	One	0.68%	21	Region	0.44%
2	Ethiopia	2.83%	12	TDF	0.64%	22	Media	0.43%
3	TPLF	1.68%	13	New	0.63%	23	Channel	0.40%
4	Abiy	1.40%	14	Like	0.62%	24	Also	0.39%
5	Amhara	1.26%	15	Ahmed	0.61%	25	Even	0.39%
6	People	1.24%	16	Music	0.55%	26	Country	0.38%
7	War	1.16%	17	Peace	0.52%	27	Never	0.35%
8	Eritrea	1.03%	18	Us	0.49%	28	Know	0.35%
9	News	1.01%	19	Support	0.49%	29	Forces	0.32%
10	Government	0.97%	20	Time	0.48%	30	Military	0.30%

Tigray War Phase 5 Word Frequency								
No	Vocabulary	Ratio	No	Vocabulary	Ratio	No	Vocabulary	Ratio
1	Tigray	3.78%	11	Peace	0.75%	21	Military	0.42%
2	TPLF	2.49%	12	Amhara	0.74%	22	Support	0.41%
3	Ethiopia	2.16%	13	One	0.67%	23	Channel	0.39%
4	War	1.71%	14	Like	0.62%	24	Fighting	0.38%
5	TDF	1.44%	15	Forces	0.59%	25	Would	0.37%
6	People	1.34%	16	Time	0.52%	26	Country	0.37%
7	Eritrea	1.10%	17	New	0.49%	27	Think	0.36%
8	Abiy	0.99%	18	Ahmed	0.48%	28	Never	0.35%
9	Government	0.91%	19	Us	0.47%	29	Want	0.35%
10	News	0.78%	20	Know	0.44%	30	Stop	0.33%

Tigray War Phase 6 Word Frequency								
No	Vocabulary	Ratio	No	Vocabulary	Ratio	No	Vocabulary	Ratio
1	Tigray	2.78%	11	TDF	0.67%	21	Ahmed	0.39%
2	TPLF	2.74%	12	Us	0.64%	22	Channel	0.38%
3	Ethiopia	2.24%	13	Amhara	0.64%	23	Two	0.37%
4	Peace	2.21%	14	Eritrea	0.64%	24	Know	0.37%
5	War	1.43%	15	Like	0.60%	25	God	0.35%
6	People	1.32%	16	Abiy	0.60%	26	Deal	0.35%
7	News	0.99%	17	Forces	0.59%	27	Military	0.35%
8	Government	0.84%	18	New	0.49%	28	Think	0.35%
9	Agreement	0.78%	19	Need	0.44%	29	South	0.33%
10	One	0.70%	20	Also	0.40%	30	Time	0.33%